Folding and Conformational Dynamics of the Hairpin Ribozyme and the Spliceosome: Combining Computational and Experimental Analyses

by

Mark A. Ditzler

A dissertation submitted in partial fulfillment
Of the requirements for the degree of
Doctor of Philosophy
(Biophysics)
in The University of Michigan
2009

Doctoral Committee:

Associate Professor Nils G. Walter, Chair
Professor Ari Gafni
Associate Professor Hashim M. Al-Hashimi
Associate Professor Mark A. Saper
Assistant Professor Kristina Hakansson

RNA

Everywhere

Acknowledgments

I thank my family, my advisor and fellow Walter lab members past and present. My parents have always provided strong and consistent support. I thank my wife Sarah and our two children Amelia and Evelyn. I also thank my father in-law, mother in-law and mother out-law for all of their support. In addition to his guidance and support I am especially thankful for the times Nils and I have been able to share pure RNA glee. I am grateful to David Rueda for his encouragement and guidance. I am indebted to the entire Walter lab for providing both a vibrant and a supportive intellectual environment. I thank in particular Sarah McDowell (Liz) for enduring many a long and strange conversation and steering me clear of complete RNA absurdity. I thank Mario Blanco and Matthew Marek for their assistance and in particular for exposing me to the full breadth of resources available through the University of Michigan library system. I am grateful to all of our collaborators and my committee members.

Table of Contents

List of Figures

Abstract

Folding and Conformational Dynamics of the Hairpin Ribozyme and the Spliceosome: Combining Computational and Experimental Analyses

by

Mark A. Ditzler

Chair: Nils G. Walter

The vital role of RNA structure and dynamics in determining biological function is increasingly appreciated throughout the life sciences. RNA-coding genes are now recognized to be far more abundant in eukaryotes than their protein-coding counterparts and are essential to the central biochemical processes within all living cells. Here, we use computational and experimental techniques in order to understand the folding and conformational dynamics of two vastly different RNA systems (the hairpin ribozyme and the spliceosome) at the single molecule level.

Large energy barriers separating misfolded and functional states are a well appreciated characteristic of RNA. By contrast, it is typically assumed that functionally folded RNA occupies a single native basin of attraction free of deeply dividing energy barriers. Here, we develop an experimental approach to isolate persistent sub-populations of a small RNA enzyme and show by single molecule fluorescence resonance energy transfer (smFRET), biochemical probing, and high-resolution mass spectrometry that commitment to one of several catalytically active folds occurs unexpectedly high on the folding energy landscape.

Despite numerous investigations, the catalytic mechanism of hairpin ribozyme self-cleavage remains elusive. To gain insight into the coupling of active site dynamics with

activity of this small catalytic RNA, we analyzed multiple molecular dynamics (MD) simulations. Our simulations suggest an important role for protonation of A38 in promoting a favorable geometry similar to that observed in transition-state analog crystal structures, and support previously proposed roles of A38, G8, and water in catalysis. Finally we discuss a plausible mechanism in which A38 acts bifunctionally and shuttles a proton directly from the 2'-OH to the 5'-oxygen.

Despite over 20 years of study, the kinetic and structural details of spliceosome assembly are not well understood. To track in real-time the conformational states through which the spliceosome takes a pre-mRNA, we have developed an smFRET based spliceosome assembly and activity assay. Our data quantify the kinetics of ATP, small nuclear RNA, and sequence dependent pre-mRNA conformational changes. We find that wild-type pre-mRNA inhabits a broad and dynamic conformational space and is specifically funneled into a constrained conformational sub-space by ATP-dependent, spliceosome induced processes.

Chapter 1:

Exploring an RNA World[a]

1.1 Introduction

We are living in an RNA world! In 1968 Leslie Orgel and Francis Crick published two speculative papers about the role of RNA in the origin of life and the genetic code (1,2). They recognized that the ability of RNA to adopt varied secondary structures could potentially endow RNA with the ability to catalyze chemical reactions, though catalytic RNA was unknown at the time and transfer RNA (the existence of which was actually predicted by Crick) was one of the few examples of structured RNA and even its three-dimensional structure would not be known until 1974 (Figure 1.1). This speculation opened the possibility that RNA may have served as the primary biological catalyst before the emergence of genomically encoded proteins and as the molecule of heredity prior to the existence of DNA. It would be another 14 years before the discovery of the catalytic capabilities of group I introns (3) and RNase P (4), coupled with the knowledge that certain viral genomes are composed entirely of RNA, would establish RNA as unique in nature for its ability to both store genetic information and catalyze chemical reactions. The dual genetic and catalytic roles of RNA lend tremendous support to what became known as the "RNA world hypothesis", which posits that purely RNA based life predated the emergence of both protein and DNA (1,2,5). In the decades that followed the importance of RNA in a myriad of modern biological processes has become increasingly clear, revealing a picture of life that could justifiably excuse an unabashedly ribo-centric approach to biology.

---

[a] Adapted from Ditzler MA, Alemán EA, Rueda D, Walter NG. Focus on function: single molecule RNA enzymology. *Biopolymers* **87** (2007) 302-316, and from Ditzler MA, Otyepka M, Sponer J, Walter NG. Molecular Dynamics and Quantum Mechanics of RNA: Conformational and Chemical Change We Can Believe In. *Accounts of Chemical Research. (in preparation)*

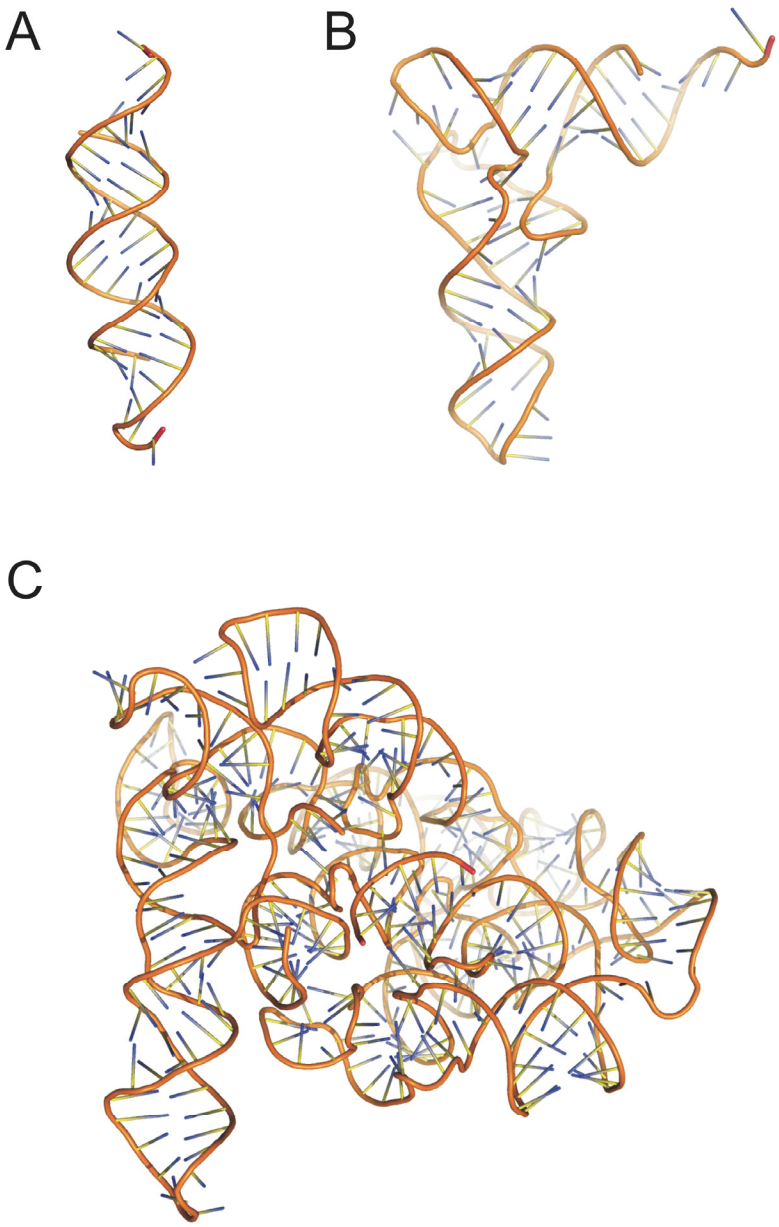


Figure 1.1: Increasingly complex RNA structures help to explain its biological function. (A) Complementary RNA sequences adopt an A-form double helix with Watson-Crick base pairing which allows RNA to serve as a molecule of heredity. A-form double helix of an siRNA is shown (6) (PDB ID 2F8S) (B) transfer RNA's structure is essential to its role in translation, provided the first example of a complex RNA structure, and inspired the speculation that RNA may be catalytic. Phenylalanine t-RNA is shown (7)(PDB ID 1EVV). (C) The recently derived structure of the group II self-splicing intron helps to explain how this RNA is able to precisely align functional groups to catalyze the removal of its own intron, and may provide insight into how the spliceosome catalyzes removal of the majority of introns which do not self-splice (8) (PDB ID 3EOG).

RNA-coding genes are now recognized to be far more abundant in eukaryotes than their protein-coding counterparts and are essential to the central biochemical processes within all living cells (9-11). RNA is responsible for the synthesis of all proteins within the cell, plays a central role in replication of many viruses, regulates gene expression in both bacteria and eukaryotes, is involved in the maintenance, processing, modification and editing of genetic information, and probably carries out a host of still unknown cellular processes. Many of these known biological roles of RNA involve catalytic RNAs (ribozymes) and, in addition to their functions in nature, catalytic RNAs have been used to derive RNA based therapeutics (12,13). Our understanding of the molecular underpinnings of organisms, and possibly the origin of life, as well as the development of new medicines therefore significantly depend on our ability to dissect the fundamental properties of non-coding RNA.

Ribozymes

Naturally occurring ribozymes can be divided into several groups based on their size: small self-cleaving RNAs (<200 nucleotides), medium-sized self-splicing introns, and larger catalytic ribonuclear-protein (RNP) complexes. The class of small ribozymes comprises the hairpin, hammerhead, hepatitis delta virus (HDV), Varkud satellite (VS), and *glmS* ribozymes. All of these ribozymes catalyze a site-specific RNA backbone cleavage reaction as well as the reverse ligation reaction. Cleavage is achieved through an S_N2-like reaction mechanism in which the 2'-hydroxyl (2'-OH) of the cleaved strand acts as the nucleophile, resulting in 2',3'-cyclic-phosphate and 5'-OH termini on the 5'– and 3'-products, respectively (Figure 1.2A) (14-16). On the other end of the spectrum, large RNPs such as RNase P, the spliceosome and the ribosome represent catalytic RNAs that recruit protein co-factors for optimal function *in vivo* (self-splicing introns are of intermediate complexity as some of them require protein co-factors, others do not). RNase P and the spliceosome carry out site-specific hydrolysis and transesterification reactions on RNA backbones, respectively, through mechanisms distinct from that of the small ribozymes. The ribosome is unique among the naturally occurring ribozymes in that it generates a product that is not itself an RNA. The ribosome catalyzes peptide bond formation between amino acids coupled to tRNA adapters and so is responsible for the

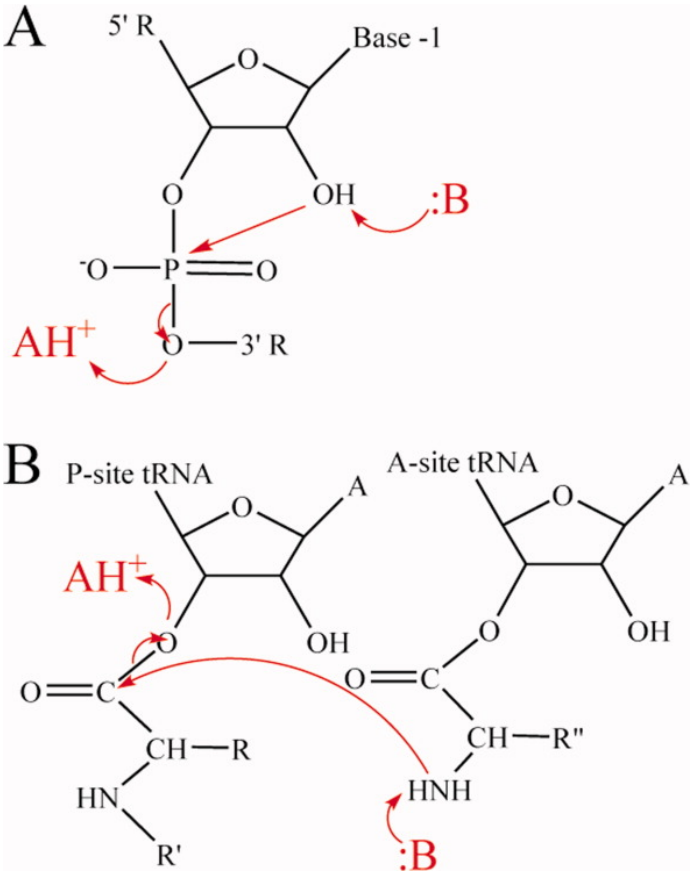


Figure 1.2: Proposed reaction mechanisms for two ribozymes of vastly different size and structure. (A) Site-specific phosphodiester transfer as catalyzed by the self-cleaving small ribozymes, including the hairpin ribozyme. A suitably positioned base B deprotonates the 2'-OH of the upstream ribose, thereby activating the 2'-oxygen for nucleophilic in line attack on the scissile phosphodiester. The 5'-oxygen leaving group is protonated by a properly positioned acid AH^+ (see Chapter 3). (B) Peptide bond formation as catalyzed by the ribosome. A suitably positioned base deprotonates the amino acid esterified with the A-site tRNA, thereby activating the amino group for nucleophilic attack on the peptidyl-tRNA ester bond on the P-site tRNA. The 3'-oxygen leaving group is protonated by a properly positioned acid AH^+.

production of all cellular protein (Figure 1.2B). Evidence that the RNA rather than protein components of RNPs are catalytic stems from the observation of catalytic competence in the absence of protein and/or an active site composed of RNA only (3,4,17,18).

The Structure-Dynamics-Function Relationship

Accompanying and largely precipitating the increasingly appreciated role of RNA structure is the rapidly growing number of available RNA and RNP high-resolution structures. These atomic resolution snap shots can provide detailed rationalization for existing biochemical data, however biological function depends of the dynamic evolution of structures along relevant functional pathways. A complete understanding of the relevant structural dynamics exhibited by RNA covers time scales from picoseconds to minutes (19). Meaningful assessment of dynamics over such wide-ranging time scales requires a correspondingly broad range of techniques (Figure 1.3), with careful consideration given to the scope and limitation of each method. Two important methods available to understanding RNA dynamics are the focus of this thesis, single-molecule FRET and molecular dynamics simulations.

1.2 Single molecule FRET of Catalytic RNA and RNPs: Focus on Function

Since their discovery a quarter-century ago, extensive investigations into the catalytic mechanisms of ribozymes have been conducted in the quest to understand and potentially exploit this essential and ubiquitous class of enzymes. Until recently, catalytic RNAs were studied in bulk solution where the number of molecules present is many orders of magnitude larger than the low copy number typical of many RNAs and RNPs in a single cell (1-10^3, up to 10^6 in case of the ribosome). Recently, it has become increasingly common to study protein and RNA enzymes using single molecule methods, offering the ability to observe short-lived mechanistic intermediates and minor sub-populations often masked in the ensemble average. Single molecule approaches to understanding RNA include atomic force microscopy, optical tweezers, and single molecule fluorescence microscopy (for review see (20,21)). Of these, single molecule fluorescence resonance

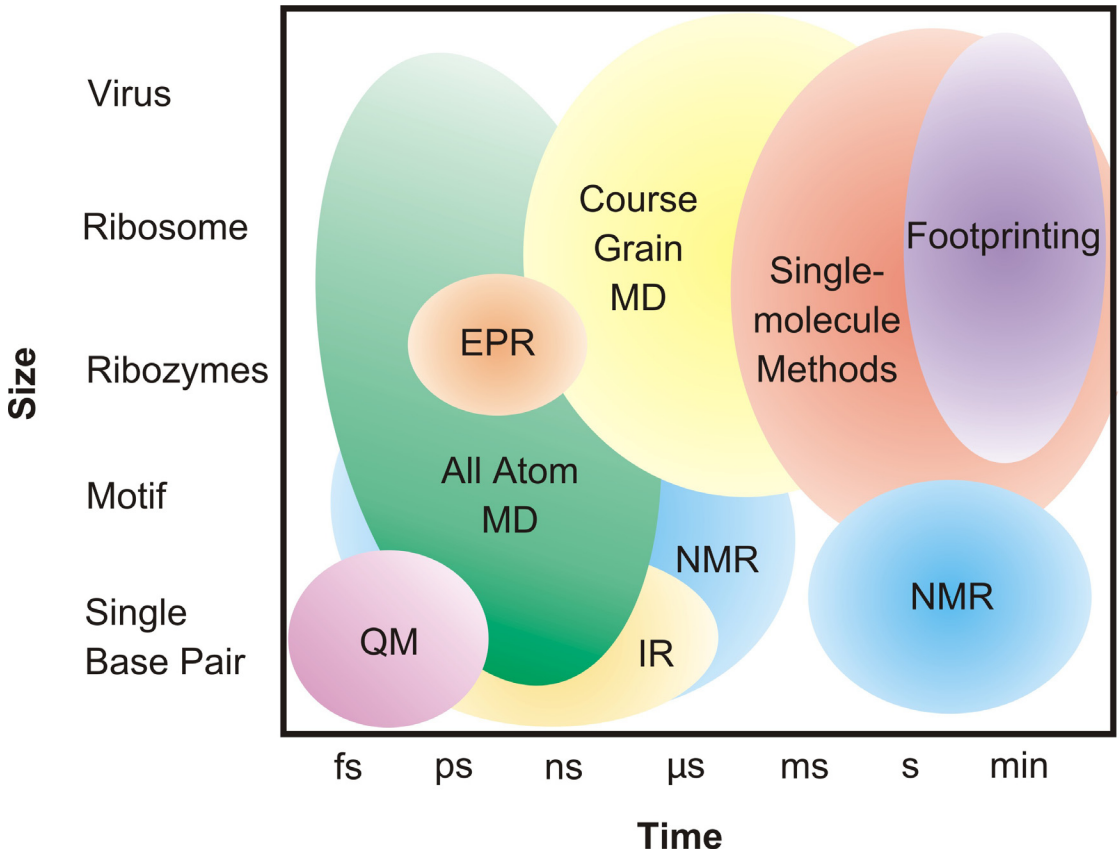


Figure 1.3: Evaluating RNA dynamics. Schematic of techniques commonly used to evaluate RNA dynamics and the system size and time scales they are used to address.

energy transfer (smFRET) has proven particularly effective in studying reaction pathways of ribozymes as smFRET assays provide information on the global dynamics of molecules under native conditions. smFRET has therefore provided researchers with the unique opportunity to quantify the (equilibrium) kinetics of both directions in reversible reactions, which are commonly found in RNA.

Examples of Single Molecule Enzymology

Currently, the primary approach used in single molecule RNA enzymology is to monitor global conformational changes associated with individual steps on or off a reaction pathway such as substrate binding, tertiary structure (un)folding, chemical catalysis, and product release. In this chapter we explore in detail single molecule investigations that highlight the scope and limitations of single molecule RNA enzymology. We focus on two significant RNA catalysts at opposite ends of the spectrum, the hairpin ribozyme and the ribosome. The hairpin ribozyme is probably the most investigated RNA in single molecule enzymology and the primary focus of this thesis (see Chapters 2 and 3). The ribosome is far more complex and has been subjected to fewer single molecule studies than the comparably simple hairpin ribozyme. However, the tremendous biological importance of protein biosynthesis has motivated substantial progress also on single molecule enzymology of the ribosome. The initial success of work with the ribosome provides great encouragement in the single molecule investigation of other complex RNPs such as the spliceosome (see Chapter 4).

The Hairpin Ribozyme: Synergy between Single Molecule and Ensemble Assays

The hairpin ribozyme (Figure 1.4A) is a small non-coding RNA that facilitates site-specific cleavage and ligation chemistry of its own backbone as part of the double-rolling circle replication of Nepovirus satellite RNAs. It serves as a convenient model system to study RNA catalysis, and a vast body of ensemble biochemical(22-28), structural(29-35), and computational data (36,37) is available, as are extensive single molecule analyses(38-44). The drive towards a complete understanding of catalysis in this system has demonstrated and exploited the power of single molecule spectroscopy to uncover short-lived intermediates, minor sub-populations, and molecular heterogeneity, which

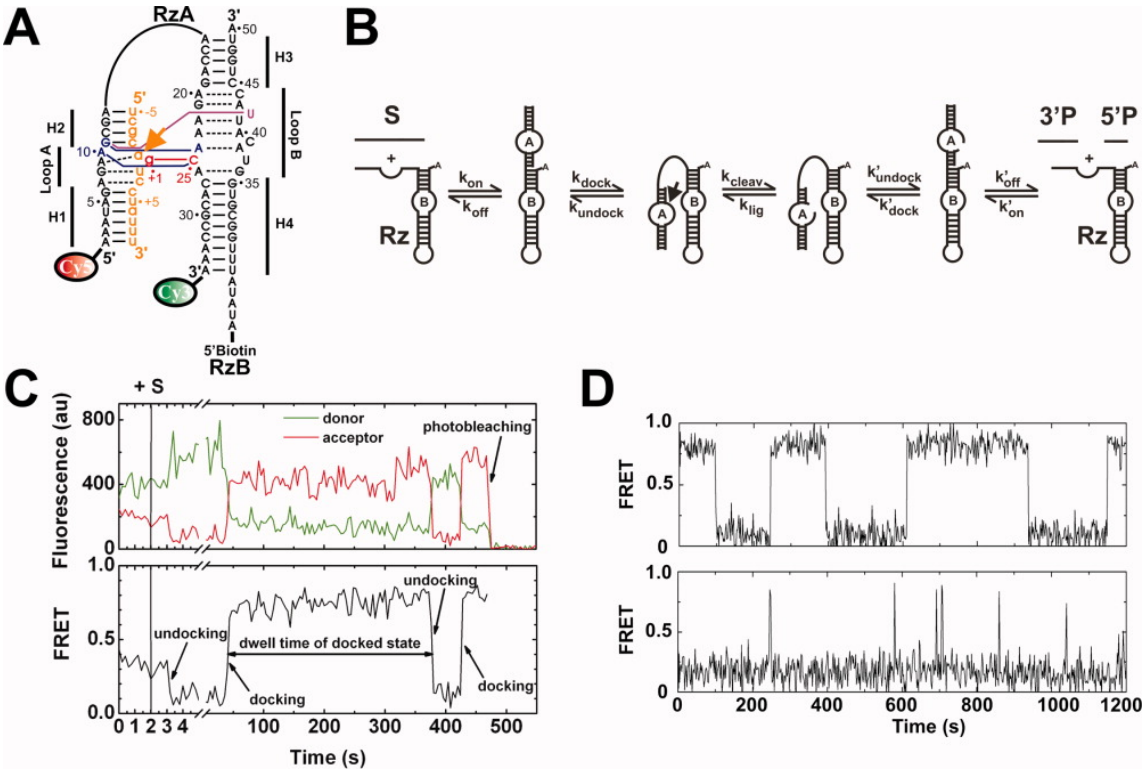


Figure 1.4: Single molecule FRET applied to hairpin ribozyme docking. (A) A two-stranded (RzA, RzB) hairpin ribozyme binds substrate (orange and small letters; arrow, cleavage site) to form internal loops A and B, each flanked by two helices (H1-H4) and connected between H2 and H3. Noncanonical base pairs are indicated by dashed lines. Tertiary structure docking occurs via a g+1:C25 Watson-Crick base pair (red), a ribose zipper (blue) and the U42 binding pocket (purple). Terminal Cy3 and Cy5 fluorophores serve as donor/acceptor FRET pair and biotin is used for surface immobilization through binding to streptavidin. (B) Multistep reaction pathway of the hairpin ribozyme with distinct kinetic steps identified by their rate constants. (C) Typical smFRET time trajectory monitoring donor and acceptor emission intensity, together with the resulting FRET = $I_A/(I_D+I_A)$ trace. Characteristic of a single molecule observation are the anti-correlated donor and acceptor signals and the single-step photobleaching; specific events are indicated. Rate constants are calculated from statistically significant numbers of state dwell times and corrected as described.(38,44) (D) Two FRET time trajectories from different molecules show dramatically different dwell times in the high-FRET docked state that reveal persistent heterogeneity between molecular sub-populations. Reproduced with permission from ref. (38).

otherwise are all hidden in the ensemble average. An effective application of single molecule techniques, conversely, requires correlation of statistically significant averages from stochastic single molecule events with observables from ensemble measurements. In fact, most successful approaches have relied on the availability of a thorough characterization of ensemble behavior in order to interpret single molecule observations with confidence.

In the case of the hairpin ribozyme, extensive insights from ensemble techniques into the ribozyme's structural and kinetic properties have formed a solid platform for probing at the single molecule level. For example, ensemble FRET experiments in solution revealed the existence of two structural states at equilibrium – the catalytically inactive undocked and the active docked conformations(22). Upon docking, the internal loops of domains A and B are brought into close contact, compacting the RNA (Figure 1.4A) (45,46). Several crystallographic studies have shown that this docked state is stabilized by a number of well characterized tertiary hydrogen bond and base stacking interactions (Figure 1.4A) (32-35). In addition, ensemble enzymology approaches were applied extensively, yet the (presumably micro-reversible) mechanism of cleavage and ligation remains debated (47). Nucleobase derived general acid-base catalysis (23,32), water assisted acid-base catalysis (35,36), and transition state charge stabilization (24-26,33,48) have all been invoked as possible mechanisms (see Chapter 3). The important contributions that a single nucleobase or even a functional group can make to proper RNA folding as well as catalysis (44) and the inherent ambiguity in the interpretation of enzymologic results (47) contribute to the difficulty of pinpointing the reaction mechanism and necessitate additional mechanistic probing tools.

smFRET based on biotin-streptavidin mediated surface immobilization and total internal reflection fluorescence microscopy (TIRFM) has been employed to dissect the reaction pathway of the hairpin ribozyme, which comprises substrate binding, interdomain docking, substrate cleavage, interdomain undocking, and finally product release (Figure 1.4B). By labeling the termini of the two interacting domains with a suitable smFRET donor/acceptor pair such as cyanine dyes Cy3/Cy5,(49) the docked, undocked and product/substrate-free (unbound) states of the ribozyme display distinguishable FRET levels (defined as $I_A/(I_D + I_A)$, where I_A and I_D are the fluorescence

signals from acceptor and donor, respectively) (Figures 1.4 and 1.5A) (38). Single-step photobleaching to background signal at the end of each smFRET time trajectory confirms that indeed a single RNA molecule is observed (Figure 1.4C). Since the cleavage products rapidly dissociate from the undocked state, cleavage and subsequent undocking result in a decrease in smFRET from the docked to the unbound state (Figure 1.5A). The good agreement between the rate of unbound state appearance in smFRET and that of product appearance as monitored by traditional (ensemble) electrophoretic separation further supports the assignment of states in and the functional validity of single molecule trajectories (Figure 1.5A) (38).

To determine (un)docking rate constants in the intact ribozyme-substrate complex unaffected by cleavage, a blocking 2'-O-methyl substitution was introduced into the active site adenosine (A-1) based on its preservation of the sugar pucker preference (in solution) and hydrogen bond acceptor capacity of the native 2'-OH (38). Substrate dissociation is slow under standard conditions (pH 7.5, 12 mM Mg^{2+}), effectively isolating the docking/undocking steps from the remaining reaction pathway. Similarly, (un)docking rate constants in the isolated ribozyme-product complex can be determined by installing a ligation blocking 3'-phosphate on the 5'-product instead of the natural 2',3'-cyclic phosphate. Rate constants can then be extracted by plotting the cumulative number, $N(t)$, of state dwell (residence) times that are shorter than time t and fitting with a multi-exponential of the form:

$$N(t) = \sum_i A_i \left(1 - e^{k_{i,obs} \cdot t}\right) \quad (1)$$

The observed rate constants $k_{i,obs}$ need to be corrected for photobleaching, which shortens the observed dwell times; while the amplitudes A_i need to be corrected if indeed multiple rate constants are observed, to avoid bias towards shorter dwell times (38,44). Corrected dwell times in the docked state then determine the undocking rate constant(s) k_{undock}, whereas corrected dwell times in the undocked state separately determine the docking rate constant(s) k_{dock} (Figure 1.4B). This ability to determine forward and reverse rate constants of a reversible reaction independently of each other is an important advantage of single molecule enzymology. In the ensemble average, only collective and synchronized relaxation of many molecules from one state to another can be observed,

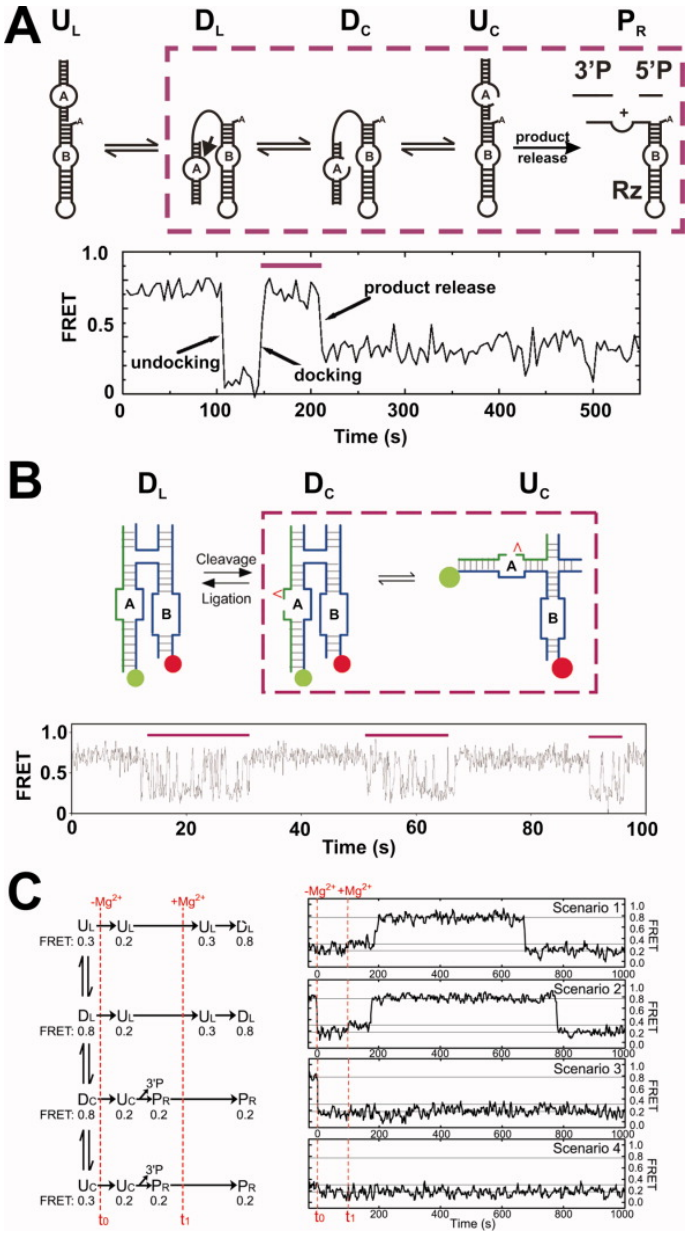


Figure 1.5: Accessing reaction chemistry of the hairpin ribozyme through single molecule FRET. (A) smFRET time trace of the two-way junction form of the hairpin ribozyme, showing the docked, undocked, and substrate/product free states at distinct FRET values.(38) The purple box and bar indicate equivalent processes on the reaction scheme and the experimental data, respectively. Individual states are indicated as U (undocked), D (docked), L (ligated), C (cleaved) and P (product). (B) smFRET time trace of the four-way junction form of the hairpin ribozyme, illustrating the difference in (un)docking dynamics before and after cleavage (i.e., in the ligated and cleaved forms). The purple box and bar highlight equivalent processes on the reaction scheme and in the experimental data, respectively.(41) (C) Schematic of the possible outcome scenarios from a double buffer-exchange experiment (first removal, then replenishment of 12 mM Mg^{2+}) on the two-way junction form of the hairpin ribozyme with the associated experimental smFRET readouts.(51) Reproduced with permission from Refs.(38,41,51)

yielding an aggregate rate constant. The fact that single molecule data often are best represented through multiple rate constants $k_{i,obs}$ is another distinction from ensemble averaging techniques. In the hairpin ribozyme (and commonly in RNA) an underlying molecular heterogeneity is apparent upon closer inspection (Figure 1.4D) (38,39,42). At least four distinct sub-populations of molecules are found in individual smFRET time trajectories (38,39,44), each of which undocks with one of the four rate constants extracted from the docking dwell times. Representatives of each sub-population are remarkably resistant to interconversion as they continue to exhibit the same undocking behavior even when probed at 3-hour intervals at 25˚C (38), in various Mg^{2+} concentrations (39), in the presence of various RNA modifications (44) or following subjection to harshly denaturing conditions (see Chapter 2). Such static heterogeneity (or molecular memory) is also observed in the context of a four-way junction form of the ribozyme, with potentially even more sub-populations exhibiting heterogeneity in both docking and undocking kinetics (40), and is independent of the strategy for RNA surface immobilization (42). The four distinctly undocking molecular sub-populations map onto fast and slow phases of biphasic ensemble cleavage (44) and folding assays (27), and so have important consequences for the interpretation of data from ensemble measurements (see Chapter 2).

The Hairpin Ribozyme: Accessing Chemistry

Whereas docking and undocking rate constants of the hairpin ribozyme can be derived directly from dwell time analyses of smFRET trajectories, catalysis itself does not result in any discernable change in smFRET signal and thus calls for less direct inference. Three different approaches have been pursued so far. In the first approach, single molecule probing of inactivated ribozyme-substrate and -product complexes is combined with ensemble cleavage assays and classic mechanistic modeling (38,44). Essentially, the intrinsic cleavage and ligation rate constants are derived by finding either a numerical (38) or analytical solution (44) to the set of differential equations that defines the reaction pathway in Figure 1.4B after substrate binding (where substrate binding is assumed to be irreversible):

$$
\begin{cases}
\dfrac{dN^{S}_{undock}}{dt} = -k_{dock}N^{S}_{undock} + k_{undock}N^{S}_{dock} \\
\dfrac{dN^{S}_{dock}}{dt} = k_{dock}N^{S}_{undock} - \left(k_{undock} + k_{cleav}\right)N^{S}_{dock} + k_{lig}N^{P}_{dock} \\
\dfrac{dN^{P}_{dock}}{dt} = k_{cleav}N^{S}_{dock} - \left(k_{lig} + k^{P}_{undock}\right)N^{P}_{dock} + k^{P}_{dock}N^{P}_{undock} \\
\dfrac{dN^{P}_{undock}}{dt} = k^{P}_{undock}N^{P}_{dock} - \left(k^{P}_{dock} + k_{off}\right)N^{P}_{undock} + k_{on}N_{diss} \\
\dfrac{dN_{diss}}{dt} = k_{off}N^{P}_{undock} - k_{on}N_{diss}
\end{cases}
\tag{2}
$$

where N denotes the population of molecules in a given state with the subscript indicating the conformation (docked or undocked) and the superscript indicating whether the ribozyme is bound to substrate (S) or product (P). Using the known overall cleavage kinetics and assuming the complete absence of interconversion between the molecular sub-populations throughout the catalytic cycle (which makes them independent of one another), a numerical fit yields upper and lower bounds for the intrinsic cleavage and ligation rate constants and an estimate of their ratio (38). Alternatively, equations 2 can be formulated in matrix notation so that an analytical simulation of the overall cleavage time course is derived by diagonalizing and solving the corresponding master equation (44). If the chemical equilibrium constant is independently determined by, for example, running a ligation reaction to completion in the presence of excess reaction product, the problem of solving the master equation for the five unimolecular reactions that describe the kinetic pathway is reduced to a single-variable fit (44). The use of substrate and product analogs and matrix-algebra assisted kinetic simulations thus enables rapid relative comparison of single functional group variants. It was discovered that functional groups far from the docking interactions and active site directly impact both docking and chemistry. These findings led to the proposal that, similar to protein enzymes (50), long range coupled molecular motions exist in ribozymes that link the overall fold to the active site and contribute to RNA function(44). Recent molecular dynamics simulations support this hypothesis (36).

It is important to note that two additional assumptions are implicitly employed in

this analysis. First, common intrinsic cleavage and ligation rate constants are assumed for all molecular sub-populations so that the derived rate constants represent averages over all molecules (a standard feature also of ensemble enzymology). A range of rate constants may exist, but kinetic modeling suggests that the chemical rate constants vary by less than three-fold between the different sub-populations (44). Second, in the analysis it is assumed that the docking and undocking rate constants obtained for the inactivated ribozyme complexes (with 2'-O-methyl modified substrate and 3'-phosphate modified 5'-product analogs) closely resemble those of the active complexes. Recent evidence from single molecule studies in the presence of cleavable substrate suggests that undocking of the substrate complex is decelerated and that of the product complex accelerated in the presence of the native 2'-OH and 2',3'-cyclic phosphate, respectively, leading to a systematic over-estimation of the intrinsic cleavage rate constant in the above analysis by ~7-fold (while docking and ligation rate constants are unaffected; see also the following discussion) (41,51). The increase in docking stability in the presence of the native 2'-OH in the ligated form is further supported by our recent MD simulations (see following section and Chapter 3).

The second single molecule approach used to access chemical rate constants of the hairpin ribozyme exploits the fact that undocking is slow in the native substrate complex, but fast in the cleaved product complex; thus, cleavage *in situ* can be fortuitously detected by an acceleration of the smFRET fluctuations between docked and undocked conformations (rather than a change in FRET level) (41). This change in dynamics is particularly pronounced at 1 mM Mg^{2+} in a four-way junction form of the ribozyme in which both the 5'- and 3'-products are extended to prevent dissociation. Significant enhancement and suppression of the docking/undocking kinetics can thus be used as signatures for cleavage and ligation, respectively (Figure 1.5B). Still, extraction of a cleavage rate constant is complicated since, first, any given transition from docked to undocked state may originate from either the substrate or product complex; second, cleavage events followed by ligation before undocking will go undetected; and, third, the observed docked state dwell times are shortened by photobleaching. Using a succession of undocking events as indication of cleavage and correcting for missed events as well as photobleaching then yields an estimate of the intrinsic cleavage rate constant. In

addition, the ligation rate constant has to be corrected for the fact that a rapidly docking/undocking ribozyme spends only part of its time in the docked state where ligation can occur. The final corrected intrinsic chemistry rate constants indicate a stronger equilibrium bias towards the ligated state than did the use of chemistry blocking modifications (41).

The third approach exploited to tease out the intrinsic chemistry rate constants also uses cleavable substrate and sets up a succession of buffer exchanges to produce distinct time sequences of smFRET signal that serve as kinetic "fingerprints" of specific catalytic intermediates (51). In concept such an approach is analogous to pulse-chase experiments widely used in ensemble enzymology, but it gains from the ability to assign a specific state to each individual molecule and count the number of representatives. Figure 1.5C illustrates how the number of molecules in the undocked (U) and docked (D) states in the presence of either ligated (L) or cleaved (C) substrate is assessed. First, chemical equilibrium is reached upon incubation of the ribozyme in standard buffer (pH 7.5, 12 mM Mg^{2+}) in the presence of a saturating excess of 3'-product. Upon addition of EDTA to remove Mg^{2+} at time t_0 and subsequent replenishment of Mg^{2+} at time t_1, distinct scenarios are observed depending on which of the four reaction intermediates U_L, D_L, D_C or U_C is observed. In particular, the two docked states undock (transit from high [0.8] to low [0.2] FRET) upon Mg^{2+} removal, while the two undocked states only slightly decrease in FRET (from 0.3 to 0.2). In addition, the two complexes involving ligated substrate will eventually dock again after the replenishment of Mg^{2+}, whereas the two cleaved complexes lose their 3'-product under these conditions (the 5'-product is covalently linked to the ribozyme) and thus can never dock after Mg^{2+} addition. Given a sufficient observation window after time t_1 (i.e., slow photobleaching) the four reaction intermediates can be unequivocally identified through their unique FRET versus time patterns (Figure 1.5C) and counted. Based on a sufficient number of molecule assignments P, yielding $P(U_L) = 21$, $P(D_L) = 591$, $P(D_C) = 47$, and $P(U_C) = 165$, the equilibrium constants of docking before and after cleavage and of internal chemistry are derived as ratios of the appropriate molecule counts (51). In conjunction with the rate constant of the very last transition from high to low FRET under standard conditions (Figure 1.5A), which is a convolution of the undocking, cleavage and ligation rate

constants, the intrinsic cleavage and ligation rate constants can be calculated.

The latter two single molecule studies yield similar intrinsic cleavage and ligation rate constants and a consistent picture of how a ribozyme is optimized for its self-cleavage function in Nepovirus replication – stable docking of the ligated ribozyme-substrate complex allows for ample time to cleave, while instable docking of the cleaved ribozyme-product complex results in rapid product release (41,51). One may note that a structural explanation for this kinetic phenomenon is still outstanding. In addition, since a properly ligated RNA is important as a replication template, one may wonder whether alternating structures switch the RNA between preferably cleaved (active) and ligated (inactive) forms. Strikingly, standard ensemble measurements of the chemistry equilibrium position do not distinguish between the docked and the undocked ligated or cleaved states and thus lead to a significant under-estimation of the ligation equilibrium constant (from $[P(U_L) + P(D_L)]/[P(D_C) + P(U_C)] \approx 2.9$) (44) compared to its true value (which is defined as $P(D_L)/P(D_C) \approx 13$) (51), and a resulting over-estimation of the intrinsic cleavage rate constant. Ensemble studies also average out important information on parallel (heterogeneous) reaction pathways, which are studied in isolation when observing single molecules, and on short-lived intermediates, which are identified by short smFRET bursts as long as they live longer than the experimental time resolution (Figure 1.4D).

In summary, single molecule enzymology studies of the hairpin ribozyme have demonstrated feasible routes towards determining rate and equilibrium constants of the chemical step in a fully reversible RNA reaction pathway that exhibits molecular heterogeneity. It thus has become possible to dissect the often surprisingly profound role(s) of individual residues and functional groups in structural dynamics and chemistry, may they be close to or far from the active site and/or tertiary structure interactions. Careful consideration needs to be given to the various types of modifications (fluorophore labeling, surface immobilization, functional group and sequence changes) that have to be introduced into the RNA to address specific scientific questions.

Investigating Complex RNA-Protein Machinery

The largest ribozyme studied so far at the single molecule level is the protein biosynthetic machinery, the ribosome, arguably the most abundant enzyme on earth. Ribosomes are very large (in bacteria ~2.5 MDa) RNA-protein complexes that universally translate the sequence of a messenger RNA (mRNA) with high fidelity into a polypeptide chain using transfer RNA (tRNA) adaptors. Ribosomes are composed of a large and a small subunit (termed 50S and 30S in bacteria, respectively). Translation is initiated by the assembly of the two subunits into the 70S ribosome on an mRNA template. Protein synthesis is catalyzed by the ribosomal RNA (rRNA) component of the large subunit by transfer of the growing peptide chain from one tRNA onto the next aminoacyl-tRNA (aa-tRNA) (52), whose selection is determined by the small subunit. A sequence of three nucleotides in the tRNA, called the anticodon, base pairs with each mRNA codon; this short hybrid is proofread with high fidelity by the rRNA component of the small subunit (53). Both subunits of the ribosome contribute to its three tRNA binding sites: the aminoacyl site (A site), the peptidyl site (P site) and the exit site (E site) (54).

The highly dynamic peptide elongation cycle is composed of at least the following steps (Figure 1.6A) (53,55-58): 1) Initial binding of the aa-tRNA to the ribosome in a ternary complex with the elongation factor Tu (EF-Tu) and guanosine 5'-triphosphate (GTP). 2) Codon recognition where base pairs form between the tRNA anticodon and mRNA codon in the A site of the 30S subunit. 3) Stimulation of the GTPase activity of EF-Tu in response to correct codon-anticodon pairing with the cognate aa-tRNA. At this stage, mismatched (non-cognate) aa-tRNAs are readily ejected from the ribosome, whereas near-cognate aa-tRNAs (only one base mismatch) remain bound (53,59). 4) GTP hydrolysis by EF-Tu. 5) Conformational change of EF-Tu coupled with dissociation of inorganic phosphate (P_i). 6) Accommodation of the A-site aa-tRNA within the peptidyl transferase center (PTC) of the large subunit, during which EF-Tu:GDP dissociates. At this point, most near-cognate aa-tRNAs are rejected in a process known as proofreading (53,60). 7) Peptidyl transfer, that is, formation of the peptide bond in the PTC, during which the polypeptide chain is rapidly transferred from the P-site to the A-site tRNA (see also Figure 1.2B). 8) The tRNAs proceed to two discernable hybrid states, where they remain bound to the mRNA in the A and P sites of

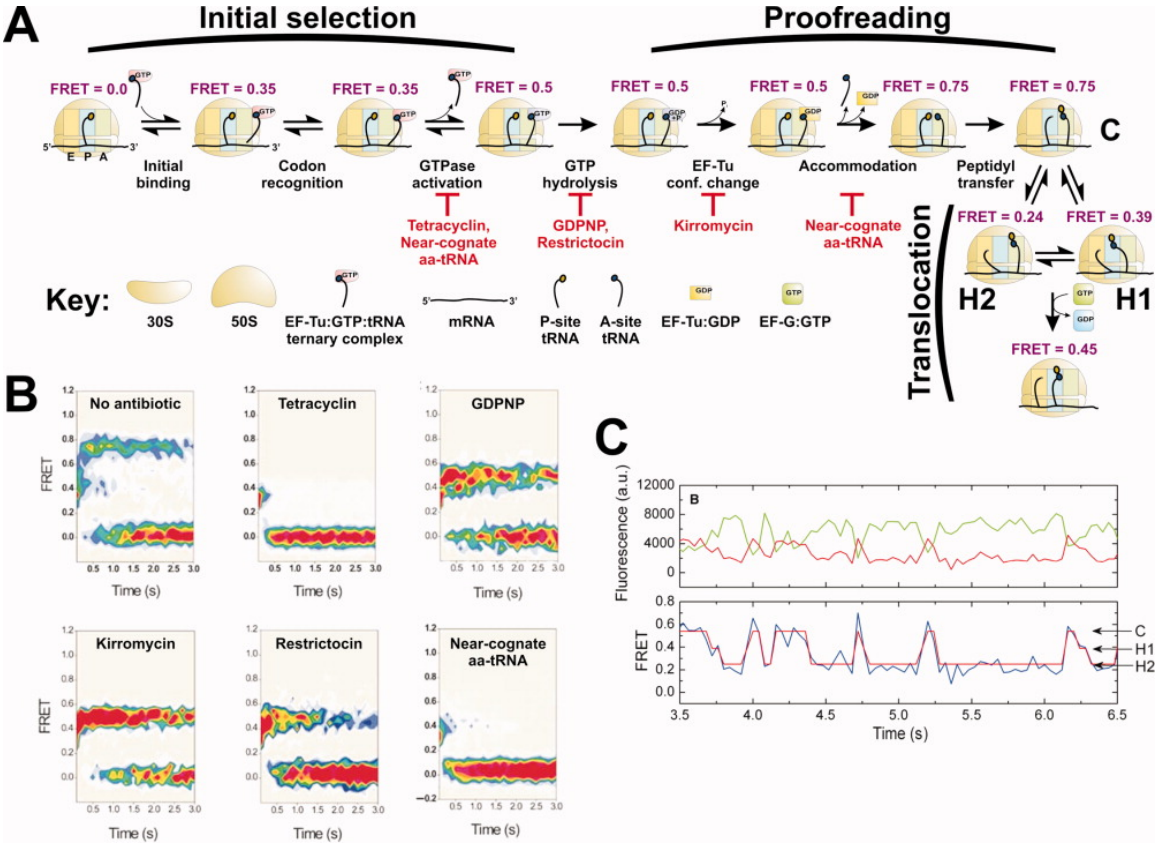


Figure 1.6: Dissecting the complex ribosomal translation cycle by single molecule FRET. (A) Schematic of the reaction cycle of the ribosome as illuminated through smFRET (FRET values are indicated in purple). The E (yellow), P (blue) and A (green) sites on the ribosome are shown as small rectangles. Specific steps can be inhibited by the antibiotics or additives indicated in red. (B) Contour plots of the postsynchronized time evolution of many individual smFRET trajectories in the presence of the indicated antibiotics and additives. Contours are plotted from tan (lowest population) to red (highest population).(56) (C) Donor and acceptor fluorescence signals and corresponding smFRET time trajectory displaying the classic (C), hybrid I (H1), and hybrid II (H2) states. The FRET data are Hidden Markov modeled (red line) to determine dwell times in the three states and derive interconversion rate constants.(58) Reproduced with permission from Refs (56,58)

the 30S subunit, while their 3'-ends bind to the P and E sites of the 50S subunit, respectively (58). 9) Full translocation of the tRNA:mRNA complex from the P and A sites to the E and P sites of the ribosome, respectively, which is accelerated by GTP hydrolysis on the elongation factor G (EF-G). The ribosome is thus ready for the next round of elongation, where the A site is prepared to accept the next aa-tRNA ternary complex (Figure 1.6A).

The elongation cycle involves extensive conformation changes and the dynamic association of multiple RNA and protein complexes, which lend themselves to FRET based distance measurements. Single-molecule analysis is particularly well suited to address such a multi-step biological process where transitions between states do not remain synchronized within the ensemble and the resulting averaged signal cannot easily be deconvoluted. Finally, the recent explosion in structural and mechanistic insights into ribosome function (reviewed in, for example, Refs. (52,53,61,62)) provides a vast knowledge base for in-depth studies of the single molecule enzymology of translation.

So far, several single molecule fluorescence studies have been conducted on reconstituted ribosomes from *Escherichia coli* (56-58,63). In particular, intermediates involved in aa-tRNA selection are resolved by combining stopped-flow mixing with the use of antibiotics that inhibit tRNA selection, a non-hydrolyzable GTP analog (GDPNP) and mutant ribosomes. Single *E. coli* ribosomes with a Cy3-labeled aa-tRNA (fMet-$tRNA^{fMet}$) in the P site are immobilized via a 5'-biotinylated mRNA on a streptavidin coated and otherwise passivated quartz slide (56). An EF-Tu:GTP:Phe-$tRNA^{Phe}$ ternary complex, where the tRNA is labeled with Cy5, is delivered to the immobilized ribosomes via stopped-flow. The FRET level of single complexes proceeds stepwise from FRET = 0.35 in the initial bound state to FRET = 0.75 in the state with a fully A-site accommodated aa-tRNA (Figure 1.6A). Due to the asynchronous nature of initial aa-tRNA binding, individual time traces were "postsynchronized" to the first FRET $\geq$ 0.25 signal. The average time from this initial FRET state to complete accommodation is ~93 ms, in agreement with previous ensemble studies on aa-tRNA selection, which validates the single molecule approach (64).

Classic approaches to stop translation at defined points along the elongation cycle provide an elegant means of assigning the FRET states and so further dissect ribosome

kinetics (57). Tetracycline is an antibiotic that inhibits selection of the ternary complex by the ribosomal A site (Figure 1.6A). smFRET traces in the presence of tetracycline show the transient sampling of a FRET = 0.35 state, with rare excursions into higher FRET (≥ 0.5) states (Figure 4B) (56). Each FRET = 0.35 event is identified as an independent attempt of a single ternary complex to enter the ribosomal A site, upon which tetracycline interrupts aa-tRNA selection. Replacing GTP with its non-hydrolyzable analog GDPNP in the ternary complex is known to efficiently stall aa-tRNA selection prior to GTP hydrolysis (Figure 1.6A). Single molecule traces observed in the presence of GDPNP rapidly transit through the FRET = 0.35 state and show a stabilized FRET = 0.5 state (Figure 1.6B). In this pre-hydrolysis state, the ternary complex is in closer contact with the ribosome than in the 0.35 FRET state. Kirromycin inhibits the conformational change of EF-Tu after GTP hydrolysis (Figures 1.6A). In single molecule traces kirromycin also stops the ribosome at a FRET = 0.5 state after readily passing through the 0.35 FRET state (Figure 1.6B). Restrictocin is an α-sarcin toxin homolog that specifically cleaves the sarcin-ricin loop (SRL) of the ribosome, which is involved in activating the GTPase activity of EF-Tu. In the presence of this toxin, smFRET time traces of the ribosome again stall in a FRET = 0.5 state before aa-tRNA accommodation (Figure 1.6B).

The fidelity of initial aa-tRNA selection and proofreading is investigated by comparing encoded mRNAs with cognate (UUU), near-cognate (CUU) and non-cognate (AAA) codons at the A site (56). Analysis of smFRET traces shows that a non-cognate aa-tRNA is effectively rejected at the codon recognition state (FRET = 0.35, Figure 1.6A). By contrast, 65% of bound cognate aa-tRNAs advance to a FRET state higher than 0.35, while only 11% of near-cognate aa-tRNAs move forward. Therefore, the FRET = 0.35 state represents an important initial selection step for correct anticodon-codon pairing that favors cognate over near-cognate aa-tRNAs by ~6:1 (Figure 1.6A). A (second) proofreading step is based on the stability of the A-site aa-tRNA bound to the rearranged ribosome after the release of the EF-Tu-GDP complex, which occurs during accommodation and is identified by smFRET as a transition from FRET = 0.5 to FRET = 0.75 (Figure 1.6A). Cognate aa-tRNAs are favorably accommodated in the PTC over near-cognate aa-tRNAs by a ratio of 24:1 (56).

After peptidyl transfer, fluctuations between FRET = 0.75 and 0.45 states are observed, which were initially attributed to a dynamic exchange between the classic (C) and translocation hybrid states (H), respectively (Figure 1.6A) (57). More recently, detailed smFRET analysis yielded evidence for the formation of two hybrid states (H1 and H2) in equilibrium with each other and the classic state (Figures 1.6A and 1.6C) (58). Rate constants for intermediate steps in aa-tRNA selection, proofreading and translocation were obtained by evaluating the FRET states and determining their corresponding single event dwell times based on the statistical evaluation afforded by Hidden-Markov modeling (Figure 1.6C and discussion below).

In summary, work on the ribosome has demonstrated how even very complex reaction cycles can be followed in real-time at the single molecule level to determine rate constants of individual reaction steps. Work on the ribosome opens the door to single-molecule analysis of additional RNP complexes such as the spliceosome (see Chapter 4) and RNase P (see Appendix A). These systems possess the necessary broad foundation of prior ensemble measurements that make them ideal candidates for single molecule approaches.

The spliceosome is a massive RNP complex that assembles to excise introns with single-nucleotide precision from pre-mRNAs as part of mRNA maturation in all eukaryotes (65-67). pre-mRNA splicing is accompanied by a complex series of conformational rearrangements of the central, presumably catalytic small nuclear (sn)RNAs, guided by a large number of protein cofactors including several RNA helicases. The presence of multiple global conformational changes along the splicing reaction pathway provides numerous opportunities to follow activity by smFRET (see Chapter 4). Additionally, the RNA helicases in the spliceosome are ATP dependent, which provides a convenient means of controlling progress along the reaction trajectory.

RNase P is an evolutionarily ancient enzyme that universally catalyzes the removal of the 5' end of pre-tRNAs as part of tRNA maturation in all organisms. Activity of the ribozyme involves the site-specific coordination of substrate, divalent metal ions, and at least one protein component by the RNase P RNA (68-70). Despite extensive study and several recent crystal structures of the RNA component, important details of the reaction pathway remain unclear (70). smFRET, which so far had only

been applied to folding of the isolated catalytic domain of RNase P RNA as described above, is ideally suited for dissecting this reaction pathway (see Appendix A).

Technical Challenges in smFRET

The examples above demonstrate how previous knowledge from ensemble experiments can be used to inform single molecule experimental design and interpretation and thus dissect multi-step enzymatic reaction pathways. smFRET observations do not require synchronicity between molecules to extract rate constants from statistical analyses of the underlying stochastic (even rare and/or brief) dwell times, making them uniquely suited for answering questions of RNA enzymology. (State dwell times often define the macroscopic rate constants since conformational transitions in RNA are typically fast, but rare. Only recently have attempts been made to measure an actual folding transition time (71).) As with any enzymatic assay, however, challenges lie in developing an assay, signal detection conditions, and strategies for extracting and interpreting experimental observables, and in deriving kinetic models. In the following we discuss some of the important challenges and highlight possible solutions.

Evaluating Dwell Times Correctly

In kinetic analyses of single molecule traces, dwell times must be measured with confidence. Large ribozymes may introduce an added level of complexity into the analysis as they may not display simple two-state FRET behavior with well defined dwell times. To deal with increasingly complex single molecule data sets a number of statistical methods have been implemented. Hidden Markov modeling (HMM), in particular, has recently been utilized to determine the number of distinct states and the transitions probabilities between states(58,72,73). HMM in principle can identify a large number of states within a single molecule trajectory in a semi-automated, reproducible fashion, thus avoiding potential bias. However, the assumptions implicit in any statistical model need to be kept in mind; for example, Hidden Markov modeling requires that the transitions be Markovian in nature, i.e., the current state is independent of past states. This assumption requires transitions within a single trajectory to be dictated by a single rate constant, which is not necessarily the case for RNA. It is therefore wise to apply

HMM independently to each single molecule trace to avoid masking any molecular heterogeneity that may be present. Furthermore, even individual trajectories may violate the assumption of Markovian behavior. Nevertheless, HMM has been successfully used to evaluate smFRET trajectories in the ribosome work discussed above (Figure 1.6C) (58), as well as in studies of Holliday junctions (72), and RecA filament assembly on single-stranded DNA (Figures 1.7A and 1.7B) (73). The extent to which HMM can be used to distinguish FRET states is demonstrated by the latter study. smFRET trajectories monitoring sequential association of up to four RecA monomers into a filament were analyzed, leading to five discernable FRET states with eight transition densities and fundamental rate constants for the stepwise binding and dissociation rates of individual monomers (Figure 1.7B) (73).

In some cases, such as in a folding study of the catalytic domain of *Bacillus subtilis* RNase P RNA, gradual structural FRET changes that lack abrupt transitions may be observed, limiting the ability to define FRET states (Figure 1.7C) (74). While the fully unfolded (0 mM Mg^{2+}) and folded (5 mM Mg^{2+}) states of the catalytic domain display narrow distributions with FRET = 0.13 and 0.85, respectively, at intermediate and physiologically relevant Mg^{2+} concentrations (at least) four FRET distributions are discerned (Figure 1.7D). Any particular RNA molecule is restricted to a limited range of FRET values (although this may in part be related to the unusually short observation window in this particular study, see also discussion below) (74). The authors therefore evaluate the folding pathway using free energy contour maps (Figure 1.7E). This approach assumes thermodynamic equilibrium since free energies are derived from probabilities based on the relative population sizes of folding states in the ensemble. The authors conclude that early folding steps in the catalytic domain of RNase P RNA involve a series of intermediates that fold under the kinetic control of local conformational rearrangements. Similar free energy contour map approaches may prove useful in evaluating progress along the reaction coordinate of RNA enzymes.

A significant limitation for any fluorescence based single molecule study is the non-ideal photophysical behavior of fluorophores; in particular, their photobleaching limits the total observation window (so that rate constants extracted from dwell times have to be corrected(44)) and long-lived dark states may persist for seconds. The donor-

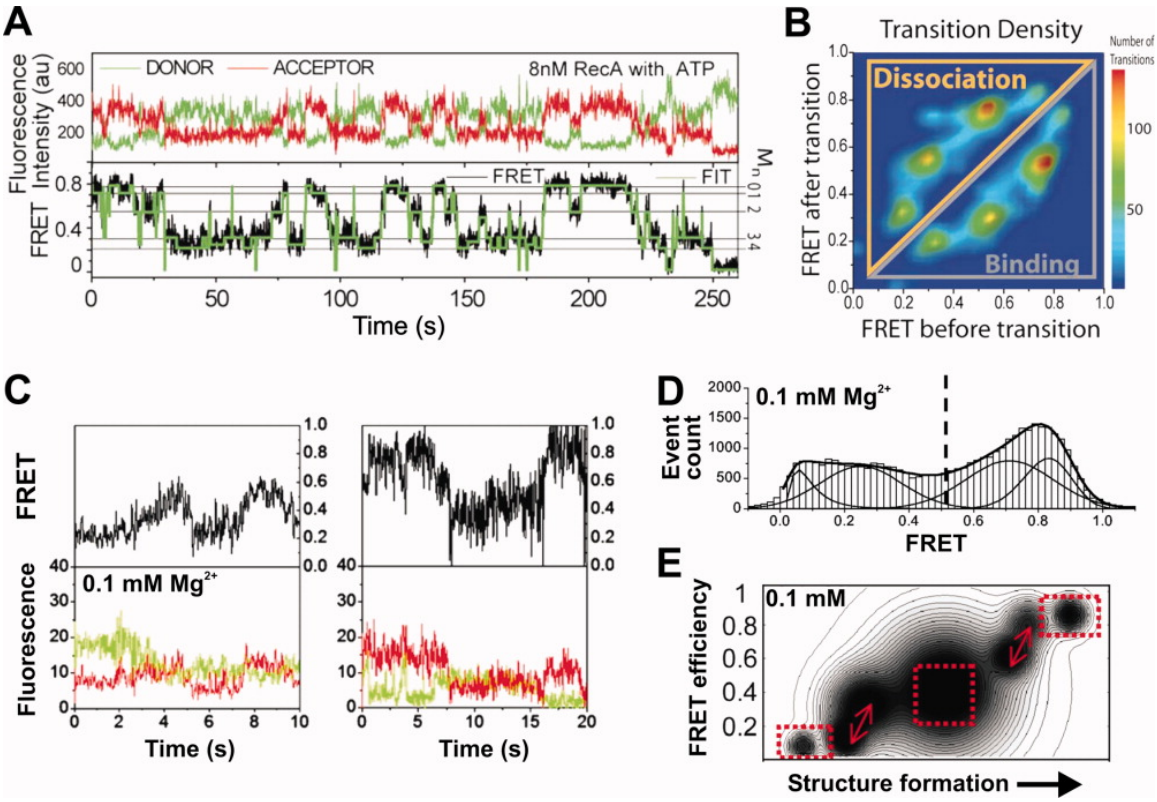


Figure 1.7: Challenges: Dealing with complex single molecule FRET kinetics. (A) Donor and acceptor signals and corresponding smFRET time trajectory upon assembling RecA in the presence of ATP onto the single-stranded 3'-extension of a double-stranded DNA. The FRET data are Hidden Markov modeled (green line) to determine dwell times in five different states (M_0 to M_4) and distinguish from acceptor dark states (FRET = 0). (B) A transition density plot of RecA binding and dissociation transitions observed on 82 DNA molecules shows five FRET states (FRET ≈ 0.2, 0.3, 0.55, 0.75, 0.85) that interconvert pairwise.(73) (C) smFRET time trajectories and donor and acceptor fluorescence signals from the catalytic domain of RNase P incubated at 0.1 mM Mg^{2+} reveal gradual transitions between poorly defined FRET states. (D) FRET distribution histogram from ~50 smFRET time trajectories of the RNase P catalytic domain observed at 0.1 mM Mg^{2+}. (E) Resulting free energy contour plot for the folding pathway of the RNase P catalytic domain as monitored by 3'- to 5'-end proximity. Two fluctuating classes, reflected by two pairs of closely connected basins (double arrows), as well as three nonfluctuating classes of smFRET states (dashed boxes) can be defined.(74) Reproduced with permission from Refs.(73,74)

acceptor pair Cy3/Cy5 is often favored in single molecule studies due to its large wavelength difference and strong FRET signal that can be prolonged by enzymatic oxygen scavenger systems (75). However, low FRET states have been observed in smFRET studies that arise from the acceptor temporarily visiting a dark state (Figure 1.7A) (44,73,76). This so called "blinking" is of concern because it can potentially be misinterpreted as a conformational change in the labeled molecule. The problem is exacerbated by the observation that the blinking kinetics vary depending on identity of the donor, inter-fluorophore distance, and buffer conditions (77,78). Fortunately, the resulting FRET = 0.0 is often sufficiently distinct from a "real" low FRET signal (Figure 1.7A). If this is not the case, rapidly alternating-laser excitation (ALEX) of the donor and acceptor fluorophores provides a solution, whereby acceptor activity is continuously probed as a control (77,79). Other promising advances toward longer smFRET observation windows may be expected from additives such as Trolox (6-hydroxy-2,5,7,8-tetramethylchroman-2–carboxylic acid), which suppress blinking and photobleaching of Cy5 (80), as well as from the development of improved fluorophores (81).

1.3 Molecular Dynamics and Quantum Mechanics of RNA: Conformational and Chemical Change We Can Believe In

Experimental studies of RNA can be complemented by computational techniques, provided they are judiciously applied to addressing issues that can be described with acceptable accuracy. Computational methods can provide insights that are not fully accessible through experimental techniques. Mere reproduction of known experimental data, though highly desirable for accuracy assessment, is not the primary goal. Simulations can reveal functionally significant stochastic fluctuations (82-84), identify problematic aspects of experimental structures (85,86), suggest whether a given base is protonated (87), and predict detailed solvent behavior (36) (88-90).

Molecular Dynamics Simulations: Overview

Explicit solvent MD is an atomistic technique dealing with a single molecule of defined starting geometry that subsequently undergoes 1-1000+ ns dynamics at room

temperature. MD provides unprecedentedly detailed information about all aspects of the time development (with sub-ps time resolution) of 3D structures, including positions of all waters, and ions. This picture, however, is limited by significant approximations. Simulations are limited by incomplete sampling of conformational space due to the short timescale of MD compared to biological processes. This limitation is being subdued step-by-step with the advance of faster computers. The truly fundamental approximation that is not waning with faster computers is the force field. The simulated systems are described by a simplistic analytic atomistic function (force fields) relating structures with potential energies (91,92). Assessment of the force field quality needs to be done case by case. The more qualitative the computational task, the more likely is the force field description sufficient. High-accuracy starting structures for MD are essential; usually x-ray diffraction or NMR data are used (93). Quantum chemistry (QM) provides a basis and rigorous benchmark by which to gauge force field quality (94-97), and coupling MD simulations to QM analysis provides a means of explicitly addressing chemical reactions (98).

Standard MD simulations can be enriched by enhanced sampling techniques (locally enhanced sampling, replica exchange or targeted MD) and free energy computations (direct or by post-processing analysis of the trajectories via Poisson-Boltzmann or Generalised Born theory). All these methods assume additional major approximations (99,100). While such calculations can provide very interesting insights, their reliability is less apparent. These computations are not included as part of this thesis.

Force Fields

Despite sophisticated parameterizations, the force field formally consists of a set of harmonic springs for bond lengths and valence angles supplemented by torsion profiles for dihedral angles (91,92,101,102). Atoms are approximated by Lennard-Jones spheres and constant point charges localized at atomic centers (101,103). Such force fields are pairwise-additive, which means they do not explicitly include polarization and charge-transfer effects. Many-body terms are by definition zeroed.

Force fields rely on the mutual compensation of errors which depends on the balance of forces in the studied system and the force field parameterization. There are two basic scenarios: i) The compensation of errors is sufficient and the force field gives basically the correct global minimum of the simulated system (96). In this case not necessarily all details are correct, but the overall description is (similar to error margins in experiments). ii) The force field does not give the correct global minimum and the simulated system sooner or later degrades (104). Such degradation may start as early as within a few nanoseconds, but in other cases it can take millisecond (or longer) to occur. In the latter case the beginning of the simulation can still rather successfully characterize the starting structure. There are two widely used nucleic acid (NA) force fields, AMBER and CHARMM (91,92). They share similar functional forms but differ in parameterization (105). AMBER has been successfully tested for many folded RNAs and non-canonical DNAs (36,82,84-90,96,97,102,104,106-109). CHARMM describes B-DNA well but has not been systematically tested for either folded RNAs or non-canonical DNAs. For A-RNA the CHARMM force field predicts unnaturally accelerated base pair breathing, which may indicate an underestimation of the strength of base pairing (107,110). Both AMBER and CHARMM offer high-quality protein force fields that allow consistent description of protein-nucleic acid complexes. The MD simulations in this thesis utilize the AMBER force field.

The AMBER parameterization describes the key electrostatic terms using atomic charges derived to reproduce the electrostatic potential around NA building blocks (101). Quantum chemical calculations show that base stacking is the best approximated term in RNA simulations (95), followed by base pairing including non-Watson-Crick base pairs utilizing the 2'-OH group (94,97). This contributes to the overall stability of RNA simulations. Description of the very flexible backbone is assumed to be inherently less accurate (102). The phosphate group is a polarizable anion and the dihedral space of backbone possesses multiple sub-states. Its description would profit from geometry-dependent electrostatics and polarization terms. Monovalent ions and solute-solvent interactions are also assumed to be reasonably described, whereas an accurate description of divalent ions is outside the applicability of the force field. Ions are represented simply as Lennard-Jones spheres with constant point charges. In reality the first ligand shell of

divalents is highly activated through polarization and charge-transfer, its waters can form very strong hydrogen-bonds, and these effects propagate beyond the first ligand shell (111). All these contributions are neglected by pair-additive force fields. Even when the divalent ions are initially placed based on experimental structures their observed dynamics are imperfect.

Good performance can be achieved by specialized force fields. These, however, cannot be used to simulate biomolecules where the need arises to catch a multitude of interactions simultaneously. Biomolecular force fields are parameterized for multipurpose use with delicate trade-offs in the parameterization. There are considerable and tremendously challenging efforts to develop more physically accurate, multipurpose force fields that would include polarization effects (92,112-115). Since such force fields will have to be more sophisticated it can be paradoxically more difficult to achieve a satisfactory balance of all parameters compared to the presently available simplistic force fields.

More Accurate Methods

In contrast to force fields, Quantum Chemistry (QM) can achieve a physically based and close to ultimate accuracy description of a chemical system. Ab initio QM methods are free of empirical parameters and offer a systematic (controllable) tuning of their quality via improving the basis sets of atomic orbitals and inclusion of electron correlation. Such calculations are, however, limited to 30-50+ atoms. QM allows evaluation of intrinsic (gas phase) interaction energies, defined as differences between the electronic energy of the dimer in a given geometry and the non-interacting monomers. This direct structure – energy relationship can be accurately calculated for any single geometry of stacking or base pairing to map the whole potential energy surface (95,97). Such energies unambiguously reflect direct forces between the interacting subsystems with no influence of the environment, making QM a genuine reference tool to parameterize and verify other computational approaches such as classical MD simulations (94,97) However, it is not straightforward to extrapolate QM data to biomolecules. Nucleic acids conformations result from a highly variable mixture of mutually compensating interactions and their balance may even vary for distinct types of nucleic acid architectures. Electrostatic

forces are substantially modulated by solvent screening effects. Specific care is needed when including NA backbone in QM studies. Small model systems in isolation favor (even for a single nucleotide) geometries that are biased by spurious intramolecular hydrogen bonds. The ionic effects associated with the phosphates typically dominate the energetics of such model calculations, especially when two or more phosphates are included (116). The problem is not quality of the QM methods per se, but incompleteness of the model systems; related experiments would suffer from the same problem.

Simulation Performance Judged by Comparison to Experimental Structures

Complementary to QM, evaluation of the overall performance of a force field for a given system can be determined by comparison to available experimental data. Evaluation of the RNA backbone is difficult for both experimental and computational approaches. The flexibility and polar nature of the backbone provide obvious difficulties for non-polarizable point-charge based force fields. Additionally the point-charges are based on the electrostatic potential of the most common C3'-endo sugar pucker and does not explicitly account for the relatively common C2'-endo pucker which may not be well represented (101). However, we usually observe surprisingly good agreement between experiment and simulation due to compensating errors, base pairing/stacking constraints and accurate starting structures.

An example of surprisingly good backbone behavior comes from our simulations of the hairpin ribozyme (see Chapter 3). Over the course of the simulations the dihedrals in a lower resolution structure switched to those observed in the higher resolution structures. Structural bioinformatics assisted in the rapid evaluation of backbone behavior during these simulations to rapidly determine if the simulated backbone adopts conformation commonly observed in experimental structures (117). However, not all simulations are correct. A textbook example of force field problems arises in simulations of guanine quadruplex DNAs (G-DNA) which consist of four-stranded stems formed by guanine quartets stabilized by tightly bound monovalent ions and complemented by single-stranded hairpin loops. The force field AMBER provides a global minimum consistent with experimental structures for the G-DNA stems but not for the loops (104) Thus, in a given simulation, different parts of a molecule can be

described with different success, depending on whether the compensation of errors is sufficient to reflect the overall balance of forces between various sub-states. Failure was unexpectedly detected also for long (15-50+ ns) B-DNA simulations, due to accumulation of irreversible incorrect backbone sub-states with concomitant progressive degradation of the whole structure. The γ backbone torsional profile was subsequently reparametrized(102). The improved force field allows stable microsecond-scale B-DNA simulations and even repairs partially degraded B-DNA structure, indicating that B-DNA is now the global minimum (96). Fortunately, the above two examples are the worst problems of the AMBER NA force field known to us. For most other systems inaccuracies of this magnitude are not encountered, with no visible instability stemming from pathological backbone substates. Specifically, backbone flip problems are not observed in RNA simulations (84,88) and considering its simplicity, the AMBER performance for RNA reported so far is surprisingly good. It is nevertheless fair to admit that the improved force field is still not satisfactory for the G-DNA loops and longer simulations of RNA tetraloops (unpublished results from Jiri Sponer) also indicate problems. Single stranded hairpin loops may be the Achilles heels for NA simulations.

Application of MD Depends on Availability of High Resolution Structures.

Provided reasonably accurate starting structures, MD can locally improve for example pairing and stacking patterns and minor backbone distortions in lower resolution experimental structures (93). Swift structural changes in the simulation usually indicate that the starting structure is high in energy and imperfect. Due to force field and sampling limitations, MD cannot predict RNA 3D structures and if the starting structure is locked in incorrect conformation via large energy barriers (>5kcal/mol), the simulation does not move away from the starting geometry (107). Unrealistic structures (mostly models) often fall apart swiftly. It is therefore necessary that high resolution structures be available for reliable evaluation of a given system by MD.

After collecting long enough simulation(s), careful inspection of the trajectory development compared with the starting experimental data should be done. Analysis of few distances of interest accompanied by the essentially uninformative RMSd plots is of limited utility and may mask considerable problems. During MD simulations the

observed structure developments result from a mixture of factors, including actual stochastic flexibility of RNA, experimental artifacts introduced through crystal contacts, disorder or modification, and finally force field artifacts. If this mixture is properly resolved then the analysis can be very insightful.

Flexibility of RNA Building Blocks

The stochastic flexibility of RNA is a key feature that is difficult to get from experiments. Stochastic fluctuations are important for function of biomolecular machines such as the ribosome but are not directly visualized by current structural experiments which show static averaged picture of the biomolecular machines (84,118). Simulations can help develop a qualitative understanding of stochastic dynamics and flexibility of key RNA building blocks (Figure 1.8) (82-84). Simulations of k-turn motifs revealed striking intrinsic dynamics and flexibility of these RNA building blocks which may contribute to the functional dynamics of the ribosome as flexible anisotropic molecular elbows. Importantly, the observation of profound dynamics of the GTPase associated center RNA (helices 42-44 of the large ribosomal subunit) requires application of an explicit solvent atomistic force field as a more course-grained normal mode analysis assuming harmonic approximation does not reflect this large-scale flexibility (82).

Experimental Artifacts

Simulations can resolve regions of limited resolution and structural effects stemming from crystal packing. A local region of lower resolution is observed close to the active site of the HDV ribozyme and MD simulations appear to improve the description of the RNA backbone near the active site (85) (Figure 1.9A). In the crystal structure the backbone cleavage site between U-1 and G+1 displays an unusual set of dihedral angles, which become more regular over the course of the simulations. Simulations ultimately reveal a U-turn motif with canonical dihedral values (85). Crystal structures of the HDV ribozyme also include an extruded guanine which participates in an inter-molecular crystal contact (119). Multiple simulations reveal the rapid loss of this conformation and suggest a possible role in promoting catalysis through novel interactions with stem P1 (86).

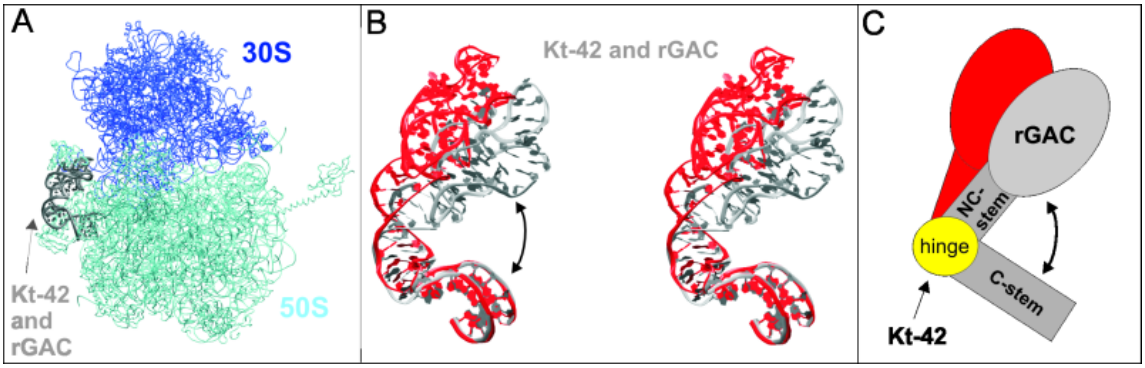


Figure 1.8: All atom, explicit solvent MD simulations can be applied to understanding global dynamics of RNA. (A) Context of the Kt-42 (the rRNA portion of the GTPase-associated center) within the ribosome. A (B) three dimensional and (C) schematic representation of global motion of Kt-42 identified through the analysis of MD trajectories. This global motion is not detected in a coarse grained normal mode analysis.

2'-deoxy or 2'-O-methyl backbone modifications or inactivating mutations are often necessary to trap ribozymes in their precatalytic structures. Simulations described in this thesis indicate that a 2'-O-methyl backbone modification of the hairpin ribozyme significantly distorts its active site. Multiple MD simulations consistently result in a change in the A-1 sugar pucker immediately 5' of the cleavage site, leading to a significant repositioning of catalytically important functional groups (Figure 1.9B)(see also Chapter 3).

Solvent Dynamics

MD can detect long-residency water molecules that occupy prominent hydration sites and stay bound for many nanoseconds, contrasting with more common individual water binding events of only ~50-500 ps. The dynamics of long-residency waters is highly variable and they can serve structural, functional and possibly catalytic roles (36,83,89,90). Simulations of the hairpin ribozyme predicted the presence of inter-domain long residency waters which was ultimately verified by the emergence of higher resolution structures (35,36). An intriguing long-residency hydration site was detected in the A-minor I tertiary interactions in k-turns 38 and 42 of the large ribosomal subunit. Their A/C base pairs dynamically oscillate between direct and water-mediated H-bond which significantly contributes to the elbow-like flexibility of the k-turns. The "static" X-ray structures show both geometries. The Kt-38 A/C interaction is water-mediated and that of Kt-42 direct.

Simulations can qualitatively describe major binding sites for monovalent ions primarily determined by electrostatics. Simulations in monovalents alone show ion density in known multivalent ion binding sites in simulations of multiple systems (36,87,88). Simulations of the HDV ribozyme exhibit monvalent binding at the cleavage site in a crystalographically resolved divalent ion binding site proximal to the 5'-O leaving group. In the simulations two Na^+ ions and accompanying first hydration sphere fill the catalytic pocket, a prediction with implications for catalysis in the absence of divalent ions. The MD predicted behavior of monovalent ions was later experimentally verified by crystal structures solved in the presence of Tl^+ (120) that reveal two Tl^+ ions at the active site. Simulations further predict a competition between ion binding and the

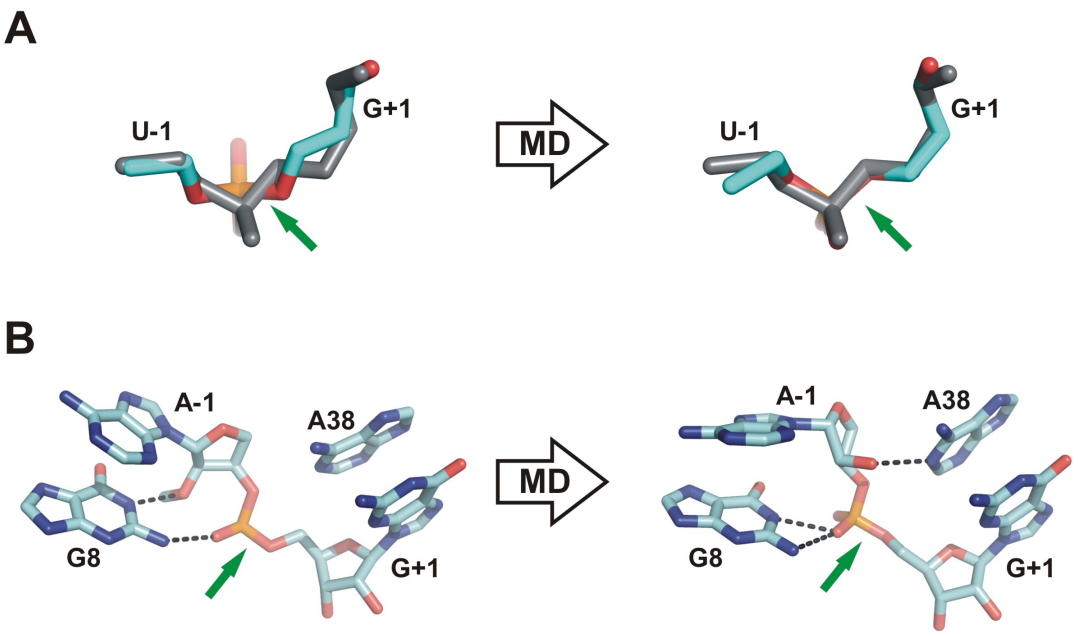


Figure 1.9: Molecular dynamics can improve backbone conformations in high resolution crystal structures. (A) Rotation of dihedral angles toward values consistent with the U-turn motif (gray sticks) occurs over the course of the HDV simulations (colored sticks) and results in what is likely an improved description of the backbone relative to the experimentally determined starting geometry. The starting geometry and final conformation of the backbone heavy atoms from the HDV ribozyme cleavage site are overlaid with a U-turn motif from a crystal structure of the hammerhead ribozyme (PDB ID 2GOZ) (B) An active site conformational change in the hairpin ribozyme reflects the loss of a catalytically inactivating 2'O-methyl modification and results in a significant rearrangement of functional group experimentally determined to be important to catalysis. The cleavage site is indicated with a green arrow for the HDV and hairpin ribozymes.

protonation state of C75, a feature not evident in the experimental structure. An additional binding site is predicted from simulations proximal to the 2'-OH nucleophile and this site is again verified by crystal structures, however the exact coordination geometry differs between experiment and simulation, likely reflecting a combination of differences between the ions used, force field approximations, and crystallographic artifacts. Other ion-binding sites can elude experimental detection, either due to low resolution or due to the fact that the ions are inherently delocalized in the pocket. Simulations of the 5SrRNA loop E and the HIV-DIS kissing complex reveal major monovalent ion binding sites that are to a significant extent delocalized (89,90,107,121).

Base Substitutions and Ionization

MD can be used to evaluate the effects of base substitutions (including protonated and deprotonated bases) at atomic level of resolution in known 3D structures. In this respect MD can efficiently complement experimental mutagenesis studies. Multiple base substitutions were modeled into experimentally determined crystal structures of the hairpin ribozyme (36). Good agreement is observed between MD and experimentally derived stability of the tertiary structure. The ddG for various hairpin ribozyme mutants derived from single-molecule FRET studies scales linearly with the number of inter-domain hydrogen bonds lost after 10 ns of simulation (36). These simulations involving the hairpin ribozyme reveal the importance of coupled molecular motions in tertiary structure stability. The loss of inter-domain stabilizing hydrogen bonds reflect long-range effects of mutations distant from the inter-domain interface, and reveal the importance of a coupled network of hydrogen bonds which includes long residency water molecules(36). For simulations of the HDV ribozyme all four standard nucleobases were modeled into the position immediately 5' of the cleavage site. For the HDV ribozyme the wild-type U-1 has the most tightly folded catalytic core in the simulations, fully consistent with the Tb(III) foot-printing data obtained in the same study (85). These simulations of the HDV reveal the importance of hydrogen-bonding of the -1 nucleobase in stabilizing a previously unnoticed U-turn motif at the cleavage site likely contributing to the increased catalytic activity. Simulations of the HDV and hairpin ribozymes are consistent with experimental data thus generating confidence in the simulated dynamics

which additionally reveal features uniquely accessible to MD thereby pushing MD beyond the simple reproduction of what was previously known about the two systems.

Simulation of both the HDV and hairpin ribozymes have been used to assess the impact of nucleobase identity and protonation state in promoting catalysis. In simulations of HDV the protonation state of C75 was varied, the simulations in which C75 is in its neutral form adopt a geometry suitable for general base catalysis (87,122), but complementary simulation of the charged C75H+ do not predict a reasonable starting geometry for C75 to act as a general acid (87). Similar to C75 in the HDV, compelling biochemical evidence for a direct catalytic role for ionization of A38 in the HP ribozyme led us to carry out simulations with the protonated and neutral forms of A38. Simulations with A38 unprotonated led to a geometry compatible with A38 acting as a general base, a role consistent with the available biochemical evidence but previously discounted based on crystallographic distances likely influenced by the backbone 2'-O-methyl modification in the crystal structures (35) (Figure 1.9B)(see also Chapter 3). Unlike the HDV, simulations with a protonated A38H+ also provide a geometry suitable for general acid catalysis. Of course, assessing a potential catalytic mechanism is limited by classical MD as bond-breaking and bond-formation are by definition not observed, and subsequent QM analysis is warranted to further evaluate the feasibility of mechanisms suggested by classical MD and biochemical data.

Accessing Chemistry: QM of Ribozymes

A wide spectrum of fast and rather accurate QM approaches have emerged recently allowing the inclusion of hundreds of atoms (123). Unfortunately, making the QM systems larger but still not complete will exemplify the errors resulting from the incompleteness of the system. However, these fast QM methods will facilitate applications of hybrid methods (QM/MM) where a smaller part of the system is treated quantum-chemically and the rest including solvent is treated classically using force fields. QM is particularly attractive for RNA enzymes, since, in contrast to force fields, QM can describe chemical reactions. The main limitations of QM/MM method come namely from insufficient sampling, quality of QM or MM methods used, treatment of boundaries and communication between QM and MM parts. In order to enhance QM/MM sampling, low

level semiempirical (AM1, SCC-DFTB etc.) and empirical (EVB) QM methods, which do not have to describe the reaction barrier accurately enough, are often used.

Calculations using hybrid QM/MM methods predict specific roles for nucleobases, divalent ions and/or electrostatic stabilization in catalyzing self-cleavage by the hammerhead, hairpin and HDV ribozymes (124) (125). MD simulations of HDV ribozyme with varied protonation state for active site cytosine (C75) provided a suitable general base geometry for the C75 unprotonated form. Combined QM/MM was further applied to determine plausibility of catalytic strategies suggested by classical MD analysis. QM/MM of the HDV ribozyme utilizing starting geometries derived from classical MD trajectories reveal a plausible role of C75 acting as a general base and Mg^{2+} as a general acid (122). The QM/MM analysis predicts an energy barrier of about 20 kcal/mol for this mechanism consistent with experimental kinetic data. For the QM scans a region made up of 80 atoms in the active site was treated quantum-mechanically using hybrid DFT method and coupled to the rest of the HDV ribozyme which was described by the classical AMBER force field. Multiple starting positions of a specifically bound Mg^{2+} were sampled, establishing a hexacordinated Mg^{2+} ion with a single innersphere contact to a cleavage site non-bridging oxygen as the most likely conformation with the Mg^{2+} acting as a Lewis acid in the cleavage mechanism. Mechanisms in which C75 acts as a general acid, as suggested by many recent biochemical studies (126,127) could not be explored due to the absence of suitable starting geometries derived from standard MD simulation. In contrast, hairpin ribozyme MD simulations described in Chapter 3 with both protonated and unprotonated forms of adenosine result in plausible catalytic geometries for A38 acting as either general acid, and/or general base respectively. These simulations reveal in part the impact of base ionization on the nature of the ground-state starting geometry, and may explain the apparent insensitivity to base ionization of initial QM/MM analysis of the ribozyme (98). Independent MD simulation and analysis prior to QM descriptions may therefore be essential in establishing starting geometries for subsequent QM analysis.

1.4 Conclusion

Current single molecule FRET microscopy approaches that detect large-scale RNA conformational changes in real-time have been previously employed to determine rate and equilibrium constants of folding as well as catalysis by ribozymes, including the small hairpin ribozyme and the large ribosome. Single molecule analysis has also exposed molecular heterogeneities and short-lived and/or rare intermediates otherwise masked in the ensemble average. In this thesis we exploit the power of single molecule RNA enzymology to provide a fresh focus on the biological function of the hairpin ribozyme and the spliceosome. In the application of these single molecule approaches care needs to be taken to develop suitable assays based on prior knowledge from ensemble studies. The preparation of proper controls, the conscientious analysis of statistically significant numbers of single molecule trajectories, and the comparison of single molecule results with ensemble data are all essential considerations.

MD simulations have proven useful in expanding on experimental structures and experimental biochemical data, providing detailed descriptions of the roles of nucleobases, ions, water, and backbone dynamics in imparting biological function to multiple systems. Experimental structural studies can profit from side by side MD analysis, using MD as an independent *a posteriori* tool to better understand the experimental structures. In this thesis we integrate our MD simulation data with previous structural studies to gain insight into the impact of active site dynamics and global stability of the hairpin ribozyme on its cleavage and ligation activity. In evaluating MD it is essential to take into account that simulations respond to averaging and error margins of the experimental data, and that there are always force field artifacts present. Unfortunately, in some instances, the force field is not sufficient to obtain meaningful results. Then it is appropriate to either fully acknowledge that (104) or to try to improve the method (102). As with any other method, the data gathered from computational efforts should always be considered in the context of all available experimental and computational data, and great care should be given to assessing limitations and to not overstating the strength of the resulting conclusions.

Powerful synergies can arise from the use of multiple alternate approaches. In

this thesis we apply smFRET, MD, and additional ensemble biochemical techniques (gel electrophoresis, mass spectrometry, footprinting, time-resolved and steady state FRET) to two systems encompassing the full range of catalytic RNAs, from the small purely RNA system of the hairpin ribozyme to the massive RNA-protein assembly that constitutes the spliceosome. Our understanding of RNA folding and conformational dynamics is advanced by the application of the methods described herein to these two systems. The future progress of our understanding of the structure-dynamics-function relationship in RNA will undoubtedly benefit from the continued investigation of both these and other RNA systems and through the careful integration of the methods used in this thesis, in particular, ensemble enzymology coupled with mutagenesis, X-ray crystallography, NMR spectroscopy, and quantum mechanical/molecular mechanical (QM/MM) calculations.

References

1. Crick, F.H.C. (1968) Origin of Genetic Code. *J Mol Biol*, **38**, 367-&.
2. Orgel, L.E. (1968) Evolution of Genetic Apparatus. *J Mol Biol*, **38**, 381-&.
3. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, **31**, 147-157.
4. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, **35**, 849-857.
5. Gesteland, R.F., Cech, T.R. and Atkins, J.F.E. (1999) *The RNA World, Second Edition*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
6. Yuan, Y.R., Pei, Y., Chen, H.Y., Tuschl, T. and Patel, D.J. (2006) A potential protein-RNA recognition event along the RISC-loading pathway from the structure of A-aeolicus argonaute with externally bound siRNA. *Structure*, **14**, 1557-1565.
7. Jovine, L., Djordjevic, S. and Rhodes, D. (2000) The crystal structure of yeast phenylalanine tRNA at 2.0 A resolution: cleavage by Mg(2+) in 15-year old crystals. *J Mol Biol*, **301**, 401-414.
8. Toor, N., Rajashankar, K., Keating, K.S. and Pyle, A.M. (2008) Structural basis for exon recognition by a group II intron. *Nat Struct Mol Biol*, **15**, 1221-1222.
9. Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C. *et al.* (2005) The transcriptional landscape of the mammalian genome. *Science*, **309**, 1559-1563.
10. Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J. *et al.* (2005) Antisense transcription in the mammalian transcriptome. *Science*, **309**, 1564-1566.
11. Taft, R.J., Pheasant, M. and Mattick, J.S. (2007) The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays*, **29**, 288-299.
12. Sullenger, B.A. and Gilboa, E. (2002) Emerging clinical applications of RNA. *Nature*, **418**, 252-258.
13. Breaker, R.R. (2004) Natural and engineered nucleic acids as tools to explore biology. *Nature*, **432**, 838-845.
14. Doudna, J.A. and Lorsch, J.R. (2005) Ribozyme catalysis: not different, just worse. *Nat Struct Mol Biol*, **12**, 395-402.
15. Fedor, M.J. and Williamson, J.R. (2005) The catalytic diversity of RNAs. *Nature Rev. Mol. Cell Biol.*, **6**, 399-412.
16. Winkler, W.C., Nahvi, A., Roth, A., Collins, J.A. and Breaker, R.R. (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature*, **428**, 281-286.
17. Ban, N., Nissen, P., Hansen, J., Moore, P.B. and Steitz, T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science*, **289**, 905-920.
18. Valadkhan, S. and Manley, J.L. (2001) Splicing-related catalysis by protein-free snRNAs. *Nature*, **413**, 701-707.

19. Al-Hashimi, H.M. and Walter, N.G. (2008) RNA dynamics: it is about time. *Curr Opin Struct Biol*, **18**, 321-329.
20. Zhuang, X. and Rief, M. (2003) Single-molecule folding. *Curr Opin Struct Biol*, **13**, 88-97.
21. Zhuang, X.W. (2005) Single-molecule RNA science. *Annual Review of Biophysics and Biomolecular Structure*, **34**, 399-414.
22. Walter, N.G., Hampel, K.J., Brown, K.M. and Burke, J.M. (1998) Tertiary structure formation in the hairpin ribozyme monitored by fluorescence resonance energy transfer. *Embo Journal*, **17**, 2378-2391.
23. Pinard, R., Hampel, K.J., Heckman, J.E., Lambert, D., Chan, P.A., Major, F. and Burke, J.M. (2001) Functional involvement of G8 in the hairpin ribozyme cleavage mechanism. *Embo Journal*, **20**, 6434-6442.
24. Lebruska, L.L., Kuzmine, I.I. and Fedor, M.J. (2002) Rescue of an abasic hairpin ribozyme by cationic nucleobases: Evidence for a novel mechanism of RNA catalysis. *Chemistry & Biology*, **9**, 465-473.
25. Kuzmin, Y., Da Costa, C., Cottrell, J. and Fedor, M. (2005) Role of an active site adenine in hairpin ribozyme catalysis. *JOURNAL OF MOLECULAR BIOLOGY*, **349**, 989-1010.
26. Kuzmin, Y., Da Costa, C. and Fedor, M. (2004) Role of an active site guanine in hairpin ribozyme catalysis probed by exogenous nucleobase rescue. *JOURNAL OF MOLECULAR BIOLOGY*, **340**, 233-251.
27. Pljevaljcic, G., Klostermeier, D. and Millar, D.P. (2005) The tertiary structure of the hairpin ribozyme is formed through a slow conformational search. *Biochemistry*, **44**, 4870-4876.
28. Cottrell, J.W., Kuzmin, Y.I. and Fedor, M.J. (2007) Functional analysis of hairpin ribozyme active site architecture. *J Biol Chem*, **282**, 13498-13507.
29. Cai, Z. and Tinoco, I., Jr. (1996) Solution structure of loop A from the hairpin ribozyme from tobacco ringspot virus satellite. *Biochemistry*, **35**, 6026-6036.
30. Butcher, S.E., Allain, F.H. and Feigon, J. (1999) Solution structure of the loop B domain from the hairpin ribozyme. *Nat Struct Biol*, **6**, 212-216.
31. Butcher, S.E., Allain, F.H. and Feigon, J. (2000) Determination of metal ion binding sites within the hairpin ribozyme domains by NMR. *Biochemistry*, **39**, 2174-2182.
32. Rupert, P.B. and Ferre-D'Amare, A.R. (2001) Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature*, **410**, 780-786.
33. Rupert, P.B., Massey, A.P., Sigurdsson, S.T. and Ferre-D'Amare, A.R. (2002) Transition state stabilization by a catalytic RNA. *Science*, **298**, 1421-1424.
34. Alam, S., Grum-Tokars, V., Krucinska, J., Kundracik, M.L. and Wedekind, J.E. (2005) Conformational heterogeneity at position U37 of an all-RNA hairpin ribozyme with implications for metal binding and the catalytic structure of the S-turn. *Biochemistry*, **44**, 14396-14408.
35. Salter, J., Krucinska, J., Alam, S., Grum-Tokars, V. and Wedekind, J.E. (2006) Water in the active site of an all-RNA hairpin ribozyme and effects of Gua8 base variants on the geometry of phosphoryl transfer. *Biochemistry*, **45**, 686-700.

36. Rhodes, M., Reblova, K., Sponer, J. and Walter, N. (2006) Trapped water molecules are essential to structural dynamics and function of a ribozyme. *P NATL ACAD SCI USA*, **103**, 13380-13385.
37. Tang, C.L., Alexov, E., Pyle, A.M. and Honig, B. (2007) Calculation of pKas in RNA: on the structural origins and functional roles of protonated nucleotides. *J Mol Biol*, **366**, 1475-1496.
38. Zhuang, X., Kim, H., Pereira, M.J., Babcock, H.P., Walter, N.G. and Chu, S. (2002) Correlating structural dynamics and function in single ribozyme molecules. *Science*, **296**, 1473-1476.
39. Bokinsky, G., Rueda, D., Misra, V.K., Rhodes, M.M., Gordus, A., Babcock, H.P., Walter, N.G. and Zhuang, X. (2003) Single-molecule transition-state analysis of RNA folding. *Proc Natl Acad Sci U S A*, **100**, 9302-9307.
40. Tan, E., Wilson, T.J., Nahas, M.K., Clegg, R.M., Lilley, D.M. and Ha, T. (2003) A four-way junction accelerates hairpin ribozyme folding via a discrete intermediate. *Proc Natl Acad Sci U S A*, **100**, 9308-9313.
41. Nahas, M.K., Wilson, T.J., Hohng, S., Jarvie, K., Lilley, D.M. and Ha, T. (2004) Observation of internal cleavage and ligation reactions of a ribozyme. *Nat Struct Mol Biol*, **11**, 1107-1113.
42. Okumus, B., Wilson, T.J., Lilley, D.M. and Ha, T. (2004) Vesicle encapsulation studies reveal that single molecule ribozyme heterogeneities are intrinsic. *Biophys J*, **87**, 2798-2806.
43. Pljevaljcic, G., Millar, D.P. and Deniz, A.A. (2004) Freely diffusing single hairpin ribozymes provide insights into the role of secondary structure and partially folded states in RNA folding. *Biophys J*, **87**, 457-467.
44. Rueda, D., Bokinsky, G., Rhodes, M.M., Rust, M.J., Zhuang, X. and Walter, N.G. (2004) Single-molecule enzymology of RNA: essential functional groups impact catalysis from a distance. *Proc Natl Acad Sci U S A*, **101**, 10066-10071.
45. Hampel, K.J., Walter, N.G. and Burke, J.M. (1998) The solvent-protected core of the hairpin ribozyme-substrate complex. *Biochemistry*, **37**, 14672-14682.
46. Pinard, R., Lambert, D., Walter, N.G., Heckman, J.E., Major, F. and Burke, J.M. (1999) Structural basis for the guanosine requirement of the hairpin ribozyme. *Biochemistry*, **38**, 16035-16039.
47. Bevilacqua, P.C. (2003) Mechanistic considerations for general acid-base catalysis by RNA: revisiting the mechanism of the hairpin ribozyme. *Biochemistry*, **42**, 2259-2265.
48. Torelli, A.T., Krucinska, J. and Wedekind, J.E. (2007) A comparison of vanadate to a 2'-5' linkage at the active site of a small ribozyme suggests a role for water in transition-state stabilization. *Rna*.
49. Walter, N.G. (2002) Probing RNA structural dynamics and function by fluorescence resonance energy transfer (FRET). *Curr. Protocols Nucleic Acid Chem.*, **11.10**, 11.10.11-11.10.23.
50. Benkovic, S.J. and Hammes-Schiffer, S. (2003) A perspective on enzyme catalysis. *Science*, **301**, 1196-1202.
51. Liu, S., Bokinsky, G., Walter, N.G. and Zhuang, X. (2007) Dissecting the multistep reaction pathway of an RNA enzyme by single-molecule kinetic "fingerprinting". *Proc Natl Acad Sci U S A*.

52. Rodnina, M.V., Beringer, M. and Wintermeyer, W. (2006) Mechanism of peptide bond formation on the ribosome. *Q Rev Biophys*, **39**, 203-225.
53. Ogle, J.M. and Ramakrishnan, V. (2005) Structural insights into translational fidelity. *Annu Rev Biochem*, **74**, 129-177.
54. Schuwirth, B.S., Borovinskaya, M.A., Hau, C.W., Zhang, W., Vila-Sanjurjo, A., Holton, J.M. and Cate, J.H. (2005) Structures of the bacterial ribosome at 3.5 A resolution. *Science*, **310**, 827-834.
55. Ramakrishnan, V. (2002) Ribosome structure and the mechanism of translation. *Cell*, **108**, 557-572.
56. Blanchard, S.C., Gonzalez, R.L., Kim, H.D., Chu, S. and Puglisi, J.D. (2004) tRNA selection and kinetic proofreading in translation. *Nat Struct Mol Biol*, **11**, 1008-1014.
57. Blanchard, S.C., Kim, H.D., Gonzalez, R.L., Jr., Puglisi, J.D. and Chu, S. (2004) tRNA dynamics on the ribosome during translation. *Proc Natl Acad Sci U S A*, **101**, 12893-12898.
58. Munro, J.B., Altman, R.B., O'Connor, N. and Blanchard, S.C. (2007) Identification of two distinct hybrid state intermediates on the ribosome. *Molecular cell*, **25**, 505-517.
59. Daviter, T., Murphy, F.V.t. and Ramakrishnan, V. (2005) Molecular biology. A renewed focus on transfer RNA. *Science*, **308**, 1123-1124.
60. Valle, M., Zavialov, A., Li, W., Stagg, S.M., Sengupta, J., Nielsen, R.C., Nissen, P., Harvey, S.C., Ehrenberg, M. and Frank, J. (2003) Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy. *Nat Struct Biol*, **10**, 899-906.
61. Noller, H.F. (2005) RNA structure: reading the ribosome. *Science*, **309**, 1508-1514.
62. Noller, H.F. (2006) Biochemical characterization of the ribosomal decoding site. *Biochimie*, **88**, 935-941.
63. Sytnik, A., Vladimirov, S., Jia, Y., Li, L., Cooperman, B.S. and Hochstrasser, R.M. (1999) Peptidyl transferase center activity observed in single ribosomes. *J Mol Biol*, **285**, 49-54.
64. Rodnina, M.V. and Wintermeyer, W. (2001) Fidelity of aminoacyl-tRNA selection on the ribosome: kinetic and structural mechanisms. *Annu Rev Biochem*, **70**, 415-435.
65. Staley, J.P. and Guthrie, C. (1998) Mechanical devices of the spliceosome: Motors, clocks, springs, and things. *Cell*, **92**, 315-326.
66. Villa, T., Pleiss, J.A. and Guthrie, C. (2002) Spliceosomal snRNAs: Mg(2+)-dependent chemistry at the catalytic core? *Cell*, **109**, 149-152.
67. Stevens, S.W. and Abelson, J. (2002) Yeast pre-mRNA splicing: methods, mechanisms, and machinery. *Methods Enzymol*, **351**, 200-220.
68. Hsieh, J., Andrews, A.J. and Fierke, C.A. (2004) Roles of protein subunits in RNA-protein complexes: lessons from ribonuclease P. *Biopolymers*, **73**, 79-89.
69. Rueda, D., Hsieh, J., Day-Storms, J.J., Fierke, C.A. and Walter, N.G. (2005) The 5' leader of precursor tRNAAsp bound to the Bacillus subtilis RNase P holoenzyme has an extended conformation. *Biochemistry*, **44**, 16130-16139.

70. Kazantsev, A.V. and Pace, N.R. (2006) Bacterial RNase P: a new view of an ancient enzyme. *Nature Reviews Microbiology*, **4**, 729-740.
71. Lee, T.H., Lapidus, L.J., Zhao, W., Travers, K.J., Herschlag, D. and Chu, S. (2007) Measuring the folding transition time of single RNA molecules. *Biophys J*, **92**, 3275-3283.
72. McKinney, S.A., Joo, C. and Ha, T. (2006) Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophysical Journal*, **91**, 1941-1951.
73. Joo, C., McKinney, S.A., Nakamura, M., Rasnik, I., Myong, S. and Ha, T. (2006) Real-time observation of RecA filament dynamics with single monomer resolution. *Cell*, **126**, 515-527.
74. Xie, Z., Srividya, N., Sosnick, T.R., Pan, T. and Scherer, N.F. (2004) Single-molecule studies highlight conformational heterogeneity in the early folding steps of a large ribozyme. *P NATL ACAD SCI USA*, **101**, 534-539.
75. Ha, T. (2001) Single-molecule fluorescence resonance energy transfer. *Methods*, **25**, 78-86.
76. Zhuang, X., Bartley, L.E., Babcock, H.P., Russell, R., Ha, T., Herschlag, D. and Chu, S. (2000) A single-molecule study of RNA catalysis and folding. *Science*, **288**, 2048-2051.
77. Sabanayagam, C.R., Eid, J.S. and Meller, A. (2005) Long time scale blinking kinetics of cyanine fluorophores conjugated to DNA and its effect on Forster resonance energy transfer. *Journal of Chemical Physics*, **123**, -.
78. Heilemann, M., Margeat, E., Kasper, R., Sauer, M. and Tinnefeld, P. (2005) Carbocyanine dyes as efficient reversible single-molecule optical switch. *Journal of the American Chemical Society*, **127**, 3801-3806.
79. Kapanidis, A.N., Laurence, T.A., Lee, N.K., Margeat, E., Kong, X. and Weiss, S. (2005) Alternating-laser excitation of single molecules. *Acc Chem Res*, **38**, 523-533.
80. Rasnik, I., McKinney, S.A. and Ha, T. (2006) Nonblinking and long-lasting single-molecule fluorescence imaging. *Nat Methods*, **3**, 891-893.
81. Conley, N.R., Pomerantz, A.K., Wang, H., Twieg, R.J. and Moerner, W.E. (2007) Bulk and Single-Molecule Characterization of an Improved Molecular Beacon Utilizing H-Dimer Excitonic Behavior. *J Phys Chem B*.
82. Razga, F., Koca, J., Mokdad, A. and Sponer, J. (2007) Elastic properties of ribosomal RNA building blocks: molecular dynamics of the GTPase-associated center rRNA. *Nucleic Acids Res*, **35**, 4007-4017.
83. Razga, F., Koca, J., Sponer, J. and Leontis, N.B. (2005) Hinge-like motions in RNA kink-turns: The role of the second A-minor motif and nominally unpaired bases. *Biophysical Journal*, **88**, 3466-3485.
84. Reblova, K., Lankas, F., Razga, F., Krasovska, M.V., Koca, J. and Sponer, J. (2006) Structure, dynamics, and elasticity of free 16s rRNA helix 44 studied by molecular dynamics simulations. *Biopolymers*, **82**, 504-520.
85. Sefcikova, J., Krasovska, M.V., Sponer, J. and Walter, N.G. (2007) The genomic HDV ribozyme utilizes a previously unnoticed U-turn motif to accomplish fast site-specific catalysis. *Nucleic Acids Res*, **35**, 1933-1946.

86. Sefcikova, J., Krasovska, M.V., Spackova, N., Sponer, J. and Walter, N.G. (2007) Impact of an extruded nucleotide on cleavage activity and dynamic catalytic core conformation of the hepatitis delta virus ribozyme. *Biopolymers*, **85**, 392-406.
87. Krasovska, M.V., Sefcikova, J., Spackova, N., Sponer, J. and Walter, N.G. (2005) Structural dynamics of precursor and product of the RNA enzyme from the hepatitis delta virus as revealed by molecular dynamics simulations. *J Mol Biol*, **351**, 731-748.
88. Krasovska, M.V., Sefcikova, J., Reblova, K., Schneider, B., Walter, N.G. and Sponer, J. (2006) Cations and hydration in catalytic RNA: molecular dynamics of the hepatitis delta virus ribozyme. *Biophys J*, **91**, 626-638.
89. Reblova, K., Spackova, N., Koca, J., Leontis, N.B. and Sponer, J. (2004) Long-residency hydration, cation binding, and dynamics of loop E/helix IV rRNA-L25 protein complex. *Biophys J*, **87**, 3397-3412.
90. Reblova, K., Spackova, N., Stefl, R., Csaszar, K., Koca, J., Leontis, N.B. and Sponer, J. (2003) Non-Watson-Crick basepairing and hydration in RNA motifs: molecular dynamics of 5S rRNA loop E. *Biophys J*, **84**, 3564-3582.
91. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *Journal of the American Chemical Society*, **117**, 5179-5197.
92. Foloppe, N., MacKerell, A.D. . (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J Comput Chem* **21**, 86-104.
93. Auffinger, P. (2006) In F, S. J. L. (ed.), *Computational studies of DNA and RNA*. Springer Verlag, pp. 283-300.
94. Sponer, J., Jurecka, P. and Hobza, P. (2004) Accurate interaction energies of hydrogen-bonded nucleic acid base pairs. *J Am Chem Soc*, **126**, 10142-10151.
95. Sponer, J., Jurecka, P., Marchan, I., Luque, F.J., Orozco, M. and Hobza, P. (2006) Nature of base stacking: reference quantum-chemical stacking energies in ten unique B-DNA base-pair steps. *Chemistry*, **12**, 2854-2865.
96. Perez, A., Luque, F.J. and Orozco, M. (2007) Dynamics of B-DNA on the microsecond time scale. *J Am Chem Soc*, **129**, 14739-14745.
97. Sponer, J.E., Reblova, K., Mokdad, A., Sychrovsky, V., Leszczynski, J. and Sponer, J. (2007) Leading RNA tertiary interactions: structures, energies, and water insertion of A-minor and P-interactions. A quantum chemical view. *J Phys Chem B*, **111**, 9153-9164.
98. Nam, K., Gao, J. and York, D.M. (2008) Quantum mechanical/molecular mechanical simulation study of the mechanism of hairpin ribozyme catalysis. *J Am Chem Soc*, **130**, 4680-4691.
99. Beck, D.A.C., White, G.W.N. and Daggett, V. (2007) Exploring the energy landscape of protein folding using replica-exchange and conventional molecular dynamics simulations. *Journal of Structural Biology*, **157**, 514-523.
100. Page, C.S. and Bates, P.A. (2006) Can MM-PBSA calculations predict the specificities of protein kinase inhibitors? *Journal of Computational Chemistry*, **27**, 1990-2007.

101. Cieplak, P., Cornell, W.D., Bayly, C. and Kollman, P.A. (1995) Application of the Multimolecule and Multiconformational Resp Methodology to Biopolymers - Charge Derivation for DNA, Rna, and Proteins. *Journal of Computational Chemistry*, **16**, 1357-1377.
102. Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T.E., 3rd, Laughton, C.A. and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys J*, **92**, 3817-3829.
103. Bayly, C.I., Cieplak, P., Cornell, W.D. and Kollman, P.A. (1993) A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *Journal of Physical Chemistry*, **97**, 10269-10280.
104. Fadrna, E., Spackova, N., Stefl, R., Koca, J., Cheatham, T.E., 3rd and Sponer, J. (2004) Molecular dynamics simulations of Guanine quadruplex loops: advances and force field limitations. *Biophys J*, **87**, 227-242.
105. Mackerell, A.D., Jr. (2004) Empirical force fields for biological macromolecules: overview and issues. *J Comput Chem*, **25**, 1584-1604.
106. Sponer, J. and Spackova, N. (2007) Molecular dynamics simulations and their application to four-stranded DNA. *Methods*, **43**, 278-290.
107. Reblova, K., Fadrna, E., Sarzynska, J., Kulinski, T., Kulhanek, P., Ennifar, E., Koca, J. and Sponer, J. (2007) Conformations of flanking bases in HIV-1 RNA DIS kissing complexes studied by molecular dynamics. *Biophys J*, **93**, 3932-3949.
108. McDowell, S.E., Spackova, N., Sponer, J. and Walter, N.G. (2007) Molecular dynamics simulations of RNA: an in silico single molecule approach. *Biopolymers*, **85**, 169-184.
109. Šponer, J., Lankaš, F. and SpringerLink (Online service). (2006). Springer, Dordrecht.
110. Pan, Y. and MacKerell, A.D., Jr. (2003) Altered structural fluctuations in duplex RNA versus DNA: a conformational switch involving base pair opening. *Nucleic Acids Res*, **31**, 7131-7140.
111. Sponer, J., Sabat, M., Gorb, L., Leszczynski, J., Lippert, B. and Hobza, P. (2000) The effect of metal binding to the N7 site of purine nucleotides on their structure, energy, and involvement in base pairing. *Journal of Physical Chemistry B*, **104**, 7535-7544.
112. Halgren, T.A. and Damm, W. (2001) Polarizable force fields. *Curr Opin Struct Biol*, **11**, 236-242.
113. Kaminski, G.A., Stern, H.A., Berne, B.J., Friesner, R.A., Cao, Y.X.X., Murphy, R.B., Zhou, R.H. and Halgren, T.A. (2002) Development of a polarizable force field for proteins via ab initio quantum chemistry: First generation model and gas phase tests. *Journal of Computational Chemistry*, **23**, 1515-1531.
114. Ren, P.Y. and Ponder, J.W. (2003) Polarizable atomic multipole water model for molecular mechanics simulation. *Journal of Physical Chemistry B*, **107**, 5933-5947.
115. Gresh, N., Sponer, J.E., Spackova, N., Leszczynski, J. and Sponer, J. (2003) Theoretical study of binding of hydrated Zn(II) and Mg(II) cations to 5 '-

guanosine monophosphate. Toward polarizable molecular mechanics for DNA and RNA. *Journal of Physical Chemistry B*, **107**, 8669-8681.
116. Svozil, D., Sponer, J.E., Marchan, I., Perez, A., Cheatham, T.E., 3rd, Forti, F., Luque, F.J., Orozco, M. and Sponer, J. (2008) Geometrical and electronic structure variability of the sugar-phosphate backbone in nucleic acids. *J Phys Chem B*, **112**, 8188-8197.
117. Richardson, J.S., Schneider, B., Murray, L.W., Kapral, G.J., Immormino, R.M., Headd, J.J., Richardson, D.C., Ham, D., Hershkovits, E., Williams, L.D. *et al.* (2008) RNA backbone: consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA*, **14**, 465-481.
118. Frank, J. and Spahn, C. (2006) The ribosome and the mechanism of protein synthesis. *Reports on Progress in Physics*, **69**, 1383-1417.
119. Ferre-D'Amare, A.R., Zhou, K. and Doudna, J.A. (1998) Crystal structure of a hepatitis delta virus ribozyme. *Nature*, **395**, 567-574.
120. Ke, A., Ding, F., Batchelor, J.D. and Doudna, J.A. (2007) Structural roles of monovalent cations in the HDV ribozyme. *Structure*, **15**, 281-287.
121. Auffinger, P., Bielecki, L. and Westhof, E. (2004) Symmetric K+ and Mg2+ ion-binding sites in the 5S rRNA loop E inferred from molecular dynamics simulations. *J Mol Biol*, **335**, 555-571.
122. Banas, P., Rulisek, L., Hanosova, V., Svozil, D., Walter, N.G., Sponer, J. and Otyepka, M. (2008) General base catalysis for cleavage by the active-site cytosine of the hepatitis delta virus ribozyme: QM/MM calculations establish chemical feasibility. *J Phys Chem B*, **112**, 11177-11187.
123. Zhao, Y. and Truhlar, D.G. (2008) Density functionals with broad applicability in chemistry. *Acc Chem Res*, **41**, 157-167.
124. Lee, T.S., Silva Lopez, C., Giambasu, G.M., Martick, M., Scott, W.G. and York, D.M. (2008) Role of Mg2+ in hammerhead ribozyme catalysis from molecular simulation. *J Am Chem Soc*, **130**, 3053-3064.
125. Liu, H., Robinet, J.J., Ananvoranich, S. and Gauld, J.W. (2007) Density functional theory investigation on the mechanism of the hepatitis delta virus ribozyme. *J Phys Chem B*, **111**, 439-445.
126. Das, S.R. and Piccirilli, J.A. (2005) General acid catalysis by the hepatitis delta virus ribozyme. *Nature chemical biology*, **1**, 45-52.
127. Cerrone-Szakal, A.L., Siegfried, N.A. and Bevilacqua, P.C. (2008) Mechanistic characterization of the HDV genomic ribozyme: solvent isotope effects and proton inventories in the absence of divalent metal ions support C75 as the general acid. *J Am Chem Soc*, **130**, 14504-14520.

Chapter 2:

A Rugged Free Energy Landscape Separates Multiple Functional RNA Folds Throughout Denaturation[b]

2.1 Introduction

Large fractions of all eukaryotic genomes are transcribed into biologically functional, non-protein coding RNAs (1), profoundly impacting our understanding of cellular function and the application of biotechnology (2). Folding of RNA is generally characterized by a rugged free energy landscape, challenging our understanding of how a given RNA sequence reaches its functional tertiary structure (3-8). Recent single molecule fluorescence resonance energy transfer (smFRET) approaches have provided unprecedented insight into the RNA folding problem by uncovering highly heterogeneous folding kinetics in a number of biologically relevant RNAs (3,7-18). Arguably the most striking example of heterogeneity comes from experiments using the hairpin ribozyme (3,10-12,17,19), a small self-cleaving RNA.

Small self-cleaving RNAs involved in the rolling-circle replication of certain plant, fungus and animal virus satellites are favored model systems to study RNA folding-function relationships as their catalytic function can serve as a practical reporter of a biologically relevant fold (20,21). Various forms of the hairpin ribozyme derived from the tobacco ringspot virus satellite RNA have provided a wealth of structural and biochemical information on how folding impacts the biological activity of RNA (3,10-12,17,19,22-28). The hairpin ribozyme catalyzes site-specific and reversible backbone

[b] Adapted from Ditzler MA, Rueda D, Mo J, Håkansson K, Walter NG. A Rugged Free Energy Landscape Separates Multiple Functional RNA Folds Throughout Denaturation. *Nucleic Acids Res.* **36 (**2008) 7088-99. FRET measurements were made by Mark Ditzler and David Rueda, footprinting and refolding assays were done by Mark Ditzler, and Mass Specta were measured by Jingjie Mo.

self-cleavage, following the dynamic formation of a docked structure in which the internal loops of two adjacent domains form well-understood tertiary contacts (Figure 2.1a) (25,26,28). Ill-understood, however, is the underlying nature of the consistently observed (3,10-12,17,19) multiple sub-populations with dramatically different folding kinetics that result in profoundly biphasic and incomplete catalytic activity (Figure 2.1b) (3,11). Given our increasing appreciation of RNA's role in biology and biotechnology and the ubiquity of heterogeneities in RNA folding and function, it is critical to ask whether the observed heterogeneities are intrinsic or not.

Here, we use smFRET, gel electrophoresis, footprinting, and mass spectrometry analyses that, taken together, demonstrate that the hairpin ribozyme's observed heterogeneity reflects a deeply furrowed folding landscape that is intrinsic to the RNA. We find that commitment to one of several functional folds occurs unexpectedly high on the RNA folding free energy landscape, resulting in partially irreversible folding. Ensemble techniques cannot easily differentiate between rapidly interchanging RNA conformers and those that are separated by deeply dividing energy barriers. The intrinsic nature of such deeply dividing energy barriers within a population of functional RNAs for even a relatively small tertiary RNA fold thus has important ramifications for the interpretation of ensemble-averaged RNA structure and function data. Finally, our observations suggest that it may not be necessary, or even possible, to completely avoid folding heterogeneity in the biological evolution and human design of functional RNAs**.**

Materials and Methods

RNA Preparation

Synthetic RNA oligonucleotides were purchased from the Howard Hughes Medical Institute Biopolymer Keck Foundation Biotechnology Resource Laboratory (Yale University, New Haven, CT), deprotected and purified by 20% (w/v) denaturing polyacrylamide gel electrophoresis (D-PAGE), in 8 M urea, 2 mM EDTA, and 89 mM Tris-Borate and C8-reverse-phase HPLC chromatography, as described previously (21,29).

Transcribed RNA was generated in vitro from partially double stranded DNA containing a T7 promoter incubated in buffer containing 120 mM HEPES-KOH pH 7.5,

30 mM $MgCl_2$, 2 mM Spermidine, 40 mM DTT, 0.01% triton x-100 (v/v) with 4 mM each of NTPs, T7 RNA polymerase, and pyrophosphatase for 3 h at 37° C. Full length transcripts were isolated using denaturing PAGE (20% acrylamide, 8 M Urea). RNA concentrations were calculated from their absorption at 260 nm correcting for absorbance by attached fluorophores when applicable. The RNA sequences are strand RzA (5'-AAA UAG AGA AGC GAA CCA GAG AAA CAC ACG CC**A AA**-3'), where the transcribed RzA contains one additional G on the 5' end and the underlined G (G8) was mutated to an A (A8) to inactivate self-cleavage during EMSA of fully transcribed ribozyme. In strand RzB (5'-**AU AUA UUU** GGC GUG GUA CAU UAC CUG GUA CCC CCU CGC AGU CCU AUU U-3') a 2' O-methyl modification is present at A-1 (underlined) for the non-cleavable construct used in most of our assays. The sequence in bold is absent from the synthetic construct prepared for mass spectrometry.

Cleavage Assays

Single-turnover ensemble cleavage assays were carried out in 50 mM Tris-HCl (pH 7.5), 12 mM $MgCl_2$, over a 26 h time course at 25°C. Ribozyme (final concentration 100 nM strand RzA with a domain terminal donor-acceptor (Cy3-Cy5) fluorophore pair and <10 nM 3'-^{32}P radiolabeled strand RzB with a 5'-biotin) was pre-annealed by heating to 70°C for 2 min and slow cooling over 5 min to room temperature. After pre-incubation for 15 min at 25°C, 1/10th volume of 10x standard buffer was added to initiate the reaction. 2-μL reaction aliquots were taken at appropriate time intervals and quenched with 13 μL stop solution (80% formamide, 0.025% xylene-cyanol, 0.025% bromophenol blue, and 10 mM EDTA). The 3' cleavage product was separated from uncleaved RzB by 20% (w/v) D-PAGE, in 8 M urea, 2 mM EDTA, and 89 mM Tris-Borate. The extent of cleavage was quantified, normalized to the sum of the substrate and product bands, and averaged over three separate cleavage assays using a PhosphorImager Storm 840 with ImageQuant software (Molecular Dynamics). Product formation was fit to the double-exponential first-order rate equation $y(t) = y0 + A1(1-e^{-kobs1*t}) + A2(1-e^{-kobs2*t})$, where A1 + A2 is the final extent of cleavage and k_{obs1} and k_{obs2} are the rate constants.

EMSA

The two strands of the hairpin ribozyme were heat annealed in native buffer containing 50 mM Tris-Acetate (pH 7.5) and either 12 mM Mg-Acetate or 50 mM Na-Acetate, and

10% (v/v) glycerol. Annealed samples were loaded onto a 10% polyacrylamide (19:1 acrylamide:bisacrylamide ratio) gel with 50 mM Tris-Acetate (pH 7.5) and 12 mM Mg-Acetate. For the comparison of cleavable and non-cleavable material the ribozyme was annealed in 50 mM Na-Acetate, and 1 mM Na-EDTA to prevent cleavage prior to electrophoresis. Non-denaturing gels were run at 20 V/cm at 4°C for 14-16 h. For FRET gels 1 μM fluorescently labeled strand RzA was annealed in the presence of 2-3 fold excess of unlabeled RzB strand. For the radioactive gel in Fig. 1 RzB was 3'-^{32}P radiolabled using pCp and T4 RNA ligase and used in trace amount for visualization in EMSA gels prepared with 10 nM labeled RzB and 500 nM unlabeled RzA. For radioactive gels of fully transcribed ribozyme as in Fig. 3 both RzA and strand RzB were 5'-^{32}P-labeled using ATP and T4 polynucleotide kinase and used in trace amounts for visualization in EMSA gels prepared with 6 μM of RzA and RzB. Non-denaturing FRET gels were imaged using FluorImager SI fluorescence scanner and ImageQuant software (Molecular Dynamics). A laser excites fluorescein at 488 nm, and the gels are scanned for fluorescence emission using a photomultiplier tube with either a 530 nm band-pass (for the donor fluorescein) or a 610 nm long-pass filter (for the acceptor tetramethylrhodamine). With the readout of donor defined as green and that of acceptor as red, the corresponding color images were superimposed using Photoshop 5.5 as described previously (5). Radioactive gels were imaged with either PhosphorImager screens or autoradiography. Quantification of native radioactive gel distributions was achieved by fitting to two Lorentzian distributions as described (6).

D-PAGE Refolding

Strands RzA and RzB were annealed at a concentration of 5.7 μM in annealing buffer containing 50 mM Tris-HCl (pH 7.5), 50 mM Na-Acetate, and 5 mM Na-EDTA, with trace amounts of 5'-^{32}P radiolabeled strands RzA and RzB. The annealing solution was heated to 70°C for 2 min before cooling to room temperature over 15 min. Annealed samples were then subjected to EMSA and imaged through autoradiography. The samples were eluted by diffusion into a volume approximately three times larger than the excised band containing 50 mM Na-EDTA (pH 8.0). The sample was then precipitated by adding 1/10th vol of 3 M Na-acetate (pH 6.5) and another 2.5 vol of 100% cold ethanol followed by centrifugation. The dried samples were resuspended in buffer

containing 50 mM Tris-Acetate (pH 7.5), 50 mM Na-Acetate, 1 mM Na-EDTA, 10% (v/v) glycerol, and subjected to 20% (w/v) D-PAGE, in 8 M urea, 2 mM EDTA, and 89 mM Tris-Borate for 3 h at ~55°C and imaged by autoradiography. The four resulting samples were eluted by diffusion into a volume of water approximately three times larger than the excised band. The sample was then precipitated by adding 1/10th vol of 3 M Na-acetate (pH 5.2) and another 2.5 volume of 100% cold ethanol, followed by centrifugation. The samples were resuspended and annealed at 70 °C in 50 mM Tris-HCl (pH 7.5), 50 mM Na-Acetate and 1 mM Na-EDTA. These samples were again subjected to EMSA, then imaged and quantified using a PhosphorImager Storm 840 with ImageQuant software.

tr-FRET

Separated material was eluted from the EMSA gel through diffusion into a standard buffer containing 12 mM $MgCl_2$ 50 mM Tris-HCl (pH 7.5) and then further exchanged into the same standard buffer using Nap-10^{TM} SephadexTM columns. A frequency-doubled Nd:YVO4 laser (Spectra-Physics Millenia Xs-P, operated at 9.0 W) pumped a frequency-doubled, mode-locked Ti:sapphire laser (Spectra-Physics Tsunami, operated at 1 W) that excited fluorescein at 490 nm with 2 ps width pulses, picked down to 4 MHz. Isotropic emission was detected at 520 nm (10 nm band-pass interference filter) in 4,096 sampling channels, with a time increment of 12 ps/channel, up to >40,000 peak counts, as previously described (30). Fluorescence decays were collected in the presence and absence of the acceptor fluorophore. The effect of the acceptor on the decay of fluorescein emission in the doubly labeled complex was then used to determine a distance distribution between the two fluorophores as previously described (30). A Förster distance of 55 Å and an orientation factor $\kappa^2 = 2/3$ were used to determine distances.

sm-FRET

For single molecule measurements of the unseparated ribozyme the RzA and RzB strands were annealed at a concentration of 500 nM. The annealing buffer containing 50 mM Tris-HCl (pH 7.5) 100 mM NaCl, and 1% 2-mercaptoethanol. The annealing solution was heated to 70°C for 2 min before cooling to room temperature over 15 min. Separated material was eluted from the EMSA gel into a buffer containing 50 mM Tris-HCl (pH 7.5), and 12 mM $MgCl_2$. Ribozyme samples were diluted to a concentration of ~25 pM

and bound to a streptavidin-coated quartz slide utilizing the biotin-streptavidin interaction as previously described (3,10,11). Fluorescence signals from donor and acceptor fluorophores were measured with either 1-s time resolution for non-cleavable EMSA separated ribozyme or 100 ms for cleavable multi-turnover ribozyme, using a total internal reflection fluorescence microscope as described (3,10,11). Single molecule measurements for non-cleavable ribozyme were carried out in 50 mM Tris-HCl (pH 7.5), 12 mM $MgCl_2$, at 22^{o}C, whereas multiple-turnover measurements of the cleavable ribozyme were carried out in 12 mM, 100 mM, or 500 mM $MgCl_2$ with an excess concentration of 1 nM substrate (the 500 mM $MgCl_2$ measurements are shown in Figure 2.1c). An oxygen scavenger system (10% glucose, 2% β-mercaptoethanol, 750 μg/ml glucose oxidase, and 90 mg/ml catalase) was used to slow down fluorophore photobleaching. Dwell times for the docked and undocked states were analyzed and fit as described previously (2, 7, 8). For the multiple-turnover measurements with cleavable ribozyme 660 dwell times in the docked state, and 682 dwell times in the undocked state from 123 smFRET time trajectories measured at 500 mM $MgCl_2$ were used for the fits. We note that increasing ionic strength has previously been shown to accelerate docking while having essentially no impact on undocking kinetics (7), thus increasing the number of events observed for each molecule. The rate constants and corresponding amplitudes for docking and undocking of the non-cleavable ribozyme used for the EMSA separation were determined by performing a global fit on all three normalized data sets (T, B, and unseparated), holding the rate constants constant across all data sets and allowing all other parameters to vary independently throughout the fit. A total of 829 dwell times in the docked state and 831 dwell times in the undocked state from 295 single-molecule FRET time trajectories were used in the global fit. In contrast to previous work, we observe two docking rate constants, however the two docking rates appear to be independent of the undocking rates and only differ by approximately one order of magnitude (or less in the case of the non-cleavable sample). We therefore determined the time-window corrected fraction of ribozymes undocking with each of the undocking rate constants separately for each of the docking rates, as described previously (2), and used the weighted average of those results to determine the final fractions given in Figure 2.4d, Figure 2.1c, and Figure 2.3. Please note that, after time window correction, the

molecule fractions are no longer identical to the amplitudes from the exponential fits due to the statistical underrepresentation of slow events, as previously described (2).

Footprinting

Tb^{3+} mediated footprinting (31) was used to partially degrade ribozyme samples separated by EMSA. ^{32}P-labeled ribozyme was separated by the EMSA and located using autoradiography. The material was cut out of the gel and eluted with the addition of buffer containing 50 mM Tris-HCl (pH 7.5), 10 mM Na-EDTA, 50 mM NaCl in a volume approximately three times larger than the excised band. The buffer was then exchanged for 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, using a Nap-10™ Sephadex™ column. Samples were concentrated using Centricon centrifugal filters at 4°C. This preparation does not result in the redistribution of either T or B species as judged by EMSA. The radiolabeled RNA (~150,000 cpm) was next incubated with RNase or Tb^{3+} to achieve partial degradation. Tb^{3+} footprinting was carried out by adding 0.1, 1 and 5 mM Tb^{3+} to the RNA samples and incubating at 25 °C for 1 h. The reaction was quenched with 50 mM EDTA in 20% (v/v) formamide, and immediately analyzed by 20% (w/v) D-PAGE, in 8 M urea, 2 mM EDTA, and 89 mM Tris-Borate, alongside T1 and alkaline ladders. The results from the reactions with 5 mM Tb^{3+} were used for Π analysis (31). For RNase V1 footprinting, the ribozyme was prepared as in Tb^{3+} footprinting but incubated instead with RNase V1 at room temperature for 15 min. Mg^{2+} was added to the V1 reaction along with the RNase to a final concentration of 100 μM. A range of RNase concentrations were used to determine optimal concentration for partial degradation. Footprinting gels were quantified using a PhosphorImager Storm 840 with ImageQuant software. The concentration of material was determined by boxing with ImageQuant. In order to normalize cleavage values and account for differences in background cleavage Π values were determined using the equation as described (10):

$$\Pi = \frac{\left(\frac{\text{band intensity at nucleotide x}}{\sum_i \text{band intensity at nucleotide i}} \right)_{5\text{mM[Tb3+]}}}{\left(\frac{\text{band intensity at nucleotide x}}{\sum_i \text{band intensity at nucleotide i}} \right)_{0\text{mM[Tb3+]}}}$$

where x is the analyzed nucleotide position and 0 mM [Tb^{3+}] is a control reaction incubated in the same manner as the Tb^{3+} containing reaction, except without the addition of Tb^{3+}. The ratio between Π values of individual nucleotides in the T and B species was determined for comparison. Final Π ratios are the result of at least two independent footprinting reactions. Π value ratios >2 or < 0.5 are considered to indicate a significant difference between the two species.

Mass Spectrometry

For mass spectrometry a construct in which helix four is truncated was annealed at a concentration of ~100 μM and separated by EMSA. Separated RNA was located using UV shadowing and eluted from the gel into a buffer containing 10 mM Na-EDTA, 50 mM Tris-HCl (pH 7.5). The RNA was then concentrated using Centricon Plus-20™ and further separated for 3 h using 20% (w/v) D-PAGE, in 8 M urea, 2 mM EDTA, and 89 mM Tris-Borate. The samples were again imaged through UV-shadowing and eluted into 50 mM Tris HCl (pH 7.5), 1 mM Na-EDTA. The RNA was desalted three times and concentrated >100-fold using Centricon Plus-20™ devices according to the manufacturer's instructions, followed by further desalting using Centricon YM-3™ filters according to the manufacturer's instructions. If significant salt adduction was still observed in mass spectra, additional desalting was performed by ethanol precipitation (once or twice; protocol modified from (32)). For this purpose the RNA was dissolved in water and 1/3 vol of 10 M ammonium acetate (pH 8) and another 2.5 vol of 100% cold ethanol (stored at -80 ºC until right before use) was added. The mixture was vortexed for a few seconds and stored at -80°C for 3 h followed by centrifugation at 12,500 rpm for 15 min. The supernatant was decanted and 400 μL cold 70% (v/v) ethanol was added to the precipitate, followed by another incubation at -80°C for 2 h. Centrifugation was once again performed at 12,500 rpm for 15 min, the supernatant was decanted, and the precipitate was dried down and re-suspended in 50 mM ammonium acetate with 25% (v/v) methanol. Desalting of the ribozyme does not result in redistribution of the B species as determined by EMSA. Reduction of salt adduction is essential since the presence of multiple salt adducts complicates spectra and spreads the total signal over multiple peaks, thereby significantly reducing signal-to-noise ratios.

Samples were electrosprayed at 50 μL/h through an external Apollo II ion source equipped with dual ion funnels in negative ion mode and analyzed by an actively shielded 7 T quadrupole-Fourier transform ion cyclotron resonance mass spectrometer (Bruker Daltonics, Billerica, MA) as described previously (33,34). All mass spectra were acquired with XMASS (version 6.1, Bruker Daltonics) in broadband mode from m/z 500 to m/z 2,500 with 512 k data points and summed over 32 or 64 scans. Data processing with one zero fill but no apodization was performed with the MIDAS analysis software (35) prior to fast Fourier transformation. Due to the high molecular weight of the RNAs, average rather than monoisotopic mass was used for spectral interpretation. Mass-to-charge (m/z) values for each observed charge state were calculated from a weighted average of all observed isotopic peaks. Masses were calculated from a weighted average of all charge states. Spectra were externally calibrated using a calibration standard (G2421A; Agilent Technologies, Palo Alto, CA) and a two-term frequency-to-mass calibration equation (36).

DNA-Assisted Strand Displacement

Following EMSA separation, the ribozyme was eluted from the EMSA gel into a buffer containing 50 mM Tris-HCl (pH 7.5), and 12 mM $MgCl_2$. The separated material was then incubated for 2 h at 37^{o}C with >10 fold excess of a DNA oligonucleotide fully complementary to one of the RNA strands. Strand invasion by the DNA oligonucleotide results in base pairing between the DNA and the complementary RNA strand (RzA or RzB depending on the DNA oligonucleotide used), efficiently separating the two ribozyme stands. This process results in one of the stands being incorporated into a fully base-paired DNA-RNA hybrid, while the other strand is converted into single-stranded form. Excess DNA oligonucleotide was then degraded by incubating with DNase RQ1 for 1 h. Finally, the ribozyme was reformed by replacing the RNA strand sequestered in the DNA-RNA hybrid with fresh strand.

Thermal Denaturation

A volume of 300 μl of annealed ribozyme at 1 μM concentration in standard buffer containing 50 mM Tris-HCl (pH 7.5), and 12 mM $MgCl_2$ was degassed followed by thermal denaturation. The sample was heated at 0.2^{o}C/s from 35-99^{o}C and UV absorbance at 260 nm was monitored using Beckman DU-640B UV–Vis

spectrophotometer. Melting temperatures were determined from two separate measurements by fitting the first derivative of the melting curves to two Gaussian distributions.

2.2 Results

Both synthetic and transcribed hairpin ribozymes harbor multiple, catalytically active molecular species

To examine the persistent folding heterogeneity of the hairpin ribozyme, we explored an ensemble electrophoretic mobility shift assay (EMSA) to isolate molecular sub-populations based on their average hydrodynamic radii. During EMSA a catalytically inactivated two-way junction (2WJ) construct (Figure 2.1a) migrates in two bands of distinct interdomain docking equilibria (Figure 2.1d). This heterogeneous mobility is observed over a wide range of annealing protocols (Figure 2.2), in the presence or absence of a domain terminal fluorophore pair, and independent of whether catalysis is blocked by a synthetic 2'-O-methyl modification of the cleavage site A-1 or a ribozyme G8A mutation in fully *in vitro* transcribed material (Figures 2.1d, 2.2, 2.4a, 2.5c). The corresponding cleavable 2WJ ribozyme cleaves to 77% completion with biphasic rate constants of 0.08 min^{-1} and 0.01 min^{-1} (Figure 2.1b), which have been linked to more and less stably docking sub-populations, respectively (3,11). During EMSA the cleavable 2WJ ribozyme cleaves to >60% completion (Figure 2.1d), thereby demonstrating native folding during EMSA and necessitating inactivating modifications for the preservation, isolation and further characterization of the two bands. Additionally, we examined smFRET trajectories of a trans-cleaving t-2WJ construct where the substrate strand is separated from the remainder of the RzB strand, such that the products generated upon backbone cleavage rapidly dissociate, giving rise to a distinct FRET signal of the substrate-free ribozyme after catalysis (Figure 2.1c). We find that smFRET trajectories of these t-2WJ ribozymes with distinct undocking behaviors repeatedly show transitions to the self-cleaved, intermediate FRET state (Figure 2.1c). We therefore infer that: (i) Multiple, catalytically active molecular species exist within both synthetic and transcribed hairpin ribozyme preparations (source independence), and (ii) our catalytically inactivating modifications allow us to separate molecular sub-populations

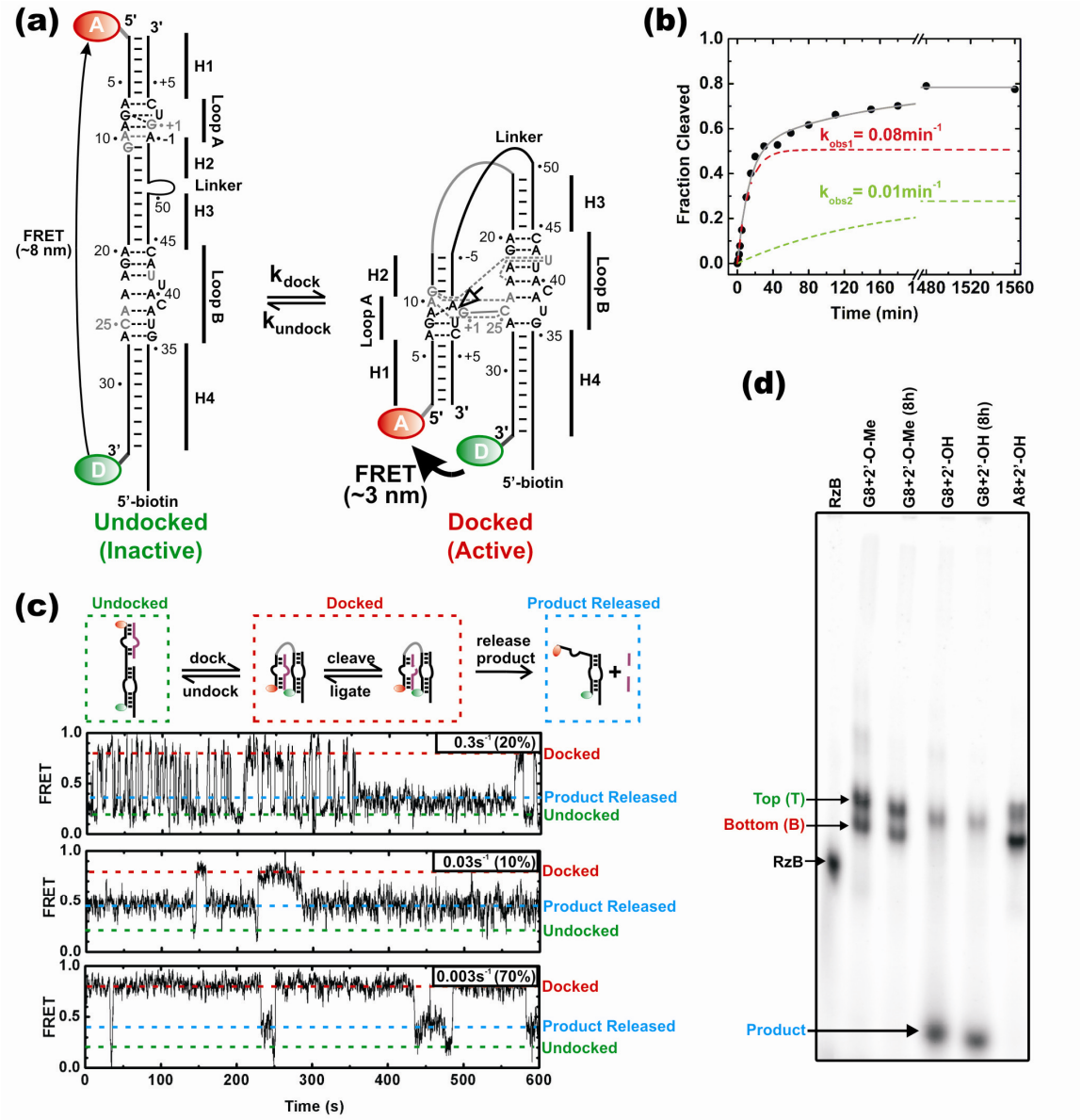

(a)
A
D
5'
3'
H1
H2
H3
H4
Linker
Loop A
Loop B
FRET
(~8 nm)
5'-biotin
Undocked
(Inactive)
k_{dock}
k_{undock}
FRET
(~3 nm)
Docked
(Active)
(b)
Fraction Cleaved
k_{obs1} = 0.08min^{-1}
k_{obs2} = 0.01min^{-1}
Time (min)
(c)
Undocked
Docked
Product Released
dock
undock
cleave
ligate
release
product
FRET
0.3s^{-1} (20%)
0.03s^{-1} (10%)
0.003s^{-1} (70%)
Time (s)
(d)
RzB
G8+2'-O-Me
G8+2'-O-Me (8h)
G8+2'-OH
G8+2'-OH (8h)
A8+2'-OH
Top (T)
Bottom (B)
RzB
Product

Figure 2.1: Heterogeneity in folding and function of the hairpin ribozyme. (*a*) Secondary structure of the 2WJ hairpin ribozyme, composed of the RzA and RzB strands. The docked and undocked conformations of the ribozyme are shown with canonical and non-canonical base pairs indicated by solid and dashed lines, respectively. Nucleotides forming interdomain hydrogen bonds are shown in gray, the cleavage site is indicated by an open arrow. The ribozyme was labeled either with a domain terminal donor(D)-acceptor(A) fluorophore pair and a 5'-biotin, and/or with a 5'- or 3'-end ^{32}P label. (*b*) Single-turnover cleavage time course of the 3'-^{32}P labeled ribozyme with a domain terminal donor-acceptor fluorophore pair and a 5'-biotin, as monitored by D-PAGE, was used to determine the indicated cleavage rate constants for the two phases of the reaction (12mM MgCl2, pH 7.5, 25C). The contribution of each phase is shown as a dashed line. (*c*) Heterogeneous undocking kinetics of the catalytically active, trans-cleaving (no Linker), t-2WJ hairpin ribozyme under multiple-turnover conditions. The docked, undocked and product released states are indicated by red, green, and blue dashed lines, respectively. Three single molecule time trajectories demonstrating catalytic proficiency of distinct sub-populations (undocking rate constants and the fraction of molecules undocking with this rate constant are given in the upper right corner of each trajectory; see also Figure 2.3). (*d*) EMSA separation of the 3'-^{32}P labeled ribozyme using a non-cleavable (2'-O-methyl) or cleavable (2'-OH) synthetic RzB strand, and inactive (transcribed with A8) or active (synthetic with fluorophores and G8) RzA strand, with and without an 8 h preincubation in native buffer. Two bands, termed T and B, are observed for ribozyme containing non-cleavable RzB (45% B) or inactive RzA (66% B). Cleavable ribozyme shows cleavage and subsequent product release for ~60% of the material during electrophoresis (with <10% B band remaining). Preincubation of the ribozyme results in additional cleavage (>70%), consistent with the final cleavage extent in solution.

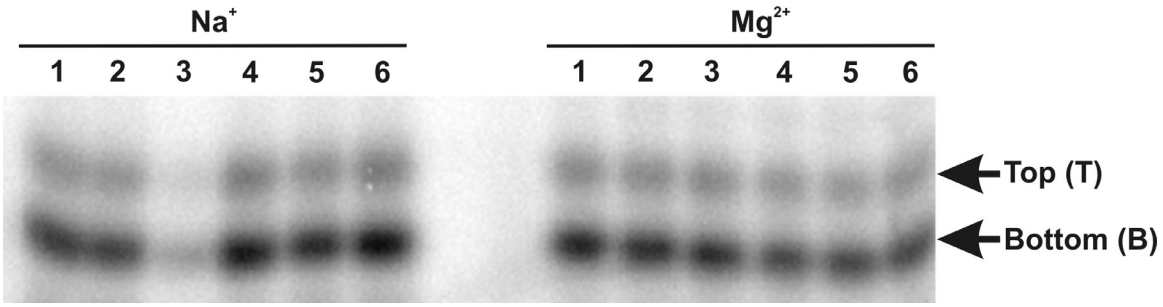


Figure 2.2: Multiple annealing conditions resulting in similar EMSA distributions. In the presence of monovalents alone (50 mM NaCl) or in the presence of divalents (12 mM $MgCl_2$) ^{32}P-labeled ribozyme (fully synthetic 2'-O-methyl construct used for mass spec analysis) was subjected to the following annealing protocols: (1) 25°C for 20 min; (2) 70°C 2 min, 25°C 20 min; (3) 90°C 2 min, 25°C 20 min; (4) 90°C 2 min, 70°C 2 min, 65°C 2 min, 25°C 20 min; (5) 90°C 2 min, 25°C 10 min repeated 3 times, then 70°C 2 min, 25°C 10 min repeated 2 times; (6) 90°C 2 min, 55°C 30 min, 25°C 20 min. The distribution between T and B species was then analyzed by EMSA.

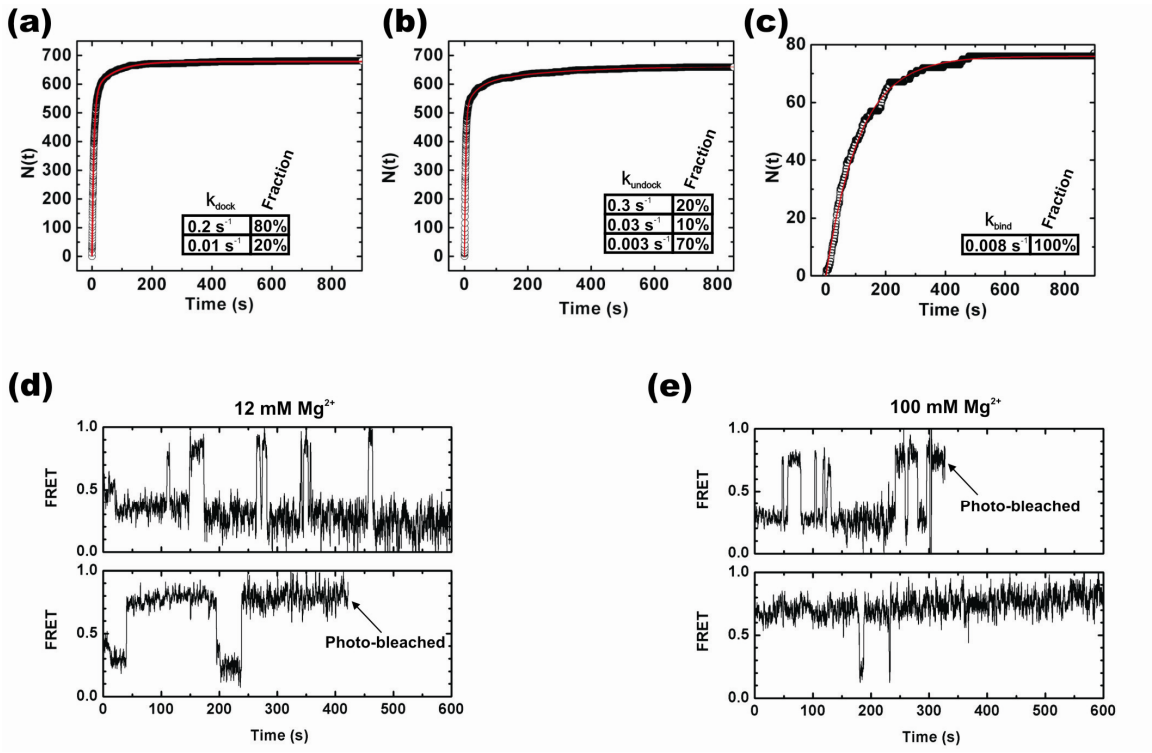


Figure 2.3: Folding heterogeneity in the catalytically active trans-cleaving t-2WJ hairpin ribozyme. Cumulative number of dwell times (N) shorter than time t in the (*a*) undocked and (*b*) docked states, and (*c*) in the substrate-free state (after product release and before renewed substrate binding) are plotted for the t-2WJ ribozyme under multiple-turnover conditions at high ionic strength (500 mM $MgCl_2$). Solid red lines indicate the (multi-)exponential fits used to determine the docking (a), undocking (*b*), and substrate binding rate constants (*c*) with their time-window corrected fractions reported in the table insets. (*d* and *e*) Examples of smFRET trajectories displaying distinct undocking behaviors in the presence of 12 mM and 100 mM $MgCl_2$, respectively.

for further study that reproduce the heterogeneous behavior of the catalytically active ribozyme.

EMSA allows for isolation of molecular species that can be further characterized by ensemble and single molecule FRET

To further examine the underlying molecular sub-populations, we eluted the top (T) and bottom (B) bands of donor-acceptor (fluorescein-tetramethylrhodamine) labeled hairpin ribozyme separately from the EMSA gel (Figure 2.4a). The isolated T and B species were subjected to time-resolved FRET (trFRET) analysis in native buffer to probe their ensemble-averaged donor-acceptor distance distributions. This analysis shows that both species access the docked conformation, but to different extents (Figure 2.4b): The more slowly migrating T species is dominated by the extended undocked fold, whereas the faster migrating B species is governed by the more compact docked fold (Figure 2.4b). These findings reveal a surprising longevity of the observed folding heterogeneity, throughout the isolation process.

Next, we employed smFRET using doubly (Cy3 and Cy5) labeled, EMSA purified ribozyme species, which demonstrates that the T species is enriched in the more rapidly undocking ribozyme sub-populations, while the B species is enriched in the more slowly undocking sub-populations (Figure 2.4c). Quantification of the relative ribozyme fractions with distinct (un)docking rate constants in the T and B species confirms this qualitative assessment (Figure 2.4d). Notably, smFRET resolves several molecular sub-populations in both EMSA species, highlighting the enhanced resolution of the single molecule approach; complementarily, ensemble EMSA allows us to physically isolate and individually characterize two kinetically distinct, enriched ribozyme species.

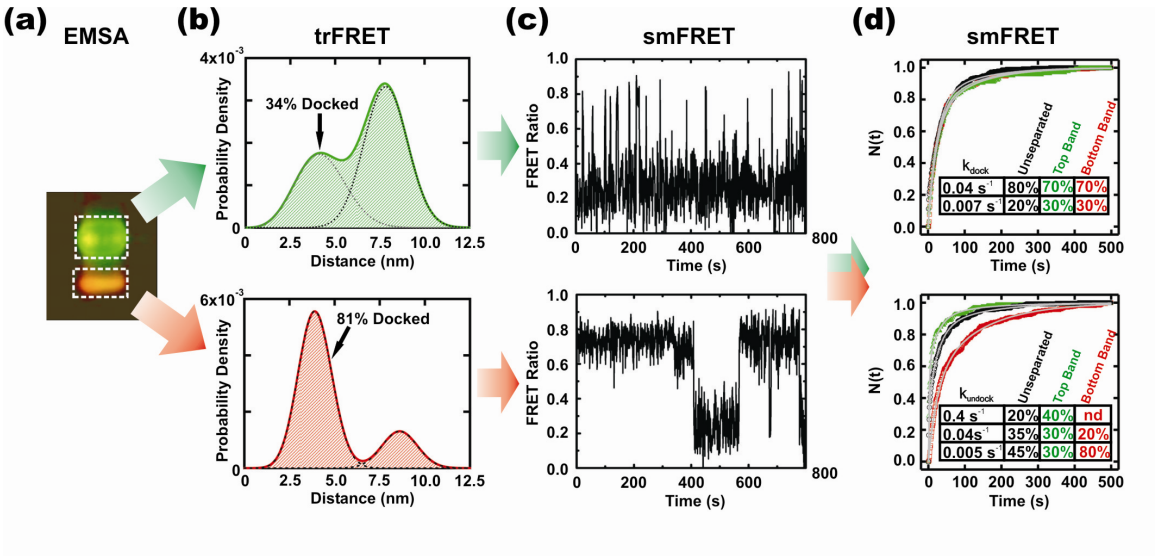


Figure 2.4. FRET characterization of EMSA separated ribozyme. (*a*) EMSA-based separation of fluorophore labeled ribozyme showing differences in gel mobility and FRET (pseudo-color, green donor, red acceptor). RNA from the boxed areas was isolated as T and B species for further analysis. (*b*) Probability distribution of inter-fluorophore distance in the T and B species as derived by trFRET analysis, with fraction of docked material indicated. (*c*) Exemplary time trajectories characteristic of the sub-populations that are most enriched in the isolated T and B species. (*d*) The T and B species exhibit significantly different undocking kinetics in single molecule dwell-time analyses. The normalized fraction of dwell times (N) shorter than time (t) in the undocked (top) and docked states (bottom) are plotted for the isolated T (green) and B (red) species as well as the unseparated ribozyme (black). Solid gray lines indicate the multi-exponential global fits used to determine the (un)docking rate constants; the time-window corrected fraction of molecules (un)docking with the given rate constant is indicated in the table insets.

The observed molecular heterogeneity remains associated with the separated composite RNA strands of each EMSA species

To further characterize the thermodynamic stability of the hairpin ribozyme's molecular sub-populations, we isolated T and B species from *in vitro* transcribed ribozyme by EMSA and then physically separated the two composite strands (RzA and RzB, Figure 2.1a) by denaturing polyacrylamide gel electrophoresis (D-PAGE, which melts all four helices of the ribozyme), followed by ethanol precipitation and re-suspension in native buffer (Figure 2.5a and b). The RzA and RzB strands from the T and B species were then recombined in all four possible permutations, heat annealed, and analyzed by EMSA (Figure 2.5c). We find that this refolding protocol results in the expected redistribution of the annealed T-RzA/T-RzB ribozyme. Strikingly, no significant redistribution of the annealed B-RzA/B-RzB ribozyme occurs, which continues to form the dominantly docked B species. Additionally, the mixed (heterologous) T-RzA/B-RzB and B-RzA/T-RzB ribozymes redistribute only partially. These observations suggest that, after folding into the stably docked B species, both the RzA and RzB strands retain a tendency to form the more active B species throughout an isolation procedure that subjects them to harshly denaturing conditions (including >2 h in 8 M urea at ~55 ^{o}C during D-PAGE). In the presence of an excess of either fresh RzA or RzB, redistribution of both the isolated T and B species can be achieved, via either heating (>50^{o}C for 2 min) (Figure 2.6) or DNA-mediated strand removal using synthetic donor-acceptor labeled ribozyme (Figure 2.7). Interconversion in the presence of excess RzA or RzB is likely trivial in nature as it allows one of the two strands in the ribozyme to be replaced, however it does provide further evidence that both strands contribute to the persistent heterogeneity. Conversion of T to B species via heating alone can be achieved in the absence of excess RzA or RzB, but we do not observe conversion of B to T, consistent with our D-PAGE refolding assay (data not shown). Additionally, we observe retention of folding heterogeneity following incubation at 37 °C for 4.5 h in a DNA mediated interconversion assay (Figure 2.7), addition of a large molar excess of EDTA followed by exchange into Mg^{2+}-free buffer in preparing material for footprinting, and after extensive desalting required for mass spectrometric analysis, as judged by EMSA. Furthermore, we observe that the sub-populations of the catalytically active trans-cleaving t-2WJ ribozyme retain distinct

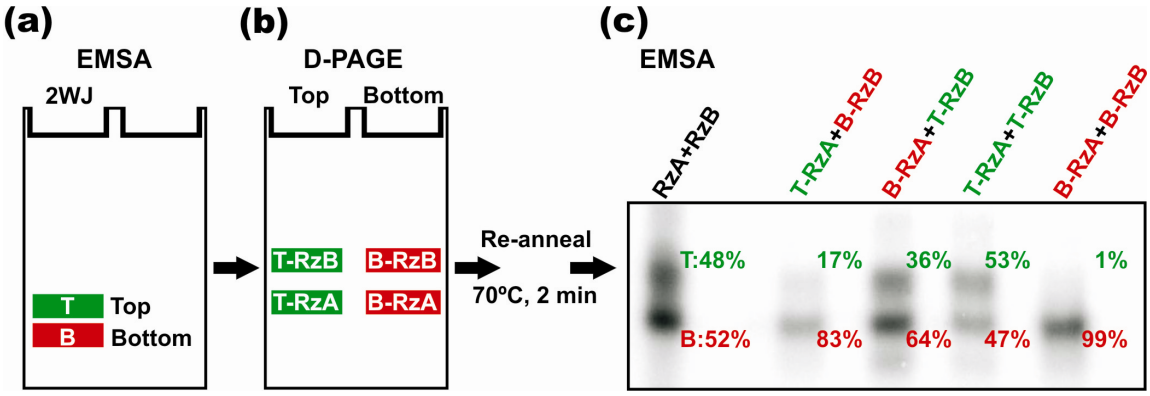


Figure 2.5: Extraordinary stability of molecular sub-populations of the hairpin ribozyme. Following EMSA separation (*a*) of fully *in vitro* transcribed ribozyme into two bands, the T and B species were further separated by D-PAGE (*b*) into their composite RzA and RzB strands. The four samples T-RzA, T-RzB, B-RzA, and B-RzB were then annealed in all four possible combinations to refold the ribozyme, and analyzed by EMSA (*c*). An image of the second EMSA gel is shown with the fraction of T and B species for each lane indicated to the right of the respective band.

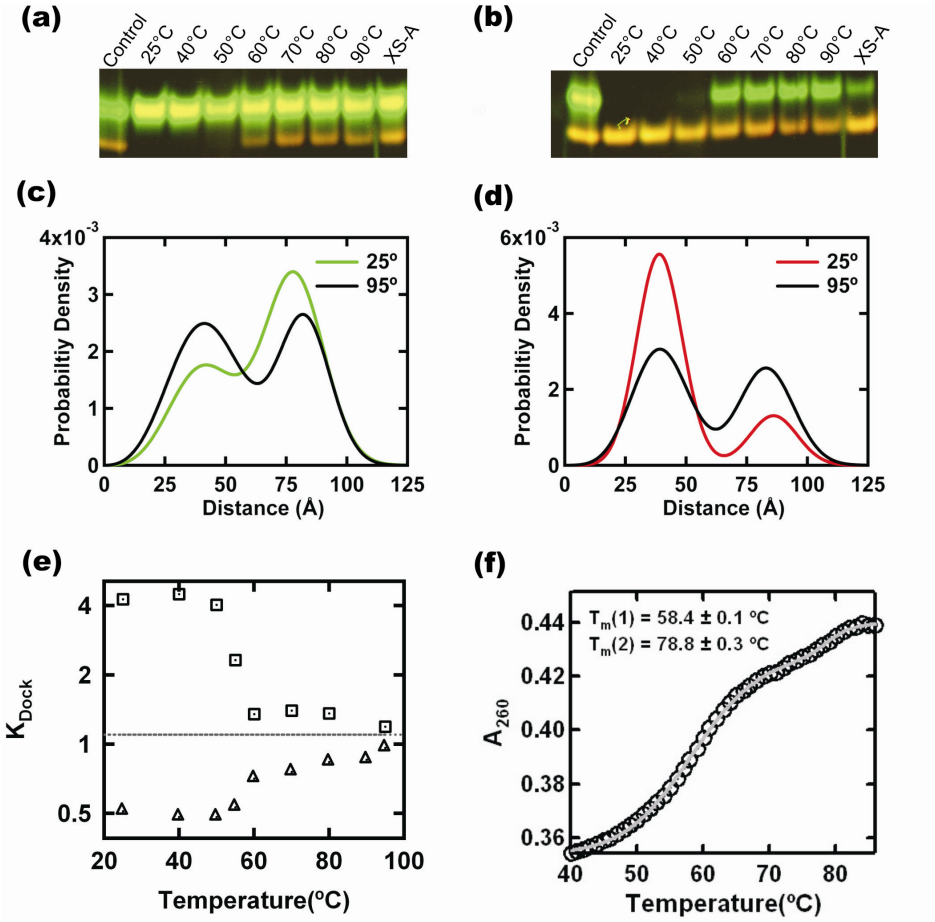


Figure 2.6: Redistribution of the fluorescein and tetramethylrhodamine doubly labeled T and B species through heat-assisted RNA strand replacement. (*a*, *b*) Redistribution of the T and B species in a background of either excess RzA (lane XS-A) or RzB (all other lanes except for control lane) as monitored by EMSA. Isolated T species (*a*) or B species (*b*) was heated to the indicated temperature for 2 min and then cooled to allow for reannealing with the excess RzA (or RzB) strand. An analytical EMSA gel was then run at 4°C and its fluorescence detected using a FluorImager SI fluorescence scanner as described above. (*c*, *d*) Redistribution of the T and B species in the presence of excess RzB as monitored by tr-FRET. Both species are shifted to similar distance distributions after heating to 95°C and then cooling back down to 25°C. The unheated controls (25°C, in color) are shown for reference. (*e*) Heating temperature dependence of the apparent docking equilibrium constant $K_{dockapp}$ = fraction(docked)/fraction(undocked) derived from tr-FRET measurements similar to those described in panels c and d. Two Gaussian distance distributions were fit to the tr-FRET data to determine the fractions of docked (smaller distance) and undocked conformation (larger distance) after heating the T (triangles) and B species (squares) in the presence of excess RzB to the indicated temperature. (*f*) UV-detected melting curve of unseparated hairpin ribozyme with apparent melting temperatures indicated. The UV-detected melting transitions, indicative of loss of secondary structure, coincide with interconversion of the T and B species in the presence of excess RzB (panel e), consistent with the notion that interconversion occurs via strand replacement.

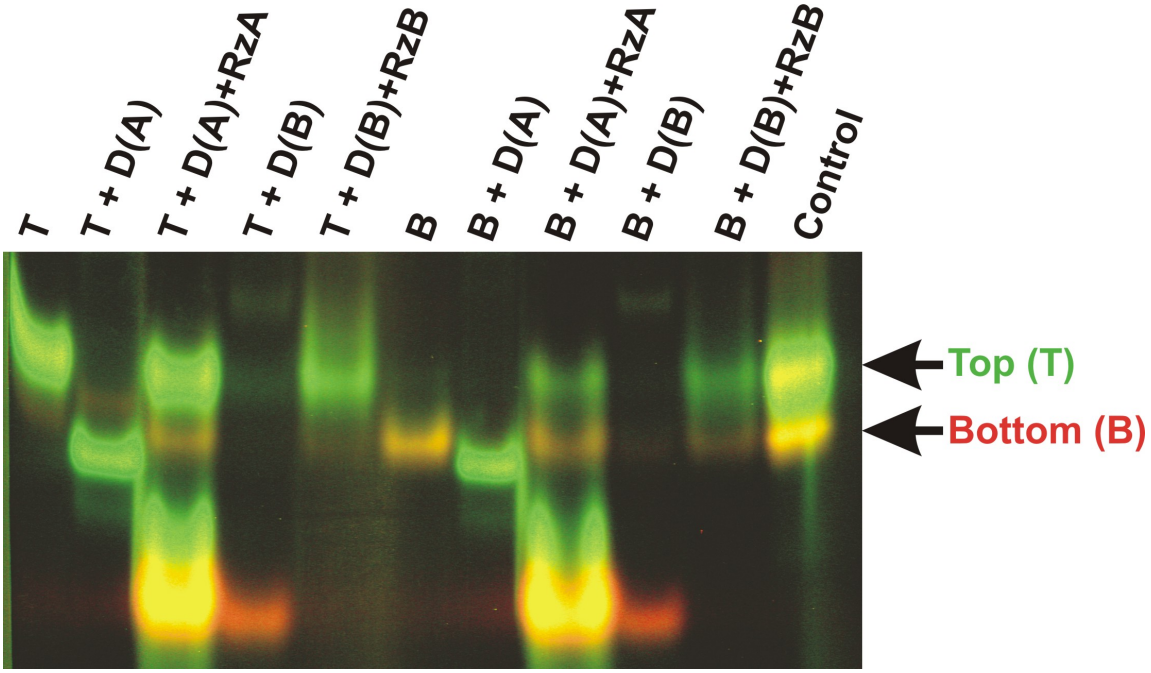


Figure 2.7: Redistribution of the isolated fluorescein and tetramethylrhodamin doubly labeled T and B species as achieved through DNA-assisted strand displacement. DNA oligonucleotides fully complementary to either RzA (lanes with D(A)) or RzB (lanes with D(B)) were used to disrupt the ribozyme's secondary structure by base pairing with their complement through strand invasion. This process results in the complementary RNA strand fully incorporating into a DNA-RNA hybrid, whereas the other RNA strand becomes single-stranded. Addition of either fresh RzA or RzB strand, as indicated for the corresponding lanes, was then used to reanneal the ribozyme and the result analyzed. The EMSA and FluorImager analysis shows that partial redistribution between the T and B species occurs under these conditions, but only in the presence of fresh RzA or RzB strand. Lanes T and B contain the isolated T and B species, respectively, and the Control lane shows the unseparated hairpin ribozyme.

undocking kinetics following product dissociation, which disrupts all inter-strand contacts in domain A (Figure 2.1c). Such a pronounced robustness of the observed molecular heterogeneity ("molecular memory" (3)) is consistent with retention of the molecular sub-populations over a wide range of ionic conditions (Figure 2.3) (10) and upon introduction of various modifications into the ribozyme (11). It is also in agreement with observations of persistent molecular heterogeneity in both less (14) and more complex tertiary folded RNAs (7,8,13).

Footprinting reveals subtle structural differences between the EMSA species

The stability of RNA helices, and the propensity to form alternate base pairing patterns, has previously been invoked as the source of the rugged folding pathways characteristic of RNA (6-8). To probe for even subtle secondary structure differences between the T and B species at single nucleotide resolution, we employed Tb^{3+} (24) and RNase V1 footprinting (37) to monitor single- and double-stranded RNA segments, respectively. To avoid trivial differences in footprinting pattern due to their distinct docking equilibria, we compared the T and B species side-by-side at low ionic strength, i.e., in the absence of docking (10), where heterogeneity is conserved. The Tb^{3+} and RNase V1 footprinting patterns are both strikingly similar for the T and B species, suggesting identical secondary structures (Figure 2.8). This finding is consistent with the paradoxical observation of extensive native structure within a deeply misfolded and partially active conformer of the Tetrahymena group I intron ribozyme, which was proposed to represent a topological isomer of the native state (7,8). For the hairpin ribozyme, however, the presence of topological isomers is inconsistent with our refolding assays, which demonstrate that the persistence of folding heterogeneity is not a consequence of strand entanglement. The most significant, albeit subtle footprinting difference between the hairpin ribozyme's T and B species we find at nucleotide C25 in loop B, which is involved in the critical G+1:C25 docking contact (10,11) (Figure 2.1a). The B species is slightly more sensitive to Tb^{3+}-mediated scission at C25 than the T species (Figure 2.8), suggesting a possible structural basis for the different folding behaviors of the T and B species: a local, persistent perturbation in the loop B structure (23).

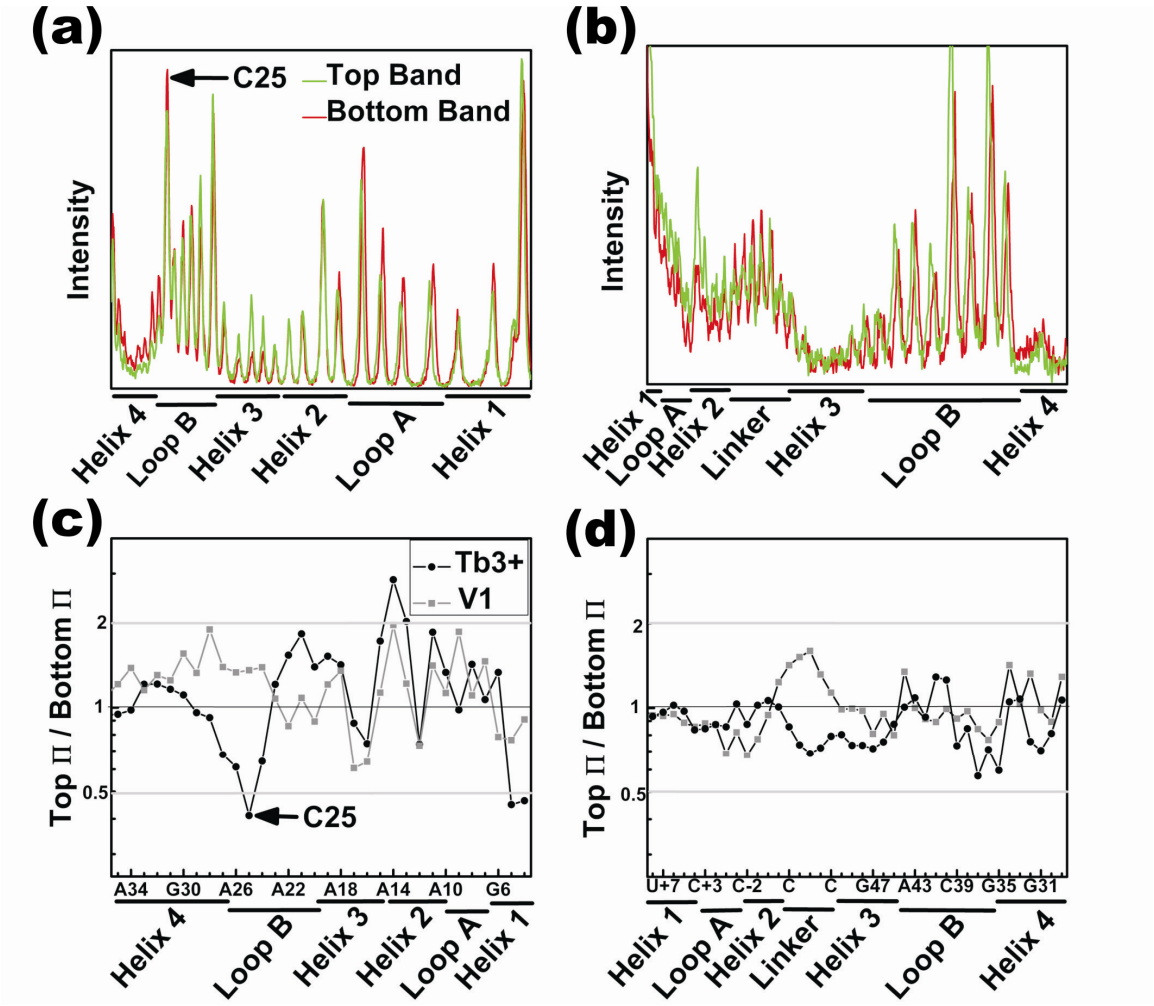


Figure 2.8 Footprinting reveals identical secondary structure in EMSA separated ribozyme. Profiles of the Tb^{3+}-mediated cleavage patterns of ^{32}P radiolabeled strand RzA (*a*) and strand RzB (*b*). The high degree of similarity between the T (top) and B (bottom) species indicates a shared secondary structure and suggests that the two species do not arise from gross misfolding of the ribozyme. Ratio of Π values, which represent averaged and normalized fractions of cleavage relative to background, are given for strands RzA (*c*) and RzB (*d*). Values less than 0.5 (gray line) indicate two-fold greater sensitivity to Tb^{3+} (circles) or RNase V1 (squares) induced cleavage in the B species. Values greater than 2 (gray line) indicate a two-fold greater sensitivity in the T species.

High-resolution mass spectrometry provides no evidence for chemical modification of the RNA

The difficulty to convert the more stably docking B species into the T species raises the possibility that the observed molecular heterogeneity originates from a chemical modification, such as retention of a protection group from chemical synthesis that favors the most stably docked sub-population(s). The fact that we observe T and B species in both chemically synthesized (Figures 2.1d, 2.4a, 2.6a, and 2.7) and *in vitro* transcribed (Figures 2.1d and 2.5) ribozyme preparations, provides evidence against some of these artifactual origins of molecular heterogeneity. To directly probe for even subtle chemical heterogeneity, we subjected the isolated RzA and RzB strands from both the T and B species to high-resolution electrospray ionization (ESI) hybrid quadrupole-Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry (33). For all four RNA samples, we observe the masses predicted from their complete RNA sequence, with less than 1 Da error, providing strong evidence against the notion that a stable adduct or chemical modification causes the observed persistent molecular heterogeneity (Figure 2.9). No additional peaks are observed in the B species, and low abundance peaks observed in the T species are of insufficient abundance and mass to account for the observed heterogeneity (see below). We cannot rule out, however, that the RNA is modified in a mass-neutral fashion, or that an RNA adduct forms that is stable under all solution conditions but dissociates in the gas phase of our mass spectrometry analysis.

Low abundance peaks in the T-RzA and T-RzB samples of Figure 2.9, marked "$", "¥", and "&", represent material that may have been present during EMSA, although backbone cleavage could also occur during the ionization process of mass spectrometry, or more generally at some point following elution from the final denaturing gel. In the T-RzA sample, the low abundance peaks marked "$" and "¥" represent material that is missing adenosine monophosphate on the 5'-end of the RzA strand. Even if this chemical heterogeneity was present during EMSA, the 5'-nucleotide of RzA is at the end of helix 1 and thus remote from the docking loops A and B and the helical junction (Figure 2.1a). In fact, in EMSA and single molecule experiments using fluorophore labeled ribozyme the donor fluorophore is present at this position, therefore in these experiments heterogeneous FRET-detected EMSA mobility can only be observed for ribozyme

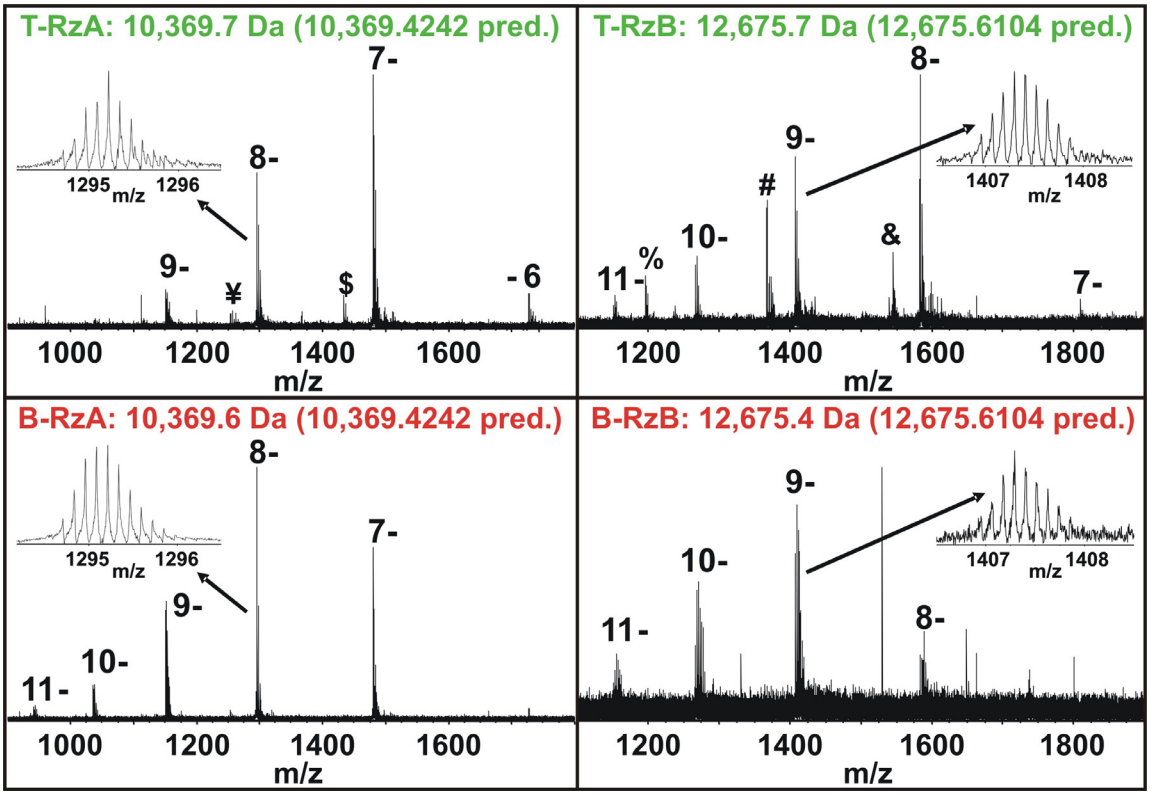


Figure 2.9: FT-ICR mass spec analysis of EMSA separated ribozyme. Negative mode ESI FT-ICR mass spectra show peaks generated from intact molecular ions of both RzA and RzB at different charge states as indicated, as well as additional peaks with relatively low abundance. The insets show zoomed in spectra of the 8- charge state for the RzA strand and the 9- charge state for the RzB strand. The experimental average masses are given above the spectra and the predicted masses are indicated in parentheses. Minor peaks ¥ and $ represent the 8- and 7- charge states, respectively, of a fragment resulting from loss of a single adenosine monophosphate from the 5'-end of T-RzA; minor peak & represents the 8- charge state of a fragment resulting from loss of a single uridine from the 3'-end of T-RzB; minor peaks % and # represent the 8- and 7- charge states, respectively, of the 5'-fragment resulting from backbone cleavage between C-2 and A-1.

containing the full length RzA. In addition, the "$" and "¥" peaks constitute only ~5% of the total signal, so are highly unlikely to induce the observed molecular heterogeneity.

In the T-RzB sample, the low abundance peak marked "&" represents uridine loss from the 3'-end of RzB, also at the end of helix 1. The abundance of this peak also constitutes only ~5% of the total signal. If we assume similar ionization efficiencies between the full length and the one-nucleotide (one-nucleoside) shorter species "$", "¥", and "&", which is expected due to their very similar size and composition, then the maximum fraction of shorter ribozymes within the T species (either in the RzA or RzB strand) that our mass spectrometry is consistent with is 10%. If chemical modifications were the cause for the appearance of the T species, our mass spectrometry analysis would leave 90% of the chemical modifications of the T species unaccounted for. While it is possible that there are additional modifications of sufficiently low abundance that have a large cumulative effect but escape detection by mass spectrometry, we do not favor this interpretation based on the body of contradicting evidence described in the discussion section and throughout the results section.

The two minor peaks labeled "%" and "#" (Figure 2.9) represent the 8- and 7-charge states, respectively, of a significantly smaller RNA fragment than the main peak (smaller by 3,097 Da, i.e., 10 nucleotides). It is the 5'-fragment of backbone cleavage between C-2 and A-1 that must have occurred at some point after EMSA and elution from the final denaturing gel, which otherwise would easily have resolved such a large mass difference. It therefore cannot be responsible for the EMSA mobility differences. The size of the peak appears to represent a larger fraction of the signal than the other minor peaks (although still less than 1/3 of the total signal), however, the significantly smaller mass of this species likely leads to a higher ionization efficiency and thus over-representation in the overall signal.

Therefore, all additional peaks in the T-RzA and T-RzB samples are of low abundance and are the result of RNA that is only slightly degraded (minus one nucleotide) at its very end (peaks "$", "¥", "&") or is a degradation product of such small mass (peaks "#", "%") that it must have been generated after EMSA and subsequent D-PAGE. We conclude that our mass spectrometry data are not consistent with the notion that chemical heterogeneity is responsible for the hairpin ribozyme's molecular

heterogeneity. We also note that no signs of covalent modification were observed in a previous tandem mass spectrometry analysis of enzymatically digested hairpin ribozyme (38), and that molecular heterogeneity is observed independently of the buffers used for gel electrophoresis and biochemical assays (data not shown). Interestingly, the higher average charge state distributions observed in our ESI mass spectrometry for B-RzA and B-RzB relative to their T species counterparts is consistent with more stably folded conformations of the B species strands (39), although they could also be due to variations in concentration and desalting efficiency.

2.3 Discussion

We show here that molecular heterogeneity is intrinsic to at least the hairpin ribozyme and common to both major routes for the *in vitro* preparation of RNA, chemical synthesis and *in vitro* transcription. We demonstrate that multiple folding sub-populations of a functional RNA are unexpectedly resistant to repartitioning following unfolding (denaturation). Furthermore, our mass spectrometry data indicate that the heterogeneity is encoded in the RNA fold and is not a result of chemical modification.

Additional, circumstantial evidence against chemical sources of molecular heterogeneity

Our EMSA analysis indicates that, if present, chemical modifications that destabilize the docked state to form the T species would need to occur either during or prior to chemical or enzymatic synthesis. This idea is rationalized as follows: If destabilizing chemical damage arose subsequent to synthesis, then we would expect to observe an accumulating conversion of B to T species following isolation of the T and B species. However, exposure of the B species to conditions identical to our handling prior to EMSA (including D-PAGE, and ethanol precipitation) does not result in (additive) conversion to T species, so that post-synthesis chemical damage cannot explain the observed heterogeneity. Furthermore, our re-folding assays are inconsistent with the presence of any form of destabilizing chemical damage that favors formation of the T species as the source of the EMSA heterogeneity. Our D-PAGE based refolding assays demonstrate that strands B-RzA and B-RzB are both fully competent to reform the B species (Figure 2.5),

so if chemical damage led to formation of the T species, then B-RzA and B-RzB have to be presumed undamaged (or at least not damaged in a way that destabilizes docking). We further infer that we would be able to interpret the fractions of T species observed in the heterologous mixtures of T-RzA+B-RzB and B-RzA+T-RzB as the fractions of damaged T-RzA and T-RzB strands, respectively. However, the sum of the fractions of top band observed for T-RzA+B-RzB and B-RzA+T-RzB is only 53% (17%+36%), far less than the 100% necessary to fully explain the T species in terms of chemical damage. From this argument follows that, if we assumed chemical damage to be responsible for the observed molecular heterogeneity, then our refolding assays would suggest that these chemical modifications be present in the B species and therefore fortuitously stabilize the docked (catalytically active) state. This possibility that chemical damage favors the predominantly docked and catalytically most active B species seems unlikely since nearly all chemical modifications of the hairpin ribozyme tested in previous studies significantly destabilize the docked state (11,22).

Potential role of S-turns in highly stable, functional RNA structures

Taking all of our observations together, the energy barrier to interconversion of the hairpin ribozyme's conformational isomers appears to approach or even exceed the barrier to backbone scission. The so-called ergodic hypothesis implies that each functional molecule, given enough time and energy, faithfully reproduces the full range of conformational and functional behaviors observed in a snapshot of the ensemble (40,41). Our observations underscore that the notion of ergodic behavior, which is generally used as a basis to interpret ensemble-averaged data, cannot be safely assumed even for RNAs as simple as the 2WJ hairpin ribozyme. While there may still exist untested conditions under which the conformations do exchange, it appears unlikely that such conditions would be compatible with RNA chemical stability since all attempts at interconversion under conditions that are extreme and yet compatible with chemical stability provide no significant signs of conversion of the B to the T species, unless one of the composite RNA strands is exchanged for fresh material. (Please note that conversion of the T to the B species does appear to occur, Figure 2.5) Our experiments therefore suggest that a given functional hairpin ribozyme molecule in solution at

standard pressure will undergo spontaneous degradation before it can sample all major populations present in a snapshot of the ensemble.

A plausible explanation for our observations is the retention of alternative, non-native, intra-strand structures that arise along the unfolding pathway of the T and B species and are resistant to unfolding under harshly denaturing conditions. While alternative explanations, including mass neutral chemical modifications such as the isomerization of the backbone to a 2',5'-phosphodiester, cannot not be entirely ruled out, we do not favor such an interpretation based on the presence of T and B species in both synthetic and *in vitro* transcribed ribozymes. A strikingly similar phenomenon was observed in studies of the Sarcin-Ricin loop (SRL) where multiple populations in native gels cannot be interconverted, following even multiple cycles of heating to 90 °C (42). The highly persistent heterogeneity in the SRL was unexpected for this small (23-nucleotide) stem loop structure, and as with the hairpin ribozyme the heterogeneity is independent of how the RNA is synthesized. Interestingly, the SRL and loop B of the hairpin ribozyme share sequence homology and an S-turn structural motif, which changes shape in the hairpin ribozyme upon interdomain docking (Figure 2.10). The S-turn thus arises as a possible source of the deeply dividing energy barriers in both RNAs.

One function, multiple folds

The capacity of RNA to adopt multiple non-exchanging functional conformations suggests that, in addition to the paradigm of one-fold-one-function (6), superior fitness of a specific RNA sequence may be achieved by its capability to adopt multiple folds that converge on a single function. Recent cross-linking experiments support the presence of multiple functional folds of the hairpin ribozyme (27). Cross-linking in the presence of cobalt-hexamine indicates the presence of U+2-G36 base stacking that is compatible with catalytic activity, but inconsistent with high-resolution x-ray crystal structures solved in the presence of cobalt hexamine (28). The crystal structures and cross-linking data appear to reveal at least two significantly different conformations compatible with catalytic activity, consistent with our single-molecule FRET analysis of the catalytically active ribozyme (Figure 2.1c). Conformational differences around G36 could account for our footprinting data at C25, which is positioned in reasonable proximity to G36 across

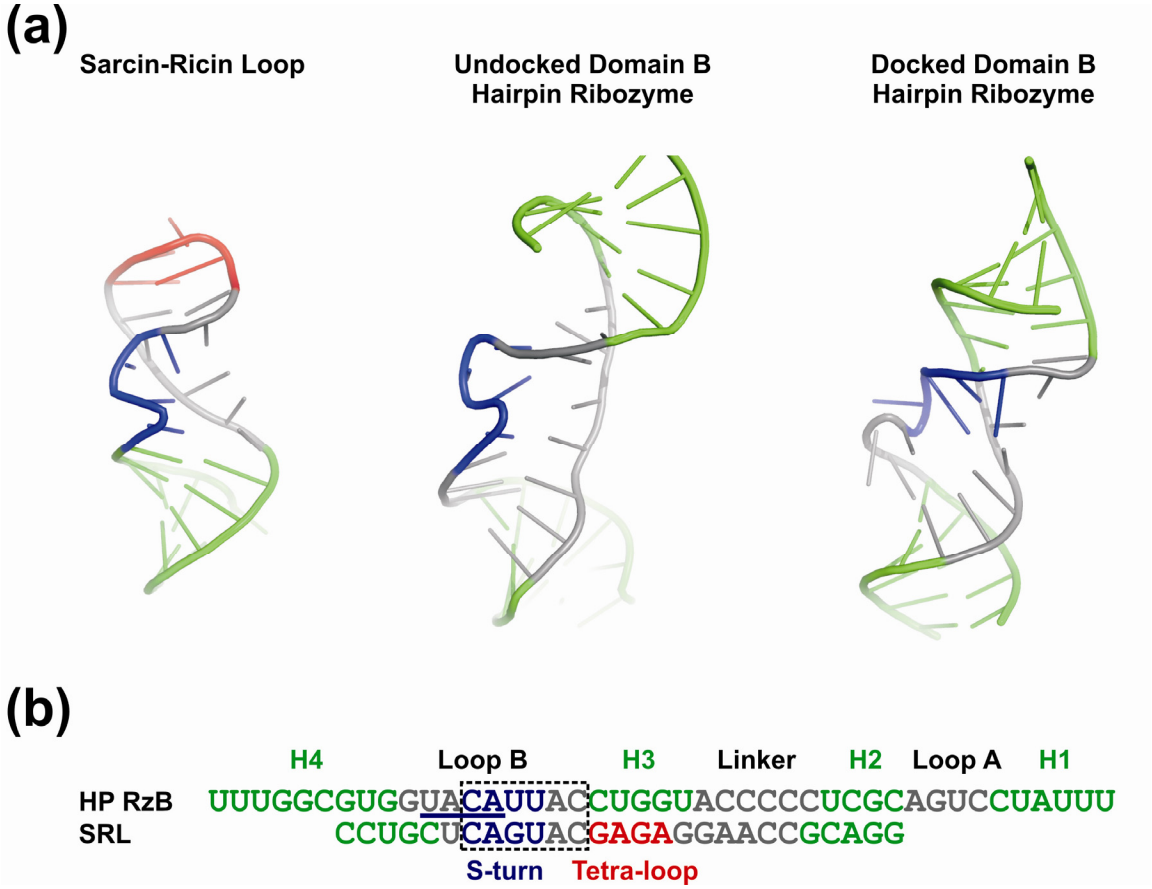


Figure 2.10: Structural and sequence homology between the sarcin-ricin loop (SRL) and loop B of the hairpin ribozyme. (*a*) Cartoon representations of the crystal structure of the SRL (PDB ID 1Q9A), the undocked solution structure of the isolated domain B (PDB ID 1B36), and the crystal structure of domain B after docking with domain A (PDB ID 2OUE). Structural elements are color coded as follows; helices are green, the GAGA tetra-loop is red, and the S-turn in the SRL and the corresponding nucleotides 39-42 in loop B are blue; in the undocked domain B structure these nucleotides compose the S-turn. The S-turn observed in the docked structure is formed by an alternative sequence (nucleotides 37-40). (*b*) Alignment of the sequence of the RzB strand used in this study with the SRL sequence from previous EMSA assays (42). The S-turns of the undocked domain B and SRL are aligned and color coded as in panel *a* for comparison, whereas the S-turn sequence in the docked domain B is underlined.

loop B. Our EMSA refolding experiments now show that the multiple functional conformations of the hairpin ribozyme are separated by unexpectedly high energy barriers.

Our experiments, in conjunction with the widespread observation of multiphasic and incomplete reaction kinetics of functional RNAs in ensemble-averaged experiments, as well as a growing body of evidence for heterogeneous single molecule behavior, suggest that deeply furrowed free energy landscapes may be an inescapable feature of RNA in biology and biotechnology.

2.4 Acknowledgements

We thank Gregory Bokinsky and Hazen Babcock for their assistance in setting up our single molecule microscope and Daniel Herschlag, Xiaowei Zhuang, Steven Chu, Andrew Feig, and Hashim al-Hashimi for helpful discussions, and an anonymous reviewer for making us aware of reference 42. This work was supported by NIH grant GM62357 to NGW and an NIH Molecular Biophysics Training Grant fellowship to MAD. Work in KH's laboratory was supported by an NSF Career award (CHE-05-47699).

References

1. Amaral, P.P., Dinger, M.E., Mercer, T.R. and Mattick, J.S. (2008) The eukaryotic genome as an RNA machine. *Science (New York, N.Y*, **319**, 1787-1789.
2. Couzin, J. (2008) MicroRNAs make big impression in disease after disease. *Science (New York, N.Y*, **319**, 1782-1784.
3. Zhuang, X., Kim, H., Pereira, M.J., Babcock, H.P., Walter, N.G. and Chu, S. (2002) Correlating structural dynamics and function in single ribozyme molecules. *Science (New York, N.Y*, **296**, 1473-1476.
4. Sosnick, T.R. and Pan, T. (2003) RNA folding: models and perspectives. *Curr. Opin. Struct. Biol.*, **13**, 309-316.
5. Pljevaljcic, G., Klostermeier, D. and Millar, D.P. (2005) The tertiary structure of the hairpin ribozyme is formed through a slow conformational search. *Biochemistry*, **44**, 4870-4876.
6. Schultes, E.A., Spasic, A., Mohanty, U. and Bartel, D.P. (2005) Compact and ordered collapse of randomly generated RNA sequences. *Nat. Struct. Mol. Biol.*, **12**, 1130-1136.
7. Russell, R., Das, R., Suh, H., Travers, K.J., Laederach, A., Engelhardt, M.A. and Herschlag, D. (2006) The paradoxical behavior of a highly structured misfolded intermediate in RNA folding. *J. Mol. Biol.*, **363**, 531-544.
8. Bhaskaran, H. and Russell, R. (2007) Kinetic redistribution of native and misfolded RNAs by a DEAD-box chaperone. *Nature*, **449**, 1014-1018.
9. Zhuang, X., Bartley, L.E., Babcock, H.P., Russell, R., Ha, T., Herschlag, D. and Chu, S. (2000) A single-molecule study of RNA catalysis and folding. *Science (New York, N.Y*, **288**, 2048-2051.
10. Bokinsky, G., Rueda, D., Misra, V.K., Rhodes, M.M., Gordus, A., Babcock, H.P., Walter, N.G. and Zhuang, X. (2003) Single-molecule transition-state analysis of RNA folding. *Proc. Natl. Acad. Sci. USA*, **100**, 9302-9307.
11. Rueda, D., Bokinsky, G., Rhodes, M.M., Rust, M.J., Zhuang, X. and Walter, N.G. (2004) Single-molecule enzymology of RNA: essential functional groups impact catalysis from a distance. *Proc. Natl. Acad. Sci. USA*, **101**, 10066-10071.
12. Okumus, B., Wilson, T.J., Lilley, D.M. and Ha, T. (2004) Vesicle encapsulation studies reveal that single molecule ribozyme heterogeneities are intrinsic. *Biophys J.*, **87**, 2798-2806.
13. Xie, Z., Srividya, N., Sosnick, T.R., Pan, T. and Scherer, N.F. (2004) Single-molecule studies highlight conformational heterogeneity in the early folding steps of a large ribozyme. *Proc. Natl. Acad. Sci. USA*, **101**, 534-539.
14. Hodak, J.H., Downey, C.D., Fiore, J.L., Pardi, A. and Nesbitt, D.J. (2005) Docking kinetics and equilibrium of a GAAA tetraloop-receptor motif probed by single-molecule FRET. *Proc. Natl. Acad. Sci. USA*, **102**, 10505-10510.
15. Brender, J.R., Dertouzos, J., Ballou, D.P., Massey, V., Palfey, B.A., Entsch, B., Steel, D.G. and Gafni, A. (2005) Conformational dynamics of the isoalloxazine in substrate-free p-hydroxybenzoate hydroxylase: single-molecule studies. *J Am Chem Soc*, **127**, 18171-18178.
16. Downey, C.D., Fiore, J.L., Stoddard, C.D., Hodak, J.H., Nesbitt, D.J. and Pardi, A. (2006) Metal ion dependence, thermodynamics, and kinetics for intramolecular

docking of a GAAA tetraloop and receptor connected by a flexible linker. *Biochemistry*, **45**, 3664-3673.
17. Liu, S., Bokinsky, G., Walter, N.G. and Zhuang, X. (2007) Dissecting the multistep reaction pathway of an RNA enzyme by single-molecule kinetic "fingerprinting". *Proc. Natl. Acad. Sci. USA*, **104**, 12634-12639.
18. Pereira, M.J.B., Nikolova, E.N., Hiley, S.L., Collins, R.A. and Walter, N.G. (2008) Single molecule FRET microscopy of a complex ribozyme reveals dynamic hierarchical folding toward catalysis. *J. Mol. Biol.*, **in press**.
19. Ditzler, M.A., Aleman, E.A., Rueda, D. and Walter, N.G. (2007) Focus on function: Single molecule RNA enzymology. *Biopolymers*, **87**, 302-316.
20. Bokinsky, G. and Zhuang, X. (2005) Single-molecule RNA folding. *Acc Chem Res*, **38**, 566-573.
21. Walter, N.G. (2001) Structural dynamics of catalytic RNA highlighted by fluorescence resonance energy transfer. *Methods*, **25**, 19-30.
22. Walter, N.G., Hampel, K.J., Brown, K.M. and Burke, J.M. (1998) Tertiary structure formation in the hairpin ribozyme monitored by fluorescence resonance energy transfer. *EMBO J.*, **17**, 2378-2391.
23. Butcher, S.E., Allain, F.H. and Feigon, J. (1999) Solution structure of the loop B domain from the hairpin ribozyme. *Nat. Struct. Biol.*, **6**, 212-216.
24. Walter, N.G., Yang, N. and Burke, J.M. (2000) Probing non-selective cation binding in the hairpin ribozyme with Tb(III). *J. Mol. Biol.*, **298**, 539-555.
25. Rupert, P.B. and Ferre-D'Amare, A.R. (2001) Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature*, **410**, 780-786.
26. Rupert, P.B., Massey, A.P., Sigurdsson, S.T. and Ferre-D'Amare, A.R. (2002) Transition state stabilization by a catalytic RNA. *Science (New York, N.Y*, **298**, 1421-1424.
27. Lambert, D., Heckman, J.E. and Burke, J.M. (2006) Cation-specific structural accommodation within a catalytic RNA. *Biochemistry*, **45**, 829-838.
28. Salter, J., Krucinska, J., Alam, S., Grum-Tokars, V. and Wedekind, J.E. (2006) Water in the active site of an all-RNA hairpin ribozyme and effects of Gua8 base variants on the geometry of phosphoryl transfer. *Biochemistry*, **45**, 686-700.
29. Walter, N.G. (2002) Probing RNA structural dynamics and function by fluorescence resonance energy transfer (FRET). *Curr. Protocols Nucleic Acid Chem. Chapter 11.10, 11.10.1-11.10.23.*
30. Pereira, M.J., Harris, D.A., Rueda, D. and Walter, N.G. (2002) Reaction pathway of the trans-acting hepatitis delta virus ribozyme: a conformational change accompanies catalysis. *Biochemistry*, **41**, 730-740.
31. Harris, D.A., Tinsley, R.A. and Walter, N.G. (2004) Terbium-mediated footprinting probes a catalytic conformational switch in the antigenomic hepatitis delta virus ribozyme. *J. Mol. Biol.*, **341**, 389-403.
32. Limbach, P.A., Crain, P.F. and McCloskey, J.A. (1995) Characterization of oligonucleotides and nucleic acids by mass spectrometry. *Curr. Opin. Biotechnol.*, **6**, 96-102.
33. Mo, J. and Hakansson, K. (2006) Characterization of nucleic acid higher order structure by high-resolution tandem mass spectrometry. *Anal. Bioanal. Chem.*, **386**, 675-681.

34. Yang, J., Mo, J., Adamson, J.T. and Hakansson, K. (2005) Characterization of oligodeoxynucleotides by electron detachment dissociation fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.*, **77**, 1876-1882.
35. Senko, M.W., Canterbury, J.D., Guan, S. and Marshall, A.G. (1996) A high-performance modular data system for Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun. Mass. Spectrom.*, **10**, 1839-1844.
36. Ledford, E.B., Jr., Rempel, D.L. and Gross, M.L. (1984) Space charge effects in Fourier transform mass spectrometry. Mass calibration. *Anal. Chem.*, **56**, 2744-2748.
37. Huntzinger, E., Boisset, S., Saveanu, C., Benito, Y., Geissmann, T., Namane, A., Lina, G., Etienne, J., Ehresmann, B., Ehresmann, C. *et al.* (2005) Staphylococcus aureus RNAIII and the endoribonuclease III coordinately regulate spa gene expression. *EMBO J.*, **24**, 824-835.
38. Thomas, J.M. and Perrin, D.M. (2006) Active site labeling of G8 in the hairpin ribozyme: implications for structure and mechanism. *J. Am. Chem. Soc.*, **128**, 16540-16545.
39. Guo, X., Bruist, M.F., Davis, D.L. and Bentzley, C.M. (2005) Secondary structural characterization of oligonucleotide strands using electrospray ionization mass spectrometry. *Nucleic acids research*, **33**, 3659-3666.
40. Neumann, J.V. (1932) Physical Applications of the Ergodic Hypothesis. *Proc. Natl. Acad. Sci. USA*, **18**, 263-266.
41. Birkhoff, G.D. and Koopman, B.O. (1932) Recent Contributions to the Ergodic Theory. *Proc. Natl. Acad. Sci. USA*, **18**, 279-282.
42. Korennykh, A.V., Plantinga, M.J., Correll, C.C. and Piccirilli, J.A. (2007) Linkage between substrate recognition and catalysis during cleavage of sarcin/ricin loop RNA by restrictocin. *Biochemistry*, **46**, 12744-12756.

Chapter 3

Molecular Dynamics Suggest Multifunctionality of an Adenine Imino Group in Acid-base Catalysis of the Hairpin Ribozyme[c]

3.1 Introduction

Over two decades of investigations into RNA catalysis have revealed numerous structural and dynamic properties of RNA that give rise to site-specific phosphoryl transfer. For example, the first ribozymes to be discovered, RNase P (1) and the self-splicing group I introns (2), utilize specifically bound metal ions as cofactors in their catalytic mechanisms (3), a strategy also used by protein-based RNases (4,5). Positioning of organic cofactors by the *glm*S ribozyme (6,7), and the group I intron, are an additional means of specifying the reactive phosphate. In analogy, early hypotheses proposed that the small self-cleaving ribozymes use specifically bound divalent metal ions as the necessary base and acid catalysts. Subsequently it was recognized that the catalytic competence of several small self-cleaving RNAs (the hairpin, hammerhead, and Varkud satellite ribozymes) in a wide range of ionic conditions, including in the presence of monovalent ions alone, supports the notion that nucleobase functional groups directly support reaction chemistry (8). These naturally occurring ribozymes are now generally considered to rely at least partially on nucleobase facilitated proton transfer to carry out their cleavage and ligation function.

High-resolution crystal structures exist for the HDV (9,10), hammerhead (11,12), *glmS* (13,14), and hairpin ribozymes (15-21). Based on these structures only the hairpin ribozyme's active site is unambiguously devoid of stably bound metal ions that could potentially participate in catalysis with the possible exception of electrostatic stabilization

[c] Adapted from Ditzler MA, Šponer J, Walter NG. Molecular Dynamics Suggest Multifunctionality of an Adenine Imino Group in Acid-base Catalysis of the Hairpin Ribozyme. *RNA* (*in press*)

which could influence chemistry from outside the active site. Furthermore, the hairpin ribozyme is highly active in cobalt(III) hexammine, which can only form chemically inert outer-sphere complexes, while in contrast the HDV ribozyme is competitively inhibited by it (9,22). This limited potential for a catalytic role of metal ions makes the hairpin ribozyme especially useful in probing the direct role of nucleobases in RNA catalysis. In addition, its small size has allowed a wide range of modifications and base substitutions to be incorporated, facilitating investigations into structure, function and dynamics (23-31). Despite intense study, much remains unclear about the chemical mechanism and the structural dynamics of the catalytically competent global conformation. Here we present the results of ~300 ns of explicit solvent molecular dynamics (MD) simulations on the hairpin ribozyme in which we examine the active site geometry together with the role of both nucleobases and structural water molecules in facilitating RNA catalysis.

MD simulation is a computational method describing the dynamics of solvated nucleic acids using classical empirical potential force fields. The outcome of simulations is determined by the quality of the force field, by the affordable simulation timescale and by the starting structures. Therefore, MD simulations are a tool primarily suitable to provide complementary analyses of known RNA structures, including testing the effects of base substitutions, modifications, protonation and deprotonation on structural dynamics and hydration. Although such a classical simulation does not allow bond breaking and creating, it can evaluate possible catalytic mechanisms by assessment of the conformational dynamics observed in the simulation (32-38). This was our goal here.

Biochemical and structural data have implicated G8 and A38 as participants in the catalysis of cleavage and ligation. The conserved guanine G8 in loop A is positioned near the active site and mutation or deletion at this site impairs activity 100- to 1000-fold without significantly disrupting the ribozyme's global structure (27,30). The position of G8 near the reactive 2'-OH in crystal structures suggests the possibility that an unprotonated N1 acts as a general base (16,18,19,30,39). By contrast, exogenous nucleobase rescue experiments suggest that the roles of G8 in catalysis lie in charge stabilization of the transition state and/or alignment of the reactive groups (27,29,40,41). Our MD simulations are consistent with the latter model as they provide evidence that G8

facilitates catalysis through stabilizing both the developing charge on the scissile phosphate and the strained backbone conformations adopted along the reaction pathway.

Recent crystallographic studies of a transition state analog of the hairpin ribozyme have suggested that localized water molecules also play a role in transition state charge stabilization (16), which is also supported by our MD simulations. Transition-state analog crystal structures place N1 of the invariant A38 close to the 5'-oxygen leaving group of the cleavage site, implicating A38 as the general acid (15,18,20,21). Abasic substitution of A38 impairs catalysis >10,000-fold and exogenous nucleobase rescue experiments indicate that the protonation state of A38(N1) plays a direct role in chemistry (23,25). It has also been suggested that the catalytic roles of A38 are to align reactive groups and to stabilize negative charge in the transition state (23,25). Our MD simulations offer support for multiple roles of A38 as the general acid, in aligning reactive groups, and in stabilizing negative charge. Contrary to the prevailing opinion, however, our results also implicate A38 as the general base in the cleavage reaction, leading us to consider a previously undescribed type of proton transfer mechanism for RNA involving a single nucleobase as bifunctional general base and acid catalyst, in analogy to protein enzymes such as serine-carboxyl peptidases, and similar to roles proposed for glucose-amine-6-phosphate in the glmS ribozyme mechanism.

3.2 Materials and Methods

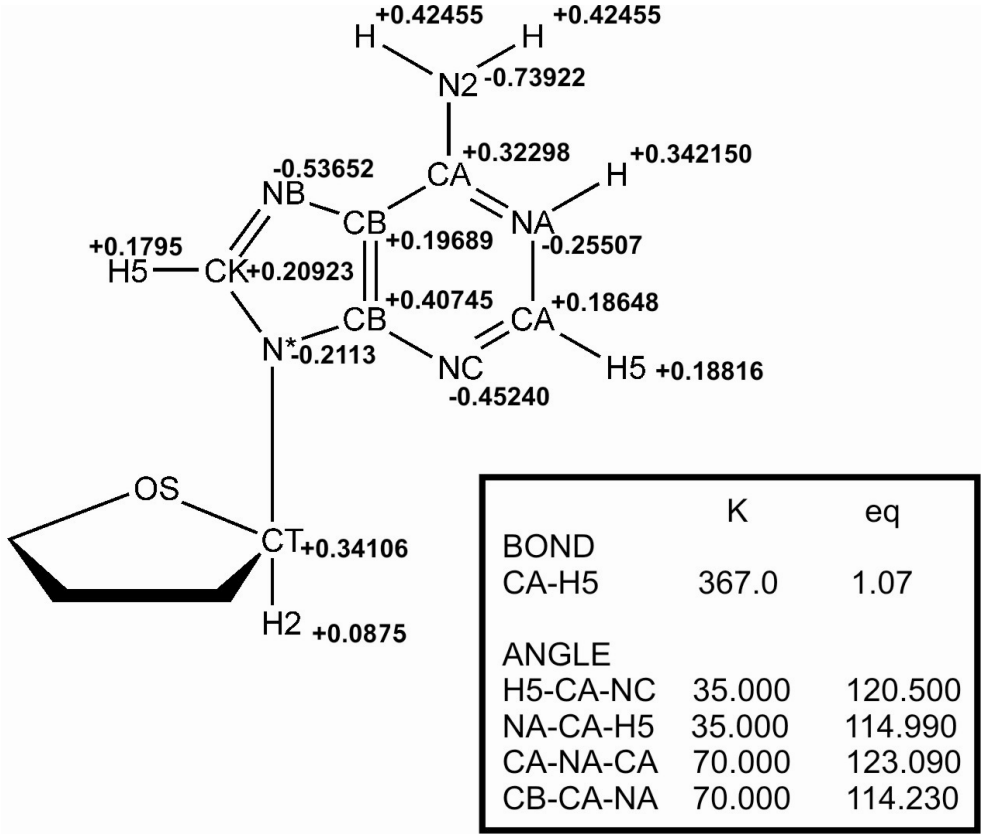
Initial MD structures, water and ion placement

Table 1 summarizes all 15 (10-40 ns long) simulations performed for this study. Simulations of the C39 ribozyme were started from the 2.05-Å resolution crystal structure of the junctionless ligated hairpin ribozyme (PDB ID 2OUE). For Simulations C39-1 and C39-2 five inter-domain water molecules were placed based on their crystallographic coordinates, a third simulation (C39-3) was initiated without these core water molecules. Eight simulations of the U39 sequence were started from the 2.65-Å resolution crystal structure of the same ribozyme (2D2K), using either the U37 sequestered (U39-S) or U37 exposed (U39-E) conformation. Half of the U39 simulations were run with a single Mg^{2+} placed based on the coordinates of a crystallographically

resolved cobalt(III) hexammine binding site (Table 1) (17). Another half of the U39 simulations (encompassing half of the Mg^{2+} containing simulations) were initiated with five inter-domain water molecules, placed by superimposing the C39 structure (for which inter-domain waters could be clearly resolved) on the U39 structure. A G8A mutant ribozyme was simulated based on the corresponding 2.4-Å resolution crystal structure (1ZFV), with inter-domain water molecules placed by superimposing the C39 structure onto the G8A structure. An additional three simulations were performed in which A38 was protonated at atom N1 starting from either the junctionless C39 crystal structure (2OUE) or the G8A structure (1ZFV), and all three used the crystallographically placed inter-domain water molecules. The LEAP module from the AMBER-8 software package was used to place additional bulk water molecules and a charge neutralizing number (55-57) of Na^+ ions at points of favorable electrostatic surface potential.

MD equilibration and simulation

All MD simulations were carried out by using the AMBER-8 software package with the parm99 Cornell *et al.* force field (42-44). The initial structures were solvated in a rectangular periodic box of TIP3P water molecules extending (at least) 10 Å from the RNA surface. The Sander module of AMBER-8 was used for the equilibration and production simulations based on our standard protocols (45-49). The particle mesh Ewald method (50) was applied with a heuristic pair list update, using a 2.0-Å non-bonded pair list buffer and a 9.0-Å cutoff. A charge grid spacing of close to 1 Å and a cubic interpolation scheme were used. The production runs were carried out at 300 K with constant-pressure boundary conditions using the Berendsen temperature coupling algorithm (51) with a time constant of 1.0 ps. SHAKE (52) was applied with a tolerance of 10^{-8} to constrain bonds involving hydrogens. Parameters for the protonated form of A38 were obtained by applying the RESP (53) fitting method to the quantum mechanically (BLY/6-31G*) derived electrostatic potential. The charges on the backbone atoms were held constant with those present in the parm99 force field, and atom types were adjusted as necessary to maintain consistency with the parm99 force field (Figure. 3.1).



	K	eq
BOND		
CA-H5	367.0	1.07
ANGLE		
H5-CA-NC	35.000	120.500
NA-CA-H5	35.000	114.990
CA-NA-CA	70.000	123.090
CB-CA-NA	70.000	114.230

Figure 3.1: Force field parameters for adenosine protonated at N1. The atom types and partial charges are indicated on the atoms, whereas nonstandard equilibrium angles and distances are boxed.

MD analysis

Production trajectories were analyzed using the ptraj modules of the AMBER-8 package to obtain inter-atomic distances, angles, pseudorotation angles, and ion density maps. To obtain a detailed inventory of all direct interdomain RNA-RNA hydrogen bonds, the ptraj module of AMBER-8 was used to track the number of hydrogen bonds formed after 9.9 to 10 ns of simulation time. Fractional hydrogen bonds were calculated from the fraction of time that the heavy-atom distances of these hydrogen bonds were <3.1Å and that the X-H…Y angle was within 60^{o} of linearity. List density plots of distances were generated from the appropriate ptraj output using Matlab. Backbone torsion angle analysis was carried out using Molprobity (54). The average backbone suiteness was determined from all complete suites, including triaged suites. Snapshots averaged over 100 ps of MD simulation time were rendered using Pymol (DeLano Scientific LLC).

Electrophoretic mobility shift assay (EMSA)

The hairpin ribozyme sequences used were Strand RzA (5'-AAA UAG AGA A**A**C GAA CCA GAG AAA CAC ACG CCA AA-3') with the underlined residue at position 8 mutated from a G to an A to prevent cleavage during electrophoresis, and Strand RzB (5'-AU AUA UUU GGC GUG GUA **Y**AU UAC CUG GUA CCC CCU CGC AGU CCU AUU U-3') with the underlined pyrimidine representing either a U in the wild-type sequence or a C in the gain-of-function mutant. RNA was generated *in vitro* from single-stranded DNA templates containing a double-stranded T7 promoter as described (55). Strand RzA was 5'-^{32}P labeled using γ-^{32}P-ATP and T4 polynucleotide kinase. The two composite strands were annealed in buffer (50 mM Tris-Acetate, pH 7.5, 50 mM Na-Acetate) by heating to 70^{o}C for 2 min followed by cooling to 25^{o}C, 10% (v/v) glycerol were added, and the sample loaded onto a non-denaturing 10% (w/v) polyacrylamide (19:1 acrylamide:bisacrylamide ratio) gel containing 50 mM Tris-Acetate, pH 7.5, 12 mM Mg-Acetate. Gels were run at 20 V/cm and 4°C for 15 h, and radioactive bands visualized using autoradiography with phosphorscreens and quantified using a PhosphorImager Storm 840 with ImageQuant software (Molecular Dynamics).

TABLE 1. MD simulations performed for the current study

Simulation	Starting Structure	Characteristics	Exp. Activity[a]	Ions	Crystal Waters[b]	Length (ns)	H-bonds after 10 ns[c]
C39-1	2OUE	U39C	++	57 Na^+	5	20	12.37
C39-2	2OUE	U39C	++	57 Na^+	5	40	11.03
C39-3	2OUE	U39C	++	57 Na^+	0	20	9.57
A38H+-1	C39-1 (4 ns)	U39C, $A38H^+$	++	56 Na^+	5	20	11.37
A38H+-2	2OUE	U39C, $A38H^+$	++	56 Na^+	5	20	10.05
U39-S1	2D2K	U37 sequestered	+	57 Na^+	5	25	7.23
U39-E1	2D2K	U37 exposed	+	57 Na^+	5	24	10.66
U39-S2	2D2K	U37 sequestered	+	55 Na^+, 1 Mg^{2+}	5	10	9.66
U39-E2	2D2K	U37 exposed	+	55 Na^+, 1 Mg^{2+}	5	13.6	8.82
U39-S3	2D2K	U37 sequestered	+	57 Na^+	0	20	7.59
U39-E3	2D2K	U37 exposed	+	57 Na^+	0	14.6	9.69
U39-S4	2D2K	U37 sequestered	+	55 Na^+, 1 Mg^{2+}	0	10	12.94
U39-E4	2D2K	U37 exposed	+	55 Na^+, 1 Mg^{2+}	0	10	6.61
A8	1ZFV	U39C, G8A	-	57 Na^+	5	20	11.18
A8 A38H+	1ZFV	U39C, G8A, $A38H^+$	-	56 Na^+	5	20	10.89

[a]Relative catalytic activity observed in prior experimental cleavage assays of this variant.
[b]Number of water molecules crystallographically resolved in the catalytic core and retained for the simulation.
[c]Hydrogen bond inventory calculated as described in Materials and Methods between 9.9 and 10 ns.

3.3 Results

Overview of our simulations

We performed 15 charge neutralized MD simulations of the hairpin ribozyme for a combined total of ~300 ns using explicit solvent, either with or without five crystallographically resolved inter-domain water molecules and either with or without a single Mg^{2+} ion in a well resolved metal binding site, all based on the 2.05-2.65 Å resolution crystal structures of junctionless forms of the hairpin ribozyme in complex with a non-cleavable substrate analog (Table 1). We simulated three different sequences based on the available crystal structures for each variant; the native sequence containing a U39 (PDB ID 2D2K), a G8A mutant (in the following termed A8) that inhibits cleavage (1ZFV), and a U39C mutant (in the following termed C39) that increases activity (56,57) (2OUE) (Figure 3.2A). We also performed three simulations in which an active site adenosine was protonated at the N1 position (in the following termed A38H+) based on recent evidence for catalytic relevance of the ionized form (23,25). Where present, a catalytically inactivating 2'-O-methyl modification of A-1 at the cleavage site was replaced with the natural 2'-OH.

MD simulations properly predict backbone conformations as defined by their suiteness and provide indirect evidence for protonation of A38 in crystal structures

Given the large number of torsion angles in the RNA backbone, a particular challenge of MD simulations is to correctly predict backbone conformations over extended simulation times (33). To evaluate the validity of our simulations, screen for potential force field artifacts, and rapidly identify regions of interest we therefore applied the recently developed program Suitename (54,58) to the backbone torsion angles in all our simulations. Overall, we observe that the conformity of the experimentally derived crystal structures with empirically inferred low-energy conformations is largely maintained in their corresponding simulations (Figures 3.2B and 3.3). Generally, backbone conformations that fall outside the empirically identified low energy conformations during the simulation also do so in the starting crystal structure, and therefore do not appear to arise from limitations of the parm99 Cornell et al. force field.

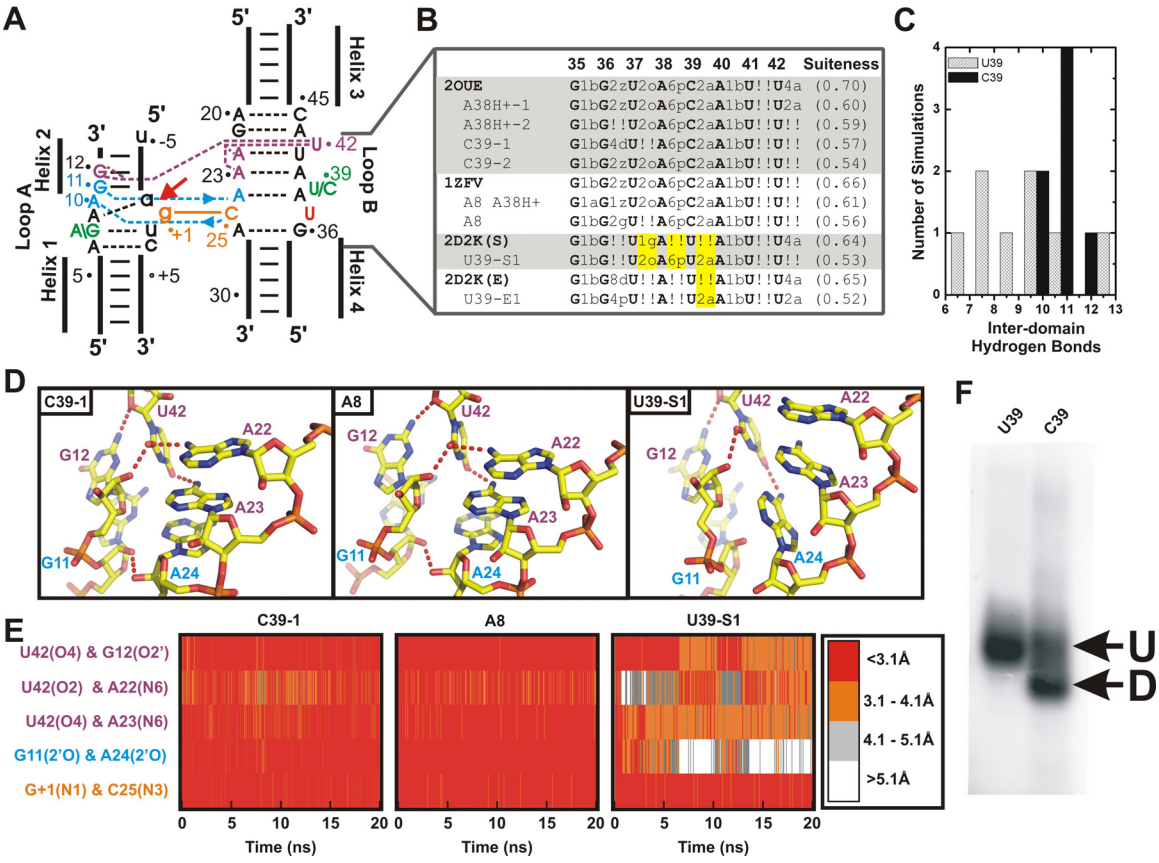

A
B
C
D
E
F
Helix 1
Helix 2
Helix 3
Helix 4
Loop A
Loop B
35 36 37 38 39 40 41 42 Suiteness
2OUE G1bG2zU2oA6pC2aA1bU!!U4a (0.70)
A38H+-1 G1bG2zU2oA6pC2aA1bU!!U2a (0.60)
A38H+-2 G1bG!!U2oA6pC2aA1bU!!U!! (0.59)
C39-1 G1bG4dU!!A6pC2aA1bU!!U!! (0.57)
C39-2 G1bG2zU2oA6pC2aA1bU!!U!! (0.54)
1ZFV G1bG2zU2oA6pC2aA1bU!!U!! (0.66)
A8 A38H+ G1aG1zU2oA6pC2aA1bU!!U!! (0.61)
A8 G1bG2gU!!A6pC2aA1bU!!U!! (0.56)
2D2K(S) G1bG!!U1gA!!U!!A1bU!!U4a (0.64)
U39-S1 G1bG!!U2oA6pU2aA1bU!!U!! (0.53)
2D2K(E) G1bG8dU!!A!!U!!A1bU!!U4a (0.65)
U39-E1 G1bG4pU!!A!!U2aA1bU!!U2a (0.52)
U39
C39
Number of Simulations
Inter-domain Hydrogen Bonds
C39-1
A8
U39-S1
U42
A22
G12
A23
G11
A24
U42(O4) & G12(O2')
U42(O2) & A22(N6)
U42(O4) & A23(N6)
G11(2'O) & A24(2'O)
G+1(N1) & C25(N3)
<3.1Å
3.1 - 4.1Å
4.1 - 5.1Å
>5.1Å
Time (ns)
U
D

Figure 3.2: Backbone conformational dynamics and global stability observed in MD simulations agree well with experimental data. (**A**) Secondary structure schematic of the minimal hairpin ribozyme used in our simulations. The cleavage site between A-1 and G+1 is indicated by a red arrow. Important inter-domain hydrogen bond interactions are indicated in blue (ribose zipper), purple (U42 binding pocket), and orange (G+1 binding pocket). (**B**) Analysis of a highly contorted region of the backbone using the software Suitename (58), whereby regions of the backbone that fall within one of the inferred low energy conformations are identified by their corresponding two character codes and outliers are identified by "!!". Crystal structures and corresponding simulations are aligned to compare backbone conformations. Two regions of the backbone that are likely more accurately represented in the simulations than the corresponding crystal structures are highlighted in yellow. Backbone conformations of simulations were determined using coordinates averaged over the last 10 ns of each simulation. The degree of conformity of the entire backbone to conformations experimentally observed for RNA in general is indicated by the average suiteness given in parentheses. Suiteness can range from 0 to 1, where 1 signifies that all suites are exactly equal to the mean dihedral values associated with their respective conformation. (**C**) Histograms of the number of C39 (black) and U39 (gray) simulations with a given number of inter-domain hydrogen bonds observed between 9.9 and 10 ns into the simulation. (**D**) Snapshots of the U42 binding pocket and part of the ribose zipper from the end of three simulations as indicated (averaged from 19.9 to 20 ns) with key hydrogen bonds indicated by dashed red tubes. (**E**) Density plots of the inter-atomic distances involved in key inter-domain hydrogen bonds (color-coded as in panel A) as tracked over the first 20 ns of each simulation. Distances are color-coded as indicated in the legend, i.e., red indicates atoms within hydrogen bonding distance. (**F**) Experimental EMSA analysis of a two-way junction form of the hairpin ribozyme indicates greater abundance of predominantly docked (D) relative to predominantly undocked (U) conformers for the C39 gain-of-function mutant compared to the U39 wild-type.

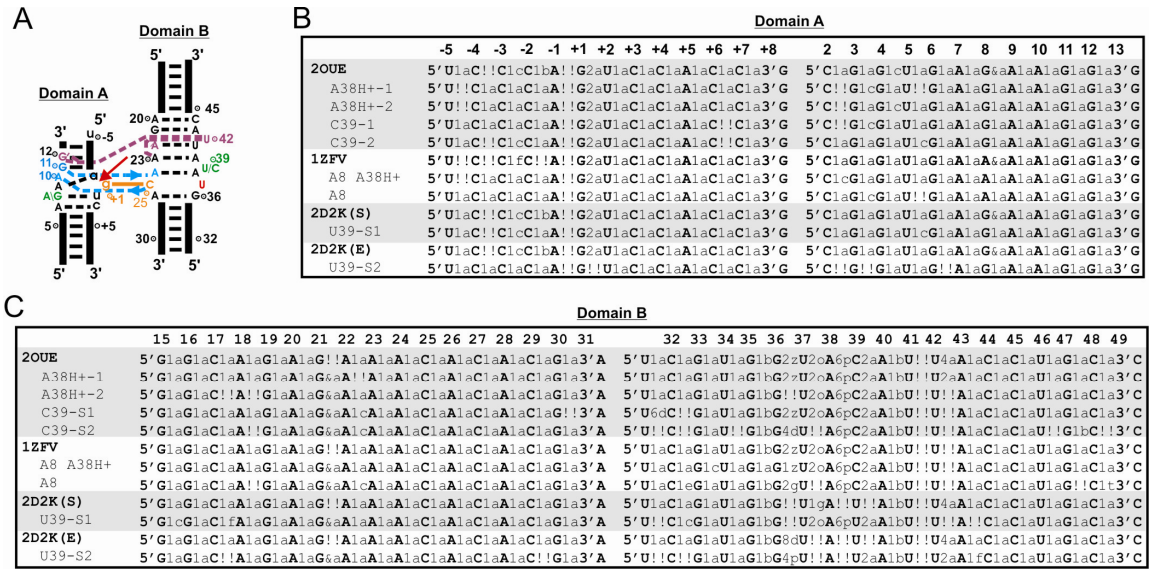


Figure 3.3: Backbone analysis of the entire backbone. (**A**) Secondary structure schematic of the minimal hairpin ribozyme used in our simulations. (**B**) Analysis of the complete domain A using the software Suitename (54,58). Regions of the backbone that fall within one of the inferred low energy conformations are identified by their corresponding two character codes; outliers are identified by !!. Crystal structures and corresponding simulations are aligned to compare backbone conformations. (**C**) Analysis of the complete domain B using Suitename.

In fact, we observe that some regions of the backbone are more accurately represented in the simulations. For example, the suite following U37 consistently shifts from 1g in the U39 crystal structure, usually observed as part of GNRA tetra loops, to the more conserved 2o suite observed in the higher resolution C39 crystal structures (highlighted in Figure 3.2B). This result is consistent with regularization of the RNA backbone in the course of previous simulations of the HDV ribozyme (45). In general, backbone conformations in the hairpin ribozyme depart from empirically identified low energy structures more significantly in the U39 simulations than in the C39 containing simulations (Figure 3.2B), potentially indicating increased dynamics of the U39 containing ribozyme.

Simulations in which N1 of A38 [A38(N1)] is not protonated diverge from the crystallographic backbone conformations more significantly than simulations in which N1 is protonated (Figure 3.2 and 3.4), providing indirect evidence that A38(N1) (unperturbed solution pK_a = 3.5 (59)) may be protonated in the crystal (pH = 6.0). The observed similarities between simulations of the G8A mutant in which A38 is protonated and the corresponding crystal structure, such as in overall backbone conformation and 2'O-P'-5'O in-line attack angle (IAA) of the scissile phosphate (simulation average = 107^o; crystal = 112^o), are consistent with protonation of A38(N1) in the G8A crystal also. Taken together, our results support previous biochemical (25), structural (15,18), and computational evidence (60) for a functionally relevant shift in the pK_a of A38.

MD simulations correctly predict enhanced docking stability of the C39 gain-of-function mutant

We have previously observed that MD simulations are effective at predicting the stability of the active docked state of the hairpin ribozyme. To assess stability, we used hydrogen bond inventories that add up all fractional hydrogen bonds as derived from the fraction of time that the heavy-atom distances of these hydrogen bonds are ≤3.1 Å and that the X-H--Y angle is within 60^o of linearity. When taken at 10 ns after equilibration such hydrogen bond inventories were found to correlate linearly with the experimentally observed decreased stability of mutants, yielding a reasonable ΔΔG of 2.4 kcal/mol per hydrogen bond lost during the simulation (49). Hydrogen bond inventories of our current

simulations predict that the inactive A8/C39 double mutant (termed A8, Table 1) should form a docked state equal in stability to the C39 gain-of-function mutant, consistent with previous electrophoretic mobility shift assays (EMSA) (30). Additionally, our hydrogen bond inventories suggest that the U39 sequence is less stably docked than the C39 mutants due to a loss of, on average, ~2 hydrogen bonds (Figure 3.2C). To pinpoint this observation more specifically, we monitored the integrity of the U42 binding pocket, ribose zipper, and the G+1 binding pocket, three interactions known to be essential to docking stability (Figures 3.2D, 3.2E, and 3.4) (31). The U39 wild-type sequence exhibits a relative loss of hydrogen bonding interactions particularly in the U42 binding pocket and the ribose zipper, which involves the 2'-OH functional groups of A10, G11, A24, and C25 (Figures 3.2D, 3.2E, 3.4, and Table 1). These observations suggest that the C39 gain-of-function mutant increases activity indirectly by stabilizing the catalytically active docked state through tertiary contacts distal from C39, attesting to a significant long-range effect of the mutation.

To further test the validity of our MD simulations, we sought to experimentally assess the relative stabilities of the U39 wild-type and C39 gain-of-function mutant by subjecting a common two-way junction form of the ribozyme to EMSA. We find that the C39 mutant is characterized by two bands of distinct mobility, whereas the U39 ribozyme lacks the more rapidly migrating band (Figure 3.2E). It has been previously shown that mutations that disturb global stability result in loss of the more rapidly migrating band that contains a predominantly docked and thus compact species (30,61,62). Our EMSA results therefore suggest that indeed the U39 wild-type is less stably docked than the C39 gain-of-function mutant, as also predicted by our simulations.

MD simulations corroborate an experimentally observed conformational heterogeneity in the U39 wild-type ribozyme

Crystal structures of the U39 wild-type ribozyme reveal two distinct conformations (termed "U37-sequestered" and "U37-exposed") at position 37 within the S-turn of loop B (17) (Figure 3.2A). We performed MD on both conformations of the U39 ribozyme (Table 1) and based on hydrogen bond inventories do not observe any clear differences in stability between the two conformations (Table 1). Remarkably, after 19 ns of one of our

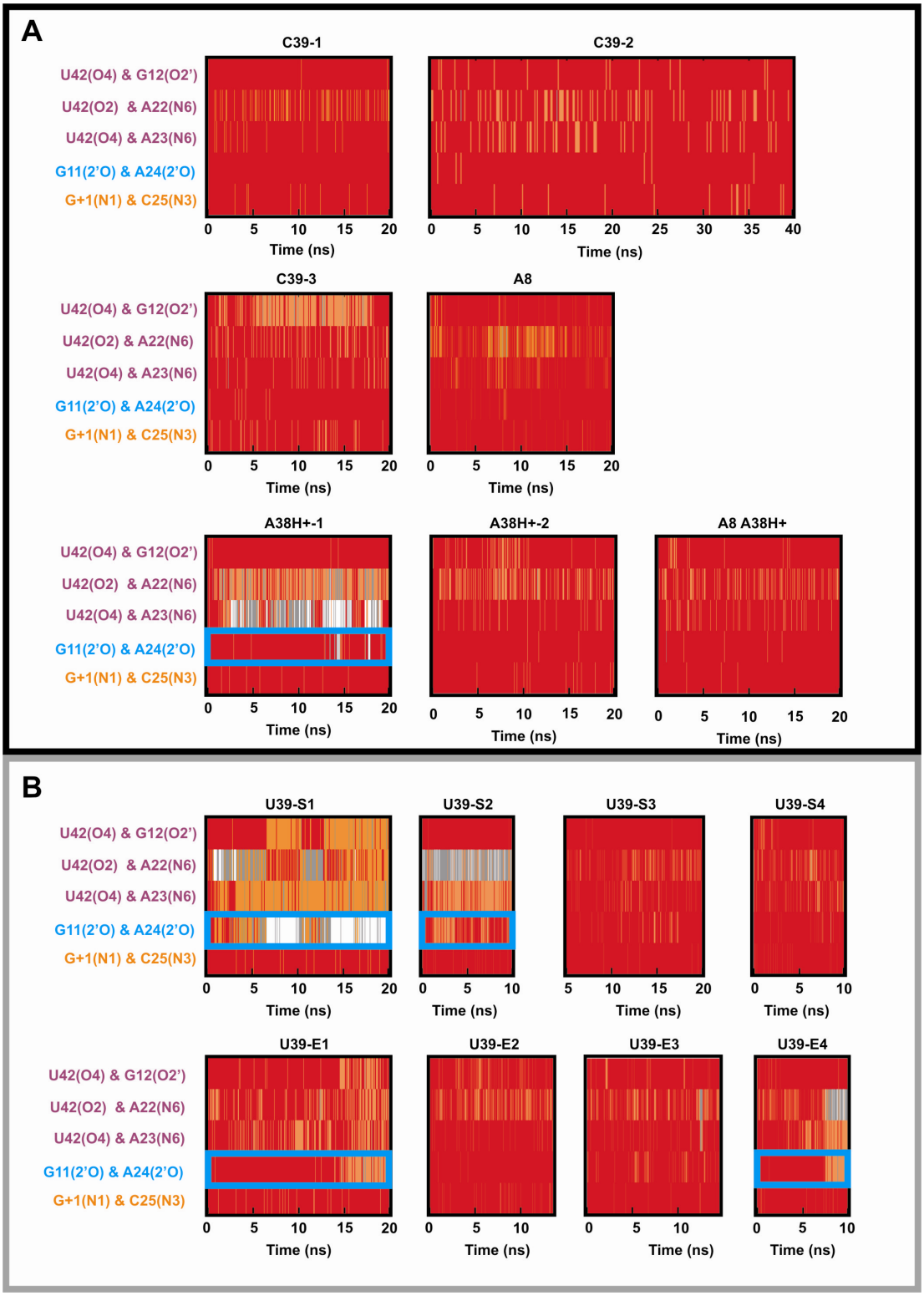


Figure 3.4: . Inter-atomic distances involved in the key inter-domain hydrogen bonds of Figures 3.2C and 3.2E, tracked for all of our simulations. Distances are color coded as indicated in Figure E. (**A**) Inter-domain distance for all simulations with C39 and (**B**) all simulations with U39. Simulations in which the ribose zipper is disrupted are indicated by blue boxes.

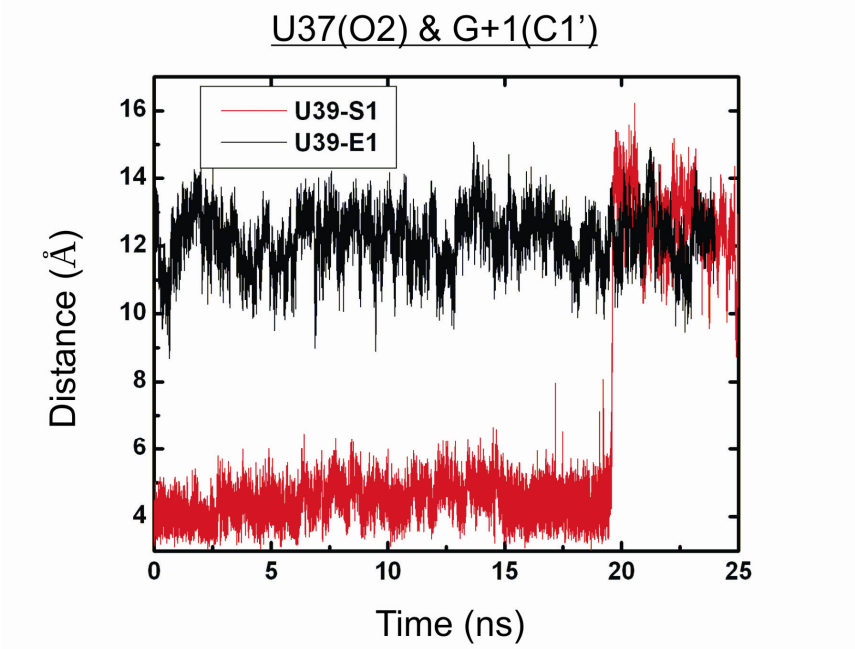


Figure 3.5: The U37 conformation transitions from the sequestered to the exposed conformation 19 ns into simulation U39-S1 (red), whereas simulation U39-E1 begins and stays in the exposed conformation (black). The U37 conformation is monitored by tracking the inter-atomic distance between O2 of U37 and C1' of G+1.

four U37-sequestered simulations the ribozyme undergoes a conformational change in which U37 moves away from the G+1 binding pocket and adopts a conformation similar to the U37-exposed conformation; extending this simulation for an additional 5 ns following this conformational change did not result in a return to the U37-sequestered conformation (Figure 3.5). By contrast, in over 100 ns of U39 simulations, the U37-exposed conformation is never observed to adopt the U37-sequestered conformation, nor is the U37-exposed conformation observed in any of our C39 simulations. These results provide evidence, within the framework of a necessarily limited set of simulations, that the U39 wild-type sequence can transition from the U37-sequestered to the U37-exposed conformation within a single docking event, whereas the C39 mutation suppresses this conformational change, consistent with the observed crystallographic heterogeneity of the U39 wild-type and its absence in the C39 gain-of-function mutant (17). It has been speculated that the enhanced conformational stability of the adjacent S-turn leads to, at least in part, the higher metal ion binding affinity and catalytic activity of the C39 mutant (17). Our data support this notion and further implicate increased docking stability as a source of catalytic enhancement of the C39 mutant.

MD simulations properly predict crystallographically resolved water molecules that are part of an extended inter-domain hydrogen bonding network

Consistent with previous X-ray crystallography (16,17) and MD studies (49), we observe long residency water molecules in an inter-domain cavity near the catalytic core (Figure 3.6). The two most stable structural water molecules are observed at the active site with residency times longer than can be characterized by our typically 20 ns simulations, while an additional set of 3-5 water molecules is present and remains relatively mobile within the inter-domain cavity, but only slowly exchanges with bulk solvent (Figure 3.6). All of these water molecules are an integral part of an extensive inter-domain hydrogen bonding network (Figure 3.6A), and notably are most stable in simulations of the C39 sequence that experimentally is most active and yields the crystal structures of highest resolution in which water is clearly identified (16). The behavior of explicit water in our simulations is therefore generally consistent with the crystal structures and previous MD simulations,

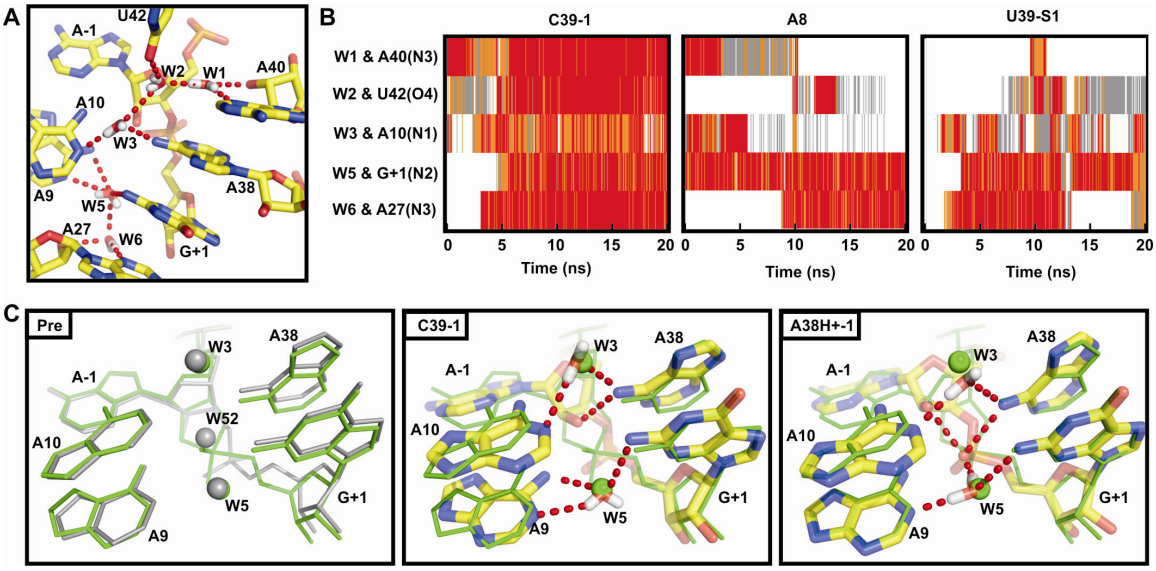


Figure 3.6: Long residency inter-domain waters are an integral part of the inter-domain hydrogen bonding network. (**A**) The positions of inter-domain waters typical of our simulations are shown. (**B**) Inter-domain waters were identified 10 ns into the simulation and inter-atomic distances of hydrogen bonding partners were tracked over the entire length of three representative simulations. Distances are color coded as in Figure 1E. (**C**) The crystal structure of the non-cleavable ribozyme-substrate analog complex (Pre, gray) (2OUE) as well as snapshots from two simulations (C39-1 and A38H+-1, color) overlaid with a vanadate TSA crystal structure (green) (2P7E). In the TSA crystal structure, a crystallographically resolved active-site water molecule in the precursor structure (W52) is displaced by the non-bridging pro-S oxygen, while flanking waters (W3 and W5) remain in nearly identical positions. In contrast, during simulations in which A38 is unprotonated (such as C39-1) the 2'-OH of A-1 takes the place of W52, while flanking waters remain in nearly identical positions. Similarly to the TSA structure, during simulations in which A38 is protonated (such as A38H+-1) the non-bridging pro-S oxygen takes the place of W52, while flanking waters remain in nearly identical positions.

providing collective evidence for an essential structural and potentially catalytic role for long residency water molecules in the hairpin ribozyme catalytic core.

More specifically, water molecules W3 and W5 in our simulations occupy nearly identical positions to water molecules resolved in crystal structures of vanadate transition state analog and non-cleavable ribozyme-substrate analog complex (2P7E and 2OUE) (Figure 3.6C). The stable binding of W5 bridging A9 and G+1 throughout our simulations regardless of protonation state of A38 and the resulting stable hydrogen bond donated to the pro-R non-bridging oxygen upon protonation of A38 support the notion of a role of structural water molecules in stabilizing the developing negative charge of the scissile phosphate in the transition state (16). In contrast, W52 is present in the non-cleavable ribozyme-substrate analog complex, but absent in the vanadate transition state analog and, accordingly, is lost during our simulations (Figure 3.6C). In our simulations the site occupied by W52 is instead occupied by A-1(2'OH) when A38 in unprotonated, and by the pro-R non-bridging oxygen in the protonated A38H+ simulations. The fact that the same water molecule (W52) is also experimentally observed to be also replaced by the same pro-R non-bridging oxygen in crystal structures of both vanadate and 2',5'-phosphodiester transition state analogs (TSA's) (2P7E and 2P7F) links A38 protonation to the ability to access the transition state geometry.

To test the impact of initial water placement on our MD simulations we performed five simulations in which the crystallographic water molecules were removed and all water was placed using the LEAP module of AMBER-8; these simulations began without any inter-cavity water (Table 1). During the course of each simulation water moved into the inter-domain cavity. Differences between simulations were most pronounced at early time points with behavior converging as the simulations progressed, thus reflecting a longer equilibration period as observed previously (49). Accordingly, we noted a tendency for a sugar pucker flip at A-1 (discussed in detail below) to occur more slowly when experimentally resolved waters were excluded from the starting structure.

MD simulations reproduce experimentally observed metal ion binding sites

Generally, we observe good agreement between the available experimental data and the behavior of metal ions in our MD simulations, as previously noted (49). For example, we

find the highest Na^+ ion density at two sites with in the E-loop (E1 and E2) of loop B, along the major groove of domain A (M), and in some simulations at the S-turn (S) in domain B (Figure 3.7A). The Na^+ density in the E-loop corresponds to two crystallographically observed calcium binding sites in the four-way junction crystal structures (Figure 3.7B) (30) and, consequently, is strongly competed during our MD simulations upon placement of a Mg^{2+} ion into the E-loop (Figure 3.7C). Notably, site E2 is only observed when the IAA is favorable and is lost upon formation of the hydrogen bond from A-1(2'OH) to A38(N1) discussed below. When present, monovalent ion density observed at the S-turn (Figure 3.7A) corresponds to a high-affinity metal binding site identified through footprinting (63). In our A8 simulations (Table 1) we observe additional high Na^+ density in close proximity to the non-bridging oxygens of the scissile phosphate (Figure 3.9A).

Most of our simulations were carried out in charge-neutralizing monovalent cations alone due to the limitations in the parameters used for divalent ions (33) and the catalytic proficiency of the hairpin ribozyme in monovalents alone (49). In four of our simulations we placed a single Mg^{2+} ion at a well-characterized multivalent outer-sphere metal ion binding site within the E-loop motif of Loop B (Figure 3.7C) (17), and again charge-neutralized with Na^+ ions (Table 1). We do not observe significant differences in either global stability or active site architecture upon inclusion of this Mg^{2+}, consistent with observations from previous simulations (49). We only note that the outer-sphere bound Mg^{2+} ion quickly loses a water molecule of its inner coordination sphere and then directly chelates the RNA at U41 (Figure 3.7C), consistent with the known bias towards inner-shell binding of Mg^{2+} ions in RNA simulations due to force field approximations (64).

A 2'-OH substitution at the cleavage site restores an active site architecture perturbed in 2'-O-methylated crystal structures

The starting coordinates for our MD simulations were derived from high-resolution crystal structures of the minimal hairpin ribozyme in complex with a non-cleavable substrate analog, solved with a cleavage blocking 2'-O-methyl substitution in place of the nucleophilic 2'-OH of residue A-1 (16,17). For our simulations we reverted back to the

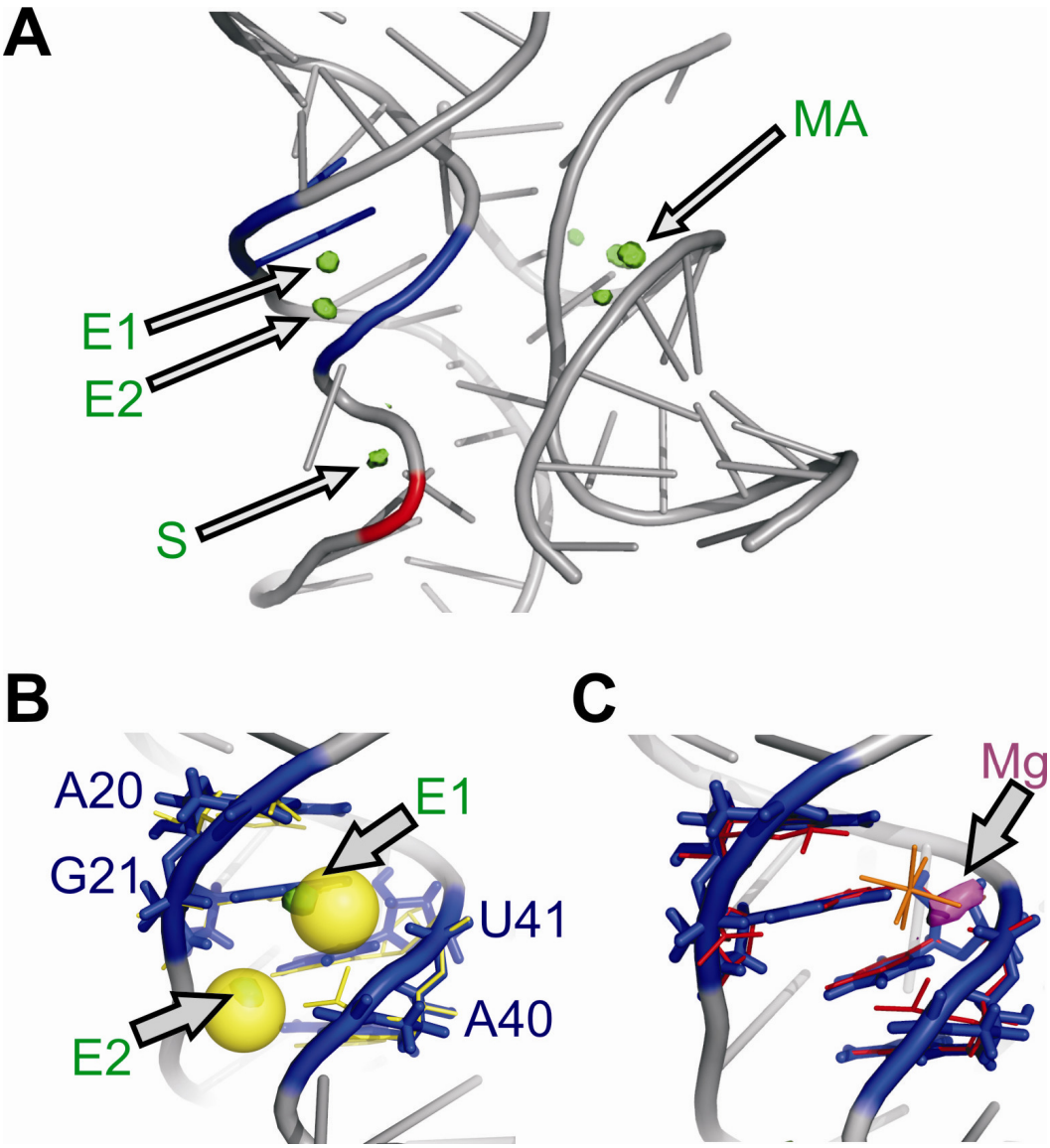


Figure 3.7: Monovalent ion behavior shows agreement with experimentally determined metal ion binding sites. (**A**) Regions of high Na^+ density observed in our simulations are shown in green. High density is observed in the major grove MA and in the E-loop at site E1 in domain B in all of our simulations. High density is observed at site E2 only when the IAA is favorable. The crystallographically observed cobalt(III) hexammine binding site (blue) (2OUE) and an ion binding site observed through solution footprinting (red) (63) are indicated for comparison. High density at site S is only occasionally observed in our simulations. (**B**) Sites E1 and E2 in the E-loop are overlaid with the corresponding calcium binding sites (yellow sticks) and the crystallogaphically resolved calcium ions (transparent yellow spheres)) from the four-way junction hairpin ribozyme precursor crystal structure (1M5K) (18). (**C**) The region of high Mg^{2+} density is shown in magenta, high sodium density is not observed at this site when Mg^{2+} is present. The Mg^{2+} binding site is overlaid with the corresponding cobalt(III) hexammine binding site (red) and the crystallographically resolved cobalt(III) hexamine (orange) from the crystal structure of the junctionless non-cleavable ribozyme-substrate analog complex (2OUE).

native 2’-OH, allowing us to test for any impact that the 2’-O-methyl substitution may have on active site architecture. In all but one of our simulations, we observe a flip in the A-1 sugar pucker from the 2'-endo/3'-exo of the ribozyme-substrate analog complex to a 2'-exo/3'-endo conformation (Figure 3.8A), usually within the first few nanoseconds of the simulation and consistent with our previous observations (49). An analysis of a total of 16 MD simulations (including four previously reported simulations (49)) yields a single-exponentially distributed rate constant for this irreversible conformational change of $k_{flip} = 0.3 \times 10^{-9}$ s^{-1} (Figure 3.8B). The change in sugar pucker is accommodated by a slight reorientation of the whole nucleotide and frequently results in a hydrogen bond from the A-1(2’-OH) to A38(N1) (Figure 3.8A). The formation of this additional, partially occupied (Figure 3.9B and 3.10), inter-domain hydrogen bond is consistent with the increased stability of the cleavable ribozyme-substrate complex (containing A-1(2’-OH)) relative to the non-cleavable ribozyme-substrate analog complex (containing A-1(2’-O-methyl)) as observed by single-molecule FRET (1.8 kcal/mol) (65). This hydrogen bond is precluded from forming in the crystal structure due to the presence of the 2’-O-methyl modification, which cannot donate a hydrogen bond and is not easily accommodated sterically by the active site when the A-1 sugar adopts the MD simulated 2'-exo/3'-endo conformation.

Notably, four high-resolution precursor crystal structures containing the native 2’-OH together with inactivating mutations exist (16), none of which exhibit the 2'-endo/3'-exo sugar pucker observed in the ribozyme-substrate analog crystal structures. Two of these four structures adopt the 2'-exo/3'-endo conformation observed in our simulations while the other two are assigned an unusual planar sugar ring (possibly representing an average of 2’-endo and 3’-endo sugar puckers), thus further supporting our proposal that a 2’-O-methyl modification of A-1 distorts the backbone geometry in the catalytic core by suppressing the favorable 2'-exo/3'-endo sugar pucker of A-1.

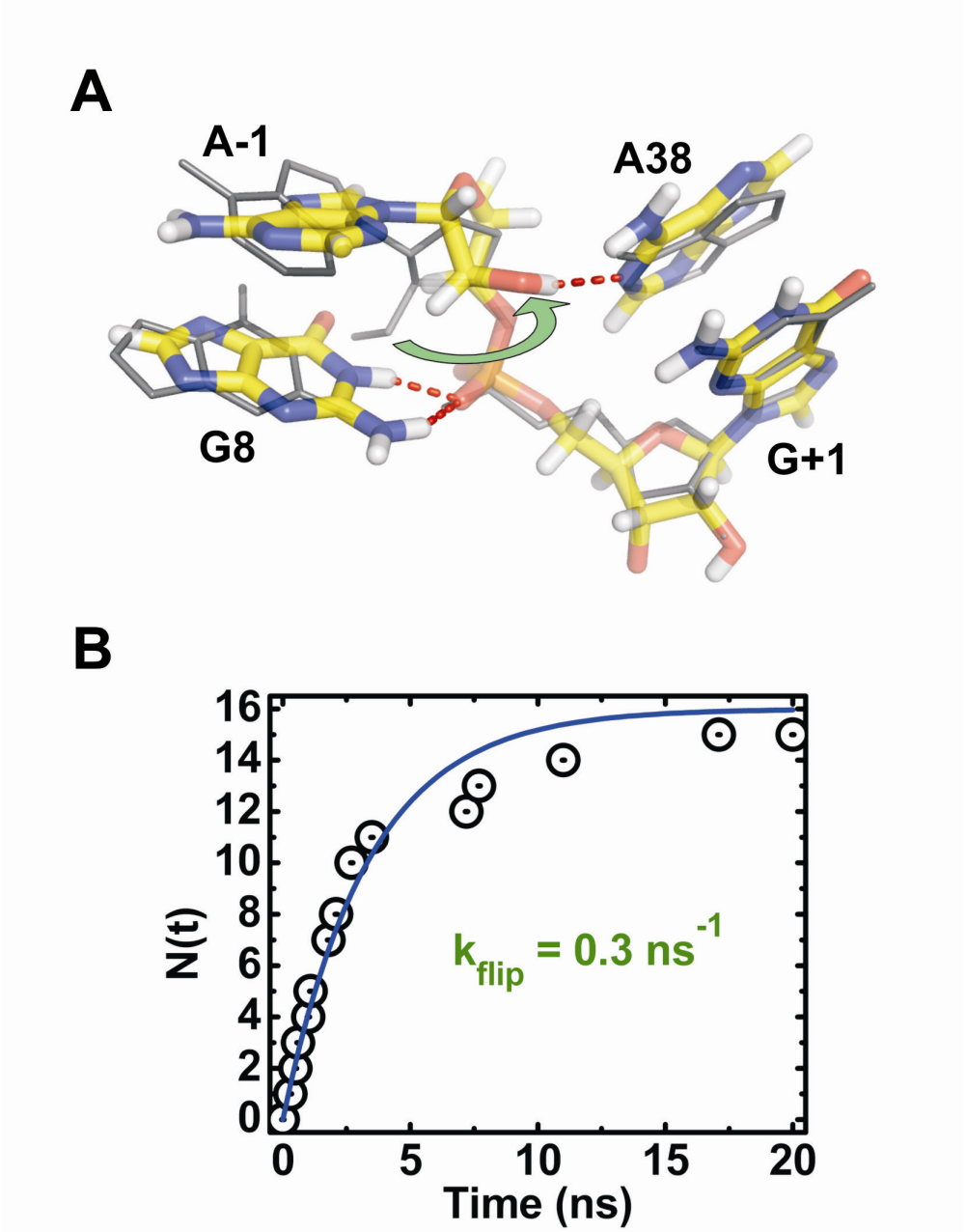


Figure 3.8: Replacement of the inactivating A-1(2'-O-methyl) modification with the native A-1(2'OH) results in a local, functionally relevant conformational change. (**A**) Active site geometry of a C39 gain-of-function mutant simulation (color) overlaid with the non-cleavable ribozyme-substrate analog complex crystal structure (PDB ID 2OUE, gray) illustrates a change in sugar pucker of A-1 that results in a hydrogen bond from A-1(2'OH) to A38(N1) and hydrogen bonds from G8 to the non-bridging pro-R oxygen (red dashed tubes). (B) The number of simulations that change sugar pucker after a given time (open circles) is well fit with a single-exponential (line).

The simulated active site conformation rationalizes the experimentally observed higher importance of A38 compared to G8 in catalysis

The prevailing conformational change in our simulations toward a 2'-exo/3'-endo A-1 sugar pucker moves the 2'-OH away from G8, thus eliminating the intra-domain G8(N1) to A-1(2'O) hydrogen bond observed in the crystal structures (Figures 3.8A, 3.9A, and 3.11A). This finding contradicts proposals that G8 either acts as a general base, or plays a direct role in the positioning and pK_a suppression of the 2'-OH of A-1 for catalysis (16,18,19,30). Instead, the observed new hydrogen bond from A-1(2'-OH) to A38(N1) supports the notions that A38 plays the role of general base and that hydrogen bonds from A38(N6) and A10(N6) to the A-1(2'-OH) provide an alternative means of pK_a suppression of the 2'-OH. This emphasis of A38 over G8 is consistent with the significantly larger catalytic rate decrease (>10,000-fold) measured upon mutation or deletion of A38 compared to that observed upon mutation or deletion of G8 (only 100- to 1,000-fold) (23,25,27,30), as well as the particularly critical role observed for the Watson-Crick face of A38 in shaping cleavage site architecture (20,21).

In simulations with a charge neutral (unprotonated) A38(N1) we find that the 2'-OH of A-1 is the most common hydrogen bond donor to A38(N1), implicating A38 as a general base in the cleavage reaction. We note, however, that this hydrogen bond is observed only when the IAA of 2'O-P'-5'O is quite unfavorable (< 130^o or even lower, Figure 3.9 and 3.10), far removed from the 180^o thought to be optimal for nucleophilic substitution (66,67), or the 162^o predicted in the transition state for model reactions of RNA backbone cleavage (68). A conformational change would thus be needed upon deprotonation of A-1(2'-OH) for cleavage to proceed, a notion that we address below. In addition to the prevalent hydrogen bond to A38(N1), we find that A-1(2'-OH) alternatively donates a hydrogen to the non-bridging pro-R oxygen, or to a long residency inter-domain water bridging the A-1(2'-OH) and A38(N1) (Figure 3.9 and 3.10). These two additional hydrogen bond acceptors have also been proposed to play a role in catalysis based on earlier simulations (49,69,70).

Coincident with the change in A-1 sugar pucker, G8(N1) moves within hydrogen bonding distance of the non-bridging phosphate oxygens, specifically the pro-R oxygen, thus supporting the previous proposal that G8 stabilizes the increased negative charge on

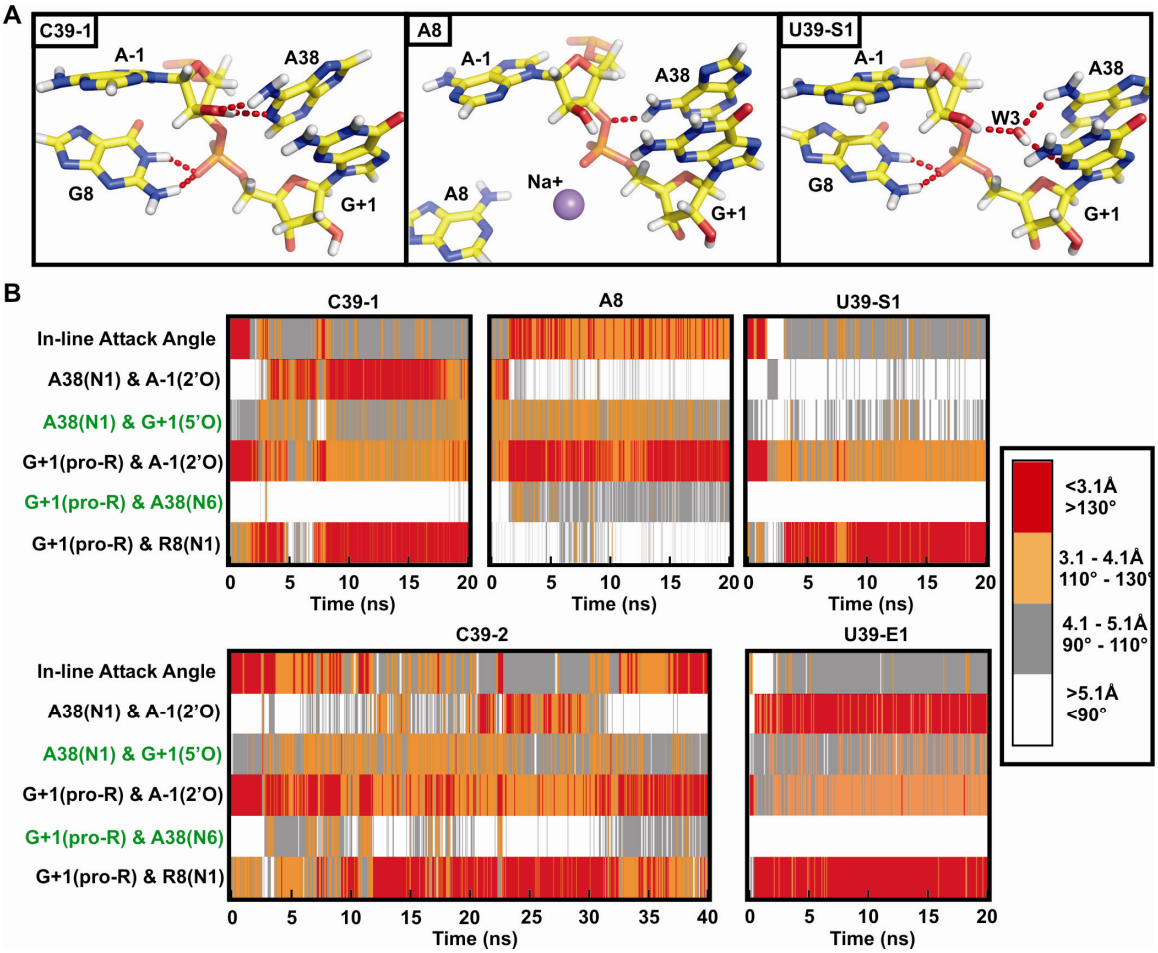


Figure 3.9: Active site dynamics provide insight into plausible roles of the functionally critical active site nucleobases G8 and A38. (**A**) Snapshots of the active site from the end of the three indicated simulations as indicated (averaged from 19.9 to 20 ns) in which N1 of A38 is unprotonated. Key hydrogen bonds are indicated by red dash tubes. (**B**) In-line attack angle and key inter-atomic distances over the course of each simulation, with distances and angles color-coded as indicated in the legend. Close proximity of A-1(2'OH) and A38(N1) is frequently observed for ribozymes known to be catalytically active (C39 and U39) and is correlated with an unfavorable in-line attack angle. Hydrogen bonds observed in TSA crystal structures (2P7E, 2P7F and 1M50, indicated in green) are generally not observed in these simulations.

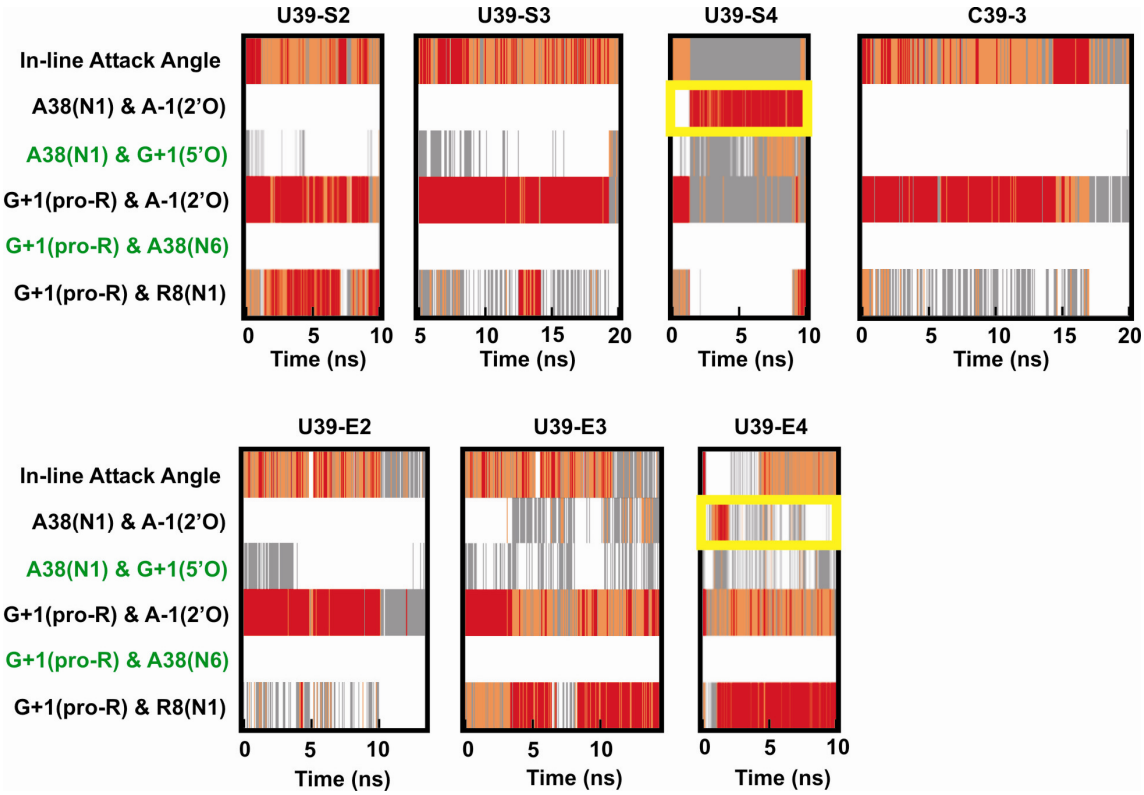


Figure 3.10: In-line attack angle and important inter-atomic distances are tracked over the course of all additional simulations not shown in Figure 3.9; distances and angles are color-coded as in Figure 3.9. Hydrogen bonding partners observed in TSA crystal structures are indicated in green, but generally not observed to form in these simulations. Simulations in which the A-1(2'OH) to A38(N1) hydrogen bond is observed are indicated by yellow boxes.

the scissile phosphate in the transition state (27,29,40). The 5'-oxygen leaving group is hydrogen bond acceptor to several transiently bound water molecules from bulk solvent that could potentially serve as specific acid catalysts in the cleavage reaction.

Protonation of A38 drives the active site toward the expected transition-state geometry

The significant biochemical and structural evidence for a shift in the pK_a of A38 in the context of the docked ribozyme (15,18,23,25) prompted us to carry out simulations in which N1 of A38 is protonated (A38H+), complementing our simulations with the charge neutral (unprotonated) A38 (Table 1). In striking contrast to the charge neutral A38 simulations, A38($N1H^+$) now primarily donates a hydrogen bond to the 5'-oxygen of G+1, the cleavage reaction's leaving group (and, to an extent, to the 3'-oxygen, Figure 3.11), placing A38 in a position to serve as the general acid during catalysis. Furthermore, we observe that proximity of A38(N1) and G+1(5'O) frequently coincides with a favorable (> 130^o) IAA (Figure 3.11). G8 moves again within hydrogen bonding distance of the non-bridging scissile phosphate oxygens, but now we observe alternating proximity to both the pro-S and the pro-R oxygens, with proximity to the pro-S oxygen strongly correlated with the ribozyme adopting a favorable IAA. These observations further support a primary role of G8 in stabilizing the negative charge of the scissile phosphate, rather than as a general base catalyst. Additionally, the non-bridging pro-R oxygen is the most common hydrogen bond acceptor of the A-1(2'-OH), consistent with the idea that it may be a suitable general base catalyst (69,70).

The hydrogen bond from A38($N1H^+$) to G+1(5'O) leaving group of the cleavage site as well as adoption of favorable in-line attack angles coincide with a rotation of the scissile phosphate, thus maximizing the number of intramolecular hydrogen bonds to the non-bridging oxygens, as expected for transition state stabilization. In particular, the scissile phosphate rotates away from the precursor structure resulting in the pro-R oxygen accepting alternating hydrogen bonds from the exocyclic amines of A9 and A38, A-1(2'OH), and the highly stable specifically bound water molecule W5, while the pro-S oxygen accepts a bifurcated hydrogen bond from N1 and N3 of G8 (Figures 3.6C and 3.11). Taken together, this dynamic conformation results in a hydrogen bonding network very similar to that observed in crystal structures of transition state analogs (TSAs)

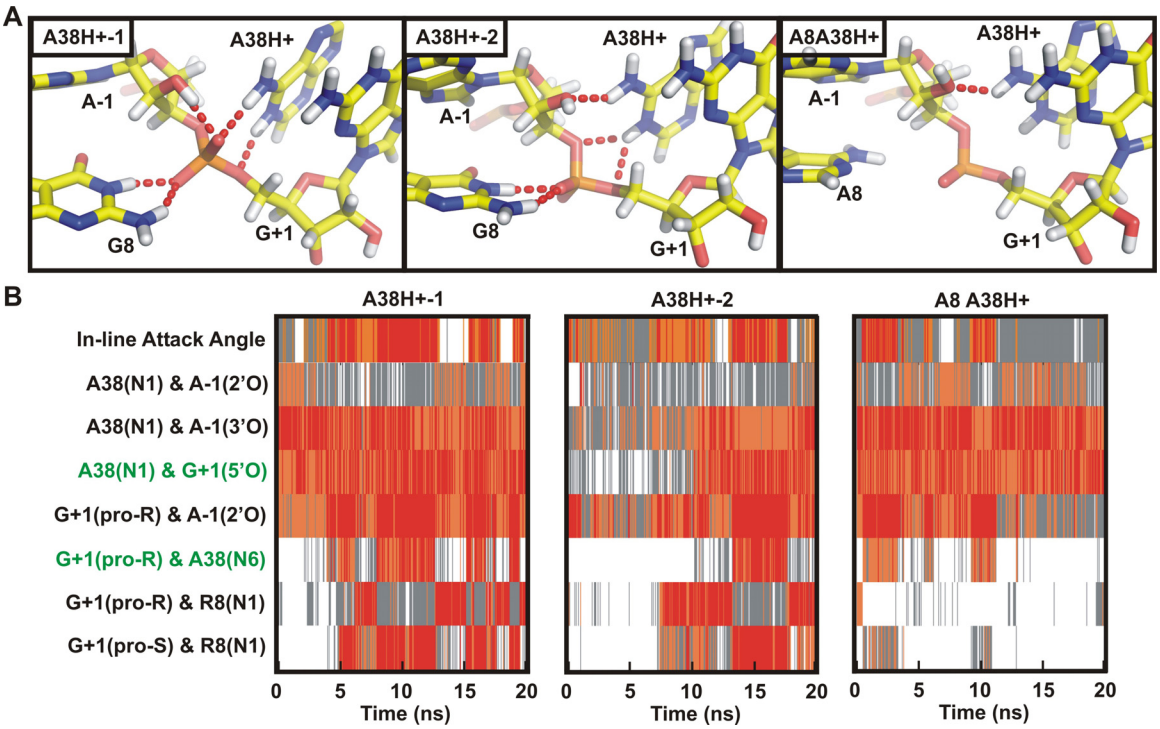


Figure 3.11: Protonation of A38(N1) results in a conformational rearrangement around the active site that establishes structural features predicted to be present within the transition state. (**A**) Snapshots of the active site from three simulations (averaged either from 9.9 to 10 ns for A38H+-1 or from 19.9 to 20 ns for A38H+-2 and A8A38H+) in which N1 of A38 is protonated. Key hydrogen bonds are indicated by red dash tubes. (**B**) In-line attack angle and important inter-atomic distances are tracked over the course of the simulation, with distances and angles color-coded as in Fgure 4. Proximity of the cleavage site G+1(5'O) and A38(N1) is frequently observed in all three simulations. A favorable IAA coincides with rotation of the non-bridging oxygens such that of G8(N1) loses its hydrogen bonding with the pro-R non-bridging oxygen in favor of a hydrogen bond with the pro-S non-bridging oxygen, as illustrated in the A38H+-1 snapshot of panel A. Hydrogen bonding partners observed in TSA crystal structures (2P7E, 2P7F and 1M50, indicated in green) are more frequently observed in these simulations.

(Figures 3.6C and 3.11) (15,18). While this TSA-like conformation is observed ~20% of the time in A38H+ simulations, it is not present in any of our simulations with a neutral A38, further supporting the catalytic relevance of the molecular dynamics observed in our A38H+ simulations.

Active site nucleobases stabilize strained backbone conformations

Not surprisingly, examination of the backbone conformations in our simulations and the starting crystal structures reveals that the cleavage site does not conform to any of the established RNA backbone conformational families (Figure 3.3) (58). Specifically the epsilon (C4'-C3'-O-P) dihedral is significantly smaller than that typically observed experimentally (58). In our MD simulations a decreased epsilon angle correlates with an increased IAA (Figure 3.12A), and inspection of multiple hairpin ribozyme crystal structures reveals the same correlation in experimental data (Figure 3.12). In order to determine how general this trend is in backbone cleavage, we evaluated all naturally occurring self-cleaving ribozymes for which high resolution crystal structures are available (hairpin, hammerhead, HDV, and *glmS* ribozymes) and find the same trend in all four ribozymes (Figure 3.12A). The general trend of lower epsilon angles accompanying the high IAA angles required for catalysis suggests that distortion of this dihedral plays an important role in defining the cleavage site in self-cleaving RNA. Importantly, protonation of A38 results in a broader range of epsilon values at the cleavage site during our simulations (Figure 3.12B), resulting in sampling of both lower epsilon and higher IAA angles than when A38 is not protonated (Figure 3.12A). Regardless of A38 protonation state the distributions of epsilon angles at the cleavage site favor smaller values than all other regions of the backbone, consistent with the distribution range of epsilon angles that is experimentally observed at the cleavage site of self-cleaving RNAs (Figure 3.12B).

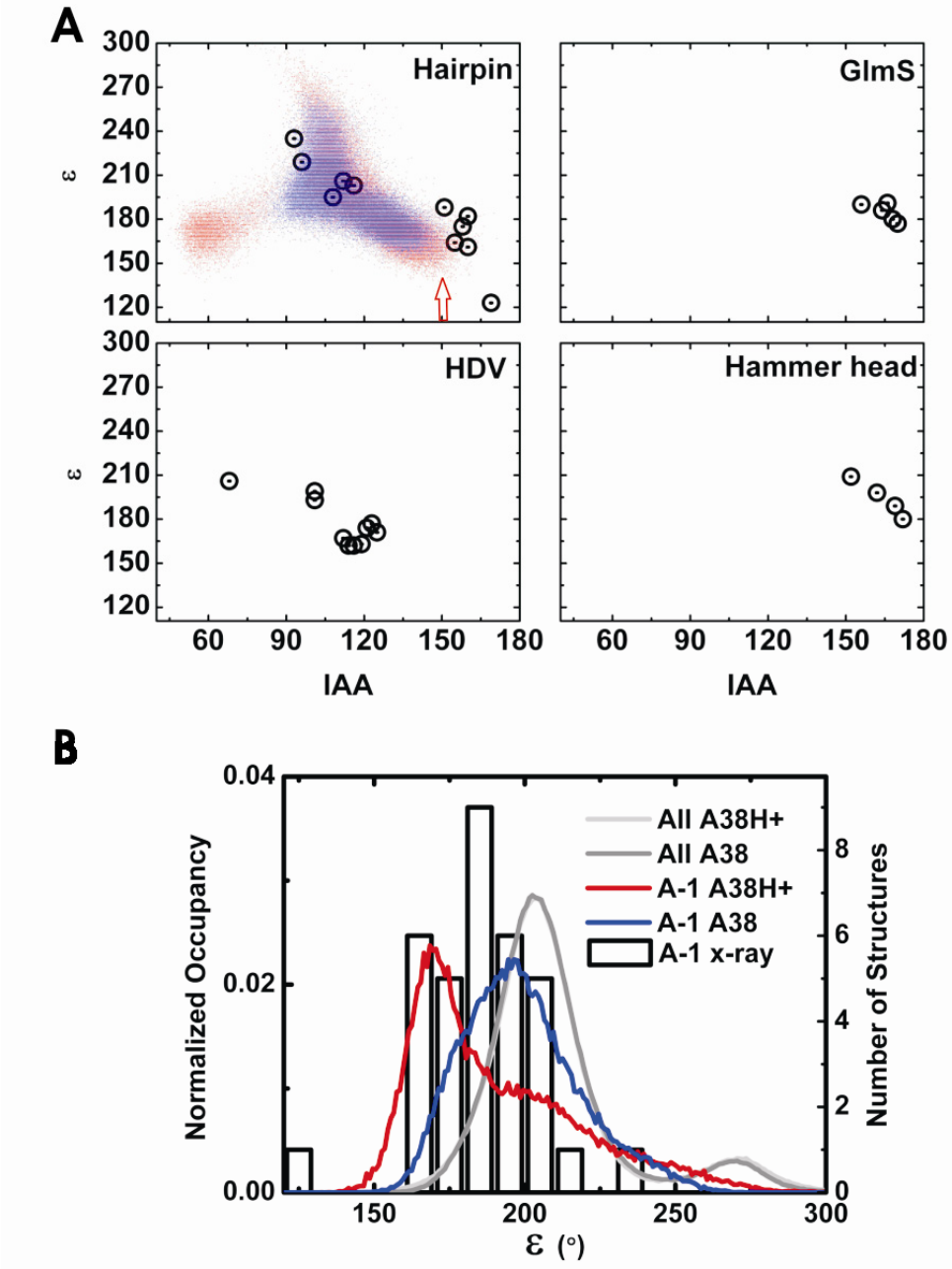

A
Hairpin
GlmS
HDV
Hammer head
ε
IAA
300
270
240
210
180
150
120
60
90
120
150
180
B
Normalized Occupancy
Number of Structures
ε (°)
All A38H+
All A38
A-1 A38H+
A-1 A38
A-1 x-ray
0.04
0.02
0.00
0
2
4
6
8
150
200
250
300

Figure 3.12: The cleavage site dynamics of the RNA backbone of our simulations share conformational features common among the small self-cleaving RNAs. (**A**) Cleavage site dihedral angle epsilon as a function of the IAA for all crystal structures with <3.1 Å resolution of the hairpin ribozyme, the glmS ribozyme bound to glucose amine 6-phosphate, the HDV ribozyme, and the full-length hammerhead ribozyme (open circles). Epsilon as a function of IAA is also shown for the A38H+ simulations (red dots) as well as the C39-1 and C39-2 simulations (blue dots). The A38H+ simulations sample a wider range of values in these plots; notably, the region of both lowest epsilon and highest IAA (red arrow) is more frequently sampled in the A38H+ simulations. (**B**) Probability densities of the cleavage site (A-1) epsilon angles as well as all backbone epsilon angles in our simulations (left axis and lines) and cleavage site (A-1) epsilon distributions among the available ribozyme crystal structures (right axis and open boxes). The epsilon angles at the cleavage site during our simulations and in the crystal structures both favor low values relative to the distribution observed in our simulations for all epsilon angles in the entire backbone. Moreover, simulations in which A38 is protonated (red line) favor even lower epsilon values than simulations where A38 is unprotonated (blue line).

3.4 Discussion

A number of catalytic mechanisms have been proposed for the hairpin ribozyme. Cleavage is generally thought to proceed through a mechanism analogous to that carried out by RNase A where two histidine residues of near-neutral pK_a act as separate base and acid catalysts and a protonated lysine residue neutralizes the increased negative charge of the scissile transition state phosphate (40,41,71,72). While the solution pK_a's of RNA functional groups are far from neutrality, there is substantial evidence for pK_a up-shifting in the context of the strong, long-range, negative electrostatic field of structured RNA (15,25,60,73,74). Biochemical and computational approaches have suggested that in the catalytically active hairpin ribozyme the pK_a's of one or all of the adenosines A10 (60,75), A22 (60), and A38 (15,18,25,60) are significantly perturbed toward neutrality . Our MD simulations indicate that A38 is positioned to concomitantly fulfill multiple catalytic functions potentially including general base catalysis (when unprotonated), transition state stabilization, and general acid catalysis (once protonated), consistent with the extraordinarily detrimental impact its replacement with an abasic site has on catalytic activity (25).

MD simulations are generally known to have significant limitations in the attainable sampling (i.e., short simulated time spans) and underlying force fields (33), however, they uniquely reveal detailed dynamics at an atomistic level that allow one to test and derive structure-based mechanistic hypotheses. Overall, we find surprisingly good agreement between the available experimental data and our multiple, independent, explicit-solvent MD simulations of the hairpin ribozyme (Table 1), suggesting that the AMBER-8 (parm99 Cornell et al.) force field performs well in such an application. In particular, the backbone conformations generally remain consistent with those observed in the underlying crystal structures or make them more canonical; the C39 gain-of function mutant is correctly predicted to dock more stably and to show less conformational heterogeneity than the U39 wild-type (Figure 3.2); and our simulations reproduce the binding sites of crystallographically resolved water molecules and cations (Figures 3.6 and 3.7). We therefore focus our discussion on aspects of our MD

simulations that provide unexpected insights into structural dynamics and catalysis of the hairpin ribozyme.

Crystallographically observed atomic distances have been used to propose that A38 and G8 function as the general acid and general base catalysts, respectively, in hairpin ribozyme catalysis (15-19,21,39). However, our simulations reveal relevant local distortions introduced into the crystal structures by the presence of a catalysis blocking 2'-O-methyl modification, and suggest alternative roles for these nucleobases in backbone cleavage. Most strikingly, the stable A-1(O2'H) to A38(N1) hydrogen bond, masked in crystal structures with A-1(2'-O-methyl) modification but stably observed in our MD simulations (Figure 3.8), opens the possibility that A38 acts as the general base. If instead G8 were the general base, as proposed previously (19,30,39,76,77), the catalytically relevant form of G8 would need to be deprotonated at position N1 (or be in a rare enol tautomeric form). However, exogenous nucleobase rescue experiments suggest that it is the N1 protonated (charge-neutral) state that activates catalysis (27,29). Furthermore, based on comparison of crystal structures of the U39C and G8A mutants it has been suggested that the role of G8 is to facilitate a favorable IAA by donating a hydrogen bond from G8(N1) and/or G8(N2) to A-1(O2'), and to lower the pK_a of the 2'-OH (16). However, this hydrogen bond is lost in our simulations and we observe a hydrogen bond with the non-bridging oxygens instead. This is consistent with the observation that activity of a ribozyme with an abasic substitution at position 8 is rescued by a cationic exogenous nucleobase (27). In addition, deprotonation of G8(N1) seems unlikely considering the pocket of deep negative electrostatic potential at the catalytic center of the ribozyme, whereas participation of the G8 tautomer is inconsistent with QM/MM simulations (69). Overall, our data support a catalytic role of G8 in facilitating the reactive conformation through its interaction with the non-bridging oxygens and in stabilizing negative charge in the transition state, but not in general base catalysis. This leaves A38 as a prime suspect to remove the proton of A-1(O2'H). The exocyclic amines of A10 and A38 donate hydrogen bonds to the 2'-OH in our simulations which, similar to the role previously proposed for G8, should help lower the pK_a of the 2'-OH.

Alternatively, recent QM/MM simulations implicate the non-bridging scissile phosphate oxygens as reasonable general base candidates in hairpin ribozyme self-

cleavage (69,70,77). The frequent hydrogen bonds between the A-1(2'OH) and the non-bridging oxygens observed in our MD simulations are consistent with such a role (Figures 3.9 and 3.11). Active site nucleobases such as G8, A9, A10, or A38 are in a position to potentially participate indirectly in catalysis by promoting proton transfer to these non-bridging oxygens (Figures 3.9 and 3.11).

The proximity of A38(N1) and G+1(O5') in several crystal structures of reaction precursor, TSA, and cleaved forms lend strong support to the proposal that A38 acts as general acid in the cleavage reaction (15-19,21), and our simulations further emphasize the stability of this juxtaposition within hydrogen bonding distance, despite rearrangement of the A-1 sugar pucker. Biochemical evidence also suggests that A38 plays a vital role in stabilizing the catalytically productive active site architecture in a way that is dependent upon the protonation state of N1 (23,25). Strikingly, our simulations begin with coordinates from an inactivated precursor but, over the course of simulations in which A38(N1) is protonated, undergo significant local structural changes in the active site to adopt characteristics exhibited by the experimentally determined TSAs (Figures. 3.6C and 3.11) (15,18). In contrast, in simulations with an unprotonated A38(N1) a similar transition of the non-bridging oxygen to a TSA-like conformation is not observed (Figure 3.9 and 3.10). Moreover, in simulations with the inactivating G8A mutation the transition is far less pronounced regardless of the A38 protonation state (Figures 3.9 and 3.11).

Taken together, our observations support a vital structural role of A38 protonation in stabilizing the catalytic geometry and provide further support for A38 as the general acid catalyst. The structural importance of A38 may explain why previous QM/MM calculations that imposed constraints on the hairpin ribozyme's active site geometry that were external to the standard force field unexpectedly predicted a similar free energy barrier to catalysis regardless of the protonation state of A38 (69), despite strong experimental evidence to the contrary (23,25). Our results indicate that maintenance of a favorable IAA requires protonation of A38 so that the protonation state of A38 is expected to significantly impact the overall energy barrier to catalysis in a way that can only be fully assessed if the dynamic ground state structure is kept unrestrained. The greater consistency between the conformations in our A38H+ simulations and the crystal

structures further suggests that the protonated A38H^+ is the predominant ionization state at the crystallographic pH of 6.0, consistent with the notion that its pK_a is perturbed from its solution value of 3.5. It should be noted, however, that the ribozyme is most active at even higher pH (with a $pK_{a,app}$ for cleavage of 6.1 that varies among mutants with the identity of the functional group in the position that A38(N1) occupies in the wild-type (25)); our simulations therefore may not reflect on the predominant protonation state at optimal cleavage conditions.

Our MD simulations depict the unprotonated A38(N1) as receiving a hydrogen bond from A-1(O2'H) (Figure 3.8A, 3.9 and 3.10), whereas the protonated A38($N1H^+$) donates a hydrogen bond to the 5'-oxygen (Figure 3,11). A recent TSA crystal structure containing a 2',5'-linkage at the cleavage site exhibits an apparently bifurcated hydrogen bond between N1 of A38 and both the 2'- and 5'-oxygens of the cleavage site (15), supporting the mechanistic relevance of our observations of both hydrogen bonds in MD simulations. Several mechanisms are consistent with these observations (Figure 3.13). First, as previously proposed A38 may serve as the general acid, with a second functional group such as A9, A10 or the non-bridging oxygens acting as general base (Figure 3.13A). In the context of this mechanism, the A-1(2'OH) to A38(N1) hydrogen bond and associated poor IAA may be attributed to an unproductive ground state, wherein a perturbed pK_a allows A38 to accept a solvent proton and subsequently aid adoption of the catalytically active conformation. Second, we propose a mechanism in which A38 acts bifunctionally as both general base and acid by shuttling a proton directly from A-1(O2'H) to G+1(O5') (Figure 3.13B). Such a mechanism is analogous to the role of, for example, an active site Glu in the first step of the reaction carried out by serine-carboxyl peptidases (Figure 3.13C) (78,79). Here, the A-1(2'OH) to A38(N1) hydrogen bond would be essential to the mechanism; following transfer of the proton from A-1(2'OH) to A38(N1) the protonated A38H+ again promotes formation of the active geometry. This mechanism requires a certain degree of conformational flexibility in that proton exchange between 2'-OH and A38(N1) must be followed by a rapid transition of the deprotonated 2'-oxyanion to a position more favorable to in-line attack, as well as by a ~0.4 Å motion of the now protonated A38($N1H^+$) toward the 5'-oxygen (Figure 8B). We can infer that such dynamics are indeed possible as we observe that the protonated A38($N1H^+$)

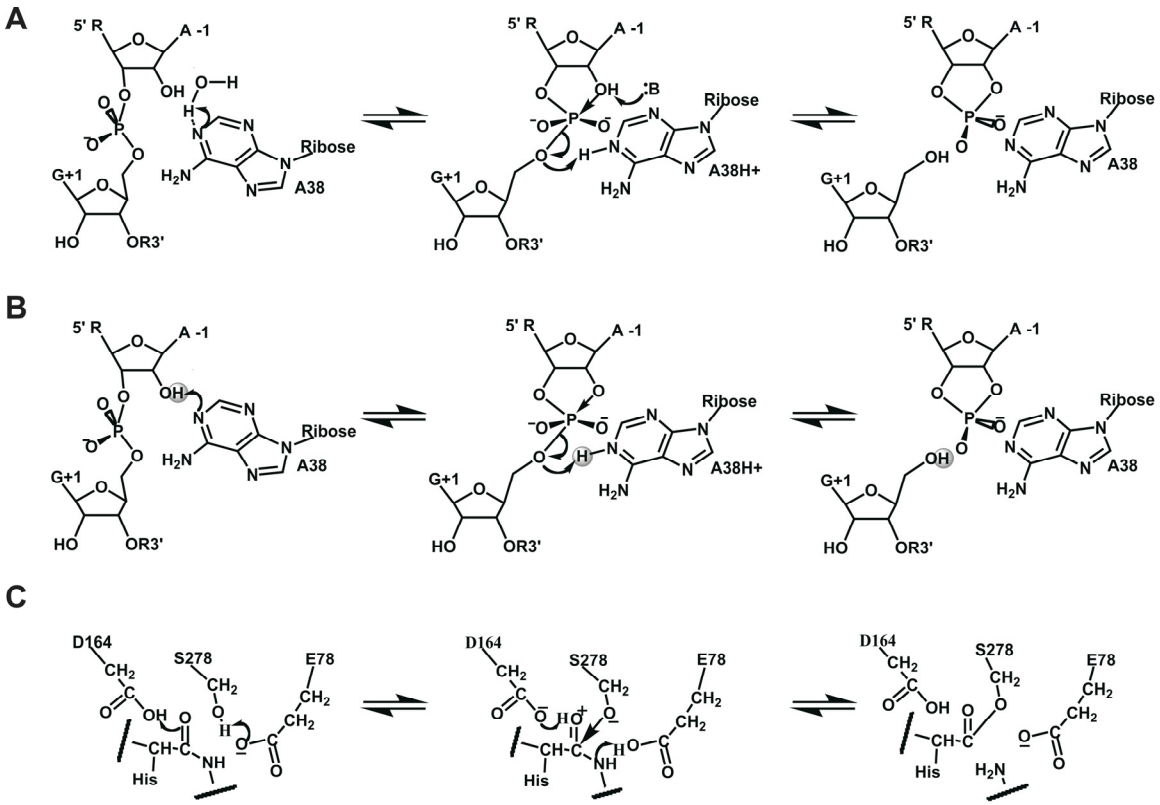


Figure 3.13: Plausible mechanisms for the reversible cleavage/ligation by the hairpin ribozyme consistent with the active site dynamics observed during MD. (**A**) Mechanism in which protonation of A38 is required to drive the active site toward the catalytic geometry and the protonated A38H+ acts as the general acid, while A9, A10, a water, or a non-bridging oxygen could serve as the base in the cleavage reaction. (**B**) Mechanism in which A38 acts as both general base and acid by shuttling a proton (grey sphere) directly from the 2'-OH of A-1 to the 5'-oxygen of G+1 for the cleavage reaction. In both mechanisms G8, A10, A38 and water 5 stabilize the negative charge accumulated in the transition state. (**C**) Mechanism proposed for the protein-based serine-carboxyl peptidase kumamolisin-As (78,79), analogous to the mechanism in panel B. A glutamate side chain with acidic solution pK_a (4.3) acts as both general base and acid by shuttling a proton directly from the serine nucleophile with solution pK_a of (~13) to the leaving group during formation of the acylated enzyme.

occasionally approaches the 2'-OH, while the unprotonated A38(N1) makes rapid excursions toward the 5'-oxygen on a picosecond time scale (Figures 3.9B and 3.11B), however we do not formally attempt to simulate the actual 2'-oxyanion, but rather maintain the 2'OH at all times. Consistent with a bifunctional A38 base-acid mechanism, the asymmetric pH activity profile observed particularly in studies involving the more structurally stable, native four-way junction form of the hairpin ribozyme suggests the presence of a single titratable group (25). In contrast, proton inventories are consistent with two protons in flight during the transition state as indicated in the first mechanism, however, the inventory data are sufficiently ambiguous that a one-proton mechanism cannot be discounted (30). In addition, the interpretation of proton inventories is complicated by the fact that changes in hydrogen bonding alone can give rise to the underlying kinetic solvent isotope effects (80).

In conclusion, our results further support the importance of nucleobase functional groups in ribozyme catalysis, and the utility of MD simulations in exploring their impact in atomistic detail. Most significantly, our simulations expose the significant structural consequences of the protonation state of the active site A38 nucleobase, providing a detailed rationalization for previous mechanistic proposals based on experimental data of the hairpin ribozyme. Finally, our simulations reveal that in addition to transition state stabilization and general acid catalysis A38 should be considered a plausible candidate for general base catalysis.

3.5 Acknowledgements

We thank Michal Otyepka for helpful discussions and Kamila Réblová for providing the electrostatic potential used to derive partial charges of A38H+. This work was supported by NIH grant GM62357 to NGW and an NIH Molecular Biophysics Training Grant fellowship to MAD. JŠ was supported by by the Academy of Sciences of the Czech Republic, grants no. AV0Z50040507 and AV0Z50040702, and by grant IAA400040802 by the Grant Agency of the Academy of Sciences of the Czech Republic.

References

1. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, **35**, 849-857.
2. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, **31**, 147-157.
3. Stahley, M.R. and Strobel, S.A. (2005) Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science*, **309**, 1587-1590.
4. Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L. and Tong, L. (2006) Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature*, **444**, 953-956.
5. Vogel, A., Schilling, O., Spath, B. and Marchfelder, A. (2005) The tRNase Z family of proteins: physiological functions, substrate specificity and structural properties. *Biol Chem*, **386**, 1253-1264.
6. McCarthy, T.J., Plog, M.A., Floy, S.A., Jansen, J.A., Soukup, J.K. and Soukup, G.A. (2005) Ligand requirements for glmS ribozyme self-cleavage. *Chem Biol*, **12**, 1221-1226.
7. Winkler, W.C., Nahvi, A., Roth, A., Collins, J.A. and Breaker, R.R. (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature*, **428**, 281-286.
8. Murray, J.B., Seyhan, A.A., Walter, N.G., Burke, J.M. and Scott, W.G. (1998) The hammerhead, hairpin and VS ribozymes are catalytically proficient in monovalent cations alone. *Chem Biol*, **5**, 587-595.
9. Ke, A., Zhou, K., Ding, F., Cate, J.H. and Doudna, J.A. (2004) A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature*, **429**, 201-205.
10. Ke, A., Ding, F., Batchelor, J.D. and Doudna, J.A. (2007) Structural roles of monovalent cations in the HDV ribozyme. *Structure*, **15**, 281-287.
11. Lee, T.S., Silva Lopez, C., Giambasu, G.M., Martick, M., Scott, W.G. and York, D.M. (2008) Role of Mg2+ in hammerhead ribozyme catalysis from molecular simulation. *J Am Chem Soc*, **130**, 3053-3064.
12. Martick, M. and Scott, W.G. (2006) Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell*, **126**, 309-320.
13. Klein, D.J. and Ferre-D'Amare, A.R. (2006) Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science*, **313**, 1752-1756.
14. Cochrane, J.C., Lipchock, S.V. and Strobel, S.A. (2007) Structural investigation of the GlmS ribozyme bound to Its catalytic cofactor. *Chem Biol*, **14**, 97-105.
15. Torelli, A.T., Krucinska, J. and Wedekind, J.E. (2007) A comparison of vanadate to a 2'-5' linkage at the active site of a small ribozyme suggests a role for water in transition-state stabilization. *Rna*, **13**, 1052-1070.
16. Salter, J., Krucinska, J., Alam, S., Grum-Tokars, V. and Wedekind, J.E. (2006) Water in the active site of an all-RNA hairpin ribozyme and effects of Gua8 base variants on the geometry of phosphoryl transfer. *Biochemistry*, **45**, 686-700.

17. Alam, S., Grum-Tokars, V., Krucinska, J., Kundracik, M.L. and Wedekind, J.E. (2005) Conformational heterogeneity at position U37 of an all-RNA hairpin ribozyme with implications for metal binding and the catalytic structure of the S-turn. *Biochemistry*, **44**, 14396-14408.
18. Rupert, P.B., Massey, A.P., Sigurdsson, S.T. and Ferre-D'Amare, A.R. (2002) Transition state stabilization by a catalytic RNA. *Science*, **298**, 1421-1424.
19. Rupert, P.B. and Ferre-D'Amare, A.R. (2001) Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature*, **410**, 780-786.
20. Torelli, A.T., Spitale, R.C., Krucinska, J. and Wedekind, J.E. (2008) Shared traits on the reaction coordinates of ribonuclease and an RNA enzyme. *Biochem Biophys Res Commun*, **371**, 154-158.
21. MacElrevey, C., Salter, J.D., Krucinska, J. and Wedekind, J.E. (2008) Structural effects of nucleobase variations at key active site residue Ade38 in the hairpin ribozyme. *RNA*, **14**, 1600-1616.
22. Hampel, A. and Cowan, J.A. (1997) A unique mechanism for RNA catalysis: the role of metal cofactors in hairpin ribozyme cleavage. *Chem Biol*, **4**, 513-517.
23. Cottrell, J.W., Kuzmin, Y.I. and Fedor, M.J. (2007) Functional analysis of hairpin ribozyme active site architecture. *J Biol Chem*, **282**, 13498-13507.
24. Lambert, D., Heckman, J.E. and Burke, J.M. (2006) Cation-specific structural accommodation within a catalytic RNA. *Biochemistry*, **45**, 829-838.
25. Kuzmin, Y., Da Costa, C., Cottrell, J. and Fedor, M. (2005) Role of an active site adenine in hairpin ribozyme catalysis. *JOURNAL OF MOLECULAR BIOLOGY*, **349**, 989-1010.
26. Rueda, D., Bokinsky, G., Rhodes, M.M., Rust, M.J., Zhuang, X. and Walter, N.G. (2004) Single-molecule enzymology of RNA: essential functional groups impact catalysis from a distance. *Proc Natl Acad Sci U S A*, **101**, 10066-10071.
27. Kuzmin, Y., Da Costa, C. and Fedor, M. (2004) Role of an active site guanine in hairpin ribozyme catalysis probed by exogenous nucleobase rescue. *J. Mol. Biol.*, **340**, 233-251.
28. Ryder, S.P. and Strobel, S.A. (2002) Comparative analysis of hairpin ribozyme structures and interference data. *Nucleic Acids Res*, **30**, 1287-1291.
29. Lebruska, L.L., Kuzmine, I.I. and Fedor, M.J. (2002) Rescue of an abasic hairpin ribozyme by cationic nucleobases: Evidence for a novel mechanism of RNA catalysis. *Chemistry & Biology*, **9**, 465-473.
30. Pinard, R., Hampel, K.J., Heckman, J.E., Lambert, D., Chan, P.A., Major, F. and Burke, J.M. (2001) Functional involvement of G8 in the hairpin ribozyme cleavage mechanism. *Embo Journal*, **20**, 6434-6442.
31. Bokinsky, G., Rueda, D., Misra, V.K., Rhodes, M.M., Gordus, A., Babcock, H.P., Walter, N.G. and Zhuang, X. (2003) Single-molecule transition-state analysis of RNA folding. *Proc Natl Acad Sci U S A*, **100**, 9302-9307.
32. Auffinger, P. and Hashem, Y. (2007) Nucleic acid solvation: from outside to insight. *Curr Opin Struct Biol*, **17**, 325-333.
33. McDowell, S.E., Spackova, N., Sponer, J. and Walter, N.G. (2007) Molecular dynamics simulations of RNA: an in silico single molecule approach. *Biopolymers*, **85**, 169-184.

34. Cheatham, T.E., 3rd. (2004) Simulation and modeling of nucleic acid structure, dynamics and interactions. *Curr Opin Struct Biol*, **14**, 360-367.
35. Razga, F., Koca, J., Mokdad, A. and Sponer, J. (2007) Elastic properties of ribosomal RNA building blocks: molecular dynamics of the GTPase-associated center rRNA. *Nucleic Acids Res*, **35**, 4007-4017.
36. Musselman, C., Al-Hashimi, H.M. and Andricioaei, I. (2007) iRED analysis of TAR RNA reveals motional coupling, long-range correlations, and a dynamical hinge. *Biophys J*, **93**, 411-422.
37. Li, W., Sengupta, J., Rath, B.K. and Frank, J. (2006) Functional conformations of the L11-ribosomal RNA complex revealed by correlative analysis of cryo-EM and molecular dynamics simulations. *Rna*, **12**, 1240-1253.
38. Deng, N.J. and Cieplak, P. (2007) Molecular dynamics and free energy study of the conformational equilibria in the UUUU RNA hairpin. *Journal of Chemical Theory and Computation*, **3**, 1435-1450.
39. Bevilacqua, P.C. (2003) Mechanistic considerations for general acid-base catalysis by RNA: Revisiting the mechanism of the hairpin ribozyme. *Biochemistry*, **42**, 2259-2265.
40. Walter, N.G. (2007) Ribozyme catalysis revisited: is water involved? *Mol Cell*, **28**, 923-929.
41. Cochrane, J.C. and Strobel, S.A. (2008) Catalytic Strategies of Self-Cleaving Ribozymes. *Acc. Chem. Res.*, **41**, 1027-1035.
42. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *Journal of the American Chemical Society*, **117**, 5179-5197.
43. Cheatham, T.E., Cieplak, P. and Kollman, P.A. (1999) A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *Journal of Biomolecular Structure & Dynamics*, **16**, 845-862.
44. Wang, J., Cieplak, P. and Kollman, P. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *JOURNAL OF COMPUTATIONAL CHEMISTRY*, **21**, 1049-1074.
45. Krasovska, M.V., Sefcikova, J., Reblova, K., Schneider, B., Walter, N.G. and Sponer, J. (2006) Cations and Hydration in Catalytic RNA. Molecular Dynamics of the Hepatitis Delta Virus Ribozyme. *Biophys. J.*, **91**, 626-638.
46. Krasovska, M.V., Sefcikova, J., Spackova, N., Sponer, J. and Walter, N.G. (2005) Structural dynamics of precursor and product of the RNA enzyme from the hepatitis delta virus as revealed by molecular dynamics simulations. *J Mol Biol*, **351**, 731-748.
47. Sefcikova, J., Krasovska, M.V., Spackova, N., Sponer, J. and Walter, N.G. (2007) Impact of an extruded nucleotide on cleavage activity and average catalytic core conformation of the HDV ribozyme. *Biopolymers*, **85**, 392-406.
48. Sefcikova, J., Krasovska, M.V., Sponer, J. and Walter, N.G. (2007) The genomic HDV ribozyme utilizes a previously unnoticed U-turn motif to accomplish fast site-specific catalysis. *Nucleic Acids Res.*, **35**, 1933-1946.

49. Rhodes, M., Reblova, K., Sponer, J. and Walter, N. (2006) Trapped water molecules are essential to structural dynamics and function of a ribozyme. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, **103**, 13380-13385.
50. Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H. and Pedersen, L.G. (1995) A Smooth Particle Mesh Ewald Method. *Journal of Chemical Physics*, **103**, 8577-8593.
51. Berendsen, H.J.C., Postma, J.P.M., Vangunsteren, W.F., Dinola, A. and Haak, J.R. (1984) Molecular-Dynamics with Coupling to an External Bath. *Journal of Chemical Physics*, **81**, 3684-3690.
52. Ryckaert, J.P., Ciccotti, G. and Berendsen, H.J.C. (1977) Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *Journal of Computational Physics*, **23**, 327-341.
53. BAYLY, C., CIEPLAK, P., CORNELL, W. and KOLLMAN, P. (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges - The RESP model. *JOURNAL OF PHYSICAL CHEMISTRY*, **97**, 10269-10280.
54. Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., 3rd, Snoeyink, J., Richardson, J.S. *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res*, **35**, W375-383.
55. Milligan, J.F., Groebe, D.R., Witherell, G.W. and Uhlenbeck, O.C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res*, **15**, 8783-8798.
56. Berzal-Herranz, A., Joseph, S., Chowrira, B.M., Butcher, S.E. and Burke, J.M. (1993) Essential nucleotide sequences and secondary structure elements of the hairpin ribozyme. *Embo J*, **12**, 2567-2573.
57. Joseph, S. and Burke, J.M. (1993) Optimization of an anti-HIV hairpin ribozyme by in vitro selection. *J Biol Chem*, **268**, 24515-24518.
58. Richardson, J.S., Schneider, B., Murray, L.W., Kapral, G.J., Immormino, R.M., Headd, J.J., Richardson, D.C., Ham, D., Hershkovits, E., Williams, L.D. *et al.* (2008) RNA backbone: Consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *Rna-a Publication of the Rna Society*, **14**, 465-481.
59. Bloomfield, V.A., Crothers, D.M. and Tinoco, I. (2000) *Nucleic Acids - Structures, Properties, and Functions*. University Science Books, Sausalito, CA.
60. Tang, C.L., Alexov, E., Pyle, A.M. and Honig, B. (2007) Calculation of pKas in RNA: on the structural origins and functional roles of protonated nucleotides. *J Mol Biol*, **366**, 1475-1496.
61. Hampel, K.J. and Burke, J.M. (2001) A conformational change in the "loop E-like" motif of the hairpin ribozyme is coincidental with domain docking and is essential for catalysis. *Biochemistry*, **40**, 3723-3729.
62. Gaur, S., Heckman, J.E. and Burke, J.M. (2008) Mutational inhibition of ligation in the hairpin ribozyme: substitutions of conserved nucleobases A9 and A10 destabilize tertiary structure and selectively promote cleavage. *RNA*, **14**, 55-65.

63. Walter, N.G., Yang, N. and Burke, J.M. (2000) Probing non-selective cation binding in the hairpin ribozyme with Tb(III). *J Mol Biol*, **298**, 539-555.
64. Reblova, K., Lankas, F., Razga, F., Krasovska, M.V., Koca, J. and Sponer, J. (2006) Structure, dynamics, and elasticity of free 16S rRNA helix 44 studied by molecular dynamics simulations. *Biopolymers*, **82**, 504-520.
65. Liu, S., Bokinsky, G., Walter, N.G. and Zhuang, X. (2007) Dissecting the multistep reaction pathway of an RNA enzyme by single-molecule kinetic "fingerprinting". *Proc Natl Acad Sci U S A*, **104**, 12634-12639.
66. Min, D., Xue, S., Li, H. and Yang, W. (2007) 'In-line attack' conformational effect plays a modest role in an enzyme-catalyzed RNA cleavage: a free energy simulation study. *Nucleic Acids Res*, **35**, 4001-4006.
67. Soukup, G.A. and Breaker, R.R. (1999) Relationship between internucleotide linkage geometry and the stability of RNA. *Rna*, **5**, 1308-1325.
68. Lopez, X., Dejaegere, A., Leclerc, F., York, D.M. and Karplus, M. (2006) Nucleophilic attack on phosphate diesters: A density functional study of in-line reactivity in dianionic, monoanionic, and neutral systems. *Journal of Physical Chemistry B*, **110**, 11525-11539.
69. Nam, K.H., Gaot, J.L. and York, D.M. (2008) Quantum mechanical/molecular mechanical simulation study of the mechanism of hairpin ribozyme catalysis. *Journal of the American Chemical Society*, **130**, 4680-4691.
70. Nam, K., Gao, J.L. and York, D.M. (2008) Electrostatic interactions in the hairpin ribozyme account for the majority of the rate acceleration without chemical participation by nucleobases. *Rna-a Publication of the Rna Society*, **14**, 1501-1507.
71. Fedor, M.J. and Williamson, J.R. (2005) The catalytic diversity of RNAs. *Nat Rev Mol Cell Biol*, **6**, 399-412.
72. Doudna, J.A. and Lorsch, J.R. (2005) Ribozyme catalysis: not different, just worse. *Nat Struct Mol Biol*, **12**, 395-402.
73. Gong, B., Chen, J.H., Chase, E., Chadalavada, D.M., Yajima, R., Golden, B.L., Bevilacqua, P.C. and Carey, P.R. (2007) Direct measurement of a pK(a) near neutrality for the catalytic cytosine in the genomic HDV ribozyme using Raman crystallography. *J Am Chem Soc*, **129**, 13335-13342.
74. Knitt, D.S., Narlikar, G.J. and Herschlag, D. (1994) Dissection of the role of the conserved G.U pair in group I RNA self-splicing. *Biochemistry*, **33**, 13864-13879.
75. Ryder, S.P., Oyelere, A.K., Padilla, J.L., Klostermeier, D., Millar, D.P. and Strobel, S.A. (2001) Investigation of adenosine base ionization in the hairpin ribozyme by nucleotide analog interference mapping. *Rna*, **7**, 1454-1463.
76. Thomas, J.M. and Perrin, D.M. (2006) Active site labeling of G8 in the hairpin ribozyme: implications for structure and mechanism. *J Am Chem Soc*, **128**, 16540-16545.
77. Wilson, T.J., Ouellet, J., Zhao, Z.Y., Harusawa, S., Araki, L., Kurihara, T. and Lilley, D.M. (2006) Nucleobase catalysis in the hairpin ribozyme. *Rna*, **12**, 980-987.
78. Wlodawer, A., Li, M., Gustchina, A., Tsuruoka, N., Ashida, M., Minakata, H., Oyama, H., Oda, K., Nishino, T. and Nakayama, T. (2004) Crystallographic and

biochemical investigations of kumamolisin-As, a serine-carboxyl peptidase with collagenase activity. *J Biol Chem*, **279**, 21500-21510.
79. Xu, Q., Guo, H.B., Wlodawer, A., Nakayama, T. and Guo, H. (2007) The QM/MM molecular dynamics and free energy simulations of the acylation reaction catalyzed by the serine-carboxyl peptidase kumamolisin-As. *Biochemistry*, **46**, 3784-3792.
80. Tinsley, R.A., Harris, D.A. and Walter, N.G. (2003) Significant kinetic solvent isotope effects in folding of the catalytic RNA from the hepatitis delta virus. *J Am Chem Soc*, **125**, 13972-13973.

Chapter 4

A Splicing Funnel: The Spliceosome Constrains Broad Conformational Dynamics of Single Pre-mRNA Molecules[d]

4.1 Introduction

Spliceosome assembly on pre-mRNA substrates in the yeast *Saccharomyces cerevisiae* has been studied for over 20 years (1,2), yet little is known about the kinetic and structural details of this complex multistep process. A clear description of the process by which the spliceosome removes introns from pre-mRNA is critical to understanding eukaryotic gene regulation. In mammals, a large fraction of protein coding genes are interrupted by non-coding intervening sequences (introns) with an average of eight introns in human protein coding genes (3). In yeast only ~4% of protein coding genes contain introns, however, these transcripts are highly expressed and carefully regulated (4). In order to ensure proper cellular function, introns must be precisely spliced out of pre-mRNAs by the spliceosome, a macromolecular complex composed of five small nuclear ribonuclear proteins (snRNPs) and numerous additional proteins (1,2). The spliceosome assembles on each pre-mRNA and carries out two sequential transesterification reactions removing the intron. Sequence homology in introns exists only in short stretches at the sites of transesterification (4). While there is some evidence for functionally significant conserved structural elements (5,6), the introns encountered by the spliceosome are highly variable in length and sequence (4). By contrast, the small nuclear RNAs (snRNAs) that comprise the RNA components of the spliceosome are

[d] Adapted from, Blanco M, Ditzler MA, Fuller F, Aravamudhan P, Guthrie C, Abelson J, Walter NG. A Splicing Funnel: The Spliceosome Constrains Broad Conformational Dynamics of Single Pre-mRNA Molecules. (*in preparation*) The identification and initial characterization of the *UBC4* intron was done by John Abelson. smFRET measurements and analysis were preformed by Mark Ditzler and Mario Blanco. Preparation of labeled substrate was done by John Abelson and Marion Blanco.

highly conserved across diverse eukaryotic species. How the spliceosome then reliably directs its diverse substrates through two consecutive sequence specific transesterification reactions is ill-understood, but essential to understanding gene regulation. Failure of the spliceosome to appropriately remove introns can have potentially catastrophic consequences for the cell and is known to be associated with multiple diseases (7).

It is generally understood that splicing *in vitro* proceeds in an ordered assembly, commencing with the ATP independent formation of a commitment complex with the association of the U1 snRNP, BBP, Mud2, and a small number of additional proteins with the pre-mRNA (8-10). In ATP dependent steps the pre-spliceosome is formed by the binding of U2 snRNP to the branch site (11) followed by complete assembly of the spliceosome by association of the U4/U6·U5 triple snRNP with the pre-mRNA. The branch point (BP) mutant used in this work is known to interfere with spliceosome assembly and blocks pre-mRNA splicing prior to the first chemical step (12). For wild type (WT) substrates U1 and U4 are released from the spliceosome and the remaining snRNPs undergo further ATP and helicase dependent conformational changes before the first transesterification reaction, which results in a 3' lariat intermediate and linear 5' exon (13,14). Further rearrangement of the spliceosomal components then takes place before the second transesterfication at the 3' splice site joining the two exons together (15). The 3' splice site (3'SS) mutant used in this work is capable of carrying out the first chemical step of splicing but blocks splicing prior to the second chemical step (12). Following the second step of splicing the spliceosome dissociates form the lariat intron and the mature mRNA. However, the role of transiently associated factors is unclear and the possibility of alternative *in vitro* assembly pathways remains open (1,16). Further adding to the overall complexity of spliceosome activity, recent work has demonstrated that both chemical steps in splicing are reversible (17). Though the general outline of the assembly pathway has been extensively studied using bulk biochemical assays, nearly nothing is known about the kinetics of each step in the pathway.

Pre-mRNA substrates must be repeatedly repositioned and base pairing interactions with spliceosomal snRNAs are continuously formed and remodeled throughout splicing. The dynamic integration of the substrate into the spliceosome and progression through two chemical transformations suggests that spliceosome assembly

and activity will be reflected in changes in global conformation and dynamics of the pre-mRNA. To track in real-time the conformational states through which the spliceosome takes a pre-mRNA without the need to isolate or synchronize reaction intermediates, we here have developed a spliceosome assembly assay based on single molecule fluorescence resonance energy transfer (smFRET). We identified the *UBC4* intron as a small, efficiently *in-vitro* spliced substrate and attached a donor-acceptor smFRET pair to the 5'- and 3'-exons. We observe spliceosome assembly in yeast whole cell extract at the single molecule level, monitoring pre-mRNA conformational dynamics over extended time periods by tracking the FRET signal from surface immobilized pre-mRNA substrates, on a PEG functionalized quartz surface. Previous single molecule studies of pre-mRNAs in yeast cell extract suggests that splicing can proceed with a substrate immobilized on a PEG functionalized surface, and that splicing proceeds with similar efficiency as standard assays with pre-mRNAs in solution (18). In order to evaluate the complex set of conformational transitions of the pre-mRNA substrate during spliceosome assembly, we used the QuB software package (http://www.qub.buffalo.edu) to identify the genuine discrete FRET states embedded in the noise inherent to single molecule detection. QuB was originally developed to evaluate electrophysiology data and has more recently been successfully applied to smFRET data from the ribosome (19). We analyzed trajectories of mutant and WT pre-mRNA substrates in splicing buffer and cell extract, and observe clear differences in the conformational dynamics between WT, 3'SS mutant and BP mutant sequences. We observe that the spliceosome encounters a pre-mRNA substrate that is characterized by an inherently broad and dynamic ensemble of conformations. We find that spliceosome assembly ultimately restricts this expansive conformational space sampled by the pre-mRNA following an initially complex mixture of ATP dependent conformational transitions. Finally, our work provides an initial framework for the quantitative evaluation of individual steps in spliceosome assembly.

4.2 Materials and Methods

Synthesis and Activity of Pre-mRNA UBC4 Substrates

Two fractions of the final pre-mRNA substrate (76 nucleotides of the 5' segment and the 59 nucleotides of the 3' segment), each containing a single 5-amino-allyl-uridine, were

synthesized and deprotected by Dharmacon. The 5' and 3' fractions were reacted with Cy5 and Cy3 N-hydroxysuccinimidyl ester, respectively. The labeled RNA strands were then ligated using T7 RNA ligase and an appropriate DNA splint as described (20). The ligation results in a shortened (135 nucleotide, or nt) version of the *UBC4* pre-mRNA with intact 95 nt intron and 20 nt 5' and 3' exons of the wild-type sequence GAACUAAGUGAUCUAGAAAGGUAUGUCUAAAGUUAUGGCCACGUUUCAAA UGCGUGCUUUUUUUUUAAAACUUAUGCUCUUAUUUACUA*A*CAAAAUCAAC AUGCUAUUGAACUA**G**AGAUCCACCUACUUCAUGUU. The intron sequence is shown in red and in the 3' splice site (3'SS) mutant, the bold underlined guanine was replaced with a cytosine, whereas in the branch point (BP) mutant the italicized underlined adenosine is replaced by a cytosine. The splicing activity of this fluorescently labeled *UBC4* pre-mRNA was determined using standard *in vitro* assays. Pre-mRNA was incubated in 40% yeast whole cell extract containing 8 mM Hepes-K^+ (pH 7.0), 2mM $MgCl_2$, 0.08 mM EDTA, 60mM $K(PO_4)$, 20mM KCl, 8% (v/v) glycerol, 3%(w/v) PEG, 0.5mM DTT supplemented with 2mM ATP, and the products were resolved via denaturing PAGE.

Preparation of Yeast Cell Extract

Spicing active whole cell extract was prepared as described previously from BJ2168 yeast cells (21). Cells were grown in 8 to 16 liters of YPD medium to an OD600 of 3.5 - 4.0. The cells were then harvested, washed and pelleted. Cell pellets were resuspended and disrupted by grinding with a mortar grinder (RM100, Retsch), according to (21). The frozen powder was thawed rapidly at room temperature and centrifuged, the supernatant was then centrifuged again and then dialyzed for 3 h against 20 mM Hepes-K^+, pH 7.0, 0.2 mM EDTA, 0.5mM DTT, 50mM KCl, 20% (v/v) glycerol. 40% (v/v) cell extract was used in all single molecule studies. ATP depletion was achieved by adding 1mM glucose to the cell extract and incubating at room temperature prior to each experiment. The U6 depletion was based on a previously described method (9,22) wherein pre-incubation of the extract with a DNA oligonucleotide complementary to U6, 2 mM ATP and 120 units of RNase H (TAKARA) results in the degradation of all endogenous U6.

Single-molecule FRET

Slides were prepared using a protocol modified from previous published methods (23,24). Quartz microscope slides were reacted with aminopropyltriethoxysilane (APTES) in acetone for 30 min to generate an amino functionalized surface, which was reacted overnight with a 10:1 mixture of succinimidyl ester functionalized O-methyl-PEG and biotin-PEG to PEGylate the surface. Sulfo-disuccinaimidlytartarate (sulfo-DST) was reacted for 30 min with the remaining unreacted amines, thus ensuring the surface does not carry a positive charge. Slide coverslips undergo a similar procedure. The slides are then rinsed, dried and a single flow channel per slide is assembled. A stock solution of 0.2 mg/ml streptavidin in buffer (50mM NaCl, 50mM Tris-HCl pH 7.5) was added to the channel and incubated for 10 min at room temperature.

The doubly labeled pre-mRNA was heat annealed to a 2'-O-methyl-RNA capture (tether) strand complementary to the 17 3' terminal nucleotides of the 3' exon by incubating at 70˚C for 2 min and cooling to room temperature for 10 min. The capture strand carries a 5' biotin, which binds to the streptavidin on the surface of the slide to immobilize the pre-mRNA. Following tether annealing, the hybrid was diluted to a concentration of ~50 pM in splicing buffer and flowed into the slide. Using a prism-based total internal reflection fluorescence (TIRF) microscope as described(25) data were collected from single molecules in splicing buffer 8 mM Hepes-K^+ (pH 7.0), 2mM $MgCl_2$, 0.08 mM EDTA, 60mM K(PO_4), 20mM KCl, 8% (v/v) glycerol, 3%(w/v) PEG, 0.5mM DTT, 40% (v/v) cell extract depleted of ATP and 40% (v/v) cell extract supplemented with 2mM ATP, both in splicing buffer. An oxygen scavenger system composed of Protocatechuate Dioxygenase (PCD) and protocatechuate (PCA) was added to splicing buffer and cell extracts to limit photobleaching (18), Trolox was also added to the solution to limit fluorophore blinking (26). The Cy3 donor was exited using a 532 nm laser and emission by the Cy3 and Cy5 fluorophores was recorded at 100 ms time resolution using an intensified CCD camera (Princeton Instruments, I-Pentamax). A FRET value was then calculated by dividing the intensity of acceptor emission by the total emission from both donor and acceptor.

Molecules with anti-correlated donor-acceptor signal were analyzed with QuB software (available at http://www.qub.buffalo.edu/soft.php). In order to evaluate noise in the data and determine the underlying FRET states, the entire data set for each condition

was analyzed by the iterative application of the Viterbi and Baum-Welch algorithms to generate idealized trajectories. The number of states assumed in the idealization was varied from 5 to 11 and the corresponding fits were evaluated using the Bayesian information criterion (BIC). The number of states that resulted in the best BIC score was used in our analysis. In the data analysis a transition rate optimization was not applied, which would result in a prohibitively high computational cost for these data sets. It is also likely that transitions in these trajectories may not be determined by a single rate constant, and although degenerate FRET states could be incorporated into the model, this would result in even greater computational cost. However, as an additional means of ensuring that the FRET transitions determined from the idealized trajectories reflect true conformational changes the Viterbi and Baum-Welch optimization was applied to the signal from the donor emission, acceptor emission, and the FRET ratio (FRET = $I_{acceptor}/(I_{donor}+I_{acceptor})$) independently, and only FRET transitions that had appropriate corresponding transitions in the donor and acceptor channels were used in our analysis. Additionally FRET transitions with a step size smaller than 0.1 were not included in our analysis. The transition plots were created using MATLAB. Histograms of FRET states were created using Origin 7.0.

4.3 Results

Identification and functional characterization of a pre-mRNA substrate

We sought to identify a small intron efficiently spliced *in vitro* using a splicing specific mircroarray (27,28). We identified multiple candidates and tested the *UBC4* intron to find that relatively efficient splicing is maintained upon truncation of the exons, ultimately resulting in a 135 nt pre-mRNA substrate (Figure 4.1A). We sought to place the fluorophores as close to the 5' and 3' splice sites as possible for maximum sensitivity in detecting distance changes between the exons without interfering with activity. The substrate used in this work has a donor (Cy3) 7 nt 5' of the 5' splice site and (Cy5) 4 nt 3' of the 3' splice site (Figure 4.1B). Introduction of 3'SS and BP mutations have the expected effect (12) of blocking splicing prior to the second and first steps of splicing, respectively (Figure 4.1A). Additionally, incubation of the wt pre-mRNA with U6

depleted extracts results in no detectable splicing (not shown). In our smFRET assays the labeled pre-mRNA was immobilized through hybridization to a short 2'-O-methyl capture RNA (Figure 4.1C) bound via biotin to a streptavidin coated and PEG functionalized quartz surface. The Cy3 donor was exited using a 532 nm laser and emission by the Cy3 and Cy5 fluorophores was recorded to determine FRET efficiency (Figure 4.1D). An enzymatic oxygen scavenger system (OSS) was used to limit photobleaching and Trolox was added to the solution in order to suppress fluorophore blinking. *In vitro* splicing showed that the labeled pre-mRNA is splicing active in the presence of capture RNA, OSS and Trolox yielding ~20% lariat intron for the 3'ss mutant and ~10% lariat and ~10% mature mRNA (not shown).

Identification of underlying conformational states and dynamics

The Förster radius of Cy3-Cy5 FRET pair is ~54 Å so that smFRET can effectively monitor small changes in inter-fluorophore distances from ~20-100 Å. Cryo-EM of snRNPs and spliceosomal complexes range in size from 200-300 Å (29,30), however footprinting indicates proximity of the two exons throughout splicing (31,32). Our smFRET trajectories monitor the dynamic positioning of the two exons within the spliceosomal complex. Consistent with the complex conformational rearrangements required for splicing, our smFRET experiments yield FRET trajectories with multiple discrete FRET states, as identified with combined Viterbi and Baum-Welch algorithms and Bayesian Information Criterion (BIC) using QuB software, transition rate constants were not explicitly optimized in this analysis. We also took advantage of the fact that donor and acceptor fluorescence are independently measured to identify real conformational transitions by fitting donor, acceptor, and FRET signal independently. A FRET transition is regarded as verified when it corresponds to anti-correlated signal changes in donor and acceptor. Alternatively, we carried out a preliminary analysis of individual trajectories in which we explicitly optimized the rates of FRET transitions as in previous single-molecule FRET analysis (19,33), from which we could also verify the presence of sequence and ATP dependent changes. However, the large number of FRET states and trajectories in our data sets and the apparent kinetic heterogeneity ultimately resulted in high computational cost and a low level of confidence in the resulting model.

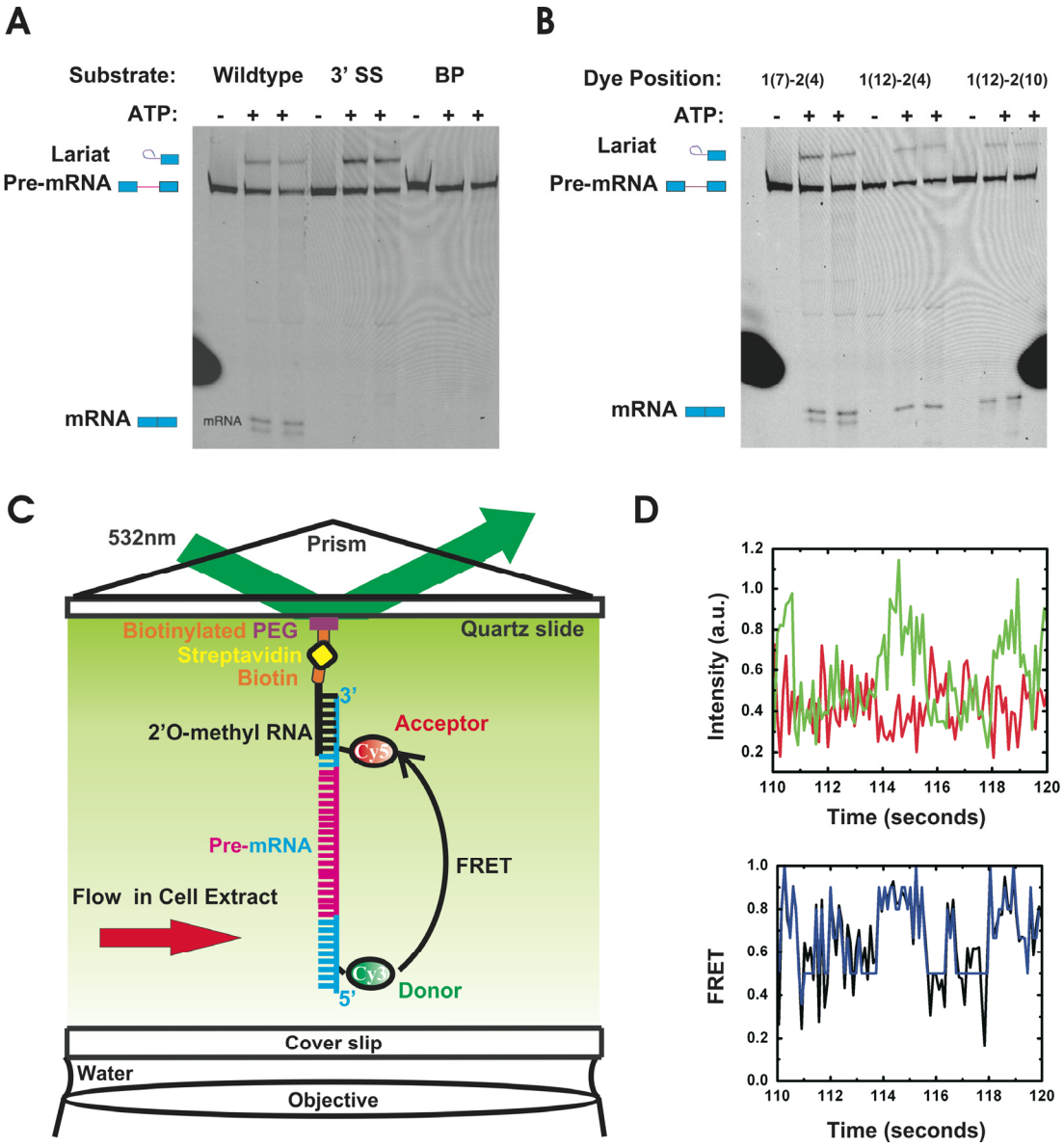


Figure 4.1: Activity of the fluorophore labeled UBC4 derived intron *in vitro*. (A) Splicing activity of the truncated and (Cy3-Cy5) labeled UBC4 derived intron using standard *in vitro* splicing assays followed by denaturing PAGE and imaged via Cy5 fluorescence. (B) Splicing of the truncated and (Cy3-Cy5) labeled *UBC4* intron with Cy3 and Cy5 at multiple positions within the exon. The substrate used in this work with Cy3 7nt from the 5'SS and Cy5 4nt from the 3'SS splices. (C) Schematic representation of the smFRET setup showing the relative orientation of the pre-mRNA, the tether and the functionalized surface. (D) Example of a single molecule FRET trajectory. Donor emission intensity is shown in green and acceptor in red. The FRET ratio is shown in black and the idealization is shown in blue.

The spliceosome encounters a dynamic pre-mRNA substrate

All three pre-mRNAs exhibit significant global dynamics with fluctuations between high and low FRET states in splicing buffer alone, in the absence of extract. The distribution of FRET states however is not identical between the pre-mRNAs, as the WT and BP mutant favor high FRET states relative to the 3'SS mutant, which spends more time in low FRET states (Figure 4.2 A). The difference in overall FRET distribution corresponds well with the ensemble of secondary structures predicted by the Vienna software package (Figure 4.2B). This software (available at http://rna.tbi.univie.ac.at/) calculates the partition function of secondary structures and predicts a distribution of compact structures for the WT and BP. By contrast, a distribution that favors less compact folding is predicted for the 3'SS mutant, which is clearly reflected in the observed distribution of FRET states (Figure 4.2A). Relating the predicted partition function with donor-acceptor distances using a previously established coarse grained model (5) to generate a distribution of predicted inter-fluorophore distances further underscores the general agreement of our smFRET data with the predicted structural diversity of the three pre-mRNAs (Figure 4.2B).

Transitions between the various FRET states are observed for all three sequences and the transition kinetics span a range of time scales (Figure 4.2C). The total number of transitions is large but several of the transitions are observed in >70% of all molecules indicating common kinetic behavior among most molecules of the same sequence. The pattern of FRET transitions varies in a way that corresponds with the differences in overall FRET distribution discussed above. Notably, the WT and BP pre-mRNAs share features distinct from the 3'SS. In particular, the WT and BP pre-mRNAs exhibit many transitions between high FRET states with a similar pattern of kinetics.

ATP-independent dynamics are dependent on substrate sequence

Spliceosomal components are known to interact with pre-mRNA in the absence of ATP. U1, BBP and Mud2 can be crosslinked to pre-mRNA in cell extracts depleted of ATP (10,34). The addition of cell extract depleted of ATP results in a shift away from the FRET distributions observed in buffer (Figures 4.2A and 4.3A). Lower FRET values are favored in all three pre-mRNAs, however, the effect is most pronounced for the WT, as

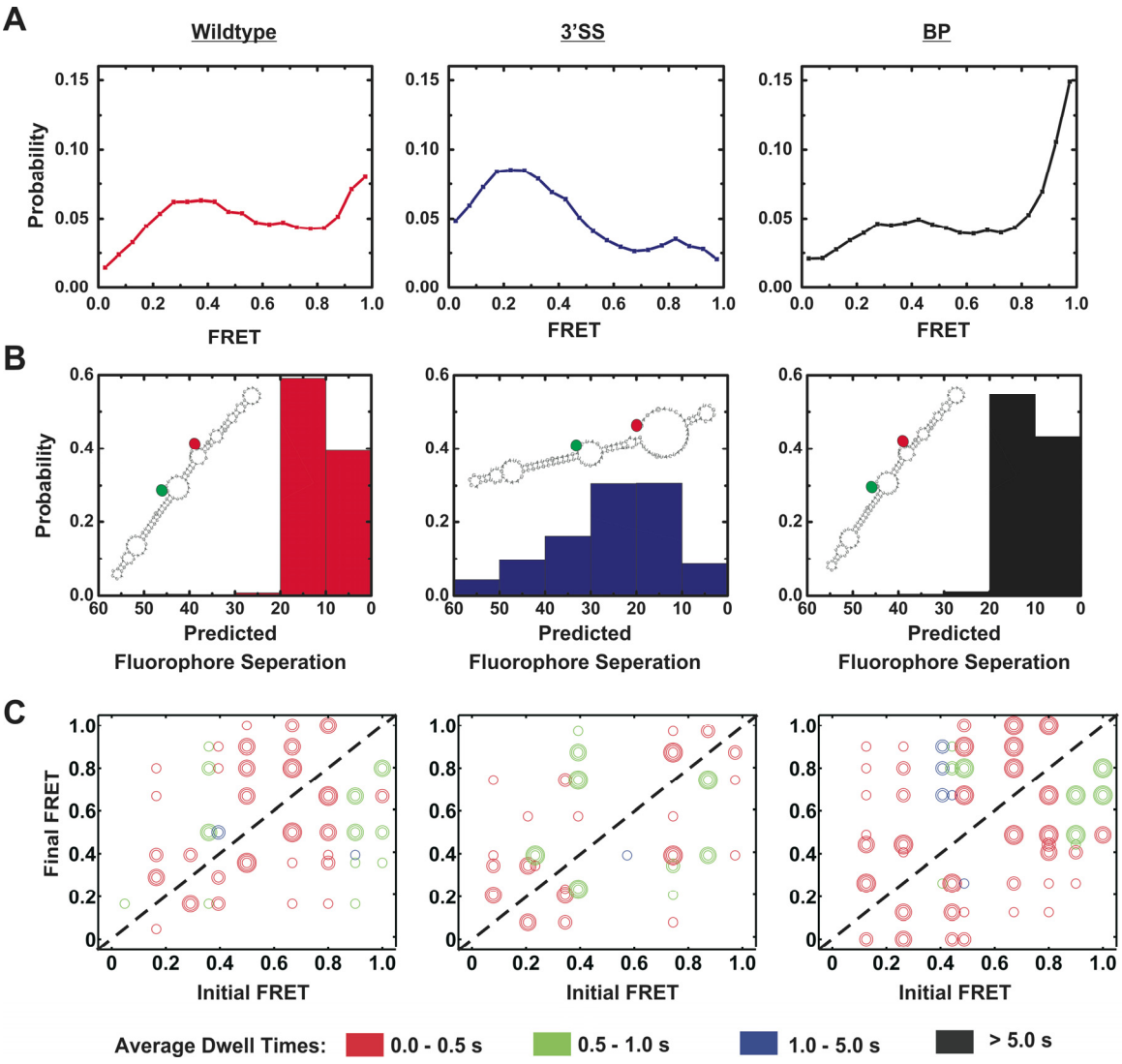


Figure 4.2: Conformational dynamics intrinsic to WT and mutant pre-mRNA substrates. Structure elements and dynamics help distinguish the three pre-mRNA substrates. (A) The probability distribution of FRET states for the three sequences are shown. The three mutants have distinct FRET distributions. (B) The distribution of predicted inter-fluorophore distances based on secondary structure analysis is show for each pre-mRNA sequence, along with the predicted lowest free energy secondary structure (inset). This correlates with the experimental FRET state distributions shown in Fig.4.2A (C) Transitions observed in the three different sequences are plotted. Each transition is indicated by a circle or set of concentric circles. The number of concentric circles for a given transition correlates with the fraction of trajectories in which that transition is observed. Transitions present in 10-30% of molecules are represented by a single circle, those in 30-50% by two concentric circles, 50-70% by three, 70%-90% by four and those present in >90% are represented by five. The transitions are also color coded with respect to the average dwell time that precedes each type of transition as indicated in the figure.

the 3'SS already favors low FRET in buffer and the BP mutant maintains a bias toward higher FRET in the presence of cell extract. The range of FRET transitions observed for the WT becomes more constrained upon addition of extract while the 3'SS and BP mutants samples a wide range of rapid FRET transitions (Figure 4.3B). It is important to note that we do not observe any significant variations in quenching/blinking or photobleaching of the fluorophores between the three pre-mRNAs in cell extract. All sequences show limited quenching/blinking and photobleaching with both phenomena exhibiting a rate constant of ~$0.01s^{-1}$, therefore suggesting that smFRET differences between the WT and mutant pre-mRNAs relate to the differences expected for their interaction with the assembling spliceosome.

We also examined the effect of depleting U6 from cell extracts on the dynamics of the WT RNA. U6 depletion is achieved by incubation with a complementary DNA strand, RNase H, and ATP to make U6 accessible to DNA-induced degradation by RNase H (35). We then depleted the U6 depleted extracts of ATP prior to addition to the immobilized WT pre-mRNA and observed smFRET states distinct from those observed in the U6 containing extracts (compare Figures 4.4A and 4.3A). This impact on FRET distribution due to the absence of U6 snRNA suggests an ATP independent association of U6 with the pre-mRNA. Still, in both U6 containing and depleted extracts the highest FRET states are diminished upon addition of extract.

ATP dependent changes in FRET distribution require an active branch site and U6 snRNA

The addition of ATP to splicing active cell extract leads to a time dependent change in the FRET distributions of the WT and 3'SS mutant pre-mRNAs, in contrast to the BP mutant (Figure 4.3A). The identity and prevalence of FRET transitions evolves differently over time for each pre-mRNA. The BP mutant does not undergo time dependent changes in overall FRET distributions and at early time points following the addition of ATP the pattern of FRET transitions remains essentially unchanged (Figure 4.3B and C), consistent with this mutant's known interference with spliceosome assembly (12). Specifically, two sets of rapid symmetric transitions between 0.3 and 0.5 FRET, and between 0.5 and 0.7 FRET are common to >90% of all molecules both before

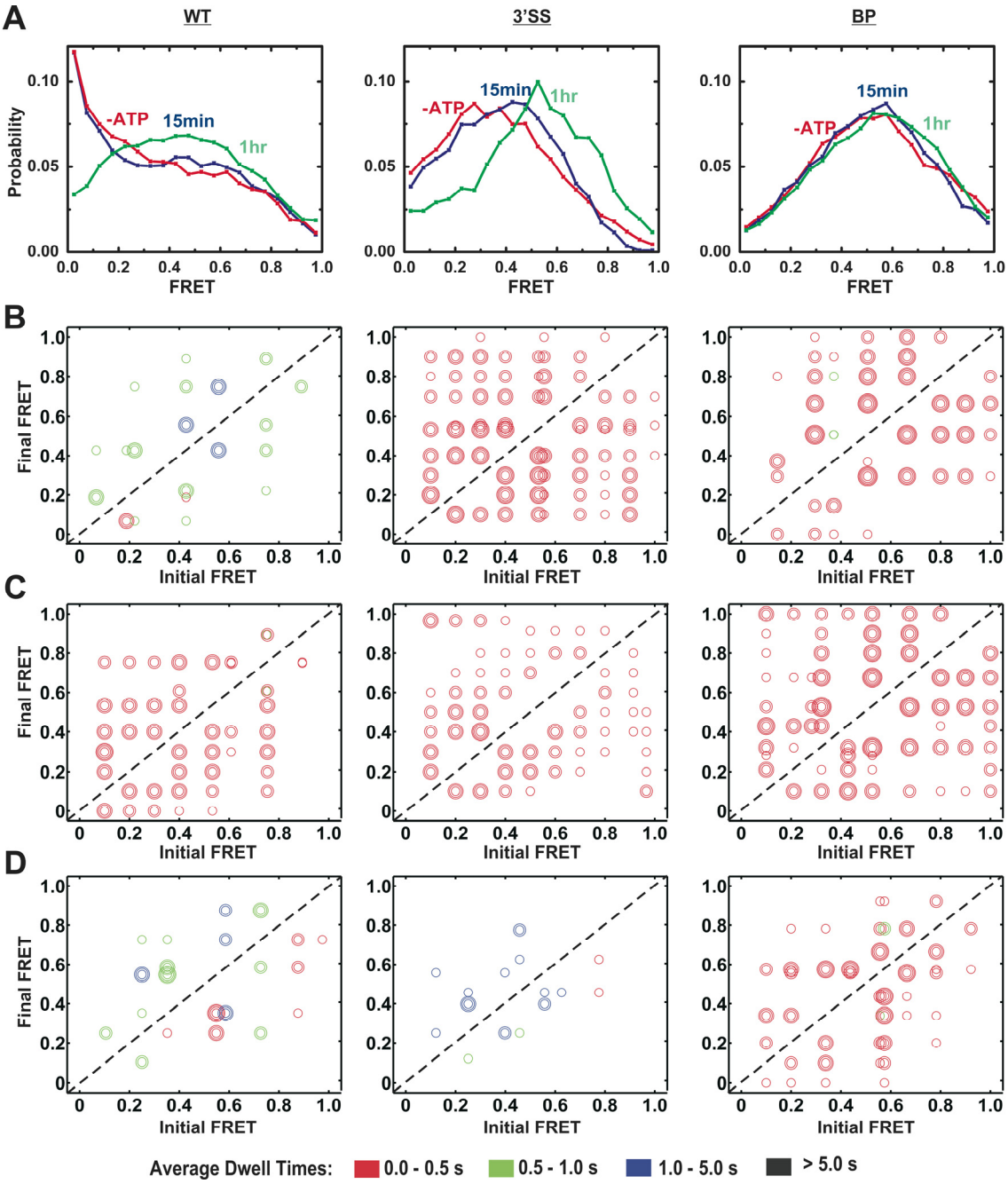


Figure 4.3: ATP dependent conformational dynamics of the WT and mutant pre-mRNA in extract. (A) The probability distribution of FRET states for the three sequences are shown. Overlaid for each sequence are the distributions observed in the absence of ATP, within 15min of the addition of ATP, and following >1hr incubation in ATP. ATP dependence is observed for the WT and 3'SS mutant, but not the BP mutant. Transitions observed for the three pre-mRNA sequences in (B) ATP depleted extract (C) within 15min of the addition of ATP supplemented extract, and (D) following >1hr incubation in the presence of ATP supplemented extract are plotted as in Figure 4.2.

and within 15 min following the addition of ATP. While little change is apparent in the overall FRET distribution of the WT pre-mRNA, the pattern of transitions changes dramatically at early time points, in striking contrast to observations on the BP mutant. For the 3'SS mutant, changes are observed in the pattern of FRET transitions, although they are less dramatic than those observed for the WT as may be expected for a partially splicing impaired mutant.

After 1 h in ATP supplemented cell extract, even the BP mutant begins to display limited ATP dependent behavior. The overall distributions for both BP and WT in U6 depleted extract continue to exhibit unchanged FRET distributions (Figures 4.3A and 4.4A), however the number and identity of transitions sampled by the BP mutant is slightly diminished and there are no longer any transitions common to >90% of all molecules, suggesting a less homogenous population. Following an hour-long incubation with ATP, the number of highly populated transitions common to >70% of all molecules increase from one to two transitions in spite of the slower kinetics, suggesting the development of a more homogenous population of molecules over time (see Discussion). In addition, the overall number of transitions observed for the WT is dramatically reduced relative to early ATP time points. For the 3'SS mutant the number of FRET transitions is also dramatically reduced, however unlike the WT, the population of molecules appears to become less homogenous (although it is worth noting that at least one transition is still present in >50% of the molecules) (Figure 4.4D). Again, we do not observe sequence dependent effects on quenching/blinking or photobleaching of the fluorophores in our cell extracts nor do we observe ATP dependent effects, although the photobleaching rate does increase about two fold at time points >1 h, which could have the effect of slightly reducing the apparent homogeneity of substrate dynamics by diminishing the observation time. Finally, we note that early experiments carried out in extracts without an oxygen scavenger system revealed that at least for the WT pre-mRNA the photobleaching rate of Cy5 is decreased upon addition of ATP (data not shown).

Control smFRET experiments in which a mature mRNA containing only the two exons was immobilized and incubated in the presence of cell extract exhibits a steady high FRET signal (Figure 4.4B). Notably, the observed steady high FRET results in a FRET distribution distinct from that observed for WT pre-mRNA incubated in ATP

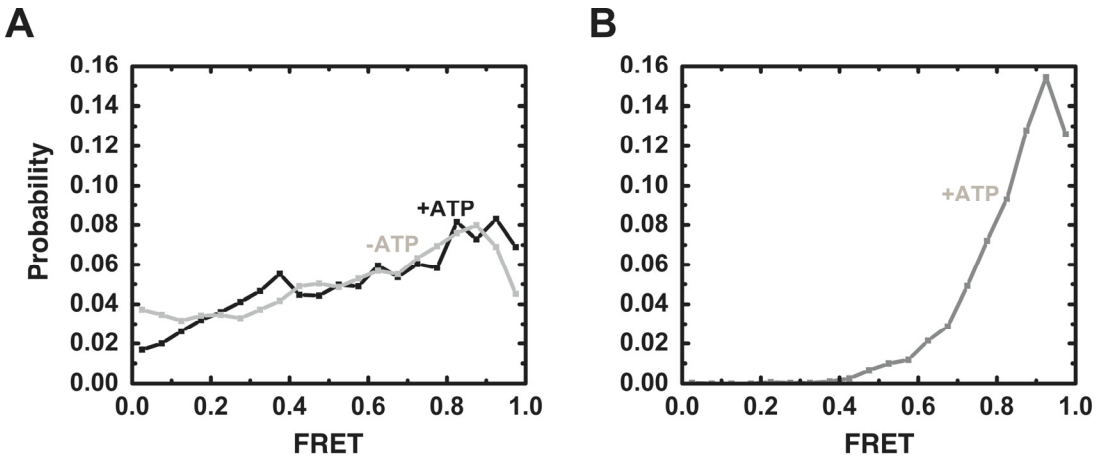


Figure 4.4: Additional FRET distributions. (A) The FRET distribution observed for the WT pre-mRNA in the presence of U6 depleted extract that is also depleted of ATP and U6 deplete extract after incubation in ATP supplemented extract for >1hr. (B) The FRET distribution observed for the intron-less control mature mRNA in ATP supplemented extract.

supplemented extract for >1 h. This finding suggests that only a fraction of the total WT pre-mRNA undergoes both steps of splicing, consistent with the ~10% of WT pre-mRNA observed to do so in standard ensemble splicing assays (Figure 4.1A).

4.4 Discussion

Here we have applied smFRET to evaluate the complex conformational dynamics inherent to intron removal by the spliceosome. Our data reveal the presence of a complex set of global conformational transitions in the small *UBC4* intron related to fluctuations in exon proximity, highlighting the necessity for single molecule approaches to study spliceosome assembly. We find that overall following an initially broad distribution of conformational states a pre-mRNA is eventually specifically funneled into a defined segment of conformational space during spliceosome assembly. Our work provides the foundation for isolating specific functional transitions and associated kinetics of single-molecules in the *UBC4* system in the future, which will be necessary for generating a comprehensive dynamic model of pre-mRNA splicing.

Our smFRET measurements in splicing buffer alone demonstrate the dependence of intron structure on its conserved branch point and 3' splice site sequences. We also observe a large degree of conformational flexibility in the WT, 3'SS and BP mutant pre-mRNAs. Heterogeneous structure and dynamics is a common feature of RNA sequences (36), and therefore likely characterizes most intron sequences in the absence of the spliceosomal machinery. In addition, one may expect to observe an increase in the variety of FRET transitions exhibited by the WT pre-mRNA after 1 h incubation with ATP, which is expected to result in a mixture of pre-mRNA, lariat intron, and mature mRNA (Figure 4.1). However, there is some evidence that spliceosomal components can maintain similar global conformations throughout splicing (31), which is strongly supported by our observation of significant reduction (funneling) of the global dynamics in our smFRET assays with WT upon extended incubation with splicing active extract.

The initial shift in FRET distribution of the WT substrate upon extract addition is consistent with ATP independent binding of U1 snRNP to the 5' splice site and the notion that branch point binding protein (BBP) and/or other proteins associating with the

branch point abrogate the frequent approaches of the two exons observed in buffer alone (32,34). The higher FRET values observed for the WT pre-mRNA in the U6 depleted cell extracts suggests that there is also an ATP independent association of additional spliceosomal components that contributes to the lower FRET distribution in the presence of U6 snRNA. Association between U1 and all four additional snRNPs in the spliceosome is known to occur in extracts even in the absence of substrate or ATP (16). ATP independent steps are less pronounced for the 3'SS mutant due to the intron structure that already keeps the exons apart, however, the exact transitions present in buffer are altered by the addition of extract. The difference in FRET distribution and transitions between WT and 3'SS is perhaps unexpected as the 3'SS is not directly involved in ATP independent assembly steps. However, if the 3' splice site is free of spliceosomal components as expected, then the variation in sequence between WT and 3'SS mutant could easily result in a different distribution of structures intrinsic to the RNA as is the case in buffer. For the BP mutant high FRET values in extract is consistent with exon proximity, which may be due to the BBP's diminished affinity for the mutant BP sequence (37), although BBP association with the BP mutant is not entirely eliminated in extracts (34). The modest change in FRET with the BP mutant pre-mRNA can be attributed to binding of U1 snRNA and additional spliceosomal components (potentially even all four additional snRNPs, see following (16)) to the unaltered 5' splice site.

The observation that U6 depleted extract still results in a small change in FRET distribution relative to buffer is consistent with the ability of the depleted extract to form a complex very early in the spliceosome assembly pathway (35). What is perhaps surprising is the fact that in the absence of ATP the FRET distribution observed for the WT pre-mRNA in U6 depleted extract is significantly different from that observed in U6 containing extracts. While it is possible that this reflects an indirect effect on assembly of U1 and BBP on the pre-mRNA it is also consistent with the proposal that the spliceosome is pre-assembled into a penta-snRNP containing all five spliceosomal snRNPs (16). In this non-canonical view of assembly U6 would disrupt the penta-snRNP, ultimately impacting all steps of assembly.

We observe clear ATP dependent changes in the dynamics of the substrates which are most pronounced for the fully splicing active WT. The significant increase in the number of states upon initial addition of ATP is consistent with the sequential assembly and conformational rearrangements of the spliceosome during splicing. It is also important to note that the early time point measurements represent the only data set acquired immediately following a perturbation to the system i.e. the addition of ATP supplemented extract. It is highly significant that the BP mutant, which is known from ensemble glycerol gradient shift assays to block the formation of any large complexes in the presence of ATP supplemented extract (12), exhibits such similar conformational dynamics following this perturbation. Importantly, for all three pre-mRNAs the overall FRET distributions are essentially unchanged immediately following the addition of ATP and therefore the pronounced changes in the FRET transitions exhibited by the WT sequence are masked in the ensemble. This observation clearly demonstrates the need for a single-molecule approach.

After the initial sampling of a broad range of FRET states, the two catalytically competent pre-mRNA sequences (WT and 3'SS) are funneled into a smaller subset of transitions upon long >1hr incubation in ATP supplemented extracts. The increase in the number of transitions present in >70% of all trajectories for the WT sequence after >1 h incubation in ATP, suggests that either the population of molecules is becoming more homogenous over time or more rapidly samples the available conformational space. The fact that the transition kinetics concomitantly become slower speaks against the second interpretation as does the observation of slightly faster photobleaching after >1 h incubation with ATP which reduces the observation time. Unlike the WT, the >1 h transition plot of the 3'SS mutant reveals no transitions common to >70% of the trajectories, which may indicate that the population becomes less homogenous over time or be a reflection of slower transition kinetics and faster photobleaching. It is therefore not readily apparent whether the BP mutant or 3'SS mutant becomes more or less homogenous over time, but relative comparison of the mutants after a 1-h incubation in ATP containing cell extract strongly suggests that the WT sequence is the most strongly directed towards homogeneity by prolonged exposure to splicing competent extract.

In addition to conformational dynamics of the pre-mRNA, the dynamics of fluorescence intensity and irreversible bleaching can be informative of spliceosome assembly. The ATP dependent increase of Cy5 lifetimes in the absence of OSS is suggestive of an ATP dependent change in the local environment of the Cy5 attached to the 3' exon. This ATP dependence is consistent with the observation that ATP independent assembly does not occur at the 3' splice site, but subsequent ATP dependent assembly of the spliceosome brings the 5' splice site closer to the large RNP assembly which could then influence the local environment of the fluorophore.

It is difficult to directly assess progression of the immobilized pre-mRNAs through the two chemical steps. The mature mRNA exhibits a FRET distribution distinct from the FRET distribution observed for the WT after incubation in ATP supplemented extracts over an extended period of time, consistent with the general observation that only a fraction of any given pre-mRNA is spliced during *in vitro* splicing assays (~10% mature mRNA). In addition, while yeast splicing does not result in the formation of an exon junction complex as observed in mammals, the splicing process may result in a final association with RNA binding proteins that are distinct from those that assemble directly on a mature splicing product. It has recently been demonstrated that truncation of the 3' exon inhibits spliceosome disassembly by blocking the binding of the helicase Prp22 (38). Therefore, it is possible that our 2'-O-methyl RNA tether, which is hybridized to most of the 3'exon, blocks the binding of Prp22 and therefore disrupts spliceosome disassembly. In this case fully spliced material would likely have a FRET signature distinct from the control and may even revert due to the recently demonstrated reversibility of splicing upon blockage of disassembly (17). Furthermore, if Prp22 were to bind to the duplex, its prossessive helicase activity may unwind the tether-pre-mRNA duplex, ultimately resulting in a loss of TIRFM-based signal indistinguishable from photobleaching. Additionally, the observation of ATP dependent survival of the fluorophores in cell extract without OSS raises the possibility that specific steps in splicing may become spectroscopically silent, although we do not observe this ATP dependence in the presence of OSS. Previous studies of surface immobilized pre-mRNA under smFRET conditions established catalytic activity by observing a loss of fluorescence from the substrate, which could instead reflect a catalytically inhibited

assembly of spliceosomal components that either quenches the fluorophore or increases its susceptibility to photo-bleaching. While further work will be required to unambiguously establish catalytic activity under single-molecule conditions, the work presented here does establish the assembly of a relatively homogenous spliceosomal complex from splicing competent extract on surface immobilized WT pre-mRNA, which is distinct from the ATP dependent trends observed for BP and 3'SS mutants and in U6 depleted extracts.

4.5 Acknowledgments

We thank Reinhard Lührmann for kindly providing us with the splicing active yeast extract used in this work.

References

1. Smith, D.J., Query, C.C. and Konarska, M.M. (2008) "Nought may endure but mutability": spliceosome dynamics and the regulation of splicing. *Mol Cell*, 30, 657-666.
2. Staley, J.P. and Guthrie, C. (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell*, 92, 315-326.
3. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, 409, 860-921.
4. Lopez, P.J. and Seraphin, B. (1999) Genomic-scale quantitative analysis of yeast pre-mRNA splicing: implications for splice-site recognition. *RNA*, 5, 1135-1137.
5. Rogic, S., Montpetit, B., Hoos, H.H., Mackworth, A.K., Ouellette, B.F. and Hieter, P. (2008) Correlation between the secondary structure of pre-mRNA introns and the efficiency of splicing in Saccharomyces cerevisiae. *BMC Genomics*, 9, 355.
6. Charpentier, B. and Rosbash, M. (1996) Intramolecular structure in yeast introns aids the early steps of in vitro spliceosome assembly. *RNA*, 2, 509-522.
7. Faustino, N.A. and Cooper, T.A. (2003) Pre-mRNA splicing and human disease. *Genes Dev*, 17, 419-437.
8. Ruby, S.W. and Abelson, J. (1988) An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly. *Science*, 242, 1028-1035.
9. Seraphin, B. and Rosbash, M. (1989) Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell*, 59, 349-358.
10. Du, H. and Rosbash, M. (2001) Yeast U1 snRNP-pre-mRNA complex formation without U1snRNA-pre-mRNA base pairing. *RNA*, 7, 133-142.
11. Parker, R., Siliciano, P.G. and Guthrie, C. (1987) Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell*, 49, 229-239.
12. Vijayraghavan, U., Parker, R., Tamm, J., Iimura, Y., Rossi, J., Abelson, J. and Guthrie, C. (1986) Mutations in conserved intron sequences affect multiple steps in the yeast splicing pathway, particularly assembly of the spliceosome. *EMBO J*, 5, 1683-1695.
13. Yean, S.L. and Lin, R.J. (1991) U4 small nuclear RNA dissociates from a yeast spliceosome and does not participate in the subsequent splicing reaction. *Mol Cell Biol*, 11, 5571-5577.
14. Makarov, E.M., Makarova, O.V., Urlaub, H., Gentzel, M., Will, C.L., Wilm, M. and Luhrmann, R. (2002) Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science*, 298, 2205-2208.
15. Schwer, B. and Guthrie, C. (1992) A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. *EMBO J*, 11, 5033-5039.

16. Stevens, S.W., Ryan, D.E., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D. and Abelson, J. (2002) Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol Cell*, 9, 31-44.
17. Tseng, C.K. and Cheng, S.C. (2008) Both catalytic steps of nuclear pre-mRNA splicing are reversible. *Science*, 320, 1782-1784.
18. Crawford, D.J., Hoskins, A.A., Friedman, L.J., Gelles, J. and Moore, M.J. (2008) Visualizing the splicing of single pre-mRNA molecules in whole cell extract. *RNA*, 14, 170-179.
19. Munro, J.B., Altman, R.B., O'Connor, N. and Blanchard, S.C. (2007) Identification of two distinct hybrid state intermediates on the ribosome. *Mol Cell*, 25, 505-517.
20. Stark, M.R., Pleiss, J.A., Deras, M., Scaringe, S.A. and Rader, S.D. (2006) An RNA ligase-mediated method for the efficient creation of large, synthetic RNAs. *RNA*, 12, 2014-2019.
21. Gottschalk, A., Neubauer, G., Banroques, J., Mann, M., Luhrmann, R. and Fabrizio, P. (1999) Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP. *EMBO J*, 18, 4535-4548.
22. Fabrizio, P. and Abelson, J. (1990) Two domains of yeast U6 small nuclear RNA required for both steps of nuclear precursor messenger RNA splicing. *Science*, 250, 404-409.
23. Ha, T., Rasnik, I., Cheng, W., Babcock, H.P., Gauss, G.H., Lohman, T.M. and Chu, S. (2002) Initiation and re-initiation of DNA unwinding by the Escherichia coli Rep helicase. *Nature*, 419, 638-641.
24. van Oijen, A.M., Blainey, P.C., Crampton, D.J., Richardson, C.C., Ellenberger, T. and Xie, X.S. (2003) Single-molecule kinetics of lambda exonuclease reveal base dependence and dynamic disorder. *Science*, 301, 1235-1238.
25. Zhuang, X., Kim, H., Pereira, M.J., Babcock, H.P., Walter, N.G. and Chu, S. (2002) Correlating structural dynamics and function in single ribozyme molecules. *Science*, 296, 1473-1476.
26. Rasnik, I., McKinney, S.A. and Ha, T. (2006) Nonblinking and long-lasting single-molecule fluorescence imaging. *Nat Methods*, 3, 891-893.
27. Clark, T.A., Sugnet, C.W. and Ares, M., Jr. (2002) Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science*, 296, 907-910.
28. Pleiss, J.A., Whitworth, G.B., Bergkessel, M. and Guthrie, C. (2007) Transcript specificity in yeast pre-mRNA splicing revealed by mutations in core spliceosomal components. *PLoS Biol*, 5, e90.
29. Furman, E. and Glitz, D.G. (1995) Purification of the spliceosome A-complex and its visualization by electron microscopy. *J Biol Chem*, 270, 15515-15522.
30. Jurica, M.S., Licklider, L.J., Gygi, S.R., Grigorieff, N. and Moore, M.J. (2002) Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA*, 8, 426-439.
31. Donmez, G., Hartmuth, K., Kastner, B., Will, C.L. and Luhrmann, R. (2007) The 5' end of U2 snRNA is in close proximity to U1 and functional sites of the pre-mRNA in early spliceosomal complexes. *Mol Cell*, 25, 399-411.

32. Abovich, N. and Rosbash, M. (1997) Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell*, 89, 403-412.
33. McKinney, S.A., Joo, C. and Ha, T. (2006) Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophysical Journal*, 91, 1941-1951.
34. McPheeters, D.S. and Muhlenkamp, P. (2003) Spatial organization of protein-RNA interactions in the branch site-3' splice site region during pre-mRNA splicing in yeast. *Mol Cell Biol*, 23, 4174-4186.
35. Fabrizio, P., McPheeters, D.S. and Abelson, J. (1989) In vitro assembly of yeast U6 snRNP: a functional assay. *Genes Dev*, 3, 2137-2150.
36. Schultes, E.A., Spasic, A., Mohanty, U. and Bartel, D.P. (2005) Compact and ordered collapse of randomly generated RNA sequences. *Nat. Struct. Mol. Biol.*, 12, 1130-1136.
37. Berglund, J.A., Chua, K., Abovich, N., Reed, R. and Rosbash, M. (1997) The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell*, 89, 781-787.
38. Schwer, B. (2008) A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. *Mol Cell*, 30, 743-754.

Chapter 5

Conclusions and Outlook

The experimental techniques described in this thesis monitor the conformational dynamics of RNA from picoseconds to hours and evaluate conformational dynamics from atomic detail in a ~20 kilodalton all-RNA system, in which all aspects of the system are defined, to following the global dynamics of a pre-mRNA in a multi-megadalton complex for which even the number and identity of the components remains unclear. Common to these approaches is the observation that any perturbation introduced into an RNA system must be evaluated in terms of its simultaneous impact on structure, dynamics, and chemistry. MD of the hairpin ribozyme demonstrates that protonation of A38 (proposed to be a general acid in the reaction) significantly impacts the structure of the active site in addition to its potential role in directly protonating the 5'O leaving group. In investigating the conformational heterogeneity of the hairpin ribozyme the presence of multiple active conformations challenges the widely accepted view of "one-function, one-fold" and tremendously complicates the interpretation of ensemble data. In evaluating splicing of single pre-mRNA molecules, mutations introduced at the branch point and 3' splice site known to interfere with chemistry at these sites were found to also have a major unanticipated impact on the dynamically exchanging structures exhibited by the pre-mRNAs in the absence of the spliceosome.

Our smFRET investigation into the dynamics of the hairpin ribozyme reveals surprisingly persistent molecular heterogeneity. Though we have implicated the S-turn as a potential motif responsible for this phenomenon, a precise structural description is still frustratingly absent. The footprinting analysis presented in this thesis provides some hints about the structural origin, and additional footprinting may be illuminating. The footprinting presented here was carried out under semi-native conditions in monovalent salt with no denaturants, however, our experiments imply that some structural feature

must persist even at elevated temperatures and in the presence of denaturants. It may therefore be possible to directly probe the retained structure by footprinting under these conditions, in the presence of urea and at high temperature. Additionally, high resolution NMR studies of the ribozmye in both the native docked form and under denaturing conditions could be used to detect and describe the folding heterogeneity in atomic detail. Furthermore, given the similar observation of persistent heterogeneity in the SRL (1), it would be informative to characterize this motif by using both biochemical footprinting and NMR spectroscopy under denaturing conditions.

One possible explanation for the molecular heterogeneity observed in the hairpin ribozyme is the presence of mass neutral chemical modifications such as backbone bond isomerisation from the native 3'-5' linkage to a 2'-5' linkage. There are several reasons why we do not favor this interpretation, most significantly the fact that the EMSA heterogeneity, which is strongly correlated with the smFRET heterogeneity, is observed for RNA produced through both chemical synthesis and *in vitro* transcription. Additionally, it has previously been shown with the native four-way junction version of the hairpin that two different types of chemical synthesis (tBDMS and ACE protected) both exhibit molecular heterogeneity at the single molecule level (2). This source independence for the hairpin ribozyme's molecular heterogeneity is also observed in the EMSA heterogeneity reported for the SRL (1). While the source independence of molecular heterogeneity is reproducible between systems, it is possible that for either the hairpin ribozyme, the SRL, or both, the RNA fold itself catalyzes bond isomerisation, in which case source independence would be expected.

There is a relatively simple means of either verifying or ruling out this specific modification. It has been shown that in the context of a DNA-RNA hybrid duplex 2'-5' linkages are significantly more susceptible to alkaline induced backbone cleavage than the native 3'-5' linkage (3), likely reflecting a more accessible in-line geometry for the 2'-5' compared to the protected 3'-5' linkage within the context of a double helix, allowing the 3'O nucleophile to more easily attach the phosphorus in the 2'-5' linkage. We could therefore separate the two populations of ^{32}P labeled hairpin ribozyme using our standard EMSA with one of the two strands labeled, and then add a DNA oligonucleotide complementary to the labeled RNA strand in molar excess. Once the

DNA oligonucleotide is annealed to one of the ribozyme strands, incubation at slightly elevated pH overnight followed by high-resolution D-PAGE should clearly reveal the presence of any 2'-5' bonds within the ribozyme.

The significant conformational dynamics imparted to RNA by its highly flexible backbone has been suggested to be a liability in catalysis (4). However in the model we derive from our MD simulations, in which we propose that A38 acts as both general acid and base, this flexibility is absolutely essential to catalysis. A similar mechanistic proposal exists for the *glmS* ribozyme (5), and it will be interesting to see if the dual utilization of functional groups emerges as a general theme in RNA catalysis, thereby permitting RNA to utilize its intrinsic backbone flexibility in carrying out its biological function. A clear limitation in MD to establishing a catalytic mechanism is that by definition bonds cannot be broken or formed, and even in the best scenario one is simply able to sample the ground state and infer possible mechanisms primarily based on proximity. However, having generated a mechanistic hypothesis through a synergistic analysis of MD, X-ray crystal structures, and kinetic data, QM/MM analysis can now be applied to evaluate the chemical feasibility of these mechanisms. Starting from the ground state conformations described in this thesis, our collaborator (Michal Otyepka) is currently carrying out such an analysis. The results will inform future biochemical and computational experiments. An iterative application of computational, structural, and biochemical analyses will be necessary to ultimately unlock the elusive mechanisms of RNA catalysis.

While there is currently no available high-resolution structure of the spliceosome, its emergence would undoubtedly immediately result in MD simulations of the complex. Structures of individual spliceosome components are currently available, however, without detailed knowledge of the structural context of the assembled spliceosome the overall utility of such simulations is unclear. However, it may be possible to take advantage of the likely evolutionary relationship between the spliceosome and the group II intron. High-resolution structures are now available for the group II intron and these structures further strengthen the hypothesis that it is related to the spliceosome (6-8). MD simulations of the group II intron could, in conjunction with the extensive

biochemical data available for this system, be of considerable value in determining the mechanism of splicing by both the group II intron and the spliceosome.

In our smFRET analysis of the spliceosome a number of important concerns about the pre-mRNA construct used need to be addressed. While it is clear that splicing can proceed in the presence of the 2'-O-methyl RNA tether it is equally clear that the impact of this tether on spliceosome disassembly must be resolved in order to more meaningfully interpret current and future smFRET data. Additionally, it is vital to establish whether or not the tether is unwound at any point during splicing. Addressing these two issues should be relatively straight forward. If the tether blocks disassembly following incubation in ATP supplemented extracts, then the spliceosome-pre-mRNA complex should be easily detected by ultracentrifugation on a glycerol gradient, as previously demonstrated in experiments in which spliceosome disassembly was blocked by removal of the 3' exon (9). Furthermore, whether or not the pre-mRNA-tether duplex is unwound at any point during splicing can also be addressed. Pre-annealing of the pre-mRNA with unlabeled 2'-O-methyl RNA tether, followed by incubation in extracts with and without ATP in the presence of an excess of ^{32}P labeled tether should reveal whether or not the initial pre-mRNA-tether duplex is unwound. If the tether is unwound by the splicing machinery, which contains many helicases, then in the ATP supplemented extracts the labeled tether will form a duplex with the substrate upon unwinding of the initial complex, while incorporation of the labeled tether would be absent from ATP depleted extracts. The degree to which the labeled tether is able to form a complex with the substrate could be detected by native PAGE. If either disassembly is blocked or the pre-mRNA-tether duplex is unwound, the data collected so far would still be of considerable value. It would, however, be necessary to establish an immobilization strategy that did not inhibit spliceosome disassembly or result in loss of substrate from the surface; once data are collected from such a substrate the comparison between the two data sets would provide information on how spliceosome disassembly and premature substrate departure from the surface bias the data presented in this thesis. For example, if spliceosome disassembly is found to be stalled, then the dynamics in the data presented here may provide a unique means of observing reverse splicing (10), and if the pre-mRNA-tether duplex is unwound at a specific step in splicing, then we will be able to establish that all

of the dynamics we observe in the current dataset occur before that point. It is possible to avoid all potential problems that arise from hybridization of the 2'-O-methyl RNA tether to the 3'exon by simply incorporating a biotin directly into the pre-mRNA either pre or post synthesis. Both of these methods would increase the overall expense of the synthesis, however, since only picomolar concentrations are necessary for smFRET analysis the costs should not be prohibitive.

It is clear that single-molecule approaches reveal a surprising level of complexity in the dynamics of pre-mRNA both in extract and buffer alone. Preliminary smFRET data from another research group investigating splicing through smFRET on a different yeast intron and using different labeling sites also reveals a wide range of FRET states for the pre-mRNA in both extracts and buffer (Aaron Hoskins, personal communication). Careful attention will need to be given to comparisons between different data sets derived from different introns and extracts. It will be important to establish whether or not the high variability in sequences and length of introns gives rise to a correspondingly wide variety of conformational dynamics intrinsic to the RNA, which could greatly complicate comparisons between multiple introns. However, secondary structure predictions do suggest the possibility of conserved structures among introns (11). The analysis of multiple introns in splicing buffer and extract will be useful in testing this hypothesis. In fact, this hypothesis could be directly tested by introducing mutations distal from the sites of transesterification and predicted to disrupt the conserved secondary structures, followed by observing their effect on conformational dynamics and splicing. It is clear that certain structural elements are incompatible with splicing, for example, the TPP riboswitch that regulates splicing in *Neurospora crassa* in part sequesters the 5' splice site (12). If intron structure is not conserved, however, as is the general consensus in the field, then each intron can be expected to exhibit its own unique set of structural transitions in buffer and quite possibly in extracts during spliceosome assembly. If this is the case it may prove much more fruitful to examine splicing at the single molecule level by labeling components of the spliceosome and not the substrate. The behavior of the spliceosome itself may exhibit conformational dynamics that are far more reproducible between different introns and even between species. To this end, our collaborators have already synthesized a Cy3-Cy5 labeled U6 snRNA.

We have only scratched the surface of kinetic information that must be established for a comprehensive understanding of spliceosome activity. We have so far only evaluated two mutant substrates. In order to understand splicing fully we must be able to identify the FRET signatures of specific intermediates. To this end it should be possible to take advantage of yeast extracts prepared from temperature sensitive mutant strains. In these strains a specific protein involved in a defined step of splicing is mutated so that it can be specifically inactivated by slightly heating the cell extract. Experiments can then be undertaken to observe spliceosome assembly up to the state that requires the inactivated protein. Then the inactivated extract can be replaced with fully active extract or supplemented with recombinant active protein and the continued progression through splicing be observed. Finally, just as kinetic analysis of the ribosome has benefited tremendously from the availability of small molecule inhibitors, the use of small molecule inhibitors of splicing such as those recently described in the human *in vitro* splicing system can be applied to smFRET studies of the spliceosome (13).

A meaningful understanding of RNA dynamics requires context and general themes must be widely demonstrated. It is therefore essential that these methods be expanded to new RNA and RNP systems. Our experimental data for the hairpin ribozyme suggest a connection between the persistent heterogeneity in the hairpin ribozyme and the SRL. This behavior is certainly more widespread than what is reported in the literature (Li Niu, Philip C. Bevilacqua, Steven Benner, personal communications), and as a growing number of techniques are applied to an expanding pool of RNAs this behavior will likely continue to be encountered. One certainly hopes that as these examples are uncovered they will be reported and attempts to determine a structural origin will be vigorously pursued, as was done here. As the number of high-resolution RNA structures continues to grow MD simulations will be able to take advantage of both the additional starting structures and an expanded database of structural information with which simulated conformations can be judged and the quality of force-fields be evaluated. While the complex kinetics apparent in initial smFRET analysis of eukaryotic splicing seem daunting, this is probably largely due to the limited context. As data from an increasing variety of substrates, fluorophore positions, and mutant extracts emerges

over the coming years from multiple laboratories, clear trends are likely to emerge. The initial quantitative analysis of smFRET is only able to provide fairly vague and qualitative information at this time, however, striving for an increasingly detailed and quantitative analysis and description of the data remains essential as we do not yet know what data features will ultimately prove most illuminating.

References

1. Korennykh, A.V., Plantinga, M.J., Correll, C.C. and Piccirilli, J.A. (2007) Linkage between substrate recognition and catalysis during cleavage of sarcin/ricin loop RNA by restrictocin. *Biochemistry*, **46**, 12744-12756.
2. Okumus, B., Wilson, T.J., Lilley, D.M. and Ha, T. (2004) Vesicle encapsulation studies reveal that single molecule ribozyme heterogeneities are intrinsic. *Biophys J.*, **87**, 2798-2806.
3. Flynn-Charlebois, A., Prior, T.K., Hoadley, K.A. and Silverman, S.K. (2003) In vitro evolution of an RNA-cleaving DNA enzyme into an RNA ligase switches the selectivity from 3'-5' to 2'-5'. *J Am Chem Soc*, **125**, 5346-5350.
4. Narlikar, G.J. and Herschlag, D. (1997) Mechanistic aspects of enzymatic catalysis: lessons from comparison of RNA and protein enzymes. *Annu Rev Biochem*, **66**, 19-59.
5. Klein, D.J. and Ferre-D'Amare, A.R. (2006) Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science*, **313**, 1752-1756.
6. Toor, N., Rajashankar, K., Keating, K.S. and Pyle, A.M. (2008) Structural basis for exon recognition by a group II intron. *Nat Struct Mol Biol*, **15**, 1221-1222.
7. Toor, N., Keating, K.S., Taylor, S.D. and Pyle, A.M. (2008) Crystal structure of a self-spliced group II intron. *Science*, **320**, 77-82.
8. Dayie, K.T. and Padgett, R.A. (2008) A glimpse into the active site of a group II intron and maybe the spliceosome, too. *RNA*, **14**, 1697-1703.
9. Schwer, B. (2008) A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. *Mol Cell*, **30**, 743-754.
10. Tseng, C.K. and Cheng, S.C. (2008) Both catalytic steps of nuclear pre-mRNA splicing are reversible. *Science*, **320**, 1782-1784.
11. Rogic, S., Montpetit, B., Hoos, H.H., Mackworth, A.K., Ouellette, B.F. and Hieter, P. (2008) Correlation between the secondary structure of pre-mRNA introns and the efficiency of splicing in Saccharomyces cerevisiae. *BMC Genomics*, **9**, 355.
12. Cheah, M.T., Wachter, A., Sudarsan, N. and Breaker, R.R. (2007) Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature*, **447**, 497-500.
13. Kuhn, A.N., van Santen, M.A., Schwienhorst, A., Urlaub, H. and Luhrmann, R. (2008) Stalling of spliceosome assembly at distinct stages by small-molecule inhibitors of protein acetylation and deacetylation. *RNA*.

Appendix A:
A Single Molecule FRET Assay to Monitor the RNase P Reaction Pathway

The ribonucleoprotein complex Ribonuclease P (RNase P) universally catalyzes the 5' maturation of precursor tRNA. The *B. subtilis* RNase P holoenzyme consists of a catalytically-active RNA component and a small protein component that directly interacts with the 5' leader of the pre-tRNA substrate and enhances metal affinity for catalysis (1). Even for the simple one-RNA-one-protein bacterial system it has been difficult to clearly identify the role of metal ions and the cationic protein component in the reaction mechanism since folding, substrate binding and catalysis are all ion dependent. Fluorescence titration and transient kinetics data indicate conformational rearrangements occur within the enzyme-substrate complex. Several labeling sites have already been explored with ensemble time-resolved FRET studies and can be used to guide development of a single molecule FRET (smFRET) assay(2).

We have developed an RNaseP construct with Cy5 and Cy3 labelling sites suitable for an smFRET assays (Figure 1A). In this construct the RNA component has a 27nt 3' extension. Hybridization of a biotinylated complimentary DNA oligo tether enables us to immobilized the construct a streptavidin coated, PEG functionalized quartz surface the via the biotin streptavidn interaction. The DNA tether also carries an amino functionality which we labelled with Cy5. We labelled the protein with Cy3 by using a single E40C mutation, the wild type protein has no cystines, thus allowing site specific labelling with the E40C mutation. We observe hybridization between the RNase P RNA and the tether, and that hybridization efficiency increasing with concentration (Figure 1B). The smFRET construct is catalytically active with only slightly diminished active (Figure 1 C).

In preliminary smFRET experiments carried out in the presence of calcium which effectively blocks catalysis but not substrate binding, we observe detectable FRET form only ~10% of molecules, and upon addition of 300nM (60 times solution K_D) pre-tRNA substrate FRET is detected for only ~5% of RNase P enzymes. It is unclear if this change results from substrate binding, future experiment in which the substrate concentration is varied should help to answer this question.

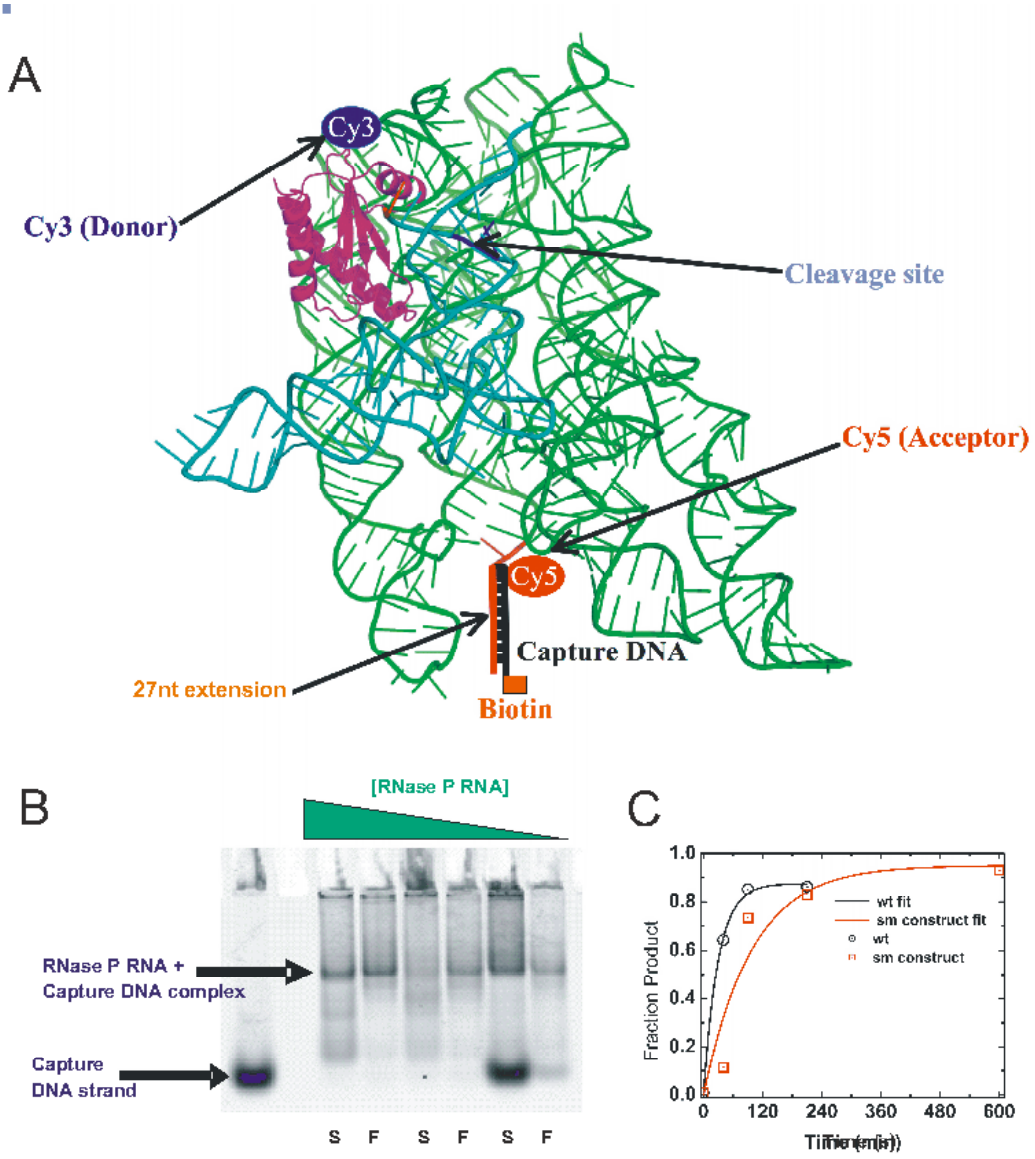


Figure 1: Three dimensional model of an smFRET RNase P construct based on model structure of RNase P (3). A 3' extension was introduced for immobilization through hybridization to a DNA oligo tether. The tether is labeled with the FRET acceptor (Cy5) and the RNase P protein is labeled with the donor (Cy3). Native PAGE shows that the tether binds to the RNA in a concentration dependent manner, and that slow cooling over several hours (S) results in degradation products not seen for fast (F) cooling (~15min). (C) Single-turnover assay reveals only slightly diminished activity in the smFRET construct.

1. Smith, J.K., Hsieh, J. and Fierke, C.A. (2007) Importance of RNA-protein interactions in bacterial ribonuclease P structure and catalysis. *Biopolymers*, **87**, 329-338.
2. Rueda, D., Hsieh, J., Day-Storms, J.J., Fierke, C.A. and Walter, N.G. (2005) The 5' leader of precursor tRNAAsp bound to the Bacillus subtilis RNase P holoenzyme has an extended conformation. *Biochemistry*, **44**, 16130-16139.
3. Niranjanakumari, S., Day-Storms, J.J., Ahmed, M., Hsieh, J., Zahler, N.H., Venters, R.A. and Fierke, C.A. (2007) Probing the architecture of the B. subtilis RNase P holoenzyme active site by cross-linking and affinity cleavage. *RNA*, **13**, 521-535.