Heterogeneous folding and function of small RNA motifs: The hairpin ribozyme and a translational riboswitch

by
Matthew Sean Marek

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Cellular and Molecular Biology)
in the University of Michigan
2014

Doctoral Committee:
Professor Nils G. Walter, Chair
Professor Hashim M. Al-Hashimi, Duke University
Professor David R. Engelke
Professor Edgar Meyhofer

2014

Dedication

This thesis would not have been possible without the guidance and patience of my mentor, Nils Walter. Much thanks also goes to the Walter Lab who was always there to solve any problem I could come up with. I'd also like to thank my family for always believing in me and supporting me, without them I wouldn't have made it. Finally, I'd like to thank all the friends I've made in grad school. I don't think I'll ever again find such a fun, intelligent, and loyal group of people that I have the privilege to call my friends.

Acknowledgements

The introductory chapter is adapted from *The Shape-Shifting Quasispecies of RNA: One Sequence, Many Functional Folds* {Marek, 2011 #212} written in collaboration with Alexander Johnson-Buck and Nils Walter. Cloning of gene *60* and its variants was carried out by Gabrielle Todd. Additionally, establishment of the S30/S100 translation system in the lab was accomplished by Gabrielle Todd. Cloning of the 1564/1563 gene and characterization of the preQ1 translational riboswitch was done by Arlie Rinaldi. Synthesis of 1M7 was done in collaboration with Benjamin Buer. Single molecule analysis was done with assistance of Paul Lund. Single molecule experiments were done with the help of Matthew Kahlscheuer, Mario Blanco, and May Dahar-Farhat.

Contents

List of Figures

Abstract

Heterogeneous folding and function are observed commonly for RNA, DNA, and protein, yet remain poorly understood. For RNA, several molecular origins have been proposed, including UV cross-linking, slow sugar repuckering, and alternate folding associated with denaturation during most purification methods. Additionally, natural alternative folds for many RNAs serve to control the biological function of the RNA. With each of these cases, there results variations in RNA fold that effect local and global behaviors of the RNA of interest.

The hairpin ribozyme is a well-studied example of a highly structured functional RNA that, despite its relatively small size of ~80 nucleotides, exhibits pronounced heterogeneity. We have assessed the relative contributions of UV cross-linking, sugar repuckering and alternative folding to molecular heterogeneity of the hairpin ribozyme and find that it is multifactorial and can be systematically suppressed. Exposure to short-wavelength (254 nm) UV irradiation for visualization during purification introduces distributive cross-links between adjacent pyrimidines and is avoided by minimizing exposure, using longer-wavelength (320 nm) irradiation, or staining with crystal violet for visualization. Differential selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) does not reveal strong evidence for slow sugar repuckering. Rather, the remaining folding heterogeneity is removed using a non-denaturing affinity purification method that maintains the co-transcriptional fold of the RNA, attesting to the risk of alternate folding on a rugged free energy landscape. These results suggest that heterogeneous folding and thus function of RNA generally may have

multiple origins and are minimized in both the test tube and the cell by the avoidance of short-range UV exposure and the segmental, 5'-to-3'-directed folding resulting from transcription.

Moving to studying natural heterogeneities necessary for biological function, we establish a ribosome expression system capable of site-specific fluorescence labeling and translation. We show that these ribosomes are capable of translating gene products *in vitro*, as well as forming initiation complexes. These assays form the basis for single molecule studies of (pre)initiation complex formation on the *Thermoanaerobacter tengcongensis* (*Tte*) 1564/1563 mRNA. The coding sequence of this mRNA is preceded by the *Tte* $preQ_1$ translational riboswitch. Acting in response to binding the small metabolite $preQ_1$, the Shine Dalgarno sequence of the mRNA is sequestered into the aptamer domain. We show that $preQ_1$ suppresses binding of the 30S ribosome to this mRNA, thus offering direct evidence for the biological function of the translational $preQ_1$ riboswitch.

Chapter 1 Introduction

1.1 Small RNAs as Model Systems

E Unus pluribum, or "Of One, Many", may be at the root of decoding the RNA sequence-structure-function relationship. RNAs facilitate the expression majority of genes in higher eukaryotes and fold in a sequence-directed fashion into three-dimensional structures that perform functions conserved across all cellular life forms, ranging from regulating to executing gene expression. While it is the most important determinant of RNA structure, the nucleotide sequence is generally not sufficient to specify a unique set of secondary and tertiary interactions due to the highly frustrated nature of RNA folding. This frustration results in an abundance of stable folded states and thus folding heterogeneity, a common phenomenon wherein a chemically homogeneous population of RNA molecules folds into multiple stable structures. Often, these alternative conformations constitute misfolds, lacking the biological activity of the natively folded RNA. Intriguingly, a number of RNAs have recently been described as capable of adopting multiple distinct conformations that all perform, or contribute to, the same function. Characteristically, these conformations interconvert slowly on the experimental timescale of minutes, suggesting that they should be regarded as distinct native states. In this dissertation, I discuss how rugged folding free energy landscapes give rise to multiple native states in the *Tetrahymena* Group I intron ribozyme, hairpin ribozyme,

sarcin-ricin loop, ribosome, and an *in vitro* selected aptamer. I further describe the varying degrees to which folding heterogeneity impacts function in these RNAs, and compare and contrast this impact with that of heterogeneities found in protein folding. Embracing that one sequence can give rise to multiple native folds, I hypothesize that this phenomenon imparts adaptive advantages on any functionally evolving RNA quasispecies. These ideas of heterogeneity are further explored experimentally in structural studies of the hairpin ribozyme and dynamics assays of ribosome initiation on a translational riboswitch.

The central dogma of molecular biology is a tried but true axiom of the direction of information flow in the cell: information in DNA passes to RNA passes to protein. (1) It is known that caveats to this basic scheme exist in the forms of reverse transcription and DNA and RNA replication, the basic principle of information flow holds true. And while Crick's original intent may have been focused on the fact that protein serves as an inescapable endpoint for genetic information, it was often RNA that was the overlooked aspect of this information flow. Serving as a mediator molecule between the information storage of DNA and its functional expression in protein, RNA held a limited role in the world of molecular biology. Even with the discovery of ribosomal RNA (rRNA) and transfer RNA (tRNA), the main purpose of RNA seemed to revolve around the production of protein. The ever expanding world of RNA function covers a wide breadth of activity, from intron splicing actions of the spliceosome to the down regulation of gene expression by microRNA (miRNA).

The discovery three decades ago that certain RNA molecules, termed ribozymes, catalyze chemical reactions in a manner similar to protein enzymes demonstrated an unexpected level of functional versatility of RNA that may have spawned life in the form of an RNA world.(2-4) Accordingly, over the last decade a large number of non-protein coding RNAs (ncRNAs) have been discovered that play essential roles in all aspects of modern life forms.(5-8) These roles include regulation of gene expression,(9, 10) post-transcriptional RNA processing,(3, 4, 11) protein biosynthesis,(12) and essential genomic processing in pathogens.(13-17) It was also discovered that some ribozyme motifs are broadly distributed among a wide set of organismal genomes.(18-21) Moreover, *in vitro*

selection has generated ribozymes with additional activities such as aminoacyl-RNA synthesis,(22) self-replication,(23) and organic synthesis,(24) all functions postulated to have played a pivotal role in the RNA world. Finally, a rapidly increasing number of crystal structures has shed light onto the impressive complexity of the underlying RNA structures.(25-27) Clearly, RNA has the capacity to assume a wide variety of functions based on the ability of its sequence to encode versatile three-dimensional structures, yet our understanding of the RNA sequence-structure-function relationship is still in its infancy.

Self-cleaving ribozymes are ideal model systems for the study of sequence-structure-function relationships in ncRNA, since their activity (and hence, proper folding) can be quickly and easily assayed.(25, 26, 28-31) These ribozymes catalyze a site-specific transesterification of the phosphate-ribose backbone resulting in the formation of two cleavage products: a 5′ product bearing a 2′-3′ cyclic phosphate, and a 3′ product bearing a 5′-OH group.

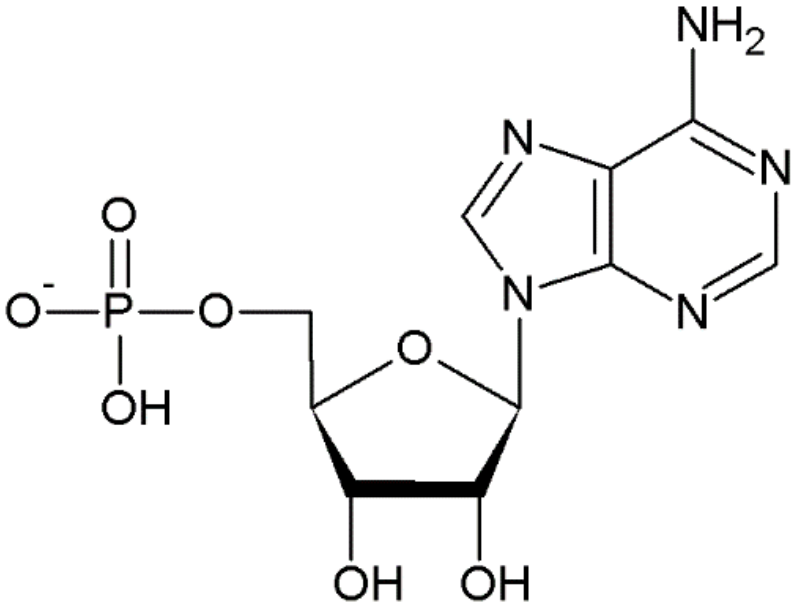


Figure 1: Struture of Adenosine Nucleotide.

The structure of the basic unit of RNA is shown with the nucleobase adenine attached at the 1' position of the ribose sugar. Hydroxyl groups occupy the 2' and 3' positions on the sugar, while a phosphate group occupies the 5' position. Connections between the 3'-hydroxyl and 5'-phosphate of a subsequent nucleotide form the backbone of RNA polymers. The resulting sequence of nucleotides controls the function of the resulting RNA by determining the structure and message contained within.

The activity of these ribozymes depends on the presence of metal cations – especially divalent cations such as Mg^{2+} – that stabilize the active tertiary structure of an RNA and may also confer a direct chemical rate enhancement.(32) Additionally, the relatively small size of most self-cleaving ribozymes allows for convenient *in vitro* transcription of large amounts of sample for biochemical assays and chemical synthesis to incorporate site-specific modifications and labels for chemogenetic and biophysical studies.

Perhaps due to the ease of relating folding to activity, studies of ribozymes have revealed a propensity of RNA to adopt alternate conformations, a phenomenon often described as conformational heterogeneity. While the first reversible misfolding of RNA was reported in leucyl-tRNA ($tRNA^{Leu}$),(33, 34) the prevalence of alternate kinetically stable structures in RNA became more apparent with studies of several self-cleaving ribozymes. In most cases, such conformational heterogeneity was attributed to "misfolded" (inactive or less active) ribozymes or long-lived folding intermediates.(35-38) In fact, virtually all ribozymes, as well as many other RNAs, are prone to this type of conformational heterogeneity, leading to a persistent view in the field that alternative folding is a nuisance to be avoided.(39, 40)

A particularly intriguing example of folding heterogeneity was recently characterized in the *Tetrahymena* group I intron (TG1I) ribozyme (**Figure 2**).(41) Upon binding of substrate to its 5′-end, the TG1I ribozyme forms a helix termed P1 that subsequently swings by ~6 nm to dock the substrate into the preformed active site and form the active complex (**Figure 2**A). Using single-molecule fluorescence resonance energy transfer (smFRET) measurements and cleavage activity assays, the authors provided evidence that subpopulations of the ribozyme exhibiting widely variable (>800-fold) docking equilibrium constants are, surprisingly, all catalytically active. In fact, 94% of all molecules within these different populations maintain the same rate constant of catalysis (**Figure 2**E,F). While only a small fraction of molecules spontaneously switches between subpopulations on an experimentally accessible time scale (i.e., the heterogeneity is relatively *static*), molecules can be induced to redistribute among subpopulations by refolding through the removal and subsequent reintroduction of Mg^{2+} ions (**Figure 2**B,C,D).(41) This finding strongly suggests that several active, or "native", states arise from conformational differences rather than changes in chemistry or local environment.

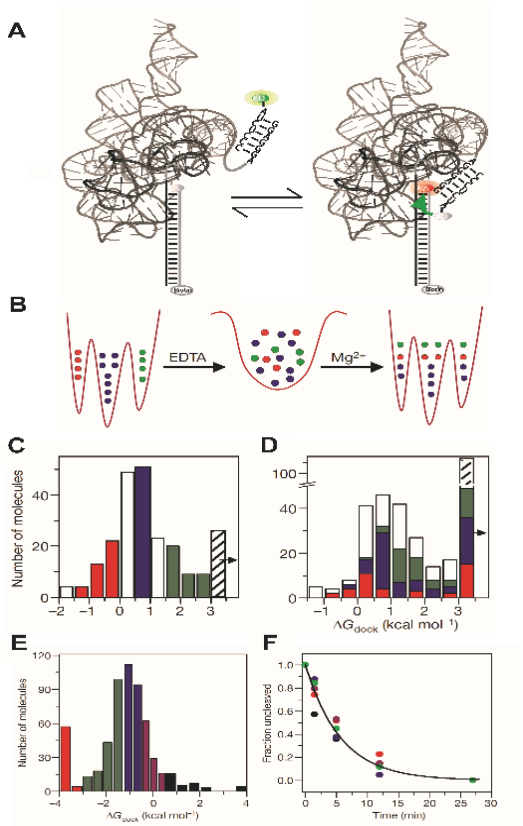


Figure 2 Single Molecule Observations of the TG1I Ribozyme Reveal Folding Heterogeneity That Manifests in Multiple Native States.

(A) Structural representation(42) of TG1I ribozyme docking as monitored by smFRET experiments. The sequence of helix P1, composed of the 5'-end of the ribozyme and the substrate RNA strand, is shown with donor fluorophore (D) attached to the 3'-end of the substrate strand. The TG1I ribozyme contains a 3'-extension that is hybridized to a DNA oligonucleotide (grey) with a 3'-acceptor fluorophore and 5'-biotin for surface immobilization and 3'-acceptor. (B) Schematic representation of the deeply furrowed folding landscape of the TG1I ribozyme. Molecules (red, blue, white and green) redistribute between three docking conformations after partial denaturation by removal (with EDTA) and addition of Mg^{2+}. (C, D) The three energy wells manifest as single molecule distributions of vastly different docking free energy (ΔG_{dock}), between which molecules (color-coded) redistribute from before (C) to after (D) denaturation. (E) The overall distribution of docking behaviors was binned into five color-coded categories. (F) Cleavage assays of molecules representing these five docking categories (color-coded) were all shown to display kinetics similar to the global average (solid line). In part adapted with permission from ref. (41).

This work on the TG1I ribozyme provides a strong impetus to revisit questions about the origin and possible biological function of folding heterogeneity in RNA. In fact, evidence of very similar behavior has been accruing for a number of functional RNAs over the past decade. In the following I will discuss how the physical properties of RNA give rise to a propensity for heterogeneous folding. Providing further examples, we will show that such folding behavior is commonplace, and in some cases clearly contributes to RNA function. Finally, I will speculate as to the significance of this behavior in the context of molecular adaptability and evolution.

1.2 The free energy landscape of RNA folding is rugged and frustrated

Among biopolymers, RNA possesses a number of characteristics that make its folding behavior unique. First, the multitude of dihedral angles in the phosphate-ribose backbone of RNA results in an immense range of possible topologies (or folds) for even relatively short RNAs. Second, the relative dominance of only a few types of base pairing interactions (Watson-Crick A·U and G·C, as well as common G·U wobble pairs) results in a "frustrated" folding landscape with a large number of nearly degenerate secondary structures. Third, the ability of RNA to form highly stable duplexes, cooperatively reinforced by a large number of hydrogen bonds and base-stacking interactions, gives its folding landscape a deeply furrowed character, resulting in alternate secondary and tertiary structures that may take very long to interconvert once formed.(43) Together, these factors give rise to a high propensity to form kinetically trapped alternate folds.

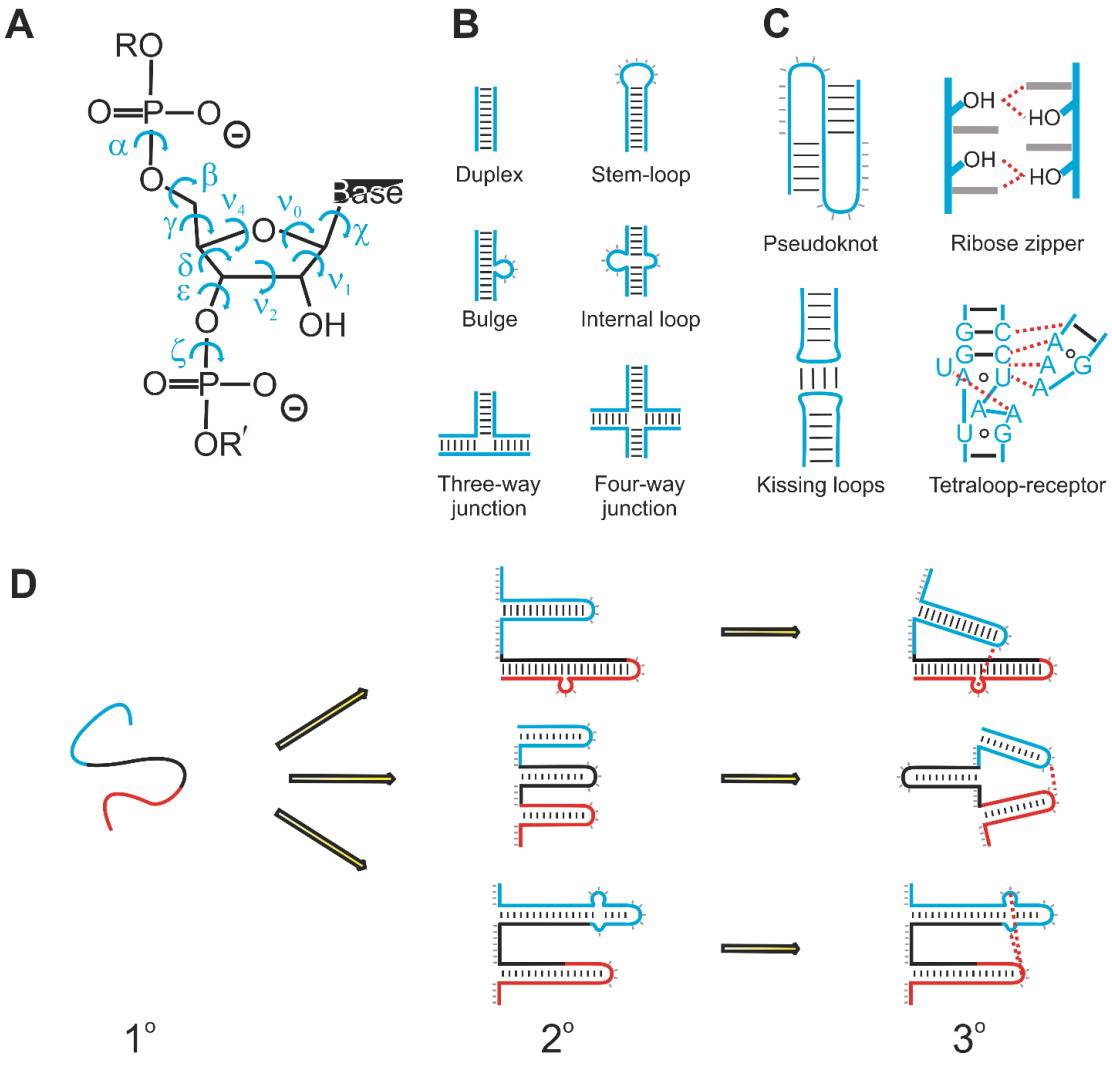


Figure 3 Diversity of RNA structures.

(A) RNA has many dihedral angles (blue) that contribute to a large number of conformational degrees of freedom per nucleotide. These include angles of torsion about the bonds of the RNA backbone (α-ζ), the nucleosidic bond (χ), and pseudo-rotation angles within the ribose ring (ν_0-ν_4). (B) Hydrogen bonding (pairing) between complementary bases gives rise to several common secondary structure motifs, consisting of the RNA backbone (blue) held in various arrangements by base pairs (black line segments) and often resulting in short segments of unpaired nucleotides (short grey lines). (C) Common tertiary structure motifs build hierarchically onto secondary structure elements. (D) The small alphabet of RNA often results in a furrowed or frustrated folding landscape with many possible conformations stabilized by alternative base-pairing. In this schematic, three segments of a linear RNA molecule are marked in blue, black, and red to illustrate the variety of possible secondary (2^o) and tertiary (3^o) structures arising from a single primary (1^o) sequence of nucleotides.

The ribose-phosphate backbone contains six rotatable dihedral angles per nucleotide (conventionally labeled α-ζ, **Figure 3**A), or even more when the constrained torsional angles of the sugar ring (ν_0-ν_4,**Figure 3**A) and the glycosidic bond dihedral (χ, Fig. 3A) are considered. This number compares to only two such angles per amino acid in proteins, resulting in a much wider range of possible conformations in RNA than in peptides of comparable size. While Watson-Crick base pairing in standard A-form helical stems places strict restraints on the possible torsional angles, single-stranded regions of RNA molecules – formally junctions, loops, and bulges (**Figure 3**B) – still have a multitude of possible conformations. As a result, the folding topology of an RNA can be rendered quite complex through the formation of multiple helical junctions and tertiary interactions such as pseudoknots, ribose zippers, kissing-loop interactions, and tetraloop-receptor interactions (**Figure 3**C).(44-49)

Several advances have been made towards classifying combinations of dihedral angles into structural motifs based on mono- or dinucleotide units, reducing the number of empirically observed conformations considerably. In one approach, backbone dihedral angles were organized into so-called "suites", where a conformation is defined between adjacent sugar residues as a group of two sets of angles: δ-ε-ζ and α-β-γ-δ.(50) Along with careful quality-based filtering of crystallographic source data, this system enables the classification of empirically observed RNA backbone conformations into 42 suites, later expanded to 46 conformers in an effort that unified a handful of other approaches.(51) Still, the number of conformations available to even a short

oligonucleotide would be staggering in the absence of other information, posing a serious obstacle to decoding the relationship between sequence and structure.

Fortunately, the complexity of RNA folding is reduced considerably by its hierarchical nature wherein secondary structure typically folds before tertiary structure (**Figure 3**D).(40, 52) Hybridization of complementary segments of an RNA sequence occurs in as little as microseconds.(29, 53) Once formed, an A-form RNA helix is stabilized by base pairing and base stacking interactions worth about -1 to -3 kcal/mol per base pair.(43, 52) As a result, even short RNA oligomers of ~10 base pairs (bp) may have half-lives of dissociation on the order of minutes or hours near room temperature.(54, 55) On the other hand, tertiary interactions (loop-loop hydrogen bonding, base-phosphate and base-sugar interactions) frequently form and interconvert on the timescale of milliseconds to seconds and are comparatively weak.(56-58) Because of this difference in kinetics and stability, secondary structure places relatively rigid constraints on accessible tertiary structures; conversely, isolated stem-loops of a larger RNA structure often fold properly even in absence of tertiary interactions.(43) While it simplifies the prediction of RNA structure, the stability of secondary structure also gives rise to a deeply furrowed free energy landscape of folding with large barriers separating different folded states.

Finally, the folding free energy landscape of RNA is described as deeply furrowed or "frustrated" by a large number of possible alternative folds (**Figure 3**D). This description derives from the tendency of RNA to become trapped in local minima and require great energy to overcome these barriers and attain "native" structure. Due to

the diminutive 4-nucleobase alphabet of RNA, there is a high probability that any two sequences of nucleotides will have coincidentally complementary regions. Even random RNA sequences containing tens of nucleotides are predicted to be approximately 50% base-paired,(59) implying the existence of numerous possible alternative secondary structures. Modern software packages use partition function approaches to predict RNA secondary structure, yielding more reliable predictions of base pairing and thermodynamic parameters than considering only the minimum-energy structure.(60-63) Clearly, the number and impact of possible alternate secondary structures are significant.

1.3 Multiple active species with slow interconversion are observed in a number of RNAs

Due to the complex and rugged folding free energy landscape of RNA, long-lived variations in structure, or so-called static heterogeneities, arise in RNAs as diverse as those found in plant virus satellites, the eukaryotic ribosome, and artificial selections of ligand-binding aptamers.(41, 56, 64-66) Misfolding is perhaps a trivial example, but it is important to consider that one function's trash may be another function's treasure. An example is the adoption of different secondary structures for the purpose of switching between active and inactive forms of a ribozyme, such as at different stages in the replication cycle of a pathogen. For instance, the hepatitis delta virus (HDV) ribozyme must be active in order to cleave concatemeric linear transcripts of HDV RNA into unit-length fragments for ligation into circular copies of the genome. Once this task is accomplished, however, the ribozyme becomes inactive by adopting alternate base

pairing patterns as part of a long, 70% self-complementary, rod-like structure of the RNA genome of HDV, presumably for the purpose of packaging and delivery of the intact genome to new host cells.(17)

Here, I focus on another type of static heterogeneity: the presence of distinct species that all contribute, in varying degrees, to the same nominal function. In investigating such phenomena, several questions must be addressed:

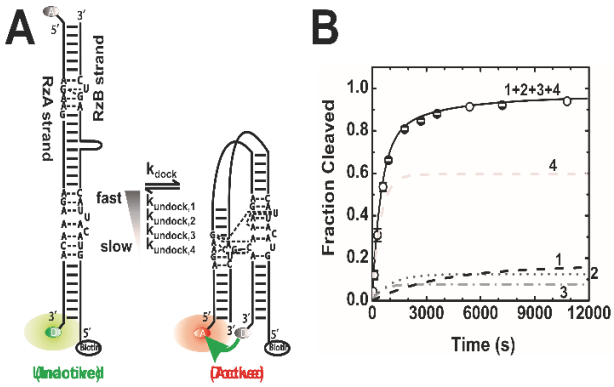
1. What impact does the structural heterogeneity have on function?
2. How robust is the heterogeneity to changes in experimental conditions?
3. Does the heterogeneity arise from purely conformational differences, i.e., can covalent modification or mutation be ruled out?
4. If the differences are conformational, do they represent alternate secondary structures or more subtle variations in tertiary structure?

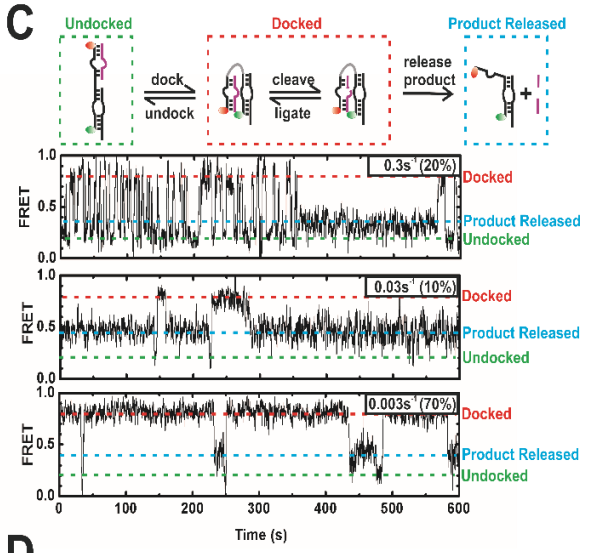
Studies of heterogeneity in model systems have begun to answer these questions. Such studies often involve a combination of ensemble and single-molecule methods, utilizing the relative strengths of each approach to deduce the nature of the heterogeneity(41, 65-67) Ensemble assays of RNA are rapid, well established for many systems, and can employ a variety of detection methods, the two most common of which are fluorescence and autoradiography. While requiring relatively large amounts of sample material, ensemble assays often require minimal modification and processing of sample material. Single-molecule fluorescence based assays generally require precise fluorescent labeling, specialized equipment, and in-depth statistical analysis, but work with low amounts of sample. Additionally, single-molecule assays allow correlations to

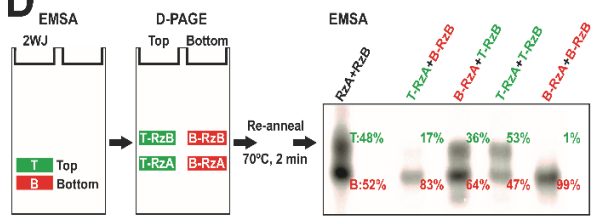
be drawn between conformational behavior and catalytic efficiency without physical separation. Only through utilizing these techniques in conjunction can the nature of these heterogeneities in RNA be addressed, as illustrated in the following examples.

1.3.1 The hairpin ribozyme

One of the most thoroughly characterized examples of static heterogeneity in an RNA is found in the hairpin ribozyme, a self-cleaving and self-ligating small ribozyme first discovered in the negative strand of the tobacco ringspot virus (TRSV) satellite RNA.(68, 69) Central to catalytic activity of the ribozyme is the docking of its two helix-loop-helix domains (A and B) by tertiary interactions between the nucleotides in bulges present in each helix to form the active site. At room temperature, in the presence of Mg^{2+} ions, this docking is readily reversible (Fig. 4A). Despite its much smaller size, the docking of a helix (domain A) into the catalytic core (domain B) of the hairpin ribozyme, enabled by a flexible hinge region, superficially resembles docking of helix P1 into the TG1I ribozyme (Compare Figs. 2A and 4A).


A
RzB strand
RzA strand
fast
slow
Biotin
B
Fraction Cleaved
Time (s)
1+2+3+4
4
1
2
3


C
Undocked
Docked
Product Released
dock
undock
cleave
ligate
release product
FRET
0.3s⁻¹ (20%)
0.03s⁻¹ (10%)
0.003s⁻¹ (70%)
Docked
Product Released
Undocked
Time (s)


D
EMSA
2WJ
Top
Bottom
D-PAGE
T-RzB
B-RzB
T-RzA
B-RzA
Re-anneal
70°C, 2 min
EMSA
RzA+RzB
T-RzA+B-RzB
B-RzA+T-RzB
T-RzA+T-RzB
B-RzA+B-RzB
T:48%
B:52%
17%
83%
36%
64%
53%
47%
1%
99%

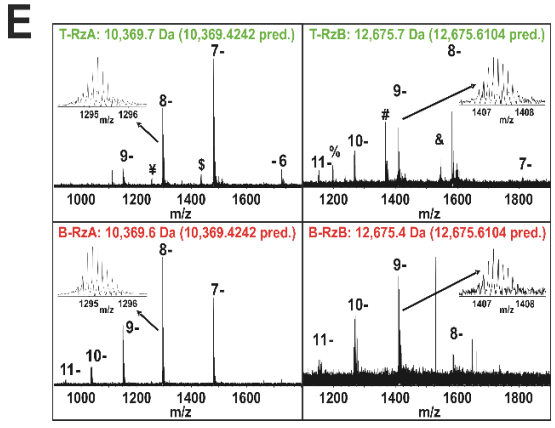

E
T-RzA: 10,369.7 Da (10,369.4242 pred.)
T-RzB: 12,675.7 Da (12,675.6104 pred.)
B-RzA: 10,369.6 Da (10,369.4242 pred.)
B-RzB: 12,675.4 Da (12,675.6104 pred.)
m/z

Figure 4: Folding Heterogeneity of the Hairpin Ribozyme.

The frustrated folding landscape of the hairpin ribozyme results in heterogeneity of structure and function. (A) smFRET assays detect docking heterogeneity of the hairpin ribozyme. Donor (D) and acceptor (A) fluorophores are attached to the 3'- and 5'-ends of the RzA strand, respectively, while the RzB strand carries a 5'-biotin for surface immobilization. Whereas all active HpRz molecules display a single docking rate constant (k_{dock}), four distinct undocking rate constants ($k_{undock,1-4}$) are observed. (B) Each docking rate constant leads to a distinct cleavage time course (numbered as in panel A), which taken together account for the overall cleavage observed in ensemble assays (1+2+3+4). (C) Reaction pathway of the hairpin ribozyme and resulting single-molecule multiple-turnover cleavage data. The docked, undocked and product released states are indicated. Three single molecule time trajectories demonstrate catalytic proficiency of each of the distinct subpopulations they represent (undocking rate constants and the fraction of molecules undocking with this rate constant are indicated). (D) Structural heterogeneity upon EMSA in which two structural forms of the hairpin ribozyme are resolved (Top and Bottom species). Separation of the two component strands RzA and RzB of each species by denaturing polyacrylamide gel electrophoresis (D-PAGE) yields four RNAs, as indicated, that were annealed in all possible combinations and analyzed by EMSA. The resulting fractions of Top and Bottom species are given. (E) High-resolution FT-ICR mass spectrometry of each of the four RNA strands isolated by the procedure described in panel D reveals identical isotope envelopes (insets) and average masses (consistent with the predicted masses) of the corresponding strands from the Top and Bottom species. In part adapted with permission from refs. (67, 70).

Utilizing single molecule fluorescence resonance energy transfer (smFRET) assays to quantify the kinetics of these docking and undocking transitions, the Walter and Chu groups first uncovered kinetic heterogeneities in the hairpin ribozyme in 2002.(57) Only by utilizing smFRET could the behaviors of individual molecules be observed without the ensemble averaging inherent to bulk assays. For smFRET detection, the opposite ends of one strand of the RNA were labeled with the fluorophores Cy3 (smFRET donor) and Cy5 (acceptor), which exhibit high FRET efficiency when proximal to each other in the docked state and low FRET efficiency in the undocked state (**Figure 4**A). FRET efficiency (E) is defined by the following equation:

$$E = \frac{1}{1+(\frac{R}{R_0})^6}$$

where R is the distance between the donor and acceptor fluorophores and R_0 is the Förster distance of the fluorophores, the distance at which there is a 50% transfer efficiency.

Samples were immobilized on a derivatized quartz slide, and Cy3 and Cy5 emission intensities of single molecules were observed by total internal reflection fluorescence (TIRF) microscopy.(57, 71) Although the rate constant of docking was invariant across all molecules, subpopulations of ribozyme molecules were observed to undock with four distinct rate constants spanning three orders of magnitude. Furthermore, individual molecules rarely (<5% of observed molecules) switched between these kinetic regimes even over 3 hours incubation – each molecule retained a "memory" of its undocking rate constant. Strikingly, the multiple undocking populations were all observed to convert to product from the docked state; i.e., the faster the undocking relative to the cleavage/ligation rate constants, the more docking/undocking cycles become necessary on average for a successful substrate turnover (**Figure 4B and 4C**). Thus, the existence of distinctly undocking, yet catalytically active, native states quantitatively explains the multi-exponential cleavage kinetics noted in ensemble assays of the hairpin ribozyme (**Figure 4A and 4B**).(57, 70)

Importantly, it was shown that this heterogeneity is not caused by the surface immobilization used in smFRET measurements. First, bulk cleavage assays of ribozymes with the same modifications (Cy3, Cy5, and biotin) yield similar kinetic parameters to those obtained by smFRET techniques.(57, 70, 72) Second, the same kinetic

heterogeneity was observed by smFRET when the Lilley and Ha groups captured ribozymes on the slide surface by encapsulation within phospholipid vesicles, rather than through direct biotin-streptavidin interaction.(73)

Surprisingly, the level of undocking heterogeneity proved not to be affected by either changes in Mg^{2+} concentration(74) or site-specific mutations or modifications that impacted the undocking rate constants; the rate constant of each subpopulation simply shifted by about the same factor.(70) In subsequent work, evidence was presented that the molecular heterogeneities of the hairpin ribozyme, while extremely long-lived, are not due to any detectable covalent modification.(67) More specifically, through the use of electrophoretic mobility shift assays (EMSAs) on polyacrylamide gels, two distinct, either slow- or fast-migrating species of the ribozyme were resolved (as top and bottom bands, respectively) . Upon elution from the gel and further analysis of the fluorophore distance distributions by time-resolved FRET (trFRET), the two species showed marked differences; while less than 40% of the slow-migrating species exists in the docked conformation at any given time, greater than 80% of the fast-migrating species exists the docked form. This shows a much greater propensity of the bottom band HpRz to be in the compact docked form, explaining the likely cause of it greater migration in gel. Analysis by smFRET showed the top band is enriched in the fastest undocking subpopulation, which is expected to be less compact, consistent with its lower mobility, whereas the bottom band is enriched in the most slowly undocking subpopulation, expected to reside largely in the compact docked conformation.(67)

This differential enrichment of the docked state within two separable species opened up an opportunity to determine whether the hairpin ribozyme subpopulations can be induced to redistribute. To this end, the individual 5'- and 3'-segment strands (termed RzA and RzB) of each species were further separated by denaturing gel electrophoresis, removing all base pairing between them. When the RzA and RzB strands of the slow-migrating species were re-annealed and re-analyzed by EMSA, the RNA redistributed into fast- and slow-migrating species as before, but the RzA and RzB strands originating from the fast-migrating species continued to preferentially form the fast-migrating species (**Figure 4**D). Accordingly, mixing RzA and RzB strands originating from different species yields intermediate levels of the two EMSA bands (Fig. 4D). This asymmetry suggests that the molecular heterogeneity leading to formation of the catalytically more active fast-migrating species is maintained even upon full denaturation of all interstrand base pairs.(67)

In the same work, we were also able to show by high-resolution (<1 amu) mass spectrometry that the RzA and RzB strands from the fast- and slow-migrating strands are identical in mass (**Figure 4**E).(67) In addition, both chemically synthesized and *in vitro* transcribed ribozyme behave similarly, and footprinting revealed only minor differences in secondary structure between the fast and slow migrating species. While these observations do not rule out mass-neutral covalent modifications as the source of conformational heterogeneity, such as certain UV-induced crosslinks, they do strongly support the notion that more subtle conformational(75) or topological differences(41) are at play.

In summary, the hairpin ribozyme folds into multiple active populations with disparate global dynamics but only subtle differences in secondary structure. Compared to the multiple native states of the TG1I ribozyme many parallels are observed, although the native states of the hairpin ribozyme are separated by larger free energy barriers and, consequently, do not interconvert quite as freely upon refolding as those of the TG1I ribozyme (**Figure 5**).

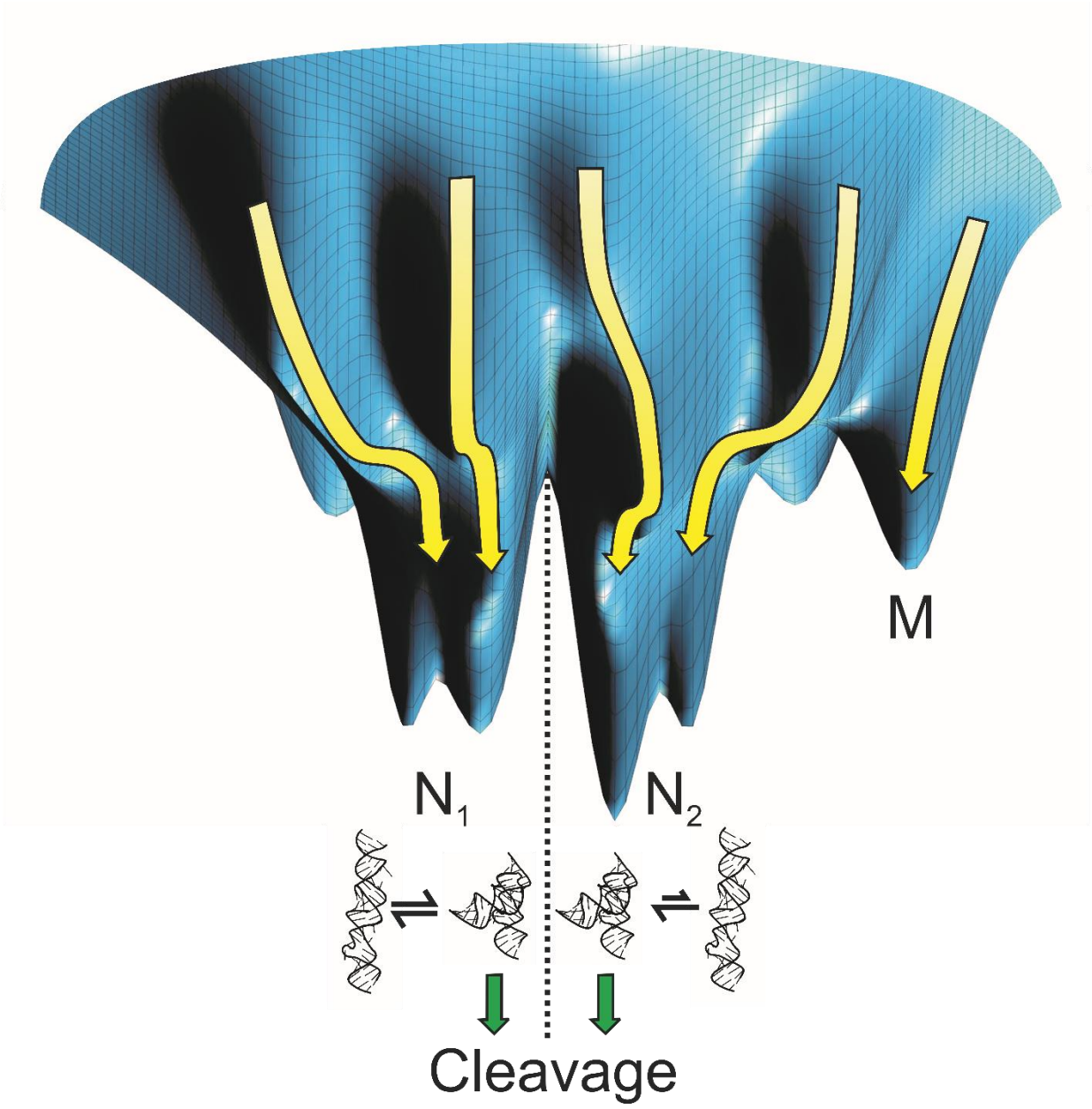


Figure 5: Schematic Representation of the Rugged Conformational Free Energy Landscape (Blue Surface) of the Hairpin Ribozyme.

The folding landscape of a biomolecule can be repsresented be a folding landscape of free energy (blue) that a molecule traverses as it folds into a stable form. An individual molecule folds along one of many possible pathways (yellow arrows) to one of multiple native states (N_1, N_2) separated by relatively large energy barriers. These native states sample similar conformations, albeit with different kinetics, and thus possess similar cleavage activity. Alternatively, the molecule may enter a trapped misfolded state (M)

that is non-functional. This diversity of folds can lead to heterogeneity of function in a given molecule.

1.3.2 The sarcin-ricin loop

Domain B of the hairpin ribozyme shares sequence homology and an RNA structural motif termed an S-turn with the toxin-sensitive sarcin-ricin loop (SRL) of the large subunit ribosomal RNA (rRNA, **Figure 6**A).(67) Intriguingly, conformational heterogeneity has also been observed in this stem-loop motif that is highly conserved across all kingdoms.(64) Several 27-nt versions of the SRL from rat were transcribed *in vitro* and analyzed by EMSA, revealing two species of identical length but with different electrophoretic mobilities (**Figure 6**B), similar to our observations on the hairpin ribozyme. No heat-induced interconversion of the two species was observed, suggesting that the heterogeneity is thermodynamically quite stable. Intriguingly, the slow-migrating species (constituting 30-50% of the total SRL) is resistant to cleavage by the endoribonuclease restrictocin (a sarcin analog) when compared to the native substrate (70-50% of total SRL, **Figure 6**).(64) As with the hairpin ribozyme, chemically synthesized SRL shows the same heterogeneity. While it is unknown whether this phenomenon persists *in vivo*, or how it might affect translation by the ribosome, i.e., which of the species represents a native state, it is tempting to speculate that the presence of multiple folds could confer an adaptive advantage through partial resistance to different toxins. Conformational heterogeneity may thus be akin to sequence variation of rRNA; in *E. coli* alone there are seven different rRNA sequences, each with minor sequence discrepancies that may confer increased adaptability to the cell.

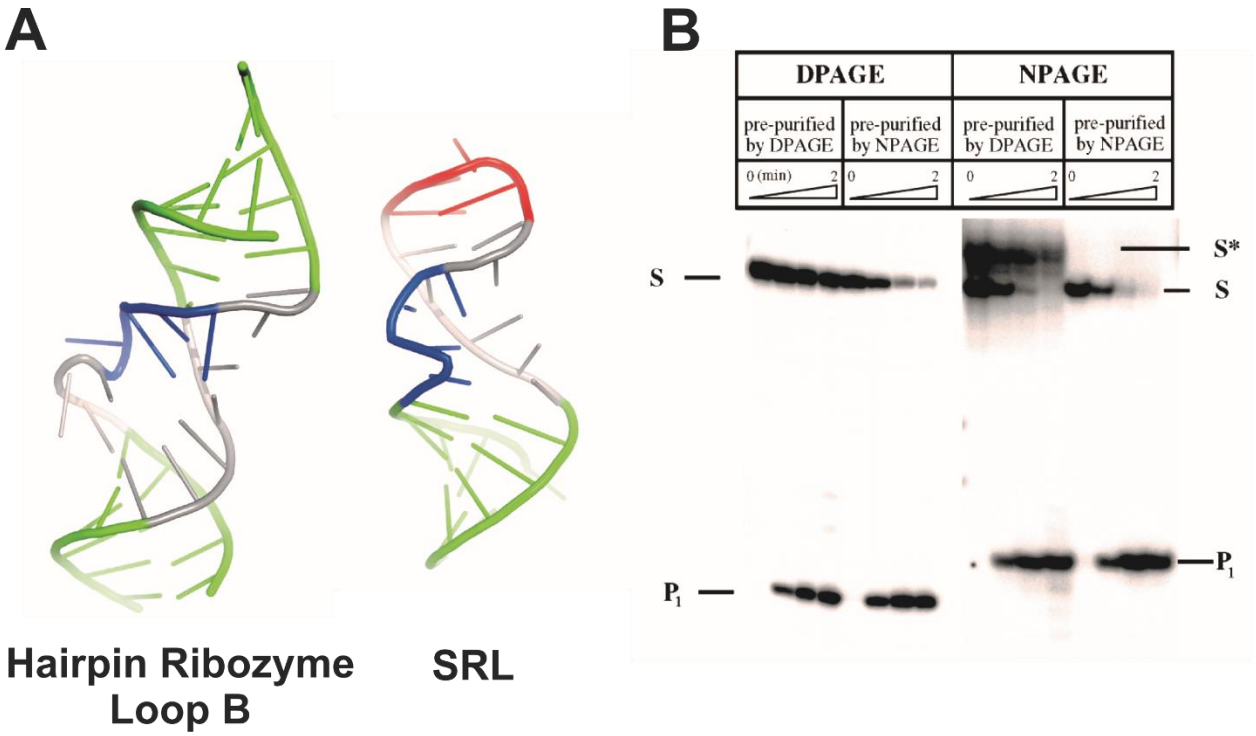


Figure 6: Folding Heterogeneity of the Sarcin-Ricin Loop (SRL).

The sarcin-ricin loop is a structure element of the ribosome which contains a similar structure and heterogeneity to the hairpin ribozyme. (A) Comparative cartoon representations of crystal structures of loop B of the docked hairpin ribozyme and the SRL reveal a common, conserved S-turn motif (blue).(67) (B) *In vitro* transcribed SRL was either purified by denaturing (DPAGE) or non-denaturing polyacrylamide gel electrophoresis (NPAGE or EMSA), subjected to restrictocin cleavage, and the products over time analyzed by gel electrophoresis, as indicated.(64) EMSA in particular reveals a non-interconvertable, slow-migrating S* species that is relatively resistant to restrictocin cleavage. Adapted with permission from refs. (64, 67).

1.3.3 The AN58 aptamer

So far, I have discussed the static heterogeneities observed in naturally occurring RNAs, yet there exists at least one artificially selected RNA aptamer that shows similar behavior. The AN58 RNA is a truncation construct of the larger aptGluR2-99 aptamer, *in vitro* selected to block the GluR2 α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic

acid (AMPA) receptor by binding to regions thought to be essential to function.(76) When AN58 was transcribed *in vitro* and purified by gel electrophoresis, two different, non-interconvertable populations (M1 and M2) were observed by EMSA (**Figure 7**). Neither of these isoforms can individually inhibit the GluR2 AMPA receptor, yet when recombined, they show inhibition comparable to that of unseparated AN58 and very similar to that of the parent aptGluR2-99 (**Figure 7**B). Sequencing by primer extension and glyoxal treatment followed by gel electrophoretic analysis showed that M1 and M2 have the same length and sequence. To elucidate any differences in secondary structure between M1 and M2, a combination of in-line probing and selective 2'-hydroxyl acetylation analyzed by primer extension (SHAPE) was utilized. Results from these footprinting assays suggested that M1 and M2 form short stem-loop structures in alternate regions of the unstructured 3′-region of the RNA (**Figure 7**D).(65) Thus, in contrast to the hairpin ribozyme, in this case alternate, very stable secondary structures appear to be responsible for the heterogeneous behavior. Yet AN58 maintains its static heterogeneity in a manner similar to both hairpin ribozyme and SRL (**Figure 7**C).(65)

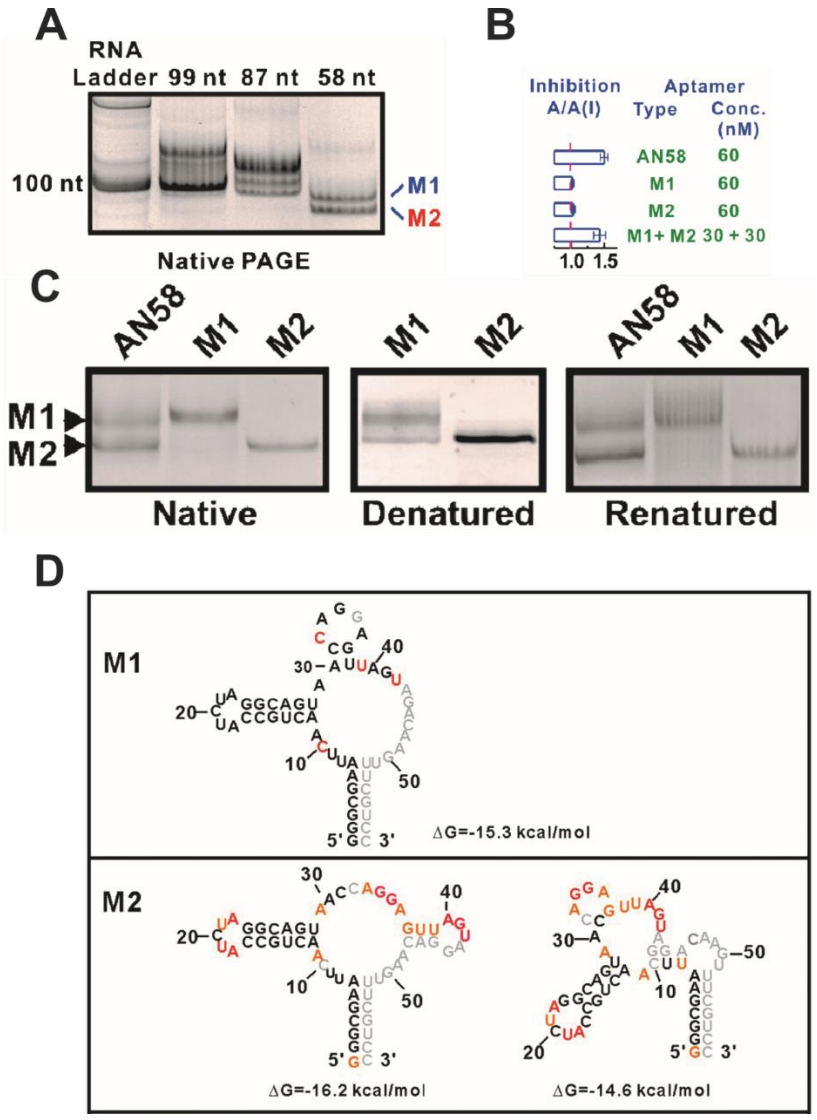


Figure 7 Folding Heterogeneity of the AN58 Aptamer.

Heterogeneity is seen in artificially derived RNA as well as those from nature. (A) AN58, a truncation of a larger AN99 aptamer, resolves as two discrete bands during EMSA (Native PAGE), termed M1 and M2, suggesting two different structures of chemically homogenous RNA. (B) While AN58 is capable of blocking the action of the GluR2 AMPA receptor channel, neither purified form (M1 or M2) can inhibit activity on its own, yet a mix of both M1 and M2 results in restoration of AN58 function. (C) M1 and M2 were separated from one another by EMSA and analyzed by EMSA (left panel), denatured by urea-containing polyacrylamide gel electrophoresis and visualized (middle panel, Denatured), then refolded and further analyzed by EMSA (right panel, Renatured), revealing a lack of interconversion between species. (D) Proposed secondary structures of M1 and M2. Adapted with permission from ref. (65).

1.3.4 The ribosome

Given that a small stem-loop of the ribosome exhibits profound folding heterogeneity, it comes as no surprise that the bacterial ribosome, at ~2.4 MDa and with three RNA and >50 protein components, the largest of all ribozymes, has recently been shown to display heterogenous intersubunit rotational dynamics in its pre-translocation complex (Fig. 8). Pre-translocation complexes (pre-complexes) occur when the acetyl(A)-site and peptidyl(P)-sites of the ribosome are occupied by tRNAs that have undergone a peptidyl transfer but have yet to translocate to the P- and E-sites, respectively. As shown in previous ensemble and single-molecule FRET based studies,(77-80) the small (30S) and large (50S) subunits of the ribosome spontaneously ratchet among the classic pre-complex (with the two tRNAs occupying the A/A-P/P sites in the 30S/50S subunits) and two hybrid conformations (Hy1: A/A-P/E, Hy2: A/P-P/E). Subsequently, elongation factor-G catalyzes translocation to the post-translocation complex (P/P-E/E). Heterogeneity arises in that only ~2/3 of all ribosomal complexes observed by smFRET display dynamic fluctuation between the classic and hybrid states (**Figure 8**).(66, 79) Of these, a roughly equal distribution of complexes exhibit transitions either between the Hy1 and the classic state or the Hy1 and the Hy2 state. The remaining ~1/3 of complexes are statically distributed among the low FRET classic and hybrid states with a high prevalence for the classic state, along with a small number of molecules occupying the high FRET POST state (**Figure 8**C).(66) The authors propose a qualitative folding free energy landscape of translocation with high energy barriers

preventing reverse translocation and smaller minima/maxima for the fluctuating and nonfluctuating FRET states, downwardly trending towards the Hy2 state that is adopted just before translocation.(66)

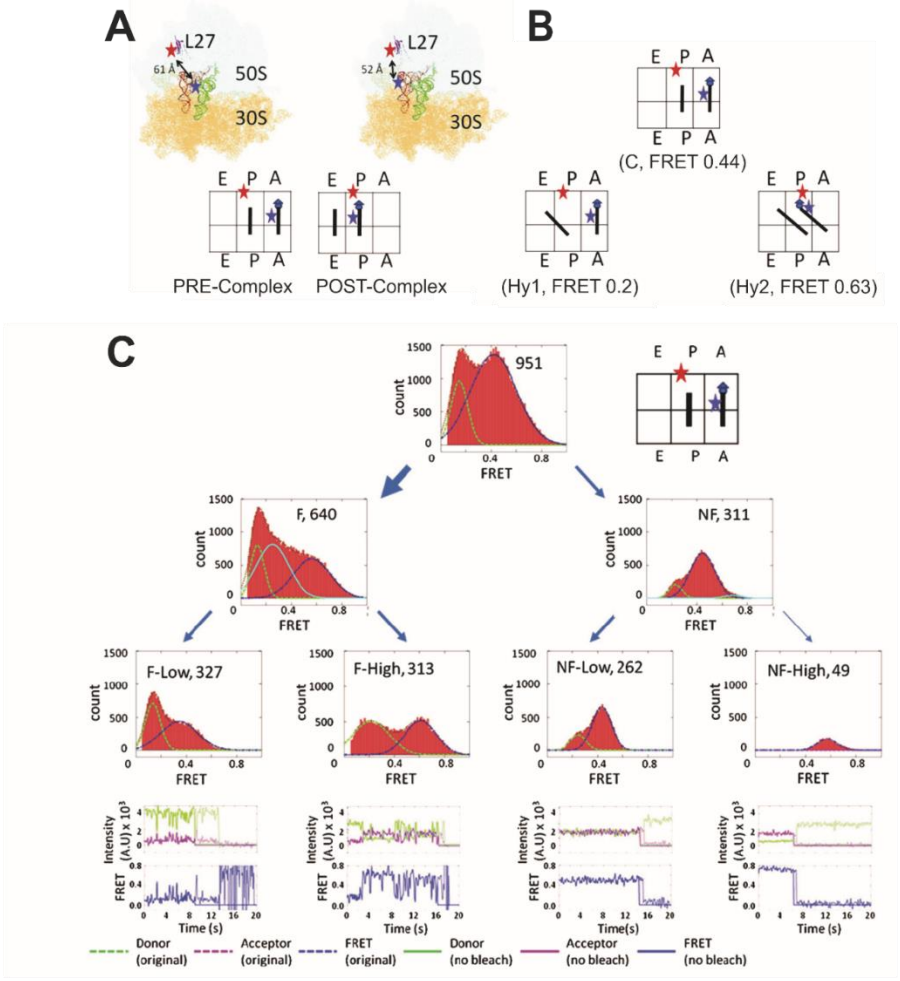


Figure 8: Folding Heterogeneity of the Bacterial Ribosome.(66)

(A) Structural and schematic representation of the fluorophore labeled *E. coli* ribosome in the pretranslocation (PRE) and posttranslocation (POST) complexes. Fluorophores attached to the L27 protein of the large subunit and A-site tRNA are used to characterize by smFRET their relative motions as the ribosome samples multiple PRE-complex conformations. (B) Ribosomes are observed in conformations including the classic state, C, along with two hybrid states, Hy1 and Hy2, with tRNA occupying various sites in the large and small subunit, each with a distinct FRET value, as indicated. (C) Ribosomes can be categorized by their dynamic (or fluctuating, F) or static (nonfluctuating, NF) occupancy of these conformations. Among the F molecules, roughly half exhibit low (0.2) to mid (0.44) FRET transitions while the other half exhibit low to high (0.63) transitions. Among the NF molecules, three categories emerge, where two categories with the low and mid FRET values of the C and Hy1 states, respectively, were described as a joint NF-Low category, whereas a high FRET category was assigned to POST complexes. Representative smFRET time trajectories accompany each of these four categories. Adapted with permission from ref. (66).

1.3.5 Riboswitches

While the ribosome itself displays multiple layers of heterogeneity in its own action, heterogeneous interactions with mRNA are an active form of gene regulation in bacterial systems. In the process of initiation, the 30S ribosomes, along with initiation factors (IFs) 1/2/3 and fMet-tRNA^{fMet}, assembles on an mRNA to form the pre-initiation complex (PIC). This is followed by the 50S ribosome binding to 30S ribosome and the dissociation of IF1 and IF3, resulting in the initiation complex (IC). While this process is a highly regulated of conversion of an mRNA message into a functional protein, it is complicated further by translational riboswitches.

Riboswitches are small, highly structured RNA motifs present in mRNA capable of modulating gene expression. Situated within an mRNA transcript, a riboswitch acts in response to the presence or absence of a small molecule ligand. These ligands are often small metabolites and their interaction with riboswitches is concentration dependent. Transcriptional riboswitches act to attenuate the transcription of mRNA by forming termination hairpins and leading to premature transcription termination. Alternately, translational riboswitches act to prevent translation by occluding the ribosomal binding site within the RNA and preventing ribosomal initiation (81).

In summary, while likely only the tip of the proverbial iceberg, the above five examples (TG1I and hairpin ribozymes, SRL, AN58 aptamer, and bacterial ribosome) arguably represent the best characterized occurrences of heterogeneous RNA behaviors. In each case, different copies of what is ostensibly a single chemical species with a

defined nucleotide sequence are capable of adopting different conformations of similar or distinguishable native functionality.

1.4. Parallels with protein folding

Often, insights into RNA structure-function relationships parallel those of proteins; it is therefore helpful to take a look at our current understanding of heterogeneities in protein folding. Like RNA, proteins fold hierarchically, with local interactions forming first and largely determining the overall structure of the folded polypeptide.(82, 83) Both RNA and proteins can adopt a given fold with very few specific sequence requirements.(20, 84) However, there are some important differences. Protein folding is mainly directed by fairly nonspecific hydrophobic collapse(83, 85) rather than by the formation of specific hydrogen bonds such as in tightly aligned (stacked) RNA base pairs. Furthermore, the native states of proteins are typically only marginally stabilized by 5-10 kcal/mol relative to their denatured states, which is comparable to the stability of an RNA duplex containing a mere 8-10 base pairs.(86, 87) In other words, a short RNA stem-loop folds with similar thermodynamic stability as an entire protein. This distinction is reflected by the fact that, whereas RNA secondary structure elements fold very stably in isolation,(43) the secondary structure of proteins is often strongly influenced by context (such as tertiary interactions).(88) Thus, one might expect to observe less profoundly heterogeneous folding in proteins than in RNA.

Ensemble kinetic experiments on protein folding have frequently observed trapped conformations. Many of these are non-native folding intermediates that generally

persist for only seconds or less, as for lysozyme(89) and staphylococcal nuclease.(90) In other cases, states with native-like activity have been observed. For instance, upon refolding from 8 M urea, the majority of dihydrofolate reductase folds transiently into an intermediate that efficiently binds a substrate analog.(91) Similar behavior was observed for RNase A, which has a long-lived folding intermediate possessing enzymatic activity similar to that of the native state in spite of some structural differences.(92) In both of these proteins, the native state ultimately forms, even though RNase A takes from one to several minutes to fold to completion.

The advent of single-molecule fluorescence spectroscopy overcame the drawbacks of ensemble averaging and revealed numerous examples of heterogeneous activity and folding of single protein enzymes, showing that this property is the rule rather than the exception. Individual lipase molecules were shown to exhibit fluctuating substrate turnover kinetics, likely explained by conformational changes on the order of tens of milliseconds.(93) Single-molecule studies of flavoenzymes, monitored by changes in the intrinsic fluorescence of the flavin cofactor in the course of its redox chemistry or electron transfer to a nearby tyrosine, found fluctuations as slow as 1 s^{-1} in both the substrate turnover kinetics and conformations of individual enzymes (Figure 9A-C),(94, 95) as did similar studies of horseradish peroxidase.(96) Longer-lived fluctuations in activity, with lifetimes on the order of minutes, were observed for single molecules of bacteriophage λ exonuclease.(97) The activity of electrophoretically purified lactate dehydrogenase molecules was found to vary by a factor of four in a manner that remained constant for a given single molecule over two hours, which the

authors suggested could be due to different stable arrangements of monomers in the homotetrameric enzyme.(98, 99) Similar findings of static heterogeneity were made for alkaline phosphatase,(100) β-galactosidase,(101) and the DNA helicase RecBCD.(102, 103) Perhaps most intriguingly, recent folding studies of GFPmut2, a triple mutant with enhanced fluorescence emission termed provide evidence of multiple native states, characterized by distinct chromophore switching kinetics, that do not interconvert over several hours unless refolded from the denatured state,(104-106) providing strong evidence of long-lived conformational heterogeneity in native proteins (Figs. 9D and 9E).

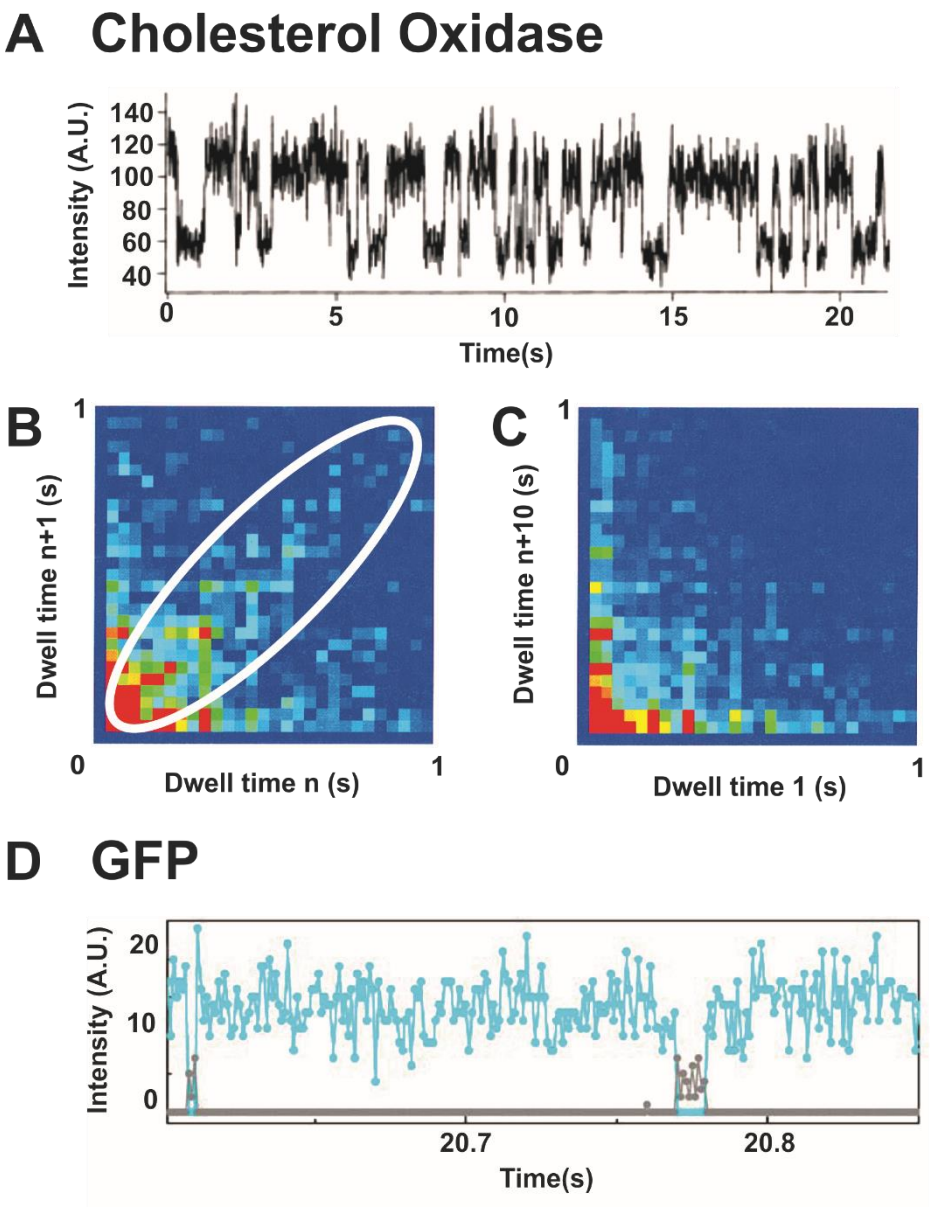


Figure 9 Examples of Folding Heterogeneity in the Proteins Cholesterol oxidase(94) and Green Fluorescent Protein (GFP).(106)

(A) Turnover of substrates by an individual molecule of cholesterol oxidase gives rise to stochastic transitions between fluorescent and non-fluorescent states. (B) Although transitions between these two states occur stochastically, dwell times in the fluorescent state for two adjacent turnovers n and n+1 are slightly correlated, as shown by the diagonal feature (encircled by a white ellipse) of a conditional probability distribution. (C) In contrast, dwell times separated by 10 turnovers are not correlated. (D) Single GFPmut2 molecules exhibit spontaneous switching between anionic (fluorescence intensity in cyan) and neutral (grey) states. Adapted with permission from ref. (94, 106).

In summary, like the RNA examples highlighted above, many proteins exhibit conformational heterogeneity that generally lasts for milliseconds to seconds, but can persist for hours in some proteins. The commonly shorter timescale of most of these protein fluctuations may reflect the less rugged conformational landscape of proteins as compared with RNA, although there are clearly exceptions. As with RNA, the microscopic origin of this heterogeneity in proteins is generally unclear. As an exception, Polakowski et al. showed that at least some long-lived heterogeneity of enzyme activity can be attributed to covalent differences such as partial degradation of the peptide or post-translational modifications such as glycosylation, which persist in crudely purified samples.(107) However, at least in the case of the GFP mutant GFPmut2, the differences between native states appear conformational in origin, as the states redistribute upon denaturation as is observed also for the TG1I ribozyme. The discovery of such behavior in both proteins and RNAs, in spite of their distinct biophysical properties, suggests that multiple similar native states may be a general feature of biopolymers of complex structure. The fact that wild-type GFP shows less conformational heterogeneity than GFPmut2 invokes the notion that natural evolution may in some cases select against it.

1.5 RNA conformational quasispecies may be natural facilitators of molecular evolution

In the context of evolution, genetic diversity within a population of organisms correlates with the fitness of that population (108); (109)]. Such diversity confers upon the

population resistance to parasites, toxins, and other environmental insults. Numerous examples for such effects have been observed, ranging from genetic resistance to certain human diseases (110); (111, 112), resistance of insects towards pesticides (113), and the appearance of antibiotic-resistant strains of bacteria (114). While a large amount of phenotypic diversity arises from genetic mutations, many other molecular sources of phenotypic variation have been elucidated over the last several decades, including the action of transcription factors and repressors [107], covalent modification of histones and DNA (115), RNA interference (116-118), riboswitches (119-123), and alternative splicing (124, 125). These mechanisms make a great degree of phenotypic diversity possible even in populations of genetically identical cells. In some cases, phenotypic variation can arise stochastically in a population, such as through the translation of low-copy-number mRNAs in cells(126, 127) or the presence of very low concentrations of transcriptional regulators.(128) Thus, stochastic single-molecule events have an important impact on the fate of an entire organism and perhaps the fitness of the entire population of organisms.(129)

In the case of rapidly replicating systems with relatively high mutation rates, such as viruses or bacteria, molecular evolution is described by the quasispecies model, in which variation and selection occur not on the level of well-defined molecular species with nearly identical genetic makeup, but rather on the level of so-called quasispecies comprising clusters of related sequences replicating according to their aggregate fitness level.(130-132) Due to high replication and mutation rates, the fitness of a single genotype, which will not likely be faithfully preserved in the offspring, becomes less

important than the overall fitness of the cluster or quasispecies. In fact, the functional diversity of the quasispecies confers enhanced adaptability to dynamic environments, allowing for example viruses to rapidly evolve resistance to vaccines and antiviral drugs.(133) To maximally exploit this evolutionary advantage, most viruses and organisms are found to maintain an inverse relationship between their mutational error rate and the length of their genome, i.e., they live close to an error threshold imposed by their genome length.(130-132) Consequently, mutation-inducing drugs can cause this error threshold to be crossed, resulting in lethality.(134, 135)

In light of the evidence presented here, the quasispecies model, formulated in the context of genotypic variation, may now need to be extended to conformational variability of RNA with a single sequence. Specifically, given the capability of RNA to form alternative active folds that are stable *in vitro* relative to the lifetime of RNA *in vivo*, I suggest that RNA might function and evolve as conformationally distinct, but functionally related conformational quasispecies. In this view, the source of functional variation is not solely provided by sequence, but by the inherent ruggedness and high degree of energetic degeneracy in the RNA folding landscape. In fit RNA quasispecies, then, alternate folding may constitute another level of adaptive phenotypic variation.

What advantages might such heterogeneity confer on quasispecies of RNA? First, it would enable molecules to achieve their function even if one particular conformer becomes a target for a toxin or nuclease. The observation of a conformational species of the sarcin-ricin loop resistant to cleavage by restrictocin(64) provides a salient example of how such heterogeneity could confer an immediate advantage, provided that the

resistant species is still biologically functional. Such conformational heterogeneity could constitute an attractive mode of adaptation, complementary to sequence variation, when it is necessary to respond to variable environmental challenges on short time scales because larger fractions of an RNA are immediately available with an altered folding behavior than are typically found to carry a specific (set of) mutation(s) leading to such behavior. Conversely, long-term exposure to a toxin or other insult would likely drive preferential selection of sequence variants that thermodynamically or kinetically prefer the formation of the resistant conformer(s).

Second, the capability of an RNA of a single sequence to adopt multiple conformations that directly or indirectly act in concert may enable short RNA oligomers to adopt more sophisticated functions, such as found in the AN58 aptamer. While this aptamer was artificially selected *in vitro*, this type of behavior could be useful in nature due to its sequence economy. Furthermore, Huang et al. suggest that this type of dual-use sequence could provide a precursor to gene duplication and phenotype divergence for functional nucleic acids.(65) Previous work, in which a single RNA sequence was designed to encode the folds and activities of both the HDV ribozyme and an RNA ligase ribozyme,(136) similarly suggests that intersections in sequence space between neutral networks of distinct functional RNAs may be common, and could give rise to new folds and functions during evolution. In fact, the simplistic single RNA-single function paradigm does not do justice to the complexity of nature, where an RNA will always have to exert multiple functions in parallel. An example is the hairpin ribozyme that, like the HDV ribozyme, needs to cleave concatemeric replication intermediates of its satellite

RNA into monomers, then ligate these into circles that function as rolling-circle replication substrates and are devoid of exonuclease-sensitive 5'- and 3'-ends so as to maintain their integrity as substrates.(58) That is, catalytic activity is essential (and defines the "native" state) for one part of the replication cycle, but catalytic inactivity is critical ("native") for another part. The existence of conformational isomers of the hairpin ribozyme with different docking-undocking equilibria may then ensure that some RNA molecules are always optimally performing one function while others optimally perform another function without losing all capacity for the former. I hypothesize that such conformational adaptability endows an RNA quasispecies with enhanced functionality in the face of dynamic evolutionary selection criteria (**Figure 10**).

Of course, essentially all studies demonstrating multiple functional folded states of RNA have been conducted *in vitro*, and it remains to be seen whether such behaviors will be recapitulated *in vivo*. The one example of obligate folding heterogeneity was observed for an artificially selected aptamer, and observations of multiple native states in the hairpin and TG1I ribozymes were made using *in vitro* transcribed or chemically synthesized RNA that had been purified at least once by denaturing polyacrylamide gel electrophoresis. In nature, by contrast, RNA folds as it is transcribed from 5'- to 3'-end, which influences folding in important ways. For example, the segmental co-transcriptional folding of circularly permuted variants of the *Tetrahymena* group I intron was found to yield a higher percentage of natively folded RNA than refolding the entire sequence at once.(137) Transcriptional speed and site-specific pausing were found to be important factors in the folding and function of the FMN riboswitch.(138) The Varkud

satellite ribozyme, shown to exhibit folding heterogeneity by smFRET(58) and EMSA, folds into a much narrower range of conformations when purified without denaturation or refolding after transcription.(139) A bioinformatic study found evidence that sequences of natural transcripts are selected for features that promote co-transcriptional folding into the correct native secondary structure.(140) Interestingly, while the hairpin ribozyme was found to fold sequentially under kinetic control during *in vitro* transcription, the relative thermodynamic stability of competing helices was a larger determinant of folding in yeast cells,(141) though kinetic traps can persist *in vivo* if they are sufficiently stable.(142) The greater preference for thermodynamically stable structures *in vivo* could be due to RNA chaperones and other RNA-binding proteins in the cell(40) that may serve to re-equilibrate kinetically trapped species via ATP-driven helicase activity or nonspecific stabilization of unfolded intermediates. In the case of CYT-19, an ATP-dependent DEAD-box helicase, there even appears to be some preference for unwinding duplexes within misfolded TG1IRz molecules, perhaps based on compactness of tertiary structure alone.(143) Another DEAD-box helicase, Mss116, has been shown to stimulate the folding of a group II intron into its near-native state by promoting the formation of unstable intermediates and dynamic sampling of structures along the folding pathway of the intron [(144); (41); (21)]. While still in their infancy, these studies of co-transcriptional RNA folding and RNA chaperone action suggest that RNA folding behavior should also be studied under conditions as similar as possible to those found in the native cellular environment.

Given the profound kinetic barriers found in some RNAs it seems likely that multiple native states of certain RNAs, either naturally evolved or engineered by humans, will persist *in vivo* even when folded co-transcriptionally in the presence of nucleic acid binding proteins. For natural RNAs, such heterogeneity may depend on the balance between energy requirements to redistribute kinetically trapped species and any (dis)advantages of maintaining a homogeneous over a heterogeneous population of native RNAs. Only *in vivo* testing will determine what roles conformational heterogeneity of RNA may have in living organisms. At least in theory, a shape-shifting RNA quasispecies, as observed *in vitro*, can be expected to impart evolutionary advantages.

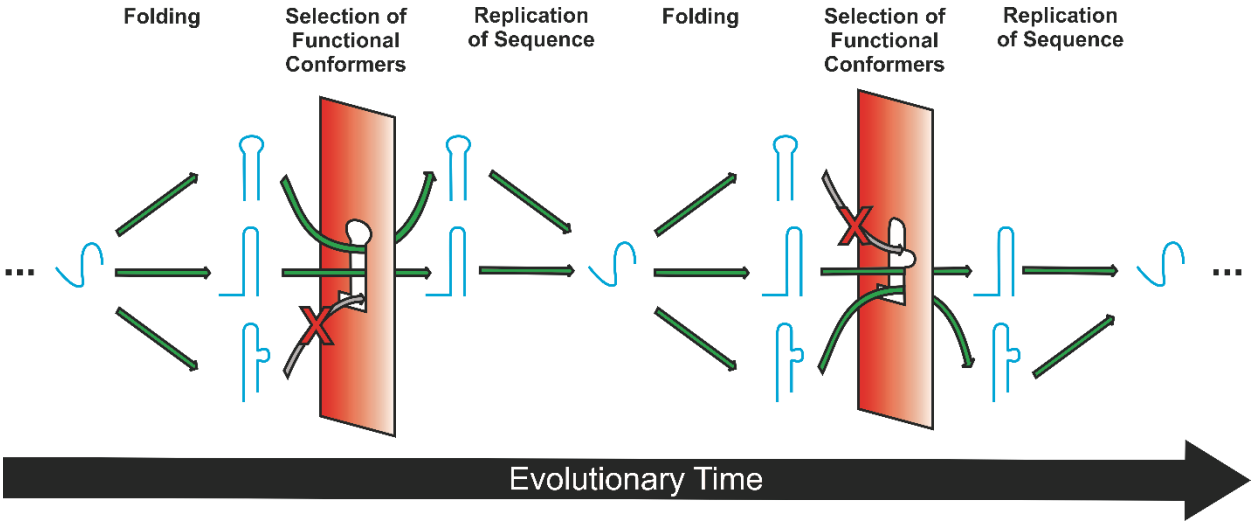


Figure 10: Molecular Quasispecies Satisfy Multiple Requirements.

Schematic representation of a possible adaptive role for conformational quasispecies of RNA under evolutionary pressure. A single RNA sequence (blue) may fold into several stable conformers, or native states, with varying functionality. Changing environmental conditions may impose certain restrictions (red) on the fitness of conformers, but the success of a subset of these conformers will enable the replication of the sequence and the evolutionary survival of all stable (and kinetically accessible) conformers. If conditions are sufficiently variable, there is a clear survival advantage to maintaining a broad quasispecies of RNA folds and functions.

In my work presented here, I will explore the heterogeneities inherent to the hairpin ribozyme. Utilizing a combination of structural probing, electrophoretic mobility shift assays, and a method of non-denaturing purification, I will show causative sources of heterogeneity and how those sources can be avoided. Additionally, I will show the structural effects of various modes of RNA damage common to in vitro purification techniques.

Further, I will show our development of a ribosome labeling, translation, and initiation system to investigate the actions of the ribosome on mRNA. These methods and experiments will serve to explore specific RNA heterogeneities and can be used to greater understand the wider world of RNA biology.

Chapter 2 Heterogeneous Behaviors of the Hairpin Ribozyme

2.1 Introduction

Structurally complex RNA molecules function to maintain, process, and spatiotemporally regulate gene expression, thus ubiquitously governing cellular life. It is estimated that 75% of the ~3 billion base pairs of the human genome are transcribed into at least 80,000 unique RNAs, many of still unknown function, far outweighing the just ~1.2% of our genome that is translated into proteins (145, 146). The secondary and tertiary structures of each of these RNA molecules are derived from the linear sequence of its four nucleotides, but given the limited number of Watson-Crick and non-canonical pairings between them, a multitude of near-isoenergetic alternative structures can fold on the resulting rugged folding free energy landscape (147, 148). Such frustrated folding has two main consequences. First, it endows an RNA with the potential to change its conformation and thus alter its function during its reaction and/or life cycle, a feature that is utilized, for example, by large RNA-based gene expression machines such as the spliceosome (149, 150) and by small parasitic satellite RNAs during their double-rolling circle replication (151). Second, it poses a challenge to obtaining high-resolution structural information by ensemble-averaging techniques such as X-ray crystallography and NMR, which both rely on uniformity of the sample.

The controlled study of RNA often requires either chemical synthesis or *in vitro* transcription, resulting in impure RNA samples containing abortive side products that need to be removed. Most commonly, denaturing polyacrylamide gel electrophoresis (D-PAGE) (152) and anion-exchange chromatography are used for the purpose, resulting in unfolding of the RNA for length separation, followed by refolding. Such refolding of full-length RNA is fundamentally different from the 5'-to-3' segmental folding that RNA undergoes upon emergence from the transcription complex, and often leads to the coexistence of alternate folds (147). While such heterogeneities traditionally have been viewed as detrimental to structural and functional studies *in vitro* (153, 154), the realization that RNAs can exhibit multiple native states has raised the possibility that such states may at least in some cases be important for proper cellular function and evolution (41, 136, 147, 155, 156). The fact that recent studies have uncovered similar structural and functional heterogeneities in DNA (157) and protein (158) further underscores their ubiquity and functional relevance.

A particularly well-studied model system for structural and functional heterogeneities in RNA is the hairpin ribozyme, found in a family of minimal transposable elements, parasitic satellite RNAs accompanying certain plant viruses (32, 151, 159). As a member of the class of reversibly self-cleaving small catalytic RNAs, the hairpin ribozyme is responsible for the accurate excision and subsequent cyclic ligation of monomeric satellite RNAs from the concatemers produced by rolling-circle replication of the parasite. Along the catalytic reaction pathway, domains A and B of the ribozyme approach to dock and form the tertiary structure that defines the active site (32, 159).

Domain docking entails substantial conformational rearrangements in both domains (160, 161) (**Figure 11***A*). Using single molecule fluorescence resonance energy transfer (smFRET), domain docking was observed to be heterogeneous as evidenced by four distinct undocking rate constants that lead to a biphasic cleavage time course (155, 156). Undocking also exhibits a "memory effect", wherein the same molecule maintains its distinct undocking rate constant throughout the experimentally accessible observation window (155, 156). In turn, this effect leads to two distinct bands in non-denaturing electrophoretic mobility shift assays (EMSAs), with higher mobility corresponding to slower exit from the more compact docked conformer (162). Isolation of the two bands and subsequent electrophoretic, smFRET and high-resolution mass spectrometric analyses showed that, first, the heterogeneity partially persists after renaturation and, second, is not due to a change from the expected mass, ruling out covalent adduct formation or degradation as molecular origins (162).

Similarly heterogeneous sub-populations of molecules were observed for the *Tetrahymena* group I ribozyme and found to only slowly interconvert (41). Their existence later was attributed in part to an extended exposure to high temperature, used to refold this large ribozyme, and short-wavelength (254 nm) UV irradiation, employed to localize the RNA during gel purification (163), in deviation from standard protocols that use a longer wavelength (320 nm) (152). In a follow-up study, excessive short-range UV irradiation in thin gels was found to cause significant photo-induced cross-linking of adjacent pyrimidine pairs (164). Of note, such cross-links are expected to be mass neutral and would not be detected by mass spectrometry (162), offering an explanation that

would also be consistent with the data available so far for the hairpin ribozyme. Another potential origin for only slowly interconverting conformational isomers has been proposed for the specificity domain of bacterial RNase P. In this case, selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) using two reagents with differential chemical reactivities identified two slowly repuckering C2'-endo sugars (165), one of which was subsequently shown to be rate-limiting for folding of the entire RNA (75).

Here, we have identified the multifactorial origins of docking heterogeneity of the hairpin ribozyme.We find that exposure to short-wavelength, but not long-wavelength, UV irradiation introduces multiple, distributive cross-links between adjacent uracils, whereas the remaining heterogeneity is caused by a denaturation/renaturation cycle. Consequently, combining minimized short-range UV exposure with a non-denaturing affinity purification protocol to maintain the co-transcriptional fold of the RNA removes all heterogeneity. Our results have important implications for the future study of an ever-expanding universe of functional RNAs.

2.2 Materials and Methods

A one strand construct of a two-way junction construct of the hairpin ribozyme was cloned into the pMCGL plasmid (139) upstream of the glmS ribozyme and cleavage site, resulting in the pHpRz plasmid. This construct was completed by adding a UUCG linker to the 3' end of RzA connecting to the RzB strand, similar to previously described constructs (162). This construct contained an AflII cleavage site between the HpRz and

glmS ribozyme and a G8A mutation within the HpRz. RNA constructs were in vitro transcribed as per previous protocols (162, 166) with minor modifications to accommodate the nondenaturing purification procedure. These modifications involved the use of additional streptavidin coated bead (100 µL to 700 µL). All transcriptions were carried out by run-off transcription from the pHpRz plasmid digested with HindIII (162). Non-denaturing purification was carried out as before (139) with minor modifications. Transcriptions of 150 µL were diluted to 700 µL after phenol/chloroform extraction and mixed with an equivalent 700 µL of washed streptavidin coated beads. Additionally, 3.5 nmols of capture strand were utilized in each 150 µL transcription. This increase in amount of beads and capture strand was done to accommodate the increased transcription yield from the optimized transcription conditions for the HpRz.

In gel purified samples, HindIII restricted plasmid was utilized to express the tandem ribozyme HpRz-glmS in the presence of 0.2 mM glucoseamine-6-phosphate to induce glmS cleavage during the transcription. Samples were PAGE purified by gel electrophoresis over 10% acrylamide (19:1 acrylamide:polyacrylamide) 8M Urea, 1X Tris-borate-EDTA (TBE, 89 mM TRIS base, 89 mM boric acid, 2 mM EDTA) denaturing PAGE (dPAGE) gels. Alternately, gels were run with TRIS-acetate-EDTA (TAE, 40mM Tris, 20mM acetic acid, and 1mM EDTA) in experiments designed to avoid possible borate additions. Body labeling of transcripts was accomplished by supplementing transcriptions with guanosine 5'triphosphate [α-^{32}P] (MP Biomedicals). ^{32}P body labeled gels were imaged by 1-30 minute exposure to x-ray film and then overlayed for band excision. Cold samples were localized initially by short 2-3 seconds

of 312nm UV, but this method was quickly replaced by use of sacrifice lanes of material which were exposed to 254nm UV shadowing and used only for localization. RNA samples were purified by crush-and-soak elution, followed by chloroform extraction and ethanol precipitation. Concentrations of cold ribozyme samples were obtained using a Beckman DU640B Spectrophotometer.

Ultraviolet shadowing of samples was carried out by two different methods utilizing two different light sources. Primarily, samples were irradiated in denaturing PAGE gels by holding the UV source at a constant distance of 7.5 cm for various time increments (0, 30, 60, 120, 300 seconds). These samples were subsequently eluted and purified from the gel material. In a single set of experiments, natively purified HpRz samples were irradiated in cleavage buffer in a 1.5 mL Eppendorf tube at the same 7.5 cm distance. The light sources used were a Spectroline model EF-180C short wave (254nm) ultraviolet lamp and a Spectroline model EB-180C medium wave (312nm) ultraviolet lamp.

Before analysis on native gels, all gel purified samples were diluted into 1X native buffer (50mM TRIS-Acetate, pH 7.5, 12mM Mg-Acetate). Samples were then heated to 70°C for 2 minutes and allowed to slowly cool to room temperature (~22°C) over 20 minutes. Electromobility shift assays were run on 10% PAGE gels in native buffer at 4°C. After the was run, radioactively labeled ^{32}P samples were imaged by overnight (12-16hr) exposure on phosphor screens before reading on a Typhon 9410. Cold samples were immediately stained with 1X SYBr-Gold (Invitrogen) for 30 minutes, destained for 30 minutes in 1X RB, and read by fluorescence on a Typhoon 9410 (488

nm excitation, 520 BP40 emission filter). Sample bands from both imaging methods were quantified using Image J software (NIH).

To synthesize 1M7, a solution of 3.16 mmoles NIA in 20 mL of dimethyl formamide (DMF) was prepared. To this, 4.14 mmoles of sodium hydride (60% in mineral oil) in an additional 20 mL of DMF is added and stirred at RT for 10 minutes. Once it has become an orange colored solution, 3.2 mmoles of methyl iodide were added and allowed to stir an additional 4 hours. The reaction was precipitated by the addition of 50 mL of ice cold 1 N hydrochloric acid (HCl). The precipitate was then filtered and washed with ice cold water. An ether wash step was initially attempted, but the loss of greater than 90% of the resultant product lead to the elimination of this step. The product was then dried on a watchglass in an oven overnight. A small sample was then dissolved in deuterated acetone ($CO(CD_3)_2$) and ^1H NMR spectrum was taken. With or without the ether wash, a very pure 1M7 product was produced. The 1H spectrum was relatively simple, with a single 3H peak for the methyl group at ~3.75 ppm and 3 1H peaks at ~8.1 (doublet), 8.2 (singlet), and 8.4 (doublet) ppm corresponding to the aromatic protons.

SHAPE footprinting was performed as per previously described (167, 168). HEX dye labeled DNA oligos complementary to the 19 nucleotides to the 3' end of the HpRz (5'- AAA TAG GAC TGC GAG GGG G -3') were ordered from Invitrogen, while a NED labeled primer with the same sequence was acquired from Applied Biosystems. Probe were ordered HPLC purified. Isatoic anhydride (IA) was obtained from Acros Organics, NMIA was obtained from Molecular Probes, while 1-methyl-7-nitroisatoic anhydride (1M7) was synthesized in house from 4-nitroisatoic anhydride from PI

Chemicals as per protocols (personal communication with Kevin Weeks' laboratory). Three different preparations of HpRz were utilized in the SHAPE footprinting assays and included natively purified HpRz, gel purified samples utilizing crystal violet staining, and samples directly exposed to 254 nm UV irradiation. A 2 capillary electrophoresis analysis was performed the University of Michigan sequencing core. Results were analyzed with QuShape software from the Weeks lab/UNC Chemistry(169).

2.3: Results and Discussion

Initial Studies of HpRz and Sarcin-Ricin Loop Heterogeneity

Previous work in the Walter lab had focused on the heterogeneity of both the structural and functional heterogeneity of the hairpin ribozyme. However, these studies had not been able to establish a molecular cause behind these observed heterogeneities. Our initial studies aimed to utilize system and expression methods previously used in the lab (67). This involved the transcription of the HpRz as a two strand construct composed of the HzA and HzB strands.

HzA: 5'-AAA UAG A<u>A</u>A AGC GAA CCA GAG AAA CAC ACG CCAAA-3'
HzB: 5'-AU AUA UUU GGC GUG GUA CAU UAC CUG GUA CCC CCU CGC AGU CCU AUU U-3'

This sequence also included the G8A mutation (underlined) which greatly reduces the cleavage activity of this ribozyme. This allows us to perform studies of structure and docking dynamics without the ribozyme proceeding to cleavage. This two strand construct was expressed by run-off transcription from partially complementary DNA

oligos ordered from Invitrogen which included a T7 promoter upstream of the HpRz coding sequence.

Early SHAPE Designs and Expression

Our first attempts to perform SHAPE experiments involved the expression of the HpRz within a SHAPE expression cassette (170). The expression cassette contained two strongly structured hairpins in either the 3' or 5' ends of the RNA of interest, in this case the HpRz. The structure to the 5' end acted as a run-off sequence for the reverse transcription (RT). As the vast majority of RT would reach the full length of the ribozyme, this portion allowed the excessively large signal to not interfere with the banding pattern in the HpRz. The extension to the 3' end acted as binding domain for the RT DNA primer. Both of these regions were strongly structured so that their presence would not likely interfere with the structure of the HpRz.

One of the major difficulties of transcribing these RNAs from single stranded templates was a low transcription rate. This was likely due to the highly structured regions before and after the hairpin ribozyme. While this high level of structure was intended to act as primer binding domain and run-off platform for the subsequent reverse transcription, they seemed to produce very little product. Fortunately, a simple solution was found in utilizing the 1 strand construct of the hairpin ribozyme, as is discussed in Section 2D.

Sources of Heterogeneity in the HpRz

A One-Strand Hairpin Ribozyme Shows Robust Docking Heterogeneity. To probe the origin of docking heterogeneity in the hairpin ribozyme, we sought a suitable variant that shows two bands in EMSAs, as a signature of folding heterogeneity, and can be easily purified and footprinted. To this end, we chose a one-strand hairpin ribozyme that is easily generated by run-off transcription (**Figure 11***A*) and rendered non-catalytic by a G8A mutation that removes one of the catalytically important residues without impacting docking (162). This ribozyme was embedded into a transcription vector with a downstream *glmS* ribozyme that can be induced by glucosamine-6-phosphate (GlcN6P) addition to site-specifically self-cleave, generating a chemically homogeneous 3' end on the hairpin ribozyme that removes transcriptional heterogeneity (139, 149). After transcription in the presence of α-^{32}P-GTP and GlcN6P and purification by standard D-PAGE, the radiolabeled ribozyme formed the two expected EMSA bands (**Figure 11***B*). A fraction of ~15% was found in the more slowly migrating (less stably docked) top band (Fig 1*B*), only slightly less than the fraction observed for previously characterized two- and three-strand ribozymes (147, 161, 162). Next, we purified larger quantities of both the top and bottom EMSA bands as previously described (147, 161, 162) for further characterization. We found that heat denaturation at 70 ^{o}C and refolding of the purified bands did not significantly interconvert them (**Figure 11***B*). Denaturation at 90 ^{o}C or 100 ^{o}C, at 90 ^{o}C in the presence of either high salt concentrations (500 mM NaCl) or 20% (w/v) polyethyleneglycol (8 kDa molecular weight) as molecular crowder, or at 100 ^{o}C in the presence of a large access of unlabeled ("cold") hairpin ribozyme did not significantly change this result (**Figure 15**). These findings suggest that the previous partial

interconversion observed for the top and bottom EMSA bands of a two-strand hairpin ribozyme is largely explained by redistribution of the 5'- and 3'-segments between molecules upon reannealing (162). Such resdistribution cannot occur with the one-strand variant unless a dimer forms, which we can exclude based on the lack of changes in banding pattern upon addition of excess ribozyme.

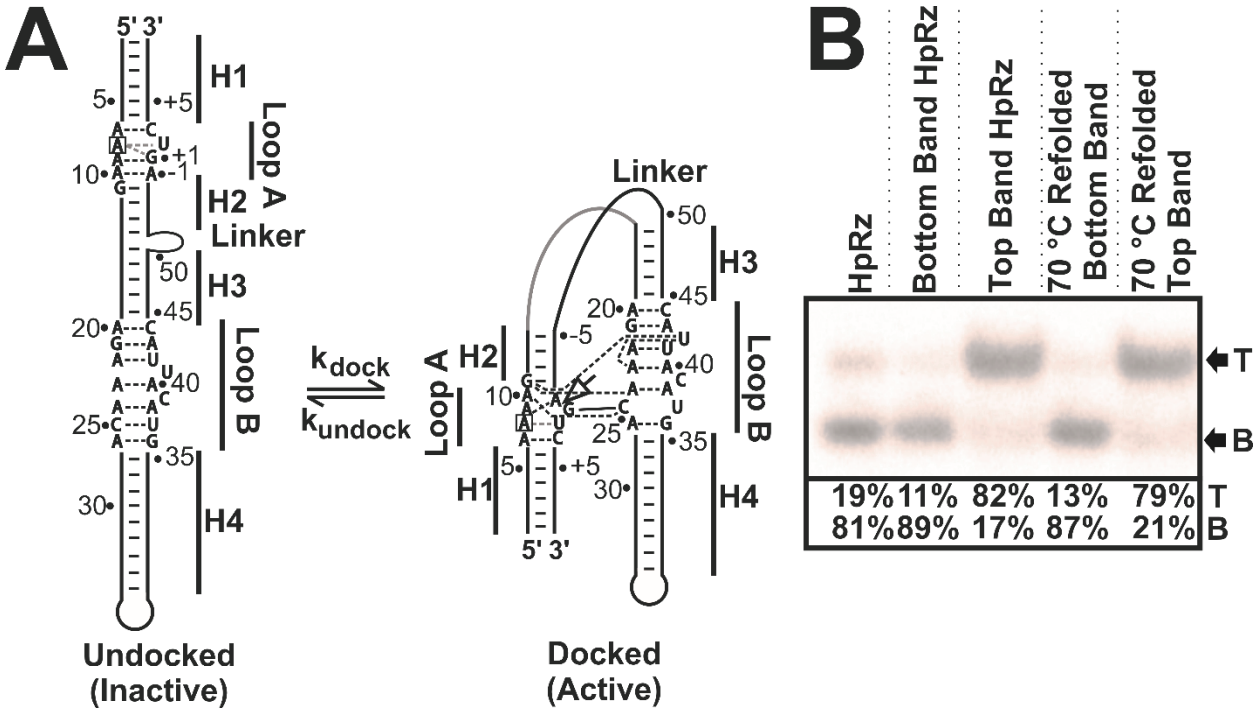


Figure 11: Hairpin Ribozyme Construct and Heterogeneity:

All of the work in this chapter is based on this one strand hairpin ribozyme and displays multiple levels of heterogeneity. (A) One strand construct of the hairpin ribozyme. The hairpin is displayed in both the extended undocked state (right) and compact docked state (right) with tertiary interactions shown as dashed lines. (B) Heterogeneity of hairpin ribozyme as assayed by 10% native PAGE gel. Gel purified HpRz show a minor shift of material in top, with the majority occupying the bottom band. Top and bottom bands maintain identity when isolated and rerun on a native gel with or without heat refolding at 70° C. (C) Effect of UV exposure on HpRz. Samples of gel purified hairpin ribozyme (+/- UV exposure) were resolved on a subsequent 20% denaturing PAGE gel. These samples also varied in refolding, with the refolded samples receiving 70 °C exposure for 90 seconds. UV exposed samples can be noted for the increased breakdown products in those lanes.

Next, we asked whether the partitioning between the top and bottom bands is influenced by a structural self-cleavage preference or the exposure time prior to GlcN6P-induced self-cleavage of the hairpin ribozyme-*glmS* ribozyme fusion RNA. To this end, we first D-PAGE purified the fused RNA, then incubated it with GlcN6P for 2 or 120 min prior to EMSA analysis to induce self-cleavage of the *glmS* ribozyme. Alternatively, GlcN6P was added at different time points during the transcription reaction to allow for 2 or 120 minutes of total self-cleavage, then the products immediately purified via D-PAGE. In each case, the resulting relative populations of top and bottom bands of the hairpin ribozyme did not change significantly (**Figure 12**). We therefore conclude that the one-strand hairpin ribozyme as generated here is a particularly robust model for

studying docking heterogeneity.

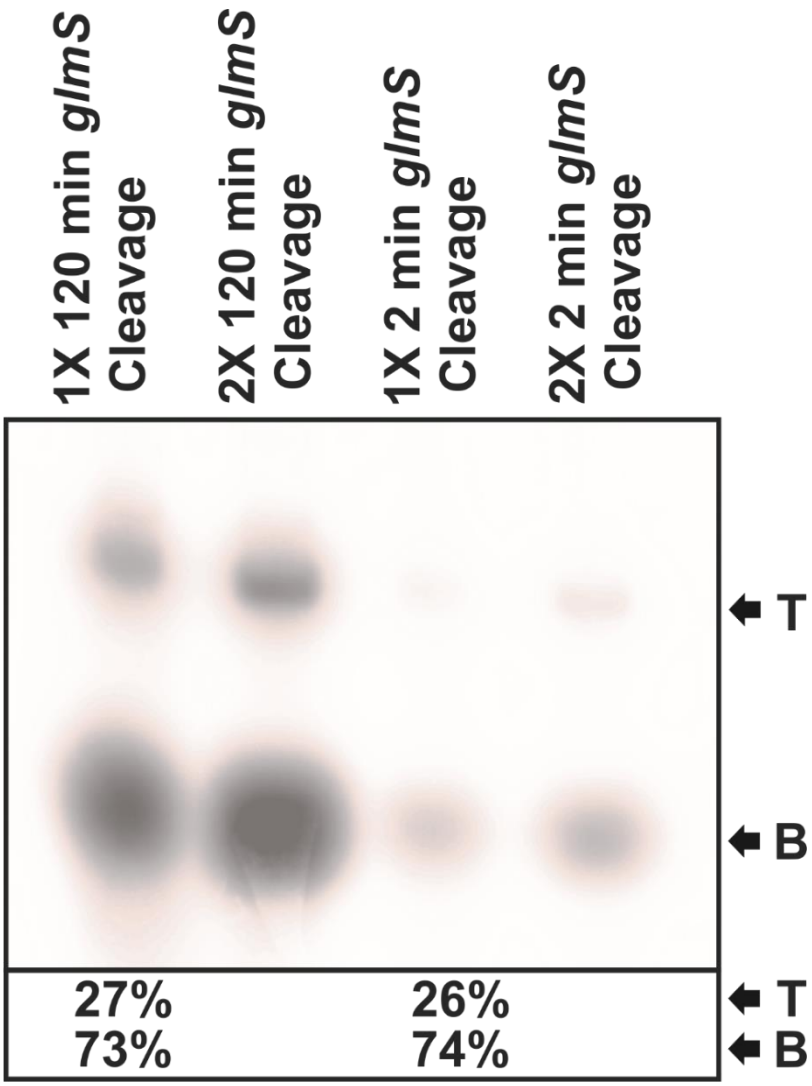


Figure 12 Cleavage Preference by glm*S* Reveals a Similar Top and Bottom Band Distribution of HpRz in Native PAGE Gels.

Samples of hairpin ribozyme-glmS ribozyme expressed in tandem were allowed to cleave for either 2 or 120 minutes to ascertain if a cleavage preference for either structure isoform by glmS existed. This variation in cleavage time could have been one explanation as to the elimination of the top band in the non-denaturing purification scheme. Whether cotranscriptional cleavage mediated by the addition of GlcN6P was carried out for 2 or 120 minutes, the resulting distribution of material between top and bottom bands was virtually identical. Samples were loaded at a 1X or 2X amount to allow for clear visualization of bands.

In gel purified samples, HindIII restricted plasmid was utilized to express the tandem ribozyme HpRz-glmS in the presence of 0.2 mM glucoseamine-6-phosphate to induce glmS cleavage during the transcription. Samples were PAGE purified by gel electrophoresis over 10% acrylamide (19:1 acrylamide:polyacrylamide) 8M Urea, 1X Tris-borate-EDTA (TBE, 89 mM TRIS base, 89 mM boric acid, 2 mM EDTA) denaturing PAGE (dPAGE) gels. Alternately, gels were run with TRIS-acetate-EDTA (TAE, 40mM Tris, 20mM acetic acid, and 1mM EDTA) in experiments designed to avoid possible borate additions. Body labeling of transcripts was accomplished by supplementing transcriptions with guanosine 5'triphosphate [α-^{32}P] (MP Biomedicals). ^{32}P body labeled gels were imaged by 1-30 minute exposure to x-ray film and then overlayed for band excision. Cold samples were localized initially by short 2-3 seconds of 312nm UV, but this method was quickly replaced by use of sacrifice lanes of material which were exposed to 254nm UV shadowing and used only for localization. RNA samples were purified by crush-and-soak elution, followed by chloroform extraction and ethanol precipitation. Concentrations of cold ribozyme samples were obtained using a Beckman DU640B Spectrophotometer.

Before analysis on native gels, all gel purified samples were diluted into 1X native buffer (50mM TRIS-Acetate, pH 7.5, 12mM Mg-Acetate). Samples were then heated to 70°C for 2 minutes and allowed to slowly cool to room temperature (~22°C) over 20 minutes. Electromobility shift assays were run on 10% PAGE gels in native buffer at 4°C. After the was run, radioactively labeled ^{32}P samples were imaged by overnight (12-16hr) exposure on phosphor screens before reading on a Typhon 9410.

Cold samples were immediately stained with 1X SYBr-Gold (Invitrogen) for 30 minutes, destained for 30 minutes in 1X RB, and read by fluorescence on a Typhoon 9410 (488 nm excitation, 520 BP40 emission filter). Sample bands from both imaging methods were quantified using Image J software (NIH).

Short-Range UV-Irradiation, But Not γ-Irradiation or Exposure to Electrophoresis Buffers, Increases Docking Heterogeneity. While high-resolution mass spectrometry did not detect any differences in the masses of top and bottom band hairpin ribozymes (162), it is formally possible that such adducts are formed only transiently and manifest in solution but not in the gas phase of a mass spectrometer. We therefore asked whether borate adducts form between the RNA and our standard electrophoresis buffer (150, 171) to result in the observed folding heterogeneity, as suggested for the P4-P6 domain from the *Tetrahymena* group I intron (163). To this end, we changed our D-PAGE buffer from Tris-borate to Tris-acetate, but via subsequent EMSA observed no change in the fraction of top band formed (**Figure 13**). This result excludes borate adducts as a source for folding heterogeneity.

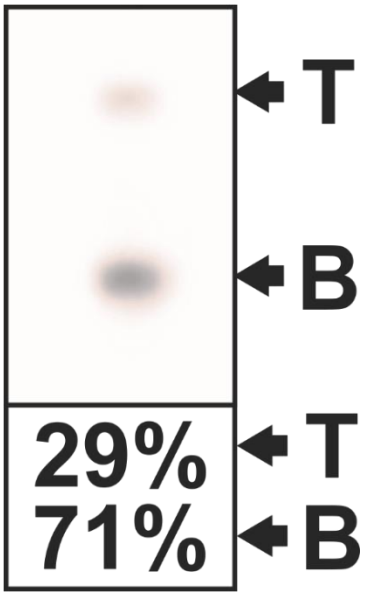


Figure 13: Varying Gel Buffering Conditions Has Little Effect on HpRz Distribution in Native PAGE Gel.
Samples of HpRz were gel purified in a denaturing TAE gel to see if the borate present in TBE gels had an effect on heterogeneity. This sample displays similar top (29%) and bottom (71%) band distribution to samples purified in TBE gels (**Figure 12**). As such, it is unlikely that borate addition from the TBE buffer is a contributing factor to HpRz heterogeneity.

Similarly, exposure to 8M urea at elevated temperatures did not change the distribution between top and bottom bands (**Figure 14**). This condition simulating the denaturing conditions experienced by the HpRz during gel purification and thus simple denaturation can be ruled out as a cause of hairpin ribozyme heterogeneity.

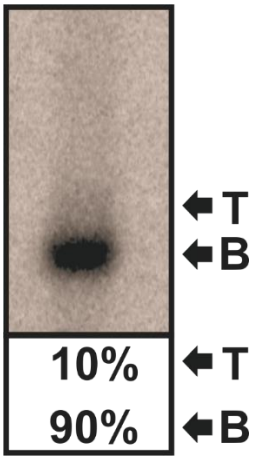


Figure 14: Exposure to 8M Urea at Elevated Temperature Doesn't Effect Heterogeneity.

As non-denaturing purification of the hairpin ribozyme avoided full chemical denaturation it was possible that exposure to urea could induce additional hetergeneity. Samples of gel purified HpRz were exposed to 8M urea at 65 °C for 30 minutes to replicate the condition the ribozyme is exposed to during gel purification. Samples were then resolved on 10% native PAGE gel, but revealed little additional heterogeneity in the form of top band material.

Next, we asked whether UV-exposure plays a role in the observed heterogeneity. Previous studies using short-wavelength (254 nm) UV irradiation revealed a long-distance photo-crosslink between G21 and U42, resulting in a slower-migrating band during D-PAGE (161, 172). While these nucleotides are juxtaposed in the structured loop B of the folded ribozyme, they are separated by ~20 nucleotides along the RNA backbone and not likely to be in close proximity in the denatured state. Accordingly, we found that short-range UV exposure led to some smearing of the RNA, but no discernible high-molecular weight crosslink product upon D-PAGE (**Figure 15**). EMSA analysis, however, showed that specifically pyrimidine dimerization caused by UV irradiation (163, 164) has become a prominent concern in the purification of RNA. While the

general effects of UV damage have long been known, the exact physical manifestations within subsequent experiments were poorly understood.

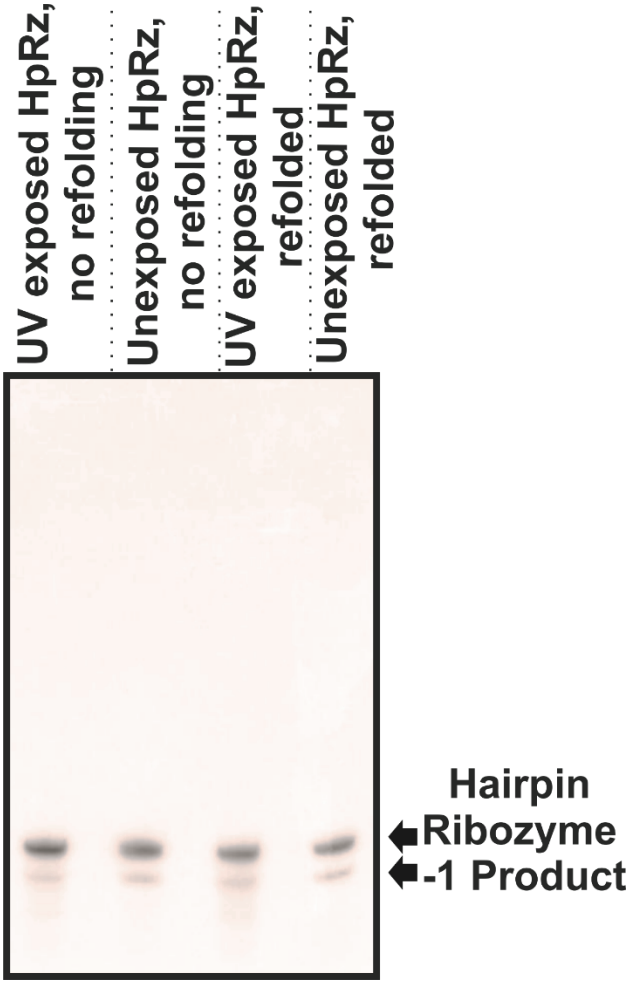


Figure 15: Effect of UV Exposure and Heat Refolding on HpRz Structure. Samples of gel purified hairpin ribozyme (+/- 300s UV exposure) were resolved on a subsequent 8M urea, 20% denaturing PAGE gel. Samples also varied in refolding, with the refolded samples receiving 70 °C exposure for 90 seconds and slow cooling to room temperature over 20 minutes. UV exposed samples can be noted for the increased breakdown products in those lanes.

In order to determine whether UV irradiation contributes to the heterogeneity seen in the hairpin ribozyme, we exposed gel-purified HpRz to a time course of 254 or 312 nm UV radiation of similar intensity within a denaturing gel matrix as described (163, 164).

After purification, samples were refolded in a native buffer previously used in numerous hairpin folding and catalysis studies (155, 162, 172), separated by native PAGE, and stained with SYBR Gold to assess heterogeneity (**Figure 16**). The fractional population of the top band increased with increased UV exposure, suggesting that photo-cross-linking contributes to the observed heterogeneity. Fitting these data with an exponential decay model revealed much faster damage upon exposure to 254 nm light, reaching a maximum of nearly 54% of sample in the top band (from an initial value of 17%) (**Figure 16A and C**). When irradiating at 312 nm, in contrast, the top band reached a maximal population of only 25% over the same 300-s period (**Figure 18B and C**). This is likely due to the lower energy and penetrance of the 312 nm irradiation as opposed to the 254 nm lamp. As such, it reaffirms our preference for using 312 nm UV lamps whenever UV shadowing is utilized for sample purification.

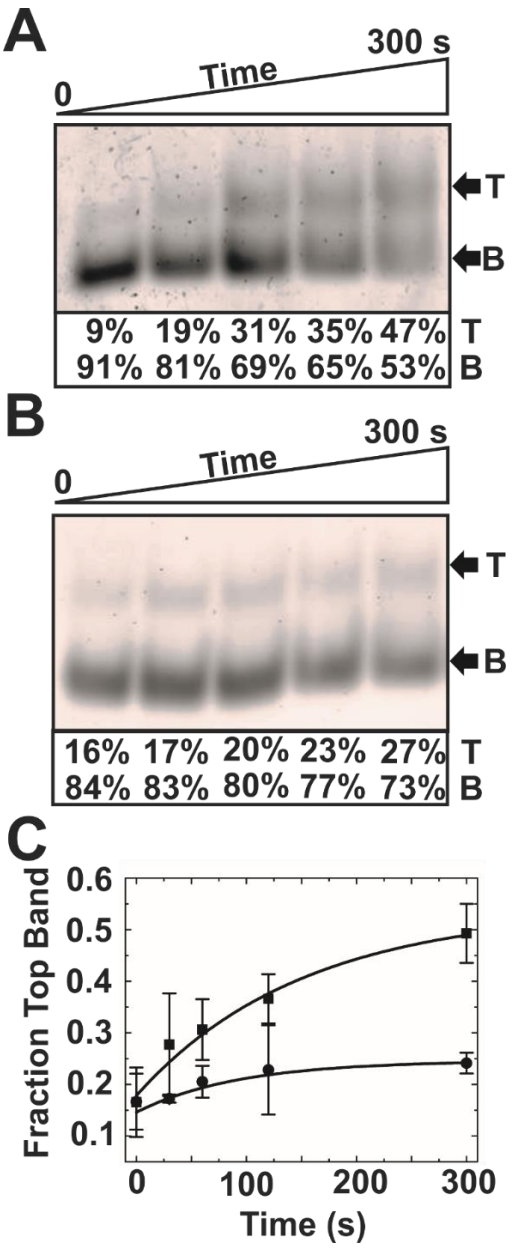


Figure 16 Effect of Ultraviolet Light on Denatured Hairpin Ribozyme
Ultraviolet irradiation on HpRz samples in a denatured state increased structural heterogeneity. (A) EMSA gels used to assess structural heterogeneity of the hairpin ribozyme. Samples were exposed to (L→R) 0, 30, 60, 120, or 300 seconds of 254nm UV light. Samples resolved into a fast migrating bottom band and a slow migrating top band and visualized after SYBr Gold staining. (B) EMSA gel as in (A) but with samples exposed to 312nm ultraviolet irradiation under the same time course. (C) Gel bands quantified, plotted, and fit to an exponential association curve. Plots for both 254 nm and 312 nm UV show a marked difference in the ability of the light sources to shift material from the bottom band to the top band.

Further, we wanted to be sure that heat refolding was not correlated with the appearance of the slower migrating band. To test if our heat refolding had a significant effect, we refolded gel purified samples of HpRz in buffer at 70 °C for 2 minutes, 90 °C for 1 minute, and 100 °C for 10 or 30 seconds. At 100 °C, the buffer replaced the standard 12 mM Mg^{2+} with 50 mM NaCl to remove divalent cations and preserve the integrity of the ribozyme at this extreme temperature. Also, we tested extremely high salt concentrations by raising NaCl to 500 mM, crowding reagents by refolding in the presence of 20% PEG 8K, and adding 100 uM cold HpRz in an attempt to create dimers and possibly alter the fraction of top band material (**Figure 17**, **Figure 18**). In all cases, the top band remained in about the same ratio with the bottom band material. This indicated that none of these conditions were significant contributors to the structural heterogeneity seen in the HpRz.

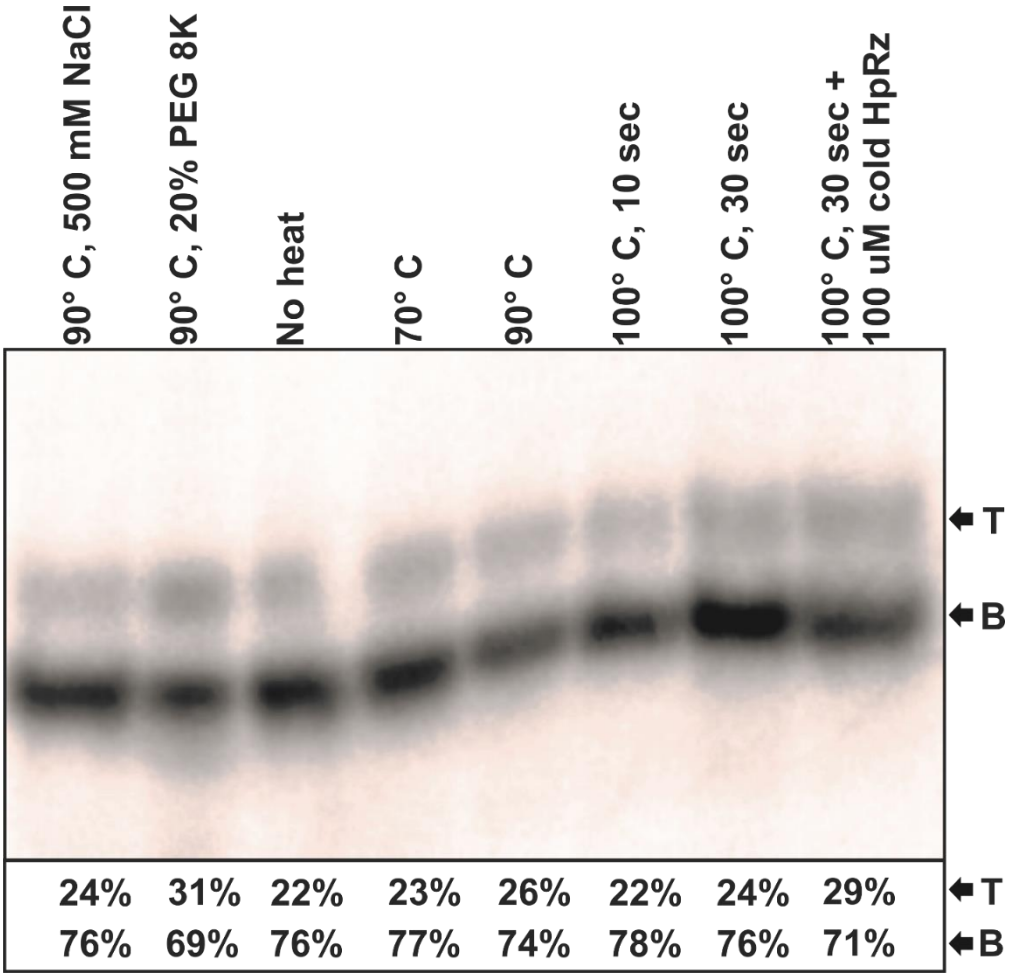


Figure 17: Various Methods of Denaturation Have Little Effect on Band Distribution in Native Gels.

To eliminate the possibility of dimers and show that various methods of denaturation had similar effects on HpRz heterogeneity, samples were exposed to an array of denaturing conditions. Samples of gel purified hairpin ribozyme were refolded at various temperatures (70, 90, 100 °C) and with various salts to examine if variability in folding affected the distribution of material between the top and bottom band. Additionally, excess cold HpRz was titrated into the folding procedure to test the concentration band distribution. In all cases, the distribution stayed about the same between top and bottom band. Samples were resolved on 10% PAGE gels.

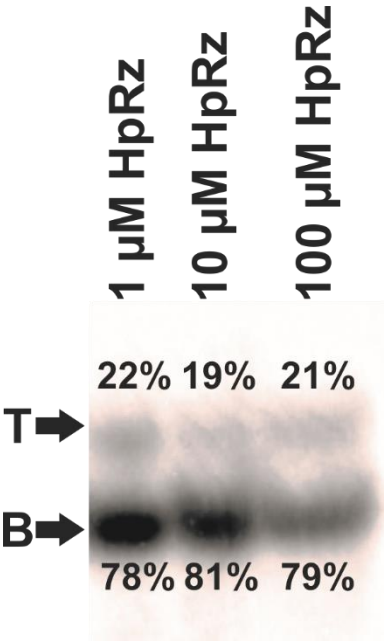


Figure 18: Excess Unlabeled HpRz Does Not Contribute to Top Band.

As exact concentrations of HpRz in radioactive samples was difficult to ascertain, the possibility that excessive RNA could lead to crowding and formation of the top band was addressed by titrating in large amounts of cold RNA. The heterogeneity of the HpRz concentration dependence of band distribution was addressed by the addition of "cold" unlabeled HpRz to ^{32}P body labeled HpRz. Samples were mixed and refolded at 100 °C for 30 seconds to ensure complete denaturation before being allowed to slowly cool and refold. After cooling to room temperature, samples were resolved on 10% native PAGE gel. Gels were then exposed to a phosphor screen for 1-4 hours and quantified on a Typhoon 9410. Band distribution changed minimally between samples, indicating that concentration dependence was not likely a major factor.

Ultraviolet shadowing of samples was carried out by two different methods utilizing two different light sources. Primarily, samples were irradiated in denaturing PAGE gels by holding the UV source at a constant distance of 7.5 cm for various time increments (0, 30, 60, 120, 300 seconds). These samples were subsequently eluted and purified from the gel material. In a single set of experiments, natively purified HpRz samples were irradiated in cleavage buffer in a 1.5 mL Eppendorf tube at the same 7.5 cm

distance. The light sources used were a Spectroline model EF-180C short wave (254 nm) ultraviolet lamp and a Spectroline model EB-180C medium wave (312 nm) ultraviolet lamp.

While previous studies resolved two populations of HpRz even without radiolabeling of the RNA (67), we investigated whether the radioactivity of 32P labeled RNA contributes to the conformational heterogeneity. Samples of HpRz body-labeled with 32P were resolved by nPAGE just after D-PAGE purification and then again after one week. These samples exhibited similar distributions of top and bottom material, with the additional week of exposure to ^{32}P having no significant effect on heterogeneity (**Figure 19**).

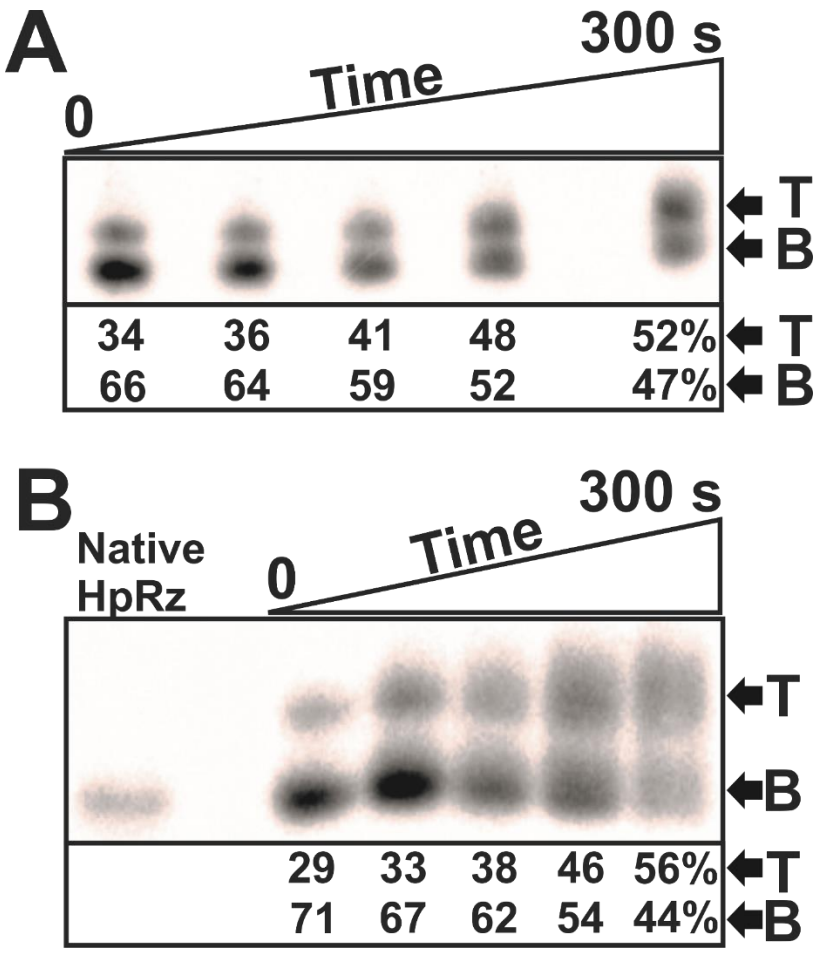


Figure 19 Sample Exposure Time to ^{32}P has Little Effect on Band Distribution in UV Exposed Samples.
As all early experiments were 32P body labeled RNA, the exposure of samples to radiation had to be ruled out as a possible contributor to heterogeneity. Above are native PAGE gels of HpRz samples exposed to 0-300 seconds of 254 nm UV. (A) Samples of HpRz were resolved immediately after initial purification over a denaturing PAGE gel. Samples displayed a typical distribution of material between top and bottom bands (Fig. 2). (B) Same samples run after 1 week of storage at -20 °C, also resolved on a native PAGE gel. The samples display a similar distribution as seen in panel A, ruling radiation from body labeling as a likely source of heterogeneity.

Synthesis of 1M7

Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) is a method used interrogate the flexibility in RNA structures. In SHAPE chemistry, reagents

preferentially react with 2'-hydroxyl groups of nucleotides in single stranded sequences of RNA. This creates adducts which add large groups to the RNA. When reacted RNAs are reverse transcribed (RT), these adducts cause premature stoppage of the reverse trasnscriptase (RTase). Utilizing fluorescently labeled DNA oligos complementary to the 3' end of the RNA, the RT provides a pattern DNA fragment corresponding to the stoppage pattern on the RNA. In order to perform SHAPE footprinting of an RNA, SHAPE reagents are necessary to react with the RNA of interest and form these adducts.

Two of the commonly used reagents were commercially available, isatoic anhydride (IA) and N-methylisatoic anhydride (NMIA). These were purchased from Sigma-Aldrich and Life Technologies, respectively. However, to perform differential SHAPE footprinting necessary to examine variation nucleotide sugar pucker, we needed a reagent with a faster half-life in water. 1-methyl-7-nitroisatoic anhydride (1M7) was used in previous publications but was not available commercially. However, it could be synthesized from an immediate precursor, 4-nitroisatoic anhydride (NIA), which we obtained from Pi Chemicals. The product was very clean according to the NMR (**Figure 20**), and recovery of product was also very good, with a recovery rate of around 90% typically accomplished.

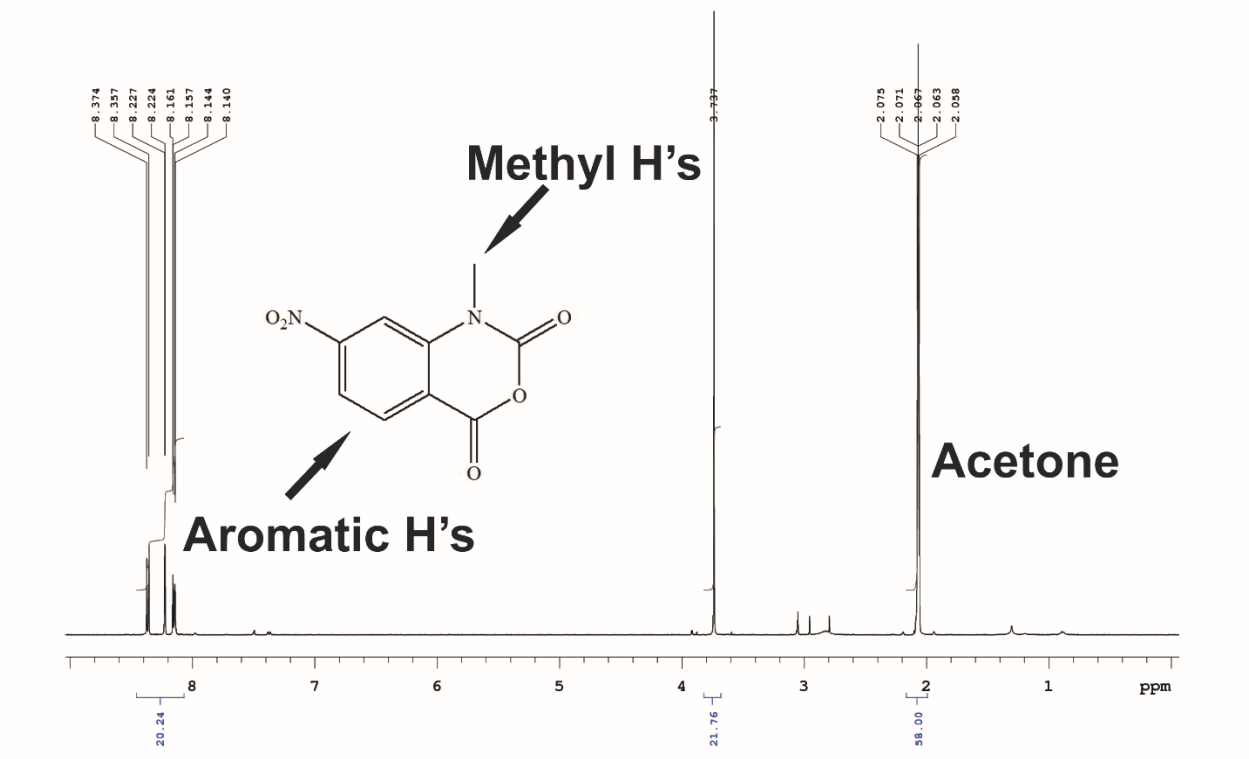


Figure 20: NMR of Synthesized 1M7.

In order to perform SHAPE footprinting on the HpRz, the SHAPE reagent 1M7 had to be chemically synthesized. After synthesis, we performed a 1H NMR corresponding to 1M7 run in deuterated acetone. The 1H spectrum was relatively simple, with a single 3H peak for the methyl group at ~3.75 ppm and 3 1H peaks at ~8.1 (doublet), 8.2 (singlet), and 8.4 (doublet) ppm corresponding to the aromatic protons. This showed a pure 1M7 product without any breakdown product.

SHAPE Footprinting of HpRz

We investigated the molecular origins of this UV-dependent perturbation of HpRz structure in greater detail using SHAPE footprinting. To isolate the impact of UV irradiation from that of dPAGE purification upon HpRz structure, we probed three different preparations of hairpin ribozyme: natively purified HpRz (N), dPAGE gel purified HpRz visualized with crystal violet rather than UV shadowing (CV), and gel

purified HpRz exposed to 300 s of 254 nm UV excitation from a distance of 7.5 cm (UV). In general, the SHAPE reactivity pattern of HpRz is consistent with the known secondary structure (Figure 1). Interestingly, nucleotide 8, (normally a G but mutated to A to dramatically reduce cleavage rate) displayed a much higher reactivity than the surrounding unpaired nucleotides. ~~E~~ssential for normal cleavage activity, this suggests unique backbone structure at this nucleotide position. Comparison between the different methods of purification revealed few differences in the secondary structures, consistent with previous Tb^{3+} footprinting results (ref). In active sites and interacting regions of RNA molecules, the ribose sugars have previously been shown to frequently display 2'-endo conformations, as opposed to 2'-exo conformations normally observed in RNA structures (75). To probe the dynamics of ribose conformations within the backbone of HpRz, we compared reactivity patterns using 2 different SHAPE reagents (isatoic anhydride 1-methyl-7-nitroisatoic anhydride). In the case of the hairpin ribozyme, no nucleotide showed significantly higher reagent-dependent reactivity (>1.0), suggesting that there weren't any likely 2'-endo conformers in the backbone. Additionally, comparing samples exposed to high energy UV irradiation to those that were not, we found that there was increased reverse transcription stop in the B-loop of the B domain, as well as U4, A5, and A8 of the A domain. However, comparing UV exposed samples against those purified without UV exposure revealed numerous premature stops. This was especially high in the B-loop (fig 3D) with both the U41-42 and C44-45 as candidates for local pyrimidine dimerization.

SHAPE footprinting was performed as per previously described (167, 168). HEX dye labeled DNA oligos complementary to the 19 nucleotides to the 3' end of the HpRz (5'-AAA TAG GAC TGC GAG GGG G -3') were ordered from Invitrogen, while a NED labeled primer with the same sequence was acquired from Applied Biosystems. Probe were ordered HPLC purified. Isatoic anhydride (IA) was obtained from Acros Organics, NMIA was obtained from Molecular Probes, while 1-methyl-7-nitroisatoic anhydride (1M7) was synthesized in house from 4-nitroisatoic anhydride from PI Chemicals as per protocols (personal communication with Kevin Weeks' laboratory). Three different preparations of HpRz were utilized in the SHAPE footprinting assays and included natively purified HpRz, gel purified samples utilizing crystal violet staining, and samples directly exposed to 254 nm UV irradiation. A 2 capillary electrophoresis analysis was performed the University of Michigan sequencing core. Results were analyzed with QuShape software from the Weeks lab/UNC Chemistry(169).

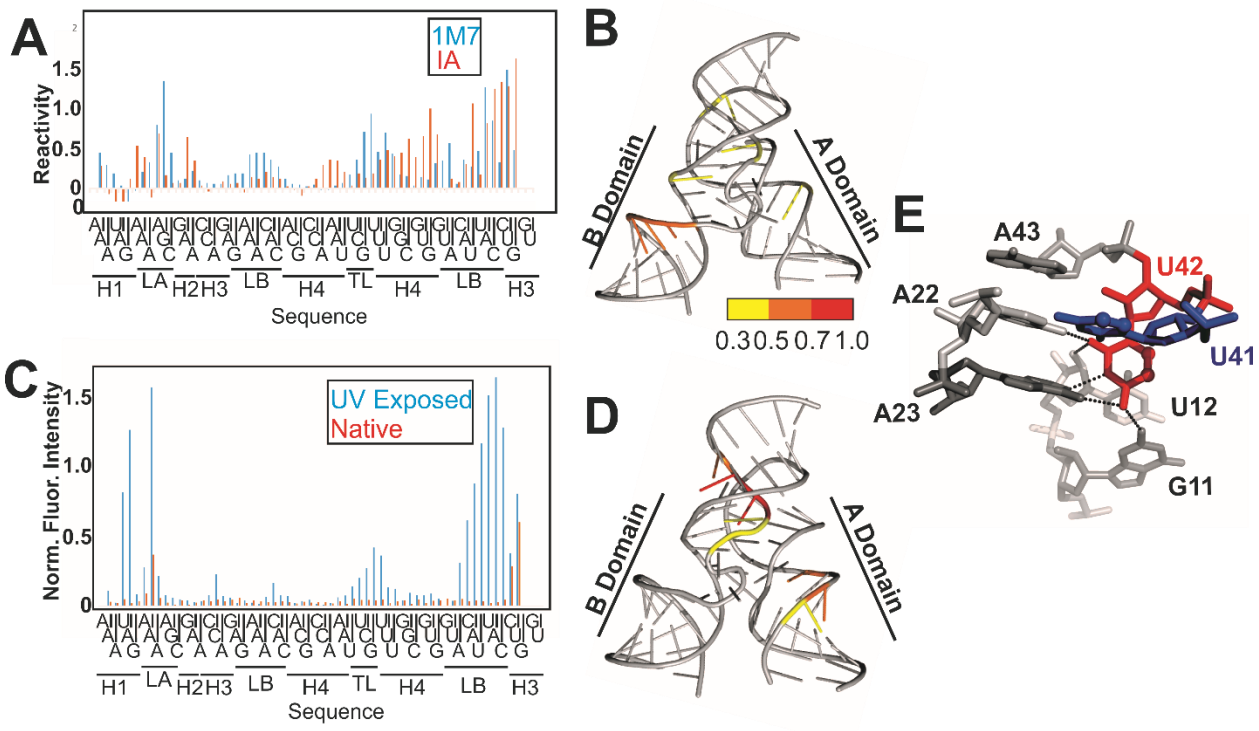


Figure 21: Structural Probing of the Hairpin Ribozyme

(A) Capillary electrophoresis elution profile of the hairpin ribozyme. Utilizing SHAPE chemistry and a two capillary approach, we are able to resolve RT readouts of hairpin ribozyme reacted with 1m7 (blue) and IA (red) reagents. Helices 1-4 are abbreviated H1-H4, loops A and B as LA and LB, and the UUCG tetraloop as TL. (B) Early reverse transcription stoppages predominate in the UV exposed samples. Samples of HpRz either unexposed (red) or exposed to 254nm UV irradiation (blue) were reverse transcribed to examine potential stoppages in due to dimerization events. The region around the B-loop results in heavy stoppages, as well as elevated stoppages around the UUCG tetraloop, both of which contain adjacent pyrimidines which are capable of dimerizing. (C) Average backbone reactivity by nucleotide. Reactivity is shown as a heat map ranging from low reactivity (yellow) to high reactivity (red). Negligibly reactive nucleotides are shown in black. The nucleotides boxed in grey indicate the primer binding for which we have no structural data. (D) Reverse transcription stops as induced by UV exposure. By subtracting the negative controls of HpRz produced by nondenaturing purification from the samples exposed to 300 seconds of UV during the gel purification. (E) Spatial representation of U42 binding pocket precluded by cross-linking. U42 (red) is shown looped out of the helical stack and forming 5 H-bonds (dashes) with G11, U12, A22, and A23 (grey). U41 (blue) is shown in the helical stack and the carbons of U41 and U42 involved in a potential pyrimidine dimer are shown as spheres. It should be noted how spatially separated these atom are from each other, as their dimerization would preclude proper H-bonds necessary for docking from occurring.

Non-denaturing Purification of HpRz

A one-strand construct of a two-way junction construct of the hairpin ribozyme was cloned into the pMCGL plasmid (139) upstream of the glmS ribozyme and cleavage site, resulting in the pHpRz plasmid. This construct was completed by adding a UUCG linker to the 3' end of RzA connecting to the RzB strand, similar to previously described constructs (162). This construct contained an AflII cleavage site between the HpRz and *glmS* ribozyme and a G8A mutation within the HpRz. RNA constructs were in vitro transcribed as per previous protocols (162, 166) with minor modifications to accommodate the nondenaturing purification procedure. All transcriptions were carried out by run-off transcription from the pHpRz plasmid digested with HindIII (162). Non-denaturing purification was carried out as before (139) with minor modifications. Transcriptions were diluted to 700 µL after phenol/chloroform extraction and mixed with an equivalent 700 µL of washed streptavidin coated beads. Additionally, 3.5 nmols of capture strand were utilized in each 150 µL transcription. This increase in amount of beads and capture strand was done to accommodate the increased transcription yield from the optimized transcription conditions for the HpRz.

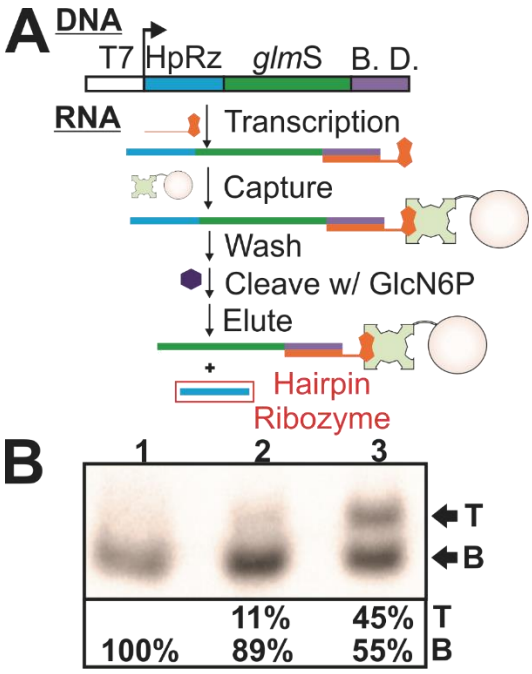


Figure 22: Non-Denaturing Purification of the HpRz.

Non-denaturing purification of RNA provides a viable alternative to gel purification which eliminates structural heterogeneity seen in the HpRz. (A) The expressions platform consists of a T7 promoter sequence (white) followed by the sequence encoding the hairpin ribozyme (blue), the glmS ribozyme (green), and the binding platform (purple). The binding platform is complementary to a biotinylated DNA capture strand (orange) is used for sample immobilization. (B) Native PAGE gel of natively purified hairpin ribozyme. As compared to gel purified material, there is virtually no top band material present in the natively purified samples (1). When repurified over a denaturing PAGE gel (2), a small fraction of top band is apparent compared to gel purified samples (3).

Having previously seen dramatic effects of dPAGE purification on the structural homogeneity other RNAs, we decided to utilize a nondenaturing purification technique in an attempt to further reduce the amount of top band material (fig 4c) (139). Strikingly, nondenaturing purification of the hairpin ribozyme revealed a single population of material when analyzed by nPAGE. In order to recapitulate the structural heterogeneity

seen in previous studies, samples were thermally re-annealed to look for a 2nd, "native like" low energy structure. Upon refolding at 70°C for 2 minutes followed by slow cooling, HpRz samples continued to migrate as a single band in nPAGE. However, HpRz transcribed in the presence of glucosamine-6-phosphate, which allowed contranscriptional cleavage of the HpRz by glmS, followed by dPAGE gel purification, shifted ~50% of material into a slower migrating top band, with or without 70°C refolding. This allowed assignment of the "bottom band" status to the natively purified material, and implicates dPAGE as one factor that introduces heterogeneity into HpRz folding. Hairipin ribozyme originally purified by non-denaturing techniques re-purified over dPAGE gel only showed a small shift back to the top band material (~8%).

Alternate Method of Avoiding Heterogeneity

Observing that any amount of UV exposure compromises the structural homogeneity of HpRz preparations, we localized dPAGE product bands by staining with crystal violet. This method of staining only requires the use of low energy ambient light to identify bands, thus avoiding exposure to intense UV irradiation altogether. Using this protocol, adapted from common cloning procedures (173), products from transcription reactions as small as 2.5 µL could be detected under ambient light. Further, using purified HpRz, we found that we could detect as little as 1 ug of RNA in a gel. Additionally, these samples resulted in only 8-10% of HpRz migrating as top band material when resolved on a subsequent native PAGE gel, suggesting that the gel purification itself was responsible for at least a fraction of the observed heterogeneity.

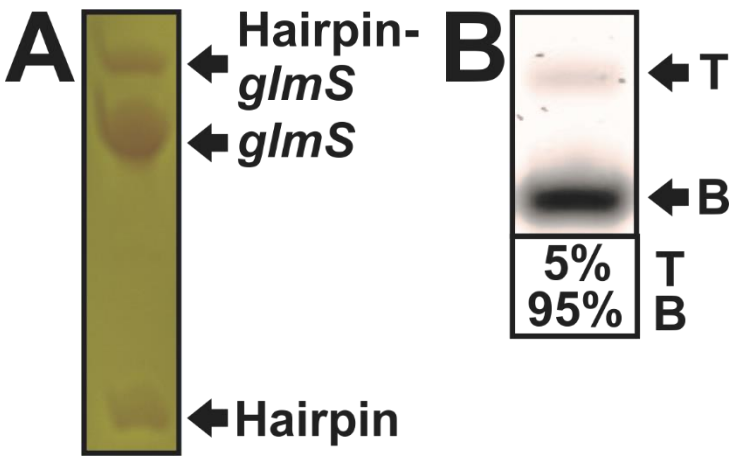


Figure 23: Crystal Violet Staining as an Alternative to UV Shadowing.

Crystal violet is a staining agent which provides the ability to isolate RNA in a dPAGE gel without exposure to UV irradiation. (A) Crystal violet staining used to image bands of hairpin ribozyme in denaturing PAGE gel. Shown is decreasing amounts of raw transcript, from 10 µL to 2.5 µL, in which all lanes show clearly visible amounts of product. (B) Native PAGE gel of gel purified HpRz using crystal violet staining. This method shows a marked decrease in the presence of the top band material compared to samples purified using UV shadowing. Average amount of top band was ~8-10% of total sample.

Utilization of crystal violet has also proven to be a very sensitive method of isolating RNA is a dPAGE gel. When working with previously gel purified HpRz of a known concentration, we were able to detect as little as 1 µg under ambient light (**Figure 24**). Realizing that we would also be utilizing this method for isolating transcription products from an in vitro transcription, we tested various volume of a typical 150 µL transcription of *glmS*-HpRz with GlcN6-P to induced *glmS* cleavage. Here, the desired HpRz transcript could be seen in as little as 2.5 µL of transcript (**Figure 25**). Crystal violet staining proved adequate to detect small amounts of both of these sources of RNA.

As such, it provides a viable alternative to the UV exposure needed for shadowing or ethidium bromide staining.

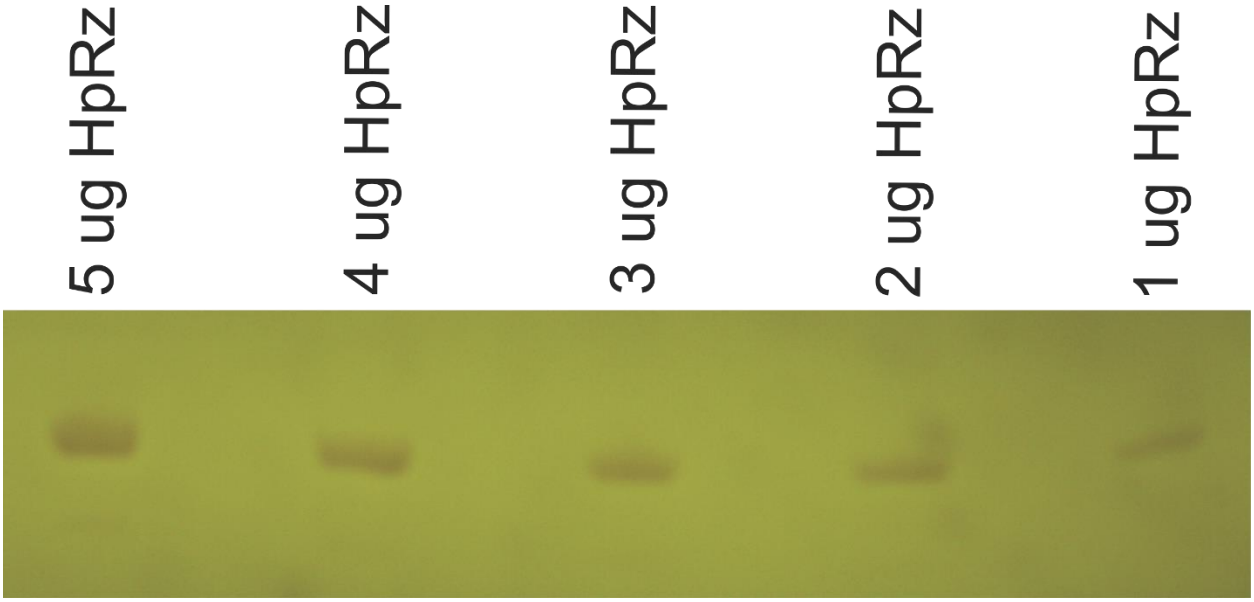


Figure 24: Crystal Violet Allows For the Detection of 1 ug of RNA in a 1.5 mm dPAGE Gel.
To assess the detection limit of crystal violet, a titration of RNA was resolved on a gel and stained with crystal violet. Samples of gel purified hairpin ribozyme were subsequently resolved on a 10% 8M urea PAGE gel. Sample amount varied from 5-0.5 ug of material. RNA could be seen with as little as 1 ug of material, but was not visible with 0.5 ug (not shown). Samples were resolved on 8M urea, 10% denaturing PAGE gel.

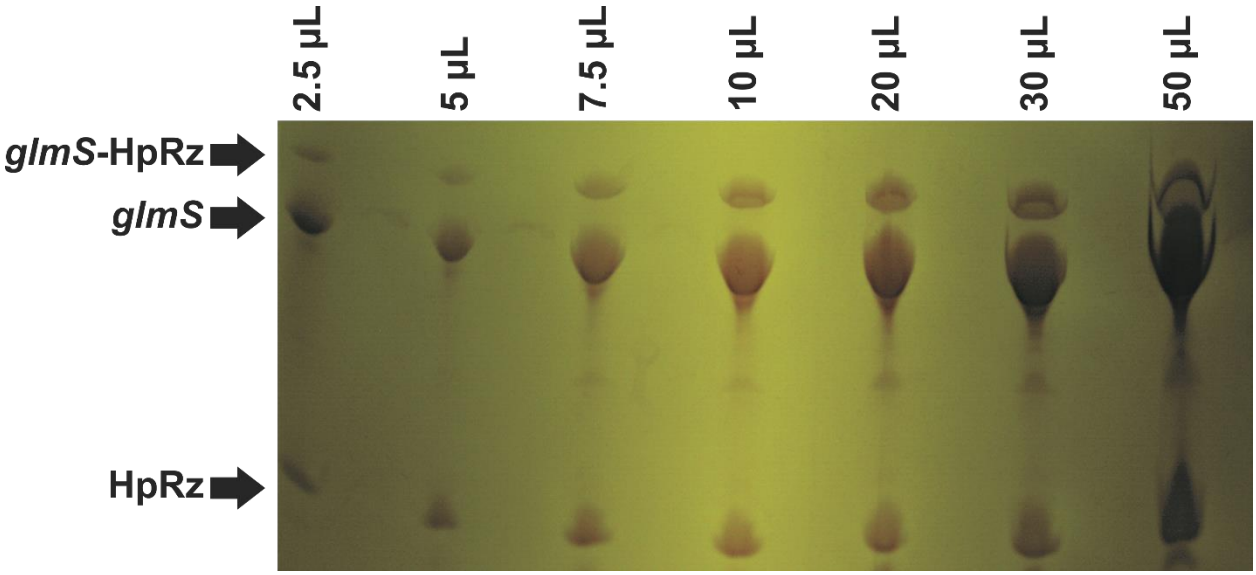


Figure 25: Detection limit of RNA Staining With Crystal Violet After Transcription. Alternatively to finding the amount of RNA that is visible under RNA staining, the volume of a typical *in vitro* transcription reaction was used as a measure of detection limit. Volumes of a transcription of *glmS*-HpRz with GlcN6-P in solution from 50 µL to 2.5 µL were visible after staining under ambient light. Samples were resolved on 8M urea, 10% denaturing PAGE gel.

Like many structured nucleic acids, the hairpin ribozyme exhibits static structural heterogeneity. Denaturing PAGE purification introduces a baseline level of structural heterogeneity to preparations of hairpin ribozyme that is increased upon UV exposure. If the observed structural heterogeneity were due to a deeply furrowed energetic landscape, typical folding temperatures ca. 70°C should be sufficient to overcome energetic barriers separating misfolded material from natively folded material. However, exposure of HpRz to elevated temperature fails to induce heterogeneity in natively purified material or remove it in dPAGE-purified material, suggesting that most of the heterogeneity is due to covalent modifications to the structure.

Of the purification techniques examined here, nondenaturing purification provides the only route to structurally homogeneous preparations. A more straightforward alternative is dPAGE purification followed by crystal violet staining, which avoids UV exposure while maintaining the ability to directly visualize a gel-purified sample without the need for expensive fluorescence scanners. Intercalating fluorescent dyes such as SYBR-Gold are often capable of excitation by visible light (488nm), and thus may offer another means of localizing bands without exposure to damaging UV irradiation. While UV shadowing with a 254 nm lamp allows for easier band localization, it damages RNA more than twice as rapidly as, and to a greater overall extent than, a 312 nm lamp. The fact that these methods plateau in damage is in agreement with the concept of the skin effect where RNA closer to the UV source shields RNA further down in the gel material. Despite previous publications asserting difficulty localizing gel purified RNA with 312 nm light (164), the contrast at this wavelength consistently proved sufficient to localize in vitro prepared transcripts of HpRz. Hence, we conclude that lower-energy UV is helpful in avoiding excessive damage to samples.

While UV damage of nucleic acids is well characterized, the molecular origins of heterogeneity introduced by dPAGE purification is unclear. The source might be covalent modification of the RNA by a component of the gel material itself, but this appears unlikely given previous results from mass spectrometry of a similar HpRz construct. Alternatively, the harsh conditions of denaturation and precipitation used during gel purification of RNA samples may be playing a role. By removing the natural co-transcriptional fold, the pathways to alternative folds within deep energy wells similar to

native folds are opened. These pathways may well be gated by highly stable multivalent cations which stabilize these dispersities of folds and result in functional variation.

If UV exposure caused the same long range cross-linking seen in previous studies (172), we would have expected to see increased early reverse transcription stops at U41, a putative site of dimerization. While we cannot distinguish between local dimerization (U41-U42) and long range dimerization seen previously by RT alone, the former seems far more likely in the denaturing conditions of the urea PAGE gel. Additionally, the slower running band seen in dPAGE gel of the long range dimerization isn't present here (**Figure 11**C), thus eliminating the long range crosslink as a possibility. But this is not the only source of structural variation seen, as simple gel purification resulted in 10-15% of material displaying a slow migration in nPAGE gels. This would suggest that the deep energy wells of the rugged energy landscape that the HpRz occupies could be responsible for this additional variation. As samples are denatured and heated during electrophoresis, only to be later heated and refolded, RNA is able to access these wells and stably occupy them. Alternately, there could be additional sources of damage from chemicals in the gel or elution.

A persistent difference between the results presented here and those previously published is as simple as the constructs themselves. While the one-strand construct used here was simply a linked version of the two-strand construct utilized previously, this presents interesting conundrums to consider in terms of purifications. The earlier characterizations of memory effects were based on 2 strand constructs which required the chemical denaturation and separation of the RzA and RzB strands. If the observed

heterogeneous behavior has a source in each of the strands, this separation and subsequent mixing allows for combinations of fast and slow strands to recombine. By limiting this additional mixing of populations, we are able to stably form populations of compact material.

Footprinting of the hairpin ribozyme allowed us to sort through the potential modes of variation. First, by comparing the reactivity difference between IA and 1M7, it seems unlikely that the difference in samples is due to a sugar pucker variation. While there were some bases with an elevated reactivity in the presence of IA, none exhibited the drastic difference indicative of a 2'-endo conformer. These regions did not show uniformity of reactivity either, leading us to believe this is a highly dynamic region. However, the increased reverse transcription stoppage in the B-loop, particularly in the U41-42 region, was indicative of a pyrimidine dimer. Under normal docking conditions, U42 flips out of the helical stack to form hydrogen bonds with G11, U12, A22, and A23, known as the U42 binding pocket (47, 174, 175). With U42 covalently linked to U41, this likely interferes with the ability of one or both bases to form proper hydrogen bonds. Previous work has shown that interfering with any one of the 5 hydrogen bonds formed by U42 in the docked conformation, the free energy of folding is increased ~2-4 kJ/mol, reducing the stability of the docked conformation (174). With this dimer at a region involved in interhelical interactions formed during docking, the perturbation of this dimer affects the ability of the ribozyme to remain in the docked conformation. Additionally, NMR studies of the loop B structure have shown U41 looped out of the helical stack and U42 in the stack, likely indicating the undocked structure is capable of undergoing

rearrangement in the B-loop (176). In either case, the two uridine residues as spatially separated, which would lead a dimer of the two to be incapable of functioning normally.

Despite the levels of heterogeneity we were able to induce in the hairpin ribozyme, the degree to which we were able to reduce and eventually eliminate this variation is great. Commonly utilized methods of sample detection include UV shadowing, ethidium bromide (EtBr) staining, and SYBR Gold or Green II staining. While some UV shadowing protocols recommend the use of longer wavelength 312 nm lamps (152), 254 nm lamps are still commonly used. While shadowing avoids exposing samples to additional chemicals, it has deleterious effects on nucleic acids (163, 164). Additionally, it is limited in sensitivity, detecting 0.3 μg of nucleic acids (A Practical Approach; Recombinant and In Vitro RNA Synthesis). Ethidium bromide provides greater sensitivity, detecting as little as 1 ng of double stranded DNA, but this is inhibited by several factors. Single stranded DNA and RNA often require 10 times this amount of material for detection. Additionally, PAGE gels reduce sensitivity 10-20 due to quenching of EtBr fluorescence by polyacrylamide. Perhaps most importantly, EtBr is known to be a potent mutagen and requires careful decontamination prior to disposal (DNA Life Technologies). To achieve greater sensitivity, SYBR dyes can be used in place of EtBr. Capable of detecting as little as 1 ng of nucleic acid, they have not shown mutagenic at concentrations used for detection (177). But with any of these dyes, they require excitement by UV illumination for direct visualization of nucleic acid. Crystal violet is frequently used in the purification of DNA, while its use in the purification of RNA is currently underutilized. Capable of detecting as little as 8 ng of nucleic acid when

used in conjunction with methyl orange, it can be seen under ambient light, avoiding UV exposure (173). Its ability to quickly and easily provide a method of sample visualization after gel purification provides a straightforward means of purification with minimal variation in sample preparation.

Further, these variations were eliminated entirely by using a non-denaturing purification technique, providing a single population of material. While the presence a 2' 3' cyclic phosphate at the 3' end of the purified RNA has not shown to adversely affect our experiments, it is possible that this would present a potential barrier in certain situations. Non-denaturing purification of RNA is likely to render bulk studies of RNA more reproducible and more accurately recapitulate biological behaviors. This has implications for any structural or functional studies of nucleic acids, but especially for single molecule studies, where any conformational heterogeneity becomes readily apparent. Additionally, the ability to remove noise or reduce the numbers of orders needed to model the behaviors of dynamic molecules allows us to more accurately study RNAs of interest.

CHAPTER 3: Development of an Active Ribosome System

3.1 Introduction

Development of single molecule tools to investigate initiation and translation events have progressed greatly in the past decade. One of the earliest single molecule observation of translation was based on the tethered particle method. In this experiment, surface immobilized ribosomes translate an mRNA with a particle tethered to the 3' end of the message. Because of the length of mRNA, the space that the particle can diffuse in restricted; the longer the message, the greater the space it can diffuse in. As the message is translated and the mRNA moves through the ribosome, the restricted diffusion is further reduced. By utilizing this method and the average size of unstructured mRNA, translocation rates of ribosomes translating poly-U messages was calculated to be 2.2 amino acids/second.

These experiments were followed by two competing methods of site specific labeling of the ribosome. From the work of Harry Noller's laboratory, labeling focused on modification of ribosomal proteins, particularly ones that were spatially close to each other. Located on the surface near the interface between the 50S and 30S subunits, large subunit protein 9 (L9) and small subunit proteins 6 and 11 (S6 and S11) were chosen for labeling. In each of these proteins, single cysteine mutants were cloned for later

expression and purification. From here, proteins were labeled and reintegrated with the ribosomal subunit. While L9 was consistently used to label the large subunit, S6 and S11 were used alternately and interchangeably to label the small subunit. In addition to providing site specific labeling, the proximity of these protein the 70S ribosome allowed for FRET between the proteins. While intersubunit rotation had been shown via EM, single molecule observations allowed real-time rotation to be observed by the increase and decrease in FRET signal as the protein moved closer together or further apart. Additionally, they were able to demonstrate the at least two distinct states (rotated and unrotated) of the ribosome existed during the act of translation.

Alternately, work from Jody Puglisi's group focused on modifying the rRNA to allow for hybridization of fluorescently labeled probes. By adding semi-stable helices of RNA to surface exposed loops of existing ribosomal helices, this provided a platform for labeling. This new helix was partially complementary resulting in a semi-stable structure with multiple mismatches that reduce the stability of the helix while maintaining the functionality of the ribosome and viability of the E. coli. Starting with the small subunit 16S rRNA, mutations to multiple location were tested. While numerous surface exposed helical loops were mutated for potential labeling locations (helix 6, 10, 17, 26, 33a, 39, and 44), only helix 6, 10, 33a, 39, and 44 proved viable (178). By separating the chemical modification from the biological system, a wide variety of modifications were possible. This provided both a site for fluorescent labeling as well as surface immobilization by use of a DNA extension and biotinylation (178, 179). Additionally, a single loop in the 23S rRNA (helix 101) proved viable for extension and labeling (180).

When helix 101 was labeled in tandem with helix 44, the two locations were capable of acting as a FRET pair providing insight into the rotation state of the ribosome, similar to previous work of Noller and co-workers (77, 78, 180).

While much of the work on ribosome dynamics has focused on the dynamic of the ribosome relative to itself, the ribosome's interactions with mRNAs is still ill-understood. As the importance of mRNA structure has become more apparent, how these structures affect ribosome assembly and subsequent translation is of great importance. Our work here was aimed to develop the necessary tools for purifying and labeling an active ribosomal system that could subsequently be utilized for single molecule studies of translation. While elements of single molecule translation systems exist, there wasn't an assembled system designed for to examine interactions of ribosomes with mRNA.

3.2 Materials and Methods

Our method of ribosome labeling required the hybridization of a fluorescently labeled DNA oligo to the extended helix 33a of the 16S rRNA. This method was pioneered in the lab of Joseph Puglisi and involved the extension surface exposed loops into semi-stable helices (178). The helix termed HP5 was found to have the best combination of labeling efficiency and translation ability once labeled.

HP5 Helical Extension: 5'-CGUCGCCUAUCCUGAUCUCCCCG-3'

This work was done in collaboration with Phil Cunningham's lab at Wayne State University. At the hx33a location, the semi-stable helix HP5 mutation was inserted into the 16S rRNA. This helix has a relatively low melting temperature of 42 °C. As initially described by Puglisi and coworkers, DNA oligonucleotides complementary to the 19 nucleotides to the 3' portion of the helix were utilized for labeling and subsequent hybridization to the ribosome(178). This was termed the "sp5" oligo. Our work used the same approach, ordering primary amine 3'-, 5'-end modified DNA oligos from Invitrogen.

SP5 Oligo: 5'-GGGAGATCAGGATA-3'

For a standard 100 uL labeling reaction, 200 µg of DNA were diluted to 86 µL by the addition of 10 µL 1 M sodium carbonate/bicarbonate, pH 8.3 and water to volume. One single reaction dye pack of Cy3/3B/5-NHS ester was dissolved in 14 µL of dimethylsulfoxide (DMSO) and added to the DNA. Samples were incubated at room temperature (~22 °C) overnight (16hrs) protected from light. Samples were then diluted to 500 µL and then desalted over a Nap-5 (prepacked Sephadex G-25 gel filtration) desalting column. After elution, 0.1 volume 3 M sodium acetate (NaOAc), 0.01 volume 100 mM GTP, and 3 volumes 100% ethanol were added and the DNA placed in -80 °C for ~30 minutes. Samples were centrifuged at 9,000 RPM in a Beckman 5314 centrifuge for 30 minutes, ethanol decanted, and the DNA pellet dried in a vacuufuge for ~30 minutes. Samples were then dissolved in 100 µL of ddH2O. A 1:10 dilution was then

run on a reverse phase HPLC C-8 or C-18 column as an analytical sample. The mobile phase was a combination of acetonitrile and triethylamine acetate (TEAA). The remainder of the labeled DNA was then purified, collected, and dried in the vacuufuge. After drying, DNA was resuspended in TE buffer (10mM Tris, pH 8.0, 1mM EDTA) and stored at -20 °C.

Ribosome labeling was accomplished by hybridizing a fluorescently labeled SP5 DNA oligo to the extended helix. The hybridization buffer consisted of 50 mM TRIS, pH 7.5, 100 mM KCl, 5 mM NH_4OAc, 0.5 mM $CaCl_2$, 5 mM Mg(OAc)2, 6 mM βME, 5 mM putrescine, 1 mM spermidine. Probe was hybridized to the ribosome at a ratio of 1:10 or 1:5, with the ribosome concentration held constant at 100 nM. Everything except the ribosome was combined and heated to 42 °C. After ribosomes were thawed, they were added to the labeling reaction for 2 minutes at 42 °C to melt the semi-stable helix, followed by 2 hours at 37 °C to allow for hybridization.

After hybridization, 30S ribosome samples were resolved on agarose-acrylamide composite native gels to observe the gel shift induced by the hybridization. This gel consisted of 0.5% (w/v) agarose, 2.75% acrylamide, 1% sucrose, 1 mM dithiothreitol (DTT), and 1X running buffer (25 mM TRIS-acetate, ph 7.5, 6 mM potassium-acetate, and 2 mM magnesium-acetate) in 200 mL of total gel volume. The agarose was prepared in 100 uL of water at a 2X concentration by heating to boiling and cooling to 37 °C. During the cooling, the rest of the gel components were prepared to a 2X concentration in 100 mL and heated to 37 °C. The two halves were combined, 200 µL each of tetramethylethlenediamine (TEMED) and 50% (w/v) ammonium persulfate (APS),

before gel casting. Samples were loaded using 50% (w/v) sucrose as 5X loading buffer. Gels were run in 1X running buffer at 110V for 2 hours at 4 °C. Gels were imaged on a Typhoon 9410 using the appropriate filters for the dye used in the experiments. Assayed by gel shift assay on a composite agarose/acrylamide, incorporation of labeled DNA into the ribosome was quantified by scanning the gel on a Typhoon 9410.

Utilizing the Inoue method, we were able to grow competent samples of the delta7 cell lines. A single colony of a cell line was picked a plate of cells and transferred to 25 mL of LB media with matching antibiotic. This addition of antibiotic was possible as we wanted to maintain the plasmid in the cells and screen out any possible contaminant cell lines. This culture was grown 8 hours at 37 °C. Then 3 1L cultures were inoculated with 10, 4, and 2 mL each and grown at ~18 °C for 14 hours. Observing A_{600}, when any one of the cultures reached 0.55, that culture was iced for 10 minutes and cells were collected by centrifugation at 2500g for 10 minutes at 4 °C. Cells were then resuspended in 80 mL of transformation buffer. Transformation buffer is composed of 10 mM PIPES, pH 6.7, 55 mM $MnCl_2$, 15 mM $CaCl_2$, and 250 mM KCl. Cells are then recollected as before. Cells were then resuspended in 20 mL of transformation buffer, 1.5 mL of DMSO was added, and the mixture was put on ice for 10 minutes. From here, 220 µL aliquots were stored in sterile 1.5 mL tubes, snap frozen in liquid nitrogen, and stored at -80 °C.

In order to produce S30 extracts to allow for in vitro translation, we utilized a protocol we adapted from multiple sources (181-183). After growing out cells on LB-agar agar plates, single colonies were grown out overnight in 76 mL of liquid LB culture.

These cultures were then used to inoculate 6L of LB liquid media with a 1:100 dilution. These cultures were grown 4 hours and immediately chilled on ice. Cells were then collected by centrifugation at 5,000 rpm for 30 minutes in a Beckman JLA 8.100 rotor. Cells were then resuspended in 40 mL of Buffer I (10 mM Tris-OAc, pH 8.0, 14 mM Mg-$(OAc)_2$, 60 mM KCl, 6 mM β-mercaptoethanol, and 50 µg/mL phenylmethylsulfonyl fluoride) and washed twice to remove any dark colored contaminants. After final wash, the pellets were resuspended in (4 mL/g cells) buffer II (10 mM Tris-OAc, pH 8.0, 14 mM Mg-$(OAc)_2$, 60 mM KCl, 1 mM dithiothreitol, and 50 µg/mL phenylmethylsulfonyl fluoride) and incubated on ice 15 minutes. Cell solutions were then passed through a cell disruptor twice to lyse cells and the resulting lysate was centrifuged at 30,000 g in a Sorvall SS-34 for 30 minutes at 4 °C. The supernatant was transferred to a new tube and centrifuged again at 30,000 g for 30 minutes at 4 °C. To this S30 extract, 0.15 volumes of preincubation mixture (0.75 M Tris-OAc, pH 8.0, 21.3 mM $Mg(OAc)_2$, 7.5 mM DTT, 75 µM amino acids, 6mM ATP, 20mg/mL phosphoenol pyruvate, and 50 units pyruvate kinase) were added and the mixture was shaken at room temperature for 70 minutes. This solution was then dialyzed against 3 changes of buffer II at 4 °C. Samples were aliquoted into 1 mL portion, snap frozen in LN_2, and stored at -80 °C.

Messenger RNAs were produced by in vitro transcription utilizing T7 RNA polymerase. Transcriptions were typically 100 µL and carried out in 40 mM TRIS, pH 8.0, 25 mM MgCl, 1 mM spermidine, 100 mM NaCl, 0.01% Triton X-100, 4 mM NTPs (each), 5 mM DTT, 5 pMols of template DNA, 0.01 U/µL PPiase, and 0.1 µg/µL T7 polymerase. The other transcription method was the one previously mentioned for the

hairpin ribozyme. Samples were then mixed with a 2X denaturing gel loading buffer (WHAT IS THIS?). Transcriptions were then resolved on a 4% denaturing polyacrylamide gel electrophoresis (dPAGE) gel at 20 watts (W) for 8 hours. Bands were visualized by UV shadowing with a Spectroline model EB-180C medium wave (312 nm) ultraviolet lamp. Bands were excised, crushed, and soaked in crush-n-soak buffer (0.5 M ammonium acetate, 0.1% (w/v) sodium dodecal sulfate, and 0.1 mM ethylenediaminetetraacetic acid) in a polyprep column overnight at 4 °C. Eluent was collected and a chloroform extraction was performed to remove the SDS. This is accomplished by the addition of 1 volume of chloroform followed by gentle mixing, and discarding the chloroform layer. An ethanol precipitation was performed on the remaining aqueous layer. After the precipitation, samples were dried in a vacuufuge, resuspended in TE, pH 7.0, and stored at -20 °C.

Plasmid swapping in SQ197 was necessary to express our modified ribosome. These cells were then grown in ampicillin containing (50 µg/mL) liquid culture for 2 weeks at 37 °C, changing media every 24 hours and plating a sample of cells. This was done to put the cells under constant growth and maintain the mutant plasmid that had been transformed in. Without kanamycin in solution and an alternate rRNA coding source on the pKK3535 plasmid, the hope was that some cells would eventually lose the k4-16 plasmid and carry only the pKK3535. This would be observed by the continued ampicillin resistance but loss of kanamycin resistance. To this end, I began screening individual colonies by repliplating on ampicillin then kanamycin plates. At each media change, a small sample of cells was plated on an ampicillin containing LB agar plate and

grown overnight at 37 °C until individual colonies could be distinguished. Then, 100-250 colonies were transferred to 1-2 gridded Amp-LB agar plates, with each colony individually numbered and grown overnight at 37 °C. Each surviving colony was then transferred to an identical grid pattern on a Kan-LB agar plate and grown overnight at 37 °C. We looked for colonies that grew on the Amp plate but not on the Kan plate, showing that it had lost the k4-16 plasmid.

Cell lines expressing the ribosomes of interest were streaked out on LB-agar plates. Single colonies were grown in 50 mL liquid cultures of LB overnight (12-16 hours) at 37 °C. These cultures were diluted 1:200 into 2 L of LB media and grown until they reached A_{600}. Cultures were cooled in an ice bath for 30 minutes and collected by centrifugation at 5,000 rpm for 30 minutes in a Beckman JLA 8.100 rotor. Pellets were resuspended in 30 mL of buffer A (20 mM Tris, pH 7.5, 100 mM NH_4Cl, 10 mM $MgCl_2$, 0.5 mM EDTA, and 6 mM βME) and passed through a cell disruptor to lyse the cells. The lysate was then cleared by 2 centrifugations at 30,000 g in a Sorvall SS-34 for 30 minutes at 4 °C. The collected supernatant was then loaded over 35 mL buffer D (1.1 M sucrose, 20 mM Tris, pH 7.5, 500 mM NH_4Cl, 1 mM $MgCl_2$, and 0.5 mM EDTA) in a Ti-45 ultracentrifuge tube. These were then centrifuged at 37,000 rpm (~160,000 g) for 16 hours to pellet the ribosomes.

While 10-40% sucrose gradients were initially poured by use of a gradient pourer, an alternative method of gradient making was later utilized at the recommendation of Dr. Lolita Piersimoni. Instead of pouring mixtures of sucrose to manually form gradients, single concentrations of sucrose could be frozen and thawed to form gradients for the

separation of ribosomal subunits (184). Gradients were prepared by pouring a 15% sucrose solution into ultra-centrifuge tubes and freezing at -80 °C overnight. Tubes were then pulled out the next morning and thawed at room-temperature for ~6 hours and stored at 4 °C until they were used for separating the ribosomal subunits.

Once cleaned ribosomes were separated from the S100 extract, the subunits had to be separated into 30S and 50S fractions. Ribosomes were resuspended in buffer E (50 mM Tris, pH 7.5, 150 mM NH_4Cl, 1 mM $MgCl_2$, 6 mM βME) and layered over the sucrose freeze gradients. These were then centrifuged at 20,000 rpm in an SW-28 swinging bucket for 18 hours at 4 °C. Gradients were then fractionated by density and A_{260} was taken. Peaks corresponding to the 30S and 50S ribosomes were separated and those fractions were pooled. Mg^{2+} was raised to 10 mM using 1 M $MgCl_2$. These were then spun down in Ti70 for 16 hours at 66,000 rpm at 4 °C. Subunit pellets were resuspended in 0.5 mL of buffer A and A_{260} of each subunit was taken. Aliquots of 10-20 µL were snap frozen and stored at -80 °C. Subunit purity was assessed by resolving 500 ng of subunit on a 1% agarose TBE gel.

3.2 Results and Discussion

Initial Ribosome Labeling and Purification Methods

For our purposes, we chose to utilize helix 33a as it is located on the leading edge, or "beak" of the 30S ribosome. This location would allow us to form a potential FRET pair with fluorophores hybridized to mRNA being processed by the ribosome. Additionally, this was one of the better characterized helices in the literature(178). In our labeling experiments, probe was hybridized to those ribosomes with the h33a extension

(Mut) as well as ribosomes lacking this helical extension (WT) (**Figure 26**). A supershift of probe signal corresponding to specific ribosome labeling was only seen in samples of probe hybridized to mutant ribosome.

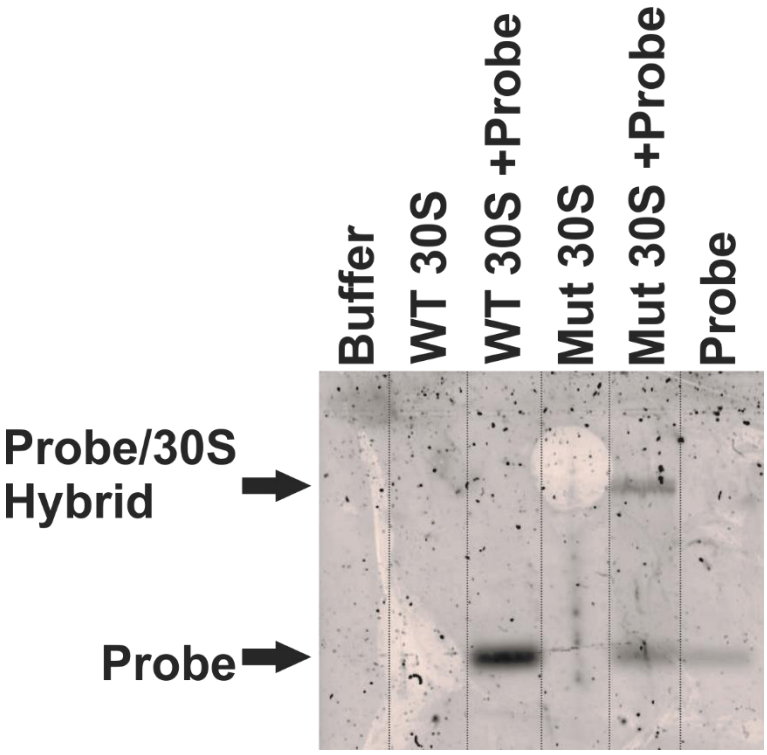


Figure 26: Specificity of 30S Ribosome Labeling.

Wild-type (WT) ribosomes underwent labeling reaction alongside mutant (Mut) ribosome with the helix 33a mutation to assess the labeling specificity of the DNA probe. The signal of Cy5 probe is seen alone in the last lane. Mock reactions without probe were also carried out. In the labeling reactions, the probe was supershifted only in the presence of the mutant ribosome. Samples are resolved on a 0.5% agarose/2.75% acrylamide composite gel run 3 hours at 110V at 4 °C.

This long incubation was deemed excessively long, so a time trial experiment was performed. In it, a standard 100 µL reaction was split into four 25 uL samples and the subsequent reactions were incubated at 37 °C for 2, 1, ½, and ¼ hours(**Figure 27**).

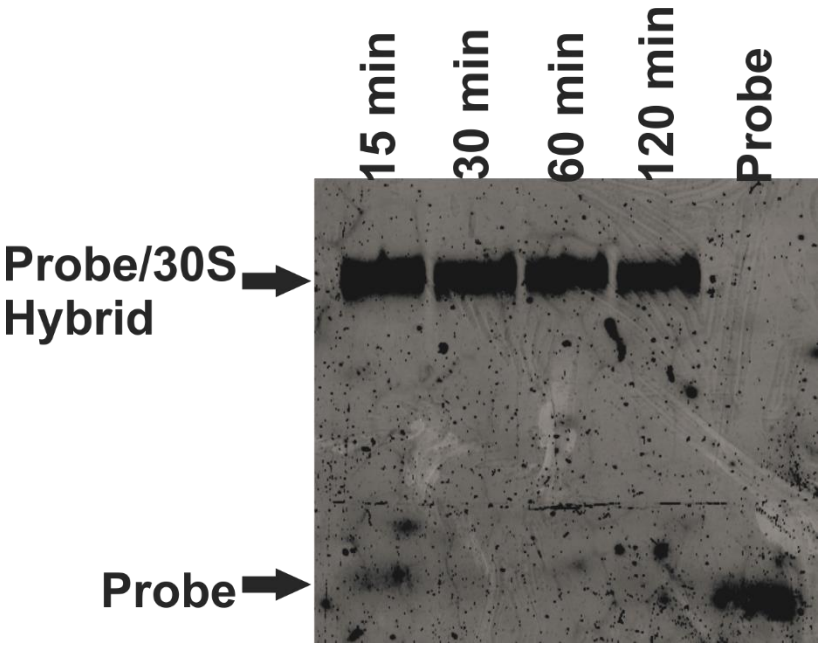


Figure 27: Timed Hybridizaiton of 30S Ribosome With Cy5-DNA Probe.

In order to reduce the hybridization time of probe to ribosome, a series of samples were hybridized for varying lengths of time. Samples of ribosome were hybridized for 15, 30, 60, and 120 minutes with the DNA probe before being resolved on a 0.5% agarose/2.75% acrylamide composite gel. We found that the majority of the probe shifted to the ribosome band by 15 minutes and all of it shifted by 30 minutes. We used 30 minute incubations for all subsequent experiments.

This experiment showed a supershift of all probe signal to the ribosome band by 30 minutes and a majority of material by 15 minutes. Finding that all of the probe was shifted by 30 minutes, all subsequent incubations were done for 30 minutes.

Initial Expression System

In order to purify appropriately modified ribosomes, we utilized an expression system from the lab of Philip Cunningham. Expression of modified ribosomes was dependent on 16S rRNA being transcribed from a modified rRNA gene carried on a plasmid. These ribosomes carried a mutant anti-Shine Dalgarno (aSD) which only

recognized mRNA which carried a complementary mutant Shine Dalgarno (SD) sequence. These mutant ribosomes were only capable of translating gene products preceded by this mutant SD sequence. To force the E. coli to transcribe these 16S rRNA and process them into mature 30S particles, the cells are exposed to chloramphenicol as they grow. Normally fatal to the cells, it can be mitigated by the expression of chloramphenicol acetyltransferase (CAT) which is capable of inactivating the chloramphenicol. Carried on a second plasmid in the cells, CAT gene is preceded by a mutant SD sequence. Thus, in order to survive in the chloramphenicol containing environment, the cells must transcribe the mutant rRNA to translate these CAT genes.

The obvious problem with this expression system is that a mixed population of ribosomes is produced, with both wild-type and mutant ribosomes. This required a unique purification system beyond how the average ribosome purification. After purifying 30S ribosomal particles, mutant ribosomes are "fished" out of the solution to separate them from wild-type ribosomes. Biotinylated DNA strands complementary to the 3' end of the mutant ribosomes (which contains the mutant SD) are added to the mixed ribosomes. After hybridizing to the mutant SD sequence, these ribosomes were immobilized on streptavidin coated beads and separated from wild-type ribosomes by centrifugation. Purified mutant ribosomes were then eluted by a low salt wash.

An aspect of this purification system is that was detrimental to future was the basis of the expression and selection: the mutant anti-Shine Dalgarno sequence. While functional for studying mutations at distal portions of the 16S rRNA, these ribosomes are unable to translate mRNA containing a wild-type Shine Dalgarno sequence. This

requires the mutation of sequences of cDNA corresponding to the mRNA of interest to include the mutant SD sequence. While not changing the translation product (as it is located before the coding sequence), it likely changes the local mRNA structure. This is detrimental for several reason. First and foremost, it has been shown that secondary structure, as well as sequence and codon usage, determines translation rate. Specifically, the structure in the first 40 nucleotides of an mRNA has a great effect on overall translation rate (185). Additionally, the wild-type sequence of SD is essential to the function of translational riboswitches. In these systems, the exposure or occlusion of the ribosome binding site (RBS) is controlled by base pairing interaction of the SD sequence and other complementary regions of the expression platform. While these interactions are regulated by the aptamer sequence, mutating the SD sequence would impede the function of the riboswitch. Despite the fact that compensatory mutations may be possible, it would likely have unexpected results and greatly change the function of the riboswitch. As such, it become apparent that our translation system would have to be changed to one with a wild-type SD sequence for translation efforts.

Reporter Systems

Numerous reporter systems have been utilized to examine the functional aspects of the ribosome over the course of my research. The earliest and simplest was also the most well studied system. The T4 phage gene *32* has been a model system of translation and initiation. Additionally, the T4 gene *60* is a unique gene encoding one subunit of a heteromeric trimer topoisomerase. The unique nature of this gene lies in the gene product, or products, as about half of the product is that of the full length protein (18.4

kDa) and half is a truncated protein (7.8 kDa). This truncation corresponds to an in-frame stop codon after Gly^{46} (186). However, this stop codon and the proceeding 50 nucleotides are often bypassed, with translation resuming on a downstream Gly codon and the full length protein translated. This sequence of RNA is often referred to as the coding gap. In work by Gabrielle Todd, it was shown that this region consists of a strong hairpin structure starting at Gly^{46} and this structure likely causes the ribosome to engage in the translational bypassing needed to translate the full length protein(168). In this work, a series of control transcripts were also generated to provide positive and negative controls of bypassing. The first, termed the Δgap, had the coding gap deleted from the sequence and translated into full length product only when translated. The other, termed the "trunc", was a truncation control, where only the first 46 amino acid codons are transcribed with the in-frame stop codon and the resulting translation only produces the truncated 7.8 kDa peptide. Messenger RNA transcripts of each of these was produced to act as templates for our initiation and translation experiments.

Early Single-Molecule Observations

To establish our system as a valid platform for further initiation and translation studies, we decided to test ribosome association with mRNA under single molecule observation. While were still working on getting an active translation system, we decided to observe the 30S ribosome association with mRNA in the absence of any other factors. During translation initiation, the 30S ribosome must scan along the mRNA to find the start site. This is accomplished by the binding of the aSD on the 30S to the SD sequence on the mRNA. Based on this interaction alone, we believed we could observe

colocalization of Cy5 labeled 30S ribosome with surface immobilized Cy3 labeled gene *60* mRNA. We were able to immobilize the mRNA by use of a biotinylated DNA capture strand complementary to the 5' end. Fluorescent labeling was accomplished by the use of a Cy3 labeled DNA oligo complementary to the 3' end. After immobilization, mRNA density was observed before 30S was flowed on and allowed to associate with the mRNA. While rare, we were able to observe colocalized signals of ribosome with mRNA in a few molecules (**Figure 28**). This finding provided us reassurance that our ribosomes were capable of binding to mRNA via the modified SD-aSD interaction. With this simple association functioning, further studies of initiation complex formation would have been possible on these mutant mRNAs.

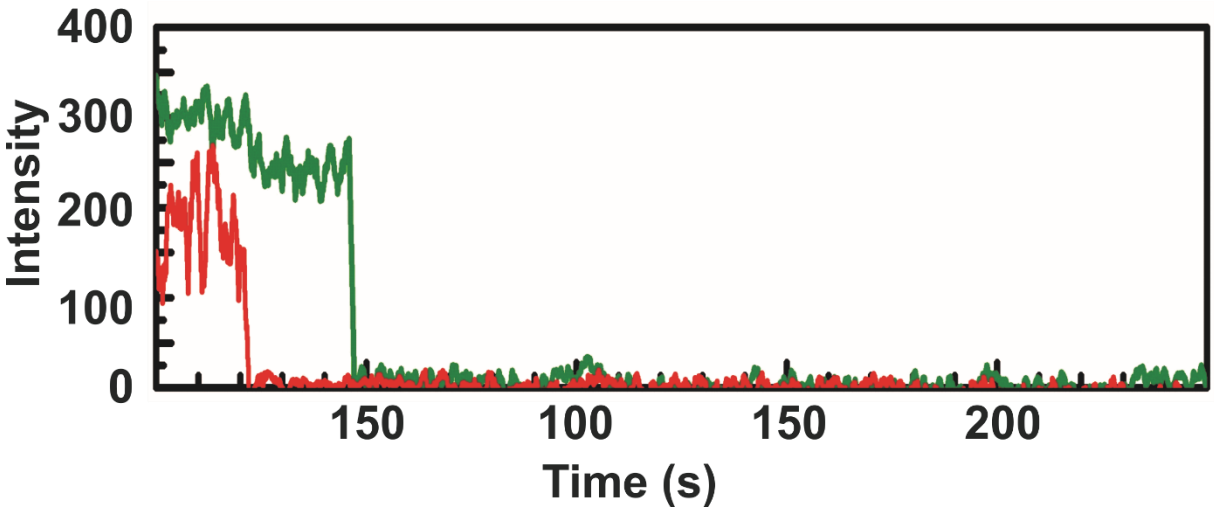


Figure 28: Colocalization of 30S Ribosome With mRNA.

With our first labeling system, we performed a 30S-mRNA association under single molecule conditions to look for biological activity. Slide surfaces had gene *60* mRNA immobilized on the surface. 30S ribsome labeled with Cy5 dyes (red) colocalized with gene *60* mRNA labeled with Cy3 (green). This colocalization was done in the absence of initiation factors. This interaction was likely due only to Shine Dalgarno/anti-Shine Dalgarno interaction. This basic assay showed that future experiments of initiation complex formation would be possible as 30S ribosome associated with mRNA.

Transition to Wild-Type Shine Dalgarno Ribosome

While our mutant ribosomes could bind to mRNA with a mutant SD, we knew that we had to utilize wild-type ribosomes to translate genes with wild-type SD sequences. This would be essential in studies of mRNAs where the SD sequence was involved in specific structural elements, such as riboswitches. In order to transition to ribosomes with wild-type SD, we had to create a template for the rRNA with the appropriate helical extension. As a basis of for this expression, we started with the pKK3535 plasmid which constitutively expresses the sequence for both 16S and 23S rRNA.

From here, we tried several different methods of introducing the necessary mutation into the sequence for the 16S rRNA. Our first attempts revolved around touchdown mutagenesis. In this, DNA primers directed to either side of helix 33a are used to amplify the plasmid. Additionally, these primers have a 5' overhang containing the 23 nucleotide hp5 extension. In each successive round of replication, copies of the mutated plasmid are doubled. As such, the annealing temperature is raised 1 degree every two rounds of replication to account for the increased amount of full length primer binding. Otherwise, mutagenesis followed the standard PFU Ultra protocol by Agilent. While we were able transform these products into DH5α competent cells, sequencing only revealed unmutated plasmids, despite DpnI digestion to remove these prior to transformation.

Next, we identified to unique restriction sites 200 nucleotides apart to either side of helix 33a. Working with undergraduate Khalil Mroue, we decided to try to create an

insert to region. Using a system that had been used previously in our lab, we designed 9 overlapping DNA segments that matched this sequence with the hp5 insert. Once these DNA sequences were annealed to each other, either end corresponded to the restricted restriction site. The pKK3535 plasmid was then restricted to remove this segment and gel purified over a 0.4% agarose TAE gel. The insert was then ligated into the plasmid backbone before being transformed into DH5α competent cells. Again, any positive transformants contatined only the wild-type ribosome sequence. One of the difficulties of this method was likely the low output of the full length insert annealing. After multiple difficulties getting the positive mutant, we contacted the Puglisi lab and they agreed to send us their copy of the mutant in pKK3535.

Cell Lines For Ribosome Expression

For numerous reasons previously mentioned, we decided to transfer to a ribosomal expression system with a wild-type Shine Dalgarno sequence. Unfortunately, this would mean we would have to change the cell lines, as we would no longer be able to separate mutant from wild-type 30S ribosomes using MRE 600 cells. To this end, we decided to utilize a series of cells lines from Catherine Squires' lab (187, 188). In normal E. coli cells, coding for ribosomal RNA is carried on 7 rRNA genes (A, B, C, D, E, G, and H). In these cells, the rRNA genes have been sequentially inactivated and replaced by a single constitutively active rRNA gene carried on a plasmid. These are known as the Δ7 cell line, as they have By utilizing plasmids carrying different antibiotic resistance, variable rRNA constructs can be swapped in and expressed in the absence of wild-type ribosomes.

Unfortunately, by the time we were ready to make this transition, the Squires lab had closed and the cell lines were no longer available. By a stroke of pure serendipity, a collection of these cell lines were preserved in Janine Maddock's lab here at the University of Michigan. They were kind enough to gift us two Δ7 cell lines, SQ 171 and SQ 193, which carried different antibiotic resistances. SQ 171 carried the pKK3535 plasmid with ampicillin resistance (Amp^R) and SQ 193 carried the k4-16 plasmid with kanamycin resistance (Kan^R). These were catalogued in the Maddock Lab as JM 5613 and JM 5614, respectively. Samples of each were grown out in the appropriate resistance LB media and samples stored at -80 °C.

Once we had these cell lines in hand, competent cells of each line had to be made. From here, the mutant pKK3535 plasmid with helix 33a, hp5 extension was then transformed into competent SQ 193 cells. From here, we went through the process of plasmid swapping to get the modified ribosome as the sole RNA expressed. This was accomplished on the 17^{th} trial (after 16 days of growth). Samples of this colony were grown out in LB-Amp and 1 mL aliquots were stored at -80 °C. Mutants were confirmed by purifying the pKK3535 plasmid and sequencing at the UM Sequencing Core.

S30 Preparation

As a method of achieving *in vitro* translation controls, we sought to purify S30 extracts to act as positive controls in *in vitro* translation systems. Utilizing a protocol based on those previously published, we set about purifying cellular extracts from MRE 600 cells (181-183). This cell line was chosen in favor of the BL21 cell as it lacks the RNase I activity that degrades rRNA, which we believed would better preserve the

ribosome (189). After my initial S30 extract was purified, numerous S30 and S100 (S30 lacking ribosomes) extracts were subsequently prepared by Gabrielle Todd, formerly of the Walter lab, utilizing MRE600 and BL21 cell lines, and remain in use as positive translation controls.

Ribosome Preparations

Once we had a cell line expressing the ribosome strain we needed for labeling, it became necessary to purify the individual 30/50S subunits. Utilizing the techniques of both the Green lab at Johns Hopkins and the Maddock lab, we were able to express and purify the ribosomal subunits. Generally, we were able to purify highly active ribosomal subunits that were consistently clean, ie, the 30S ribosome subunits were free of 50S and vice versa.

Noller Ribosomes for Protein Labeling

While we were pleased to have a ribosomal system that we could express and site specifically label for single molecule studies, we were presented with the opportunity of adding another tool to our toolbox. Working in collaboration with Dimitri Ermolenko, his lab was able to provide us with strains of bacteria that lacked the S6 protein in their 30S ribosomes. Additionally, we were also provided with plasmids coding for a single cysteine variant of S6 (77, 78). These were both items which had originally been developed and implemented for fluorescent studies of the ribosome in the lab of Harry Noller. With these in hand, and working with fellow Walter lab member Paul Lund, we set about purifying the ribosomes and proteins separately and later reintegrating fluorescently labeled S6 into 30S ribosomes lacking this protein. Purification of the

ribosome occurred as previously stated for the h33a mutants. Samples of 30S and 50S ribosomes were snap frozen and stored at -80 °C. In addition to this method for labeling the 30S ribosome, we were also provided a cell line for ribosome expression and clones for L9 protein expression. This provides with a potential method for labeling the large ribosomal subunit. While neither of these methods have been utilized for single molecule assays to this point, it is another tool we have to investigate the ribosome. To this point, we have accomplished the fluorophore labeling of the small subunit of the ribosome. This can now be utilized for our future studies of ribosome interactions with mRNA.

CHAPTER 4: Single Molecule Studies of Ribosome Assembly on a Translational Riboswitch

4.1: Introduction

Composed of two domains, riboswitches consist of an aptamer domain and an expression platform. The aptamer domain acts to sense a small metabolite by binding it, while the expression platform acts to control the action of transcription or translation. This action is accomplished by the binding of a ligand binding to the aptamer domain inducing a structural rearrangement of the expression platform. The $preQ_1$ family of riboswitches include both a transcriptional (from *Bacilus subtilis, Bsu*) and translational (from *Thermoanaerobacter tengcongensis, Tte*) variant, the $preQ_1$ riboswitches detect the presence of the metabolite $preQ_1$ (7-amino-7-deazaguanine) (190, 191). Under the conditions of a near-physiological buffered system, these riboswitches display two distinct but interconverting structural states, termed the pre-folded (ligand free) and folded (ligand bound) states. Titration with the $preQ_1$ ligand in both cases stabilized the folded state. Interestingly, the mechanism of ligand-induced folding varied between the riboswitches, with the *Bsu* riboswitch behavior following a conformational selection mechanism, with $preQ_1$ binding preceding folding, while *Tte* follows an induced fit mechanism, with ligand binding and folding happening simultaneously. While the

structural rearrangements of both have been extensively studied *in vitro*, their actions in a more biologically relevant setting have yet to be tested (81). The *Tte* riboswitch is expressed upstream of the nested 1564/1563 gene and acts to sequester the SD sequence in the presence of preQ1 (190, 191). To this end, our assembled toolbox of single molecule translation components (purified initiation factors, labeled ribosomes, and tRNAs) could be used to observe the binding effects that the presence of $preQ_1$ has on ribosome assembly on the *Thermoanaerobacter tengcongensis* 1564/1563 mRNA. By testing ribosome association with this mRNA in the presence and absence of $preQ_1$, we can better understand its effect on gene regulation in bacteria. By developing a single molecule initiation assay, we can directly observe interactions of ribosome and mRNA and see the effects of ligand with a riboswitch.

4.2 Materials and Methods

In order to establish an entirely *in vitro* method of translation initiation, we turned to a recently published set of initiation factors from the Gonzalez lab(192). IF1, 2, and 3 were independently cloned into the ampicillin resistant backbone vector of pProEx-HTb (pProEx-HTb/IF1, pProEx-HTb/IF2, and pProEx-HTb/IF3). The parent vector contains a 6xHis binding domain and Tev cleavage site upstream of a multiple cloning site into which the initiation factor sequence was cloned. The initiation factors were all from wild-type E. coli. A detailed protocol is available in the Methods in Enzymology chapter written by Fei, et al (192). To perform the Tev cleavage in the purification of the IFs, Tev protease was produced in house from a plasmid courtesy of the Gonazalez lab. The

Tev protease gene was in the parent pMal-Tev vector and expressed as per the published protocol (192).

To form initiation complexes (IC), 70S ribosomes were assembled on gene *60* mRNA in the presence of initiation factors (IFs), fMet-tRNA^{fMet}, and GTP. First, mRNA was refolded by heating at 95 °C for 2 minutes and then snap cooled on ice. Separately, the rest of the components of the reaction were assembled in a Master Mix. These reactions are carried out in 1X polymix buffer, which is prepared as a 10X solution (50 mM $Mg(OAc)_2$, 5 mM $CaCl_2$, 80 mM putrescine, 10 mM spermidine, 50 mM KPO_4, 950 mM KCl, 50mM NH_4Cl, 10 mM DTT, final pH 7.5-7.6). To this, 1 M K_2HPO_4 is added to a final concentration of 10 mM. IFs, which are stored between 100 and 200 µM, are added to the reaction at a final concentration 3 µM. Additionally, GTP and fMet-tRNA^{fMet} were added to final concentrations of 2 and 2.5 µM, respectively. Finally 70S ribosome were added to a final concentration of 2 µM to complete the Master Mix. The denatured mRNA was added to the Master Mix and incubated at 37 °C for 30 minutes. After the incubation, $MgCl_2$ was added to a final concentration of 20 mM to stabilize the ribosome on the mRNA. This solution was layered over a 1.1 M sucrose solution and centrifuged at 69,000 rpm in a TLA 100.3 rotor for 2 hours at 4 °C to pellet the IC. The sucrose solution was decanted and pellets resuspended in 200 µL of 1X polymix buffer. Samples were then ready for either *in vitro* translation or for surface immobilization for single molecule studies.

Single molecule studies were carried out as per previous protocols (Rinaldi, Blanco, Lund, and Walter, publication pending). The first variation was the inclusion of

a “surf-n-turf” blocking step after streptavidin functionalization of the slide surface, prior to mRNA binding. The second was the inclusion of a full PIC forming complex, as above in the IC formation, though lacking the 50S ribosome and raising Mg^{2+} concentration to 20 mM. For all conditions containing ligand, a near saturating concentration of 1.62 µM was utilized as a means of preserving our finite supply of $preQ_1$.

4.3 Results and Discussion

Initiation Factor Expression

In order to perform preinitiation complex/initiation complex formation assays, it was essential to have all other initiation components. Aside from the 30S/50S ribosome, initiation factors would be essential to study the initiation of ribosomes on various messenger RNAs. This was initially accomplished in collaboration with an undergraduate in our lab, Kimberly Haupt, and later with Paul Lund. We were able to purify each of the factors, typically several mL at 100-300 µM concentrations.

Ribosome labeling: DNA vs LNA

One of the major limitations of hybridizing a modified DNA strand to the ribosome was its low melting temperature and stability. Hoping to be able to label a greater proportion of 30S ribosomes, we turned to newly developed linked nucleic acid (LNA) technology. A product of Exiqon, LNAs involve a 2’-O, 4’-C linkage which locks a nucleotide sugar in the 3’-endo conformation, which is optimal for Watson Crick base pairing. Made entirely of LNA, a probe to the HP5 extension (SP5L) could be shortened while still increasing binding stability.

SP5L – 5’- ATCAGGATA-3’

This change in probes drastically increased the predicted melting temperature of the ribosome/probe binding. The original 14 nucleotide DNA probe was predicted to have a melting temperature of 55 °C, which while stable, could be improved upon. One of the great advantages of LNA labeling was the specificity of LNA for its complementary DNA or RNA. Utilizing the new 9 nucleotide LNA, it was predicted to bind at the much higher temperature of 93 °C. Ordered with 5’ and 3’ amine modifications, probes were fluorescently labeled with Cy5 as we had done previously (Section 3A).

In vitro Translation with Labeled Ribosomes

To test the viability of the purified ribosomes, *in vitro* translation assays were undertaken to test their ability to translate mRNA messages. In addition to labeled and unlabeled ribosomes, we utilized the well characterized gene *60* system to test not only the translation ability, but also the ability of the ribosome to translate structurally complex messages. *In vitro* translation was carried out as has been done previously in our lab (168). We began by using S30 extracts for translation as they had previously proved to be our most robust (if somewhat dirty) translation system. This reaction produced the expected gene products of the full length 18.4 kDa protein and the truncated 7.8 kDa product. Additionally, when we hybridized DNA oligos necessary for single molecule experiments to the gene *60* mRNA, we again saw translation at a similar output

to samples with probes. This was an unsurprising finding, as the probes are outside of the coding region of the mRNA and likely have little influence on the translating ribosome.

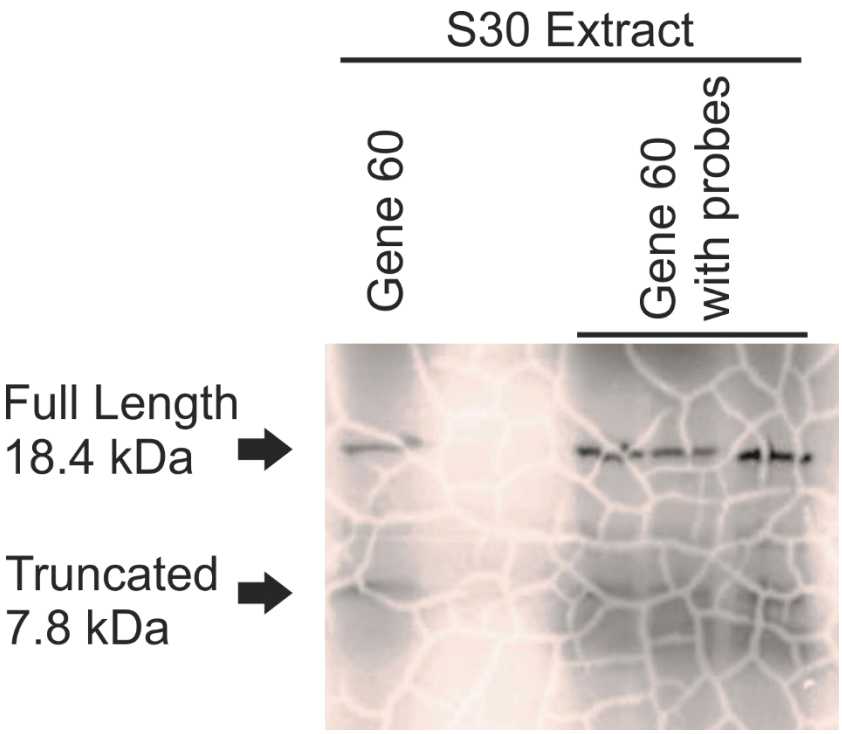


Figure 29: Translation of Gene *60* using S30 Cellular Extracts.

In vitro translation of gene *60* mRNA in the presence and absence of probes hybridized to the mRNA. Translation utilized an S30 extract as the source of biological translation materials. Translation of gene *60* produces the expected full length (18.4 kDa) and truncated (7.8 kDa) proteins. Translation efficiency with S30 extract is unaffected by the presence of labeling and immobilization DNA probes hybridized to gene *60*. Translation products were resolved on a 16% SDS-PAGE gel.

Given the robust translation seen in the S30 extract, we sought a more controllable system without the interference of endogenous mRNA, ribosomes, or RNases. While our lab had previously employed S100 extracts to remove endogenous ribosomes (168), it still left us with other factors we wanted to eliminate. This is when we turned to a commercial translation system, the PURExpress from New England

Biolabs (NEB). Based on the work of Shimizu, et. al, this system depended entirely on purified translation components and did not involve the use of cellular extracts (193). This would not only allow us to use our own ribosome (necessary for fluorescent labeling), but would eliminate any other biological contaminants. We tested this system on the gene *60* mRNA as we had done for the S30 extracts and found a similarly capable system. Both full length and truncated gene products were produced, indicating that we had a system that.

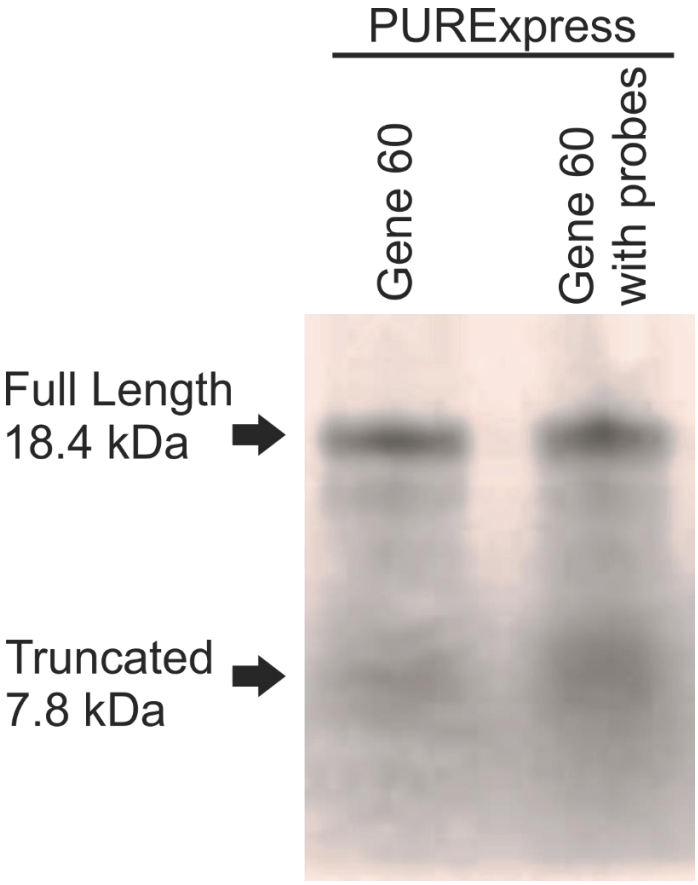


Figure 30: Translation of Gene *60* using PURExpress Translation Kit.

To allow for a more pure and controllable system, translations were carried out in PURExpress translation kit containing only purified translation components. In vitro translation of gene *60* mRNA in solutions contain ^{35}S methionine were carried out for 30 minutes at 37°C before resolving on an SDS-PAGE gel. Translation of gene *60* produces

the expected full length (18.4 kDa) and truncated (7.8 kDa) proteins. Translation efficiency in the PURExpress kit is unaffected by the presence of labeling and immobilization probes hybridized to gene *60*. Translation products were resolved on a 16% SDS-PAGE gel.

When S30 extracts were compared directly to the PURExpress, translation efficiency of the S30 extracts was evident. Drastically greater signal was seen in samples expressed by the S30 extract. This could have been due to the translation components being present at a much higher concentration in the cellular extracts.

Further, we tested the ability of preformed initiation complexes to produce translation products (**Figure 31**). While the translation products from preformed ICs were minimal, translation did occur after the IC was formed. This provided reassurance that we could lead surface immobilized ICs or possibly PICs into translation for later single molecule observation. By chasing ICs into translation, this type of assay could provide insight into how the ribosome processes through structured mRNAs on their first round of translation.

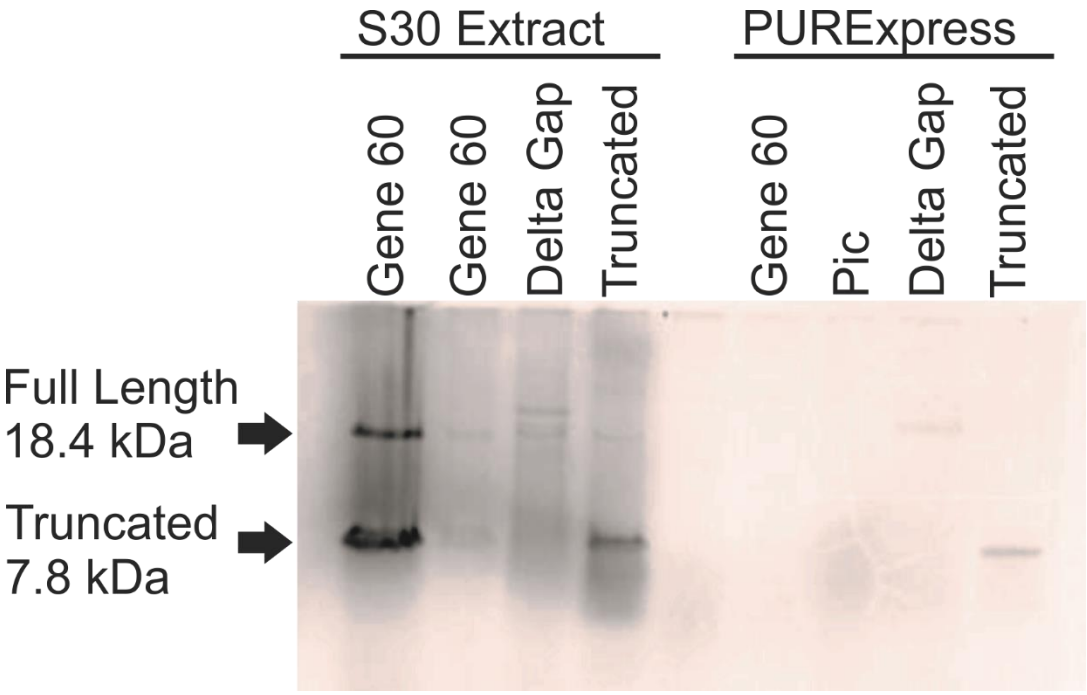


Figure 31: PURExpress Produces Reduced Translation Efficiency.

Comparing S30 extracts to PURExpress translation systems allowed the translation efficiency of the two systems to be compared. Translation was carried out in either S30 cellular extracts or with the PURExpress translation kit and in house purified ribosomes. Translation products were resolved on a 16% SDS-PAGE gel. Translation efficiency was much lower in the PURExpress system, but as this could easily be due to a much higher ribosome concentration in the S30 extract.

Single Molecule Experiments

With the single molecule tools we have assembled, we were able to begin studies of the initiation step of translation. We decided to expand on the work of Arlie Rinaldi with the nested 1564/1563 mRNA and Tte ribososwitch. Dr. Rinaldi had previously looked at the association of a fluorescently labeled RNA oligo matching the sequence of the aSD to the mRNA. Further, this was done in the presence and absence of $preQ_1$ to see the effect the riboswitch had on SD availability. As the $preQ_1$ ligand binds to the

riboswitch at low nM affinity and changes the conformation of the aptamer, with the SD binding sequence sequestered in the aptamer domain and unavailable for 30S ribosome binding(81, 191, 194). As such, we expected that in the presence of $preQ_1$, we would see far less colocalization of mRNA and ribosome signal, and that the signals we did see would be of a much shorter duration.

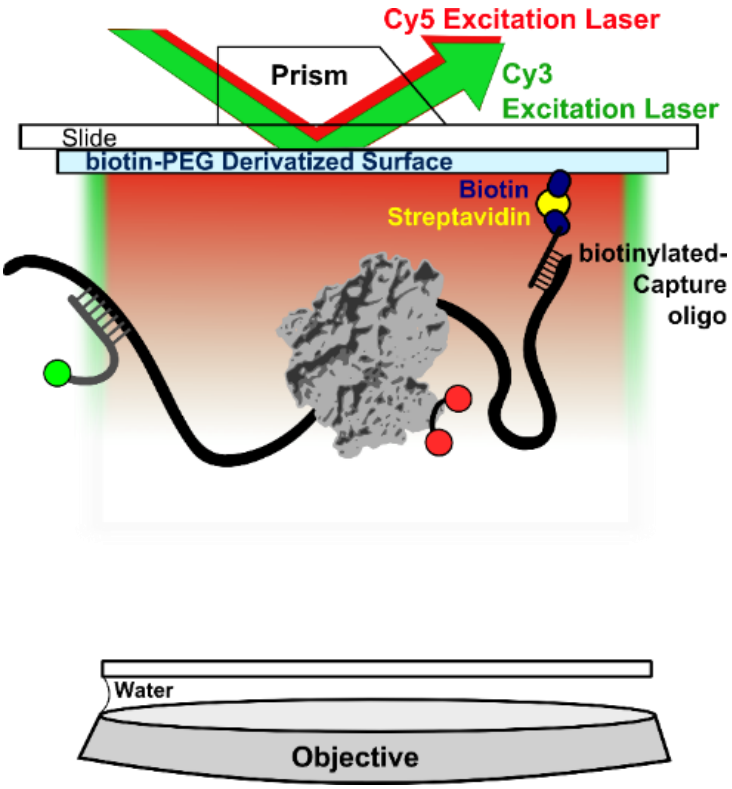


Figure 32: Schematic of Single Molecule Ribosome Initiation Assay

To perform single molecule studies of ribosome association with mRNA, the primary components must be fluorescently labeled and one must be surface immobilized. In our experiments, 1564/1563 mRNA (black) has a biotinylated DNA (blue) hybridized to the 5' end and Cy3 labeled LNA (green) to the 3' end of the 1564 gene and surface immobilized on a streptaviding (yellow). The 30S ribosome (grey) labeled with Cy5 (red) can then be flowed onto the slide and allowed to associate with the mRNA. To observe these components, constant excitation of Cy3 and Cy5 by red and green laser illumination is carried out.

Primarily, we wanted to see if we could perform similar association assays using fluorescently labeled 30S ribosome as to those we had performed with the mutant 30S ribosome (**Figure 32**). We Cy3 labeled and immobilized our mRNA by use of a biotinylated DNA probe to the 5' of the 1564/1563 gene and a Cy3 labeled LNA probe to the 3' end of the 1564 gene.

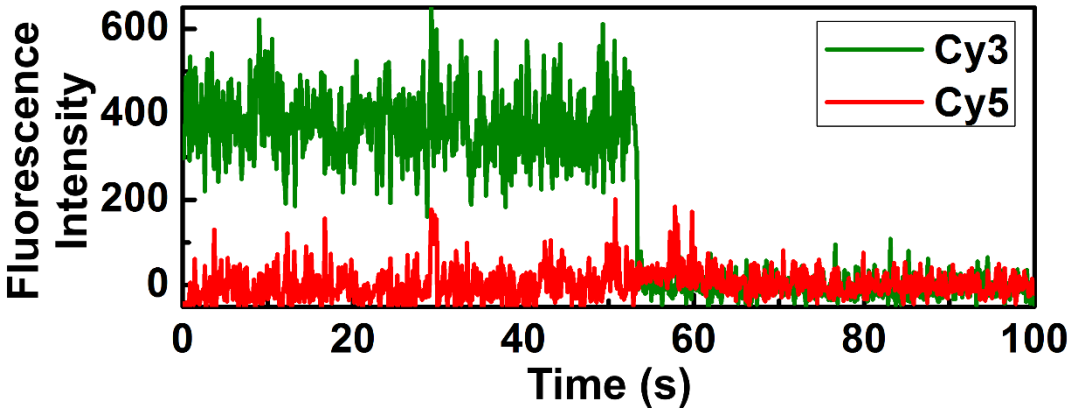


Figure 33: Immobilization of 1564/1563 mRNA-Cy3.

Single molecule experiments of mRNA with immobilization and labeling probes hybridized were surfaced immobilized to test labeling strategy. The above single molecule trace is of Cy3 labeled 1564/1563 mRNA. Samples are immobilized by the use of a biotin-DNA strand complementary to the 5' end of the 1564 gene. The Cy3 label is attached to an LNA strand complementary to the 3' end of the 1564 gene. This trace showed that our labeling scheme was a viable method for the basis of future smFRET experiments in which the 1564/1563 mRNA needed to be immobilized.

Looking at the mRNA, we were very commonly able to stable Cy3 signal for 1-2 minutes as the mRNA was immobilized on the surface. In samples of mRNA lacking a capture strand, we didn't see any long lived, stable Cy3 association with the surface. This is due in part to the stringent surface blocking protocol that we follow. Called "Surf-n-Turf", this involves the passivation of salmon sperm DNA ("surf") and bovine

serum albumin (BSA, "turf") to block non-specific binding. Before our biological samples were immobilized on the slide, BSA (0.1 mg/mL) and salmon sperm DNA (0.5 mg/mL) were incubated for 10 minutes to prevent any non-specific association of mRNA or ribosome.

After we were able to immobilize mRNA at an appropriate surface density to image individual molecules, we next turned to 30S initiation complex formation. For this, we flowed on samples of Cy5 labeled 30S ribosome in the presence of the components necessary for IC formation (IFs, fMet-tRNA^{fMet}, GTP) and OSS. These components were allowed to incubate on the mRNA coated slide for 5 minutes before observation. Under direct excitement by green (Cy3) and red (Cy5) laser, we were able to see numerous colocalization events (**Figure 34**). These colocalized signals were interpreted as stable association of 30S ribosome with mRNA. Interestingly, while simple colocalizations did occur (**Figure 34**A), evidence of FRET between Cy3 and Cy5 appeared even under direct excitation of both fluorophores (**Figure 34**B). This traces shows several instances of anti-correlated signal change between the Cy3 and Cy5. This suggests that, even though they are spatially separated by ~600 nucleotides in the primary structure, the ribosome binding site can potentially come very close to the 3' end of the 1564 gene were the Cy3 LNA probe is situated. This FRET behavior is further confirmed by excitation by the green laser only. If no FRET were occurring, we would only be able to see signal from the Cy3 emission. Instead, in some traces we are able to see both Cy3 and Cy5 signals, which is only possible if the fluorophores are close enough (<100Å) for FRET to occur (**Figure 34**C).

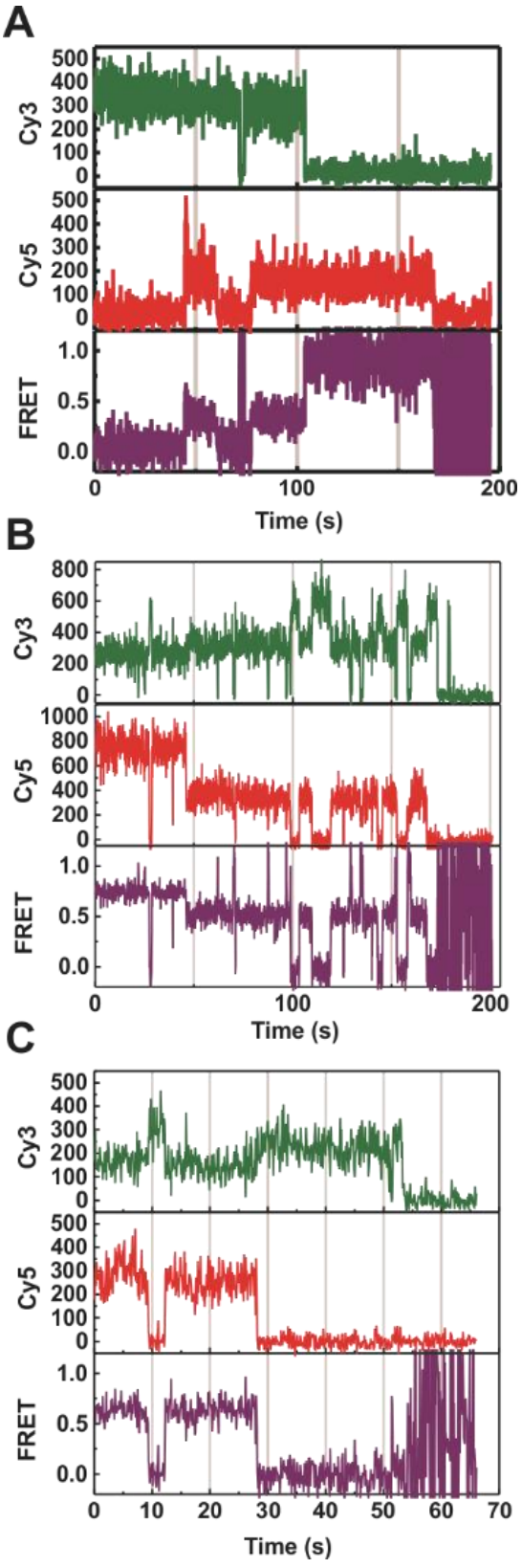

A
Cy3
Cy5
FRET
Time (s)
B
Cy3
Cy5
FRET
Time (s)
C
Cy3
Cy5
FRET
Time (s)

Figure 34: Colocalization Traces Reveal FRET Between Ribosome and mRNA.

Ribosome-Cy5 (red trace) is allowed to incubate on slide with immobilized mRNA-Cy3 (green trace). Samples were illuminated with red and green lasers in A and B but only green laser in C. (A) Immobilized mRNA displays transient interaction with ribosome before stable binding. (B) Colocalized ribosome-mRNA show FRET via anti-correlated signal change despite red laser excitation of Cy5. (C) Under only green laser excitation only, 30S ribosome-Cy5 signal caused by FRET with mRNA-Cy3 is observed, along with anti-correlated signals as Cy5 blinks (10s) and dies (28s). Despite the distal location in the primary structure, mRNA secondary structure likely brings the 3' and 5' ends of the 1564 gene into close proximity.

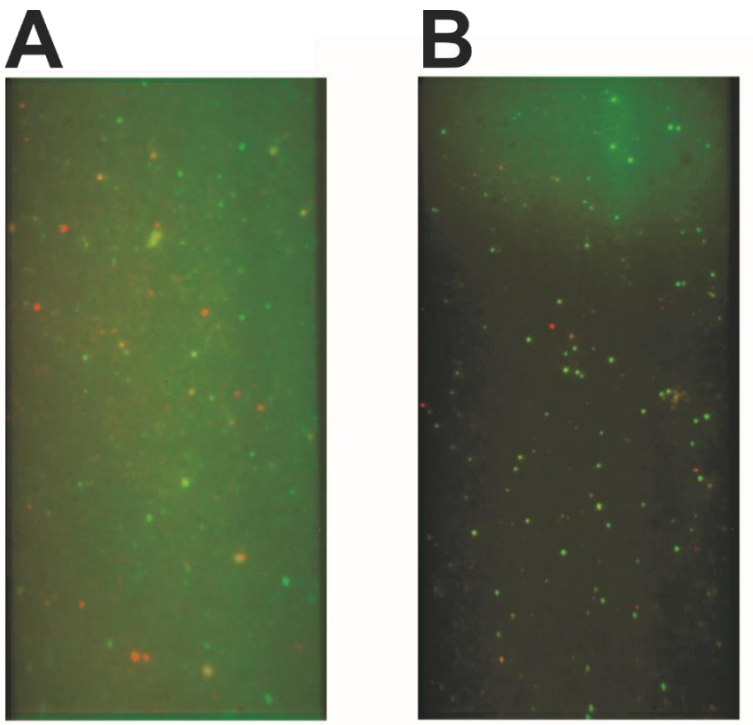


Figure 35: Field of View Overlay of Ribosome-mRNA Colocalization With (B) and Without (A) preQ$_1$ Ligand.

When ribosomes and mRNA are incubated together under the presence of absence of preQ$_1$ ligand, the time averaged fields of view for each signal can be overlayed. The Cy3 signal of mRNA is seen in green while Cy5 signal of the ribosome is seen in red. Colocalized signals of mRNA and ribosome are seen in orange/yellow. A much higher level of colocalization is seen in samples lacking the preQ$_1$ ligand, indicating that the ligand reduces this colocalization.

Once we knew that colocalization was occurring, we wanted to assay the effect of $preQ_1$ binding to the mRNA. Another sample was prepared as before, but this one included 1.62 μM $preQ_1$. At this concentration, the vast majority of riboswitches should be in their bound, SD occluded state (81). When looking at simple signal overlays of Cy3 and Cy5 density, a much greater proportion of colocalization is seen the sample lacking (**Figure 35**A) $preQ_1$ than those containing it. This effect was further quantified by simple counting of traces. For all recorded traces, those showing a single step Cy3 photobleach were counted as single mRNA molecules. From these, those traces that also showed a discreet 1 or 2 step photobleach of Cy5 signal were counted as colocalized. By dividing the latter by the former, I was able to determine the percentage of mRNA showing ribosome association under PIC forming conditions (**Figure 36**). This was also done for condition of 30S association with immobilized mRNA in the absence of other factors normally necessary for initiation complex formation. This was meant to provide a more direct assessment of SD accessibility by removing the other stabilizing factors. In both sets of experiments, the presence of ligand reduced the fraction of colocalization that occurred. Oddly, this difference was much more striking in the presence of IFs than in their absence.

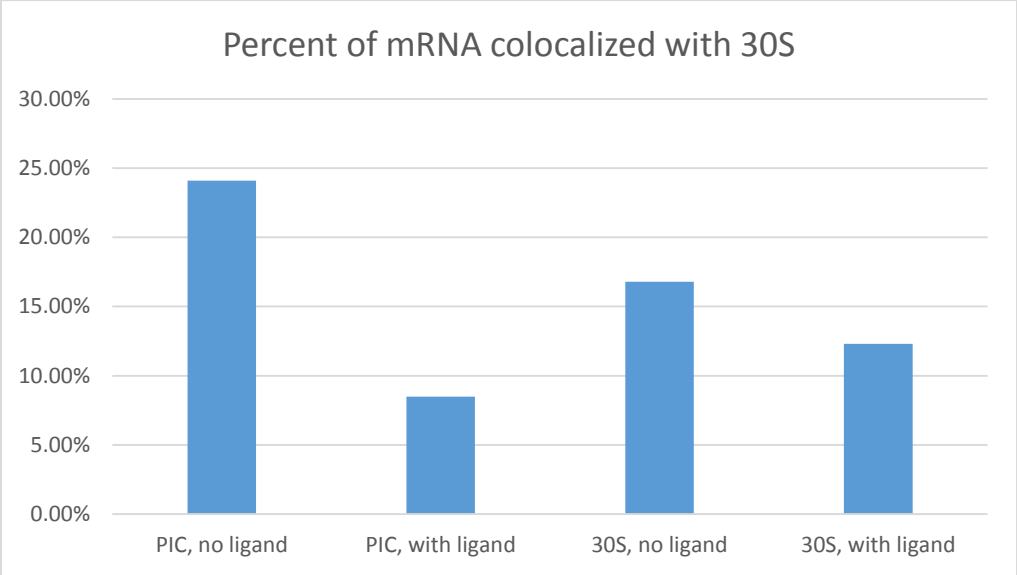


Figure 36: Percent of Single mRNA Molecules Colocalized with 30S Ribosome.

Counting the total number of Cy3 signals displaying a single photobleaching event, we then counted all those also showing a Cy5 signal. This Cy5 was indicative of a 30S ribosome colocalization. This was done for samples in the presence (ligand) or absence (no ligand)of preQ$_1$ ligand and with (PIC) or without (30S) the full complement of initiation factors and fMet-tRNA^{fMet}.

Taking this analysis a step further, we fit those traces showing colocalization to 2-state idealizations using QuB. We then integrated the total time the Cy5 ribosome spent in the high (bound) state for all traces. This total "on" time was then divided by the number of traces to provide an average time of colocalization/trace (**Figure 37**).

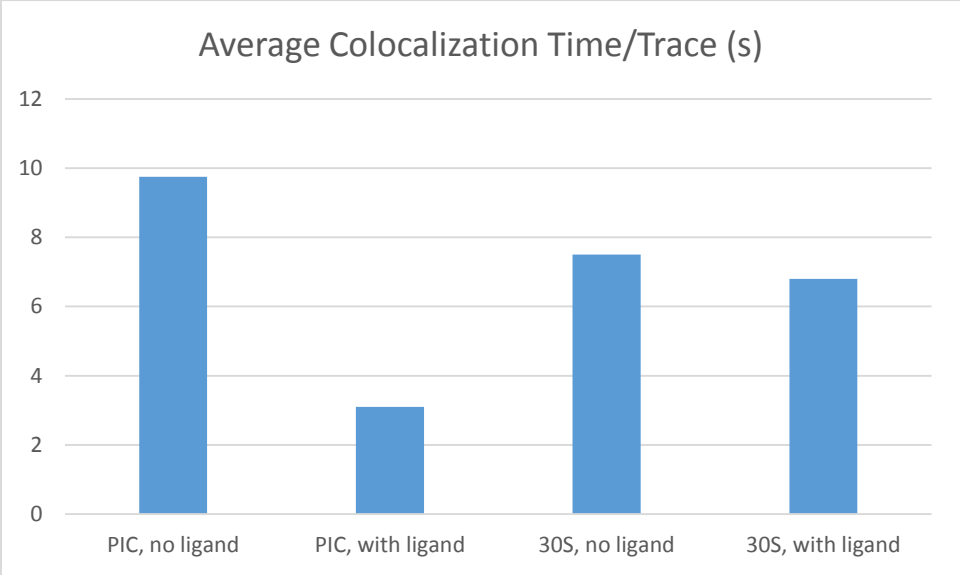


Figure 37: Average Total Ribosome Dwell Time on mRNA/Trace.

Idealized traces were integrated to find total time of the high signal from Cy5. Total time in the high state for all traces was divided by the number of traces to provide a total "on" time for the ribosome on the mRNA.

In all conditions, those lacking ligand showed the longest association times, 9.75s for the PIC and 7.5s for the 30S only conditions. While these range from ~3-fold to almost insignificantly higher than their ligand containing counterparts (3.1s for the PIC, 6.8s for the 30S only), actual association time could be much longer. Photobleaching of the fluorophores was indistinguishable from dissociation, leaving this as an area that needs to be improved upon before further experiments should be undertaken. Also somewhat surprising was the minimal effect of the $preQ_1$ on changing the binding behaviors of the 30S ribosome in the absence of other factors. However, despite this limitation, the general trend of ligand limiting 30S ribosome association with mRNA was followed.

fMet-tRNA Labeling

In an effort to look at an alternative method of labeling components of the PIC materials, we decided to label the fMet-tRNA prior to charging it with methionine. Using a previously published protocol, we performed a Cy5 lableing of this tRNA (195). Simply, fMet-tRNA was reduced by incubation with $NaBH_4$ in 40 mM Tris, pH 7.5 on ice for an hour. The tRNA was then ethanol precipitated to remove excess $NaBH_4$. Labeling was then carried out by dissolving dried tRNA in 0.1M sodium formate and adding Cy5-hydrazide in DMSO. After incubating at 37 °C for 2 hours, the reaction was diluted into 400 uL of 50 mM sodium acetate, pH 6.0. Three rounds of phenol-chloroform extractions were carried out to remove excess dye. Samples were then ethanol precipitated, resuspended in TE buffer, and purified using reverse phase chromatography over a C18 column. Labeled and unlabeled fractions were collected, pooled, and again ethanol precipitated. While labeling efficiency was low (~5%), this was expected and unlabeled material could be subsequently relabeled.

Labeling this translation component allows us to look at multiple different assembly events. First, if we were to label the mRNA with Cy3 and the fMet-$tRNA^{fMet}$ with Cy5, we would be able to do a series of experiments similar to what we have already done with labeled mRNA and ribosome observing assembly. Under these conditions, we would observe a constant immobilized signal from the mRNA-Cy3 and watch for fMet-$tRNA^{fMet}$-Cy5 association. Using this system, we would be able to ascertain the average time to tRNA association and whether or not such association is transient or stable. Comparing the times to those we find with 30S ribosome association could allow us to

determine the order of assembly in translation initiation. Alternately, we could also label the fMet-tRNA^{fMet} and the 30S ribosome simultaneously and observe their stable association with an mRNA immobilized on a slide surface. While proper controls for immobilization in the absence of mRNA must be run, this would provide the most direct approach to the order of assembly, if an order exists.

Our work developing this translation initiation system has allowed us to begin characterizing the interaction of ribosomes with a riboswitch. While preliminary, the reduced interactions of 30S ribosome with a riboswitch in the presence of it ligand begins to explain the reduced translation levels seen in vivo. Further, this system make possible future studies of ribosome/mRNA interactions on the single molecule level, allowing the effects of any number of translation components to be observed.

CHAPTER 5: SUMMARIES, CONCLUSIONS, FUTURE DIRECTIONS

Heterogeneity exists is biological systems. This much is apparent from our studies of the hairpin ribozyme and ribosome/riboswitch interactions. Some of these heterogeneities artifacts of how we study these systems, introduced errors and mistakes we incur. But nature introduces many of these to biological systems for a variety of reasons. Heterogeneity introduces the ability of an RNA or protein to modulate behaviors to meet the demand of the environment. In the case of ribozymes, this is most simply seen in life cycle of the hairpin ribozyme. During rolling circle replication, the hairpin must be catalytically active to promote the cleavage of multimers into monomers. Once individual satellite viruses are formed, the hairpin must act to catalyze the ligation of circular satellites. As there are formed, the hairpin must assume another role as the circles must remain intact for the next round of rolling circle replication. This multitude of roles is aided by the inherent heterogeneous nature of the hairpin ribozyme. These variations of structure, sequence, and behavior serve to codify our continued research of heterogeneities in biological systems.

These heterogeneous biological system may exist for any multitude of reasons. Variations as simple as point mutations can have a drastic effect on the resulting molecules in both proteins and RNA. While numerous mutations can be silent or easily compensated for in either type of biomolecule, point mutations at active sites or inter- or

intramolecular interaction sites have the greatest impact. We have studied here a system in the hairpin ribozyme where a single point mutation of G8A incurs a drastic reduction in the cleavage rate of the enzyme. Located at the active site in the docked hairpin ribozyme, mutating the natural guanosine into an adenosine removes its ability to help catalyze the self-cleavage of the ribozyme. (196-198) While mutations in structural elements can be overcome by non-Watson-Crick base pairings, the specificity of a nucleobase, or amino acid in a protein, is necessary for normal catalytic function.

While these sorts of mutations are thought of as "deleterious" because of their inhibition of measurable functions, their actual implications are much are complicated. It is unfortunate that such a negative connotation to the idea of heterogeneities, but in truth, they offer advantages to natural systems. The problem we must face is that the cell does not always need to be in the "most active" state. The variations in short or long term environmental changes can place stresses on the cell that require the modulation of function in a timeframe much too rapid for evolution of a system. In these scenarios, the inherent heterogeneities of macromolecular machinery is essential to the continued survival of a cell. By being able to enrich for a certain behavior or variant structure, the cell is able to utilize heterogeneity to overcome stresses. Our further study and understanding of these systems and how heterogeneities are harnessed as advantages and not as "deleterious" abnormalities, will allow for the greater understanding of the natural world and challenge our notions of "native" forms of enzymes and ribozymes.

Future Directions

While order of assembly for 30S ribosome and fMet-tRNA^{fMet} are easily within our reach, this could very easily be extended to the order of assembly for 30S and the initiation factors, particularly IF1 and IF3. These initiation factors are responsible for stabilizing the binding of the 30S subunit to mRNA in the PIC. With different pairs of materials (IF1/30S, IF3/30S, mRNA/IF1, mRNA/IF3, etc) fluorescently labeled, we could look for order of these associations. Kinetically, factor association with the 30S is well understood and seems to be an ordered process (199).

Additionally, if we were to label the 50S ribosome, we would be able to detect order of departure of the initiation factors relative to the final IC formation. This labeling could be accomplished by use of the L9 labeling scheme utilized by the Noller and Ermolenko labs (77, 78). This would involve expressing and purifying the clones for the L9 protein we already have in our possession. We could then fluorescently label it and incorporate it into the 50S ribosome. Alternately, we could utilize the rRNA modification scheme similar to what we used for the 16S h33a labeling. In the 50S ribosome, helix 101 was identified as a viable platform for semi-stable helical extension which can be expressed, purified, and labeled in the same manner our work has been done to this point (180). Additionally, Paul Lund of our lab has been working to create single cysteine mutants of all three initiation factors. Once expressed, these will be able to be fluorescently labeled and utilized in studies of PIC/IC formation. Whether we are looking at assembly or departure order, this will soon become a full tool box for examining IC formation.

Of particular importance, long observation times of both the ribosome and mRNA are necessary for extended observations of translation events. To be able to clearly distinguish between ribosome dissociation and fluorophore bleaching. Often in experiments thus far, we have been limited in observation windows of 2-3 minutes before fluorophores bleached. Current methods of extending lifetimes are based on reducing oxygen levels and include simple degassing of buffers and use of a three part oxygen scavenging system (OSS). Our OSS is composed of protocatechuic acid (PCA), protocatechuate-3,4-dioxygenase (PCD), and 6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid (Trolox) (200, 201). PCD catalyzes the conversion of PCA into *β*-carboxy-*cis, cis*-muconic acid and utilizes O_2 in the process. Trolox is an antioxidant derivative of vitamin E which further removes dissolved oxygen and prevents photobleaching. Utilized together, they form the basis of our OSS system.

While our standard OSS does an excellent job of extending fluorescent lifetimes to minutes, its components can be utilized in alternate ways to improve lifetimes. Recent work from Scott Blanchard's lab has developed a method of bonding Trolox, along with cyclooctatetraene (COT) and 3-nitrobenzyl alcohol (NBA), directly to cyanine dyes(202). By coupling fluorophores with antioxidants, greatly increased lifetimes to as much as 8-fold are seen in *in vitro* studies, along with reduced blinking of fluorophores. Using the primary amine modified LNAs directed to the ribosome we already have in our possession, we can label these in order to modify them with these long lived dyes. Once labeled, we can hybridize these to the h33a modified 30S ribosomal subunits to create ribosomes with extended lifetimes. This would allow us to more directly observe the

dynamics of ribosome assembly and translation processes. With ribosome doubly labeled with stable fluorophores, distinguishing between ribosome dissociation (1 step signal loss) and long lived association (2 step photobleach) would be possible over tens of minutes.

For even longer lived conditions, we can use time skip movie observations. Previously used by members of the lab to study splicing at the single molecule level, skip movies involve the alternating short excitation and capture with long dark periods. While this method is incapable of observing fast transitions, associations or dissociations that take place over tens of minutes could easily be captured (unpublished data).

As our system becomes better understood, the ability to observe additional components of initiation and translation will allow us a more complete picture of the biological process. Of the other components that we could label, the initiation factors represent convenient and mutable options.

REFERENCES

1. Crick F (1970) Central Dogma of Molecular Biology. *Nature* 227(5258):561-&.
2. Cech TR, Zaug AJ, & Grabowski PJ (1981) Invitro Splicing of the Ribosomal-Rna Precursor of Tetrahymena - Involvement of a Guanosine Nucleotide in the Excision of the Intervening Sequence. *Cell* 27(3):487-496.
3. Kruger K*, et al.* (1982) Self-Splicing Rna - Auto-Excision and Auto-Cyclization of the Ribosomal-Rna Intervening Sequence of Tetrahymena. *Cell* 31(1):147-157.
4. Guerrier-Takada C, Gardiner K, Marsh T, Pace N, & Altman S (1983) The Rna Moiety of Ribonuclease-P Is the Catalytic Subunit of the Enzyme. *Cell* 35(3):849-857.
5. Amaral PP, Dinger ME, Mercer TR, & Mattick JS (2008) The eukaryotic genome as an RNA machine. *Science* 319(5871):1787-1789.
6. Walter NG, Woodson SA, & Batey RT (2009) *Non-protein coding RNAs* (Springer, Heidelberg).
7. Moazed D (2009) Small RNAs in transcriptional gene silencing and genome defence. *Nature* 457(7228):413-420.
8. Mercer TR, Dinger ME, & Mattick JS (2009) Long non-coding RNAs: insights into functions. *Nature reviews. Genetics* 10(3):155-159.
9. Mello CC & Conte D (2004) Revealing the world of RNA interference. *Nature* 431(7006):338-342.
10. Winkler WC, Nahvi A, Roth A, Collins JA, & Breaker RR (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* 428(6980):281-286.
11. Toor N, Rajashankar K, Keating KS, & Pyle AM (2008) Structural basis for exon recognition by a group II intron. *Nat. Struct. Mol. Biol.* 15(11):1221-1222.
12. Nissen P, Hansen J, Ban N, Moore PB, & Steitz TA (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science* 289(5481):920-930.
13. Prody GA, Bakos JT, Buzayan JM, Schneider IR, & Bruening G (1986) Autolytic Processing of Dimeric Plant-Virus Satellite Rna. *Science* 231(4745):1577-1580.
14. Forster AC & Symons RH (1987) Self-Cleavage of Virusoid Rna Is Performed by the Proposed 55-Nucleotide Active-Site. *Cell* 50(1):9-16.
15. Hampel A & Tritz R (1989) Rna Catalytic Properties of the Minimum (-)Strsv Sequence. *Biochemistry* 28(12):4929-4933.
16. Saville BJ & Collins RA (1990) A Site-Specific Self-Cleavage Reaction Performed by a Novel Rna in Neurospora Mitochondria. *Cell* 61(4):685-696.

17. Lai MMC (1995) The Molecular-Biology of Hepatitis-Delta Virus. *Annu Rev Biochem* 64:259-286.
18. Salehi-Ashtiani K, Luptak A, Litovchick A, & Szostak JW (2006) A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science* 313(5794):1788-1792.
19. Martick M, Horan LH, Noller HF, & Scott WG (2008) A discontinuous hammerhead ribozyme embedded in a mammalian messenger RNA. *Nature* 454(7206):899-U857.
20. Webb CH, Riccitelli NJ, Ruminski DJ, & Luptak A (2009) Widespread occurrence of self-cleaving ribozymes. *Science* 326(5955):953.
21. de la Pena M & Garcia-Robles I (2010) Ubiquitous presence of the hammerhead ribozyme motif along the tree of life. *Rna* 16(10):1943-1950.
22. Illangasekare M & Yarus M (1999) Specific, rapid synthesis of Phe-RNA by RNA. *P Natl Acad Sci USA* 96(10):5470-5475.
23. Lincoln TA & Joyce GF (2009) Self-Sustained Replication of an RNA Enzyme. *Science* 323(5918):1229-1232.
24. Agresti JJ, Kelly BT, Jaschke A, & Griffiths AD (2005) Selection of ribozymes that catalyse multiple-turnover Diels-Alder cycloadditions by using in vitro compartmentalization. *P Natl Acad Sci USA* 102(45):16170-16175.
25. Scott WG (2007) Ribozymes. *Current opinion in structural biology* 17(3):280-286.
26. Serganov A & Patel DJ (2007) Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nature reviews* 8(10):776-790.
27. Korostelev A & Noller HF (2007) The ribosome in focus: new structures bring new insights. *Trends Biochem. Sci.* 32(9):434-441.
28. Walter NG (2007) Ribozyme catalysis revisited: is water involved? *Molecular cell* 28(6):923-929.
29. Al-Hashimi HM & Walter NG (2008) RNA dynamics: it is about time. *Current opinion in structural biology* 18(3):321-329.
30. Cochrane JC & Strobel SA (2008) Catalytic strategies of self-cleaving ribozymes. *Acc. Chem. Res.* 41(8):1027-1035.
31. Walter NG & Perumal S (2009) The small ribozymes: Common and diverse features observed through the FRET lens. In: Non-Protein Coding RNAs. *Non-Protein Coding RNAs*, eds Walter NG, Woodson SA, & Batey RT (Springer, Berlin), Vol 13, pp 103-127.
32. Johnson-Buck AE, McDowell SE, & Walter NG (2011) Metal ions: supporting actors in the playbook of small ribozymes. *Metal ions in life sciences* 9:175-196.
33. Gartland WJ & Sueoka N (1966) 2 Interconvertible Forms of Tryptophanyl Srna in E Coli. *P Natl Acad Sci USA* 55(4):948-&.
34. Lindahl T, Adams A, & Fresco JR (1966) Renaturation of Transfer Ribonucleic Acids through Site Binding of Magnesium. *P Natl Acad Sci USA* 55(4):941-&.
35. Walstrum SA & Uhlenbeck OC (1990) The Self-Splicing Rna of Tetrahymena Is Trapped in a Less Active Conformation by Gel Purification. *Biochemistry* 29(46):10573-10576.
36. Esteban JA, Banerjee AR, & Burke JM (1997) Kinetic mechanism of the hairpin ribozyme - Identification and characterization of two nonexchangeable conformations. *J. Biol. Chem.* 272(21):13629-13639.

37. Pan T & Sosnick TR (1997) Intermediates and kinetic traps in the folding of a large ribozyme revealed by circular dichroism and UV absorbance spectroscopies and catalytic activity. *Nat. Struct. Biol.* 4(11):931-938.
38. Chadalavada DM, Knudsen SM, Nakano S, & Bevilacqua PC (2000) A role for upstream RNA structure in facilitating the catalytic fold of the genomic hepatitis delta virus ribozyme. *J. Mol. Biol.* 301(2):349-367.
39. Treiber DK & Williamson JR (1999) Exposing the kinetic traps in RNA folding. *Current opinion in structural biology* 9(3):339-345.
40. Russell R (2008) RNA misfolding and the action of chaperones. *Front Biosci* 13:1-20.
41. Solomatin SV, Greenfeld M, Chu S, & Herschlag D (2010) Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature* 463(7281):681-684.
42. Guo F, Gooding AR, & Cech TR (2004) Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site. *Molecular cell* 16(3):351-362.
43. Sigler PB (1975) Analysis of Structure of Transfer-Rna. *Annu Rev Biophys Bio* 4:477-527.
44. Rastogi T, Beattie TL, Olive JE, & Collins RA (1996) A long-range pseudoknot is required for activity of the Neurospora VS ribozyme. *Embo J* 15(11):2820-2825.
45. Cate JH*, et al.* (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* 273(5282):1678-1685.
46. Ferre-D'Amare AR, Zhou KH, & Doudna JA (1998) Crystal structure of a hepatitis delta virus ribozyme. *Nature* 395(6702):567-574.
47. Rupert PB & Ferre-D'Amare AR (2001) Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature* 410(6830):780-786.
48. Klein DJ & Ferre-D'Amare AR (2006) Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science* 313(5794):1752-1756.
49. Lipfert J, Ouellet J, Norman DG, Doniach S, & Lilley DM (2008) The complete VS ribozyme in solution studied by small-angle X-ray scattering. *Structure* 16(9):1357-1367.
50. Murray LJW, Arendall WB, Richardson DC, & Richardson JS (2003) RNA backbone is rotameric. *P Natl Acad Sci USA* 100(24):13904-13909.
51. Richardson JS*, et al.* (2008) RNA backbone: Consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *Rna* 14(3):465-481.
52. Brion P & Westhof E (1997) Hierarchy and dynamics of RNA folding. *Annu Rev Bioph Biom* 26:113-137.
53. Porschke D & Eigen M (1971) Co-operative non-enzymic base recognition. 3. Kinetics of the helix-coil transition of the oligoribouridylic--oligoriboadenylic acid system and of oligoriboadenylic acid alone at acidic pH. *J. Mol. Biol.* 62(2):361-381.
54. Turner DH, Sugimoto N, & Freier SM (1990) Thermodynamics and Kinetics of Base-Pairing and of DNA and RNA Self-Assembly and Helix-Coil Transition. *Nucleic Acids,* Landolt-Bornstein, ed Saenger W (Springer, Berlin), pp 201-227.
55. Herschlag D (1995) RNA Chaperones and the RNA Folding Problem. *J. Biol. Chem.* 270(36):20871-20874.
56. Zhuang XW*, et al.* (2000) A single-molecule study of RNA catalysis and folding. *Science* 288(5473):2048-+.
57. Zhuang XW*, et al.* (2002) Correlating structural dynamics and function in single ribozyme molecules. *Science* 296(5572):1473-1476.

58. Pereira MJB*, et al.* (2008) Single VS ribozyme molecules reveal dynamic and hierarchical folding toward catalysis. *J. Mol. Biol.* 382(2):496-509.
59. Gralla J & Delisi C (1974) Messenger-Rna Is Expected to Form Stable Secondary Structures. *Nature* 248(5446):330-332.
60. Zuker M (1989) On Finding All Suboptimal Foldings of an Rna Molecule. *Science* 244(4900):48-52.
61. Mccaskill JS (1990) The Equilibrium Partition-Function and Base Pair Binding Probabilities for Rna Secondary Structure. *Biopolymers* 29(6-7):1105-1119.
62. Mathews DH (2004) Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *Rna* 10(8):1178-1190.
63. Markham NR & Zuker M (2005) DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res.* 33:W577-W581.
64. Korennykh AV, Plantinga MJ, Correll CC, & Piccirilli JA (2007) Linkage between substrate recognition and catalysis during cleavage of sarcin/ricin loop RNA by restrictocin. *Biochemistry* 46(44):12744-12756.
65. Huang Z*, et al.* (2009) One RNA aptamer sequence, two structures: a collaborating pair that inhibits AMPA receptors. *Nucleic Acids Res.* 37(12):4022-4032.
66. Altuntop ME, Ly CT, & Wang YH (2010) Single-Molecule Study of Ribosome Hierarchic Dynamics at the Peptidyl Transferase Center. *Biophys J* 99(9):3002-3009.
67. Ditzler MA, Rueda D, Mo JJ, Hakansson K, & Walter NG (2008) A rugged free energy landscape separates multiple functional RNA folds throughout denaturation. *Nucleic Acids Res.* 36(22):7088-7099.
68. Buzayan JM, Gerlach WL, & Bruening G (1986) Satellite tobacco ringspot virus RNA: A subset of the RNA sequence is sufficient for autolytic processing. *P Natl Acad Sci USA* 83(23):8859-8862.
69. Buzayan JM, Gerlach WL, Bruening G, Keese P, & Gould AR (1986) Nucleotide sequence of satellite tobacco ringspot virus RNA and its relationship to multimeric forms. *Virology* 151(2):186-199.
70. Rueda D*, et al.* (2004) Single-molecule enzymology of RNA: essential functional groups impact catalysis from a distance. *P Natl Acad Sci USA* 101(27):10066-10071.
71. Walter NG, Huang CY, Manzo AJ, & Sobhy MA (2008) Do-it-yourself guide: how to use the modern single-molecule toolkit. *Nat. Methods* 5(6):475-489.
72. Liu S, Bokinsky G, Walter NG, & Zhuang X (2007) Dissecting the multistep reaction pathway of an RNA enzyme by single-molecule kinetic "fingerprinting". *P Natl Acad Sci USA* 104(31):12634-12639.
73. Okumus B, Wilson TJ, Lilley DMJ, & Ha T (2004) Vesicle encapsulation studies reveal that single molecule ribozyme heterogeneities are intrinsic. *Biophys J* 87(4):2798-2806.
74. Bokinsky G*, et al.* (2003) Single-molecule transition-state analysis of RNA folding. *P Natl Acad Sci USA* 100(16):9302-9307.
75. Mortimer SA & Weeks KM (2009) C2'-endo nucleotides as molecular timers suggested by the folding of an RNA domain. *Proceedings of the National Academy of Sciences of the United States of America* 106(37):15622-15627.
76. Huang Z, Pei W, Jayaseelan S, Shi H, & Niu L (2007) RNA aptamers selected against the GluR2 glutamate receptor channel. *Biochemistry* 46(44):12648-12655.

77. Ermolenko DN, *et al.* (2007) Observation of intersubunit movement of the ribosome in solution using FRET. *J. Mol. Biol.* 370(3):530-540.
78. Ermolenko DN, *et al.* (2007) The antibiotic viomycin traps the ribosome in an intermediate state of translocation. *Nat. Struct. Mol. Biol.* 14(6):493-497.
79. Munro JB, Altman RB, O'Connor N, & Blanchard SC (2007) Identification of two distinct hybrid state intermediates on the ribosome. *Molecular cell* 25(4):505-517.
80. Cornish PV, Ermolenko DN, Noller HF, & Ha T (2008) Spontaneous intersubunit rotation in single ribosomes. *Molecular cell* 30(5):578-588.
81. Suddala KC, *et al.* (2013) Single transcriptional and translational preQ1 riboswitches adopt similar pre-folded ensembles that follow distinct folding pathways into the same ligand-bound structure. *Nucleic acids research* 41(22):10462-10475.
82. Baldwin RL & Rose GD (1999) Is protein folding hierarchic? I. Local structure and peptide folding. *Trends Biochem. Sci.* 24(1):26-33.
83. Dill KA, Ozkan SB, Shell MS, & Weikl TR (2008) The protein folding problem. *Annu. Rev. Biophys.* 37:289-316.
84. Dalal S & Regan L (2000) Understanding the sequence determinants of conformational switching using protein design. *Protein Sci.* 9(9):1651-1659.
85. Matthews BW (1995) Studies on protein stability with T4 lysozyme. *Adv. Protein Chem.* 46:249-278.
86. Freier SM, *et al.* (1986) Improved free-energy parameters for predictions of RNA duplex stability. *P Natl Acad Sci USA* 83(24):9373-9377.
87. Markham NR & Zuker M (2008) UNAFold: software for nucleic acid folding and hybridization. *Methods Mol. Biol.* 453:3-31.
88. Minor DL, Jr. & Kim PS (1994) Context is a major determinant of beta-sheet propensity. *Nature* 371(6494):264-267.
89. Kiefhaber T (1995) Kinetic traps in lysozyme folding. *P Natl Acad Sci USA* 92(20):9029-9033.
90. Kamagata K, Sawano Y, Tanokura M, & Kuwajima K (2003) Multiple parallel-pathway folding of proline-free Staphylococcal nuclease. *J. Mol. Biol.* 332(5):1143-1153.
91. Jennings PA, Finn BE, Jones BE, & Matthews CR (1993) A reexamination of the folding mechanism of dihydrofolate reductase from Escherichia coli: verification and refinement of a four-channel model. *Biochemistry* 32(14):3783-3789.
92. Schmid FX & Blaschek H (1981) A native-like intermediate on the ribonuclease A folding pathway. 2. Comparison of its properties to native ribonuclease A. *Eur. J. Biochem.* 114(1):111-117.
93. Flomenbom O, *et al.* (2005) Stretched exponential decay and correlations in the catalytic activity of fluctuating single lipase molecules. *P Natl Acad Sci USA* 102(7):2368-2372.
94. Lu HP, Xun LY, & Xie XS (1998) Single-molecule enzymatic dynamics. *Science* 282(5395):1877-1882.
95. Yang H, *et al.* (2003) Protein conformational dynamics probed by single-molecule electron transfer. *Science* 302(5643):262-266.
96. Edman L, Foldes-Papp Z, Wennmalm S, & Rigler R (1999) The fluctuating enzyme: a single molecule approach. *Chem Phys* 247(1):11-22.
97. van Oijen AM, *et al.* (2003) Single-molecule kinetics of lambda exonuclease reveal base dependence and dynamic disorder. *Science* 301(5637):1235-1238.

98. Xue Q & Yeung ES (1995) Differences in the chemical reactivity of individual molecules of an enzyme. *Nature* 373(6516):681-683.
99. Tan WH & Yeung ES (1997) Monitoring the reactions of single enzyme molecules and single metal ions. *Anal Chem* 69(20):4242-4248.
100. Craig DB, Arriaga EA, Wong JCY, Lu H, & Dovichi NJ (1996) Studies on single alkaline phosphatase molecules: Reaction rate and activation energy of a reaction catalyzed by a single molecule and the effect of thermal denaturation - The death of an enzyme. *J Am Chem Soc* 118(22):5245-5253.
101. Craig DB & Dovichi NJ (1998) Escherichia coli beta-galactosidase is heterogeneous with respect to the activity of individual molecules. *Can J Chem* 76(5):623-626.
102. Handa N, Bianco PR, Baskin RJ, & Kowalczykowski SC (2005) Direct visualization of RecBCD movement reveals cotranslocation of the RecD motor after chi recognition. *Molecular cell* 17(5):745-750.
103. Bianco PR & Kowalczykowski SC (2000) Translocation step size and mechanism of the RecBC DNA helicase. *Nature* 405(6784):368-372.
104. Baldini G, Cannone F, & Chirico G (2005) Pre-unfolding resonant oscillations of single green fluorescent protein molecules. *Science* 309(5737):1096-1100.
105. Cannone F*, et al.* (2005) Tracking unfolding and refolding of single GFPmut2 molecules. *Biophys J* 89(3):2033-2045.
106. Baldini G*, et al.* (2007) Evidence of discrete substates and unfolding pathways in green fluorescent protein. *Biophys J* 92(5):1724-1731.
107. Polakowski R, Craig DB, Skelley A, & Dovichi NJ (2000) Single molecules of highly purified bacterial alkaline phosphatase have identical activity. *J Am Chem Soc* 122(20):4853-4855.
108. Reed DH & Frankham R (2003) Correlation between fitness and genetic diversity. *Conserv Biol* 17(1):230-237.
109. Franklin I ed (1980) *Conservation Biology: an evolutionary-ecological perspective* (Sinauer Associates, Sunderland, Massachusetts), pp 135-150.
110. Allison AC (1957) Malaria in carriers of the sickle-cell trait and in newborn children. *Experimental parasitology* 6(4):418-447.
111. Dean M*, et al.* (1996) Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science* 273(5283):1856-1862.
112. Robertson DL, Hahn BH, & Sharp PM (1995) Recombination in AIDS viruses. *Journal of molecular evolution* 40(3):249-259.
113. Uyenoyama MK (1986) *Pesticide Resistance: Strategies and Tactics for Management* (National Academy Press, Washington, D. C.).
114. Lyon BR & Skurray R (1987) Antimicrobial resistance of Staphylococcus aureus: genetic basis. *Microbiological reviews* 51(1):88-134.
115. Wolffe AP & Matzke MA (1999) Epigenetics: regulation through repression. *Science* 286(5439):481-486.
116. Hannon GJ (2002) RNA interference. *Nature* 418(6894):244-251.
117. Grimson A*, et al.* (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* 27(1):91-105.

118. Han J*, et al.* (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* 125(5):887-901.
119. Mandal M & Breaker RR (2004) Gene regulation by riboswitches. *Nature reviews. Molecular cell biology* 5(6):451-463.
120. Nahvi A*, et al.* (2002) Genetic control by a metabolite binding mRNA. *Chemistry & biology* 9(9):1043.
121. Mironov AS*, et al.* (2002) Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell* 111(5):747-756.
122. Winkler W, Nahvi A, & Breaker RR (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* 419(6910):952-956.
123. Winkler WC, Cohen-Chalamish S, & Breaker RR (2002) An mRNA structure that controls gene expression by binding FMN. *Proceedings of the National Academy of Sciences of the United States of America* 99(25):15908-15913.
124. Blencowe BJ (2006) Alternative splicing: new insights from global analyses. *Cell* 126(1):37-47.
125. Graveley BR (2001) Alternative splicing: increasing diversity in the proteomic world. *Trends in genetics : TIG* 17(2):100-107.
126. Cai L, Friedman N, & Xie XS (2006) Stochastic protein expression in individual cells at the single molecule level. *Nature* 440(7082):358-362.
127. Taniguchi Y*, et al.* (2010) Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 329(5991):533-538.
128. Choi PJ, Cai L, Frieda K, & Xie XS (2008) A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* 322(5900):442-446.
129. Dekel E & Alon U (2005) Optimality and evolutionary tuning of the expression level of a protein. *Nature* 436(7050):588-592.
130. Eigen M, Mccaskill J, & Schuster P (1988) Molecular Quasi-Species. *J Phys Chem-Us* 92(24):6881-6891.
131. Schuster P (1993) RNA based evolutionary optimization. *Orig Life Evol Biosph* 23(5-6):373-391.
132. Biebricher CK & Eigen M (2006) What is a quasispecies? *Curr Top Microbiol Immunol* 299:1-31.
133. Lauring AS & Andino R (2010) Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog.* 6(7):e1001005.
134. Anderson JP, Daifuku R, & Loeb LA (2004) Viral error catastrophe by mutagenic nucleosides. *Annu Rev Microbiol* 58:183-205.
135. Bull JJ, Meyers LA, & Lachmann M (2005) Quasispecies made simple. *PLoS Comput Biol* 1(6):e61.
136. Schultes EA & Bartel DP (2000) One sequence, two ribozymes: Implications for the emergence of new ribozyme folds. *Science* 289(5478):448-452.
137. Heilman-Miller SL & Woodson SA (2003) Effect of transcription on folding of the Tetrahymena ribozyme. *Rna* 9(6):722-733.
138. Wickiser JK, Winkler WC, Breaker RR, & Crothers DM (2005) The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Molecular cell* 18(1):49-60.

139. Pereira MJ, Behera V, & Walter NG (2010) Nondenaturing purification of co-transcriptionally folded RNA avoids common folding heterogeneity. *PloS one* 5(9):e12953.
140. Meyer IM & Miklos I (2004) Co-transcriptional folding is encoded within RNA genes. *BMC Mol. Biol.* 5:10.
141. Mahen EM, Harger JW, Calderon EM, & Fedor MJ (2005) Kinetics and thermodynamics make different contributions to RNA folding in vitro and in yeast. *Molecular cell* 19(1):27-37.
142. Mahen EM, Watson PY, Cottrell JW, & Fedor MJ (2010) mRNA Secondary Structures Fold Sequentially But Exchange Rapidly In Vivo. *Plos Biol* 8(2):e1000307.
143. Bhaskaran H & Russell R (2007) Kinetic redistribution of native and misfolded RNAs by a DEAD-box chaperone. *Nature* 449(7165):1014-1018.
144. Fedorova O, Solem A, & Pyle AM (2010) Protein-facilitated folding of group II intron ribozymes. *Journal of molecular biology* 397(3):799-813.
145. Djebali S*, et al.* (2012) Landscape of transcription in human cells. *Nature* 489(7414):101-108.
146. Pitchiaya S, Heinicke LA, Custer TC, & Walter NG (2014) Single Molecule Fluorescence Approaches Shed Light on Intracellular RNAs. *Chem Rev* 114(6):3224-3265.
147. Sargueil B, Hampel KJ, Lambert D, & Burke JM (2003) In vitro selection of second site revertants analysis of the hairpin ribozyme active site. *J Biol Chem* 278(52):52783-52791.
148. Shcherbakova I, Mitra S, Laederach A, & Brenowitz M (2008) Energy barriers, pathways, and dynamics during folding of large, multidomain RNAs. *Curr Opin Chem Biol* 12(6):655-666.
149. Tinsley RA, Furchak JR, & Walter NG (2007) Trans-acting glmS catalytic riboswitch: locked and loaded. *RNA* 13(4):468-477.
150. Kim DH, Faull KF, Norris AJ, & Eckhert CD (2004) Borate-nucleotide complex formation depends on charge and phosphorylation state. *J Mass Spectrom* 39(7):743-751.
151. Flores R*, et al.* (2011) Rolling-circle replication of viroids, viroid-like satellite RNAs and hepatitis delta virus: variations on a theme. *RNA Biol* 8(2):200-206.
152. Walter NG (2003) Probing RNA structural dynamics and function by fluorescence resonance energy transfer (FRET). *Current protocols in nucleic acid chemistry / edited by Serge L. Beaucage ... [et al.]* Chapter 11:Unit 11 10.
153. Fedor MJ & Westhof E (2002) Ribozymes: the first 20 years. *Mol Cell* 10(4):703-704.
154. Uhlenbeck OC (1995) Keeping RNA happy. *RNA* 1(1):4-6.
155. Zhuang X*, et al.* (2002) Correlating structural dynamics and function in single ribozyme molecules. *Science* 296(5572):1473-1476.
156. Rueda D*, et al.* (2004) Single-molecule enzymology of RNA: essential functional groups impact catalysis from a distance. *Proc Natl Acad Sci U S A* 101(27):10066-10071.
157. Hyeon C, Lee J, Yoon J, Hohng S, & Thirumalai D (2012) Hidden complexity in the isomerization dynamics of Holliday junctions. *Nature chemistry* 4(11):907-914.
158. Liu B, Baskin RJ, & Kowalczykowski SC (2013) DNA unwinding heterogeneity by RecBCD results from static molecules able to equilibrate. *Nature* 500(7463):482-485.
159. Walter NG & Perumal S (2009) The Small Ribozymes: Common and Diverse Features Observed through the FRET Lens. *Springer Ser Biophys* 13:103-127.

160. Ferre-D'amare A R & Rupert PB (2002) The hairpin ribozyme: from crystal structure to function. *Biochem Soc Trans* 30(Pt 6):1105-1109.
161. Hampel KJ & Burke JM (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(12):3723-3729.
162. Ditzler MA, Rueda D, Mo J, Hakansson K, & Walter NG (2008) A rugged free energy landscape separates multiple functional RNA folds throughout denaturation. *Nucleic Acids Res* 36(22):7088-7099.
163. Greenfeld M, Solomatin SV, & Herschlag D (2011) Removal of covalent heterogeneity reveals simple folding behavior for P4-P6 RNA. *The Journal of biological chemistry* 286(22):19872-19879.
164. Kladwang W, Hum J, & Das R (2012) Ultraviolet shadowing of RNA can cause significant chemical damage in seconds. *Scientific reports* 2:517.
165. Gherghe CM, Mortimer SA, Krahn JM, Thompson NL, & Weeks KM (2008) Slow conformational dynamics at C2'-endo nucleotides in RNA. *J Am Chem Soc* 130(28):8884-8885.
166. Porschke D, Burke JM, & Walter NG (1999) Global structure and flexibility of hairpin ribozymes with extended terminal helices. *Journal of molecular biology* 289(4):799-813.
167. Wilkinson KA, Merino EJ, & Weeks KM (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nature protocols* 1(3):1610-1616.
168. Todd GC & Walter NG (2013) Secondary structure of bacteriophage T4 gene 60 mRNA: implications for translational bypassing. *Rna* 19(5):685-700.
169. Karabiber F, McGinnis JL, Favorov OV, & Weeks KM (2013) QuShape: rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *Rna* 19(1):63-73.
170. Merino EJ, Wilkinson KA, Coughlan JL, & Weeks KM (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* 127(12):4223-4231.
171. Stellwagen NC, Gelfi C, & Righetti PG (2000) DNA and buffers: the hidden danger of complex formation. *Biopolymers* 54(2):137-142.
172. Butcher SE & Burke JM (1994) A photo-cross-linkable tertiary structure motif found in functionally distinct RNA molecules is essential for catalytic function of the hairpin ribozyme. *Biochemistry* 33(4):992-999.
173. Yang Y, Jung DW, Bai DG, Yoo GS, & Choi JK (2001) Counterion-dye staining method for DNA in agarose gels using crystal violet and methyl orange. *Electrophoresis* 22(5):855-859.
174. Klostermeier D & Millar DP (2002) Energetics of hydrogen bond networks in RNA: Hydrogen bonds surrounding G+1 and U42 are the major determinants for the tertiary structure stability of the hairpin ribozyme. *Biochemistry* 41(48):14095-14102.
175. Rhodes MM, Reblova K, Sponer J, & Walter NG (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(36):13380-13385.
176. Butcher SE, Allain FH, & Feigon J (1999) Solution structure of the loop B domain from the hairpin ribozyme. *Nature structural biology* 6(3):212-216.

177. Kirsanov KI, Lesovaya EA, Yakubovskaya MG, & Belitsky GA (2010) SYBR Gold and SYBR Green II are not mutagenic in the Ames test. *Mutation research* 699(1-2):1-4.
178. Dorywalska M*, et al.* (2005) Site-specific labeling of the ribosome for single-molecule spectroscopy. *Nucleic acids research* 33(1):182-189.
179. Uemura S*, et al.* (2008) Single-molecule imaging of full protein synthesis by immobilized ribosomes. *Nucleic acids research* 36(12):e70.
180. Marshall RA, Dorywalska M, & Puglisi JD (2008) Irreversible chemical steps control intersubunit dynamics during translation. *Proceedings of the National Academy of Sciences of the United States of America* 105(40):15364-15369.
181. Lesley SA, Brow MA, & Burgess RR (1991) Use of in vitro protein synthesis from polymerase chain reaction-generated templates to study interaction of Escherichia coli transcription factors with core RNA polymerase and for epitope mapping of monoclonal antibodies. *The Journal of biological chemistry* 266(4):2632-2638.
182. Robertson JM & Wintermeyer W (1981) Effect of translocation on topology and conformation of anticodon and D loops of tRNAPhe. *Journal of molecular biology* 151(1):57-79.
183. Powers T & Noller HF (1991) A functional pseudoknot in 16S ribosomal RNA. *Embo J* 10(8):2203-2214.
184. Davis PB & Pearson CK (1978) Characterization of density gradients prepared by freezing and thawing a sucrose solution. *Analytical biochemistry* 91(1):343-349.
185. Kudla G, Murray AW, Tollervey D, & Plotkin JB (2009) Coding-sequence determinants of gene expression in Escherichia coli. *Science* 324(5924):255-258.
186. Huang WM*, et al.* (1988) A persistent untranslated sequence within bacteriophage T4 DNA topoisomerase gene 60. *Science* 239(4843):1005-1012.
187. Asai T*, et al.* (1999) Construction and initial characterization of Escherichia coli strains with few or no intact chromosomal rRNA operons. *J Bacteriol* 181(12):3803-3809.
188. Asai T, Zaporojets D, Squires C, & Squires CL (1999) An Escherichia coli strain with all chromosomal rRNA operons inactivated: Complete exchange of rRNA genes between bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 96(5):1971-1976.
189. Cammack KA & Wade HE (1965) The sedimentation behaviour of ribonuclease-active and -inactive ribosomes from bacteria. *The Biochemical journal* 96(3):671-680.
190. Spitale RC, Torelli AT, Krucinska J, Bandarian V, & Wedekind JE (2009) The structural basis for recognition of the PreQ0 metabolite by an unusually small riboswitch aptamer domain. *The Journal of biological chemistry* 284(17):11012-11016.
191. Jenkins JL, Krucinska J, McCarty RM, Bandarian V, & Wedekind JE (2011) Comparison of a preQ1 riboswitch aptamer in metabolite-bound and free states with implications for gene regulation. *The Journal of biological chemistry* 286(28):24626-24637.
192. Fei JY*, et al.* (2010) A Highly Purified, Fluorescently Labeled in Vitro Translation System for Single-Molecule Studies of Protein Synthesis. *Method Enzymol* 472:221-259.
193. Shimizu Y, Kanamori T, & Ueda T (2005) Protein synthesis by pure translation systems. *Methods* 36(3):299-304.
194. Roth A*, et al.* (2007) A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. *Nature structural & molecular biology* 14(4):308-317.

195. Pan D, Qin H, & Cooperman BS (2009) Synthesis and functional activity of tRNAs labeled with fluorescent hydrazides in the D-loop. *Rna* 15(2):346-354.
196. Grasby JA, Mersmann K, Singh M, & Gait MJ (1995) Purine functional groups in essential residues of the hairpin ribozyme required for catalytic cleavage of RNA. *Biochemistry* 34(12):4068-4076.
197. Shippy R, Siwkowski A, & Hampel A (1998) Mutational analysis of loops 1 and 5 of the hairpin ribozyme. *Biochemistry* 37(2):564-570.
198. Pinard R*, et al.* (2001) Functional involvement of G8 in the hairpin ribozyme cleavage mechanism. *Embo J* 20(22):6434-6442.
199. Milon P, Maracci C, Filonava L, Gualerzi CO, & Rodnina MV (2012) Real-time assembly landscape of bacterial 30S translation initiation complex. *Nature structural & molecular biology* 19(6):609-615.
200. Rasnik I, McKinney SA, & Ha T (2006) Nonblinking and long-lasting single-molecule fluorescence imaging. *Nature methods* 3(11):891-893.
201. Aitken CE, Marshall RA, & Puglisi JD (2008) An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys J* 94(5):1826-1835.
202. Altman RB*, et al.* (2012) Cyanine fluorophore derivatives with enhanced photostability. *Nature methods* 9(1):68-U178.