Title
Discovery and biological function of self-cleaving ribozymes

Author
Jimenez, Randi Maria

Publication Date
2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Discovery and biological function of self-cleaving ribozymes

DISSERTATION

submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Randi Maria Jimenez

Dissertation Committee:
Professor Andrej Lupták, Chair
Professor Celia Goulding
Professor Tom Poulos

2015

DEDICATION

To

My mother Renee Valadez Jimenez and my father Thomas Jimenez.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to acknowledge my advisor, Andrej Lupták, for his guidance and encouragement without which none of this was possible.

Thank you to my committee members: Drs Celia Goulding and Sheryl Tsai.

I would like to thank past and current members of the Lupták research group for their unending support, cooperativity, and company throughout my time at UC Irvine. The individuals deserving of special thanks: Dr Judy Webb for her mentorship, Drs Cassandra Burke, Dana Ruminski, and Kelsey Pobanz for their leadership and skills, and Julio Polanco for his steadiness and support. Their encouragement and friendship kept me realistic and focused and for that they have my deepest gratitude.

I extend thanks to American Society for Biochemistry and Molecular Biology, Inc. and Elsevier Inc. for permission to use copyright material in my dissertation.

My dissertation work received funding from UC Irvine Ayala School of Biological Sciences and the National Institutes of Health Ruth L. Kirschstein National Research Service Award predoctoral fellowship.

I would finally like to thank my family for their encouragement to pursue my degree. I would like to thank my mama for setting me on the right track all those years ago by teaching me the alphabet and its phonic sounds. Thanks to my parents I learned to be a hard worker, be a team player, and most importantly be myself. Thanks to my siblings Jesse, Bobbi, and Billy for making me into an “alpha”. Thanks to my “half an orange”, Kevin for the support, encouragement, and friendship over the years.

CURRICULUM VITAE

Randi M. Jimenez

EDUCATION

(2010 – 2015)	University of California, Irvine Department of Molecular Biology and Biochemistry Graduate Program
(2009-2010)	University of California, Irvine Medicinal Chemistry and Pharmacology Gateway Program
2009	B.S. in Biochemistry, University of California, Riverside

RESEARCH EXPERIENCE

January 2010 - present
University of California, Irvine
Advisor: Andrej Lupták, Ph.D. Department of Pharmaceutical Sciences, Department of Chemistry, Department of Molecular Biology and Biochemistry

September - December 2009
Department of Pharmacology, University of California, Irvine
Advisor: Kelvin W. Gee, Ph.D., UCI School of Medicine Department of Pharmacology

June - August 2006 and June - August 2007
Undergraduate Training Program, Loma Linda University, Loma Linda, CA
Mentor: Eileen Brantley, Ph.D. Center for Health Disparities and Molecular Medicine, LLU School of Medicine Department of Basic Sciences

TEACHING EXPERIENCE

Teaching Assistant University of California, Irvine
Department of Molecular Biology & Biochemistry
- Molecular Biology 99 discussion (2011-2012)
- Biochemistry 98 discussion (2012)
- Microbiology 118L lab (2011-2013)

Department of Pharmaceutical Sciences
- Molecular Pharmacology 107A discussion (2011)
- Physiology lab 120L (2012)

Department of Chemistry
- General Chemistry 1C lab (2010)

AWARDS AND HONORS

2013-2015	NIH Ruth L. Kirschstein National Research Service Award for individual predoctoral fellowships to promote diversity in health-related research (F31)
2012	UC—Irvine School of Biological Sciences, William D. Redfield Graduate Fellowship
2012	UC—Irvine Bioinformatics Training Program predoctoral fellowship

PUBLICATIONS AND PRESENTATIONS

Jimenez RM, Polanco JA, Lupták A. Chemistry and biology of self-cleaving ribozymes. Trends in Biochemical Sciences. Review (submitted).

Rampášek L, **Jimenez RM**, Lupták A, Brejová B, Vinař T. RNA Motif Search With Data-Driven Element Ordering. Bioinformatics (submitted).

Ho B, Polanco J, **Jimenez R**, Lupták A. Discovering human RNA aptamers by structure-based bioinformatics and genome-based *in vitro* selection. Methods Enzymol. 2014; 549:29-46.

Jimenez RM, Lupták A. Structure-based search and in vitro analysis of self-cleaving ribozymes. Methods Mol Biol. 2012; 848:131-43.

Jimenez RM, Rampášek L, Brejová B, Vinař T, Lupták A. Discovery of RNA motifs using a computational pipeline that allows insertions in paired regions and filtering of candidate sequences. Methods Mol Biol. 2012; 848:145-58.

***Jimenez RM**, Delwart E, Lupták A. Structure-based search reveals hammerhead ribozymes in the human microbiome. Journal of Biological Chemistry. 2011 Mar 11; 286(10):7737-43
*Paper of the week and "Best of JBC 2011".

McLean L, Soto U, Agama K, Francis J, **Jimenez R**, Pommier Y, Sowers L, Brantley E. Aminoflavone induces oxidative DNA damage and reactive oxidative species-mediated apoptosis in breast cancer cells. Int J Cancer. 2008 Apr 1;122(7):1665-74.

Advancement to Ph.D. candidacy examination

2012 Discovery and characterization of self-cleaving ribozyme structural variants. UC Irvine, Department of Molecular Biology & Biochemistry
Committee members: Andrej Lupták, Sheryl Tsai, Rachel Martin, Suzanne Sandmeyer, Klemens Hertel

Posters

†2014 Biological impact of transcripts terminated by HDV-like ribozymes. **Randi Jimenez**, Dana Ruminski, Andrej Lupták. RNA 2014: Annual RNA Society Meeting, Quebec, Canada

†2012 Application of new tools for functional RNA discovery. **Randi Jimenez**, Ladislav Rampášek, Tomas Vinař, Bronislav Brejová, Andrej Lupták. Annual International Conference for Intelligent Systems in Molecular Biology (ISMB), Integrative RNA Biology SIG, Long Beach, CA

2011 RNArobo: A software package for RNA structural motif search and filtering. Ladislav Rampášek, **Randi M. Jimenez**, Bronislav Brejová, Andrej Lupták. UC—Irvine, Annual International Student Organization Symposium.

2007 Aminoflavone induced anticancer activity is attenuated in MDA-MB-468 human breast cancer cells pretreated with the DNA-methyltransferase 5-aza-2′-deoxycytidine. **Randi Jimenez**, Louisa Amis, Willie Davis, Eileen Brantley. Loma Linda University, Health Disparities Research Symposium

2006 Aminoflavone modulates reactive oxygen species formation in a cytochrome P450 1A-dependent manner in estrogen receptor negative MDA-MB-468 breast cancer cells. **Randi Jimenez**, Lancelot McLean, Louisa Amis, Eileen Brantley. Loma Linda University, Health Disparities Research Symposium

† Includes corresponding oral presentation.

Talks

2014 Biological impact of transcripts terminated by HDV-like ribozymes. RNA 2014: Annual RNA Society Meeting, Quebec, Canada

2013 Characterization of metabolite responsive hepatitis delta virus-like ribozymes. UC—Irvine, Department of Molecular Biol. & Biochem., department retreat, Lake Arrowhead, CA

2012 Application of new tools for functional RNA discovery. Annual international conference for Intelligent Systems in Molecular Biology (ISMB), Integrative RNA Biology SIG, Long

Beach, CA

2012 Characterization of metabolite responsive hepatitis delta virus-like ribozymes. UC—Irvine, Department of Molecular Biol. & Biochem., weekly graduate research seminar

2012 Hammerhead ribozymes from the human microbiome. UC—Irvine, Department of Molecular Biol. & Biochem., annual graduate research seminar; UC—Irvine, Department of Pharmaceutical Sciences, Annual Vertex Pharmaceuticals Symposium

ABSTRACT OF THE DISSERTATION

Discovery and biological function of self-cleaving ribozymes

by

Randi Maria Jimenez

Doctor of Philosophy in Biological Sciences

with a concentration in Molecular Biology & Biochemistry

University of California, Irvine, 2015

Professor Andrej Lupták, Chair

RNA plays diverse biological roles, including in regulation and catalysis. Transcriptome analysis reveals transcription from throughout the genome, including many non-coding RNAs (ncRNA), distinct from housekeeping RNAs. The complexity and diversity of these ncRNAs presents an enormous task in relating RNA structure and function. The widespread distribution of self-cleaving RNAs (ribozymes) strongly suggests a biological significance. The ribozymes have been discovered in highly diverse genomic contexts throughout nature, from viroids to vertebrates. Their biological roles include self-scission during rolling-circle replication of RNA genomes, co-transcriptional processing of eukaryotic retrotransposons, and metabolite-dependent gene expression regulation in bacteria. Other examples, including highly conserved mammalian ribozymes, suggest that many new biological roles are yet to be discovered. The work presented here focuses on two families of ribozyme (hammerhead and HDV-like) which are widespread throughout all kingdoms of life, including in the human genome, but for which the biological roles remain unclear. Through the discovery of new ribozymes and characterization of their biochemical properties, we aim to elucidate the biological roles of self-cleaving ribozymes.

Deep sequencing of viral or bacterial nucleic acids monitors the presence and diversity of microbes in select populations and locations. Metagenomic study of mammalian viromes can help trace paths of viral transmissions within or between species. High-throughput sequencing of patient and untreated sewage microbiomes showed many sequences with no similarity to genomic sequences of known function or origin. To estimate the distribution of functional RNAs in these microbiomes, we used the hammerhead ribozyme motif to search for sequences capable of assuming its three-way-junction fold. While only two of

the three possible natural HHR topologies had been known, our analysis revealed highly active ribozymes that terminated in any of the three stems. Altogether, thirteen ribozymes were confirmed, the most abundant of these are type II HHRs, one of which is the fastest natural cis-acting HHR yet discovered. We demonstrate that a structure-based search for a known functional RNA is a powerful tool for analysis of metagenomic datasets, complementing sequence alignments.

The hepatitis delta virus-like ribozyme (HDV) ribozyme was first discovered in the genome of the human pathogen, where self-scission by the ribozyme processes RNA concatamers during replication. The nested double-pseudoknot secondary structure of the ribozyme was used in bioinformatic searches to identify the widespread occurrence of this class of self-cleaving ribozyme throughout nature. Previous work using structure-based searches identified a HDV-like ribozyme located immediately upstream of the *glmM* gene, encoding glucosamine mutase, in the genome of the human gut bacterium *Faecalibacterium prausnitzii*. Recent searches identified an active HDV ribozyme located immediately downstream of the same *glmM* gene in the *F. prausnitzii* genome. Three additional ribozymes from three different bacterial genomes were also identified through bioinformatics approaches, however they do not map near *glmM* genes in the respective genomes. These five ribozymes are among a handful of HDV-like ribozymes mapping to bacterial genomes and their location suggests they play a role in regulating expression of a downstream gene. Our work demonstrates that *in vitro* catalysis by these ribozymes is sensitive to the presence of amine sugars such as glucosamine 6-phosphate, the substrate of the glmM enzyme. *In vivo* assays in *E.coli* demonstrate that expression of a downstream ORF is influenced by the ribozyme and its state of self-cleavage. The *F. prausnitzii* ribozymes represent a unique arrangement in which ribozyme self-cleavage may facilitate the regulation of the downstream gene.

On the other hand, the occurrence of HDV-like ribozymes in eukaryotic genomes falls to diverse genomic loci, however in some cases their location hinted at potential biological functions. Many ribozymes mapped to the 5′ end of non-long terminal repeat retrotransposons and leads to a model where the ribozyme apparently plays several roles in the retrotransposition cycle. Of particular interest is the ribozyme mapping to the R2 retrotransposable element in *Drosophila.* The predicted coding sequence for the first three amino acids of the R2 protein maps within the ribozyme structure. Previous work showed that these HDV-like ribozymes promote translation initiation both *in vitro* and *in vivo*, but the precise

location of the first translated codon remained unknown. Here we investigate the mechanism of translation initiation of ribozyme-terminated mRNAs. Our results indicate that the correct folding of the ribozyme core is important in promoting translation. Translation occurs in the absence of a cap, start codon, and poly-A tail. Taken together our data suggest that this translation is distinct from other cap-independent mechanisms. This work highlights the significance of self-cleaving ribozymes in both bacterial and eukaryotic genomes.

ABBREVIATIONS

HHR-hammerhead ribozyme

HDV – hepatitis delta virus

glmS – glucosamine 6-phosphate synthase

glmM – glucosamine 6-phosphate mutase

HBV – hepatitis B virus

ORF – open reading frame

mRNA – messenger RNA

rRNA – ribosomal RNA

CPEB3 – cytoplasmic polyadenylation element-binding protein 3

EST – expressed sequence tag

RT-qPCR – reverse transcriptase-polymerase chain reaction

SNP – single-nucleotide polymorphism

tRNA – transfer RNA

RT – reverse transcriptase

IRES – internal ribosomal entry site

UTR – untranslated region

Fluc – Firefly luciferase

Rluc – Renilla lucinerase

CIV – *Chilo* Iridescent Virus

HCV – hepatitis C virus

LTR – long terminal repeat

LINE – long interspersed elements

SINE – short interspersed elements

WT – wild-type

m7G – 7methylguanosine

CITE – cap-independent translation enhancer

RRL – rabbit reticulocyte lysate

Chapter 1

Chemistry and biology of self-cleaving ribozymes

Foundations for ribozyme exploration

Small nucleolytic ribozymes carry out site-specific phosphodiester scission without the need for protein chaperones or enzymes. There are six ribozyme families found in nature, consisting of: hairpin, hammerhead, hepatitis delta virus-like, glmS, *Neurospora* Varkud satellite, and twister motifs. All rely on a combination of catalytic strategies to complete self-scission in an active site formed by the secondary and tertiary structures unique to each family. For these ribozymes, cleavage involves a nucleophilic attack by a 2′ hydroxyl on an adjacent phosphodiester bond, yielding a 2′ 3′-cyclic phosphate and a 5′-hydroxyl product. Crystal structures have been essential to illuminate how functional groups participate in the catalytic mechanism used by each family of ribozyme. Shifted pKa values resulting from the local electronic environments within the active site allow nucleobases and cofactors to participate in general acid-base catalysis. Despite the plethora of information regarding the genomic distributions, structures, and mechanisms, the biological roles of the three most widespread families (hammerhead, HDV-like, and twister) remain largely elusive. The study of small ribozymes has provided a platform for discovering new catalytic RNAs and new roles for non-coding RNA, as well as aided in the design of new molecules for synthetic biology. We summarize the chemical and biological properties of the six types of self-cleaving ribozymes.

Hammerhead ribozyme

The hammerhead ribozyme (HHR) was originally discovered in viroids and virusoids where it functions in the processing of rolling circle transcripts (Fig.1-1A) (1-3). Over the next two decades, isolated instances were found in various newt species (4), cave crickets (5), and the human blood fluke *Schistosoma mansoni* (6). HHRs found in the genomes of various schistosomes map to short interspersed element-like (SINE) retrotransposable element (Fig.1-1B) (6,7). The conservation of these SINE-associated HHRs suggests vertical transmission from a common ancestor of schistosomes. The HHR is perhaps the most widespread ribozyme family, found to exist in all three types across all domains of life (4-6,8-12). Most eukaryotic occurrences map to satellite transcripts and retroelements; however, a conserved group in amniotes map to the introns of specific genes (9,10,13). This conservation in introns implicates a role for these ribozymes in pre-mRNA processing. It seems that the function of phosphodiester scission that has been preserved in propagating mobile elements has been adapted to introducing diversity in more complex genetic controls, but more research is needed to decipher the potential role of HHRs in intron cleavage and pre-mRNA processing. The efficient self-scission of HHRs *in vitro* and *in vivo* suggests the activity is central to their biological significance.

The HHR secondary structure is comprised of three helices (stems I, II, III) branching from a catalytic core containing fifteen nucleotides essential to catalysis (Fig. 1-2A). The ribozyme exists naturally in three circularly permuted topologies (types I, II, and III) and has also been found in a discontinuous type I format conserved in the 3′ UTR of mouse C-type lectin type II (*Clec2*) and some mammalian *Clec2-like* genes (14).

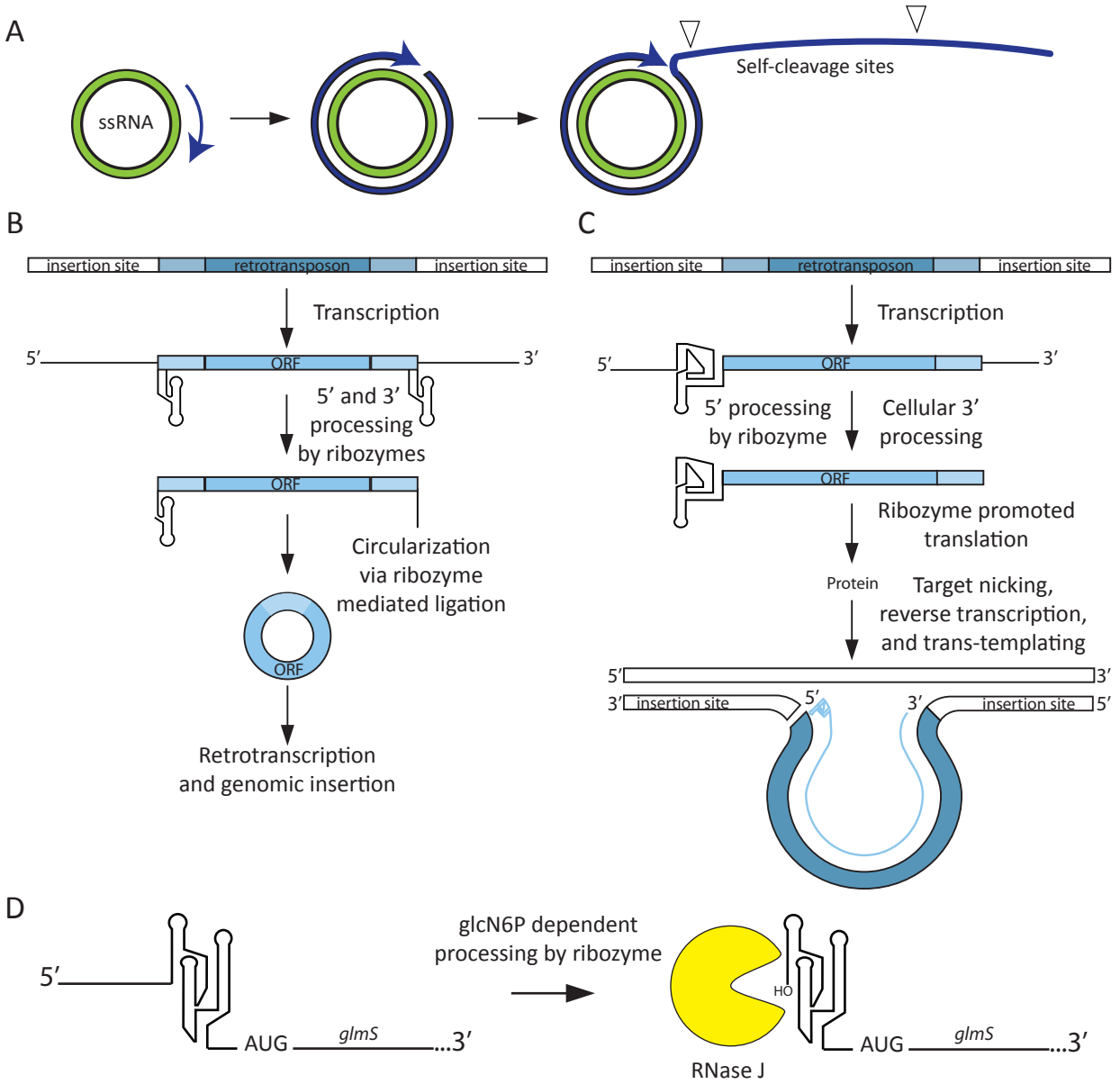


Figure 1. Roles of self-cleaving ribozymes in different biological systems. (A) Most instances of self-cleaving ribozymes found in viral genomes are involved in replication. During rolling-circle replication of a single-stranded RNA genome, concatameric copies of the opposite polarity are generated. The self-cleavage activity of ribozymes generates unit length copies that must be circularized and replicated to complete the cycle. (B) Hammerhead ribozymes mapping to non-autonomous retrotransposons (SINEs) serve to mobilize the retroelement. Complementarity of the 5′ and 3′ ends facilitates ribozyme self-ligation resulting in circularization of the RNA; this feature enhances the reverse transcription and genomic insertion processes that complete the retrotransposition cycle. (C) Similarly, retrotransposon-associated HDV-like ribozymes liberate the 5′ end of the retroelement from the full-length transcript. The ribozyme structure in the 5′ UTR promotes translation of the downstream ORF. Further, the HDV-like ribozyme on the 5′ end of the RNA promotes the template switching necessary to complete genomic insertion of the autonomous (LINE) element. (D) The glmS ribozyme is located in the 5′ UTR of the *glmS* gene, encoding the glutamine-fructose 6-phosphate aminotransferase enzyme responsible for generating glucosamine 6-phosphate (glcN6P), necessary for bacterial cell wall synthesis. The ribozyme requires glcN6P as a cofactor for catalysis. The resulting 5′-hydroxyl makes the processed transcript a substrate for the 5′-3′ exonuclease RNase J. Therefore, the glmS ribozyme is a riboswitch that turns off glcN6P synthase expression in response to glcN6P.

The HHR was the first ribozyme to have its crystal structure solved (15,16). These early crystallographic analyses took advantage of the potentially compact architecture of HHRs, as biochemical evidence suggested that removing structured domains peripheral to the catalytic core was not detrimental to activity (1). The early crystal structures revealed the minimal HHR trapped in an open or pre-catalytic state that would require significant rearrangement to bring the essential core residues into position for in-line attack. The following years of conflicting mechanistic and structural data were resolved by the characterization of a distal tertiary contact between stems I and II (17,18). This contact results in significantly faster catalytic rates, presumably because it stabilizes the catalytically competent conformation of the active site. The global structure of the ribozyme is Y-shaped, comprising a coaxial stack of stems III and II that pack next to stem I. The crystal structure of the full-length ribozyme reveals the optimal positioning of residues implicated in general acid-base catalysis, as well as correct positioning of the nucleophile for attack on the scissile phosphate (19).

The currently proposed model of catalysis by a HHR involves only ribonucleotides and no divalent metal ions participating directly in catalysis. Biochemical and crystallographic evidence supports C_{17} nucleophile attacking the adjacent scissile phosphate of $C_{1.1}$. Invariant residues G_8 and G_{12} serve as general acid and general base, respectively (20). A recent crystal structure reveals three Na+ cations in the HHR active site, one of which binds to G_{12} and may be involved in stabilizing the negative charge of the deprotonated N1 position (21). A role for monovalents in perturbing the pK_a of the general base has been suggested previously (22,23). The N1 position of the ionized form of G_{12} is thought to abstract the 2′ hydrogen from the nucleophilic oxygen

of C_{17}, the 2′-O is then positioned for in-line attack on the phosphate of $C_{1.1}$ (24). The ribose of G_8 hydrogen-bonds to the 5′-oxyanion leaving group and donates a proton during cleavage of the phosphodiester backbone (19,24).

Hepatitis delta virus family of ribozymes

The hepatitis delta virus life cycle depends on coinfection with hepatitis B virus (HBV) (25) and it relies on host cellular machinery for replication of its single-stranded RNA genome (26). The HDV genome encodes two related, but distinct ribozymes that were first discovered through their function in processing concatameric copies of the viral RNA genome during rolling circle transcription (Fig.1-1A) (27). More recently, this ribozyme motif was discovered in the human genome using an *in vitro* selection form a genomic library and sequence analysis showed that it is highly conserved in mammals (28). When the conserved nested double-pseudoknot secondary structure was used to identify the motif in other genomes, the family of HDV-like ribozymes was found to exist in nearly all branches of life (29). The mapping of HDV-like ribozymes to the 5′ UTR of autonomous (LINE) retrotransposons points to a role in the processing of retrotransposons (Fig.1-1C). Besides processing the 5′ end of the retroelement out of the parent transcript, ribozymes may be facilitating the insertion of the retrotransposon. The R2 LINE retrotransposon relies on the R2 reverse transcriptase protein to switch from the RNA to the cDNA as a template, thus completing the retrotransposition cycle (30,31). This process is more efficient when the RNA template contains a 5′-hydroxyl, such as that which would result from ribozyme self-cleavage.

Expression analysis and 5′ processing studies support a model in which HDV-like ribozymes function in non-long terminal repeat (LTR) retrotransposition (32-34). RT-

qPCR experiments reveal the extent of self-cleavage varies by sex and life stage in the mosquito *Anopheles gambiae* (29), suggesting environmental sensitivity and regulation. Further, the HDV-like ribozyme found in an intron of the cytoplasmic polyadenylation element binding protein 3 (*CPEB3*) gene is highly conserved in mammals (28), and people homozygous for a SNP known to affect CPEB3 ribozyme activity show differences in episodic memory (35). Other ribozymes are located at unrelated genomic loci and in both coding and non-coding transcripts, suggesting multiple biological functions. HDV-like ribozymes have fast self-cleavage kinetics and have a remarkably stable structure, demonstrating activity in up to 18 M formamide (36,37).

The HDV-like ribozyme secondary structure is a composed of five helical regions (P1.1, P1, P2, P3, P4) forming two coaxial stacks (P1, P1.1 stack on P4 and P2 stacks on P3) that are joined by single stranded regions (J1/2, L3, J4/2) (Fig. 1-2B). There are six conserved nucleotides fulfilling functional or structural roles in the active site. The ribozyme exists naturally in a minimal (lacking P4) or extended (extension of J1/2 and P4) form.

Crystal structures of the genomic HDV ribozymes revealed the secondary structure forms a nested double-pseudoknot (Fig. 1-2B) (38-40). The active site is buried in the junction of P1, P1.1 and P3. C_{75} participates not only in the reaction mechanism but is also important to the active site architecture, as structural analysis of $C_{75}U$ inactive mutants reveal significant distortions (39). The exocyclic amine of C_{75} forms hydrogen bonds with the *pro-*R_p oxygens of C_{22} and the scissile phosphate. The protonated form of C_{75} is thought to be stabilized through interactions with the scissile phosphate (40) and may interact electrostatically with the metal ion bound in the active site (41).

Divalent metals have a profound impact on catalysis and recent evidence reveals a Mg^{2+} ion in a metal binding pocket in the ribozyme active site (40,42). This negatively charged pocket is formed by a reverse G●U wobble pair (G_{25} and U_{20}), the scissile phosphate, and the phosphate of U_{23} (42).

The proposed mechanism of catalysis suggests that HDV-like ribozymes are multi-channel, employing different strategies in the presence and absence of divalent metal ions. In one scenario, the protonated form of C_{75} donates a proton to the 5′-O of G_1. The HDV-like ribozymes are metalloenzymes under biological conditions with the divalent metal cation thought to stabilize the developing negative charge on the nucleophile (42). Metal binding to the nucleophile U_{-1} facilitates deprotonation of a general base, likely a hydroxide ion, and also likely stabilizes the pentavalent phosphorane transition state (43). An attack on the scissile phosphate of G_1 results in a phosphorane intermediate that is resolved into the two products, in a manner similar HHRs. There is little divalent ion specificity and catalysis can proceed in high concentrations of monovalent cations alone (44). In the absence of divalent cations, the dramatic shift in pKa of the catalytic nucleobase C_{75} may be enough to stabilize the negatively charged nucleophile and acid-base catalysis proceeds without mediation by a divalent metal cation (45,46).

Hairpin ribozymes

There are only three known examples of this ribozyme family, residing in the satellite RNAs of the tobacco ringspot (1), chicory yellow mottle (47), and arabis mosaic viruses (48). Each occurrence coincides with the appearance of a HHR on the RNA of opposite polarity and together these ribozymes function in processing multimeric RNA during rolling circle replication of the viral genome. The hairpin ribozyme secondary structure is

composed of four helical stems (A-D) anchored in a four-way junction and forming two coaxial stacks (D stacks on A and C stacks on B). Stems A and B contain internal bulges housing the scissile phosphate (stem A) and nucleotides essential to catalysis, whereas stems C and D provide structural stability and are not necessary for ribozyme activity.

Crystal structures of the hairpin ribozyme have been determined with and without stems C and D, the junction (49) and hinge forms (50), respectively. The structures support biochemical evidence for the intimate interaction of loops A and B that form the active site of the ribozyme (Fig. 1-2C) (49-51). A_{38} and G_8 are ideally positioned to serve as general acid and general base, respectively. G_8 is on the strand opposite of the scissile phosphate in loop A and A_{38} is in loop B. Single molecule studies revealed that the loop-loop interaction is dynamic and oscillates between an antiparallel (active form) and parallel state (52,53). Analysis of the roles of A_{38} and G_8 by phosphorothioate substitutions supports ionization of the two groups but further suggests that a conformational change is limiting the ribozyme activity (54). Significant alterations in backbone geometry results in widening of the minor grooves in both stems A and B, providing a landscape for protruding bases to participate in interhelical interactions.

G_{+1} is extruded from the minor groove of stem A, and is involved in an network of interactions with both stem A and B, together forming the G_{+1}-binding pocket (50). Any mutations interfering with these interactions completely abolish ribozyme activity (55). The highly stabilized extrusion of G_{+1} constrains the riboses of A_{-1} and G_{+1} enabling the in-line arrangement of the nucleophile and leaving group (50). The mechanism of catalysis for hairpin ribozymes, based on the crystal structure of a phosphorane mimic,

vanadate, points to a role of A_{38} and G_8 acting as general acid and general base, although they may also participate in transition state stabilization (51). In the crystal structures, G_8 is hydrogen-bonded to the 2′-O (through N1) and the *proS* non-bridging O of the scissile phosphate, whereas A_{38} is hydrogen-bonded to the 5′-O and the *proR* O (51). These roles are supported by functional group modifications and nucleobase substitution experiments (56-59). The N1 of G_8 donates a H-bond to the nucleophilic oxyanion of A_{-1}, which attacks the adjacent phosphate 5′ of G_{+1}. A_{38} acts as general acid by protonating the leaving group, resulting in scission of the RNA backbone.

Neurospora Varkud satellite ribozyme

The Varkud satellite (VS) ribozyme is important in replication of a single-stranded RNA satellite found in some strains of *Neurospora* (60). The secondary structure of the VS ribozyme is composed of seven helices (I-VII) that form three three-way junctions (Fig. 2D). A kissing loop interaction between stems I and V (61) facilitates docking of the internal loop of stem I with the internal loop of stem VI which forms the active site.

There is currently no published crystal structure solved for the VS ribozyme, however x-ray scattering experiments reveal A_{756} (stem loop VI) and G_{638} (stem loop I) are positioned in close proximity to the scissile phosphate. Substitution of G_{638} by diaminopurine shifts the pH profile for ribozyme catalysis suggesting that the guanosine is involved in proton transfer (62). Phosphorothioate substitution at the scissile phosphate rescues activity of the $A_{756}G$ impaired mutant, suggesting that A_{756} participates as general acid in ribozyme catalysis (63) which points to G_{638} acting as general base. This is similar to the mechanism proposed for the hairpin ribozyme wherein a guanosine and adenosine facilitate general acid-base catalysis.

The kissing loop interaction of SLI and SLV is important to the active conformation of the ribozyme, promoting catalysis, and structural stability. It was first predicted based on site-directed mutagenesis and chemical modification analysis (64,65). The kissing loop interaction involves a conformational change (shift) in SLI that is necessary for ribozyme activity (66-68). The SLI/SLV interaction facilitates the docking of SLI with SLVI thereby forming the active site (67,69,70).The SLI/SLV kissing loop NMR structure has been solved (71-73) and supporting isothermal titration calorimetry experiments demonstrate that the SLI/SLV kissing-loop interaction is a major thermodynamic barrier and is therefore likely regulating ribozyme activity (61). The formation of the essential SLI/SLV kissing loop interaction relies upon divalent metal cations however there is no evidence for divalent cations being involved in catalysis. The mechanism of catalysis for the VS ribozyme is that G_{638} acts as base by abstracting the proton from the 2′-OH of G_{620}. The 2′-O then attacks the leaving group 5′-O of A_{621}. A_{756} acts by protonating the leaving group and the final 5′-OH and 2′ 3′-cyclic phosphate products are resolved.

Glucosamine 6-phosphate synthase (glmS) ribozyme

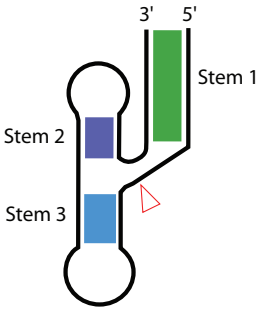
The glmS ribozyme was first discovered in 2004 as a conserved motif located in the 5′ UTR of the *glmS* gene, encoding for glutamine-fructose-6-phosphate amidotransferase, in many Gram-positive bacteria (74). It is the only example of a metabolite responsive self-cleaving ribozyme, utilizing glcN6P as a cofactor necessary for catalysis. It is also a catalytic riboswitch, regulating the expression of the *glmS* gene by a negative feedback mechanism whereby scission of the mRNA results in degradation of the message and a decrease in synthesis of the enzyme generating the glcN6P. Self-cleavage is completely dependent on the presence of amine-containing

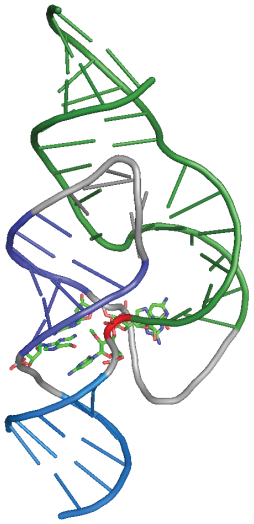
ligands, and the pKa of the amine functionality impacts reactivity but not binding affinity (75).

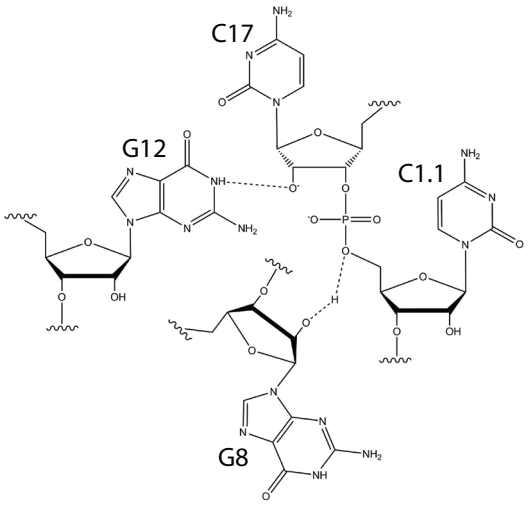
The structure of the glmS ribozyme is composed of three coaxial stacks (P1 stacks P3.1, P4 stacks P4.1, and P2.1), which are packed in a nearly parallel fashion (Fig. 2E). The core of the ribozyme is doubly pseudoknotted (P2.1 and P2.2 are pseudoknots). P2.2 contains the scissile phosphate and helps form the binding site for the glucosamine 6-phosphate (glcN6P) cofactor. P3 and P4 are not necessary for activity; however, they enhance cleavage by providing structural stability (74,76). Ribozyme activity depends on the presence of the cofactor which is involved in both the catalytic and regulatory roles of the riboswitch. Crystal structures of both pre- and post-cleavage states revealed there is no conformational change upon ligand binding (77).

The crystal structures of the glmS ribozyme were important to refining the structural model predicted from covariation analysis. In addition, the location of a divalent metal, the metabolite binding pocket, and the active site provided strong evidence for the role of glcN6P in the catalytic mechanism. The major grooves of P2.1 and P2.2 form the roof of the active site and the metabolite binding pocket. Both glcN6P (78,79) and the competitive inhibitor glc6P (77) are buried in the metabolite binding pocket which lies behind the active site. Both the sugar and phosphate of the metabolite cofactor are recognized in the pocket: the G_1 nucleobase stacks the sugar ring, N1 H-bonds to the phosphate, which also coordinates a Mg^{2+} cation. The 1-hydroxyl of glc6P donates a H-bond to the pro-*R*p oxygen of the scissile phosphate A_{-1} and receives an H-bond from N1 of G_{65} (77). The 2-OH of glc6P forms a H-bond to the 5′-O of G_1 (77), whereas in the

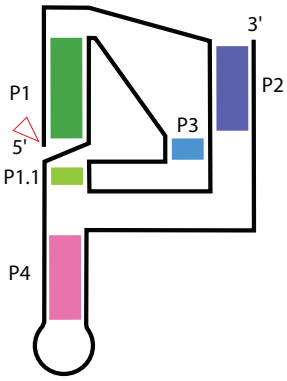
A

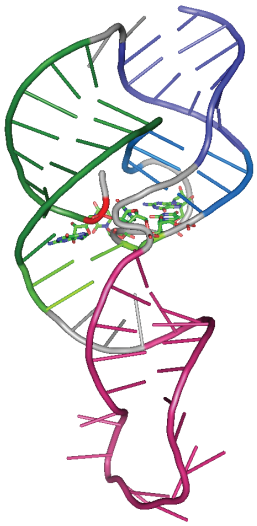

3'
5'
Stem 1
Stem 2
Stem 3

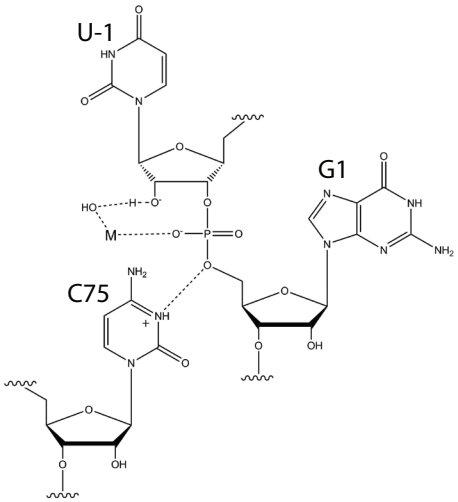



C17
G12
C1.1
G8

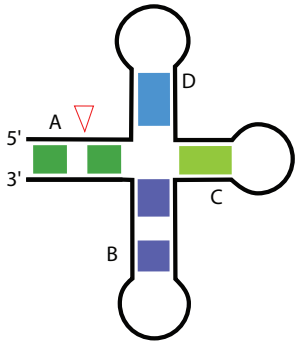
B

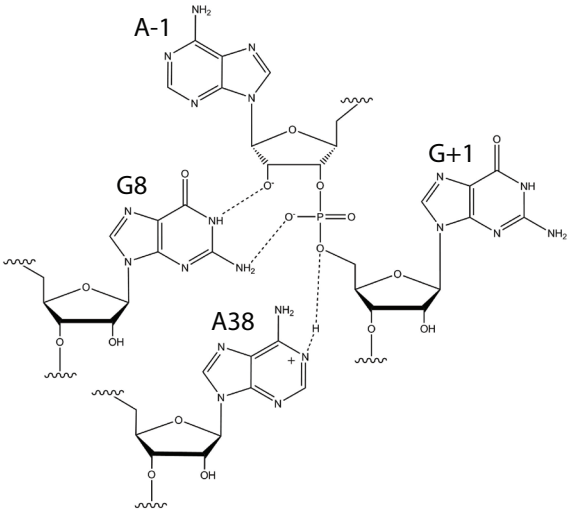

3'
P1
P2
P3
5'
P1.1
P4




U-1
G1
C75

C


D
A
5'
3'
C
B


A-1
G8
G+1
A38

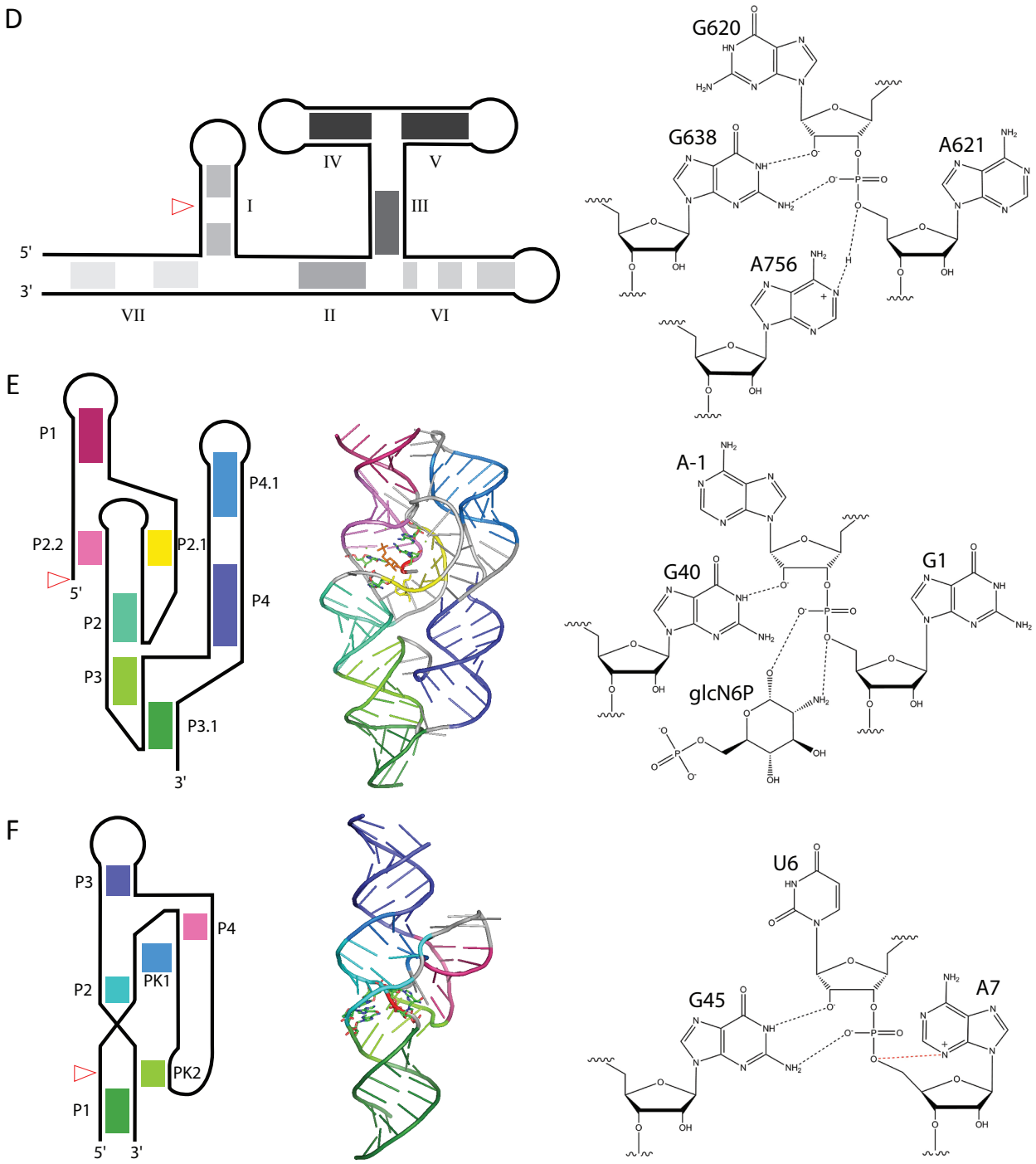


Figure 2. Overview of self-cleaving ribozymes: structural features and proposed cleavage mechanisms. The secondary structures of the (A) hammerhead (B) hepatitis delta virus-like (C) hairpin (D) *Neurospora* Varkud satellite (E) glmS, and (F) twister ribozymes are shown on the left column. The corresponding crystal structures for each ribozyme are shown on the middle column with colored helices to emphasize helical stacks and pseudoknots. The site of scission is pointed out by the red arrowheads in the secondary structures and highlighted in red in the respective crystal structures. A model of the cleavage site for each ribozyme is shown on the right column illustrating interactions promoting the cleavage event. Interactions proposed to stabilize the transition state of the reaction are shown for all ribozymes. For the twister ribozyme, a hypothetical interaction (dashed red line) between N3 of A_7 and the scissile phosphate suggests that A_7 acts as the general acid during self-cleavage. PDBID: 3ZD5, ISJ3, IM5K, 2NZ4 4QJH

presence of glcN6P, the amine moiety is making that bond (78). There are additional interactions between the cofactor and nucleobases A_{50} and C_2.

At the active site, G_{65} forms an irregular base-pair with A_{-1}. Together, G_{39} and G_{65} position the scissile phosphate by binding its non-bridging oxygens. These interactions, along with the 2′-endo ribose pucker of A_{-1}, twist the RNA backbone into a nearly perfect in-line conformation. The nucleotides 5′ and 3′ of the scissile phosphate are unstacked and splayed apart. Overall, the active site is remarkably rigid, and the glmS ribozyme is the only other enzyme besides the ribosome to be active in a crystallized form.

G_{40} and the metabolite are in proper position to serve in general acid-base catalysis. $G_{40}A$ and G_{65} mutants, interacting with leaving group and cofactor, respectively, completely abolish ribozyme activity. Raman crystallography revealed the ionization states of the glcN6P cofactor and supports the role of G_{40} in tuning the pKa of both the phosphate and amine of the ligand (80). This is remarkable considering the amine of free glcN6P is neutral. Thus the positively charged amine is positioned to serve in proton transfer. The relationship between the ionization state of the cofactor and the ribozyme activity suggest the cofactor participates directly in proton transfer (81). However, the only evidence for the respective roles of G_{40} and the glcN6P cofactor in catalysis are their positions within the active site. The current proposed model for catalysis by the glmS ribozyme suggests that the amino group of glcN6P (~3 Å from the 5′-O leaving group) serves as general acid and N1 of G_{40} (within 3.2 Å of the nucleophile) serves as general base (77,78). Therein, G_{40} abstracts a proton from the 2′-OH of A_{-1}, which attacks the 5′-O of G_1 which is protonated by glcN6P.

Twister ribozyme

The twister ribozyme was discovered in 2013 through bioinformatics searches for conserved RNA structural motifs (82). The secondary structure is composed of three stems (P1, P2, P4) joined by internal and terminal loops (L1, L2, L4). There are two pseudoknots (T1 formed between L1 and L4, T2 formed by L2 and L4). The ribozyme can exist in an extended format with additional stem loops P3 and P5 creating a three- or four-way junction at L2. In addition, the twister ribozyme can exist in three circularly permuted varieties (P1, P3, or P5). Twister ribozymes demonstrate robust cleavage kinetics both *in vitro* (k_{obs} > 1000 min^{-1} at physiological conditions) and *in vivo* (82). Biochemical and crystallographic evidence support the roles of ten conserved nucleotides at the active site.

The crystal structure of the *Oryza sativa* (Osa-1-4) P1 type twister ribozyme lacking P3 and P5 was solved to 2.3 Å resolution by Y. Liu *et. al*. (83). The structure correlates well with that predicted by covariation and in-line probing analysis (82). Helices P1, T1, P2, and T2 are coaxially stacked and P4 lies approximately parallel to the stack. Every nucleotide in L4 is involved in base-pairing interactions and substitutions to any one is detrimental to catalytic activity. Some of this base-pairing involved in T1 formation is facilitated by a large helical twist between T1 and P2. The active site is located in the major groove of the T1-P2 helix. The A_8-G_{45} pair at the base of P2 stacks on top of the A_{28}-A_{46} pair at the top of T1. G_{45} donates a hydrogen-bond to the *proR*-O of the scissile phosphate of A7 (in P1) and has been implicated in proton transfer during catalysis (83). Another crystal structure of a P1-type twister from an environmental sequence solved to 2.9 Å resolution reveals a Mg^{2+} cation at the scissile phosphate (84). As in

many other ribozyme active sites, the nucleotides flanking the scissile phosphate are splayed apart, promoting the in-line arrangement necessary for catalysis. The resolution of a divalent cation along with the ordering and in-line arrangement of the U_6 nucleophile and A_7 leaving group is unique to this study and provides further elucidation of the factors contributing to the remarkable rates of catalysis by twister ribozymes.

Two additional crystal structures of type-P3 twister ribozymes from *O. sativa* and an environmental sequence further suggest that a conserved adenosine is involved in stabilizing the negative charge on a non-bridging oxygen at the cleavage site (85). The N6 of A_{63} (corresponding to A_{46} in the Osa-1-4 in the above mentioned crystal structure), is positioned to participate in transition state stabilization. Similar to the hairpin ribozyme, the twister crystal structures demonstrate the utilization of a transition state stabilization strategy. Unlike the hairpin ribozyme, twister has a pre-organized active site due to the rigidity provided by extensive base-pairing in such a short sequence space; only two out of the ten conserved nucleotides in the active site remain unpaired. Remarkably, there are no nucleobases proposed to play the role of general acid, however, active site rearrangement may be revealed upon crystallization of a transition state analog of the ribozyme.

The six families of self-cleaving ribozyme represent some of the smallest catalytic molecules known to biology. The intricate and elegant 3D structures unique to each family of ribozyme depict different means of accomplishing the same reaction. This is useful to synthetic biology; where ribozymes can be used as platforms conveying information about ligand recognition and molecular targeting. Equally exciting and consuming is the idea that there are many more RNA structures capable of catalysis yet

to be discovered in nature. The vast distribution for some of the ribozyme families presented here leads to many intriguing questions regarding their biological relevance.

Biological significance

Self-cleaving ribozymes are remarkably efficient enzymes, and we are continuing to discover their importance in biological systems. The biology of these common ribozymes remains one of the most unexplored areas of research. The glmS ribozyme is unique in its role as a riboswitch controlling the flux of glcN6P in many Gram-positive bacteria. For the hairpin and VS ribozymes, each known instance is associated with a defined biological role in viral genome replication, although their true distribution may reach further than we currently understand. After all, it took nearly 20 years to uncover more than a handful of examples for either HHR or HDV-like ribozymes. The HHR, HDV-like, and twister ribozyme families are widespread and found in diverse biological contexts. Some isolated occurrences have been studied for their particular influence on genetic controls (13,32).

The presence of self-cleaving ribozymes in intronic regions presents an exciting avenue of research with relevance to pre-mRNA processing and alternative splicing. A number of HHRs are found in introns of a conserved set of genes in amniotes (9) and the ultraconserved HDV-like ribozyme mapping to the *CPEB3* gene in mammals is also located in an intron (28). A number of studies have shown these ribozymes to be active *in vivo*. For the CPEB3 and R2 retrotransposon-associated HDV-like ribozymes, RT-qPCR analyses demonstrate that the levels of cleaved product vary by tissue or life-stage, respectively (28,29). These observations suggest that the catalytic activity of these ribozymes is regulated, lending additional support to the idea that ribozymes are

presenting an additional level of control for the expression of genes in which they reside. Interestingly, engineering self-cleaving ribozymes into introns or exons is a strategy used to understand co-transcriptional pre-mRNA processing (86). These studies suggest a kinetic competition between splicing and ribozyme cleavage, the latter of which can be disruptive to splicing (87), particularly if ribozyme catalytic rates are “fast” enough so as to interfere with spliceosome assembly (86).

In some natural cases, multiple ribozymes are found near each other in the same genomic locus (8,29,82), suggesting a cooperative influence on transcript processing or function. For example, some eukaryotic HHRs that are found associated with a particular family of retrotransposon, the Penelope-like elements (PLEs), display an organization suggestive of dimerization (88). The dimeric form, containing two sites of self-cleavage, is possible when HHRs occur in tandem and has been shown to lead to a more stable active structure compared to monomeric species of the ribozyme (89). In contrast, the discontinuous HHR present in the 3′ UTR of rodent *Clec2* genes is a single HHR in which invariant regions comprising the ribozyme core are separated by up to hundreds of nucleotides. The self-cleavage activity of these Clec2 HHRs removes the polyadenylation signal from the 3′ end of the mRNA, leading to a reduction in protein expression *in vivo* (14).

Very few known self-cleaving ribozymes are highly conserved amongst eukaryotic genomes. In rare instances, such as the *Clec2*-associated HHRs and the *CPEB3* HDV-like ribozymes, a single ribozyme sequence is conserved across multiple organisms (14,28), demonstrating that the self-scission activity was preserved and propagated through evolutionary lineages. Whereas many instances of HHR, HDV-like, and twister

ribozymes are found associated with retrotransposable elements (29,88,90), the extensive distribution of these ribozymes suggests they are common genomic features associated with multiple biological functions. While the field has gained a wealth of knowledge from studying independent examples of genetic elements harboring self-cleaving ribozymes, the biological significance of many other instances remains elusive.

The discovery of new self-cleaving ribozymes has transitioned from uncovering autocatalysis by single transcripts to high throughput bioinformatic approaches to identifying conserved RNA structures. As the power of computational and high-throughput RNA structure probing methods continues to grow, undoubtedly new complex and intricate RNA motifs will be revealed, among them possibly new families of self-cleaving ribozymes. *In vitro* selections from genomic and synthetic DNA pools suggest that there are other motifs capable of catalyzing self-scission and its reverse, ligation (28,91,92). In addition, ribozyme catalysis can be studied in a trans-cleaving format, where the bisection of secondary structure allows substrate recognition to regulate the rate of catalysis. The ease with which robust trans-cleaving ribozymes can be designed and constructed suggests that these intermolecular interactions might also exist in nature. Similar to ribozymes, riboswitches are typically cis-acting, controlling transcription, translation, or stability of the proximal mRNA. Characterization of the SAM riboswitch elements in the genome of *Listeria monocytogenes* identified the first RNA behaving as a riboswitch in cis and as a regulatory ncRNA in trans (93). This result highlights the possibility of functional RNAs such as ribozymes and riboswitches to be responsible for performing multiple tasks. In this regard, engineering of natural ribozymes has provided additional insight.

Finally, merging the ribozyme and riboswitch fields has created new tools to artificially control gene expression (94,95). Aptazymes are engineered RNAs composed of both a ligand binding domain and a ribozyme domain. They were first engineered as allosteric ribozymes via the fusion of an ATP aptamer and a HHR, whereby the sensing of ATP by the aptamer domain resulted in tuning of the HHR rate of catalysis (96). The ability to adapt ribozymes into unnatural and unprecedented functional platforms attests to the versatility of these RNAs (97). The functional flexibility of ribozymes and the dynamic range in catalytic rates makes these RNAs versatile tools in the synthetic biologist toolbox. Perhaps some of the tricks we ask ribozymes to perform *in vitro* or *in vivo* are really characteristics that nature already uses but remain for us to be discovered.

Chapter 2

Structure-based searches reveal hammerhead ribozymes in microbial genomes

Introduction

RNAs fulfill diverse biological roles, including regulation and catalysis. The discovery of catalytic RNAs over the last 30 years supports the RNA world hypothesis, which proposes that RNA predated proteins as the information carrier and the catalytic macromolecule (98). Biological catalytic RNAs include phosphotransferases and the ribosomal peptidyl transferase (99,100). Phosphotransferases include the six families of small ribozymes and additionally the group I intron-like ribozymes (101). Early work identified these self-cleaving RNAs through analysis of single gene transcripts and pathogen genomes (102). In addition, *in vitro* selection experiments have identified a variety of self-cleaving ribozymes (91), including the hammerhead motif, which was found independently several times (92) and the first HDV-like ribozyme found in mammalian genomes (28).

Hammerhead ribozymes (HHRs) have simple structural requirements, robust and tunable biochemical activity, and rather low information content, leading to the possibility of several independent appearances (92). The increasing prevalence of natural HHRs may imply multiple biological roles that evade detection through sequence alignment approaches. Considering that functional RNAs are generally conserved in secondary structure and not necessarily in sequence, they can be described as a set of base-paired domains of variable sequence connected by single-stranded regions. The unpaired strands are typically sequence-conserved in active sites, binding pockets, or tertiary contacts, and highly variable in connecting regions. A functional RNA, for

example a known ribozyme, can thus be described in terms of its secondary structure and conserved single-stranded elements (103,104). A computational search for sequences capable of assuming the same secondary structure and possessing all conserved sequences in the prescribed positions can thus reveal new examples of such RNA, even if its sequence is vastly different from previously identified members of the family. One application of this approach is in identification of known ribozymes in new genomes, as has been demonstrated for the HHR and HDV-like ribozymes (6,14,29,105,106).

The HHR structure consists of three helices (stems I, II, and III) anchored in the 11-nt catalytic core. They are capable of both self-scission and ligation, with ligation occurring at a rate about one hundred times slower than scission (107). In self-cleaving (*cis*-acting) HHRs, two of the stems are capped by loops, resulting in a topology resembling a hammerhead. HHRs can be grouped into three types according to the location of the open-ended stem containing the 5′ and 3′ ends of the ribozyme (types I, II, and III) (Fig. 2-1). Biochemical and crystallographic studies have revealed a tertiary interaction, which can enhance catalytic rates by up to 500-fold when compared to the minimal ribozyme under physiological conditions (17-19).

One powerful application of structure-based searches is in analysis of metagenomic data. These often originate from heterogeneous samples of potentially unknown biological composition. One of the benefits of metagenomic analysis is to monitor the genetic diversity and geographic presence of known or novel microbes. For example, numerous human viral pathogens have been detected in untreated sewage (108). Metagenomic data consist of a large number of short sequence reads that may lack

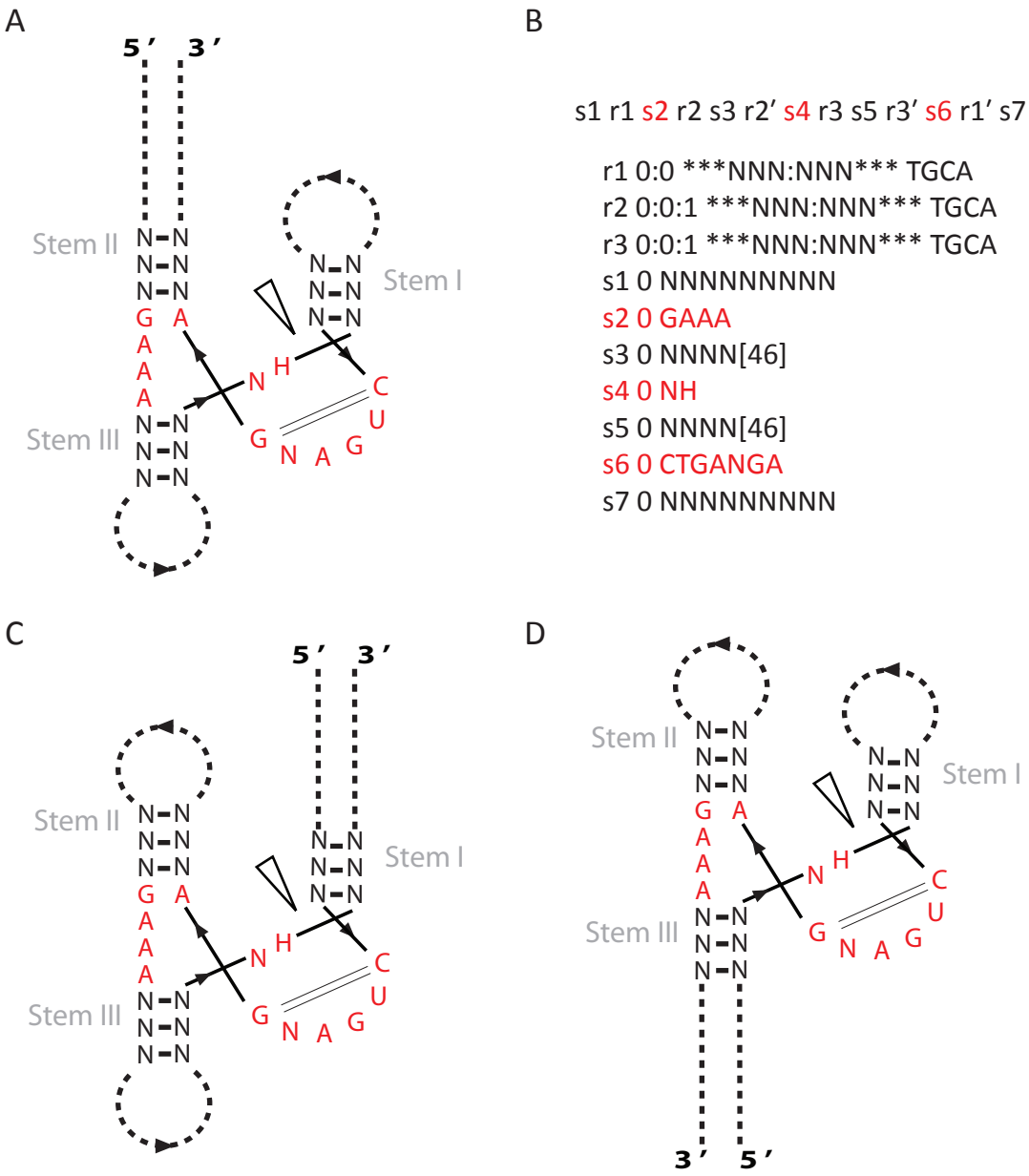


Figure 2-1. Descriptors defining the three possible topologies of the hammerhead ribozyme. (A) HHR type II secondary structure is color coded to indicate corresponding elements in (B) the descriptor used for the computational search. (C) HHR type I secondary structure. (D) HHR type III secondary structure. For each structure, the catalytic core is indicated in red. The open arrowhead indicates the site of self-cleavage. Small black arrowheads indicate directionality in joining single-stranded regions.

sufficient coverage for reconstruction of individual genomes. However, the datasets can be used to estimate a sample's species diversity or to propose the existence of novel species, depending on whether a read can be matched to a known genome. Both absence of genomic context and lack of similarity to known genomic sequences decrease the utility of such sequences, resulting in dearth of assignable function. To test such sequences for functional significance requires other approaches, one of which is a motif search for known structured, functional RNAs.

Given the strong base-pairing and relatively low sequence requirements of self-cleaving ribozymes, we decided to map the HDV and HHR motifs to the metagenomic sequences. While a search for HDV-like ribozymes did not reveal any sequences capable of folding into the double-pseudoknot fold, we found several sequences that can independently assume the three topologies of HHRs. We prepared the HHR constructs, flanked by several nucleotides of the surrounding sequences, and measured the *in vitro* self-cleavage activity. To test whether the incidence is higher or lower than would be predicted by chance, we performed structure-based searches for HHR through randomly-generated sequences of the same length (2.2×10^8 nt) and nucleotide content as our metagenomic data set, which contained about 8.15×10^5 sequences. The results of these searches agree with the predicted low distribution of HHR motifs in random sequences (103,105), providing a degree of significance to the incidence of HHRs in the metagenomic data set. To estimate the distribution of functional RNAs in these microbiomes, we used the HHR motif to search for sequences capable of assuming its three-way-junction fold. We searched through a set of metagenomic sequences from human samples including: stools from children with non-poliovirus

acute flaccid paralysis (109), cerebrospinal fluids from unexplained cases of encephalitis, plasma from unexplained cases of hepatitis, untreated sewage, attenuated viral vaccines (110), and stools from North American bats (111). While only two of the three possible natural HHR topologies had been known, our analysis revealed highly active ribozymes that terminated in any of the three stems. The most abundant of these are type II HHRs, one of which is the fastest natural *cis*-acting HHR yet discovered. Altogether, thirteen ribozymes were confirmed *in vitro*, but only one showed sequence similarity to previously described HHRs. Sequences surrounding the ribozymes do not generally show similarity to known genes, except in one case, where a ribozyme is immediately preceded by a bacterial *RadC* gene. We demonstrate that a structure-based search for a known functional RNA is a powerful tool for analysis of metagenomic datasets, complementing sequence alignments.

Motif searches identify type I, II, and III HHRs in diverse metagenomic sequences

Motif searches revealed a total of 20 candidate ribozymes of all three HHR topologies, of which 13 were confirmed to be active *in vitro* (Table 2-1). Our descriptors (Fig. 2-1) were similar to those used by Martick *et al.* to identify the *CLEC2* hammerhead ribozyme in rodents (14), except that we allowed for positions 7 and 17 of the catalytic core to be occupied by any nucleotide. Utilizing the rnabob algorithm, the permissiveness of a given descriptor can result in a larger output, but with a greater incidence of false positives (104). We therefore, used the RNAfold algorithm of the ViennaRNA package (112) to screen for correctly folding stem-loops. This filtering resulted in 17 candidates showing stable predicted folding. We found two constructs differing by a single point insertion in the flanking single stranded regions (hhrII-

SewR3_00868s-1 and hhrII-SewR3_00560s-1). The motifs reside in the same location within the context of their respective deep sequencing data reads; however, the two reads do not seem to vary in sequence composition (90 % identity) enough to rule out differences due to sequencing errors. In addition, hhrII-SewR3_02495s-1 and hhrII-SewR3_00810s-1 also reside in the same location; however, there is enough difference (73 % identity) between the reads to consider these two HHRs independent finds. All 20 original candidates were tested for *in vitro* catalytic activity.

Self-scission of microbial ribozymes

The catalytic activity of 13 novel HHRs identified through structure-based searches of deep-sequencing data from metagenomic sampling was verified through *in vitro* self-cleavage experiments. The observed cleavage rate constants are given in Table 2-1. Figure 2-2 shows a self-cleavage experiment for two type II HHRs in 10 mM Mg^{2+} under single-turnover conditions. The ribozymes cleave to about 80 %, with the remaining ~20 % representing the lower limit on cleavage yield. Ligation reactions of hammerhead ribozymes have been shown to be efficient *in vitro*, plateauing at ~23 % in *S. mansoni* HHR (107). However, self-scission reaction plateaus cannot be directly interpreted as equilibrium with ligation reactions, as the kinetics is complicated by the presence of inactive species. Future kinetics experiments will determine ligation rates for these type II HHRs.

All of the discovered ribozymes display robust catalytic activity *in vitro.* The cleavage kinetics data were fit to a single exponential decay and uncleaved residuals model, giving a rate constant for self-scission, $k_{obs,cleave}$, reported in Table 2-1. To examine the

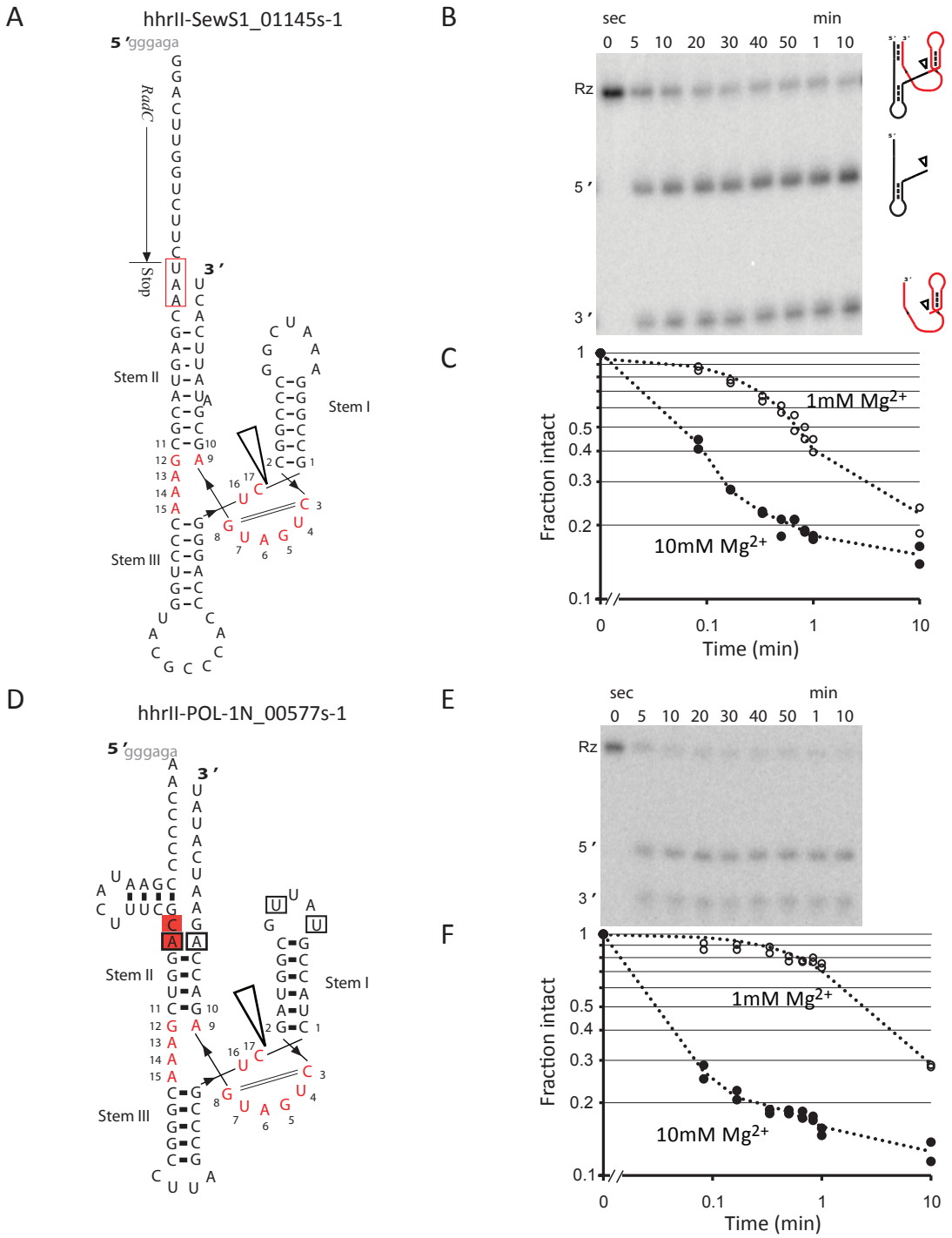


Figure 2-2. Secondary structure and activity of type II HHRs. (A) Predicted secondary structure of ribozyme hhII-SewS1_01145s-1. Catalytic core is numbered according to the nomenclature previously proposed and shown in red. The open arrowhead indicates the cleavage site and the solid arrowheads indicate the direction of the RNA strand. Predicted tertiary contacts between stems I and II are boxed. Conserved CA nts at the end of loop 2 are shaded orange. Synthetic sequences used to promote transcription are shown in grey lowercase at the beginning of most constructs. (B) *In vitro* self-cleavage activity of the ^{32}P-labeled ribozyme incubated in 10 mM $MgCl_2$ at 37˚C. (C) Log-log graph of ribozyme activities at 10 mM $MgCl_2$ (●) and 1 mM $MgCl_2$ (○), both at 37˚C. (D) The predicted secondary structure of ribozyme hhII-POL-1N_00577s-1. (E) *In vitro* self-cleavage activity, measured under the same conditions as (B). (F) Log-log graph of ribozyme activities at 37˚C.

Table 1. Catalytic rate constants of HHRs with confirmed *in vitro* activity.

Type	Identity	Isolate origin	k_{obs} (min^{-1}) 10 mM Mg^{2+}	1 mM Mg^{2+}
Type I	hhrI-SewB1_01643s-1	raw sewage	27.6 ± 0.1	1.7 ± 0.2
	hhrI-454-SetG2_00452-1	bat guano	20.60 ± 0.01	1.2 ± 0.1
	hhrI-pak6178-j_00497s-1	patient feces	7 ± 3	0.15 ± 0.02
Type II	hhrII-POL_IN_00577s-1	attenuated vaccine (Polio)	37 ± 2	0.54 ± 0.05
	hhrII-SewR3_00810s-1	raw sewage	25 ± 8	23 ± 9
	hhrII-SewS1_01145s-1	raw sewage	20 ± 2	1.4 ± 0.1
	hhrII-SewR3_02495s-1	raw sewage	18 ± 2	15 ± 0.01
	hhrII-OC1_0744s-1	patient plasma	13.3 ± 0.1	0.04 ± 0.16
	hhrII-SewR3_00868s-1	raw sewage	13 ± 3	0.10 ± 0.02
	hhrII-SewR3_00560s-1	raw sewage	11 ± 6	0.12 ± 0.02
Type III	hhrIII-ll-74_04911s-1	patient cerebrospinal fluid	21 ± 3	0.50 ± 0.02
	hhrIII-G3_26_03977s-1	patient cerebrospinal fluid	14.3 ± 0.7	0.74 ± 0.03
	hhrIII-SewR3_01770s-1	raw sewage	0.15 ± 0.01	0.008 ± 0.010

All experiments were performed at 140 mM KCl, 10 mM NaCl, 10 mM Tris-HCl, pH 7.4, and incubated at 37˚C at the indicated Mg^{2+} concentration. The reported k_{obs} values are the average rate constants ± the average deviation of at least two *in vitro* measurements. *Fit to bi-exponential decay and uncleaved residuals model. All other data were fit to a single exponential and uncleaved residuals model.

ribozyme activity at physiological-like conditions, self-cleavage experiments were also carried out at 1 mM Mg^{2+}.

Comparisons to known HHRs

Tertiary interactions between peripheral domains have previously been shown to greatly affect the *in vitro* activity of HHRs. For example, the presence of tertiary interactions in the type I *S. mansoni* HHR is known to enhance the rate of scission by 50 to 500-fold under physiological-like conditions (113,114). To allow for loop-loop contacts, all of our constructs were designed to include at least 9 nucleotides upstream and downstream of the ribozyme motif. It has previously been suggested that ribozymes with different metal dependencies experience differences in the rate-limiting step, with tertiary stabilizations placing the molecule further along the reaction pathway (115). Many of our ribozymes exhibit fast self-scission, suggesting that their catalytic cores are stabilized by tertiary interactions. In addition, although the kinetics experiments were performed at only two Mg^{2+} concentrations, the 13 HHRs displayed a range of Mg^{2+} dependencies, with a ten-fold difference in Mg^{2+} concentration resulting in a ~300 times enhanced cleavage in the case of hhrII-OC1_0744s-1 (Table 2-1). This large cooperativity also suggests that at 10 mM Mg^{2+} the ribozyme activity is far from maximal. Interestingly, both hhrII-OC1_0744s-1 and hhrII-POL_IN_00577s-1, show high Mg^{2+} dependence and fast kinetics, while neither is predicted by RNAfold (112) to fold into the correct HHR structure within the context of their respective ~500 nt genomic sequence read. All the remaining ribozymes are predicted to form a stable catalytic core even when embedded within their full sequence reads.

Many naturally occurring HHRs contain conserved motifs in loops 1 and 2, which have been identified as essential features for loop-loop interactions (116). Base pairing between the 5′ U of loop 1 and the 3′-A of loop 2 as well as the interaction of an extrahelical (penultimate) pyrimidine of loop 1 with a 5′ purine of loop 2 can be predicted for 7 of our HHRs (Fig. 2-2 and 2-3, boxed). Two others display the latter interaction only. In addition, 4 of the HHRs predicted to have tertiary interactions, contained the last two CA dinucleotides of loop 2, which are conserved in chrysanthemum chlorotic mottle viroid (CChMVd), but also occur in other natural HHRs (116). These two interactions are predicted to be conserved in over 20 naturally occurring HHRs including those of insects, *Arabidopsis*, CChMVd, and tobacco ringspot virus satellite RNA (116).

Several of our ribozymes did not contain conserved loop sequences involved in known tertiary interactions. To predict novel tertiary contacts, we used the DotKnot program (117) to detect potential pseudoknots and kissing loops, such as those known to exist for the *S. mansoni* HHR (113). DotKnot predicted a pseudoknot between loop 1 and the flanking sequence of stem II in hhrII-SewR3_00810s-1 and hhrII-SewR3_02495s-1 (Fig. 2-3). This type of tertiary interaction has not previously been proposed in an HHR and points to a potential diversity of tertiary contacts that can stabilize peripheral regions in these ribozymes. These two related ribozymes exhibit the highest rate constants at physiological Mg^{2+} (1 mM) and vanishing Mg^{2+} dependence between 1 and 10 mM (Table 2-1). This result supports the existence of a strong, Mg^{2+}-independent tertiary interaction, which may be achieved by base-pairing in a pseudoknot. Overall, there is no apparent trend in the catalytic rates between our HHRs

that do and do not display predictable tertiary contacts, further supporting the hypothesis that novel tertiary interactions stabilize the fast-cleaving ribozymes.

Among the 13 ribozymes, 10 cleave a UC sequence, which is identical to the *S. mansoni* HHR. This sequence is known to have higher activity than other combinations of nucleotides in engineered hammerhead motifs (118). Many other known viral and viroid HHRs, as well as the recently discovered bacterial and eukaryotic HHRs possess UC or UA at their cleavage sites (8,9,119). The sequence conservation of this site suggests that high cleavage activity serves a biologically important function.

Considering that our descriptors allowed large stem-loops to occur, the active HHRs are relatively small, with stem lengths averaging between 3 and 7 bps and loops between 4 and 8 nts (Fig. 2-2 and 2-3). Some of the common features include long stem I (6-8 nts) and short stem II (3-4 nts), with the exception of hhrII-SewS1_01145s-1, which has a predicted 8 bp stem II. These lengths follow closely with stem and loop sizes seen in over 20 other naturally occurring HHRs (116,119).

The remarkable catalytic rates for our fastest HHRs are greater than the rates of other natural HHRs tested in the same format (*cis*) and similar reaction conditions. One of the most extensively studied natural *cis*-HHRs is the *S. mansoni* ribozyme with a k_{obs}=0.22-0.36 min^{-1} (30˚C, 10 mM Mg^{2+}, 40 mM Tris, pH 8) (6). To compare cleavage kinetics under identical conditions, we designed construct hhrI-Sman after the *S. mansoni* ribozyme. Our hhrI-Sman construct displayed k_{obs}=6.7±0.2 min^{-1} (10 mM Mg^{2+}, 10 mM Tris, pH 7.4). However, it should be noted that a chimera of arabis mosaic virus small satellite RNA (satArMV+) in *cis*-format displayed k_{obs} > 750 min^{-1} (25˚C, 10 mM Mg^{2+}, 50 mM MES buffer, pH 7.3) (119). Although the diversity of reaction conditions

and HHR constructs throughout the literature prevent direct comparison, in the natural *cis*-format, 12 of our HHRs are among the fastest reported to date.

HHR incidence in a randomly generated sequence file

To estimate whether the incidence of the HHRs in the metagenomic data is higher or lower than would be expected by chance, we generated a random sequence of identical length and nucleotide content as our metagenomic sequence file. Given the low probability of finding a hammerhead motif (1 per 10^{13} nt) (6), we would not expect to find stable HHRs in this random sequence, and indeed this is the case. Structure-based searches produced 4 times fewer hits, all of which failed to be confirmed by subsequent secondary structure prediction. By comparison, the HHR motif is over-represented in the 2.2×10^{8} nt metagenomic sequences and the occurrences are unevenly distributed among the sequenced samples. This result implies that HHRs are highly selected for in certain microbiomes and suggests that the samples contain novel HHR-harboring microbes.

The catalytic core of the hammerhead ribozyme can accommodate additional nucleotides between stems I and II. HHRs with this extended core sequence have previously been identified through bioinformatics searches (120). To identify such ribozymes, additional structure-based searches allowing extended core regions were performed with our metagenomic sequences. Three output sequences representing extended core motifs were predicted to have stable secondary structure and were tested for *in vitro* self-scission; however, none displayed catalytic activity in these assays.

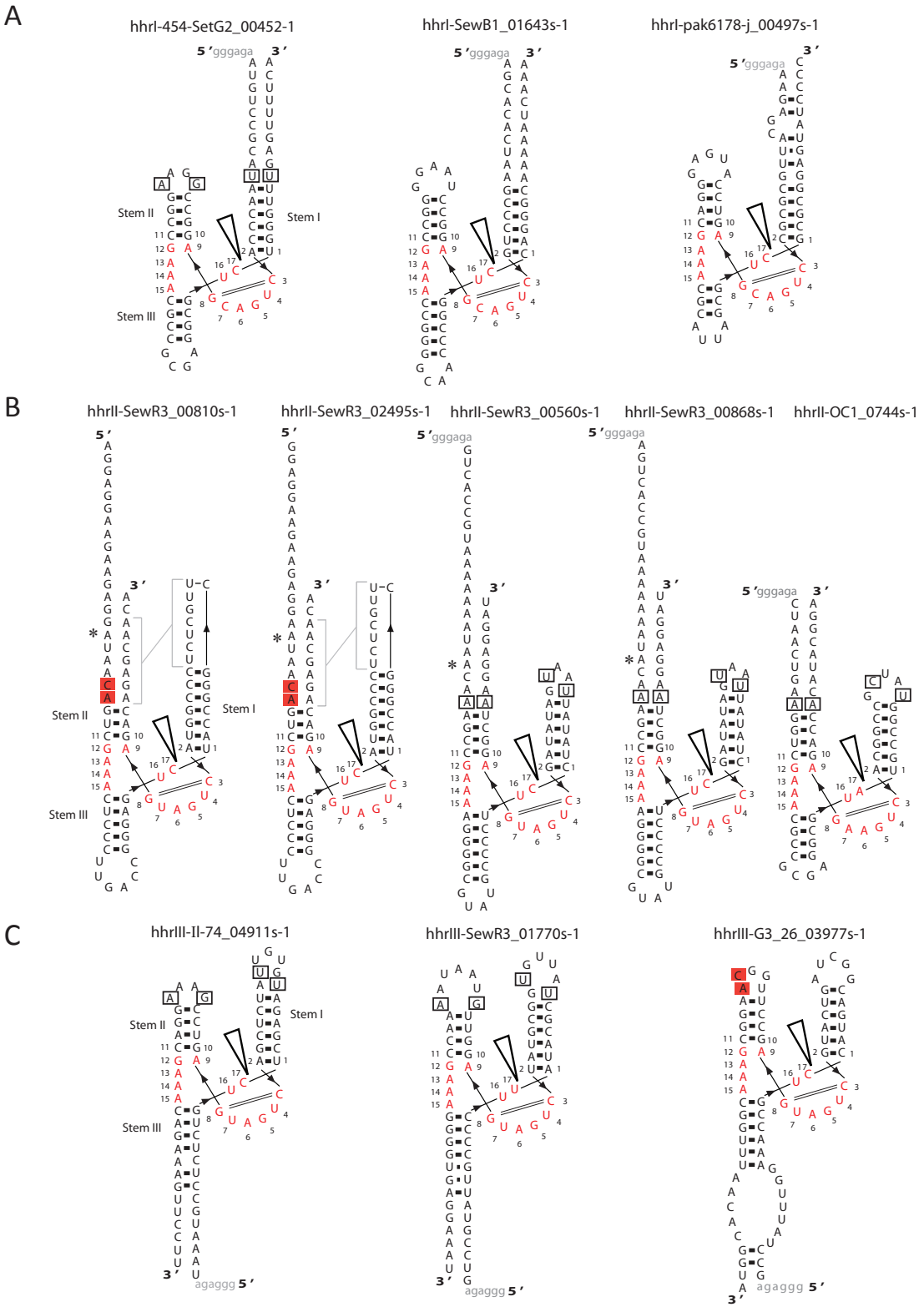


Figure 2-3. HHRs discovered in a metagenomic dataset. Secondary structure of type I (A), II (B), and III (C) HHRs with *in vitro* cleavage rate constants reported in Table 1. All labels and symbols are the same as in Fig. 1. Asterisks indicate differences between ribozymes hhrII-SewR3_00810s-1 and hhrII-SewR3_02495s-1, and between ribozymes hhrII-SewR3_00560s-1 and hhrII-SewR3_00868s-1. Grey brackets indicate the putative pseudoknot interactions between loops 1 and 2 in hhrII-SewR3_00810s-1 and hhrII-SewR3_02495s-1.

Discussion

Our work shows that HHRs are highly enriched in a diverse set of subject and environmental samples, including human stool and sewage. Identifying structured RNAs is one way to characterize new genomic information and categorize genetic variants or novel enteric viruses. Many of our metagenomic sequences did not align closely with any known genomes. While the origin of the sample isolates is known, further investigation is needed to identify the viral or bacterial isolate origin of these ribozymes. Two ribozymes were found to bear sequence similarity to previously analyzed HHRs. Of the 13 HHRs, BLASTn (121) revealed only one HHR similar to a genomic ribozyme: hhrIII-ll-74_04911s-1 has a high sequence similarity to the hammerhead ribozymes studied in many viroid-like RNAs and virus satellite RNAs, particularly chicory yellow mottle virus satellite RNA (sChYMV) (E score 3e-4). It has previously been suggested that viroids and satellite RNAs originated from the transcripts of repetitive sequences when the transcripts parasitize viral replication machinery and use viruses as vectors to jump from one organism to another (105). The occurrence of HHRs in repetitive DNA of many species suggests that the catalytic activity is important to the existence of these transcripts. hhrIII-ll-74_04911s-1 may represent another example of such activity, although its flanking sequences do not map to known satellite sequences.

Interestingly, hhrI-454-SetG2_00452-1, isolated from bat guano, bears 89 % identity to an artificial RNA construct used to make a 3D model of HHR based on FRET measurements (GenBank accession 1RMN_A) (122,123). It exhibits fast kinetics (Table 2-1) and contains both predicted tertiary interactions between stems I and II.

In the case of hhrII-SewS1_011455-1, BLAST searches did not identify another ribozyme with similar sequence, but they did reveal that the upstream sequence is highly similar to the *RadC* gene (RadC is a DNA repair protein) of the thermophilic photosynthetic bacterium *Roseiflexus castenholzii DSM* (E value 7e-13). The 5′ end of the HHR construct contains a UAA stop codon, which corresponds to the end of the *RadC* gene (Fig. 2-2). The *R.castenholzii RadC* gene does not have a hammerhead motif at its 3′ end, and a search through its entire genome (GenBank CP000804) did not reveal any HHRs, suggesting that our isolate represents a novel genomic arrangement of *RadC* and HHR. The close proximity of the ribozyme to the 3′ terminus of the *RadC* coding region suggests that the ribozyme is either involved in mRNA processing or that it forms a 5′ terminus of a novel genetic element close enough to the *RadC* gene to form a single transcriptional unit with the mRNA.

Conclusions and future directions

It will be interesting to learn whether the viral or bacterial sources of these 13 ribozymes display pathogenicity or are merely commensal microbes in the human microbiome. The presence of the ribozymes identified on the RNA strands of yet uncharacterized viral genomes will require further studies. The sequence reads containing these ribozymes will be extended by overlapping with other unclassifiable reads, potentially forming contigs that can be analyzed for signatures of viral genomes. Deeper sequencing of those microbiomes harboring our ribozymes is needed in order to assemble new microbial genomes. Given that metagenomic sampling can be used to monitor genetic variation over time and location, functional HHRs may be evolutionarily important to these viral or bacterial genomes. In general, although the biological

functions of all HHRs have not been identified, the association with satellites and RT ORFs suggests a role in retrotransposition (8) which has been suggested for HDV ribozymes as well (29,33).

We continued using structure-based searching programs rnabob and rnarobo to find new HHRs. Structure descriptors, similar to that given in Fig. 2-1, were used to search through other available genomes, including the yeast *Yarrowia lipolytica* CLIB122 (NC_006067.1-NC_006072.1) genome (124). There are several type-I HHRs known to exist in this genome (10). Both rnabob and rnarobo found a single HHR (Yli-1-3 position 1037945-1037850) using a descriptor with a strict requirement for at least four base pairs in each helix. A rnarobo search using the same descriptor, but allowing single-nucleotide insertions in each of the three helices, increased the number of unique hits from one to fifteen. Among these hits were Yli-1-3, and also Yli-1-4 through Yli-1-11 ribozymes (4). The search allowing insertions, utilizing rnarobo, was necessary to finding the eight additional ribozymes and this data attest to the power of the new rnarobo algorithm to find structured RNAs, as single nucleotide insertions in helices are common features in natural RNAs.

Decreasing the requirement in base pairing of helices from four to three base-pairs resulted in a total of four hits from both rnabob and rnarobo searches (hhr1-3bp), including Yli-1-3 and Yli1-13, the latter of which was not found when helices were required to be at least four base pairs long. Using rnarobo and allowing for single nucleotide insertions in any of the three helices of the HHR significantly increased the number of hits returned by rnarobo from four to fifty-four, including the same known ribozymes (Yli-1-3 through Yli-1-11 and Yli-1-13). A sequence alignment of the output

reveals that many sequences can be grouped into two main families. The first family includes the known “Yli” ribozymes, and the second includes previously unidentified putative ribozymes. Further analysis of this second family was done using the fold filter script included with the rnarobo-2.0 package (125).

Each of the ten sequences in the second family, although not necessarily unique in sequence, is unique in the genomic location, making each of these ten putative ribozymes independent findings. It is very likely that these putative ribozymes will demonstrate activity *in vitro*. In this instance, rnarobo may have found a previously unidentified family of type I HHRs in the *Y. lipolytica* genome. This is significant in that no other sequence conserved family of HHRs is known to exist in this genome and is similar to the large family of HHRs in the *Schistosoma mansoni* genome (6,10) and also the discontinuous HHRs found in mammalian *Clec2* and *Clec2*-like genes (14). Future work should verify that these new findings are *bona fide* ribozymes.

In this study, structure-based searches revealed fast natural HHRs, the first natural type II HHRs, and a putative novel pseudoknot stabilizing the most robust type II ribozymes. Pseudoknots stabilizing HHR tertiary contacts have been confirmed in more recent HHR discoveries (10). In the case of metagenomic sequences, identifying the presence of catalytic motifs can be used to evaluate the distribution of functional RNAs in a sample and possibly assist in the identification of novel viruses. In addition, we uncovered a putative new family of HHRs in the genome of the yeast *Y. lipolytica*. The expansion of coding and non-coding RNAs libraries will ultimately help characterize new genomes. The discovery of new ribozymes will further help us understand their distribution in nature and their biological importance.

Chapter 3

Discovery of bacterial HDV-like ribozymes sensing sugar metabolites and the implications in gene expression

Introduction

Ribonucleic acids have been assigned diverse biological roles, including silencing (siRNA), translation inhibition (miRNA), metabolite-dependent gene regulation (riboswitches), and numerous catalytic activities (ribozymes) such as in protein synthesis (rRNA) and viral genome processing. Among ribozymes there are six families of self-cleaving ribozymes fulfilling diverse biological roles throughout all branches of life. In general the full distributions and biological functions of the most widespread ribozyme families are not well defined. On the contrary, there are currently 24 families of riboswitches governing the expression of genes that are specific to each family. In bacteria, riboswitches regulate gene expression at the transcriptional or translational level in response to cellular ligand concentrations (2 % of *Bacillus subtilis* genes are regulated by riboswitches) (126). The glmS ribozyme family is also a riboswitch family, turning off the expression of the *glmS* gene in response to glucosamine 6-phosphate (74). Given the expansive distribution of certain catalytic RNAs and the predominate use of riboswitches to control bacterial gene expression, we expect these non-coding RNAs to exist in a higher abundance than is currently known. Only a limited number of self-cleaving ribozymes have been linked to a biological function. The family of hepatitis delta virus (HDV)-like ribozymes was first discovered for their role in rolling circle replication of the virus genome (27). The discovery of the first HDV-like ribozyme in

mammals (28) motivated the search for the nested double-pseudoknot structural motif throughout available genome databases.

The conserved secondary structure of HDV-like ribozymes was used to uncover the extensive distribution of this ribozyme family through all kingdoms of life (29). Considering the multitude of available bacterial genomes, the HDV-like ribozyme is significantly underrepresented. A single ribozyme (drz-Fpra-1) was found in bacteria and maps immediately upstream of the *glmM* gene in the genome of *Faecalibacterium prausnitzii*, a commensal bacterium found in the human gut (29,127). The proximity of this ribozyme suggests it may be involved in regulation of the downstream ORF. While a role for HDV-like ribozymes in expression of downstream ORFs has previously been shown in eukaryotic systems (32), we hypothesize that in bacteria regulation by the ribozyme is responsive to metabolites, similar to the mode of regulation by riboswitches.

There are two types of self-cleaving ribozymes known to respond to fluctuations in cellular levels of specific metabolites: group I intron-like (GIR) and *glmS* ribozymes. The activity of all GIRs requires GTP as a cofactor for self-scission or self-splicing activity (101,128). Recently, a GIR with regulated activity was discovered in bacteria. The activity is triggered by an allosteric change that promotes the correct structural conformation necessary for self-scission to occur, while still utilizing GTP as a cofactor (129). This allosteric ribozyme was later verified to be part of a tandem riboswitch-ribozyme complex in which the activity of the ribozyme is regulated by the binding activity of the upstream c-di-GMP aptamer domain of a class II c-di-GMP riboswitch which governs a splice site junction and affects expression of the downstream gene (130).

The glmS ribozyme maps to the 5′ UTR of the *glmS* gene in many Gram-positive bacteria genomes (74). The *glmS* gene codes for glutamine:fructose 6-phosphate aminotransferase, the enzyme responsible for glucosamine 6-phosphate (glcN6P) synthesis. This self-cleaving ribozyme is only active in the presence of its glcN6P cofactor that is necessary for its catalysis (77,79,81,131). The scission of the phosphodiester backbone by the ribozyme makes the message a target for degradation by RNase J (131). Therefore the glmS ribozyme acts as a riboswitch by reducing expression of the glmS enzyme when glcN6P levels are elevated (74). Contrary to the glmS ribozyme relying on a cofactor for catalysis, HDV-like ribozymes are metalloenzymes under biological conditions (42).

Multiple lines of evidence show that HDV-like ribozyme activity is sensitive to the cellular environment. For HDV-like ribozymes mapping to the genome of the mosquito *Anopheles gambiae*, RT-qPCR experiments reveal the extent of self-cleavage varies by sex and life stage (29). The HDV-like ribozyme mapping to an intron of the cytoplasmic polyadenylation element binding protein 3 (*CPEB3*) gene in mammals displays levels of self-cleavage that vary by tissue type in mouse (28). The mechanisms behind this observed environmental sensitivity have yet to be elucidated. The crystal structure of the HDV ribozyme reveals a compact architecture where the scissile phosphate and catalytic cytosine nucleobase are “tucked” within the active site (38-40). The hydrated metal cation coordinated in the active site serves to neutralize the developing negative charge on the nucleophile (42) but may also be involved in transition-state stabilization (43). The complexity of the active site limits the ability to accommodate small molecules, making it unlikely for ligands to have a role in the catalytic mechanism. If a small

molecule were to have affinity for the HDV ribozyme active site, it would likely be detrimental to catalysis. Though it is unlikely that a population of HDV-like ribozymes evolved to rely on some small molecule cofactor for self-cleavage, it is possible that environmental sensitivity involves an allosteric mechanism. In this situation, small molecules can influence folding of the ribozyme, thereby enhancing the rate of catalysis.

The HDV-like ribozyme drz-Fpra-1 maps to the 5′ end of the *glmM* gene, coding for phosphoglucosamine mutase. This mutase is an isomerase that converts glucosamine 6-phosphate (glcN6P) to glucosamine 1-phosphate (glcN1P). This process is part of the hexosamine biosynthetic pathway that generates precursors for the synthesis of various macromolecules in both bacteria and eukaryotes. The final product UDP-N-acetyl D-glucosamine (UDP-GlcNAc) is critical to bacterial cell wall synthesis, and the biosynthetic pathway is a major target in the development of antibiotics. The major mechanism of regulating flux of this pathway differs between Gram-positive and Gram-negative bacteria, but both focus on controlling the conversion of fructose 6-phosphate to glucosamine 6-phosphate by the L-glutamine D-fructose-6-phosphate aminotransferase (glmS synthase) enzyme. In *E. coli*, levels of glcN6P are mediated via small trans-encoded RNA (sRNA) glmY and glmZ (132). GlmZ activates translation of *glmS* mRNA by base-pairing, providing stabilization, and the cellular levels of glmZ are tightly controlled by RNase E which requires the protein factor RapZ. When glcN6P is limiting, glmY acts a decoy, occupying RapZ, allowing glmZ to drive translation of the *glmS* mRNA. The levels of glmY increase with a decrease in cellular levels of glcN6P; however, the mechanism of this accumulation is still a mystery. Whereas in *E. coli*, homeostasis of glcN6P follows the pattern of glmZ levels, Gram-positive bacteria use a

more direct feedback mechanism. In most Gram-positive bacteria, the major regulatory step is controlled by the glmS ribozyme/riboswitch (74). Of the many bacterial genomes harboring glmS ribozymes, the genome containing the HDV-like ribozymes (*F. prausnitzii* SL3/3 draft genome) do not.

While the collection of each ribozyme family is ever expanding, it is possible that a structural variant of the glmS ribozyme might exist in the *F. prausnitzii* SL3/3 genome, especially since two other *F. prausnitzii* genomes do contain the glmS ribozyme (133). The alternative is that instead of a glmS riboswitch regulating flux of the hexosamine biosynthetic pathway, another type of self-cleaving ribozyme is fulfilling that role. The *F. prausnitzii* HDV-like ribozymes represent a unique arrangement in which the self-cleavage activity may facilitate regulation of the *glmM* gene. The *F. prausnitzii* bacterium is a firmicute that Gram stains negative despite phylogenetically belonging to the class *Clostridia* which are mainly Gram positive (134). Little is known about transcription of the *glmM* gene in *F. prausnitzii;* however, Northern blot analysis of the operon containing the *glmM* gene in *E. coli* and *P. aeroginosa* suggests the gene is cotranscribed as part of a 2, 3 or 5-gene containing transcript (135). It is possible that the HDV-like ribozymes found surrounding the *glmM* gene in *F. prausnitzii* serve to liberate the glmM message from a polycistronic transcript. This processing would likely have repercussions on the expression of the message.

We hypothesize that bacterial HDV-like ribozymes may recognize a cellular signal, which affects the lifetime of the transcript or its translation, thereby regulating expression in a manner dependent on the concentration of the signal in the cell. Our findings suggest that HDV-like ribozymes may be able to recognize small molecules

that affect their activity. In addition, the catalytic activity of the ribozyme is shown to influence the expression of a downstream ORF in *E. coli*.

Discovery of new HDV-like ribozymes in bacterial genomes

Non-coding RNAs were traditionally found by careful analysis of genes, viral genomes, and transcriptome sequencing. Subsequent *in vitro* structural analyses of these ncRNAs is very low-throughput, analyzing only single sequences of up to only a few hundred bases at a time. This makes it difficult to determine the biological impact of structured RNAs on a genome-wide scale. In addition, the growth of sequence databases demands high-throughput computational RNA prediction tools to help expand the field of comparative genomics. Algorithms have been developed which can identify RNA structure based on sequence and sequence covariation. Currently, there are a variety of software packages available to detect new instances of known structured RNAs or to discover new structural motifs. Programs employing structure-based searches are able to identify new instances of conserved structural motifs.

Structure-based searching programs and other descriptor-based tools specify important features of a known motif in the descriptor, allowing any functional RNA to be described by its secondary structure and conserved single-stranded elements (103,104) (Fig. 3-1A). The descriptor file describes the occurrence of helical or single-stranded elements in a linear space (5′ - 3′) as well as the specific sequence or length requirements each individual element must meet (Fig. 3-1B). Sequence data are then threaded through the descriptor and any sequence matching the designated secondary structure is returned as a putative hit. This technique has led to the discovery of HHRs and HDV-like ribozymes in many new genomes (11,12,29,90).

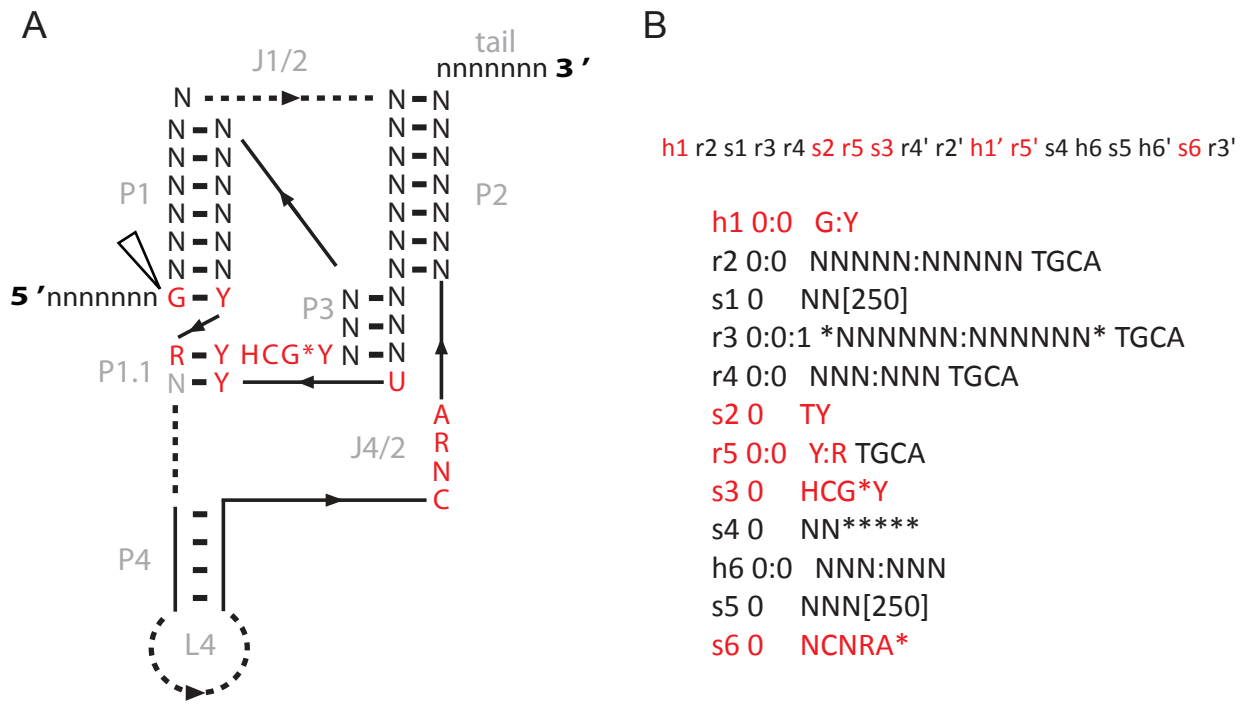


Figure 3-1. A structure descriptor for HDV-like ribozymes. (A) The secondary structure of HDV-like ribozymes shows the conserved nucleotides necessary for catalytic activity in red, coordinating with (B) the structure descriptor for use with the searching algorithm rnarobo. The red asterisk in the secondary structure signifies the permissivity for a single nucleotide insertion in known HDV-like ribozymes. The site of self-cleavage is indicated by the open arrowhead. The names of structural features are given.

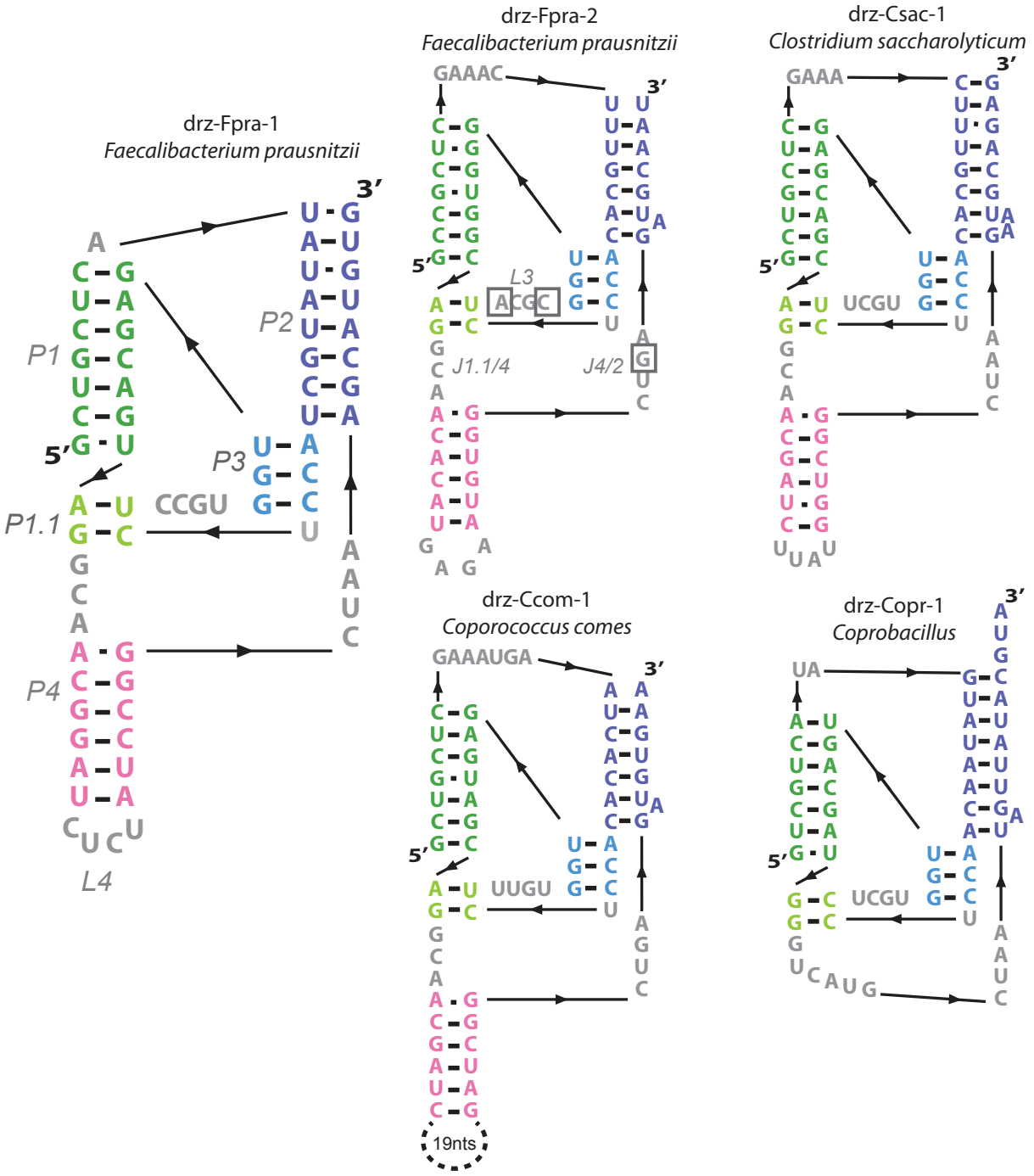


Figure 3-2. Secondary structures of known bacterial HDV-like ribozymes. Boxed nucleotides on the drz-Fpra-2 ribozyme indicate positions that differ from drz-Fpra-1 in the HDV ribozyme core. The ribozyme name follows the nomenclature previously described where "drz" refers to the family of HDV-like ribozymes and is followed by an abbreviation of the organism name (*Faecalibacterium prausnitzii*=Fpra) and a number to distinguish multiple ribozymes in the same genome.

Previous work using structure-based searches identified a HDV-like ribozyme (drz-Fpra-1-1, Fig. 3-2) immediately upstream of the *glmM* gene in the genome of the bacterium *F. prausnitzii* (29) and it was the only known example of a bacterial genome harboring a HDV-like ribozyme. Subsequent searches, utilizing a descriptor (Fig. 3-1B) used with the RNArobo software package, identified a number of putative HDV-like ribozymes. Remarkably, one of these new ribozymes (drz-Fpra-2) mapped to the same genomic locus as drz-Fpra-1, located immediately downstream of the same *glmM* gene in the *F. prausnitzii* genome.

We investigated the incidence of HDV-like ribozymes at the 5′ and 3′ ends of *glmM* genes in other bacterial genomes however no sequences matched the canonical nested double-pseudoknot of HDV-like ribozymes. Putative ribozymes upstream of the *glmM* gene can also be found in the *F. prausnitzii* M21/2 and L2/6 genomes, but they have not been validated *in vitro*. Sequence-based searches for *F. prausnitzii*-like ribozymes resulted in tens of matches with good Expect (E) values (lower than 0.01 %). By utilizing a computational pipeline published by our group (125,136), structure-based searches and sequence homology identified three new ribozymes mapping to three different bacterial genomes.

Five HDV-like ribozyme constructs, including any stable structured elements immediately 5′ to the ribozyme, were tested for catalytic activity from purified RNA generated by *in vitro* transcription (Table 3-1). One construct from *Eubacterium rectale* is identical in sequence to *Coprococcus comes* ribozyme but with a G-A mismatch at the base of the P3 helix; this mutation caused the construct to be inactive under the experimental conditions. In addition, the ribozyme mapping to the genome of

Table 3-1. Genome, location, and *in vitro* activity of bacterial HDV-like ribozymes.

Ribozyme	Organism	Genomic Location	Purified RNA k_{obs} (min^{-1})		Cotranscription k_{obs} (min^{-1})	k_{rel}
			10 mM Mg^{2+}	1 mM Mg^{2+}	1 mM Mg^{2+}	10 mM glcN6P
drz-Fpra-1	*Faecalibacterium prausnitzii*	upstream *glmM*	0.9*	0.57	0.37 ± 0.05	1.26
drz-Fpra-2	*Faecalibacterium prausnitzii*	downstream *glmM*	1.7 ± 0.3	0.01	0.035 ± 0.020	0.50
drz-Csac	*Clostridium saccharolyticum*	upstream hyp protein	0.50 ± 0.05	0.21	0.195 ± 0.075	0.46
drz-Ccom	*Coprococcus comes*	upstream hyp protein	0.12	-	-	-
drz-Copr	*Coprobacillus*	-	2.2	0.75	0.23 ± 0.01†	-

* Rate determined by lab member Webb, C.H. †Rate determined by lab member Riccitelli, N.

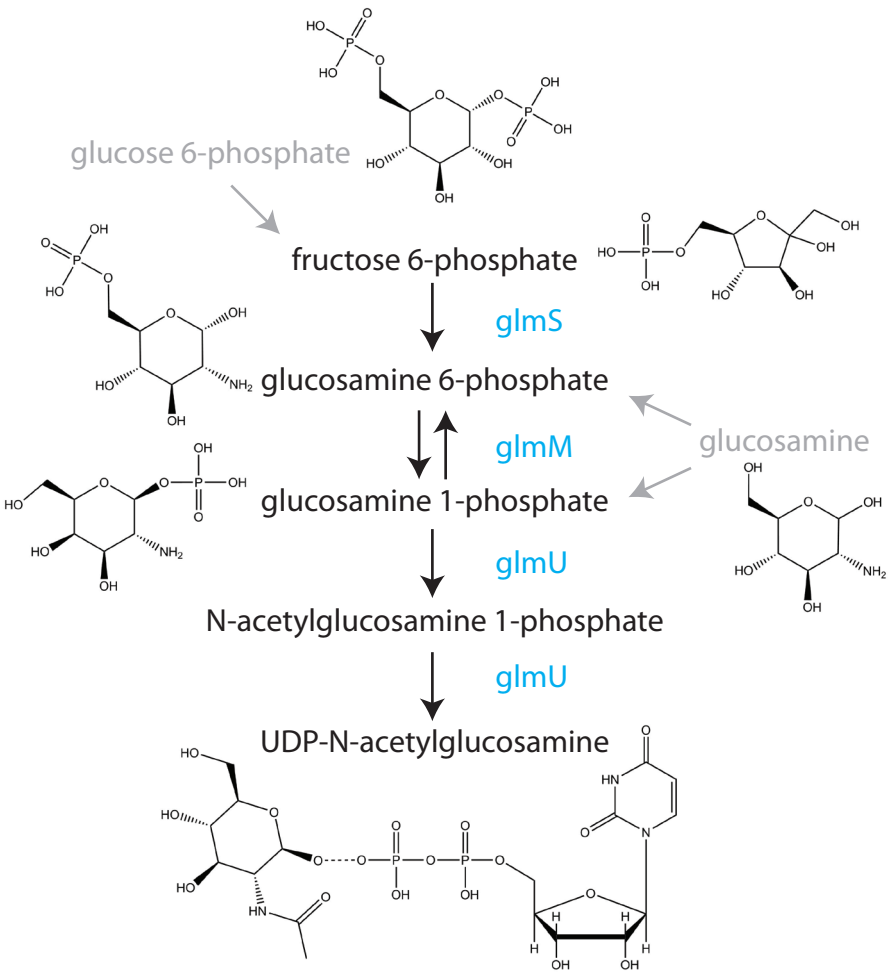


Figure 3-3. The hexosamine biosynthetic pathway in bacteria. This pathway leads to the production of uridine-diphosphate N-acetylglucosamine, an important precursor in bacterial cell wall synthesis. Other sugars indirectly contribute to flux through this biosynthetic pathway: glucose 6-phosphate is converted to fructose 6-phosphate and glucosamine gets phosphorylated into either glucosamine 6-phosphate or glucosamine 1-phosphate. The structures for the sugars used during i*n vitro* self-cleavage assays are given.

Table 3-2. Metabolites used during *in vitro* self-cleavage assays.

Metabolite	Significance
glucosamine 1-phosphate (glcN1P)	the product of the isomerization reaction catalyzed by the glmM enzyme
glucosamine (glcN)	a structural analog missing both phosphate and amine groups
fructose 6-phosphate (fruc6P)	a precursor of glcN6P
glucose 1,6-bisphosphate (glc1,6-bP)	a structural analog of the intermediate in the isomerization between glcN6P and glcN1P
uridine diphosphate *N*-acetylglucosamine (UDP-N-AcglcN)	the end product of the hexosamine biosynthetic pathway

The name of the metabolite is given with its abbreviation. The significance for the use of the metabolite in self-cleavage assays pertains to a function in the hexosamine biosynthetic pathway or as a structural analog of sugars that are part of the pathway.

Coprobacillus was identified by structure-based searches for minimal HDV-like ribozymes, which also uncovered numerous other ribozymes in microbial genomes found in metagenomic samples (137). The new ribozymes (Fig. 3-2) described here are not located near the *glmM* genes in the respective bacterial genomes; however, there is a relatively high degree of sequence similarity to the original *F. prausnitzii* ribozyme.

Cotranscriptional self-scission of bacterial HDV-like ribozymes in vitro

In order to gain an appreciation for how these ribozymes might behave in a cellular environment, an *in vitro* cotranscriptional assay was employed. The intention is to observe how these RNAs behave as the nascent transcript is being generated and folded. This approach avoids the denaturation steps that occur during a typical RNA gel purification. The observed rates of cleavage determined cotranscriptionally are predicted to be different than previous k_{obs} determined for purified RNA, as gel purification methods create biases towards a slower behaving population. Cotranscriptional assays have been described for self-cleaving ribozymes (138,139) and they have been shown to reveal fast-reacting species that are lost during extensive handling of *in vitro* transcribed RNA (140). The rates of self-cleavage under cotranscriptional conditions for all the bacterial ribozymes are shown in Table 3-1.

In addition to observing native catalytic rates under physiological-like conditions *in vitro*, the sensitivity of the HDV-like ribozymes to small molecules was also tested. The small molecules were selected based on their relevance to the glmM enzyme and the hexosamine biosynthetic pathway (Fig. 3-3). Glucosamine 6-phosphate (glcN6P) is the substrate of the glmM enzyme and is the same small molecule utilized by the glmS ribozyme as a cofactor, eliciting ribozyme activity and down-regulation. Glucose 1-

phosphate is a structural analog, lacking the amine moiety. Other HDV-like ribozymes were used as controls for self-cleavage. The antigenomic HDV (aHDV), and HDV-like ribozymes from mouse *Mus musculus* (drz-Mmus CPEB3) and purple sea urchin *Strongylocentrotus purpuratus* (drz-Spur-3) have well-characterized biochemical activity *in vitro* and under cotranscriptional conditions. In addition, these ribozymes are not known to interact with any small molecules. The sensitivity of all the ribozymes to the small molecules is shown in Fig. 3-4.

While most ribozymes tested display slight sensitivity to the metabolites, the concentrations of the metabolites used in the assay are higher than the calculated physiological concentrations in *E. coli*. For example, the calculated intracellular concentration of glcN6P is 1.2 mM in glucose-fed exponentially growing *E. coli* (141). Therefore, although significant perturbations to cotranscriptional cleavage rates are only observed at high concentrations, there are specific effects that negate a general sensitivity of the assay to the presence of the tested small molecules. These data suggest that self-cleavage by HDV-like ribozymes is influenced by some feature of the tested sugars. Interestingly, there are some sequence specific effects: whereas aHDV and drz-Spur-3 are not affected by glcN6P, drz-Mmus experienced a decrease in self-cleavage with k_{rel} ~ 0.5 (Fig. 3-4). Only drz-Fpra-1 and drz-Fpra-2 display an enhancement of self-cleavage in the presence of glcN6P. In contrast, high concentrations of glc6P attenuate activity of nearly every ribozyme tested with the exception of drz-Spur-3 (Fig. 3-4). While the sensitivity of the ribozymes tested to glc6P suggests some general characteristic for HDV-like ribozymes, more work is needed to establish how the small molecules are being detected. The difference between the

structure of glcN6P and glc6P is an amine group, and the data suggest that it is important for promoting enhanced self-cleavage. The amine functional group may be necessary in either directly promoting the transesterification reaction leading to scission (as in the *glmS* ribozyme) or in stabilizing the active structure of the ribozyme. The glucose phosphate portion of the metabolite is important to the activity. Titrations of glcN6P and glc6P reveal that they have no impact on k_{rel} for drz-Fpra-1 at concentrations below ~2 mM (Fig. 3-5), roughly the physiological concentrations of either small molecule in *E. coli*. We expect that the phosphate group is not interfering with the cotranscriptional assay by chelating Mg^{2+} because different phosphate sugars should have had similar impacts on catalysis. Further, if the phosphate group was chelating Mg^{2+}, lower concentrations of metabolite would have a more pronounced effect on catalysis. Our data suggest that large fluctuations of cellular concentrations will lead to differential activity of the ribozyme, which in turn influences expression of the downstream ORF *in vivo*.

In order to further probe the structural requirements of a metabolite recognized by the *F. prausnitzii* ribozymes, we measured the responsiveness of drz-Fpra-1 and drz-Fpra-2 against a panel of metabolites. These data suggest that each ribozyme is recognizing multiple metabolites, presumably in the same binding pocket.

The results demonstrate that the activities of drz-Fpra-1 and -2 are influenced by most of the metabolites tested (Fig. 3-6). The catalysis by drz-Fpra-1 is enhanced by the amine containing sugars glcN6P and glcN1P. However, catalysis is attenuated in the presence of glc6P and glc1,6bp and glcN has almost no impact on the k_{obs}. The influence of fruc6P and UDP-N-AcglcN were not tested with drz-Fpra-1. The catalysis

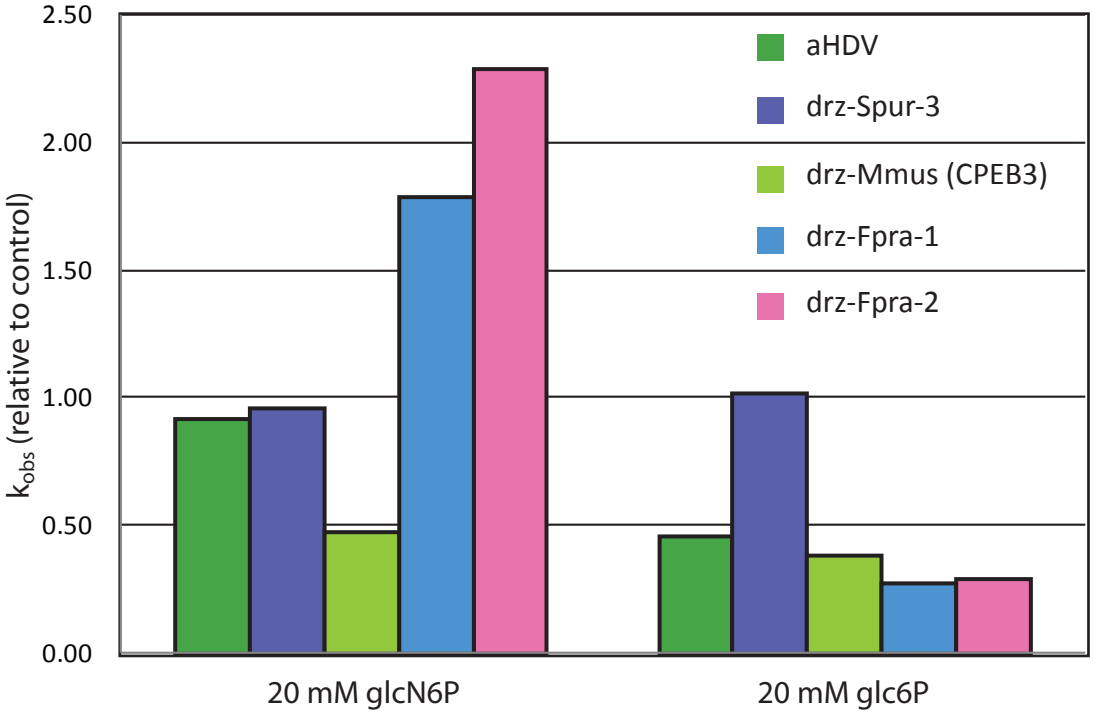


Figure 3-4. The activities of HDV-like ribozymes in the presence of sugars. *In vitro* self-cleavage assays were used to determine the k_{obs} for the ribozymes given. Only the *Faecalibacterium prausnitzii* (Fpra) ribozymes have been implicated in sensing metabolites from the hexosamine biosynthetic pathway. The activities were monitored in the presence of 20 mM glucosamine 6-phosphate (glcN6P) or 20 mM glucose 6-phosphate (glc6P); the sugars differing only in the amine group. The k_{obs} were normalized to control reactions where no metabolites were present.

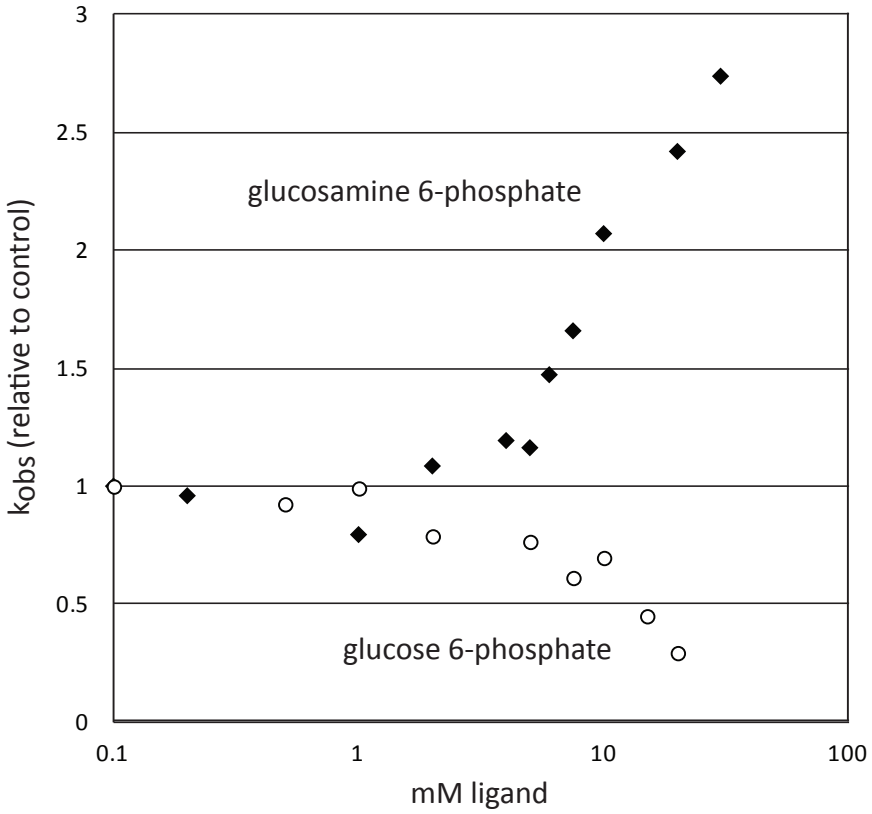


Figure 3-5. The sensitivity of drz-Fpra-1 to glucosamine 6-phosphate and glucose 6-phosphate. *In vitro* self-cleavage assays were used to determine the k_{obs} for drz-Fpra-1 activity in the presence of sugars. The resulting values were normalized to control reactions in which no sugars were present. Increasing concentrations of glucose 6-phosphate result in a decrease in the observed rates of self-cleavage, whereas increasing the concentration of glucosamine 6-phosphate enhances the observed rates of self-cleavage.

by drz-Fpra-2 is enhanced by the amine containing sugars glcN6P and glcN. Catalysis by drz-Fpra-2 is attenuated in the presence of all other sugars except UDP-N-AcGlcN which has no impact on the k_{obs}. Considering the remarkable conservation of the intricate nested double-pseudoknot structure of HDV-like ribozymes, it is curious that select sequences would have such a specific response to metabolites. This is an intriguing finding that suggests that despite conserved secondary and tertiary structure, the folding and catalysis by individual sequences may be differentially influenced by the environment. This finding is significant in providing a possible mechanism for the environmental sensitivity observed for HDV-like ribozymes *in vivo* (28,29).

The Fpra-like ribozymes, displaying up to 80 % sequence identity to core regions in drz-Fpra-1, have not yet been tested for sensitivity to the small molecules during cotranscriptional self-cleavage experiments. The conservation of the bacterial HDV-like ribozymes lies in the P3, L3, P1.1, J1.1/4, and J4/2 regions which are important in formation of the active site (Fig. 1-2). In fact, there are only three nucleotides that differ between drz-Fpra-1 and drz-Fpra-2 in the above described regions (boxed in Fig. 3-2). Future work should investigate how these three positions may contribute to the differential sensitivities to the tested metabolites. Further, cotranscriptional self-cleavage analysis of the bacterial HDV-like ribozymes will inform us as to the importance of this conservation and if it has a role in sensitivity to the small molecules glcN6P or glcN1P.

Cotranscriptional self-scission of drz-Fpra-1 under consistent ionic strength conditions

During the course of carrying out *in vitro* cotranscriptional assays measuring the k_{obs} for HDV-like ribozymes in the presence of different metabolites, various inconsistencies

begin to arise. The most worrisome of which was the drastic change in k_{obs} for drz-Fpra-1 in different buffer systems and in the presence of various monovalent ions, in particular high concentrations of K^+ attenuate the rate self-cleavage. In addition, high concentrations of the different metabolites introduced varying amounts of monovalents from the commercial products, and also overwhelmed the buffering capacity of the Tris buffer originally used for cotranscriptional assays. While the data obtained so far was subject to any combination of these influences, the sensitivity of only the *F. prausnitzii* ribozymes to the metabolites tested suggests the effect is specific. In order to more accurately measure the enhancement of self-cleavage by the metabolites, we designed a new protocol in which pH and ionic strength are tightly controlled. In addition, previous work characterizing the specificity of the glmS ribozyme-glcN6P cofactor interactions revealed that Tris (2-amino-2-hydroxy-methyl-propane-1,3-diol) is able to promote catalysis, albeit with a k_{obs} four orders of magnitude below that of glcN6P (81). We therefore eliminated Tris from both the transcription and reaction buffers and switched to HEPES (2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid)-based buffers. The k_{obs} for the drz-Fpra and drz-Fpra-like ribozymes are given in Table 3-1.

The reaction rate profile for drz-Fpra-1 was determined under the new controlled ionic strength. The data obtained are given in Fig. 3-7 and show the Mg^{2+} response curve for drz-Fpra-1. The trend follows that of a typical HDV-like ribozyme, where high concentrations of Mg^{2+} result in attenuation in ribozyme activity, presumably by causing misfolding (137). However, in the presence of 20 mM glcN6P, this attenuation is overcome and the reaction rates continue to increase with increasing Mg^{2+}. The enhancement in catalysis reaches up to 2-fold at high concentrations of both ligand and

Mg^{2+}. In addition, the presence of glcN1P has no impact on ribozyme catalysis suggesting that this molecule does not bind the ribozyme. Taking previous experiments into consideration, it seems that the amine group of glcN6P is necessary in enhancing catalysis. It is possible that the sterics between the amine and phosphate groups in glcN1P do not allow the molecule to fit into a binding site. On the other hand, the molecule may be bound but the sterics prohibit the amine group from participating in the chemistry necessary to enhance ribozyme activity. Previous data suggested that glc6P binds the ribozyme but the absence of the amine functionality had a negative impact on catalysis. If the influence of glc6P on ribozyme catalysis is reproducible under the new experimental conditions, then it is possible that the amine group is participating in electrostatic interactions facilitating folding of the ribozyme core. This can be determined by *in vitro* cotranscriptional experiments: 1) testing different structural analogs of glcN6P under the new conditions, 2) making point mutations in the ribozyme core, 3) a pH profile, or 4) designing a *trans*-ribozyme and measuring the rates of folding and catalysis.

A number of studies have been published demonstrating that HDV ribozyme activity can be inhibited by antibiotics (142-144), but the *F. prausnitzii* ribozymes are the first to show an enhancement of catalysis in the presence of a ligand. While the effect we observe is specific to the Fpra-1 and -2 ribozymes with amine containing sugars, the possibility remains that other HDV-like ribozymes are influenced by other small molecules not tested in this study. Small molecules affecting ribozyme catalysis can be acting in two ways: 1) they may be participating selecting the folding pathway for the *F. prausnitzii* ribozymes or 2) they are assisting nucleobases in fulfilling catalysis by

facilitating charge neutralization. The mechanism of catalysis by HDV-like ribozymes is well characterized but this study sheds light on an interesting facet of ribozyme research. Discovering new ribozymes involves verifying self-cleavage of the transcript of interest under physiological-like conditions with only divalent cations to support RNA folding. What if some natural ribozymes have evolved to rely on cellular factors to facilitate folding or catalysis? These transcripts would assume a consensus secondary/tertiary fold that would not be “autocatalytic”. In this case *in vitro* transcription and self-cleavage assays would be unable to verify new ribozymes. Regardless of catalytic competency, structural or chemical probing techniques can identify structural rearrangements or ligand binding.

Additional constructs of drz-Fpra-1 were designed in order to examine the influence on the surrounding transcript on catalytic rates. These were designed by changing the length of leader and tail regions surrounding the ribozyme structure. The following give the size of leader and tail sequences for the extended drz-Fpra-1 constructs: Original length construct: -43/6; SL: -43/6; LS: -105/6; LM: -105/61; and LL: -105/101.

The presence of 20 mM glcN6P had variable impact on the observed rates of self-cleavage for these four contructs. The k_{rel} values demonstrate that changing the flanking transcript impacts the sensitivity to glcN6P from the control construct. These experiments were performed under conditions of consistent ionic strength with the addition of 5 mM Mg^{2+}. The k_{obs} calculated for all reactions in the presence of 20 mM glcN6P as normalized to control. In the case of LS and LL constructs, k_{rel} values dropped to 0.70 and 0.76, respectively, whereas SL and LM demonstrate k_{rel} values of 1.27 and 1.00, respectively. The original construct demonstrates a k_{rel} of 1.38.

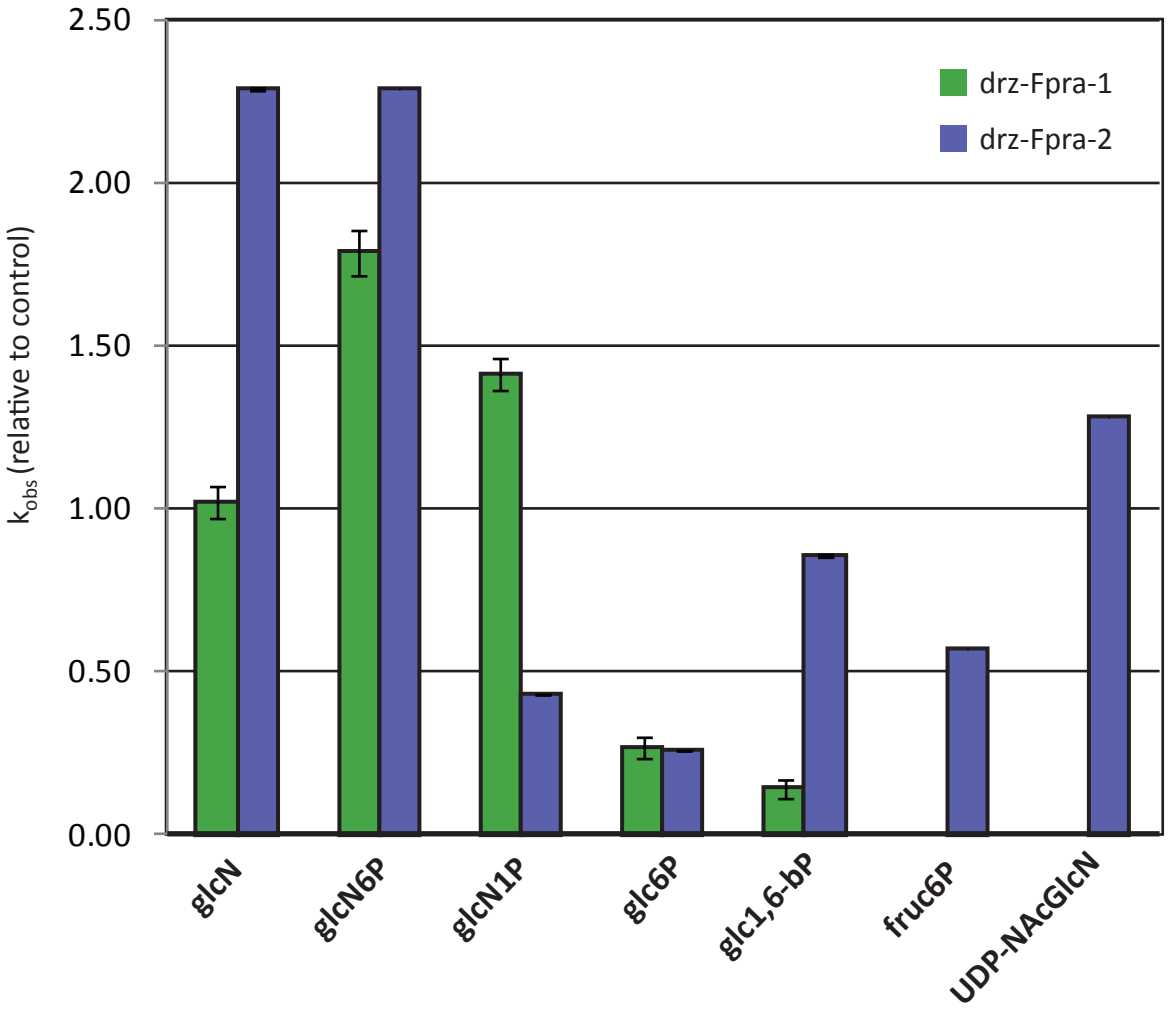


Figure 3-6. The sensitivity of drz-Fpra-1 and drz-Fpra-2 to a panel of metabolites. *In vitro* self-cleavage assays were used to monitor the change in k_{obs} for both ribozymes in the presence of various sugars related to flux of the hexosamine biosynthetic pathway (glucosamine, glucosamine 6-phosphate, glucosamine 1-phosphate, glucose 6-phosphate, glucose 1,6-bisphosphate, fructose 6-phosphate, and uridine diphosphate N-acetylglucosamine). The calculated k_{obs} values were normalized to control reactions which did not contain any sugars. Values for k_{rel} greater than one indicate the enhancement of self-cleavage in the presence of the indicated sugars and values less than one indicate an attenuation of self-cleavage. Data represents the k_{rel} ± the standard error of at least duplicate experiments. The k_{obs} for drz-Fpra-1 was not measured for self-cleavage in the presence of fructose 6-phosphate or UDP-N-acetylglucosamine.

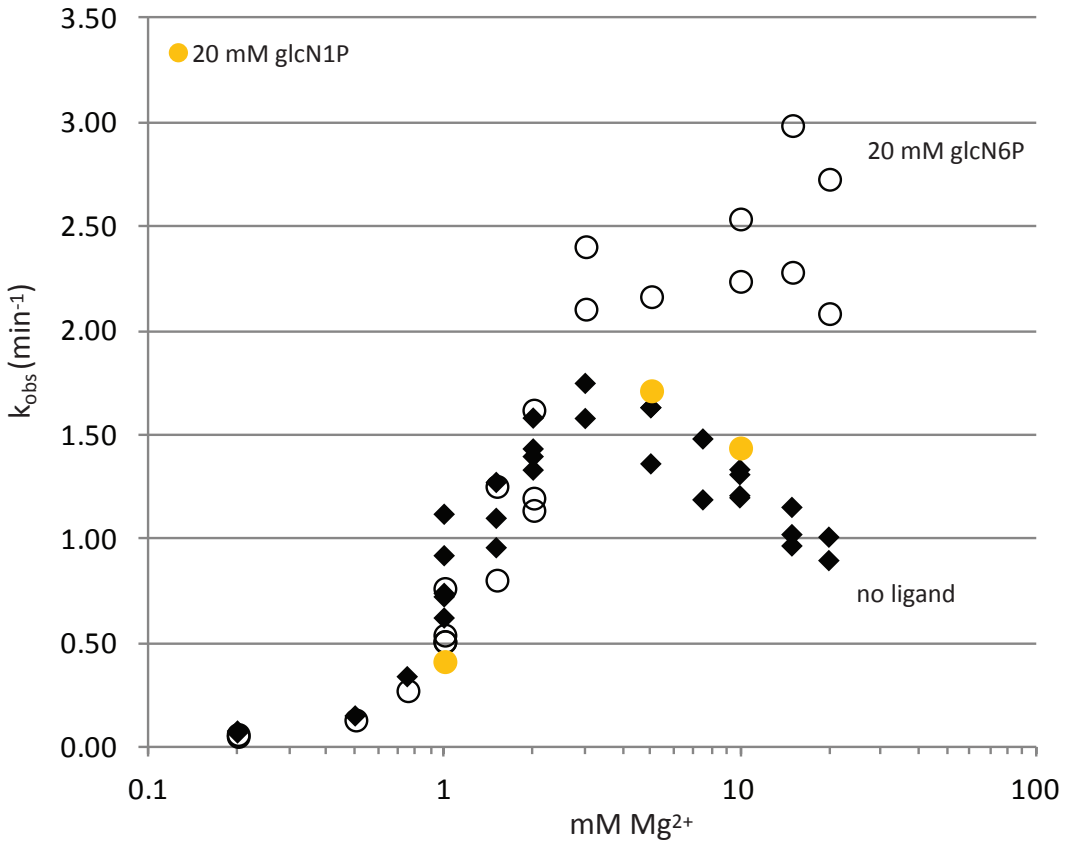


Figure 3-7. A Mg^{2+} response curve for drz-Fpra-1 and the impact of amine sugars. The k_{obs} for cotranscriptional *in vitro* self-cleavage under controlled ionic strength of drz-Fpra-1 was determined across a range of Mg^{2+} concentrations in the absence (black diamonds) and presence of either 20 mM glucosamine 6-phosphate (open circles) or 20 mM glucosamine 1-phosphate (orange circles). The data demonstrate that at high concentrations of Mg^{2+}, 20 mM glucosamine 6-phosphate enhances the observed rate of cleavage by up to approximately 2-fold.

The data suggest that the metabolite is influencing the rate of catalysis by interfering with or promoting a specific folding pathway. Lengthening the transcript containing the active ribozyme appears to be detrimental to glcN6P sensitivity implying that the ribozyme core is inherently sensitive to the metabolite. This supports the hypothesis that glcN6P is enhancing catalysis by drz-Fpra-1 via an allosteric mechanism involving the ribozyme core.

Structure probing and metabolite binding by drz-Fpra-1

To investigate metabolite binding by HDV-like ribozymes, we employed structure probing techniques to unmask the metabolite binding site and characterize the affinity of the interaction. In-line probing is a technique that takes advantage of the implicit tendency of RNA to non-enzymatically degrade over time based on its structure. Specifically, the correct “in-line” orientation of the 2′ oxygen, phosphorus center, and 5′ oxygen, allows the 2′ oxygen to act as a nucleophile and displace the 5′ oxygen, cleaving the RNA backbone (145). Experiments monitor the pattern of RNA degradation in the presence of increasing concentrations of ligand, which is resolved by high resolution PAGE. This technique can be used to verify some structural features of the ribozyme-ligand complex, ascertain the ligand binding site, and reveal the K_D of the interaction.

An in-line degradation experiment of drz-Fpra-1 (Fig. 3-8) shows how certain features in the ribozyme core are progressively protected with increasing concentrations of glcN6P. Notably, regions L3 and J1.1/4 are stabilized in the presence of the metabolite thus resulting in a reduction in scission of the phosphodiester backbone at those positions (boxed). This protection pattern may result from direct interaction with

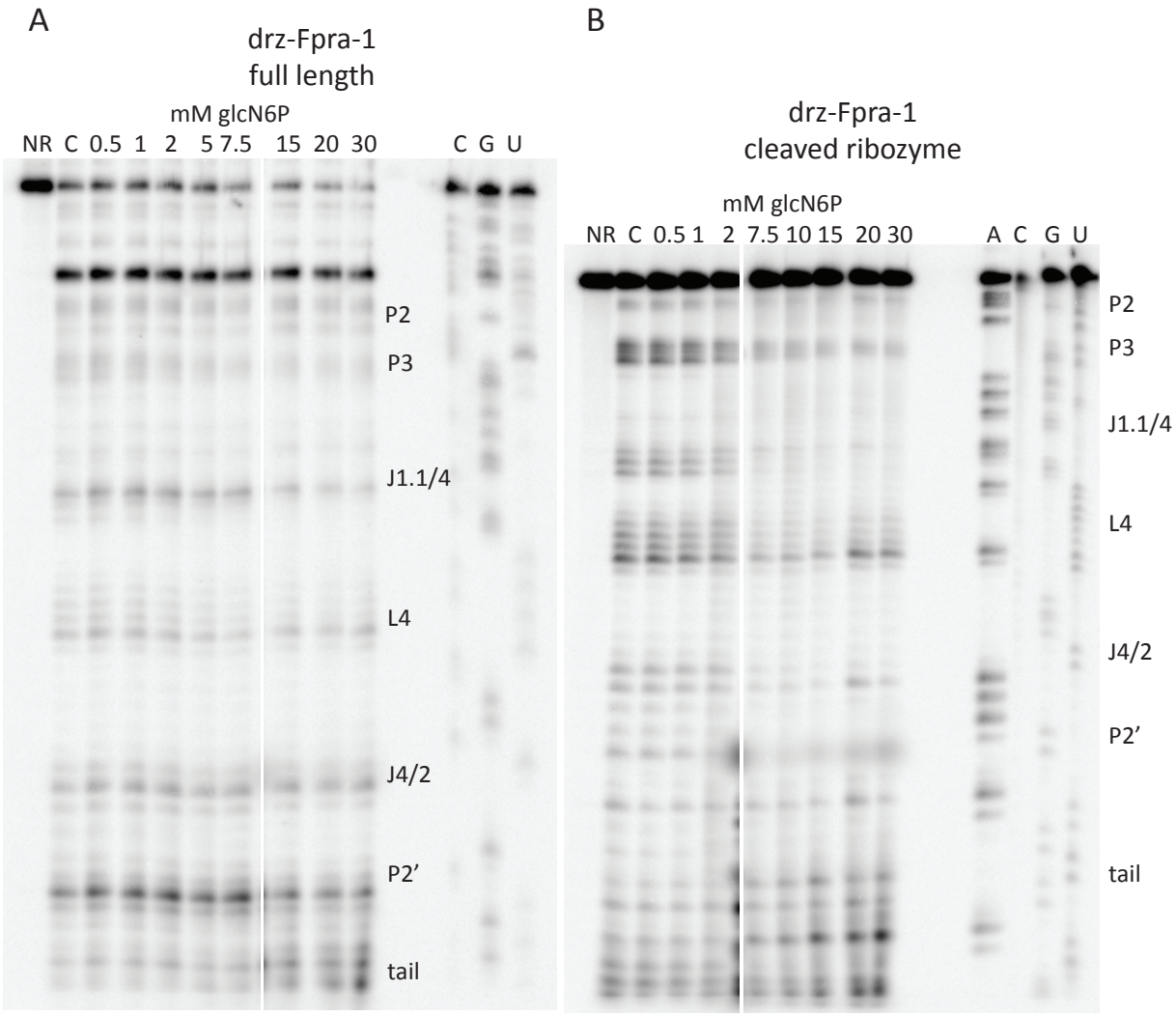


Figure 3-8. In-line degradation experiments of Fpra-1 in the presence of glucosamine 6-phosphate. The in-line probing experiments were resolved to single nucleotide resolution by denaturing PAGE. The lanes denoted "NR" are "no-reaction" lanes to serve a size control for full-length, purified, 3' end-labeled RNAs. The lanes denoted "C" are "control" lanes to provide a base line degradation pattern of the RNA in the absence of glucosamine 6-phosphate. Both the full-length transcript containing drz-Fpra-1 (A) and the product of self-cleavage containing the drz-Fpra-1 ribozyme structure (B) were subjected to in-line probing. There are various sites of protection, observed as diminishing intensity bands, with increasing concentrations of glucosamine 6-phosphate. There are almost no sites that experience an increase in degradation, observed as an increase in intensity, with increasing concentrations of glucosamine 6-phosphate. The sensitivity to in-line degradation in the presence of glucosamine 6-phosphate is more intense for the self-cleaved product than the full-length transcript which may indicate a tighter association of the sugar to the catalytic form of the ribozyme.

the ligand; however it might also correspond to structural rearrangements upon ligand binding at some distal position. Considering that the pattern of protection in these regions is more drastic in an in-line experiment for the cleaved ribozyme (Fig. 3-8B) versus the full-length transcript (Fig. 3-8A), we can say that the presence of glcN6P is leading to stabilization of the correctly folded ribozyme. Together with cotranscriptional self-cleavage assays, we hypothesize that the metabolite is stabilizing the catalytically competent form of the ribozyme. However, structure probing alone cannot determine how it is accomplishing this in the local electrostatic environment.

Future work will further characterize the ribozyme-ligand complex. The above described in-line probing experiments were performed under conditions where increasing concentration of glcN6P was resulting in a change of both ionic strength and pH. Therefore, we cannot rule out the possibility that the sites of protection were a result of local electrostatic perturbations resulting from those *in vitro* conditions. Therefore, further structural characterization of the drz-Fpra-1 and glcN6P complex will utilize selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) analysis under *in vitro* conditions where the only variable between reactions is the concentration of ligand.

Future in vitro cotranscriptional characterization of drz-Fpra-1 biochemistry

In addition, a number of drz-Fpra-1 mutants have been generated in order to decipher the influence of the metabolites on ribozyme folding. Based on HDV ribozyme crystal structures, making a single mutation at the catalytic cytosine residue abolishes all catalytic activity but maintains the correct secondary and tertiary structure. If these metabolites are assisting in the catalytic mechanism through electrostatic interactions, the presence of the correct cofactor might rescue activity of an inactive mutant. This

scenario can also be tested through point mutations to other important functional groups in the active site including: scissile phosphate or any of the nucleotides in L3. As mentioned previously, probing some of the nucleotide positions that vary between drz-Fpra-1 and drz-Fpra-2 might unlock the residues responsible for metabolite interactions which can account for the different sensitivities observed. Considering that drz-Csac-1 (Fig. 3-2) has such high sequence similarity to drz-Fpra-1 yet does not demonstrate an enhanced rate of self-cleavage in the presence of glcN6P, various chimeric constructs were designed by swapping structured elements between the two ribozymes. Specifically, six constructs were designed, three for each of the two ribozymes (Appendix A). It will be interesting to see which, if any changes result in drz-Csac-1 gaining sensitivity to glcN6P and conversely, which changes result in drz-Fpra-1 losing sensitivity. Together with SHAPE analysis, these experiments will reveal the binding pocket for the ligand.

In summary, data collected from *in vitro* cotranscriptional assays demonstrate the HDV-like ribozymes drz-Fpra-1 and drz-Fpra-2 are sensitive to glucosamine phosphate sugars. This sensitivity is modulated by the functional groups on the sugar. Under conditions where ionic strength is controlled, 20 mM of glcN6P leads to two-fold enhancements in the rate of self-cleavage for drz-Fpra-1 when the concentration of Mg^{2+} is above 5 mM. These data suggest that the ligand is stabilizing the catalytically competent form of the ribozyme. If the ligand was participating in the catalytic mechanism, an enhancement over control would be observed for all concentrations of Mg^{2+}. Though the effects on the observed rates of catalysis are modest, they are specific for this ribozyme, and we anticipate that drz-Fpra-2 will demonstrate similar

reactivity under the ionic strength controlled conditions. For drz-Fpra-1 we observe that lower concentrations of glcN6P at various Mg^{2+} concentrations lead to variable effects on the rates of self-cleavage; it appears that glcN6P enhances the rate of self-cleavage only when it is in excess to the concentration of Mg^{2+} (data not shown). When the concentration of Mg^{2+} is in excess to the concentration of glcN6P, the presence of the ligand attenuates the observed rate of self-cleavage, suggesting that the glcN6P is interacting with or displacing Mg^{2+}. Considering that HDV-like ribozymes are "hybrid engines" (42), utilizing both a divalent metal ion and nucleobases for the catalysis, the complexity in the *in vitro* cotranscriptional experiments is difficult to isolate and analyze in parts. Future work should be in designing a trans-acting version of Fpra-1 to separate the kinetics of folding from that of catalysis. These experiments in parallel with isothermal calorimetry measurements should inform us if glcN6P promotes folding of the ribozyme core and how glcN6P-Mg^{2+} interactions can account for the *in vitro* self-cleavage observations. The apparent sensitivity of drz-Fpra-1 to different metabolites motivated the design of an *in vivo* assay to investigate the impact of drz-Fpra-1 on expression of the downstream ORF.

Designing a dual reporter construct for in vivo activity assays

The approach to investigating the activity of drz-Fpra-1 *in vivo* monitored the influence of ribozyme activity on a dual reporter system. The construct to be expressed in *E. coli* is bicistronic in which the ribozyme and surrounding genomic sequence are inserted between the *Renilla luciferase* (Rluc) and *Firefly luciferase* (Fluc) genes. In this format, the upstream Rluc acts as a control and the downstream Fluc reports any effect the ribozyme causes to expression of the ORF. The expression of Rluc is controlled by

a lac promoter and the transcript includes a Shine Dalgarno sequence that begins seven nucleotides upstream of the AUG start codon. For the drz-Fpra-1 construct, the intergenic region between *Fluc* and *Rluc* is designed to be nearly identical to the intergenic region containing drz-Fpra-1 from the *F. prausnitzii* SL3/3 genome. This intergenic region contains a Shine-Dalgarno sequence that begins 14 nucleotides upstream of the AUG start codon for the *glmM* ORF. In addition, the length of the intergenic region in the *F. prausnitzii* genome (142 nts) is identical to the length in the bicistronic construct, corresponding to the two reporter ORFs being out of frame with each other. Besides a short hairpin structure immediately 5′ to the ribozyme, there are no predicted secondary structures formed by this intergenic region. A full list of constructs used for these studies can be found in Appendix A.

In order to test the responsiveness of the Fpra1 constructs to fluctuations in metabolism, we designed four different approaches to measure *in vivo* expression from the dual reporter system. The first approach was to measure steady-state levels during exponential growth of the *E. coli* cell culture, and the growth conditions included a minimal media utilizing different carbon sources. The second approach was to measure the response to stress, during which time the metabolism of the cells in culture is undergoing drastic changes. The stress conditions are caused by an abrupt change in media from one very rich in nutrients (LB) into the minimal media. We expanded upon this “pulse” of starvation and followed up with a “chase” back into rich media. The third approach, pulse-chase, was intended to give us a metabolic timeline for the responsiveness of ribozyme driven expression. The results of these experiments would provide evidence for the environmental sensitivity of ribozyme driven expression

thereby demonstrating a new role for HDV-like ribozymes as genetic controls in bacteria. Finally, the *E. coli* cell line was changed so that it was possible to induce expression from the lac promoter.

E. coli steady-state expression of a reporter gene containing drz-Fpra-1 in the 5' UTR

To test the hypothesis that the *F. prausnitzii* HDV-like ribozymes are involved in regulating the gene they surround, we cloned the above listed constructs downstream of the Renilla luciferase gene and upstream of the Firefly luciferase gene (Fig. 3-9A). For all constructs tested, the reporter activities are given as a ratio of Firefly/Renilla measured levels in random fluorescent units (RFU). *E. coli* cultures are grown overnight in LB media. The next day cells are pelleted, washed and resuspended in minimal media and incubated overnight at 37˚C with agitation. The following day the OD_{600} measurements are taken. Cultures are normalized to the same OD_{600} of 0.4 into minimal media and incubated for 1hr at 37˚C with agitation. Cultures are normalized to the same OD_{600} of 0.2 into minimal media and incubated at 37˚C with agitation. Timepoints denote the length of time the new cultures are incubated at 37˚C following normalization to OD_{600} of 0.2. At the indicated timepoints, 1 mL aliquots of culture are pelleted, lysed using chicken egg white lysozyme (Sigma), and immediately stored at -80˚C. To measure the levels of Fluc and Rluc, two 10 µL aliquots of the resulting cell lysates are plated in a black bottom 96-well plate. Each aliquot receives 50 µL of either Luciferase Activity Reagent (Promega) containing luciferin to measure Fluc levels, or 55 µM coelenterazine (Promega) to measure Rluc levels.

Our data shows the Fluc/Rluc ratio is greater than one for both the Fpra1-WT and Fpra1-C_{58}U mutant. This ratio is greater when a self-cleaving ribozyme is located upstream of Fluc and increases by the five hour timepoint (Fig. 3-9B). When an active ribozyme is present, the two ORFs are separated and become independent messages. When the inactive mutant is present, the two ORFs remain together as a bicistronic transcript. Our data suggests that Fluc is sensitive to the cleaved/uncleaved state of the ribozyme, determined by experiments with an inactive variant Fpra1-C_{58}U. These data suggest that a HDV-like ribozyme on the 5′ end of the transcript can either serve to protect the RNA from degradation or to promote translation of the downstream ORF.

We also tested the expression from our bicistronic construct under different growth conditions. Minimal media was supplemented with glycerol (control), glucose, or glucosamine as carbon sources. Given previous *in vitro* data suggesting that the *F. prausnitzii* ribozymes are sensitive to various amine sugars, we sought to change the flux of the hexosamine pathway by providing carbon sources that would drive production by varying degrees (Fig. 3-3). Glycerol does not get metabolized though any hexose sugar intermediates and so should have no influence on this pathway. Glucose would get converted to fructose 6-phosphate which can be fed directly into the hexosamine biosynthetic pathway or to glycolysis, thereby increasing cellular levels of glcN6P. Glucosamine will get phosphorylated, thereby feeding directly into this pathway by increasing cellular levels of glucosamine 6- or 1-phosphate. We find that Fluc/Rluc ratios from Fpra1-WT and Fpra1-C_{58}U remain the same, regardless of carbon source (Fig. 3-10). We observe that with any of the carbon sources tested, the Fluc/Rluc ratio increases by the five hour timepoint in the case of Fpra1-WT.

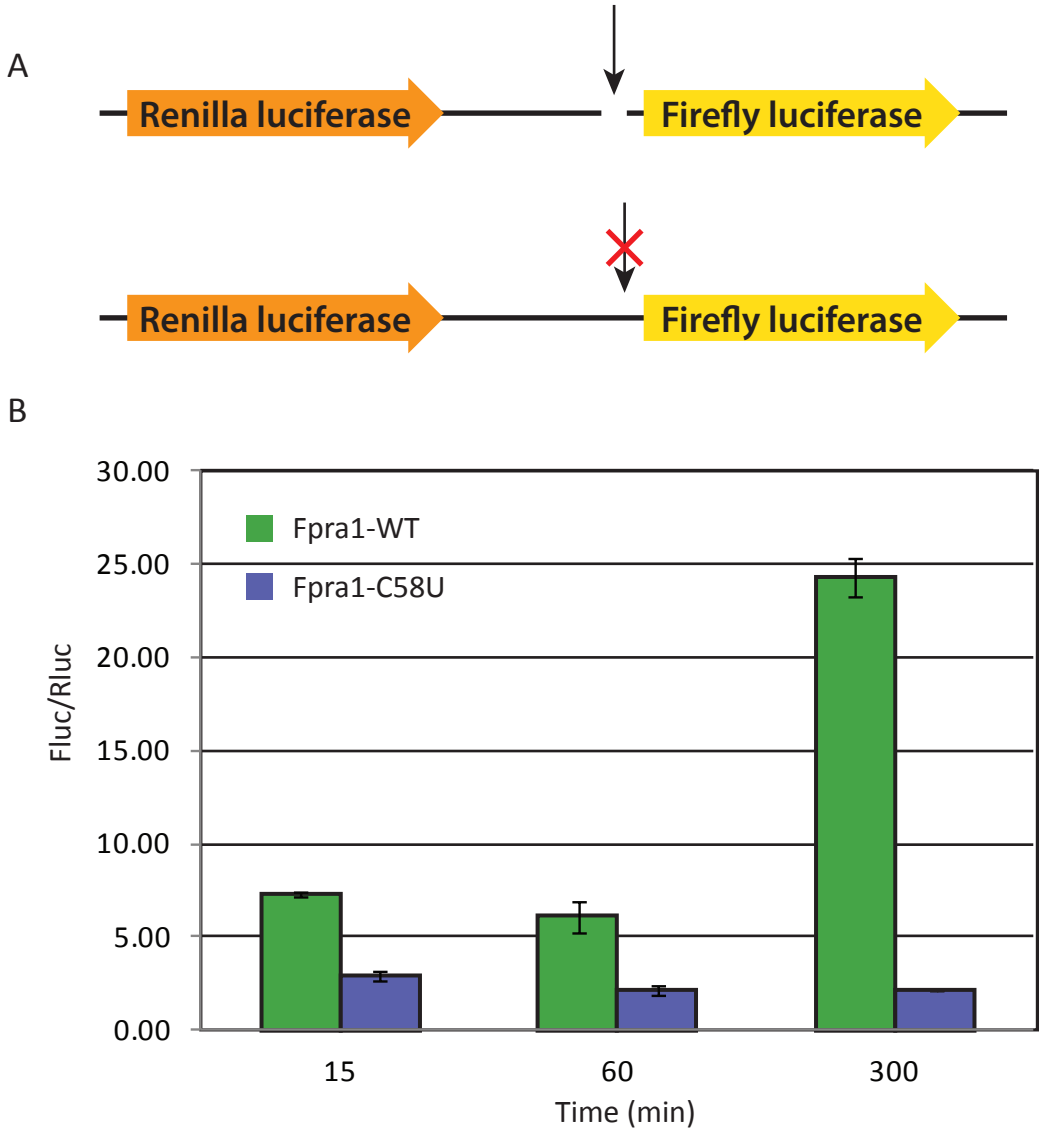


Figure 3-9. The influence of drz-Fpra-1 on translation in *E.coli* during steady-state. (A) A schematic of the bicistronic reporter constructs used during *in vivo* assay. *Renillla luciferase* is encoded upstream of a *Firefly luciferase* gene. The intergenic region was used as a cloning site for all constructs. The presence of Fpra1-WT, the active version of drz-Fpra-1, would result in the separation of the two ORFs upon self-cleavage by the ribozyme. The presence of Fpra1-C_{58}U, the inactive version of drz-Fpra-1, would maintain a bicistronic message. (B) The Firefly (Fluc) and Renilla (Rluc) activity ratios for WT and C_{58}U constructs are plotted. The activities were monitored at the indicated times beginning when cultures were normalized to an OD_{600} of ~0.4 into minimal media containing glycerol as a carbon source. Data represents the average values ± the standard error of three biological replicates.

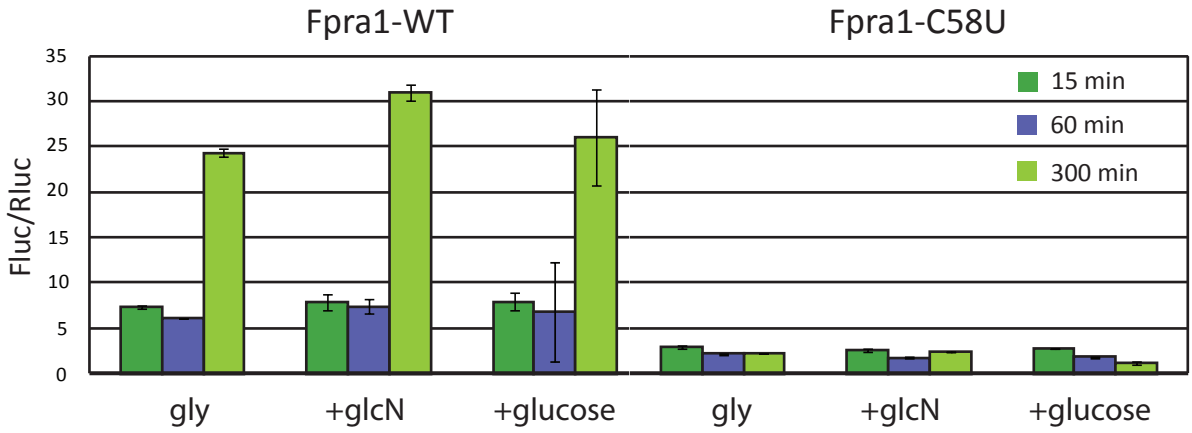


Figure 3-10. The influence of drz-Fpra-1 on translation in *E.coli* grown on different carbon sources. The Fluc/Rluc ratios are plotted for WT and $C_{58}U$ constructs and do not change with different carbon sources (gly=glycerol, glcN=glucosamine, and glucose=glucose) . The WT construct experiences an enhancement in Fluc/Rluc ratios by the 300 min timepoint. Data represents the average values ± the standard error of three biological replicates.

This *in vivo* data suggests that the presence of a self-cleaving ribozyme influences expression of the downstream ORF. It is not necessarily the secondary structure but the bisection of the bicistronic transcript that promotes expression of the downstream ORF. Our data suggests that the HDV-like ribozyme is not behaving as a ribosomal binding site due to the differential expression pattern between Fpra1-WT and Fpra1-C_{58}U. We further investigated whether the enhancement in Fluc levels is due to the stability of the message capped by the ribozyme secondary structure or the ability to promote translation of the ORF.

The impact of stress-induction on reporter gene levels in E. coli

In order to characterize the mechanism of reporter expression from a transcript capped by a HDV-like ribozyme, we created construct Fpra1-noSD. This construct does not contain the Shine Dalgarno (SD) ribosomal binding site between the ribozyme and the start codon of the downstream ORF. This construct can be used to decipher whether the structure of the cleaved ribozyme is promoting translation. Both Fpra1-WT and Fpra1-noSD contain the active form of the ribozyme which means that in both cases, the bicistronic transcript gets separated into two independent messages. In one case, canonical translation occurs when the 30S ribosomal subunit binds the SD sequence and translation begins when the translation initiation complex encounters the AUG start codon. In the second case, the absence of the SD sequence should prevent expression of the Fluc protein, resulting in Fluc/Rluc ratios <1.

E. coli cultures are grown overnight in LB media. The next day cells are pelleted, washed and resuspended in minimal media and incubated for 6 hrs at 37˚C with agitation. The OD_{600} measurements are taken and cultures are normalized to the same

OD_{600} of 0.2 into fresh minimal media and incubated at 37˚C with agitation. Timepoints denote the length of time the new cultures are incubated at 37˚C following normalization. We consider this experiment to sample the expression of the dual reporter system under stressful conditions because cell cultures were not dilute into minimal media and in fact the OD_{600} reading is >1. Upon dilution into fresh minimal media during normalization, the starved cells quickly use up the nutrients in the media and become stressed during the time course of the experiment. The can be inferred from the raw data for Fluc and Rluc levels (Fig. 3-11B and 3-11C).

Contrary to the steady-state experiment where Rluc levels steadily decline during the time course of the experiment, Rluc levels remain relatively steady and then suddenly decrease during this experiment. While the Rluc levels are identical for both Fpra1-WT and Fpra1-noSD, the Fluc levels are strikingly different. In the case of Fpra1-WT, Fluc levels remain steady and then suddenly decrease, similar to the Rluc levels. However, the Fpra1-noSD Fluc levels steadily increase during the time course. The data shows that the Fluc/Rluc ratio (Fig. 3-11A) for Fpra1-WT drops nearly two-fold and then remains constant. The Fluc/Rluc ratio for Fpra1-noSD is steady then increases about two-fold. These data suggest that during stress, canonical translation is being halted because any ORFs relying on an SD sequence experience a decrease in expression. However, the presence of a HDV-like ribozyme on the 5′ end of the message allows the downstream ORF to continue getting expressed. These data suggest that the HDV-like ribozyme is promoting translation and not necessarily prolonging the lifetime of the transcript. RT-qPCR experiments designed to quantify the levels of RNA generated from our bicistronic construct are discussed later.

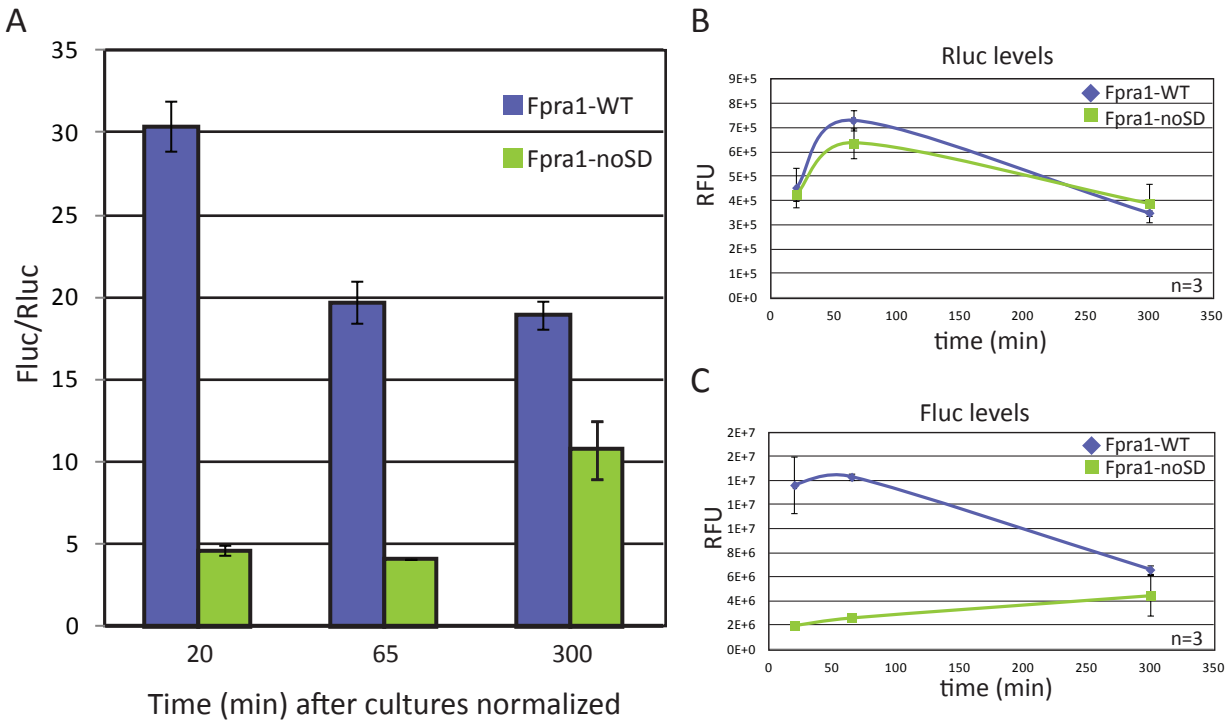


Figure 3-11. The impact of stress on the expression of a bicistronic message separated into independent ORFs by the self-cleavage of drz-Fpra-1. (A) The Firefly (Fluc) and Renilla (Rluc) activity ratios for WT and noSD (no Shine Dalgarno) are plotted. Activities were monitored at the indicated time points denoting the length of time cultures were incubated at 37˚C following normalization into minimal media after being grown in LB broth. The switch in media from a high complexity nutrient source to a minimal media results in stress that can be observed in (B) showing the activity of Rluc at each timepoint. Rluc activity serves as a control, informing the general levels of protein expression from a *lac* promoter based on the state of the culture. (C) Fluc activity at each timepoint demonstrate that there is a significant different in expression by the two constructs, where a SD site accounts for a majority of the Fluc activity measured from the WT construct. However, under stressful conditions the Fpra ribozyme can maintain Fluc activity in the absence of a SD site (noSD).

To further investigate the cellular conditions promoting ribozyme-driven translation we designed a pulse-chase experiment.

The results of pulse-chase assay on reporter gene levels in E. coli

In order to understand the background expression from a bicistronic message we determined the Fluc/Rluc ratio for the Fpra1-noRBZ construct during a pulse-chase experiment. The pulse refers the transfer of cell from an LB culture into minimal media. The chase refers to the transfer of cells back into LB culture. The results of this experiment, data not shown, reveal that the SD site located upstream of the Fluc ORF is contributing to most of the expression at that ORF. Under these experimental conditions, all four constructs tested appear to behave proportionately. The data suggests that upon transfer into different media, each construct experiences the same fold decrease or increase in Fluc/Rluc ratio. However, looking at the raw RFU values for Fluc and Rluc independently, we see that the Rluc reporter control behaves very differently for each of the four constructs during the course of the experiment (with up to a four-fold spread in RFU levels among constructs). This suggests 1) that the construction of the intergenic region between Rluc and Fluc is affecting the upstream ORF or 2) that translation of culture grown in rich LB broth is very sensitive to handling. In order to gain control over the reporter system we switched into an inducible system.

Inducing expression of bicistronic constructs

The incentive for switching to an inducible system is to control the timing of induction at the lac promoter for the bicistronic dual reporter system. We can then observe the production of both Rluc and Fluc during the early stages of expression. This eliminates the variability that may arise from a constitutive system where it is difficult to decipher

effects arising from the life-time of the transcript or the competency of the transcript for translation. The data presented here is preliminary because it has not been replicated. The Fluc and Rluc levels are determined for four constructs Fpra1-WT, Fpra1-$C_{58}U$, Fpra1-noSD, and Fpra1-noRBZ. For all constructs the Fluc and Rluc levels were determined for *E. coli* cultures in LB media with and without induction by 500 μM IPTG. Cultures were grown overnight at 37˚C with agitation. The next morning all cultures were normalized to an OD_{600} of 0.4 into fresh LB media and allowed to grow for 45 min at 37˚C with agitation. After that time, the appropriate cultures were induced by the addition of IPTG. The indicated times refer to the length of time cultures were allowed to grow following induction. Aliquots were lysed using lysozyme as previously described and immediately stored at -80˚C. The resulting cell lysates are plated in a 96-well black bottom plate and LAR and coelenterazine used to measure Fluc and Rluc levels, respectively.

The uninduced cultures are used to independently normalize the Fluc and Rluc levels measured for the induced culture. These normalized values are then used to calculate the Fluc/Rluc values for each construct, shown in Fig. 3-12. It should be noted that the reporter genes did not demonstrate induction of expression until the 35 min time point. The data suggests that the presence of a SD site (Fpra1-noRBZ) in an intergenic region of a bicistronic message causes the Fluc/Rluc ratio to decrease, demonstrating the preference for ribosome assembly on the 5′ end of the bicistronic transcript. The presence of a self-cleaving ribozyme in the absence of a SD site (Fpra1-noSD) results in a 3-fold enhancement in expression of the downstream ORF when the system is induced. This suggests that separation of the ORFs in the transcript is beneficial to

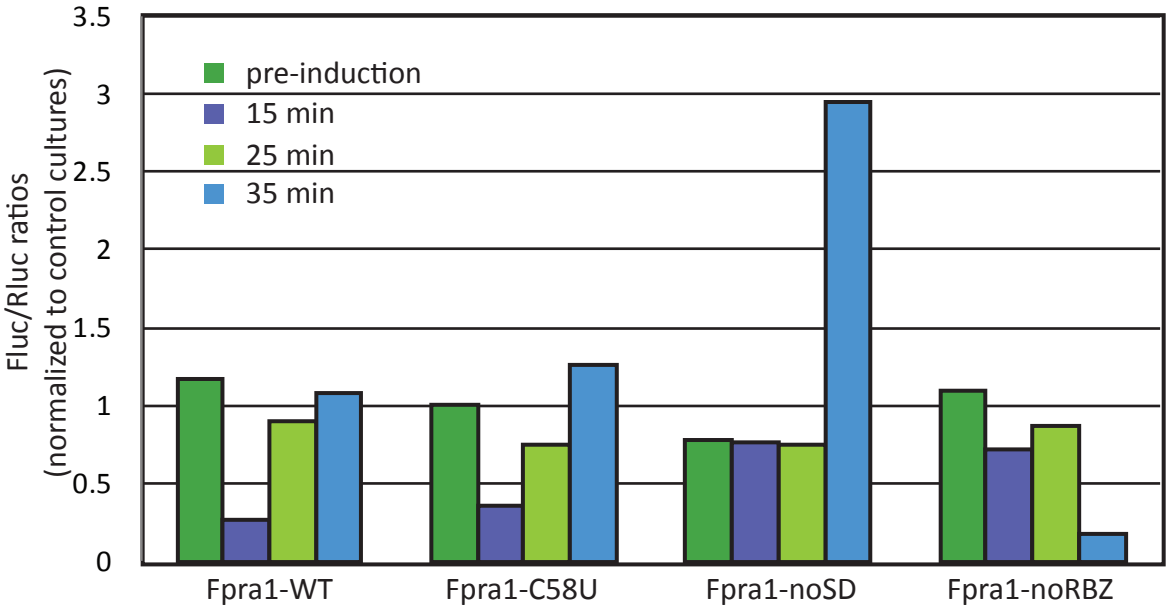


Figure 3-12. The influence of Fpra constructs on translation from an inducible system in *E.coli*. The Firefly (Fluc) and Renilla (Rluc) activity ratios for WT, $C_{58}U$, noSD, and noRBZ (lacking drz-Fpra1) are plotted. These ratios are normalized to the cultures not induced by IPTG. Induction was detected by 35 min after the addition of IPTG (increase in Rluc levels, data not shown). At this timepoint, we observe that the noSD construct leads to a higher Fluc/Rluc ratio than all other constructs which contain a SD site. This suggests that the mechanism of expression varies between messages with a SD site or HDV-like ribozyme in the 5' UTR.

ribosome assembly on the 5′ end of the Fluc ORF. The absence of any SD site suggests the ribozyme is playing a role in recruiting ribosomes. The analysis of data from these two above mentioned constructs helps to explain the observed results for Fpra1-WT and Fpra1-C_{58}U. In this experiment, Fpra1-WT and Fpra1-C_{58}U appear to behave similarly. The results for Fpra1-C_{58}U can be explained by the following: induction of the bicistronic message allows accumulation of ribosomes on the 5′ end leading to an increase in Rluc over Fluc levels, however the presence of the ribozyme upstream of *Fluc* leads to an increase in Fluc levels and therefore we do not observe a change upon induction of the *E. coli* culture. The observed results for Fpra1-WT can explained in a similar way, except that separation of the two ORFs does not alter ribosome accumulation under these experimental conditions. Comparing the results for Fpra1-WT and Fpra1-noSD is promising in that it rules out any non-specific ribosomal assembly on the 5′ end of transcripts. The observations can be explained by the confliction of ribosomal assembly to use the SD site or the alternative mode of recruitment provided by the ribozyme. While it appears we have isolated the influence of the HDV-like ribozyme alone on expression of the downstream ORF, an additional construct (Fpra1-C_{58}U-noSD) is necessary to confirm that the enhancement of expression is dependent on separation of the bicistronic message. Previous data collected from culture in minimal media during steady-state suggests that it is necessary for maintaining steady expression levels of the downstream ORF.

Discussion

The presence of a HDV-like ribozyme located in the 5′ UTR of a bacterial gene suggests a role for the ribozyme in gene expression regulation. The presence of

ribozymes located on either end of the gene, which is suggested to be expressed as part of a polycistronic transcript, implicates the self-cleaving ribozymes in liberating a single message out of the parent transcript. This role for HDV-like ribozymes has previously been shown for HDV-like ribozymes associated with retrotransposons in eukaryotes (32,33). While the liberation of the retroelement message is necessary for the continuation of the retrotransposition cycle, it is also necessary for the maturation of 28S rRNA. This gives a biological role to the act of self-scission and additionally the presence of the HDV-ribozyme structure on the end of the message has been shown to promote expression of the downstream ORF *in vitro* and *in vivo* (32,33). Though there is some evidence to suggest HDV-like ribozymes are sensitive to environmental differences during development or across different tissues, the bacterial *F. prausnitzii* ribozymes are the first to have a specific metabolite sensitivity characterized. During *in vivo* experiments, we observed different expression profiles for the active Fpra1-WT and inactive Fpra1-C_{58}U bicistronic constructs. Therefore, expression of the downstream ORF is influenced by the presence of a self-cleaving ribozyme, and further the enhancement of expression provided by the active ribozyme is affected by different bacteria culture conditions. However, we cannot conclude that this is a result of flux through the hexosamine biosynthetic pathway. Through this study, we have provided evidence for the environmental sensitivity of the *F. prausnitzii* ribozymes both *in vitro* and *in vivo*. Further, we observe an interesting interplay between the active ribozyme and a Shine Dalgarno site located just tens of nucleotides downstream from the end of the ribozyme secondary structure. Our data suggests that while canonical translation occurs from the SD site, under certain conditions there is a preference for expression

promoted from the ribozyme. This is interesting in that the ribozyme is behaving as a genetic control without occluding the SD site, the mechanism by which many riboswitches turn off translation. It can be speculated that the ribozyme is able to sustain expression of the downstream ORF under conditions where canonical translation is being halted. This is significant in that it may be the first incidence of an ON/OFF control in bacteria that does not involve, or is not predicted to involve, significant rearrangements in RNA structure. The mechanism for choosing ribozyme or SD site is the predominant focus of future directions.

Conclusions and future directions

Our data suggests that an active HDV-like ribozyme is able to promote expression of the downstream ORF under certain conditions. Considerable effort was put into examining the RNA levels and observing how they compare to the expression of the Fluc and Rluc reporters. The intention was to determine if changes in expression of Fluc were the result of scission and consequential degradation of either piece of the message. The case where the RNA and reporter levels do not correlate would suggest the mechanism of regulation is at the level of translation, whereas if they directly correlate then the mechanism is through alterations to the lifetime of the message. We were unable to perform reliable RT-qPCR experiments from RNA extracted from the MachT1 *E. coli* cell line. The possible reasons for this include: 1) the inability to safely extract RNA without a considerable amount of RNase degradation, or 2) the high-copy number plasmid used in designing constructs did not promote a high amount of transcript generated from the lac promoter. Based on the comparable levels of the Rluc reporter across different constructs tested *in vivo*, we predict that self-cleavage does not

lead to degradation of the Rluc ORF. The Rluc piece of the transcript would contain a 2′-3′ cyclic phosphate on the 3′ end. All *E. coli* ribonucleases digest in the 3′-5′ direction and most exoribonucleases are nonspecific for their substrates (reviewed in (146)). In *E. coli* RNase E has been implicated in processing RNA in the 5′-3′ direction, however RNase J1, first identified in *Bacillus subtilis,* is the first bacterial exoribonuclease to demonstrate robust 5′-3′ activity (147). In fact, RNase J is the ribonuclease recognizing and degrading the glmS message upon scission by the glmS ribozyme (131). In addition, the mechanism of action by RNase J involves recognition of 5′-OH at the end of any transcript, and this effect can be reconstituted in *E. coli* when RNase J is introduced; and it should be noted that *E. coli* do not express RNase J. There is a gene encoding RNase J in the genome of *F. prausnitzii*, and this reveals an innate flaw of the *in vivo* studies. Our *in vivo* expression data informed us of the impact on expression of a downstream ORF by a ribozyme in a bacterial system, but the transcript was never exposed to RNase J. Therefore, the results generated from a cell line expressing RNase J will characterize the true biological impact of the *F. prausnitzii* ribozymes in conditions similar to their native context. Given that RNase J is active in *E. coli*, a feasible future direction would be to coexpress the bicistronic dual reporter system with the gene encoding RNase J. If RNase J is required for destabilization of mRNA with a 5′-OH, should we expect then for the Fpra1 constructs to behave as “suicide” switches, such as the glmS riboswitch?

The answer is no. Unlike the glmS ribozyme, the HDV-like ribozymes are autocatalytic, requiring only divalent metal ions to facilitate catalysis. Our studies suggest that catalysis by HDV-like ribozymes can be accelerated by small molecules.

Therefore, the activity of the ribozyme is impacting expression of the downstream ORF under any cellular conditions. We assume that RNase J will act on the cleaved transcript containing a 5′-OH. However, the expression levels of RNase J itself must fluctuate under different cellular conditions. The fluctuation of RNase J levels may or may not correlate with the fluctuation in environmental conditions that enhance *F. prausnitzii* HDV-like ribozyme activity. Therefore, we may be able to elucidate the mechanism of the ON/OFF regulation of an ORF located downstream of a ribozyme that does not rely on a cofactor for activity. The gene will be ON when levels of RNase J are low. When RNase J levels are low and the ribozyme activity is high, this may lead to an even greater enhancement in expression, as has been suggested by our *in vivo* experiments in *E. coli*. The gene will be OFF when RNase J levels are high and the ribozyme activity is enhanced. Future work should focus on understanding the dynamic relationship between ribozyme activity and cellular levels of RNase J. The possibility remains that the exceptional stability of the nested double-pseudoknot of HDV-like ribozymes is resistant to degradation by RNase J.

The work presented here is significant in defining a new role for HDV-like ribozymes in bacteria. This work provides a new platform for genetic engineering in bacteria and may be able to guide rational design of antibiotic therapies. This work provides insight into the chemical capacity and biological relevance of a well-characterized structured RNA family.

Chapter 4

Elucidating the mechanism of HDV-like ribozyme-promoted translation of a downstream ORF

Introduction

The central dogma of molecular biology teaches us that DNA directs its transcription into RNA, which dictates translation into proteins. The unidirectional flow of information accounts for a great number of phenomena we observe in biological systems. Over time, it became increasingly clear that proteins are not the sole class of macromolecules dominating the hierarchy of essential catalytic cellular events. In fact, RNA fulfills informational and catalytic roles that are vital to biology. Our current understanding of RNA places it at the heart of such processes as mRNA processing (spliceosome), telomere maintenance (telomerase), X-chromosome inactivation (Xist lncRNA), and protein synthesis (ribosome), to name a few. The root of these dynamic roles lies in the ability to fold into unique and stable structures. While folding is predominantly governed by the thermostability of base-stacking in local environments, proteins and trans-encoded RNAs can alter or disrupt the native structures. Ribozymes are a catalytic class of RNAs that must achieve the correct global structure in order to perform chemistry. Large ribozymes are part of RNP complexes and the structural support provided by proteins facilitates the multi-turnover capacity of these machines. Self-cleaving ribozymes are comparatively smaller in size and global architecture and do not rely on proteins for catalysis. However, accumulating evidence suggests that these structured RNAs may have a life beyond self-scission. While the structure of these ribozymes is necessary for catalysis, it may also be important in molecular recognition.

In this case ribozymes are at the core of two different signaling cascades: one arising from scission of the phosphodiester backbone and the second from intermolecular interactions with the cleaved or uncleaved version of the ribozyme. This concept is perhaps why it has been difficult to ascertain a primary biological role for HDV-like ribozymes in eukaryotes.

The ribozymes first discovered in the human pathogen hepatitis delta virus (HDV) self-cleave to process concatamers to unit-length RNA genomes during rolling-circle viral replication. This class of self-cleaving ribozyme is widespread throughout nature (29) and HDV-like ribozymes map to diverse genomic loci, suggesting multiple biological functions. For example, the human cytoplasmic element binding protein 3 (*CPEB3*) HDV-like ribozyme is conserved in mammals. It is located in intron 2 and positioned 10-14 kbp upstream of the second coding exon in mammals (excluding marsupials) (28). Interestingly, the ribozyme has a single nucleotide polymorphism (SNP) in humans. The $U_{36}C$ SNP confers enhanced rates of self-cleavage for the ribozyme *in vitro* potentially resulting in a larger fraction of truncated (self-cleaved) transcripts *in vivo*. This observation is important when considering the timeline of transcription and of the splicing events that must occur before the ribozyme truncates the transcript. The time for the RNA polymerase to move from the ribozyme to the 5′ exon junction is estimated to be ~10 min (148). The slow k_{obs} 0.01 min^{-1} observed *in vitro* initially suggested that the ribozyme was evolutionarily selected to be slow in order to accommodate splicing (28). However, recent evidence demonstrates that the human *CPEB3* ribozyme is intrinsically fast-reacting when the flanking single-stranded sequences are changed, suggesting that activity may be regulated by the ability of the

surrounding transcript to interfere with ribozyme folding, which can in turn be influenced by RNA binding proteins (140). Given that ESTs mapping to the cleaved ribozyme have been detected and RT-qPCR and 5′ RACE data demonstrate that the ribozyme is active *in vivo* (28), it is possible that chaperones are assisting in folding of the *CPEB3* ribozyme. Most intriguing is the observation that the levels of cleaved ribozyme vary by tissue with up to 50 % of ribozyme containing transcripts being cleaved in mouse brain tissue (28). In the brain, the predicted role for CPEB proteins is in synaptic plasticity and memory. Behavioral tests showed that the presence of the $U_{36}C$ SNP impacts performance in an episodic memory task (35). The authors speculated that the decrease in performance may be attributed to decreased levels of CPEB3 protein resulting from the enhanced ribozyme cleavage rates that cause truncated mRNA. It is interesting that a number of hammerhead ribozymes in eukaryotes also map to the introns of a conserved set of genes (9). These ribozymes may similarly play roles in pre-mRNA processing and perhaps this will emerge as a primary role for self-cleavage in eukaryotic biology. However, not all HDV-like ribozymes map to introns and another well-studied example is the set of *Drosophila* R2 retrotransposon-associated ribozymes.

A large number of HDV-like ribozymes map to retrotransposons of various types (32,149). Of particular interest is the ribozyme mapping to the 5′ UTR of the R2 retrotransposable element in *Drosophila.* R2 elements are non-LTR retrotransposable elements that are cotranscribed with 28S rRNA in many insect genomes (150) and must be processed out in order to retrotranspose into another locus in the genome (151).

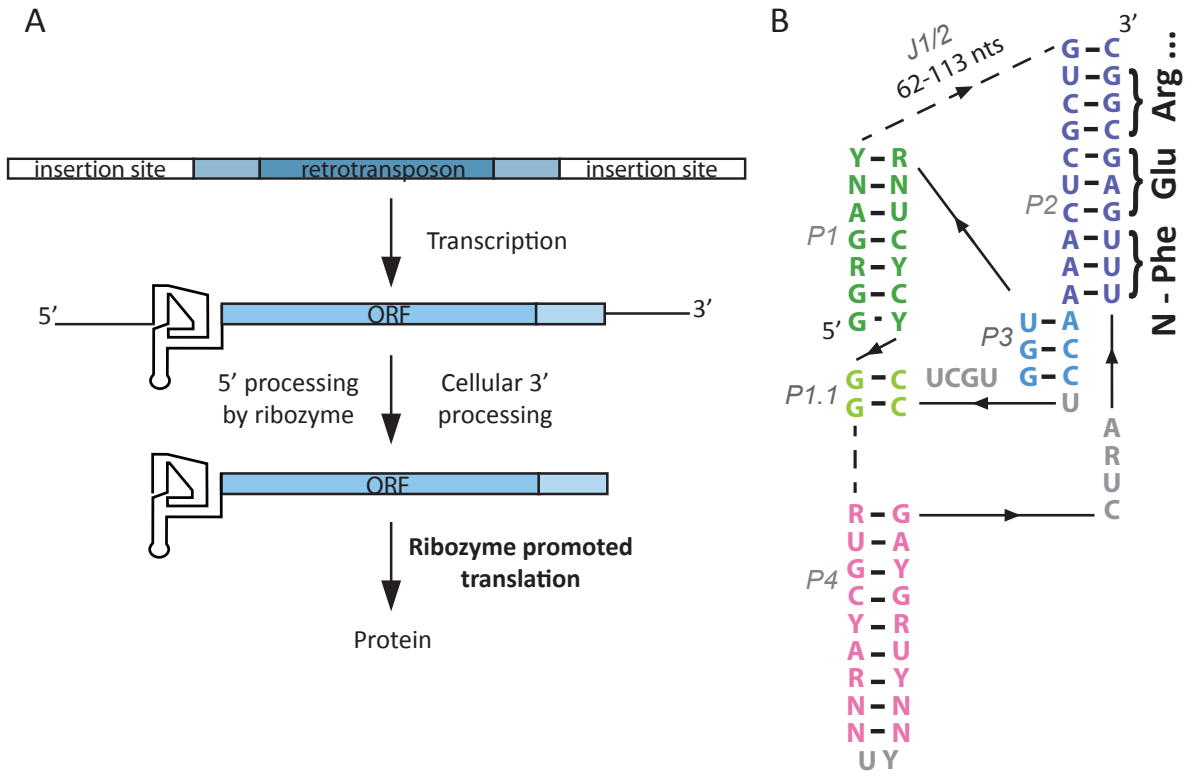


Figure 4-1. The proposed model of retrotransposon RNA processing by drz-Dsim-1. (A) The R2 retrotransposon is co-transcribed with 28S rRNA, and the 5' terminus is generated by the self-cleaving ribozyme in *Drosophila*. The ribozyme forms the 5' UTR of the retrotransposon ORF and aids in the efficiency of translation. (B) The first three proposed amino acids for the R2 protein are predicted to be encoded within the drz-Dsim secondary structure.

Although the R2 is probably the best characterized non-LTR element, the mechanism of cotranscriptional processing has only recently been elucidated. The presence of self-cleaving ribozymes at the 5′ end of retrotransposons leads to a model in which the ribozyme plays several roles in the retrotransposition cycle (32-34), including liberating the 5′ end of the element from the parent transcript. The R2 element terminates with a 5′-OH and the ribozyme forms a very stable structure in which the 5′ terminus is "tucked" within the structure. The ribozyme secondary structure is stable in up to 95 % formamide (37), potentially providing sufficient stability to be resistant to 5′-exonuclease activity. The cleaved ribozyme comprises the entire 5′ UTR of the R2 transcript, suggesting a role in translation of the R2 ORF. The ORF encoded in the R2 transcript is essential to the completion of the retrotransposition cycle; however the mechanism of translation has not been characterized.

The protein translated from the R2 ORF contains a restriction enzyme-like endonuclease domain and a reverse transcriptase domain, both of which are necessary for the continuation of the retrotransposition cycle (152,153). There is no conserved AUG found before the first predicted nucleic acid binding domain (154), however there are non-productive start and stop codons (32,154). Because the product of ribozyme cleavage is a 5′-OH and not a triphosphate, the transcript is not likely to be capped. The absence of a 5′-m7G cap on the R2 message led to the prediction that translation initiation is non-canonical and is similar to the mode of translation used by IRES elements (reviewed in (155)). The presence of a complex pseudoknotted structure together with the observation that the ribozyme comprises the entire 5′ UTR of the R2 element in *Drosophila*, leads to the hypothesis that the ribozyme is directly involved in

translation (Fig. 4-1A). Previous work showed that HDV-like ribozymes promote translation both *in vitro* and *in vivo* (32). Here we investigate the mechanism of translation initiation of ribozyme-terminated mRNAs. Our results indicate that the correct folding of the ribozyme core and composition of the transcript's 5′ end are important in promoting translation. Translation occurs in the absence of a cap, start codon, and 3′ UTR. Taken together our data suggest that this translation is distinct from other cap-independent mechanisms.

Translation initiation *in vitro*

Construct design and expression system

The approach to investigating the impact of HDV-like ribozymes on translation *in vitro* used the expression of a *Firefly luciferase* (Fluc) reporter. The constructs were made by cloning ribozyme and leader sequence upstream of the Fluc ORF. *In vitro* transcription is performed using T7 RNA polymerase and the resulting transcripts are translated *in vitro* using rabbit reticulocyte lysate (RRL). Many of the ribozyme constructs and additionally, the hepatitis C virus (HCV) IRES constructs were made by former lab member Dana Ruminski, Ph.D. The Fluc ORF contains an AUG start codon except in the constructs indicated. In addition, most constructs contain the simian virus 40 (SV40) 3′ UTR, except when indicated. The HCV IRES is used as a positive control for cap independent translation; however it should be noted that the construct contains both an AUG within the structure of the IRES, as it exists in the viral transcript, and the AUG found at the beginning of the Fluc ORF that was cloned into the vector. The distance between the end of the ribozyme secondary structure and the AUG start codon of the Fluc ORF varied for different constructs, ranging from 8-9 nucleotides.

Appendix A lists all constructs which were based on HDV-like ribozymes. drz-Dsim-1: Maps to the 5′ UTR of the R2 retrotransposon in the genome of the fruit fly *Drosophila simulans (29)*. drz-CIV-1: Maps to the genome of the *Chilo* iridescent virus where it is found immediately upstream of a RNA polymerase gene (29). drz-Hsap (CPEB3): Maps to the genome of *Homo sapiens* where it is found in the second intron of the cytoplasmic polyadenylation element binding protein 3 (*CPEB3*) gene (28). hhrII-SewS1_01145s-1: A type II HHR that has been identified in metagenomic sequencing data originating from raw sewage samples (11). HCV-IRES: The HCV-IRES sequence and construct is based on that previously described (156).

Elucidating the mechanism of ribozyme-promoted translation

To understand the mechanism of translation promoted by HDV-like ribozymes, various structural features of the transcript were probed for the influence on Fluc levels generated by *in vitro* translation. In general, different HDV-like ribozymes promote expression of the downstream ORF, but to varying degrees (32,157). In addition, the inactive versions of the ribozymes also promote translation at levels near or above those of their WT counterparts. These observations suggest that the functional group on the 5′ end of the transcript does not impact the ability of the ribozyme structure to promote translation of the downstream ORF. To validate this observation, we examined the difference in a Dsim construct where instead of the 5′-OH that results from self-cleavage, there is a 5′-triphosphate (Dsim-Δleader). The production of this construct is accomplished by having the G_1 position be the first nucleotide transcribed. The results of *in vitro* translation for this new construct demonstrates a 65 % drop in Fluc levels when there is a triphosphate on the 5′ end of the transcript (157). The crystal structures

of HDV ribozymes show that the scissile phosphate is “tucked” into the active site, and so the 5′ end of the transcript may be inaccessible (38,40). NMR data suggest that a single phosphate at the 5′ end disrupts the structure of the ribozyme (158). Therefore, these data suggest that disruption of the HDV ribozyme structure leads to a reduction of *in vitro* translation of the downstream ORF. Other experiments show that adding a 5′-m7G cap to transcripts *in vitro* does not influence ribozyme-promoted translation. It appears that translation is independent of the functional group on the 5′ end of the transcript unless it interferes with ribozyme structure, in which case it can be detrimental to translation.

To decipher which structural features of HDV-like ribozymes are directly influencing the levels of Fluc activity, peripheral domains outside of the ribozyme core were systematically tested. The J1/2 region linking P1 and P2 can vary from a linker of only a few nucleotides to large stable helices (32,149). The R2-associated ribozymes typically contain long structured J1/2 elements, other ribozymes such as drz-Hsap-CPEB3 and drz-CIV do not, and yet they are still very efficient at promoting translation *in vitro*. The P4 domain is typically a structured helix that can vary in length in different ribozymes (32,149). However, smaller HDV-like ribozymes missing this helix have been found in nature and they portray slow rates of self-cleavage *in vitro* (137). Deletion of the P4 helix (Dsim-ΔP4) maintains Fluc activity at levels similar to Dsim-WT (157). The ability of HDV-like ribozymes to promote translation when structural features peripheral to the catalytic core are absent suggests the core itself is necessary for recognition by the translation machinery.

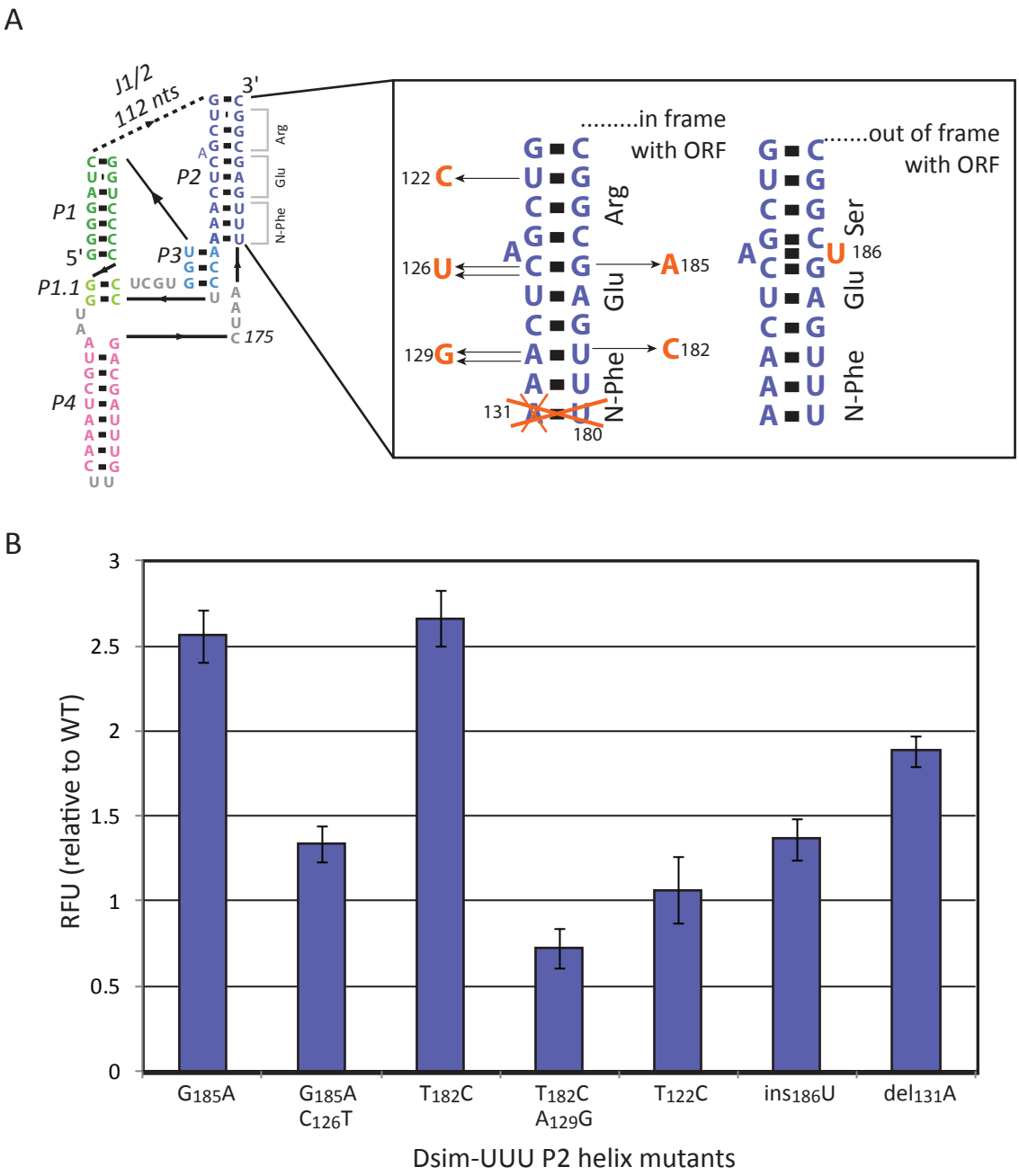


Figure 4-2. *In vitro* translation promoted by drz-Dsim-1 constructs with mutations in the P2 helix. (A) A series of single and double point mutants were created for drz-Dsim-UUU which contains the WT ribozyme but the start codon for the downstream Fluc ORF has been mutated to UUU. The single arrows denote a point mutation and the double arrows denote the second mutation in a series to restore base-pairing at the indicated positions. All the mutants displayed self-cleavage activity during *in vitro* transcription. The mutants were designed to disrupt or enhance the thermostability of the P2 helix. The P2 helix is predicted to encode the first three amino acids of the R2 protein in *Drosophila*. An additional single nucleotide insertion was predicted to disrupt the register leading to an absence of *Firefly luciferase* (Fluc) activity. (B) All the P2 helix mutants promoted the translation of Fluc during *in vitro* translation. Destabilization of the P2 helix was accomplished by disrupting base-pairs and this lead to an increase in Fluc activity over WT. In some cases, increasing the predicted thermostability of the P2 helix dropped Fluc activity below WT.

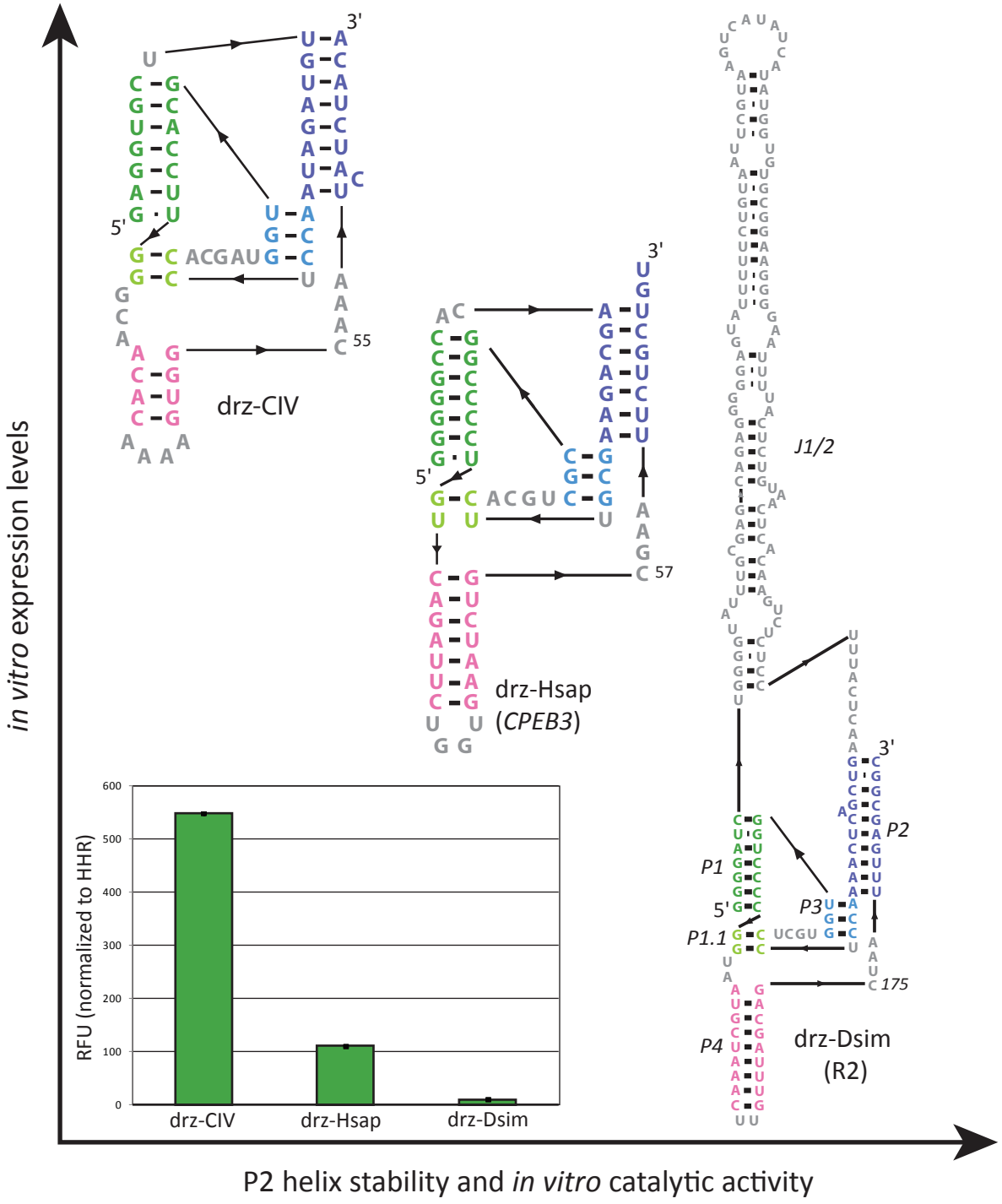


Figure 4-3. The efficiency to promote i*n vitro* translation and *in vitro* self-cleavage are inversely correlated for HDV-like ribozymes. The accumulation of *in vitro* self-cleavage experiments for known ribozymes has demonstrated that stronger thermodynamic stability, in some cases provided by extensive structured peripheral domains, such as for rbz-Dsim-1, lead to faster rates of catalysis. Our *in vitro* translation experiments reveal that HDV-like ribozymes the strongest thermodynamic stability produce the lowest levels of Fluc activity (inset). All data are average values ± the average deviation of duplicate experiments.

The current model of translation promoted by ribozymes found in the 5′ UTR of R2 retrotransposons suggests that the first amino acids of the R2 ORF are encoded within the P2 helix of the HDV-like ribozymes (Fig. 4-2A). To investigate the role of P2 in promoting translation of the downstream ORF a variety of mutants were designed to test aspects of ribozyme structural stability and also the role of the ribozyme in dictating the site of translation initiation. *In vitro* translation experiments reveal a correlation between the thermostability of the P2 helix and the levels of Fluc activity (Fig. 4-2B). When the stability of the P2 helix is reduced by a point mutation disrupting the formation of a base pair, as in the case of Dsim-g185a, Dsim-t182c, or Dsim-Δa131, there is an appreciable increase in Fluc activity when compared to Dsim-WT. A double mutation restoring the base-pairing but with a different base pair reduces Fluc activity back down near WT levels but in a manner dependent on the change in stability of the respective G●C or A-U base pair. If the replacement base-pair is more stable than WT, we observed a slight reduction in Fluc activity, as in the case of Dsim-t182c-a129g. The exceptions to this general trend are Dsim-t122c and Dsim-185ins, both of which were designed with the intention to further stabilize the P2 helix above WT. Both of these constructs lead to Fluc activity near WT levels.

These data can be interpreted in two ways: 1) destabilizing the ribozyme leads to enhancement of translation, perhaps by reducing catalysis, or 2) unwinding of the P2 helix, thought to encode the first amino acids of the R2 ORF in the natural context, is necessary for promoting translation. Previous experiments demonstrated that in some cases the inactive version (C/U mutant) of some ribozymes lead to higher levels of Fluc activity compared to their WT counterparts (32) and so we cannot attribute the changes

in Fluc activity by point mutations to be dependent on catalysis by the ribozyme. Furthermore, generating constructs with multiple mutations that further destabilized the P2 helix led to variable affects on Fluc activity generated by *in vitro* translation (data not shown). Those results are attributed to complete disruption of the correct formation of the HDV-like ribozyme core.

Multiple lines of evidence suggest that ribozyme-promoted translation is enhanced when the structure of the ribozyme displays weaker thermostability. However, there is a tipping point where weakening of the ribozyme structure disrupts overall folding of the catalytically competent form of the ribozyme. The accumulation of *in vitro* translation data suggest that the correct formation of a HDV-like ribozyme core is essential to ribozyme-promoted translation; however, the stability of that core structure inversely correlates with the amount of Fluc activity generated by *in vitro* translation assays (Fig. 4-3). This trend may have significant implications on the molecular interactions between the ribozyme and translation machinery. The fact that stability of the ribozyme tunes the level of translation suggests 1) the ribozyme is directly interacting with factors necessary for translation or 2) the structure of the ribozyme is influencing the surrounding transcript, thereby impacting translation. We probed other features of the Fluc transcript to distinguish which model can account for our observations.

The ability of the HDV-like ribozyme constructs to promote translation of the downstream ORF in a cap-independent manner suggests a non-canonical mode of translation initiation. While cap-dependent translation exclusively utilizes a scanning mechanism to find the AUG start codon, some IRESs also rely on ribosomal scanning but others do not utilize a canonical start codon at all (159). To investigate the mode of

translation initiation in ribozyme-terminated mRNAs, various mutations were made to the AUG start codon at the beginning of the Fluc ORF. Some HDV-like ribozymes such as drz-Fpra-1 and drz-CIV map immediately upstream of ORFs containing canonical AUG start codons, but others such as drz-Hsap-CPEB3 and the R2-associated ribozymes are not positioned near conserved AUG start codons (29). In most cases, the first codon for the R2 ORF in *Drosophila* is predicted to be UUU (Fig. 4-1B) (154). *In vitro* translation of constructs containing any mutation to the AUG codon results in ~90 % decrease in Fluc levels (Fig. 4-4) (157). There is one exception to this generalization: drz-CIV-UUU shows enhanced Fluc activity over drz-CIV-AUG (Fig. 4-4). The observation that mutations of the start codon cause a decrease in Fluc activity suggests that translation is enhanced by an AUG start codon, but that it can be initiated without one, albeit at a cost to expression efficiency. While we cannot conclude that HDV-like ribozymes are behaving as IRESs, translation from our constructs does portray characteristics that are distinct from CITEs. CITEs are cap-independent translation enhancers that are found in the 3′ UTR of plant viruses and are thought to promote translation by delivering translation machinery to the 5′ UTR (160). This mode of cap-independent translation relies on ribosomal scanning for a canonical AUG start codon.

To further characterize the mode of ribozyme-promoted translation we analyzed the influence of the 3′ UTR on *in vitro* translation. The constructs used in our expression assays contain a 3′ UTR composed of the simian virus 40 (SV40) polyadenylation signal. Poly-A tails are important structural and functional elements of eukaryotic mRNA. They are involved in mRNA stability (161) and signaling for mRNA export from the nucleus (162). They have been shown to impact transcription termination (163) and

also splicing (164). Polyadenylation is a hallmark of mRNA maturation that can enhance translation efficiency (165). By preventing the incorporation of the viral 3′ UTR on the Fluc ORF containing transcript, we observe that Fluc activity drops by an order of magnitude in all cases (Fig. 4-4). The construct HHR2-WT-AUG served as negative control and the absence of the 3' UTR (HHR2-WT-AUG-Δ3'UTR) did not impact translation. These data suggest that ribozyme-promoted translation may be enhanced by polyadenylation or a stable 3′ UTR. The fact that translation in this system does not depend on a polyadenylation signal suggests that the UTR is stabilizing the message, perhaps by providing protection from 3′-5′ exonucleases present in RRL.

Critical analysis of the structural features of the RNA transcript influencing translation of the luciferase reporter has provided insight to ribozyme-promoted translation. The non-canonical mode of translation from this system is independent of 5′ cap and 3′ UTR or polyadenylation signal. Translation relies on correct formation of the HDV ribozyme core and prefers, but does not rely on, an AUG start codon at the beginning of the ORF. Making various point mutations impacts translation efficiency, suggesting the ribozyme is doing more than just extending the lifetime of the transcript. Additionally, those mutations bring to light a correlation between ribozyme thermostability and Fluc activity. Data from the P2 mutants together with the observation that Dsim-C175U promotes translation greater than Dsim-WT suggests that the thermostability and not necessarily self-cleavage is impacting the kinetics of translation initiation. This finding leads to the hypothesis that the ribozyme core is involved in some molecular recognition and that an optimal level of unwinding of the P2 helix is beneficial to the mode of translation.

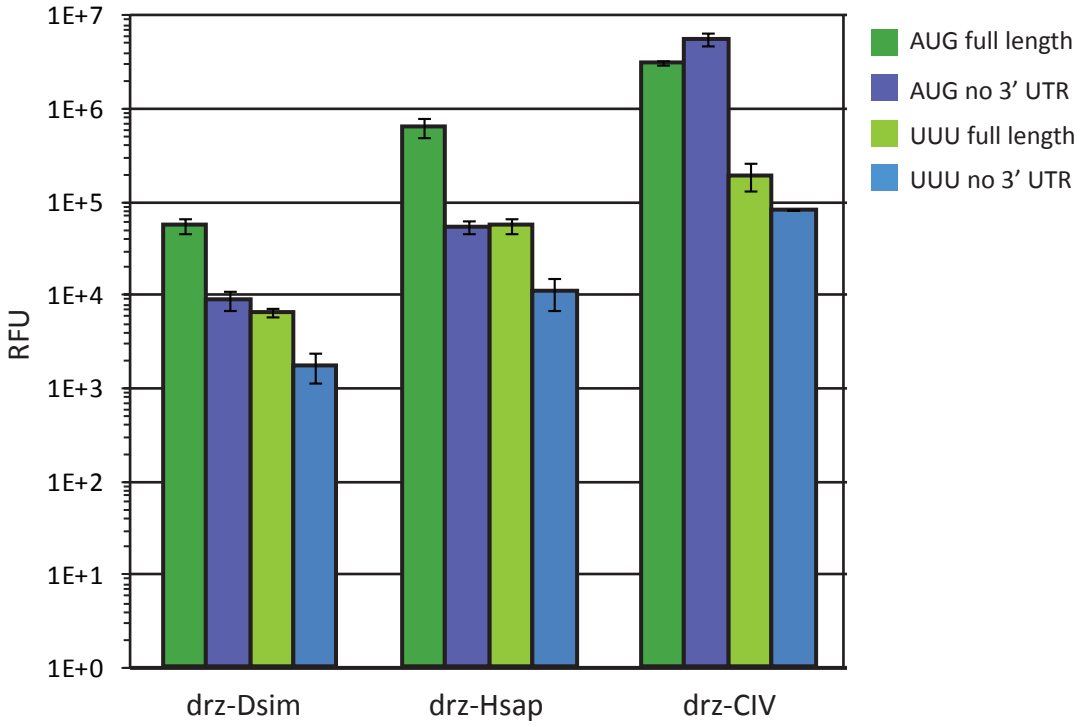


Figure 4-4. *In vitro* translation promoted by HDV-like ribozymes. Different mutations of the transcript containing the indicated ribozymes and the *Firefly luciferase* (Fluc) ORF were used to characterize the features utilized by ribozyme-promoted translation. Deletion of the 3' UTR leads to a ~10-fold reduction in measured Fluc activity, except in the case of drz-CIV-AUG. Deletion of the start codon also leads to a ~10-fold reduction in measured Fluc activity, however translation occurs without a canonical AUG start codon. All data are average values ± the average deviation of duplicate experiments.

To investigate the recognition of HDV-like ribozymes by a molecular factor in RRL, we used toeprinting assays. Toeprinting analysis is a technique used to map the protein binding site on RNA of interest. This method employs a 5′ end labeled primer in a reverse transcription (RT) reaction where protein binding sites can be observed as stops in the RT when run on a high resolution denaturing PAGE (166,167). Toeprint analysis of Dsim-WT and CPEB3-WT with various start codon mutations demonstrate hard stops of RT corresponding to the beginning of the ORF (157). These stops are more pronounced with the canonical AUG start codon. Additional stops are seen within the linker region or at the 3′ end of the ribozyme structure for Dsim-WT and CPEB3-WT ribozymes, respectively. These observations can be attributed to the ribozyme or a ribozyme binding factor. The proportion of full-length RT product suggests that the RT enzyme is able to move through the ribozyme structure and so the toeprint pattern is not attributed to RT stops at the ribozyme. To determine if ribosomes are responsible for the toeprinting patterns observed, the antibiotic puromycin was used to prevent 60S subunit joining, thereby stalling bound 40S subunits. When puromycin is present the same toeprinting pattern occurs, suggesting that ribosomal subunits are not responsible for the observations (157). All toeprint analysis was carried out on translation reactions producing Fluc, meaning that the transcripts present were competent for translation. The presence of a factor remaining bound to the ribozyme during both the translation and RT reactions may suggest that the interaction is strong or that the factor remains bound during translation. The latter supporting the observation that ribosomes are not responsible for the toeprint patterns observed. This means that the non-canonical mode

of translation promoted by HDV-like ribozymes relies on a factor present in RRL to recruit translation machinery.

The analysis of toeprinting patterns for HDV-like ribozymes located upstream of the Fluc ORF suggests that there is a factor in RRL binding ribozymes: 1) a factor recognizing the ribozyme remains bound during active translation and 2) the ribozyme plays a role in the site of translation initiation. Additional toeprint assays demonstrate that the same toeprint pattern occurs regardless of mutations to the AUG start codon (157). This observation suggests that the mode of translation promoted by ribozymes relies on the ribozyme defining the register in which the ORF is read, since ribosomal scanning mechanisms would not lead to Fluc production when AUG is mutated. While the toeprinting data generally suggest that there is a factor in RRL binding HDV-like ribozymes, the specificity of the interaction and its role in translation was unclear.

In addition to data that demonstrate the ribozyme promotes translation regardless of linker length and despite the absence of a canonical start codon, data from the P2 helix mutants support the idea that the ribozyme does not define the translation start site but are involved in defining the register. The P2 helix insertion mutant Dsim-185ins promotes translation at WT levels (Fig. 4-2B). This mutant has a single nucleotide insertion that would disrupt the translation register if the first three amino acids are encoded within the P2 helix. Given that translation of Fluc occurs with different constructs varying by linker region lengths between the start codon and ribozyme, and also that single nucleotide insertions within the structure do not interfere with translation, we hypothesize that the ribozyme is presenting a translation start site that is not encoded within the ribozyme structure. Toeprinting assays demonstrate that the

ribozyme is bound by a factor in RRL that may act to recruit translation machinery. Therefore, we attempted to identify the factors in RRL recognizing HDV-like ribozymes in order to further characterize the mode of translation initiation.

Characterizing ribozyme-protein interactions *in vitro*

NatEMSA method and construct design

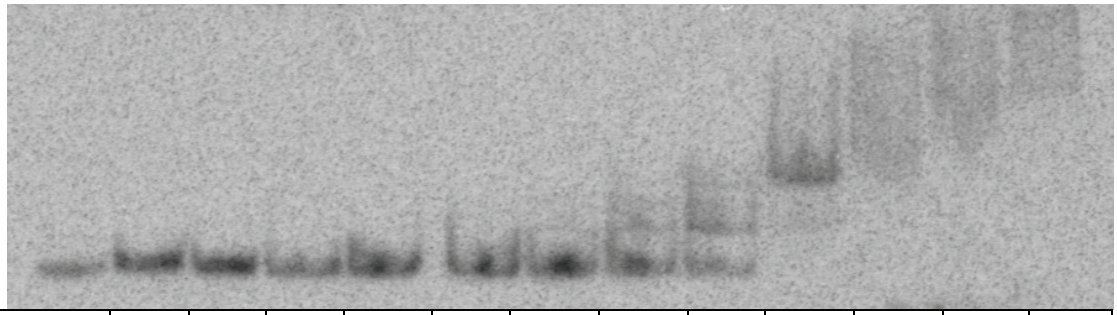
We used binding assays analyzed by electrophoretic mobility shift to elucidate the mechanism of translation machinery recruitment to the 5′ UTR of the reporter ORF. Electrophoretic mobility shift assays under native conditions (natEMSA) can be used to observe RNA-protein interactions by allowing visualization of high affinity binding resulting in retardation of the radiolabeled RNA of interest on a non-denaturing polyacrylamide gel (168). Toeprint analyses of full length reporter constructs showed that the RT experiences a hard stop when it encounters the ribozyme structure; however this assay does not define the exact cause for the stop. To provide direct evidence for specific protein binding we designed a natEMSA protocol to test protein binding to cleaved ribozymes with short unstructured tails that do not contain an AUG or other start codon. These constructs were transcribed by T7 RNA polymerase and gel purified. The self-cleaved product was excised from the gel and end-labeled using ^{32}P-pCp. In some cases, the RNA was body-labeled during transcription using ^{32}P-ATP. The RNAs used in the assay contained a 5′-OH and a short tail, the length given with the construct name. Binding conditions are the same as those used during *in vitro* translation, except that amino acids are left out and tRNA is added. Excess of tRNA serves to eliminate non-specific RNA-protein interactions. The composition of the RRL

was left intact in most cases, but fractionation of the RRL was also employed to enrich for the initial factor binding the ribozyme.

NatEMSA demonstrates ribozyme-specific protein binding in RRL

We predict that the factor(s) in RRL binding HDV-like ribozymes can be isolated by affinity methods. Protein binding assays can be used to observe protein assembly and characterize the affinity of interactions with the RNA of interest. Qualitative experiments with different ribozyme constructs demonstrate differential electrophoretic mobility shift in the presence of RRL. An example of this shift is seen in Fig. 4-5 for rbz-Dsim-7. At a fixed concentration of RNA a gradual shift upwards occurs with increasing amount of RRL. This effect is specific for the HDV-like ribozymes tested, and the gradual steps upward are most distinct for the rbz-Dsim-7 construct; therefore, it was chosen as a model for EMSA analysis. The justification for this choice is that the earliest protein binding events are the best resolved for this ribozyme in these assays. The specific recognition of HDV-like ribozymes by proteins in RRL suggests that the ribozymes are driving translation by participating in protein assembly on the 5′ end of the transcript. These observations validate those made in the toeprint analysis but are able to exclude non-specific RNA-protein interactions and RT stops caused by RNA secondary structure. Further, the observation of sequential binding of proteins into larger complexes, resulting in slower electrophoretic mobility, suggests that the ribozyme or the initial ribozyme binding proteins recruit additional macromolecules, potentially translation machinery.

To characterize the stability of the complexes we observe in the natEMSAs, we monitored the ability of the ribozyme to remain in complex under various conditions.



45 mM KCl	X											
2 mM $MgCl_2$		X										
Translation buffer			X	X	X	X	X	X	X	X	X	X
800 nM tRNA				X	X	X	X	X	X	X	X	X
RRL (v/v)					0.1%	0.3%	0.5%	1%	5%	10%	25%	50%

Figure 4-5. Native electromobility shift assays demonstrate macromolecules in RRL bind rbz-Dsim-7. Rabbit reticulocyte lysate (RRL) is the system used for in vitro translation experiments. The 3' end radiolabeled RNA containing the product of self-cleavage by rbz-Dsim-7 was used in a binding assay with RRL. Various salt conditions demonstrate that binding by macromolecules and not the conditions of the binding assay are responsible for the retardation in electromobility by rbz-Dsim-7 in the presence of RRL. Increasing amount (v/v) of RRL results in different complexes with different electromobility on the 10 % native polyacrylamide gel.

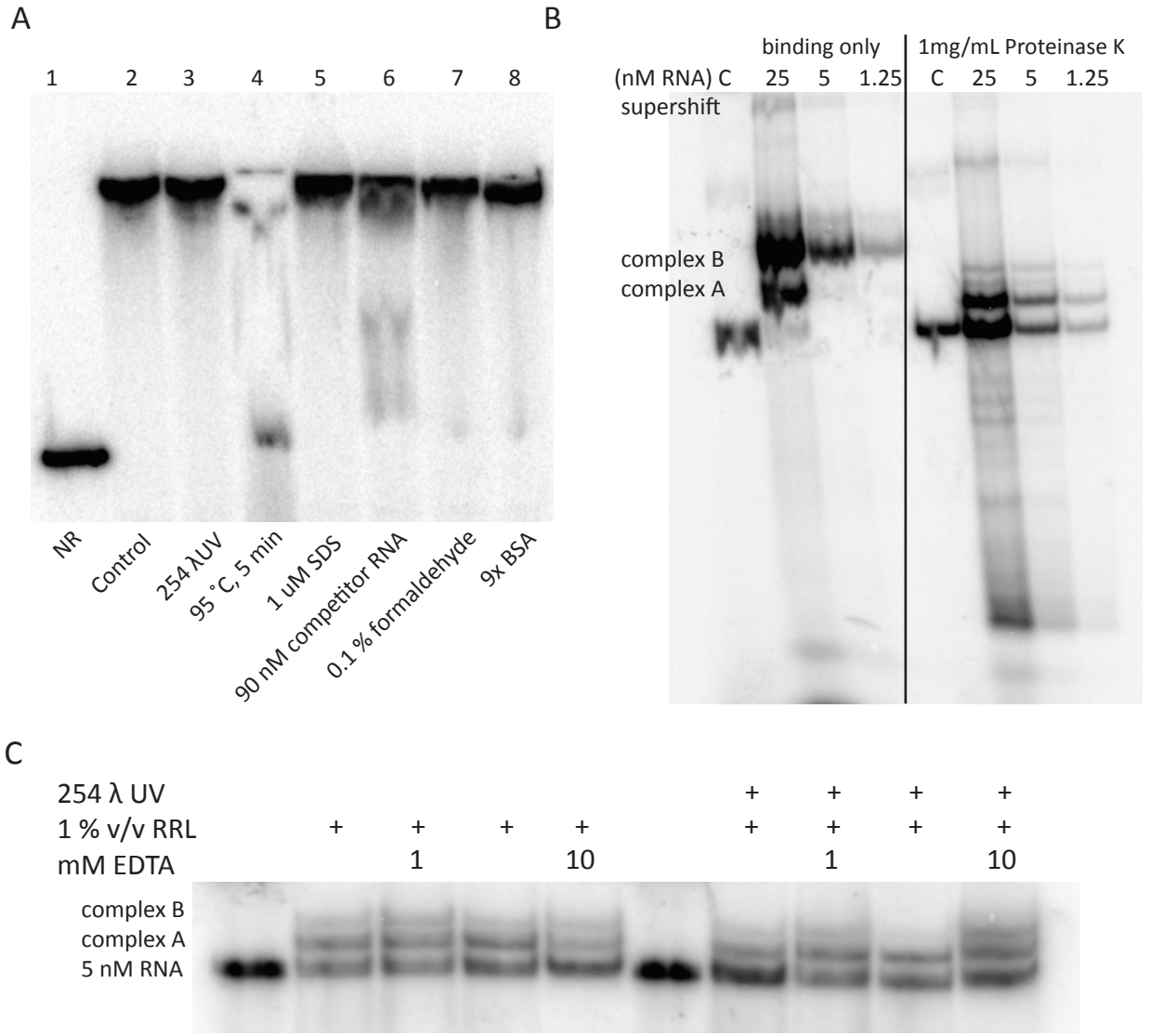


Figure 4-6. Highly denaturing conditions confirm proteins in RRL specifically recognize rbz-Dsim-7. (A) The integrity of the complexes formed by interaction between rbz-Dsim-7 and factors in RRL was tested under the given conditions and run on a denaturing polyacrylamide gel. Some conditions were designed to crosslink the macromolecules to prevent separation on denaturing PAGE (UV and formaldehyde), which was not necessary as the control lane demonstrates that binding survives electrophoresis in a 7 M polyacrylamide gel. Low concentration of SDS was not efficient at denaturing the interaction. However, complex assembly was disrupted by either 95 ˚C incubation for 5 min prior to loading or by competition with excess unlabeled rbz-Dsim-7 competitor RNA. Bovine serum albumin (BSA) is used as a crowding agent for in vitro reactions and did not seem to impact complex assembly. (B) Digestion with proteinase K, resolved on a non-denaturing polyacrylamide gel, demonstrates that the supershift and complex B are composed of proteins whereas complex A is protected from proteinase digestion, or is not a protein. (C) However, a non-denaturing gel resolving complex assembly of rbz-Dsim-7 in 1 % v/v RRL demonstrate that the addition of EDTA, a chelating agent, does not destroy complex A suggesting it is not made up of trans-RNA or RNA:DNA interactions. Subjecting binding reactions to various harsh conditions demonstrates that the complexes observed on natEMSA polyacrylamide gels are ribozyme-protein complexes.

Under partially denaturing conditions, we observe that when the bound complex is placed in 3 M urea, partially denaturing conditions, and then ran on a 7 M, denaturing gel, high concentrations of RRL result in slower electrophoretic mobility for rbz-Dsim-7 (Fig. 4-6A). This result suggests that the affinity of the ribozyme-protein interaction is very strong and other data demonstrate that some complexes can only be denatured by incubation at 95˚C or competition with a HDV-like ribozyme (Fig. 4-6A). In addition, protein digestion with proteinase K also destroys most complex assembly with the exception of complex A (Fig. 4-6B). It appears that the ribozyme is providing some protection against proteinase K digestion to the factor directly bound to it. Together, data from the exposure of these complexes to various denaturing conditions demonstrate that all complexes apart from complex A are proteins.

Further, we probed the capacity for complex formation with chelating agents and by limiting the flexibility of ribozyme structure. In order to investigate the impact of a rigid/fixed RNA secondary and tertiary structure on protein binding we performed cross-linking studies where the RNA is crosslinked via 254 λUV prior to incubation with RRL. There were no observable differences in natEMSA pattern between the cross-linked and native versions of rbz-Dsim-7. However, the presence of EDTA during the binding incubation seems to cause a favorable shift towards complex B assembly when compared to uncrosslinked control, as seen at 1 % and 10 % v/v RRL (Fig. 4-6C). In addition, 10 mM EDTA did not disrupt complex A or B assembly suggesting that neither complex is the result of *trans*-RNA interactions; we therefore refer to all binding events observed during natEMSAs as ribozyme-protein interactions.

A number of HDV-like ribozymes demonstrate protein binding in our assay including rbz-CPEB3, rbz-Fpra-1, full-Dsim-C158U-7, and rbz-CIV. In addition to the wild-type, self-cleaved rbz-Dsim-7, we also observed protein binding for a number of Dsim mutants. One set of Dsim mutants include truncated versions of the ribozyme that do not necessarily yield an active ribozyme structure, but maintain different structural features of the HDV ribozyme core. There was no electrophoretic mobility shift observed for the constructs Dsim-P2-7 and Dsim-P1.1-7, neither of which produces an active ribozyme. Binding was only observed for Dsim-Δleader-7 and rbz-Dsim-ΔP4-7, both on which should fold into the correct HDV-like ribozyme core, the latter demonstrating self-cleavage *in vitro*. In addition, these mutants were also shown to compete for binding against the rbz-Dsim-7 WT (Fig. 4-7). These data further suggest that Dsim-Δleader-7 is initially bound by proteins with the same affinity as the WT counterpart. While we have not measured the relative K_D for any of the binding we observed, the qualitatively similar EMSA pattern observed with titrating RNA against 10 % v/v RRL (Fig. 4-8A and 4-8B) suggests a similar affinity. This result is surprising given the significant difference seen during *in vitro* translation, where the Dsim-Δleader construct displayed significantly low Fluc activity.

These data suggest that the instability of the HDV core caused by the presence of a 5′-triphosphate, as in the case of Dsim-Δleader-7 (Fig. 4-8B), does not interrupt ribozyme recognition but it does negatively affect a downstream protein binding event leading to a reduction in translation. This finding suggests that the initial protein(s) recognizing and binding the ribozyme are not part of the translation machinery but perhaps serve as a scaffold for a translation competent protein complex.

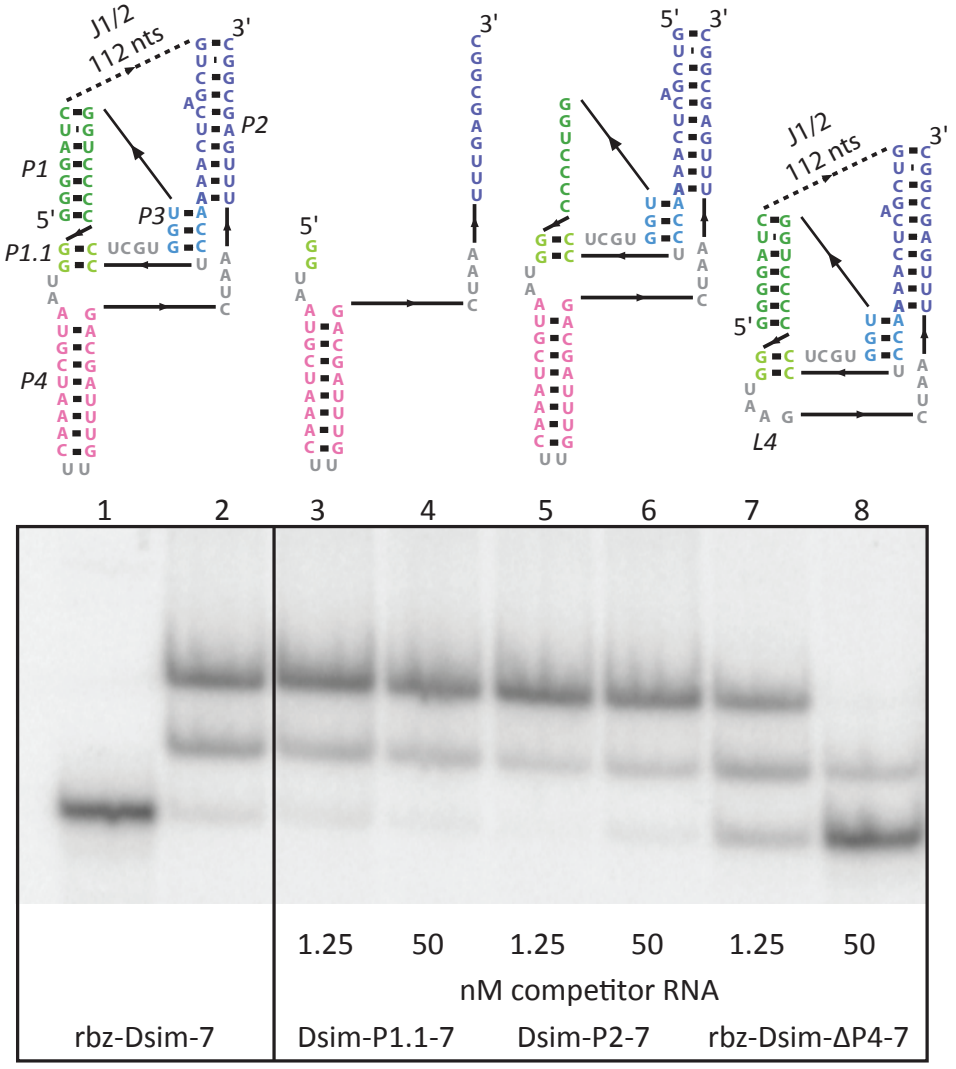
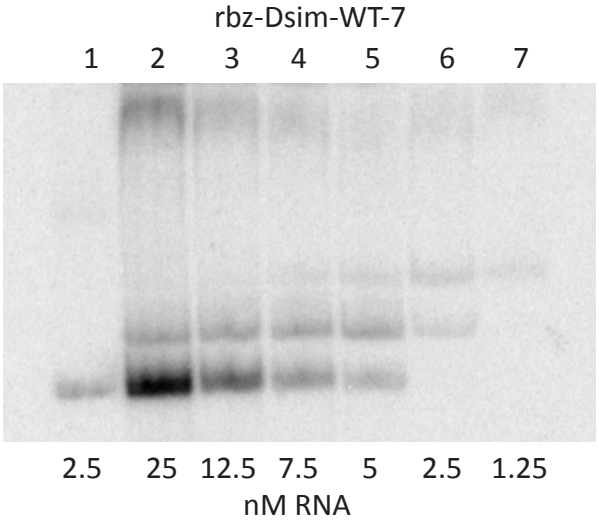
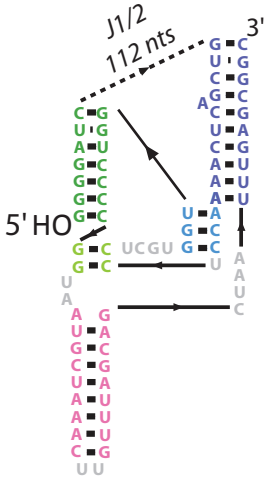


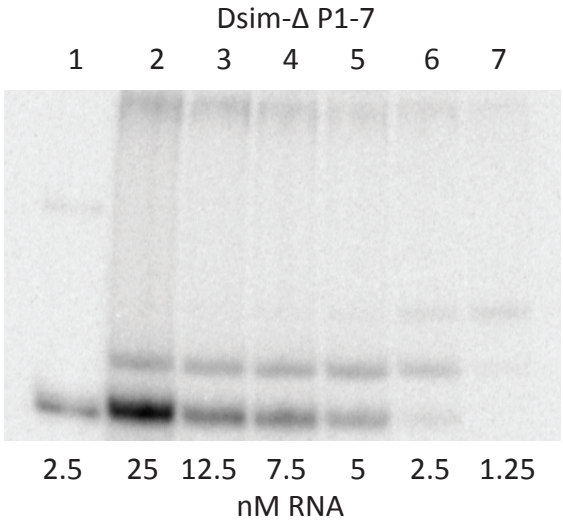
Figure 4-7. Interference assay demonstrates complex assembly with rbz-Dsim-7 can be disrupted by the presence rbz-Dsim-ΔP4-7 but not sequences from the rbz-Dsim core. These three rbz-Dsim-based constructs were designed to test which features of the ribozyme are recognized by factors in RRL during natEMSAs. The concentration of 3' end radiolabeled rbz-Dsim-7 is 5 nM in all reactions. The deletion of the P4 helix (rbz-Dsim-ΔP4-7) is still an active ribozyme in vitro and is able to compete for binding. Constructs Dsim-P1.1 and Dsim-P2 are not active ribozymes and the inability to compete for binding against rbz-Dsim-7 demonstrates that it is the structure of the ribozyme core and not sequences within it that are being recognized by factors in RRL leading to complex assembly in natEMSAs.

A





B



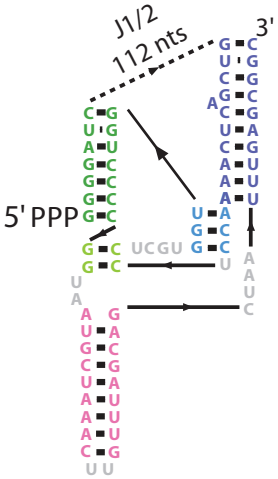


Figure 4-8. NatEMSAs demonstrate similar protein binding affinities for rbz-Dsim-7 and rbz-Dsim-Δleader-7. These two constructs differ by the functional group on the 5' end which is (A) a hydroxyl for the self-cleaving ribozyme and (B) a triphosphate for the Δleader construct. The similar protein affinity for the two constructs is assumed based on the qualitatively similar natEMSA patterns observed with increasing concentrations of RNA at 10 % v/v RRL.

The unanticipated similarity in protein binding between Dsim-Δleader-7 and Dsim-7 together with the difference in Fluc activities suggests that multiple proteins are making direct contact with the ribozyme or that the initial protein recognizing the ribozyme is acting on it, in which case some structural perturbations would impede this kinetic event.

NatEMSA patterns similar to rbz-Dsim-7 were observed for rbz-Dsim-19, rbz-Dsim-g185a-19, rbz-Dsim-t182c-19, and rbz-Dsim-del131a-19. The three Dsim mutants contain single nucleotide changes in the P2 helix and lead to enhanced Fluc expression *in vitro*. However, based on the protein-binding assay, we cannot conclude that the differences in Fluc activity generated by the mutants are due to enhanced affinity for the protein(s) recognizing the ribozyme. The observations suggest, similarly to Dsim-Δleader, that slight perturbations in ribozyme stability do not affect the ability of the recognition protein to bind the ribozyme, but those perturbations influence some downstream event leading to translation. The favorable shift towards complex B assembly when rbz-Dsim-7 was crosslinked and exposed to EDTA suggests that the protein(s) in complex B is/are responsible for sensing the stability of the ribozyme structure. Future work should test other Dsim mutants demonstrating reduced ability to promote translation in combination with EDTA titrations. The results of those binding assays will provide further evidence for a specific protein sensitive to ribozyme thermostability and that this factor might ultimately be the rate limiting step in translation initiation promoted by HDV-like ribozymes.

It appears that the ribozyme-protein interactions are strong considering complexes survive denaturing conditions and protein digestion. We utilized RNA structure probing

techniques in order to characterize ribozyme-protein interactions. In-line probing is a common technique used to probe features of RNA structure and ligand interactions. Every RNA sequence will have a unique degradation pattern, with highly structured features less prone to degradation. In addition, the pattern changes upon structural rearrangement or ligand binding. We designed in-line degradation experiments for rbz-Dsim-7 and Dsim-Δleader-7. These two constructs have been shown to promote different levels of Fluc activity during *in vitro* translation but portray similar protein binding affinities. As expected, the in-line degradation patterns for the two constructs are slightly different within L3 and P1.1 regions (Fig. 4-9A). In addition, the in-line degradation patterns for both constructs change slightly in the presence of RRL, however RRL seems to speed up degradation and there are no sites of protection, possibly due to residual RNase activity in RRL. Therefore, we cannot conclude that the change of degradation in the presence of RRL is due to protein interactions, as an increase in divalent metal cations present in RRL would also cause this effect. However, when the reaction is treated with proteinase K prior to extended incubation, allowing for the in-line degradation to proceed, new sites of protection are observed. Previous experiments showed that the ribozyme protected the protein(s) in complex A from proteinase K digestion. By treating with proteinase K the in-line degradation pattern should reflect protection or RNA structural rearrangements attributable to the protein(s) during early complex formation. The sites of protection are the same for both rbz-Dsim-7 and Dsim-Δleader-7; however there is a region in Dsim-Δleader-7 that experiences significant rearrangement in the presence of RRL after proteinase K treatment. In that region, two sites are protected while a site in the middle becomes more susceptible to

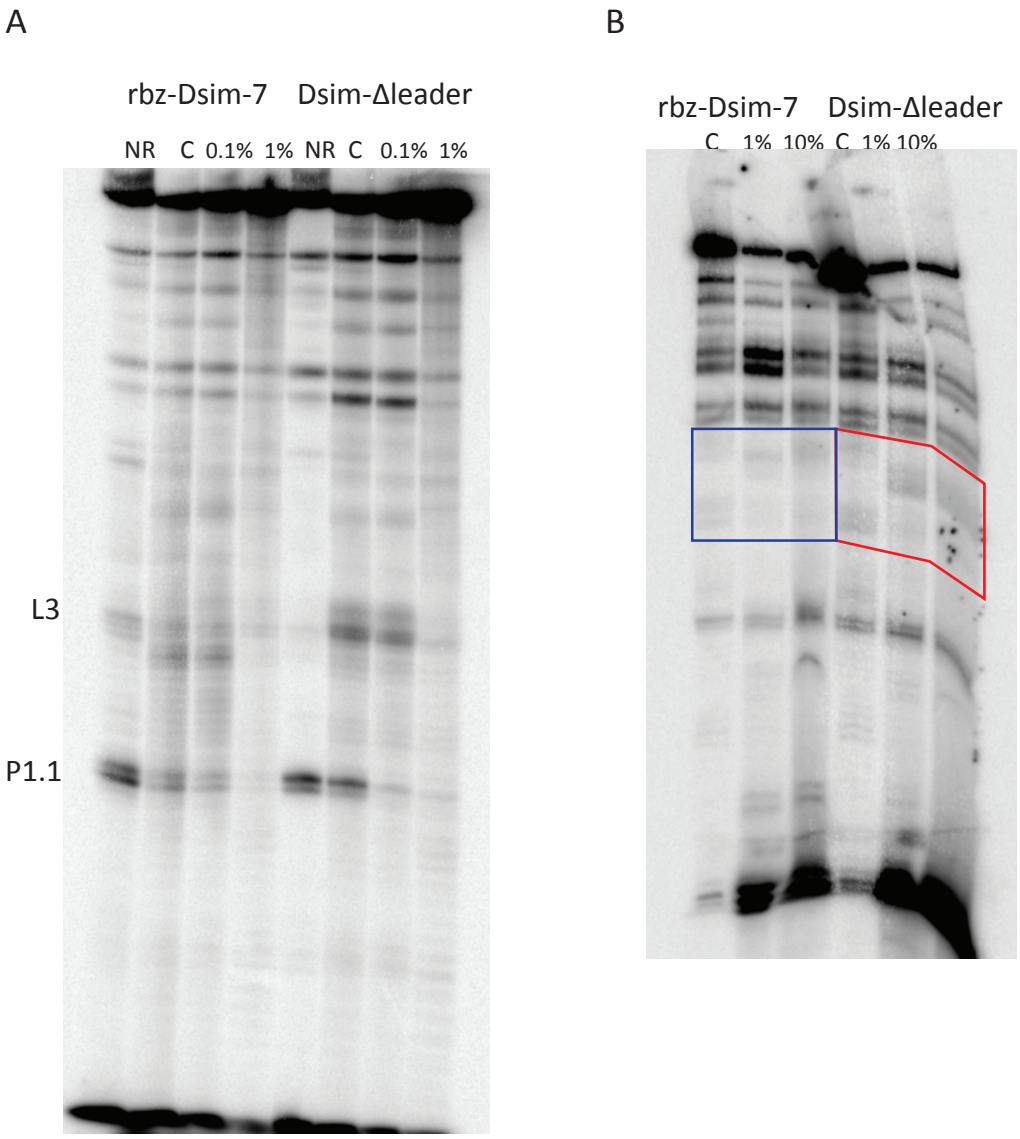


Figure 4-9. In-line probing experiments reveal subtle differences in structural stability between rbz-Dsim-7 and rbz-Dsim-Δleader-7 in RRL. In-line probing experiments were performed by incubating the RNA or RNA+RRL for 24 hrs at 37˚C. (A)The strongest regions of changes in degradation pattern in the presence of RRL are annotated as structural features L3 and P1.1 of the HDV-like ribozyme. The variation in degradation between rbz-Dsim-7 and rbz-Dsim-Δleader-7 is expected due to the destabilization of the HDV core structure when there is a 5'-triphosphate. It is difficult to see a trend with increasing amounts of RRL as it seems the RNA was being protected from degradation, based on the increasing intensity of full-length RNA with higher amounts of RRL. (B) Proteinase K digestion was employed to eliminate non-specific RNA binding. Previous experiments demonstrate that the proteins causing complex A during natEMSAs is protected from proteinase K digest. Therefore, we expect the degradation pattern to reflect binding of complex A proteins. The regions of significant changes in susceptibility to degradation are boxed, in blue for rbz-Dsim-7 and in red for rbz-Dsim-Δleader-7.

degradation (Fig. 4-9B). This pattern of protection and susceptibility to degradation does not occur for rbz-Dsim-7. Future work will map this region back to the ribozyme secondary structure. This region of the ribozyme must be important for protein interactions and may explain how slight differences in thermostability of the HDV-like ribozyme core lead to dramatic differences in the ability of the ribozyme to promote translation *in vitro*.

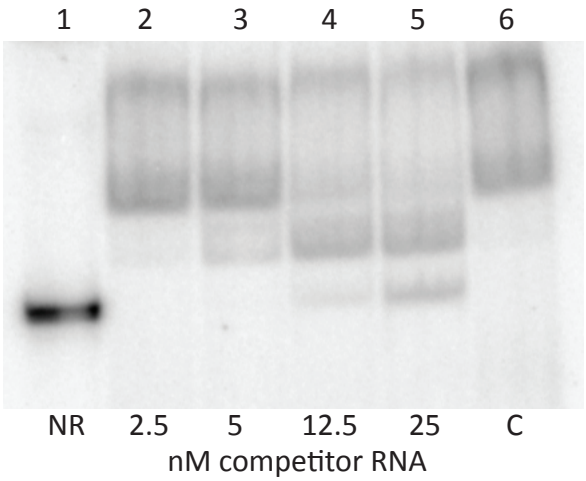
To further characterize the initial ribozyme-protein binding events, we attempted to understand the stoichiometry of the RNA-protein complexes we observe in our natEMSAs. We chose to study the titration of RNA at 10 % v/v RRL because this demonstrated the best resolution of early protein binding events for rbz-Dsim-7. When the amount of RRL is held constant and the RNA is titrated, we observe a differential titration for certain RNA-protein complexes. An example for rbz-Dsim-7 is shown in Fig. 4-8A. These data demonstrate that at very low concentrations of RNA, all the RNA is bound by a macromolecular factor from RRL. Increasing concentration of RNA does not lead to progressive protein complex assembly, such as when RRL is titrated against a fixed amount of RNA. Instead it seems that the protein(s) causing complex A is/are saturated. Therefore, an increasing amount of RNA does not cause a larger amount of complex A as determined by densitometry. On the other hand, complex B is titrated away and we observe an increase in the amount of a supershifted complex and in unbound RNA. These observations suggest that complex B leads to the production of supershifted complexes and thereby, presumably, to translation machinery assembly. These titration experiments also suggest that the protein causing complex A is the limiting factor in supershifted complex assembly. This titration may be due to low

concentrations of the protein in 10 % RRL or to a high dissociation rate for complex A. Considering that initial complex assembly is only observed at less than or equal to 10 % v/v of RRL and low nM RNA, it is more likely that the protein is in very low abundance in those dilutions of RRL. Based on the current RNA titration data, we predict that the concentration of the protein causing complex A to be approximately ~5 nM in 10 % v/v of RRL, suggesting the concentration to be approximately ~50 nM in 100 % RRL. This concentration is within a typical range for proteins in the cell and suggests the protein is not generally in low abundance. Further, we calculate the stoichiometric value at which 50 % of the RNA will be bound in any complex to be ~51 nM at 10 % v/v RRL. This ten-fold difference suggests that the protein causing complex A is limiting further complex assembly.

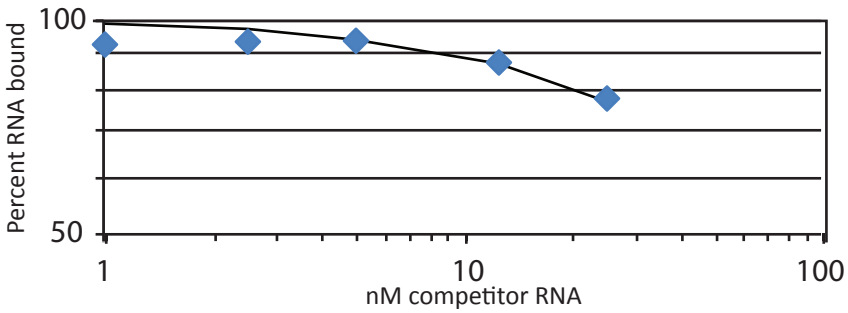
In addition, we performed competition experiments to determine the K_I of protein binding (Fig. 4-10A). The data were plotted as a fraction of RNA bound versus concentration of competitor RNA (nM) and it was described by the equation "$Activity=Activity_0/(1+[competitor]/K_{1/2})$" (Fig. 4-10B). For rbz-Dsim-7 the K_I is determined to be ~85 nM at 10 % v/v of RRL when the competitor RNA is also rbz-Dsim-7. The K_I for protein binding by rbz-Dsim-7 gives us further insight to the stoichiometry of binding (50-100 nM RNA) in 10 % v/v RRL and further supports that complex A is limiting subsequent complex assembly.

In order to enrich for the protein(s) leading to early binding complexes A and B, we tested binding of rbz-Dsim-7 against fractions of RRL generated by size exclusion. Fractionation of RRL was performed using Amicon size exclusion microcentrifuge filters.

A


1
2
3
4
5
6
NR
2.5
5
12.5
25
C
nM competitor RNA

B


100
50
Percent RNA bound
1
10
100
nM competitor RNA

The resulting fractions were then used to run natEMSAs with rbz-Dsim-7. The results show that the protein(s) resulting in early complexes A and B can be found in fractions corresponding to the molecular weight range of 50-100 kDa (not shown). This information is useful for reducing the complexity of the RRL and thereby reducing background in protein capture assays.

Data from natEMSAs demonstrate that HDV-like ribozymes are specifically recognized by a protein in RRL and that the interaction with that protein is limiting assembly of the translation complex assembly. The integrity of complex A suggests that the interaction with the ribozymes is strong and that the RNA may be protecting the protein from proteinase K digestion. The interaction with proteins forming complexes A does not seem to be sensitive to the thermostability of the ribozyme, based on data from Dsim-Δleader-7 and the P2 mutation constructs. Crosslinking and EDTA experiments suggest the protein forming complex B may be sensing the ribozyme thermostability, however we do not yet have kinetic or activity data to verify this claim. Future work will focus on the generation of chimeric constructs designed to further investigate the ribozyme structural features essential to protein recognition and further understand the kinetics of protein binding which may be responsible for the variation in Fluc activity levels demonstrated by different ribozymes during *in vitro* translation experiments. The integrity of the interaction leading to complex A motivated the use of protein capture assays in order to identify the protein recognizing HDV-like ribozymes.

Protein capture by RNA affinity chromatography

Both transcriptional and translational controls involve RNA-protein interactions. The regulatory proteins involved in these controls can be isolated via their affinity for the

nucleic acid targets. A number of strategies exist for isolating RNA binding proteins and identifying them via mass spectrometry (169). The key concept is to immobilize RNA as bait on a solid support, and the RNA binding proteins can be captured under native conditions. Protein complexes can be eluted by salt gradients, competitively eluted when certain tags are employed, or digested in-column for MS analysis. One major drawback of RNA pull-down assays is the high abundance of RNA binding proteins which cause high background binding (170). Various approaches can be used to circumvent non-specific binding and we employ some of these in our protocol.

For our RNA affinity chromatography, we relied on biotin tags to immobilize the RNA of interest on an agarose matrix. The RNA constructs included full-Dsim-WT-45 and rbz-Dsim-WT-45. A commercial mix of tRNA was used as a control. The RNA was 3′ end labeled with biotin via ligation with biotin-pCp or biotin was incorporated during *in vitro* transcription by T7 RNA polymerase. The RRL was fractionated to enrich for early binding protein(s) (greater than 50 kDa MWCO) and binding conditions were similar to those used during *in vitro* translation experiments. In addition, the lysate was pre-cleared for non-specific affinity to the resin used as a matrix. The agarose matrix was blocked at unbound sites by salmon sperm DNA. After allowing the fractionated, pre-cleared RRL to incubate on the column, the flow-through and all subsequent washes and elutions were collected and run on SDS-PAGE with proteins visualized by Coomassie staining. The streptavidin-agarose matrix was used a negative control. Two fractions from all experiments were further processed for MS analysis: the first partially denaturing wash (PD1) and the elution (E) that corresponds to column incubation in elevated pH to hydrolyze the RNA and release any bound proteins. Proteins in these

fractions were precipitated with 80 % acetone, resuspended in 20 mM ammonium bicarbonate, desalted, and then digested for peptide mass determination. Proteins were digested for MALDI-TOF following the protocol provided with Pierce MS-grade trypsin. The resulting peptide mixtures were desalted and concentrated using C18 ZipTips. MALDI was performed using AB SCIEX TOF/TOF 5800 System and protein determination was done using Access Matrix Mascot Server. The complete list of MALDI-TOF results is given in the Appendix.

Previous data show that the ribozyme-protein interaction is resistant to denaturing conditions and protein digestions. After proteins were allowed to bind the RNA on the column, the column was washed until no proteins are detected via Coomassie stained SDS-PAGE gels. In an effort to enrich for the initial binding protein(s) leading to complexes A and B, we utilized 3 M partially denaturing washes. We suggest that the proteins that were eluted by the partially denaturing washes correspond to those from a large complex which was recruited by the ribozyme recognition protein(s). The final elution was by digesting the RNA off the beads with either NaOH or NH4OH. Fig. 4-11 is an example of the fractions collected from a protein capture experiment for rbz-Dsim-45 run on an SDS-PAGE gel. This gel demonstrates a bump in proteins eluted off the column during PD1 and again by the elution.

The control experiments reveal that a number of non-specific proteins are consistently captured by our matrix materials. These proteins include: 39S ribosomal protein L12, serotransferrin, DNA-directed RNA polymerase III subunit RPC5, and methyl-CpG-binding domain protein 3-like 1. Some of the interesting hits specific to the PD1 fractions when rbz-Dsim-45 used as bait include: H/ACA ribonucleoprotein

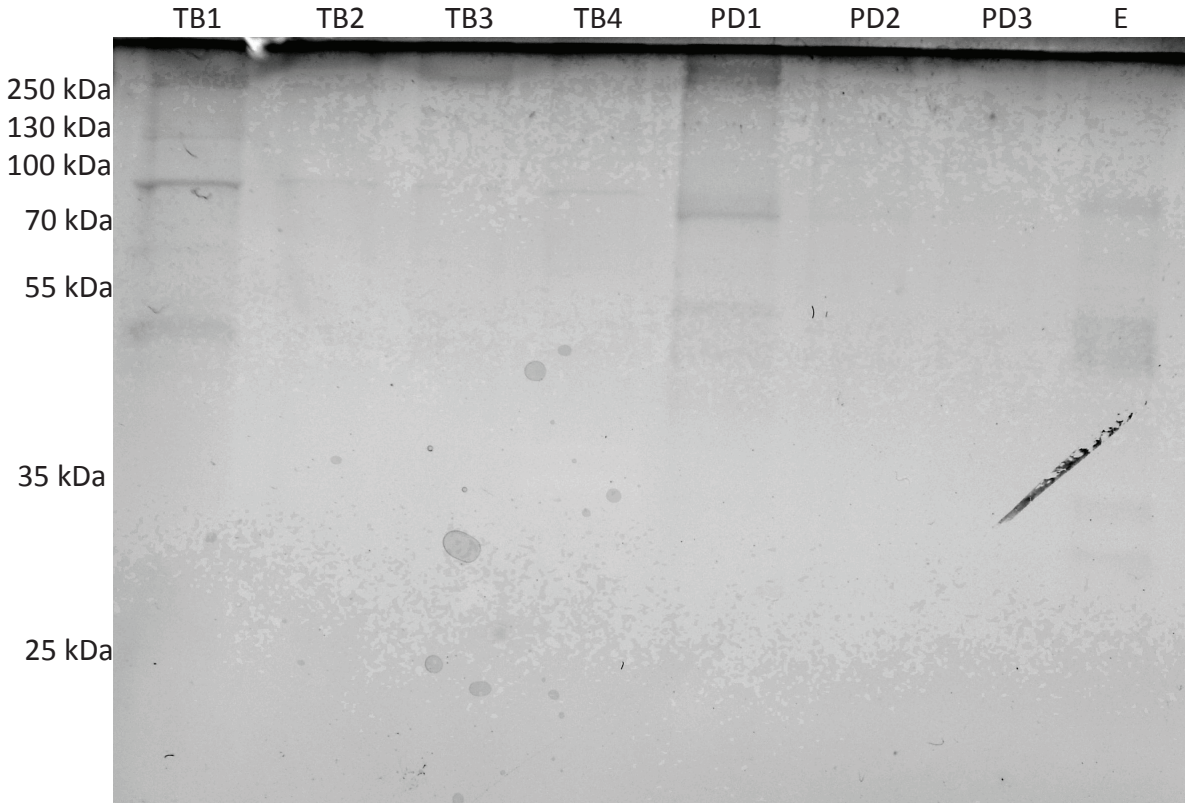


Figure 4-11. SDS-PAGE of fractions collected from RNA affinity chromatography experiments. The biotinylated rbz-Dsim-45 construct was used as bait on a streptavidin-agarose matrix and fractionated, pre-cleared RRL was used at 100 %. Non-specific binding proteins were washed away during the translation buffer (TB) washes. Some proteins were eluted by partially denaturing (PD) washes, with a majority eluted with the first wash. The final elution was completed by hydroxide treatment of the column to destroy the RNA thereby releasing specific binding proteins.

complex subunit 1, nuclear transition protein 2, ribonuclease P protein subunit p29, elongation factor 1-alpha1, and transcription factor AP-1 to name a few. All these interesting hits are proteins that are known to interact with or act on RNA. Some of the hits specific to the E fraction include: nuclear transition protein 2, guanine nucleotide-binding protein G(k) subunit alpha, high mobility group nucleosome-binding domain-containing protein 5, 28S ribosomal protein S10, ribonuclease P protein subunit p29, and eukaryotic translation initiation factor 2 subunit 2 to list a few. In addition, the elution fraction was run on a 12 % SDS-PAGE and the three predominant bands were excised and subjected to in-gel trypsin digestion prior to MALDI-TOF. The MS results verify the previously observed results from the raw elution fraction: elongation factor-1 and nuclear transition factor 2. Nuclear transition factor 2 is found in the nucleus of sperm and can be attributed to the use of salmon sperm DNA to block sites on the matrix that do not harbor the bait RNA. Elongation factor-1 (eEF1A) is responsible for the GTP-dependent binding of aminoacyl-tRNA to the A-site of ribosomes during protein biosynthesis. It is known to directly interact with RNA and has recently been shown to play a role in RNA virus genome replication (171). eEF1A has a mass around 50 kDa which supports natEMSA data demonstrating that the earliest proteins binding rbz-Dsim-7 are between 50-100 kDa. However, eEF1A is considered a high abundance protein which is perhaps how it has been adapted by viral replication (171). Interestingly, eEF1A has been shown to directly interact with structured RNA in viral UTRs but also to a region (not the delta ribozymes) of the HDV RNA genome (172). Future work should verify the interaction of eEF1A with rbz-Dsim by the use of

antibodies in natEMSA assays. The binding of the antibody would further retard the migration of a potential ribozyme-eEF1A complex.

Another protein that was identified by multiple independent assays includes ribonuclease P p29 which was detected in both PD1 and E fractions. Ribonuclease P is a ribozyme that functions in the maturation of tRNA (173) however, it has also been implicated in transcription of other non-coding RNA (reviewed in (174)). The protein subunit p29 (RRP29) can also be found as part of other ribonucleoprotein complexes localized in the nucleus. More work is needed to first verify that RPP29 is recognizing HDV-like ribozymes and then to elucidate a role for this complex assembly in promoting translation.

To provide further validation of MALDI-TOF data collected from RNA affinity chromatography experiments, we excised complexes directly from natEMSA 7.5 % native polyacrylamide gels. In particular, we performed in-gel trypsin digest of complex B. The most interesting top hit was ribosome biogenesis protein TSR3 and the telomerase holoenzyme component WD repeat-containing protein 53, which are predicted to be functional partners identified in the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database for known and predicted protein-protein interactions (175). Another predicted functional partner is radical S-adenosyl methionine (SAM) domain-containing protein 2 which was also identified through the in-gel trypsin digest of proteins isolated from the E fraction for rbz-Dsim-45. This finding gives some validation to the plethora of MALDI-TOF data where functionally interacting proteins are being isolated. This data suggest that the protein(s) forming complex B in natEMSAs is being eluted from the RNA affinity chromatography experiment.

MS data from two different sample preparations identified Gar1, a protein belonging to the family of DEAD-box containing proteins. DEAD-box proteins exist in both bacteria and eukaryotes and play roles in RNA metabolism (reviewed in (176)). In addition they have been shown to have RNA helicase activity by enhancing the susceptibility of nucleic acids to nucleases (177). Also, some evidence suggests DEAD-box proteins function to rearrange RNP complexes in addition to modifying RNA structures (178). Some of these biochemical activities contribute towards such biological functions as pre-mRNA processing, ribosome biogenesis, and translation. While the identification of a DEAD-box protein interacting with a ribozyme-protein complex is fascinating, more work is needed to understand the role of this protein in ribozyme-promoted translation. Future work will focus on identifying more proteins in the ribozyme-protein complex and most importantly, the protein recognizing HDV-like ribozymes. Identification of these proteins is critical to defining the mechanism of ribozyme-promoted translation.

Towards this end, we have prepared samples for LC-MS/MS in collaboration with Paul Gershon, Ph.D. at UCI. We expect for proteins identified by MALDI-TOF MS to be present, but more important we expect other proteins to be identified with high confidence. So far, the only results verify the existence of protein contaminants from the environment (keratin, trypsin, streptavidin, etc.) or non-specific adherence to the streptavidin-agarose matrix (serotransferrin, hemoglobin, etc.) The continuation of RNA affinity assays and MS will identify proteins leading to a translation initiation complex, the assembly of which is promoted by HDV-like ribozymes. Protein identification is critical in defining the mechanism of cap-independent translation promoted by ribozymes.

Discussion and significance

The work presented here sheds light on the mechanism of translation initiation by ribozyme-terminated mRNAs. Our results indicate that the correct formation of the ribozyme core is important in promoting translation. Translation occurs in the absence of a cap, start codon, and 5′ UTR or poly-A tail. In addition, translation occurs regardless of the catalytic activity of the ribozyme which suggests that the structure of the ribozyme core, not necessarily the product of self-cleavage is significant in promoting translation. Taken together, our data suggest that this translation is distinct from other cap-independent mechanisms. Unlike the HCV IRES, ribosomes are not being directly recruited by the HDV-like ribozymes. Also, there is no need for an AUG start codon which suggests ribosomal scanning is not being used to initiate translation. The accumulation of *in vitro* translation data leads to a model where HDV-like ribozymes promote non-canonical translation in a eukaryotic system without the need for 5′-m7G cap, 3′ UTR, or start codon.

Protein binding assays demonstrate that there is a protein present in RRL that specifically recognizes the ribozyme structure. Our data suggest that this protein-RNA interaction is strong and that perhaps the protein remains bound to the ribozyme during active translation. This observation suggests that the recognition protein is responsible for recruiting translation machinery; however, we have not yet verified the identity of this protein. Therefore, we have only just begun to elucidate the pathway of protein recruitment and assembly leading to translation promoted by HDV-like ribozymes. The continued use of RNA pull-down assays and MS will be essential to defining this mode of cap-independent translation.

Both the speed of ribozyme catalysis and the relative abundance of a specific recognition factor are sources of regulation that might be at play *in vivo*. Our work demonstrates that an additional source of regulation may come from the overall thermostability of the HDV-like ribozyme nested double-pseudoknot structure. The thermostability governs both the rates of catalysis and the efficiency of the ribozyme to promote translation of the downstream ORF. There is an inverse relationship between the thermostability of the ribozyme and the efficiency of translation. The potentially dynamic regulation of translation promoted by HDV-like ribozymes is suggested to rely on both the implicit ribozyme thermostability and the availability of the protein recognizing the ribozyme. While the former would account for the variation in levels of processed RNA across tissues or life stages (28,29), the latter elicits a signaling cascade we have just begun to understand. This concept illuminates the differences in biological function associated to ribozymes in different genomes, where in some cases faster catalysis is beneficial to the system than the ability to promote translation. Perhaps there is a trend to the evolutionary tree of HDV-like ribozymes that we have not previously realized.

Conclusions and future directions

Our data suggest that HDV-like ribozymes are defining the register of translation but that the start site of translation is not necessarily encoded within the ribozyme structure. Considerable effort was put into developing a tagged version of the Firefly luciferase enzyme generated by ribozyme-promoted translation in RRL. The protein was unable to be purified by IMAC when a His-tag was appended to the C-terminus of the protein. The construct suffered from poor protein yield because the transcript unnecessarily lacked a

3′ UTR. But overall, the His-tagged protein did not demonstrate good affinity to the metal chromatography columns employed for purification. In the elution fractions, the protein could not be detected by Coomassie stained SDS-PAGE. MALDI-TOF of the resulting elution fractions detected only fragments of the protein and we were unable to map the N-terminus. We redesigned the construct to contain a 3xFLAG-tag on the C-terminus of Fluc, this construct also lacking a 3′ UTR. The resulting tagged protein demonstrated good affinity to the anti-FLAG column used for purification. LC-MS/MS of the resulting elutions was performed by Dr. Paul Gershon in the Department of Molecular Biology and Biochemistry at the University of California, Irvine. The N-terminus of the protein was not detected and therefore we were unable to assign the start position of translation. Future work should be focused on scaling up the expression of a 3xFLAG-tagged Fluc generated by HDV-like ribozyme translation. The yield of protein can be improved by the incorporation of a 3′ UTR, a ribozyme other than Dsim, an inactive version of the ribozyme, and by scaling up the *in vitro* translation. Further, the RRL system used for *in vitro* translation has been used to express proteins for purification (179), however, we experienced difficulty eliminating hemoglobin completely from our protein preparations. We consider ribozyme-promoted translation to express low levels of proteins and therefore, contamination by high abundance proteins will continue to dampen efforts to purify proteins from RRL. The best expression system would be an *in vivo* expression from yeast that would allow for larger scale protein preparations. Mapping the N-terminus of the protein would elucidate the mechanism of translation by ribozyme-promoted translation. Together with the identification of the

protein specifically recognizing HDV-like ribozyme structure, we could define a novel mode of cap-independent translation.

The work presented here characterized the mechanism of ribozyme-promoted translation and paved the way for the discovery of a eukaryotic protein specifically recognizing the HDV-like ribozyme structure. We have discovered the inverse correlation between ribozyme thermostability and translation efficiency which would have implications in tracing ribozyme evolutionary lineages. The foundation presented here for a model of HDV-like ribozymes in promoting translation brings to light another layer of complexity in genetic controls of eukaryotic biological systems. The bimodal biological function of HDV-like ribozymes brings significance to the apparent random localization of this family of ribozymes across different genetic elements: annotated and unannotated, coding and non-coding. Perhaps the “random” distribution deserves deeper investigation.

CHAPTER 5

Materials and methods

In vitro RNA transcription

RNA was transcribed at 37˚C for one hour in a 20 µL volume containing 10 mM DTT; 20 mM spermidine; 2.5 mM each GTP, UTP and CTP; 250 µM ATP; 4.5 µCi [α-^{32}P]-ATP (Perkin Elmer, Waltham, MA); 7.75 mM $MgCl_2$; 20 µM of inhibitor oligonucleotide (specific for each construct); 1 unit of T7 RNA polymerase, and 0.5 pmole of DNA template. Constructs were transcribed *in vitro* in the presence of limited Mg^{2+} and 20 µM of cleavage-site inhibitor oligonucleotide to prevent co-transcriptional self-scission. Transcripts were purified by denaturing PAGE.

In-vitro cleavage kinetics

In vitro self-scission reactions were initiated with the addition of Mg^{2+} (to a final concentration of 1 mM or 10 mM) to solutions containing the ^{32}P-labeled ribozyme precursor in 140 mM KCl, 10 mM NaCl, 10 mM Tris buffer, pH 7.4 (final concentrations), at 37 ˚C and terminated by adding equal volume of stop buffer containing 20 mM EDTA, 5 mM Tris pH 7.4, 8 M urea, with xylene cyanol and bromophenol blue loading dyes. The denaturing PAGE gel of self-cleavage products was exposed to phosphorimage screens and analyzed using Typhoon phosphorimager and ImageQuant software (GE Healthcare).

In vitro cotranscriptional cleavage kinetics

In vitro transcription is done similar to above with the following modifications: 1) without the addition of inhibitor oligonucleotide; 2) 5 mM $MgCl_2$; 3) 1mM of each GTP, UTP, CTP; and 4) HEPES pH 7.4. The 10 µL reaction was initiated by the addition of DNA and incubated at 24 ˚C for 10 min. After that time, a 0.5 µL aliquot of the reaction was terminated by the addition of urea loading buffer. The remaining 4.5 µL volume was dilute 20-fold into the following buffer: 50 mM HEPES pH 7.4, 10 mM NaCl, 140 mM KCl, and the desired concentration of $MgCl_2$. For conditions requiring consistent ionic strength, the buffer and metabolite stocks were pH adjusted by the addition of KOH and the contribution of K^+ from the stocks was tracked. The concentration of K^+ was adjusted by the addition of KCl for a final reaction concentration of 140 mM. Timepoints were taken at the indicated times following the dilution of the transcript into the new buffer at 37 ˚C and the 5 µL aliquots were terminated by adding 5 µL volume of stop buffer containing 20 mM EDTA, 5 mM Tris pH 7.4, 8 M urea, with xylene cyanol and bromophenol blue loading dyes. The

denaturing PAGE gel of self-cleavage products was exposed to phosphorimage screens and analyzed using Typhoon phosphorimager and ImageQuant software (GE Healthcare).

RNA 3′-terminus labeling

RNA was *in vitro* transcribed in the absence of [α-^{32}P]-ATP and PAGE purified. The appropriate RNA species was excised, precipitated, and resuspended in water. RNA was then ligated at 37°C for 3 hours in a volume of 10 µL, containing RNA ligase buffer (New England Biolabs), 2 µCi [5′-^{32}P] cytidine 3′, 5′-bisphosphate (Perkin Elmer) and one unit of T4 RNA ligase (NEB) and PAGE purified again.

In-line probing

The 3′-end labeled RNA was incubated with varying amounts of ligand for up to 2 days at 37°C in a buffer containing 140 mM KCl, 10 mM NaCl, 20 mM Tris chloride, pH 7.9, 1 mM $MgCl_2$, and 1 mM spermidine. The partially hydrolyzed RNAs were resolved using denaturing PAGE, exposed to phosphorimage screens (Molecular Dynamics/GE Healthcare, Pittsburgh, PA, USA), and scanned by GE Typhoon phosphorimager. The sequences in the degradation pattern were confirmed by running α-phosphorothioate nucleotide incorporated RNA cleaved by treatment with iodoethanol.

Cloning ribozyme bicistronic reporter constructs

The vector used for bicistronic constructs contains a *Renilla luciferase* gene upstream of a *Firefly luciferase* in a pCR2.1-TOPO plasmid (ref. DJR). The vector was linearized at AvrII and XhoI sites at the intergenic region. The appropriate constructs were generated by PCR and digested by XhoI and BsaI to produce sticky ends. The insert was ligated to the vector using T4 DNA ligase. The resulting plasmid was transformed into competent cells and the resulting E.coli colonies were sequenced to check the insertion and luciferase sequences.

***In vivo* expression analysis**

E.coli cultures are grown overnight in LB media. The next day cells are pelleted, washed and resuspended in minimal media and incubated overnight at 37˚C with agitation. The following day the OD_{600} measurements are taken. Cultures are normalized to the same OD_{600} of 0.4 into minimal media (1x M9 salts, 2 mM $MgSO_4$, 0.5 % carbon source, 0.1 mM $CaCl_2$) and incubated for 1hr at 37˚C with agitation. Cultures are normalized to the same OD_{600} of 0.2 into minimal media and incubated at 37˚C with agitation. Timepoints denote the length of time the new cultures are incubated at 37˚C following

normalization to OD_{600} of 0.2. At the indicated timepoints, 1 mL aliquots of culture are pelleted, lysed using chicken egg white lysozyme (Sigma) resuspended in 1X PBS, and immediately stored at -80˚C. To measure the levels of Fluc and Rluc, two 10 µL aliquots of the resulting cell lysates are plated in a black bottom 96-well plate. Each aliquot receives 50 µL of either Luciferase Activity Reagent (Promega) containing luciferin to measure Fluc levels, or 55 µM coelenterazine (Promega) to measure Rluc levels.

Site-directed mutagenesis

To generate ribozymes with a mutation, the QuikChange mutagenesis kit (Stratagene) was used with oligonucleotides containing the desired mutation. All constructs were sequenced to check the ribozyme and luciferase sequence. In some cases, multiple rounds of mutagenesis were necessary for independent mutation sites.

***In vitro* transcription for luciferase activity assays**

For *in vitro* and *in vivo* translation assays, nonradioactive RNA was transcribed at 37ºC for 3 hours in a 50 µL volume containing 10 mM of DTT, 2.5 mM each GTP, UTP, CTP, and ATP, 35 mM MgCl2, 40 units of RNasin® Plus RNase Inhibitor (Promega), 10 units of T7 RNA polymerase, and at least 1000 ng of restriction enzyme digested plasmid DNA. After treatment with DNase I (RNase-free, Promega, 1 unit/µg DNA for 15 min at 37ºC), RNA was purified by RNA Clean & Concentrate-25 columns (Zymo). Concentrations were determined by absorbance at 260 nm and adjusted based on densitometry measurements of full length RNA from a SYBR Gold (Invitrogen) stained 1% agarose gel.

***In vitro* translation**

Promega's standard nuclease-treated Rabbit Reticulocyte Lysate System was used for 10 µL *in vitro* translation reactions containing 50% (v/v) RRL, 13 units of RNasin Plus RNase Inhibitor (Promega), 1X Complete Protease Inhibitor Cocktail (Roche: one tablet in 1 mL used as 50X), 2 mM DTT, 20 µM amino acids, 1.8 mM $MgCl_2$, 45 mM KCl, and 26 mM KOAc, to obtain final salt concentrations of 2.2 mM Mg^{2+}, 45 mM KCl, and 90 mM KOAc that improve the fidelity of translation initiation. The reactions were initiated by addition of at least 10 ng *in vitro* transcribed RNA. Reactions were incubated at 30 ºC for 90 min, stopped on ice, and immediately examined for luciferase activity by mixing with 100 µL Luciferase Assay Reagent (Promega) in black 96-well plates (BD Falcon) and imaged using the IVIS® Lumina II system (Caliper Life Sciences). Translation efficiency was measured based on integrated density of

luminescence from each reaction and reported as the mean of at least two independent translation experiments.

Native electromobility shift assays (natEMSA)

Ribozyme constructs were in vitro transcribed as described previously and PAGE purified. Either the band corresponding to the full-length transcript or 3′ product of self-cleavage was excised, precipitated, and resuspended in water and quantified by absorbance at 260 nm. The RNA was then 3′ end-labeled as described previously and unincorporated [5′-^{32}P] cytidine 3′, 5′-bisphosphate was removed by a Sephadex G-25 column (Sigma). The appropriate concentration of end-labeled RNA was used in a binding reaction containing: 13 units of RNasin Plus RNase Inhibitor (Promega), 1X Complete Protease Inhibitor Cocktail (Roche: one tablet in 1 mL used as 50X), 2 mM DTT, 0.825 μM tRNA, 2.2 mM Mg^{2+}, 45 mM KCl, 90 mM KOAc, and the appropriate dilution of rabbit reticulocyte lysate (RRL-Promega). Reactions were incubated for 10 min at 30 ˚C then dilute 2-fold into the same buffer plus 3 % glycerol. Immediately, samples were loaded onto a polyacrylamide gel made with 1x THM buffer: 33 mM Tris chloride, 66 mM HEPES, pH 7.1, 2 mM $MgCl_2$, and 20 mM KCl. The running buffer was also 1x THM buffer pH 7.1 but without KCl. The gel was run for 2.5 hrs at 10 W, exposed to phosphorimage screens (Molecular Dynamics/GE Healthcare, Pittsburgh, PA, USA), and scanned by GE Typhoon phosphorimager.

***In vitro* transcription for biotinylated RNA**

In vitro transcription was performed as previously described except that UTP was dropped to 250 μM to promote incorporation of biotin-16-aminoallyluridine-5'- triphosphate (TriLink) which was added to a final reaction concentration of 2 mM. The transcription was initiated by the addition of 1 unit of T7 RNA polymerase and incubated for 2-4 hours at 37 ˚C. The reaction was terminated by the addition of equal volume 8 M urea loading buffer and the RNA resolved by 7.5 % denaturing PAGE. Only bands corresponding to the product of self-cleavage were excised to ensure the random incorporation of biotin did not interfere with ribozyme folding.

RNA affinity chromatography

Streptavidin-agarose beads (Invitrogen) are used as the support matrix for RNA pull-down assays. The beads are blocked by the addition of 1 mg/mL salmon sperm DNA (Invitrogen) for 10 min at 25 ˚C and then equilibrated into 1x TB buffer, pH 7.4: 2 mM DTT, 1.8 mM $MgCl_2$, 45 mM KCl, 26 mM KOAc,

and 0.8 μM tRNA. Biotinylated RNA is bound to the column for 1 hr at 25 ˚C with agitation. Rabbit reticulocyte lysate is fractionated and concentrated with a 50 kDa MWCO Amicon column. The lysate is resuspended back to 100 % with the addition of 1x TB. The fractionated lysate is then pre-cleared by incubation on streptavidin-agarose beads for 10 min at 25 ˚C. The resulting flow-through is applied to the RNA bound column and incubated for about 1 hr at 25 ˚C. The flow through and three washes with 1x TB are collected. Next, the column is washed three times with 1x PD buffer: 2 mM DTT, 1.8 mM $MgCl_2$, 45 mM KCl, 26 mM KOAc, and 3 M urea. Last, the proteins bound to the RNA are eluted off the column with 10 % NH_4OH. All fractions were resolved by 10 % SDS-PAGE stained by Coomassie blue R. For mass spectrometry, desired fractions are precipitated by 80 % acetone at -20 ˚C overnight. The next day the pellet resulting from centrifugation is air dried and resuspended in 20 mM NH_4HCO_3, pH 8.

Trypsin digestion

A 10 μL volume of each sample in 20 mM NH_4HCO_3, pH 8 is aliquoted into fresh PCR tubes following desalting by a Sephadex G-25 column. The proteins in the samples are reduced by the addition of 20 mM dithiothreitol (DTT) and incubated at 60 ˚C for 1 hr. The samples are then alkylated by the addition of 40 mM iodoacetamide and incubation for 30 min at room temperature. Trypsin is then added at a ratio of 1:20 w/w following the instructions by the manufacturer (Thermo Scientific Pierce Trypsin Protease, MS Grade) and incubated overnight at 37 ˚C. the resulting digest products are cleaned and concentrated using C18 ZipTips (Millipore) following the instructions from the manufacturer.

MALDI-TOF

The matrix used is α-cycano-4-hydroxycinnamic acid (~ 10 mg/mL) in a 1:1 acetonitrile: water mixture with 1 % trifluoroacetic acid (TFA). This mixture containing the matrix is used to elute of the C18 ZipTip onto a spot on the matrix plate. The MALDI spectra are collected on AB SCIEX TOF/TOF 5800 System. The resulting mass lists are analyzed by Access Mascot Server (Matrix Science).

REFERENCES

1. Prody, G. A., Bakos, J. T., Buzayan, J. M., Schneider, I. R. & Bruening, G. Autolytic processing of dimeric plant virus satellite RNA. Science 231, 1577-1580, (1986).

2. Forster, A. C. & Symons, R. H. Self-Cleavage of Virusoid Rna Is Performed by the Proposed 55-Nucleotide Active-Site. Cell 50, 9-16, (1987).

3. Forster, A. C. & Symons, R. H. Self-cleavage of plus and minus RNAs of a virusoid and a structural model for the active sites. Cell 49, 211-220, (1987).

4. Epstein, L. M. & Gall, J. G. Transcripts of newt satellite DNA self-cleave in vitro. Cold Spring Harb Symp Quant Biol 52, 261-265, (1987).

5. Rojas, A. A. et al. Hammerhead-mediated processing of satellite pDo500 family transcripts from Dolichopoda cave crickets. Nucleic Acids Res 28, 4037-4043, (2000).

6. Ferbeyre, G., Smith, J. M. & Cedergren, R. Schistosome satellite DNA encodes active hammerhead ribozymes. Mol Cell Biol 18, 3880-3888, (1998).

7. Laha, T., McManus, D. P., Loukas, A. & Brindley, P. J. Sjalpha elements, short interspersed element-like retroposons bearing a hammerhead ribozyme motif from the genome of the oriental blood fluke Schistosoma japonicum. Biochim Biophys Acta 1492, 477-482, (2000).

8. de la Pena, M. & Garcia-Robles, I. Ubiquitous presence of the hammerhead ribozyme motif along the tree of life. RNA 16, 1943-1950, (2010).

9. de la Pena, M. & Garcia-Robles, I. Intronic hammerhead ribozymes are ultraconserved in the human genome. EMBO Rep 11, 711-716, (2010).

10. Perreault, J. et al. Identification of Hammerhead Ribozymes in All Domains of Life Reveals Novel Structural Variations. Plos Comput Biol 7, (2011).

11. Jimenez, R. M., Delwart, E. & Luptak, A. Structure-based search reveals hammerhead ribozymes in the human microbiome. J Biol Chem 286, 7737-7743, (2011).

12. Seehafer, C., Kalweit, A., Steger, G., Graf, S. & Hammann, C. From alpaca to zebrafish: hammerhead ribozymes wherever you look. RNA 17, 21-26, (2011).

13. Garcia-Robles, I., Sanchez-Navarro, J. & de la Pena, M. Intronic hammerhead ribozymes in mRNA biogenesis. Biol Chem 393, 1317-1326, (2012).

14. Martick, M., Horan, L. H., Noller, H. F. & Scott, W. G. A discontinuous hammerhead ribozyme embedded in a mammalian messenger RNA. Nature 454, 899-U857, (2008).

15. Pley, H. W., Flaherty, K. M. & McKay, D. B. Three-dimensional structure of a hammerhead ribozyme. Nature 372, 68-74, (1994).

16. Scott, W. G., Murray, J. B., Arnold, J. R., Stoddard, B. L. & Klug, A. Capturing the structure of a catalytic RNA intermediate: the hammerhead ribozyme. Science 274, 2065-2069, (1996).

17. Khvorova, A., Lescoute, A., Westhof, E. & Jayasena, S. D. Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. Nat Struct Biol 10, 708-712, (2003).

18. De la Pena, M., Gago, S. & Flores, R. Peripheral regions of natural hammerhead ribozymes greatly increase their self-cleavage activity. EMBO J 22, 5561-5570, (2003).

19. Martick, M. & Scott, W. G. Tertiary contacts distant from the active site prime a ribozyme for catalysis. Cell 126, 309-320, (2006).

20. Ferre-D'Amare, A. R. & Scott, W. G. Small self-cleaving ribozymes. Cold Spring Harb Perspect Biol 2, a003574, (2010).

21. Anderson, M., Schultz, E. P., Martick, M. & Scott, W. G. Active-site monovalent cations revealed in a 1.55-A-resolution hammerhead ribozyme structure. J Mol Biol 425, 3790-3798, (2013).

22. Han, J. & Burke, J. M. Model for general acid-base catalysis by the hammerhead ribozyme: pH-activity relationships of G8 and G12 variants at the putative active site. Biochemistry 44, 7864-7870, (2005).

23. Roychowdhury-Saha, M. & Burke, D. H. Distinct reaction pathway promoted by non-divalent-metal cations in a tertiary stabilized hammerhead ribozyme. RNA 13, 841-848, (2007).

24. Chi, Y. I. et al. Capturing hammerhead ribozyme structures in action by modulating general base catalysis. Plos Biol 6, 2060-2068, (2008).

25. Lai, M. M. The molecular biology of hepatitis delta virus. Annu Rev Biochem 64, 259-286, (1995).

26. Taylor, J. M. Hepatitis delta virus. Intervirology 42, 173-178, (1999).

27. Wu, H. N. et al. Human Hepatitis-Delta Virus-Rna Subfragments Contain an Autocleavage Activity. P Natl Acad Sci USA 86, 1831-1835, (1989).

28. Salehi-Ashtiani, K., Luptak, A., Litovchick, A. & Szostak, J. W. A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. Science 313, 1788-1792, (2006).

29. Webb, C. H., Riccitelli, N. J., Ruminski, D. J. & Luptak, A. Widespread occurrence of self-cleaving ribozymes. Science 326, 953, (2009).

30. Bibillo, A. & Eickbush, T. H. The reverse transcriptase of the R2 non-LTR retrotransposon: continuous synthesis of cDNA on non-continuous RNA templates. J Mol Biol 316, 459-473, (2002).

31. Bibillo, A. & Eickbush, T. H. End-to-end template jumping by the reverse transcriptase encoded by the R2 retrotransposon. J Biol Chem 279, 14945-14953, (2004).

32. Ruminski, D. J., Webb, C. H., Riccitelli, N. J. & Luptak, A. Processing and translation initiation of non-long terminal repeat retrotransposons by hepatitis delta virus (HDV)-like self-cleaving ribozymes. J Biol Chem 286, 41286-41295, (2011).

33. Eickbush, D. G. & Eickbush, T. H. R2 retrotransposons encode a self-cleaving ribozyme for processing from an rRNA cotranscript. Mol Cell Biol 30, 3142-3150, (2010).

34. Sanchez-Luque, F. J., Lopez, M. C., Macias, F., Alonso, C. & Thomas, M. C. Identification of an hepatitis delta virus-like ribozyme at the mRNA 5 '-end of the L1Tc retrotransposon from Trypanosoma cruzi. Nucleic Acids Research 39, 8065-8077, (2011).

35. Vogler, C. et al. CPEB3 is associated with human episodic memory. Front Behav Neurosci 3, (2009).

36. Smith, J. B. & Dinter-Gottlieb, G. Antigenomic Hepatitis delta virus ribozymes self-cleave in 18 M formamide. Nucleic Acids Res 19, 1285-1289, (1991).

37. Duhamel, J. et al. Secondary structure content of the HDV ribozyme in 95% formamide. Nucleic Acids Res 24, 3911-3917, (1996).

38. Ferre-D'Amare, A. R., Zhou, K. & Doudna, J. A. Crystal structure of a hepatitis delta virus ribozyme. Nature 395, 567-574, (1998).

39. Ke, A., Zhou, K., Ding, F., Cate, J. H. & Doudna, J. A. A conformational switch controls hepatitis delta virus ribozyme catalysis. Nature 429, 201-205, (2004).

40. Chen, J. H. et al. A 1.9 A crystal structure of the HDV ribozyme precleavage suggests both Lewis acid and general acid mechanisms contribute to phosphodiester cleavage. Biochemistry 49, 6508-6518, (2010).

41. Gong, B., Chen, J. H., Bevilacqua, P. C., Golden, B. L. & Carey, P. R. Competition between Co(NH(3)(6)3+ and inner sphere Mg2+ ions in the HDV ribozyme. Biochemistry 48, 11961-11970, (2009).

42. Golden, B. L. Two distinct catalytic strategies in the hepatitis delta virus ribozyme cleavage reaction. Biochemistry 50, 9424-9433, (2011).

43. Thaplyal, P., Ganguly, A., Hammes-Schiffer, S. & Bevilacqua, P. C. Inverse Thio Effects in the Hepatitis Delta Virus Ribozyme Reveal that the Reaction Pathway Is Controlled by Metal Ion Charge Density. Biochemistry 54, 2160-2175, (2015).

44. Perrotta, A. T. & Been, M. D. HDV ribozyme activity in monovalent cations. Biochemistry 45, 11357-11365, (2006).

45. Luptak, A., Ferre-D'Amare, A. R., Zhou, K., Zilm, K. W. & Doudna, J. A. Direct pK(a) measurement of the active-site cytosine in a genomic hepatitis delta virus ribozyme. J Am Chem Soc 123, 8447-8452, (2001).

46. Gong, B. et al. Direct measurement of a pK(a) near neutrality for the catalytic cytosine in the genomic HDV ribozyme using Raman crystallography. J Am Chem Soc 129, 13335-13342, (2007).

47. Rubino, L., Tousignant, M. E., Steger, G. & Kaper, J. M. Nucleotide sequence and structural analysis of two satellite RNAs associated with chicory yellow mottle virus. J Gen Virol 71 (Pt 9), 1897-1903, (1990).

48. Kaper, J. M., Tousignant, M. E. & Steger, G. Nucleotide sequence predicts circularity and self-cleavage of 300-ribonucleotide satellite of arabis mosaic virus. Biochem Biophys Res Commun 154, 318-325, (1988).

49. Grum-Tokars, V., Milovanovic, M. & Wedekind, J. E. Crystallization and X-ray diffraction analysis of an all-RNA U39C mutant of the minimal hairpin ribozyme. Acta Crystallogr D Biol Crystallogr 59, 142-145, (2003).

50. Rupert, P. B. & Ferre-D'Amare, A. R. Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. Nature 410, 780-786, (2001).

51. Rupert, P. B., Massey, A. P., Sigurdsson, S. T. & Ferre-D'Amare, A. R. Transition state stabilization by a catalytic RNA. Science 298, 1421-1424, (2002).

52. Tan, E. et al. A four-way junction accelerates hairpin ribozyme folding via a discrete intermediate. Proc Natl Acad Sci U S A 100, 9308-9313, (2003).

53. Nahas, M. K. et al. Observation of internal cleavage and ligation reactions of a ribozyme. Nat Struct Mol Biol 11, 1107-1113, (2004).

54. Kath-Schorr, S. et al. General acid-base catalysis mediated by nucleobases in the hairpin ribozyme. J Am Chem Soc 134, 16717-16724, (2012).

55. Grasby, J. A., Mersmann, K., Singh, M. & Gait, M. J. Purine functional groups in essential residues of the hairpin ribozyme required for catalytic cleavage of RNA. Biochemistry 34, 4068-4076, (1995).

56. Kuzmin, Y. I., Da Costa, C. P. & Fedor, M. J. Role of an active site guanine in hairpin ribozyme catalysis probed by exogenous nucleobase rescue. J Mol Biol 340, 233-251, (2004).

57. Kuzmin, Y. I., Da Costa, C. P., Cottrell, J. W. & Fedor, M. J. Role of an active site adenine in hairpin ribozyme catalysis. J Mol Biol 349, 989-1010, (2005).

58. Wilson, T. J. et al. Nucleobase catalysis in the hairpin ribozyme. RNA 12, 980-987, (2006).

59. Pinard, R. et al. Functional involvement of G8 in the hairpin ribozyme cleavage mechanism. EMBO J 20, 6434-6442, (2001).

60. Collins, R. A. The Neurospora Varkud satellite ribozyme. Biochem Soc Trans 30, 1122-1126, (2002).

61. Bouchard, P. & Legault, P. A remarkably stable kissing-loop interaction defines substrate recognition by the Neurospora Varkud Satellite ribozyme. RNA 20, 1451-1464, (2014).

62. Wilson, T. J., McLeod, A. C. & Lilley, D. M. A guanine nucleobase important for catalysis by the VS ribozyme. EMBO J 26, 2489-2500, (2007).

63. Wilson, T. J. et al. Nucleobase-mediated general acid-base catalysis in the Varkud satellite ribozyme. Proc Natl Acad Sci U S A 107, 11751-11756, (2010).

64. Beattie, T. L., Olive, J. E. & Collins, R. A. A secondary-structure model for the self-cleaving region of Neurospora VS RNA. Proc Natl Acad Sci U S A 92, 4686-4690, (1995).

65. Rastogi, T., Beattie, T. L., Olive, J. E. & Collins, R. A. A long-range pseudoknot is required for activity of the Neurospora VS ribozyme. Embo Journal 15, 2820-2825, (1996).

66. Andersen, A. A. & Collins, R. A. Rearrangement of a stable RNA secondary structure during VS ribozyme catalysis. Mol Cell 5, 469-478, (2000).

67. Andersen, A. A. & Collins, R. A. Intramolecular secondary structure rearrangement by the kissing interaction of the Neurospora VS ribozyme. Proc Natl Acad Sci U S A 98, 7730-7735, (2001).

68. Flinders, J. & Dieckmann, T. A pH controlled conformational switch in the cleavage site of the VS ribozyme substrate RNA. J Mol Biol 308, 665-679, (2001).

69. Hiley, S. L. & Collins, R. A. Rapid formation of a solvent-inaccessible core in the Neurospora Varkud satellite ribozyme. EMBO J 20, 5461-5469, (2001).

70. Hiley, S. L., Sood, V. D., Fan, J. & Collins, R. A. 4-thio-U cross-linking identifies the active site of the VS ribozyme. EMBO J 21, 4691-4698, (2002).

71. Bouchard, P. & Legault, P. Structural insights into substrate recognition by the Neurospora Varkud satellite ribozyme: importance of U-turns at the kissing-loop junction. Biochemistry 53, 258-269, (2014).

72. Michiels, P. J. A., Schouten, C. H. J., Hilbers, C. W. & Heus, H. A. Structure of the ribozyme substrate hairpin of Neurospora VS RNA: A close look at the cleavage site. Rna-a Publication of the Rna Society 6, 1821-1832, (2000).

73. Hoffmann, B. et al. NMR structure of the active conformation of the Varkud satellite ribozyme cleavage site. Proc Natl Acad Sci U S A 100, 7003-7008, (2003).

74. Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A. & Breaker, R. R. Control of gene expression by a natural metabolite-responsive ribozyme. Nature 428, 281-286, (2004).

75. McCarthy, T. J. et al. Ligand requirements for glmS ribozyme self-cleavage. Chem Biol 12, 1221-1226, (2005).

76. Roth, A., Nahvi, A., Lee, M., Jona, I. & Breaker, R. R. Characteristics of the glmS ribozyme suggest only structural roles for divalent metal ions. RNA 12, 607-619, (2006).

77. Klein, D. J. & Ferre-D'Amare, A. R. Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. Science 313, 1752-1756, (2006).

78. Cochrane, J. C., Lipchock, S. V. & Strobel, S. A. Structural investigation of the GlmS ribozyme bound to Its catalytic cofactor. Chem Biol 14, 97-105, (2007).

79. Cochrane, J. C., Lipchock, S. V., Smith, K. D. & Strobel, S. A. Structural and chemical basis for glucosamine 6-phosphate binding and activation of the glmS ribozyme. Biochemistry 48, 3239-3246, (2009).

80. Gong, B., Klein, D. J., Ferre-D'Amare, A. R. & Carey, P. R. The glmS Ribozyme Tunes the Catalytically Critical pK(a) of Its Coenzyme Glucosamine-6-phosphate. J Am Chem Soc 133, 14188-14191, (2011).

81. Viladoms, J. & Fedor, M. J. The glmS Ribozyme Cofactor is a General Acid-Base Catalyst. J Am Chem Soc 134, 19043-19049, (2012).

82. Roth, A. et al. A widespread self-cleaving ribozyme class is revealed by bioinformatics. Nature Chemical Biology 10, 56-U92, (2014).

83. Liu, Y., Wilson, T. J., McPhee, S. A. & Lilley, D. M. Crystal structure and mechanistic investigation of the twister ribozyme. Nat Chem Biol 10, 739-744, (2014).

84. Ren, A. M. et al. In-line alignment and Mg2+ coordination at the cleavage site of the env22 twister ribozyme. Nat Commun 5, (2014).

85. Eiler, D., Wang, J. & Steitz, T. A. Structural basis for the fast self-cleavage reaction catalyzed by the twister ribozyme. Proc Natl Acad Sci U S A 111, 13028-13033, (2014).

86. Fong, N., Ohman, M. & Bentley, D. L. Fast ribozyme cleavage releases transcripts from RNA polymerase II and aborts co-transcriptional pre-mRNA processing. Nat Struct Mol Biol 16, 916-922, (2009).

87. Lacadie, S. A., Tardiff, D. F., Kadener, S. & Rosbash, M. In vivo commitment to yeast cotranscriptional splicing is sensitive to transcription elongation mutants. Genes Dev 20, 2055-2066, (2006).

88. Cervera, A. & De la Pena, M. Eukaryotic penelope-like retroelements encode hammerhead ribozyme motifs. Mol Biol Evol 31, 2941-2947, (2014).

89. Forster, A. C., Davies, C., Sheldon, C. C., Jeffries, A. C. & Symons, R. H. Self-Cleaving Viroid and Newt Rnas May Only Be Active as Dimers. Nature 334, 265-267, (1988).

90. Hammann, C., Luptak, A., Perreault, J. & de la Pena, M. The ubiquitous hammerhead ribozyme. RNA 18, 871-885, (2012).

91. Tang, J. & Breaker, R. R. Structural diversity of self-cleaving ribozymes. Proc Natl Acad Sci U S A 97, 5784-5789, (2000).

92. Salehi-Ashtiani, K. & Szostak, J. W. In vitro evolution suggests multiple origins for the hammerhead ribozyme. Nature 414, 82-84, (2001).

93. Loh, E. et al. A trans-acting riboswitch controls expression of the virulence regulator PrfA in Listeria monocytogenes. Cell 139, 770-779, (2009).

94. Vinkenborg, J. L., Karnowski, N. & Famulok, M. Aptamers for allosteric regulation. Nat Chem Biol 7, 519-527, (2011).

95. Famulok, M., Hartig, J. S. & Mayer, G. Functional aptamers and aptazymes in biotechnology, diagnostics, and therapy. Chem Rev 107, 3715-3743, (2007).

96. Tang, J. & Breaker, R. R. Rational design of allosteric ribozymes. Chem Biol 4, 453-459, (1997).

97. Carothers, J. M., Goler, J. A., Juminaga, D. & Keasling, J. D. Model-driven engineering of RNA devices to quantitatively program gene expression. Science 334, 1716-1719, (2011).

98. Gesteland, R. F., Cech, T. & Atkins, J. F. The RNA world : the nature of modern RNA suggests a prebiotic RNA world. 3rd edn, (Cold Spring Harbor Laboratory Press, 2006).

99. Scott, W. G., Martick, M. & Chi, Y. I. Structure and function of regulatory RNA elements: Ribozymes that regulate gene expression. Bba-Gene Regul Mech 1789, 634-641, (2009).

100. Fedor, M. J. & Williamson, J. R. The catalytic diversity of RNAs. Nat Rev Mol Cell Biol 6, 399-412, (2005).

101. Nielsen, H., Westhof, E. & Johansen, S. An mRNA is capped by a 2 ',5 ' lariat catalyzed by a group I-like ribozyme. Science 309, 1584-1587, (2005).

102. Cech, T. R. Ribozymes, the first 20 years. Biochem Soc Trans 30, 1162-1166, (2002).

103. Bourdeau, V., Ferbeyre, G., Pageau, M., Paquin, B. & Cedergren, R. The distribution of RNA motifs in natural sequences. Nucleic Acids Res 27, 4457-4467, (1999).

104. Riccitelli, N. J. & Luptak, A. Computational discovery of folded RNA domains in genomes and in vitro selected libraries. Methods 52, 133-140, (2010).

105. Ferbeyre, G., Bourdeau, V., Pageau, M., Miramontes, P. & Cedergren, R. Distribution of hammerhead and hammerhead-like RNA motifs through the GenBank. Genome Res 10, 1011-1019, (2000).

106. Seehafer, C., Kalweit, A., Steger, G., Graf, S. & Hammann, C. From alpaca to zebrafish: Hammerhead ribozymes wherever you look. RNA 17, 21-26, (2010).

107. Canny, M. D., Jucker, F. M. & Pardi, A. Efficient ligation of the Schistosoma hammerhead ribozyme. Biochemistry 46, 3826-3834, (2007).

108. Blinkova, O. et al. Frequent detection of highly diverse variants of cardiovirus, cosavirus, bocavirus, and circovirus in sewage samples collected in the United States. J Clin Microbiol 47, 3507-3513, (2009).

109. Victoria, J. G. et al. Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. J Virol 83, 4642-4651, (2009).

110. Victoria, J. G. et al. Viral nucleic acids in live-attenuated vaccines: detection of minority variants and an adventitious virus. J Virol 84, 6033-6040, (2010).

111. Li, L. et al. Bat guano virome: predominance of dietary viruses from insects and plants plus novel mammalian viruses. J Virol 84, 6955-6965, (2010).

112. Hofacker, I. L. Vienna RNA secondary structure server. Nucleic Acids Res 31, 3429-3431, (2003).

113. Canny, M. D. et al. Fast cleavage kinetics of a natural hammerhead ribozyme. J Am Chem Soc 126, 10848-10849, (2004).

114. Nelson, J. A. & Uhlenbeck, O. C. When to believe what you see. Molecular Cell 23, 447-450, (2006).

115. Roychowdhury-Saha, M. & Burke, D. H. Extraordinary rates of transition metal ion-mediated ribozyme catalysis. RNA 12, 1846-1852, (2006).

116. Dufour, D., de la Pena, M., Gago, S., Flores, R. & Gallego, J. Structure-function analysis of the ribozymes of chrysanthemum chlorotic mottle viroid: a loop-loop interaction motif conserved in most natural hammerheads. Nucleic Acids Res 37, 368-381, (2009).

117. Sperschneider, J. & Datta, A. DotKnot: pseudoknot prediction using the probability dot plot under a refined energy model. Nucleic Acids Research 38, e103, (2010).

118. Kore, A. R., Vaish, N. K., Kutzke, U. & Eckstein, F. Sequence specificity of the hammerhead ribozyme revisited; the NHH rule. Nucleic Acids Res 26, 4116-4120, (1998).

119. Shepotinovskaya, I. V. & Uhlenbeck, O. C. Catalytic diversity of extended hammerhead ribozymes. Biochemistry 47, 7034-7042, (2008).

120. De la Pena, M. & Flores, R. An extra nucleotide in the consensus catalytic core of a viroid hammerhead ribozyme - Implications for the design of more efficient ribozymes. J Biol Chem 276, 34586-34593, (2001).

121. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25, 3389-3402, (1997).

122. Menger, M., Eckstein, F. & Porschke, D. Multiple conformational states of the hammerhead ribozyme, broad time range of relaxation and topology of dynamics. Nucleic Acids Res 28, 4428-4434, (2000).

123. Tuschl, T., Gohlke, C., Jovin, T. M., Westhof, E. & Eckstein, F. A three-dimensional model for the hammerhead ribozyme based on fluorescence measurements. Science 266, 785-789, (1994).

124. Dujon, B. et al. Genome evolution in yeasts. Nature 430, 35-44, (2004).

125. Jimenez, R. M., Rampasek, L., Brejova, B., Vinar, T. & Luptak, A. Discovery of RNA motifs using a computational pipeline that allows insertions in paired regions and filtering of candidate sequences. Methods Mol Biol 848, 145-158, (2012).

126. Mandal, M., Boese, B., Barrick, J. E., Winkler, W. C. & Breaker, R. R. Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. Cell 113, 577-586, (2003).

127. Miquel, S. et al. Faecalibacterium prausnitzii and human intestinal health. Curr Opin Microbiol 16, 255-261, (2013).

128. Nielsen, H. & Johansen, S. D. Group I introns: Moving in new directions. RNA Biol 6, 375-383, (2009).

129. Lee, E. R., Baker, J. L., Weinberg, Z., Sudarsan, N. & Breaker, R. R. An allosteric self-splicing ribozyme triggered by a bacterial second messenger. Science 329, 845-848, (2010).

130. Chen, A. G., Sudarsan, N. & Breaker, R. R. Mechanism for gene control by a natural allosteric group I ribozyme. RNA 17, 1967-1972, (2011).

131. Collins, J. A., Irnov, I., Baker, S. & Winkler, W. C. Mechanism of mRNA destabilization by the glmS ribozyme. Genes Dev 21, 3356-3368, (2007).

132. Gorke, B. & Vogel, J. Noncoding RNA control of the making and breaking of sugars. Gene Dev 22, 2914-2925, (2008).

133. McCown, P. J., Roth, A. & Breaker, R. R. An expanded collection and refined consensus model of glmS ribozymes. RNA 17, 728-736, (2011).

134. De Vos, P. & ebrary Inc. (Springer, New York, 2009).

135. Tavares, I. M., Leitao, J. H. & Sa-Correia, I. Chromosomal organization and transcription analysis of genes in the vicinity of Pseudomonas aeruginosa glmM gene encoding phosphoglucosamine mutase. Biochem Biophys Res Commun 302, 363-371, (2003).

136. Jimenez, R. M. & Luptak, A. Structure-based search and in vitro analysis of self-cleaving ribozymes. Methods Mol Biol 848, 131-143, (2012).

137. Riccitelli, N. J., Delwart, E. & Luptak, A. Identification of minimal HDV-like ribozymes with unique divalent metal ion dependence in the human microbiome. Biochemistry 53, 1616-1626, (2014).

138. Chadalavada, D. M., Cerrone-Szakal, A. L., Wilcox, J. L., Siegfried, N. A. & Bevilacqua, P. C. Mechanistic analysis of the hepatitis delta virus (HDV) ribozyme: methods for RNA preparation, structure mapping, solvent isotope effects, and co-transcriptional cleavage. Methods Mol Biol 848, 21-40, (2012).

139. Shepotinovskaya, I. & Uhlenbeck, O. C. Enhanced product stability in the hammerhead ribozyme. Biochemistry 49, 4494-4500, (2010).

140. Chadalavada, D. M., Gratton, E. A. & Bevilacqua, P. C. The Human HDV-like CPEB3 Ribozyme Is Intrinsically Fast-Reacting. Biochemistry 49, 5321-5330, (2010).

141. Bennett, B. D. et al. Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. Nat Chem Biol 5, 593-599, (2009).

142. Rogers, J., Chang, A. H., von Ahsen, U., Schroeder, R. & Davies, J. Inhibition of the self-cleavage reaction of the human hepatitis delta virus ribozyme by antibiotics. J Mol Biol 259, 916-925, (1996).

143. Chia, J. S., Wu, H. L., Wang, H. W., Chen, D. S. & Chen, P. J. Inhibition of Hepatitis Delta Virus Genomic Ribozyme Self-Cleavage by Aminoglycosides. J Biomed Sci 4, 208-216, (1997).

144. Stokowa-Soltys, K. et al. High affinity of copper(II) towards amoxicillin, apramycin and ristomycin. Effect of these complexes on the catalytic activity of HDV ribozyme. J Inorg Biochem 124, 26-34, (2013).

145. Regulski, E. E. & Breaker, R. R. In-line probing analysis of riboswitches. Methods Mol Biol 419, 53-67, (2008).

146. Bandyra, K. J. & Luisi, B. F. Licensing and due process in the turnover of bacterial RNA. Rna Biology 10, 627-635, (2013).

147. Mathy, N. et al. 5 '-to-3 ' exoribonuclease activity in Bacteria: Role of RNase J1 in rRNA maturation and 5 ' stability of mRNA. Cell 129, 681-692, (2007).

148. Ucker, D. S. & Yamamoto, K. R. Early events in the stimulation of mammary tumor virus RNA synthesis by glucocorticoids. Novel assays of transcription rates. J Biol Chem 259, 7416-7420, (1984).

149. Webb, C. H. & Luptak, A. HDV-like self-cleaving ribozymes. RNA Biol 8, (2011).

150. Jakubczak, J. L., Burke, W. D. & Eickbush, T. H. Retrotransposable Elements R1 and R2 Interrupt the Ribosomal-Rna Genes of Most Insects. P Natl Acad Sci USA 88, 3295-3299, (1991).

151. Jamrich, M. & Miller, O. L., Jr. The rare transcripts of interrupted rRNA genes in Drosophila melanogaster are processed or degraded during synthesis. EMBO J 3, 1541-1545, (1984).

152. Luan, D. D., Korman, M. H., Jakubczak, J. L. & Eickbush, T. H. Reverse Transcription of R2bm Rna Is Primed by a Nick at the Chromosomal Target Site - a Mechanism for Non-Ltr Retrotransposition. Cell 72, 595-605, (1993).

153. Christensen, S. M., Ye, J. & Eickbush, T. H. RNA from the 5' end of the R2 retrotransposon controls R2 protein binding to and cleavage of its DNA target site. Proc Natl Acad Sci U S A 103, 17602-17607, (2006).

154. George, J. A. & Eickbush, T. H. Conserved features at the 5 ' end of Drosophila R2 retrotransposable elements: implications for transcription and translation. Insect Mol Biol 8, 3-10, (1999).

155. Lozano, G. & Martinez-Salas, E. Structural insights into viral IRES-dependent translation mechanisms. Curr Opin Virol 12, 113-120, (2015).

156. Berry, K. E., Waghray, S. & Doudna, J. A. The HCV IRES pseudoknot positions the initiation codon on the 40S ribosomal subunit. RNA 16, 1559-1569, (2010).

157. Ruminski, D. Characterization of biological roles for HDV-like self-cleaving ribozymes. Thesis, University of California, Irvine, (2013).

158. Luptak, A. Structure-function relationship in RNA: bacterial group I intron folding and active site protonation in a genomic HDV ribozyme,. Thesis, Yale University, (2002).

159. Thompson, S. R. Tricks an IRES uses to enslave ribosomes. Trends Microbiol 20, 558-566, (2012).

160. Shatsky, I. N., Dmitriev, S. E., Terenin, I. M. & Andreev, D. E. Cap- and IRES-independent scanning mechanism of translation initiation as an alternative to the concept of cellular IRESs. Mol Cells 30, 285-293, (2010).

161. Wang, Z., Day, N., Trifillis, P. & Kiledjian, M. An mRNA stability complex functions with poly(A)-binding protein to stabilize mRNA in vitro. Mol Cell Biol 19, 4552-4560, (1999).

162. Dower, K., Kuperwasser, N., Merrikh, H. & Rosbash, M. A synthetic A tail rescues yeast nuclear accumulation of a ribozyme-terminated transcript. RNA 10, 1888-1899, (2004).

163. Yeung, G. et al. Poly(A)-driven and poly(A)-assisted termination: two different modes of poly(A)-dependent transcription termination. Mol Cell Biol 18, 276-289, (1998).

164. Cooke, C., Hans, H. & Alwine, J. C. Utilization of splicing elements and polyadenylation signal elements in the coupling of polyadenylation and last-intron removal. Mol Cell Biol 19, 4971-4979, (1999).

165. Paek, K. Y. et al. Translation initiation mediated by RNA looping. Proc Natl Acad Sci U S A 112, 1041-1046, (2015).

166. Hartz, D., McPheeters, D. S., Traut, R. & Gold, L. Extension inhibition analysis of translation initiation complexes. Methods Enzymol 164, 419-425, (1988).

167. Hartz, D., McPheeters, D. S. & Gold, L. Selection of the initiator tRNA by Escherichia coli initiation factors. Genes Dev 3, 1899-1912, (1989).

168. Rio, D. C. Electrophoretic mobility shift assays for RNA-protein complexes. Cold Spring Harb Protoc 2014, 435-440, (2014).

169. Tacheny, A., Dieu, M., Arnould, T. & Renard, P. Mass spectrometry-based identification of proteins interacting with nucleic acids. J Proteomics 94, 89-109, (2013).

170. Mittal, N., Roy, N., Babu, M. M. & Janga, S. C. Dissecting the expression dynamics of RNA-binding proteins in posttranscriptional regulatory networks. Proc Natl Acad Sci U S A 106, 20300-20305, (2009).

171. Li, D., Wei, T., Abbott, C. M. & Harrich, D. The unexpected roles of eukaryotic translation elongation factors in RNA virus replication and pathogenesis. Microbiol Mol Biol Rev 77, 253-266, (2013).

172. Sikora, D., Greco-Stewart, V. S., Miron, P. & Pelchat, M. The hepatitis delta virus RNA genome interacts with eEF1A1, p54(nrb), hnRNP-L, GAPDH and ASF/SF2. Virology 390, 71-78, (2009).

173. Guerriertakada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The Rna Moiety of Ribonuclease-P Is the Catalytic Subunit of the Enzyme. Cell 35, 849-857, (1983).

174. Esakova, O. & Krasilnikov, A. S. Of proteins and RNA: the RNase P/MRP family. RNA 16, 1725-1747, (2010).

175. Szklarczyk, D. et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 43, D447-452, (2015).

176. Ozgur, S. et al. The conformational plasticity of eukaryotic RNA-dependent ATPases. FEBS J 282, 850-863, (2015).

177. Ray, B. K. et al. ATP-dependent unwinding of messenger RNA structure by eukaryotic initiation factors. J Biol Chem 260, 7651-7658, (1985).

178. Chen, J. Y. F. et al. Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. Molecular Cell 7, 227-232, (2001).

179. Morley, S. J. & Hershey, J. W. A fractionated reticulocyte lysate retains high efficiency for protein synthesis. Biochimie 72, 259-264, (1990).

APPENDIX A

Chapter 3

Constructs designed for future *in vitro* self-cleavage assays

1) drz-Csac-1-L3

2) drz-Csac-1-P1

3) drz-Csac-1-L3/P1

4) drz-Fpra-1-L3

5) drz-Fpra-1-P1

6) drz-Fpra-1-L3/P1

Where L3 refers to an exchange of the four nucleotides between P1.1 and the base of P3, P1 refers to an exchange of the entire P1 helix, and L3/P1 means that both changes were made.

Constructs used during *in vivo* studies

Fpra1-WT: The wild type construct is the active form of drz-Fpra-1 in an intergenic region very close to the wild type sequence.

Fpra1-C_{58}U: The point mutation creates an inactive form of drz-Fpra-1.

Fpra1-noSD: The active form of the ribozyme where the Shine Dalgarno sequence has been disrupted by a serious of nucleotide changes.

Fpra1-noRBZ: The ribozyme secondary structure has been deleted. The intergenic region is shortened to 73 nts and still contains the Shine Dalgarno sequence.

*Fpra1-del3, Fpra1-del6, and Fpra1-del9: The noSD construct with the series of deletions between the ribozyme secondary structure and the AUG start codon of the *Fluc* ORF.

*Fpra-spacer6: The noSD construct with the space between the ribozyme secondary structure and the AUG start codon of the *Fluc* ORF reduced from 36 to 6 nts.

*not tested

Cell lines used for *in vivo* studies

The expression of the bicistronic transcript is under the control of the lac promoter. The predominant cell line used for *in vivo* assays is Mach1-T1R (Invitrogen). This cell line is typically used for cloning purposes and is fast growing with a doubling time of approximately 50 min in LB media. The parental *E.coli* strain is

non-K-12, wild-type W strain (ATCC #9637, S. A. Waksman). The entire *lac* operon has been deleted which means that there is no lac repressor expressed in this cell line. This means that there is constitutive expression from the lac promoter. The genotype for Mach1 *E.coli* is given below:

Mach1-T1R (Invitrogen)

F– Φ80*lac*ZΔM15 Δ*lac*X74 *hsd*R(rK–, mK+) Δ*rec*A1398 *end*A1 *ton*A

We also used an inducible system in order to more closely measure the expression of both reporters. Using an inducible system for *in vivo* assays serves the benefit of observing the production of protein as a direct result of transcript production. In this new system we can control the culture conditions and decide when to induce expression of our bicistronic construct. In addition, we can better understand the natural course of transcript lifetime and the correlation with the observed levels of reporter gene product. The XL-1 *E.coli* strain overexpresses the lac repressor; therefore isopropyl β-D-1-thiogalactopyranoside (IPTG) can be used to induce expression from the lac promoter. The genotype for the XL-1 blue *E.coli* is given below:

XL-1 blue (Agilent)

endA1 gyrA96(nal^{R}) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB^{+} lacI^{q} Δ(lacZ)M15] hsdR17(r_K^- m_K^+)

APPENDIX B

Chapter 4

Constructs used for *in vitro* translation assays

rbz-Dsim constructs

Dsim-WT-AUG: The downstream Fluc ORF starts with an AUG start codon. The remaining constructs in this group also utilize the AUG start codon.

Dsim-C175U: A point mutant of drz-Dsim-1 leading to an inactive form of the ribozyme.

Dsim-Δleader: The first nucleotide transcribed is G1 located at the 5′ end of drz-Dsim-1. While normally this would be the site of self-cleavage resulting in a 5′-OH, this construct contains a 5′-triphosphate.

Dsim-ΔP4: A Dsim-WT variant in which the entire P4 helix has been deleted and there is a four nucleotide linker between P1.1 and the catalytic C175. This form of the ribozyme is still active *in vitro* although with a reduced kobs.

Dsim-P2: An inactive form of Dsim-WT in which transcription begins at the top of P2 (essentially the 5' side of the P1 helix and the J1/2 region are deleted). We have not validated the secondary structure the construct actually assumes *in vitro*.

Dsim-P1.1: An inactive form of Dsim-WT in which transcription begins at the 3′ side of P1.1. This transcript contains the sequence corresponding to J1.1/4, P4, active site, and 3′ side of P2 of the Dsim ribozyme.

Dsim-P4: The transcript of this construct contains the sequence corresponding to the 3′ side of P4, active site, and the 3′ side of P2.

rbz-CIV constructs

CIV-WT-AUG

CIV-WT-UUU

CIV-WT-AUG-Δ3'UTR

CIV-WT-UUU-Δ3'UTR

rbz-Hsap (CPEB3) constructs

CPEB3-WT-AUG

CPEB3-WT-UUU

CPEB3-WT-AUG-Δ3'UTR

CPEB3-WT-UUU-Δ3'UTR

Other structured RNA constructs

HHR2-WT-AUG

HHR2-WT-UUU

HHR2-WT-AUG-Δ3'UTR

HHR2-WT-UUU-Δ3'UTR

HCV-IRES-AUG

rbz-Dsim mutants

Dsim-WT-UUU: The active form of drz-Dsim-1 and the downstream Fluc ORF start codon has been mutated to UUU. The remaining constructs in this group also utilize the UUU codon.

Below are various point mutations (changes, insertions, or deletions) and double mutations. These mutations are in the P2 helix. The predicted changes in ΔG, referring to the thermostability of the entire nested double-pseudoknot secondary structure without leader or tail, are given for each construct. The calculated changes in ΔG for the P2 helix only are also given.

Construct name	Change in ΔG from WT (kcall/mol)	Change in from WT (kcal/mol) P2 only
Dsim-g185a	7.09	4.65
Dsim-g185a-c126t	5.49	2.36
Dsim-t182c	5.09	2.43
Dsim-t182c-a129g	1.29	-1.37
Dsim-t122c	0.36	-2.7
Dsim-185ins	-1.91	-4.99
Dsim-Δa131	0.56	0.83
Dsim-Δa131-Δt180	N/A	N/A
Dsim-t182c-Δa131	N/A	N/A
Dsim-g185a-Δa131	N/A	N/A

Constructs used for natEMSA

(rbz): The 3′ product of HDV-like ribozyme self-cleavage.

(full): The full length transcript for an HDV-like ribozyme with an approximately 70 nt long leader.

Each construct is named by a description for the 5′ end of the transcript (described above) and the name for the HDV-like ribozyme, followed by the length of the tail.

rbz-CPEB3-12

full-CPEB3-12

rbz-CPEB3-7

rbz-Fpra-1-6

rbz-Dsim-7

rbz-Dsim-45

rbz-Dsim-127

full-Dsim-C158U-7

rbz-CIV-8

full-CIV-8

Dsim- Δleader-7

Dsim-P2-7

Dsim-P1.1-7

Dsim-ΔP4-7

rbz-Dsim-19

rbz-Dsim-g185a-19

rbz-Dsim-t182c-19

rbz-Dsim-del131a-19

HCV IRES-UUU

MALDI TOF results from MASCOT web server for peptide mass fingerprint data analysis.

Complex B was isolated from a natEMSA experiment and subjected to in-gel trypsin digest prior to MALDI-TOF.

1. TSR3_HUMAN Mass: 34145 Score: 39 Expect: 9 Matches: 16

   Ribosome biogenesis protein TSR3 homolog OS=Homo sapiens GN=TSR3 PE=1 SV=1

2. HOME3_RAT Mass: 39867 Score: 34 Expect: 26 Matches: 15

   Homer protein homolog 3 OS=Rattus norvegicus GN=Homer3 PE=2 SV=2

3. ST5_HUMAN Mass: 127603 Score: 34 Expect: 29 Matches: 35

   Suppression of tumorigenicity 5 protein OS=Homo sapiens GN=ST5 PE=1 SV=3

4. WDR53_MOUSE Mass: 39245 Score: 32 Expect: 42 Matches: 17

WD repeat-containing protein 53 OS=Mus musculus GN=Wdr53 PE=2 SV=1

5. CRIP2_BOVIN Mass: 23417 Score: 31 Expect: 50 Matches: 12

Cysteine-rich protein 2 OS=Bos taurus GN=CRIP2 PE=2 SV=1

6. TFPI2_HUMAN Mass: 27942 Score: 31 Expect: 58 Matches: 14

Tissue factor pathway inhibitor 2 OS=Homo sapiens GN=TFPI2 PE=1 SV=1

7. UBE3A_HUMAN Mass: 101593 Score: 30 Expect: 59 Matches: 32

Ubiquitin-protein ligase E3A OS=Homo sapiens GN=UBE3A PE=1 SV=4

8. BRF2_MACFA Mass: 42590 Score: 29 Expect: 76 Matches: 15

Transcription factor IIIB 50 kDa subunit OS=Macaca fascicularis GN=BRF2 PE=2 SV=1

These are a second MALDI-TOF of complex B (same prep as above).

1. GTPB1_BOVIN Mass: 73223 Score: 50 Expect: 0.68 Matches: 53

GTP-binding protein 1 OS=Bos taurus GN=GTPBP1 PE=2 SV=2

2. GTPB1_MOUSE Mass: 72939 Score: 50 Expect: 0.74 Matches: 53

GTP-binding protein 1 OS=Mus musculus GN=Gtpbp1 PE=1 SV=2

3. GTPB1_RAT Mass: 73070 Score: 49 Expect: 0.9 Matches: 53

GTP-binding protein 1 OS=Rattus norvegicus GN=Gtpbp1 PE=1 SV=1

4. GTPB1_HUMAN Mass: 73035 Score: 46 Expect: 1.6 Matches: 57

GTP-binding protein 1 OS=Homo sapiens GN=GTPBP1 PE=1 SV=3

5. AT2B2_MOUSE Mass: 133701 Score: 33 Expect: 35 Matches: 64

Plasma membrane calcium-transporting ATPase 2 OS=Mus musculus GN=Atp2b2 PE=1 SV=2
AT2B2_RAT Mass: 137922 Score: 32 Expect: 46 Matches: 64

Plasma membrane calcium-transporting ATPase 2 OS=Rattus norvegicus GN=Atp2b2 PE=2 SV=2
S10A8_BOVIN Mass: 10567 Score: 22 Expect: 4.1e+02 Matches: 8

Protein S100-A8 OS=Bos taurus GN=S100A8 PE=1 SV=2

6. T2FB_BOVIN Mass: 28486 Score: 31 Expect: 58 Matches: 18

General transcription factor IIF subunit 2 OS=Bos taurus GN=GTF2F2 PE=2 SV=1

7. IL12B_MESAU Mass: 37759 Score: 29 Expect: 87 Matches: 19

Interleukin-12 subunit beta OS=Mesocricetus auratus GN=IL12B PE=2 SV=1

8. CUTC_MOUSE Mass: 29384 Score: 28 Expect: 1e+02 Matches: 21

Copper homeostasis protein cutC homolog OS=Mus musculus GN=Cutc PE=2 SV=1

---

9. MIPO1_MOUSE Mass: 32073 Score: 28 Expect: 1.1e+02 Matches: 26

Mirror-image polydactyly gene 1 protein homolog OS=Mus musculus GN=Mipol1 PE=2 SV=1

---

10. PRDC1_HUMAN Mass: 25828 Score: 28 Expect: 1.1e+02 Matches: 22

Phosphoribosyltransferase domain-containing protein 1 OS=Homo sapiens GN=PRTFDC1 PE=1 SV=1

---

11. S10A4_MOUSE Mass: 11942 Score: 28 Expect: 1.2e+02 Matches: 12
Protein S100-A4 OS=Mus musculus GN=S100a4 PE=1 SV=1

---

12. AT2B2_HUMAN Mass: 137987 Score: 27 Expect: 1.4e+02 Matches: 62

Plasma membrane calcium-transporting ATPase 2 OS=Homo sapiens GN=ATP2B2 PE=1 SV=2

---

13. DPYS_MOUSE Mass: 57202 Score: 27 Expect: 1.4e+02 Matches: 30

Dihydropyrimidinase OS=Mus musculus GN=Dpys PE=1 SV=2

---

14. LRC51_BOVIN Mass: 22151 Score: 26 Expect: 1.6e+02 Matches: 16

Leucine-rich repeat-containing protein 51 OS=Bos taurus GN=LRRC51 PE=2 SV=1

---

15. CE047_HUMAN Mass: 19365 Score: 26 Expect: 1.6e+02 Matches: 26
Uncharacterized protein C5orf47 OS=Homo sapiens GN=C5orf47 PE=2 SV=2

---

16. RN150_HUMAN Mass: 48896 Score: 25 Expect: 2.1e+02 Matches: 26
RING finger protein 150 OS=Homo sapiens GN=RNF150 PE=2 SV=2

These results are from the first partially denaturing wash (PD1) from RNA affinity chromatography experiment with the matrix as a target and fractionated RRL as the source.

1. PSD3_HUMAN Mass: 116646 Score: 34 Expect: 25 Matches: 30

PH and SEC7 domain-containing protein 3 OS=Homo sapiens GN=PSD3 PE=1 SV=2

---

2. TCAM2_MOUSE Mass: 26574 Score: 34 Expect: 27 Matches: 17

TIR domain-containing adapter molecule 2 OS=Mus musculus GN=Ticam2 PE=1 SV=1

---

3. GLRX2_PONAB Mass: 18104 Score: 34 Expect: 28 Matches: 13

Glutaredoxin-2, mitochondrial OS=Pongo abelii GN=GLRX2 PE=2 SV=1

---

4. PARVB_HUMAN Mass: 41745 Score: 34 Expect: 30 Matches: 22

Beta-parvin OS=Homo sapiens GN=PARVB PE=1 SV=1

---

5. HPRT_CRIGR Mass: 24856 Score: 33 Expect: 31 Matches: 16

Hypoxanthine-guanine phosphoribosyltransferase OS=Cricetulus griseus GN=HPRT1 PE=2 SV=2

---

6. PNDC1_PONAB Mass: 60570 Score: 33 Expect: 34 Matches: 20

Poly(A)-specific ribonuclease PARN-like domain-containing protein 1 OS=Pongo abelii GN=PNLDC1 PE=2 SV=1

---

7. GPR31_MOUSE Mass: 36246 Score: 32 Expect: 39 Matches: 16

12-(S)-hydroxy-5,8,10,14-eicosatetraenoic acid receptor OS=Mus musculus GN=Gpr31 PE=3 SV=1

---

8. CLK2_MOUSE Mass: 60463 Score: 32 Expect: 39 Matches: 23

Dual specificity protein kinase CLK2 OS=Mus musculus GN=Clk2 PE=1 SV=2

---

9. CLK2_HUMAN Mass: 60509 Score: 32 Expect: 42 Matches: 25

Dual specificity protein kinase CLK2 OS=Homo sapiens GN=CLK2 PE=1 SV=1

---

10. CX056_PONAB Mass: 25908 Score: 32 Expect: 43 Matches: 18

UPF0428 protein CXorf56 homolog OS=Pongo abelii PE=2 SV=1
CX056_HUMAN Mass: 26007 Score: 30 Expect: 68 Matches: 18

UPF0428 protein CXorf56 OS=Homo sapiens GN=CXorf56 PE=1 SV=1

---

11. PDLI3_PIG Mass: 40091 Score: 32 Expect: 47 Matches: 14

PDZ and LIM domain protein 3 OS=Sus scrofa GN=PDLIM3 PE=2 SV=1

---

12. TRAM1_RAT Mass: 43174 Score: 32 Expect: 47 Matches: 21

Translocating chain-associated membrane protein 1 OS=Rattus norvegicus GN=Tram1 PE=2 SV=3
TRAM1_MOUSE Mass: 43183 Score: 27 Expect: 1.4e+02 Matches: 20

Translocating chain-associated membrane protein 1 OS=Mus musculus GN=Tram1 PE=1 SV=3

---

13. ATP8_ELEMA Mass: 7944 Score: 31 Expect: 50 Matches: 13

ATP synthase protein 8 OS=Elephas maximus GN=MT-ATP8 PE=3 SV=1
ATP8_LOXAF Mass: 7944 Score: 31 Expect: 50 Matches: 13

ATP synthase protein 8 OS=Loxodonta africana GN=MT-ATP8 PE=3 SV=1

---

14. SCD5_BOVIN Mass: 38362 Score: 31 Expect: 52 Matches: 17

Stearoyl-CoA desaturase 5 OS=Bos taurus GN=SCD5 PE=2 SV=1

---

15. C1QB_BOVIN Mass: 26611 Score: 31 Expect: 54 Matches: 12

Complement C1q subcomponent subunit B OS=Bos taurus GN=C1QB PE=1 SV=1

---

16. RM12_CRICR Mass: 21694 Score: 31 Expect: 54 Matches: 15

39S ribosomal protein L12, mitochondrial OS=Cricetus cricetus GN=MRPL12 PE=2 SV=1
RM12_MOUSE Mass: 21809 Score: 27 Expect: 1.3e+02 Matches: 15

39S ribosomal protein L12, mitochondrial OS=Mus musculus GN=Mrpl12 PE=1 SV=2

---

17. ACADS_PIG Mass: 45221 Score: 30 Expect: 59 Matches: 20

Short-chain specific acyl-CoA dehydrogenase, mitochondrial OS=Sus scrofa GN=ACADS PE=2 SV=1

---

18. TA2R7_RAT Mass: 36447 Score: 30 Expect: 61 Matches: 19

Taste receptor type 2 member 7 OS=Rattus norvegicus GN=Tas2r7 PE=1 SV=2

---

19. IGF1_CAPHI Mass: 17641 Score: 30 Expect: 63 Matches: 12

Insulin-like growth factor I OS=Capra hircus GN=IGF1 PE=2 SV=2

---

20. Z280B_HUMAN Mass: 62514 Score: 30 Expect: 73 Matches: 22

Zinc finger protein 280B OS=Homo sapiens GN=ZNF280B PE=1 SV=2

These results are from the elution of the RNA affinity chromatography experiment with only the matrix as a target and fractionated RRL as the source.

1. CO2_BOVIN Mass: 84164 Score: 38 Expect: 10 Matches: 27

Complement C2 OS=Bos taurus GN=C2 PE=2 SV=1

---

2. ONCM_BOVIN Mass: 27878 Score: 37 Expect: 12 Matches: 17

Oncostatin-M OS=Bos taurus GN=OSM PE=3 SV=1

---

3. B2MG_CACME Mass: 13764 Score: 37 Expect: 13 Matches: 11

Beta-2-microglobulin OS=Cacajao melanocephalus GN=B2M PE=3 SV=1

---

4. TRFE_RABIT Mass: 78901 Score: 34 Expect: 25 Matches: 25

Serotransferrin OS=Oryctolagus cuniculus GN=TF PE=1 SV=4

---

5. IFT74_HUMAN Mass: 69310 Score: 34 Expect: 25 Matches: 25

Intraflagellar transport protein 74 homolog OS=Homo sapiens GN=IFT74 PE=1 SV=1

---

6. SOX_MOUSE Mass: 44446 Score: 32 Expect: 43 Matches: 17

Peroxisomal sarcosine oxidase OS=Mus musculus GN=Pipox PE=1 SV=1

---

7. PDLI3_PIG Mass: 40091 Score: 31 Expect: 48 Matches: 16

PDZ and LIM domain protein 3 OS=Sus scrofa GN=PDLIM3 PE=2 SV=1

---

8. MB3L1_MOUSE Mass: 20948 Score: 31 Expect: 48 Matches: 13

Methyl-CpG-binding domain protein 3-like 1 OS=Mus musculus GN=Mbd3l1 PE=2 SV=1

9. PDIP2_MOUSE Mass: 42072 Score: 31 Expect: 54 Matches: 17

Polymerase delta-interacting protein 2 OS=Mus musculus GN=Poldip2 PE=2 SV=1

10. HPCA_BOVIN Mass: 22527 Score: 30 Expect: 62 Matches: 12

Neuron-specific calcium-binding protein hippocalcin OS=Bos taurus GN=HPCA PE=2 SV=3
HPCA_HUMAN Mass: 22527 Score: 30 Expect: 62 Matches: 12

Neuron-specific calcium-binding protein hippocalcin OS=Homo sapiens GN=HPCA PE=2 SV=2
HPCA_MOUSE Mass: 22527 Score: 30 Expect: 62 Matches: 12

Neuron-specific calcium-binding protein hippocalcin OS=Mus musculus GN=Hpca PE=1 SV=2
HPCA_PIG Mass: 22527 Score: 30 Expect: 62 Matches: 12

Neuron-specific calcium-binding protein hippocalcin OS=Sus scrofa GN=HPCA PE=2 SV=3
HPCA_RAT Mass: 22527 Score: 30 Expect: 62 Matches: 12

Neuron-specific calcium-binding protein hippocalcin OS=Rattus norvegicus GN=Hpca PE=1 SV=2

11. RPC5_HUMAN Mass: 80532 Score: 29 Expect: 76 Matches: 23

DNA-directed RNA polymerase III subunit RPC5 OS=Homo sapiens GN=POLR3E PE=1 SV=1

12. RFX5_HUMAN Mass: 65682 Score: 29 Expect: 86 Matches: 26

DNA-binding protein RFX5 OS=Homo sapiens GN=RFX5 PE=1 SV=1

13. IL8_PIG Mass: 11910 Score: 29 Expect: 87 Matches: 8

Interleukin-8 OS=Sus scrofa GN=CXCL8 PE=1 SV=1

The first partially denaturing wash (PD1) from RNA affinity chromatography beads with rbz-Dsim-45 as a target and fractionated, pre-cleared RRL as source.

1. REEP3_HUMAN Mass: 29245 Score: 40 Expect: 7.1 Matches: 20

Receptor expression-enhancing protein 3 OS=Homo sapiens GN=REEP3 PE=1 SV=1

2. DMRT2_MOUSE Mass: 62400 Score: 35 Expect: 20 Matches: 27

Doublesex- and mab-3-related transcription factor 2 OS=Mus musculus GN=Dmrt2 PE=1 SV=1
DMRT2_HUMAN Mass: 62687 Score: 30 Expect: 73 Matches: 26

Doublesex- and mab-3-related transcription factor 2 OS=Homo sapiens GN=DMRT2 PE=2 SV=2

3. CLK2_HUMAN Mass: 60509 Score: 33 Expect: 33 Matches: 23

Dual specificity protein kinase CLK2 OS=Homo sapiens GN=CLK2 PE=1 SV=1

CLK2_MOUSE Mass: 60463 Score: 27 Expect: 1.3e+02 Matches: 21

Dual specificity protein kinase CLK2 OS=Mus musculus GN=Clk2 PE=1 SV=2

4. SCD5_BOVIN Mass: 38362 Score: 30 Expect: 61 Matches: 17

Stearoyl-CoA desaturase 5 OS=Bos taurus GN=SCD5 PE=2 SV=1

5. KCND3_RAT Mass: 74378 Score: 30 Expect: 65 Matches: 25

Potassium voltage-gated channel subfamily D member 3 OS=Rattus norvegicus GN=Kcnd3 PE=1 SV=2

6. RND3_PIG Mass: 27892 Score: 29 Expect: 80 Matches: 14

Rho-related GTP-binding protein RhoE OS=Sus scrofa GN=RND3 PE=2 SV=1

7. RM12_HUMAN Mass: 21563 Score: 29 Expect: 82 Matches: 12

39S ribosomal protein L12, mitochondrial OS=Homo sapiens GN=MRPL12 PE=1 SV=2

8. ST1E1_CAVPO Mass: 35564 Score: 29 Expect: 90 Matches: 17

Estrogen sulfotransferase OS=Cavia porcellus GN=SULT1E1 PE=1 SV=1

9. CSN4_RAT Mass: 46546 Score: 28 Expect: 1e+02 Matches: 17

COP9 signalosome complex subunit 4 OS=Rattus norvegicus GN=Cops4 PE=1 SV=1
CSN4_HUMAN Mass: 46525 Score: 26 Expect: 1.6e+02 Matches: 17

COP9 signalosome complex subunit 4 OS=Homo sapiens GN=COPS4 PE=1 SV=1
CSN4_MACFA Mass: 46525 Score: 26 Expect: 1.6e+02 Matches: 17

COP9 signalosome complex subunit 4 OS=Macaca fascicularis GN=COPS4 PE=2 SV=1
CSN4_MOUSE Mass: 46541 Score: 26 Expect: 1.6e+02 Matches: 17

COP9 signalosome complex subunit 4 OS=Mus musculus GN=Cops4 PE=1 SV=1
CSN4_PIG Mass: 46525 Score: 26 Expect: 1.6e+02 Matches: 17

COP9 signalosome complex subunit 4 OS=Sus scrofa GN=COPS4 PE=2 SV=1
CSN4_PONAB Mass: 46511 Score: 26 Expect: 1.6e+02 Matches: 17

COP9 signalosome complex subunit 4 OS=Pongo abelii GN=COPS4 PE=2 SV=1

10. SVS5_MOUSE Mass: 13011 Score: 28 Expect: 1e+02 Matches: 10

Seminal vesicle secretory protein 5 OS=Mus musculus GN=Svs5 PE=2 SV=1

11. PDE6D_BOVIN Mass: 17493 Score: 28 Expect: 1.1e+02 Matches: 14

Retinal rod rhodopsin-sensitive cGMP 3',5'-cyclic phosphodiesterase subunit delta OS=Bos taurus GN=PDE6D PE=1 SV=1
PDE6D_HUMAN Mass: 17523 Score: 28 Expect: 1.1e+02 Matches: 14

Retinal rod rhodopsin-sensitive cGMP 3',5'-cyclic phosphodiesterase subunit delta OS=Homo sapiens GN=PDE6D PE=1 SV=1

12. PDE6D_MOUSE Mass: 17451 Score: 28 Expect: 1.1e+02 Matches: 14

Retinal rod rhodopsin-sensitive cGMP 3',5'-cyclic phosphodiesterase subunit delta OS=Mus musculus GN=Pde6d PE=1 SV=1

13. CS043_HUMAN Mass: 18465 Score: 28 Expect: 1.2e+02 Matches: 15
Uncharacterized protein C19orf43 OS=Homo sapiens GN=C19orf43 PE=1 SV=1

14. GAR1_HUMAN Mass: 22505 Score: 27 Expect: 1.2e+02 Matches: 8

H/ACA ribonucleoprotein complex subunit 1 OS=Homo sapiens GN=GAR1 PE=1 SV=1

15. ZN449_GORGO Mass: 61178 Score: 27 Expect: 1.3e+02 Matches: 16
Zinc finger protein 449 OS=Gorilla gorilla gorilla GN=ZNF449 PE=3 SV=1
ZN449_HUMAN Mass: 61148 Score: 27 Expect: 1.3e+02 Matches: 16
Zinc finger protein 449 OS=Homo sapiens GN=ZNF449 PE=1 SV=3
ZN449_PANPA Mass: 61219 Score: 27 Expect: 1.3e+02 Matches: 16
Zinc finger protein 449 OS=Pan paniscus GN=ZNF449 PE=3 SV=1
ZN449_PANTR Mass: 61233 Score: 27 Expect: 1.3e+02 Matches: 16
Zinc finger protein 449 OS=Pan troglodytes GN=ZNF449 PE=3 SV=1

16. SSLP1_MOUSE Mass: 11718 Score: 27 Expect: 1.3e+02 Matches: 7

Secreted seminal-vesicle Ly-6 protein 1 OS=Mus musculus GN=Sslp1 PE=1 SV=1

17. MOTI_MACMU Mass: 12813 Score: 27 Expect: 1.3e+02 Matches: 10

Promotilin OS=Macaca mulatta GN=MLN PE=3 SV=1

18. PDLI2_HUMAN Mass: 37835 Score: 27 Expect: 1.5e+02 Matches: 14
PDZ and LIM domain protein 2 OS=Homo sapiens GN=PDLIM2 PE=1 SV=1
PDLI2_MACFA Mass: 37768 Score: 27 Expect: 1.5e+02 Matches: 14
PDZ and LIM domain protein 2 OS=Macaca fascicularis GN=PDLIM2 PE=2 SV=1

These results are from the elution from the RNA affinity chromatography experiment with rbz-Dsim-45 as a target and fractionated, pre-cleared RRL as source.

1. STP2_MACFA Mass: 16109 Score: 42 Expect: 3.7 Matches: 22

Nuclear transition protein 2 OS=Macaca fascicularis GN=TNP2 PE=2 SV=2
STP2_MACMU Mass: 15690 Score: 31 Expect: 52 Matches: 19

Nuclear transition protein 2 (Fragment) OS=Macaca mulatta GN=TNP2 PE=2 SV=1

2. RM12_HUMAN Mass: 21563 Score: 42 Expect: 4.4 Matches: 14

39S ribosomal protein L12, mitochondrial OS=Homo sapiens GN=MRPL12 PE=1 SV=2

3. MYL9_RAT Mass: 19765 Score: 41 Expect: 4.9 Matches: 13

Myosin regulatory light polypeptide 9 OS=Rattus norvegicus GN=Myl9 PE=1 SV=2

4. RM12_CRICR Mass: 21694 Score: 41 Expect: 5.2 Matches: 15

39S ribosomal protein L12, mitochondrial OS=Cricetus cricetus GN=MRPL12 PE=2 SV=1
RM12_MOUSE Mass: 21809 Score: 31 Expect: 55 Matches: 14

39S ribosomal protein L12, mitochondrial OS=Mus musculus GN=Mrpl12 PE=1 SV=2

---

5. PHF13_HUMAN Mass: 34302 Score: 35 Expect: 19 Matches: 19

PHD finger protein 13 OS=Homo sapiens GN=PHF13 PE=1 SV=2

---

6. GNAI3_HUMAN Mass: 41076 Score: 35 Expect: 21 Matches: 25

Guanine nucleotide-binding protein G(k) subunit alpha OS=Homo sapiens GN=GNAI3 PE=1 SV=3

---

7. F188B_BOVIN Mass: 85250 Score: 35 Expect: 22 Matches: 27

Protein FAM188B OS=Bos taurus GN=FAM188B PE=2 SV=1

---

8. RPC5_HUMAN Mass: 80532 Score: 35 Expect: 22 Matches: 22

DNA-directed RNA polymerase III subunit RPC5 OS=Homo sapiens GN=POLR3E PE=1 SV=1

---

9. PDIA5_HUMAN Mass: 60297 Score: 34 Expect: 24 Matches: 28

Protein disulfide-isomerase A5 OS=Homo sapiens GN=PDIA5 PE=1 SV=1

---

10. RND3_HUMAN Mass: 27864 Score: 34 Expect: 25 Matches: 17

Rho-related GTP-binding protein RhoE OS=Homo sapiens GN=RND3 PE=1 SV=1
RND3_MOUSE Mass: 27864 Score: 34 Expect: 25 Matches: 17

Rho-related GTP-binding protein RhoE OS=Mus musculus GN=Rnd3 PE=1 SV=1
RND3_PIG Mass: 27892 Score: 34 Expect: 25 Matches: 17

Rho-related GTP-binding protein RhoE OS=Sus scrofa GN=RND3 PE=2 SV=1
RND3_RAT Mass: 27864 Score: 34 Expect: 25 Matches: 17

Rho-related GTP-binding protein RhoE OS=Rattus norvegicus GN=Rnd3 PE=2 SV=1

---

11. ANR45_MOUSE Mass: 27945 Score: 34 Expect: 25 Matches: 14

Ankyrin repeat domain-containing protein 45 OS=Mus musculus GN=Ankrd45 PE=2 SV=1

---

12. GRB14_MOUSE Mass: 61162 Score: 33 Expect: 30 Matches: 22

Growth factor receptor-bound protein 14 OS=Mus musculus GN=Grb14 PE=1 SV=1

---

13. MB3L1_MOUSE Mass: 20948 Score: 33 Expect: 31 Matches: 13

Methyl-CpG-binding domain protein 3-like 1 OS=Mus musculus GN=Mbd3l1 PE=2 SV=1

---

14. RM12_BOVIN Mass: 21558 Score: 33 Expect: 31 Matches: 14

39S ribosomal protein L12, mitochondrial OS=Bos taurus GN=MRPL12 PE=1 SV=1

---

15. HMGN5_HUMAN Mass: 31506 Score: 33 Expect: 32 Matches: 27

High mobility group nucleosome-binding domain-containing protein 5 OS=Homo sapiens GN=HMGN5 PE=1 SV=1

---

16. GNAI3_MOUSE Mass: 41082 Score: 33 Expect: 34 Matches: 25

Guanine nucleotide-binding protein G(k) subunit alpha OS=Mus musculus GN=Gnai3 PE=1 SV=3
GNAI3_RAT Mass: 41066 Score: 33 Expect: 34 Matches: 25

Guanine nucleotide-binding protein G(k) subunit alpha OS=Rattus norvegicus GN=Gnai3 PE=1 SV=3

---

17. NOV_HUMAN Mass: 41473 Score: 33 Expect: 34 Matches: 18

Protein NOV homolog OS=Homo sapiens GN=NOV PE=1 SV=1

---

18. CUL2_MOUSE Mass: 87448 Score: 32 Expect: 37 Matches: 26

Cullin-2 OS=Mus musculus GN=Cul2 PE=1 SV=2

---

19. VAX1_HUMAN Mass: 35034 Score: 32 Expect: 39 Matches: 21

Ventral anterior homeobox 1 OS=Homo sapiens GN=VAX1 PE=1 SV=1
VAX1_MOUSE Mass: 35487 Score: 30 Expect: 63 Matches: 18

Ventral anterior homeobox 1 OS=Mus musculus GN=Vax1 PE=2 SV=2

---

20. RT10_PONAB Mass: 23075 Score: 32 Expect: 39 Matches: 11

28S ribosomal protein S10, mitochondrial OS=Pongo abelii GN=MRPS10 PE=2 SV=1

These results are from the PD1 from the RNA affinity chromatography experiment with rbz-Dsim-45 as a target and fractionated, pre-cleared RRL as source. The target and source were crosslinked by UV λ 254 prior to capture on the streptavidin matrix.

1. STP2_MACFA Mass: 16109 Score: 45 Expect: 1.9 Matches: 27

Nuclear transition protein 2 OS=Macaca fascicularis GN=TNP2 PE=2 SV=2
STP2_MACMU Mass: 15690 Score: 34 Expect: 25 Matches: 24

Nuclear transition protein 2 (Fragment) OS=Macaca mulatta GN=TNP2 PE=2 SV=1

---

2. FJX1_HUMAN Mass: 48705 Score: 45 Expect: 2.2 Matches: 30

Four-jointed box protein 1 OS=Homo sapiens GN=FJX1 PE=2 SV=1

---

3. TFPI1_RAT Mass: 35615 Score: 42 Expect: 4.7 Matches: 31

Tissue factor pathway inhibitor OS=Rattus norvegicus GN=Tfpi PE=2 SV=1

---

4. TA2R7_RAT Mass: 36447 Score: 39 Expect: 8.4 Matches: 24

Taste receptor type 2 member 7 OS=Rattus norvegicus GN=Tas2r7 PE=1 SV=2

5. RND3_HUMAN Mass: 27864 Score: 38 Expect: 9.8 Matches: 21

Rho-related GTP-binding protein RhoE OS=Homo sapiens GN=RND3 PE=1 SV=1
RND3_MOUSE Mass: 27864 Score: 38 Expect: 9.8 Matches: 21

Rho-related GTP-binding protein RhoE OS=Mus musculus GN=Rnd3 PE=1 SV=1
RND3_PIG Mass: 27892 Score: 38 Expect: 9.8 Matches: 21

Rho-related GTP-binding protein RhoE OS=Sus scrofa GN=RND3 PE=2 SV=1
RND3_RAT Mass: 27864 Score: 38 Expect: 9.8 Matches: 21

Rho-related GTP-binding protein RhoE OS=Rattus norvegicus GN=Rnd3 PE=2 SV=1
RND3_PONAB Mass: 27765 Score: 32 Expect: 41 Matches: 18

Rho-related GTP-binding protein RhoE OS=Pongo abelii GN=RND3 PE=2 SV=1

6. MYL9_RAT Mass: 19765 Score: 38 Expect: 11 Matches: 15

Myosin regulatory light polypeptide 9 OS=Rattus norvegicus GN=Myl9 PE=1 SV=2

7. RPP29_MOUSE Mass: 25849 Score: 37 Expect: 13 Matches: 26

Ribonuclease P protein subunit p29 OS=Mus musculus GN=Pop4 PE=2 SV=1

8. PHF13_HUMAN Mass: 34302 Score: 36 Expect: 15 Matches: 23

PHD finger protein 13 OS=Homo sapiens GN=PHF13 PE=1 SV=2

9. MTBP_MOUSE Mass: 101619 Score: 36 Expect: 16 Matches: 46

Mdm2-binding protein OS=Mus musculus GN=Mtbp PE=1 SV=1

10. ZNAS2_HUMAN Mass: 21060 Score: 36 Expect: 17 Matches: 17

Putative uncharacterized protein encoded by ZNF503-AS2 OS=Homo sapiens GN=ZNF503-AS2 PE=5 SV=1

11. PHAR3_MOUSE Mass: 62842 Score: 34 Expect: 24 Matches: 29

Phosphatase and actin regulator 3 OS=Mus musculus GN=Phactr3 PE=1 SV=2

12. EF1A1_BOVIN Mass: 50451 Score: 34 Expect: 24 Matches: 26

Elongation factor 1-alpha 1 OS=Bos taurus GN=EEF1A1 PE=1 SV=1
EF1A1_CRIGR Mass: 50424 Score: 34 Expect: 24 Matches: 26

Elongation factor 1-alpha 1 OS=Cricetulus griseus GN=EEF1A1 PE=2 SV=1
EF1A1_FELCA Mass: 50451 Score: 34 Expect: 24 Matches: 26

Elongation factor 1-alpha 1 OS=Felis catus GN=EEF1A1 PE=2 SV=1
EF1A1_HORSE Mass: 50435 Score: 34 Expect: 24 Matches: 26

Elongation factor 1-alpha 1 OS=Equus caballus GN=EEF1A1 PE=2 SV=1
EF1A1_HUMAN Mass: 50451 Score: 34 Expect: 24 Matches: 26

Elongation factor 1-alpha 1 OS=Homo sapiens GN=EEF1A1 PE=1 SV=1
EF1A1_MOUSE Mass: 50424 Score: 34 Expect: 24 Matches: 26

Elongation factor 1-alpha 1 OS=Mus musculus GN=Eef1a1 PE=1 SV=3
EF1A1_PANTR Mass: 50451 Score: 34 Expect: 24 Matches: 26

Elongation factor 1-alpha 1 OS=Pan troglodytes GN=EEF1A1 PE=2 SV=1
EF1A1_PONAB Mass: 50451 Score: 34 Expect: 24 Matches: 26

Elongation factor 1-alpha 1 OS=Pongo abelii GN=EEF1A1 PE=2 SV=2
EF1A1_RABIT Mass: 50451 Score: 34 Expect: 24 Matches: 26

Elongation factor 1-alpha 1 OS=Oryctolagus cuniculus GN=EEF1A1 PE=1 SV=1
EF1A1_RAT Mass: 50424 Score: 34 Expect: 24 Matches: 26

Elongation factor 1-alpha 1 OS=Rattus norvegicus GN=Eef1a1 PE=2 SV=1
EF1A3_HUMAN Mass: 50495 Score: 31 Expect: 52 Matches: 26

Putative elongation factor 1-alpha-like 3 OS=Homo sapiens GN=EEF1A1P5 PE=5 SV=1

---

13. RF1ML_HUMAN Mass: 43687 Score: 34 Expect: 28 Matches: 23

Peptide chain release factor 1-like, mitochondrial OS=Homo sapiens GN=MTRF1L PE=1 SV=1

---

14. TSYL5_HUMAN Mass: 45344 Score: 33 Expect: 36 Matches: 31

Testis-specific Y-encoded-like protein 5 OS=Homo sapiens GN=TSPYL5 PE=1 SV=2

---

15. FA13A_BOVIN Mass: 80392 Score: 32 Expect: 40 Matches: 37

Protein FAM13A OS=Bos taurus GN=FAM13A PE=2 SV=1

---

16. WBS22_HUMAN Mass: 32202 Score: 32 Expect: 40 Matches: 22

Probable 18S rRNA (guanine-N(7))-methyltransferase OS=Homo sapiens GN=WBSCR22 PE=1 SV=2

---

17. BAP29_HUMAN Mass: 28416 Score: 32 Expect: 41 Matches: 20

B-cell receptor-associated protein 29 OS=Homo sapiens GN=BCAP29 PE=1 SV=2
BAP29_PONAB Mass: 28415 Score: 32 Expect: 41 Matches: 20

B-cell receptor-associated protein 29 OS=Pongo abelii GN=BCAP29 PE=2 SV=1

---

18. JUN_HUMAN Mass: 35824 Score: 32 Expect: 42 Matches: 23

Transcription factor AP-1 OS=Homo sapiens GN=JUN PE=1 SV=2
JUN_PIG Mass: 35778 Score: 32 Expect: 42 Matches: 23

Transcription factor AP-1 OS=Sus scrofa GN=JUN PE=2 SV=1
JUN_MOUSE Mass: 36092 Score: 32 Expect: 45 Matches: 23

Transcription factor AP-1 OS=Mus musculus GN=Jun PE=1 SV=3
JUN_RAT Mass: 36149 Score: 32 Expect: 45 Matches: 23

Transcription factor AP-1 OS=Rattus norvegicus GN=Jun PE=1 SV=1

JUN_BOVIN Mass: 36233 Score: 32 Expect: 47 Matches: 23

Transcription factor AP-1 OS=Bos taurus GN=JUN PE=2 SV=2

These results are from the elution from the RNA affinity chromatography experiment with rbz-Dsim-45 as a target and fractionated, pre-cleared RRL as source. The target and source were crosslinked by UV λ 254 prior to capture on the streptavidin matrix.

1. TAL2_MOUSE Mass: 12308 Score: 55 Expect: 0.23 Matches: 16

T-cell acute lymphocytic leukemia protein 2 homolog OS=Mus musculus GN=Tal2 PE=3 SV=1
TAL2_HUMAN Mass: 12284 Score: 38 Expect: 11 Matches: 13

T-cell acute lymphocytic leukemia protein 2 OS=Homo sapiens GN=TAL2 PE=1 SV=1

---

2. KCNJ1_HUMAN Mass: 45279 Score: 44 Expect: 2.9 Matches: 17

ATP-sensitive inward rectifier potassium channel 1 OS=Homo sapiens GN=KCNJ1 PE=1 SV=1

---

3. TDRD3_RAT Mass: 73212 Score: 38 Expect: 9.4 Matches: 23

Tudor domain-containing protein 3 OS=Rattus norvegicus GN=Tdrd3 PE=2 SV=1

---

4. AT11C_MOUSE Mass: 130754 Score: 38 Expect: 10 Matches: 33

Phospholipid-transporting ATPase 11C OS=Mus musculus GN=Atp11c PE=2 SV=2

---

5. COLI_MOUSE Mass: 26975 Score: 38 Expect: 12 Matches: 15

Pro-opiomelanocortin OS=Mus musculus GN=Pomc PE=2 SV=1

---

6. GCSH_HUMAN Mass: 19101 Score: 38 Expect: 12 Matches: 12

Glycine cleavage system H protein, mitochondrial OS=Homo sapiens GN=GCSH PE=1 SV=2

---

7. TDRD3_MOUSE Mass: 82602 Score: 36 Expect: 17 Matches: 24

Tudor domain-containing protein 3 OS=Mus musculus GN=Tdrd3 PE=1 SV=4

---

8. MT21E_BOVIN Mass: 33599 Score: 33 Expect: 36 Matches: 15

Protein-lysine methyltransferase METTL21E OS=Bos taurus GN=METTL21E PE=2 SV=1

---

9. RAB21_MOUSE Mass: 24490 Score: 31 Expect: 52 Matches: 10

Ras-related protein Rab-21 OS=Mus musculus GN=Rab21 PE=1 SV=4
RAB21_RAT Mass: 24547 Score: 31 Expect: 53 Matches: 10

Ras-related protein Rab-21 OS=Rattus norvegicus GN=Rab21 PE=2 SV=1

---

10. TRFE_RABIT Mass: 78901 Score: 31 Expect: 54 Matches: 17

Serotransferrin OS=Oryctolagus cuniculus GN=TF PE=1 SV=4

---

11. CUL2_HUMAN Mass: 87554 Score: 31 Expect: 54 Matches: 24

Cullin-2 OS=Homo sapiens GN=CUL2 PE=1 SV=2
CUL2_PONAB Mass: 87554 Score: 31 Expect: 54 Matches: 24

Cullin-2 OS=Pongo abelii GN=CUL2 PE=2 SV=1

---

12. NEK8_RAT Mass: 76574 Score: 30 Expect: 59 Matches: 20

Serine/threonine-protein kinase Nek8 OS=Rattus norvegicus GN=Nek8 PE=1 SV=1

---

13. CUL2_MOUSE Mass: 87448 Score: 30 Expect: 61 Matches: 24

Cullin-2 OS=Mus musculus GN=Cul2 PE=1 SV=2

---

14. GCSH_RABIT Mass: 19028 Score: 30 Expect: 62 Matches: 12

Glycine cleavage system H protein, mitochondrial OS=Oryctolagus cuniculus GN=GCSH PE=2 SV=1

---

15. TA2R7_PAPHA Mass: 37137 Score: 30 Expect: 62 Matches: 15

Taste receptor type 2 member 7 OS=Papio hamadryas GN=TAS2R7 PE=3 SV=1
TA2R7_MACMU Mass: 37245 Score: 30 Expect: 69 Matches: 15

Taste receptor type 2 member 7 OS=Macaca mulatta GN=TAS2R7 PE=3 SV=1

---

16. CALD1_RAT Mass: 60662 Score: 30 Expect: 68 Matches: 20

Non-muscle caldesmon OS=Rattus norvegicus GN=Cald1 PE=1 SV=1

---

17. IL8_SHEEP Mass: 11570 Score: 30 Expect: 69 Matches: 10

Interleukin-8 OS=Ovis aries GN=CXCL8 PE=3 SV=1

---

18. IF2B_HUMAN Mass: 38706 Score: 30 Expect: 74 Matches: 16

Eukaryotic translation initiation factor 2 subunit 2 OS=Homo sapiens GN=EIF2S2 PE=1 SV=2
IF2B_PONAB Mass: 38678 Score: 30 Expect: 74 Matches: 16

Eukaryotic translation initiation factor 2 subunit 2 OS=Pongo abelii GN=EIF2S2 PE=2 SV=1
IF2B_RABIT Mass: 38644 Score: 30 Expect: 74 Matches: 16

Eukaryotic translation initiation factor 2 subunit 2 OS=Oryctolagus cuniculus GN=EIF2S2 PE=1 SV=1

---

19. COLI_RAT Mass: 26808 Score: 29 Expect: 76 Matches: 13

Pro-opiomelanocortin OS=Rattus norvegicus GN=Pomc PE=1 SV=1

---

20. RPP29_MOUSE Mass: 25849 Score: 27 Expect: 1.4e+02 Matches: 16
Ribonuclease P protein subunit p29 OS=Mus musculus GN=Pop4 PE=2 SV=1

The elution fraction from the RNA affinity chromatography experiment with rbz-Dsim-45 as target was run on a 12 % SDS-PAGE. The three predominant bands (by Coomassie staining) were excised. These results are from the in-gel trypsin digest of the highest band.

1. CXCL7_HUMAN Mass: 14171 Score: 40 Expect: 6.3 Matches: 13

Platelet basic protein OS=Homo sapiens GN=PPBP PE=1 SV=3

---

2. RM23_HUMAN Mass: 17770 Score: 38 Expect: 11 Matches: 12

39S ribosomal protein L23, mitochondrial OS=Homo sapiens GN=MRPL23 PE=1 SV=1

---

3. MTG1_BOVIN Mass: 37380 Score: 37 Expect: 13 Matches: 19

Mitochondrial ribosome-associated GTPase 1 OS=Bos taurus GN=MTG1 PE=2 SV=2

---

4. RSAD2_PIG Mass: 42350 Score: 33 Expect: 34 Matches: 22

Radical S-adenosyl methionine domain-containing protein 2 OS=Sus scrofa GN=RSAD2 PE=2 SV=1

---

5. MIB2_MOUSE Mass: 107719 Score: 32 Expect: 39 Matches: 38

E3 ubiquitin-protein ligase MIB2 OS=Mus musculus GN=Mib2 PE=1 SV=2

---

6. SOMA_BALPH Mass: 24749 Score: 31 Expect: 48 Matches: 7

Somatotropin OS=Balaenoptera physalus GN=GH1 PE=3 SV=1
SOMA_DELDE Mass: 24779 Score: 31 Expect: 48 Matches: 7

Somatotropin OS=Delphinus delphis GN=GH1 PE=3 SV=1
SOMA_HIPAM Mass: 24747 Score: 31 Expect: 48 Matches: 7

Somatotropin OS=Hippopotamus amphibius GN=GH1 PE=3 SV=1

---

7. EF1D_HUMAN Mass: 31217 Score: 31 Expect: 55 Matches: 10

Elongation factor 1-delta OS=Homo sapiens GN=EEF1D PE=1 SV=5
EF1D_MACFA Mass: 31285 Score: 31 Expect: 55 Matches: 10

Elongation factor 1-delta OS=Macaca fascicularis GN=EEF1D PE=2 SV=2

---

8. FGF3_HUMAN Mass: 26984 Score: 30 Expect: 69 Matches: 13

Fibroblast growth factor 3 OS=Homo sapiens GN=FGF3 PE=1 SV=1

---

9. GRAP2_MOUSE Mass: 36844 Score: 30 Expect: 69 Matches: 11

GRB2-related adaptor protein 2 OS=Mus musculus GN=Grap2 PE=1 SV=1

---

10. SOMA_SPAEH Mass: 24897 Score: 29 Expect: 76 Matches: 7

Somatotropin OS=Spalax ehrenbergi GN=GH1 PE=3 SV=1

---

11. ECP_MACNE Mass: 19011 Score: 29 Expect: 80 Matches: 10

Eosinophil cationic protein OS=Macaca nemestrina GN=RNASE3 PE=3 SV=1

12. EF1D_RAT Mass: 31425 Score: 29 Expect: 86 Matches: 12

Elongation factor 1-delta OS=Rattus norvegicus GN=Eef1d PE=1 SV=2

13. NSUN7_MACFA Mass: 80052 Score: 29 Expect: 87 Matches: 21

Putative methyltransferase NSUN7 OS=Macaca fascicularis GN=NSUN7 PE=2 SV=1

14. INP5E_RAT Mass: 72763 Score: 29 Expect: 92 Matches: 28

72 kDa inositol polyphosphate 5-phosphatase OS=Rattus norvegicus GN=Inpp5e PE=1 SV=2

15. MYP0_HORSE Mass: 27638 Score: 28 Expect: 98 Matches: 11

Myelin protein P0 OS=Equus caballus GN=MPZ PE=2 SV=1

The elution fraction from the RNA affinity chromatography experiment with rbz-Dsim-45 as target was run on a 12 % SDS-PAGE. The three predominant bands (by Coomassie staining) were excised. These results are from the in-gel trypsin digest of the middle band.

1. U2AF1_BOVIN Mass: 28197 Score: 42 Expect: 4.3 Matches: 18

Splicing factor U2AF 35 kDa subunit OS=Bos taurus GN=U2AF1 PE=2 SV=1
U2AF1_MOUSE Mass: 28311 Score: 29 Expect: 76 Matches: 16

Splicing factor U2AF 35 kDa subunit OS=Mus musculus GN=U2af1 PE=1 SV=4
U2AF1_HUMAN Mass: 28368 Score: 29 Expect: 76 Matches: 16

Splicing factor U2AF 35 kDa subunit OS=Homo sapiens GN=U2AF1 PE=1 SV=3

2. SCP2D_BOVIN Mass: 17477 Score: 37 Expect: 12 Matches: 14

SCP2 sterol-binding domain-containing protein 1 OS=Bos taurus GN=SCP2D1 PE=2 SV=1

3. FGF16_MOUSE Mass: 23877 Score: 34 Expect: 27 Matches: 15

Fibroblast growth factor 16 OS=Mus musculus GN=Fgf16 PE=2 SV=2
FGF16_HUMAN Mass: 23858 Score: 28 Expect: 1.1e+02 Matches: 13
Fibroblast growth factor 16 OS=Homo sapiens GN=FGF16 PE=1 SV=1
FGF16_RAT Mass: 23852 Score: 27 Expect: 1.3e+02 Matches: 14

Fibroblast growth factor 16 OS=Rattus norvegicus GN=Fgf16 PE=2 SV=1

4. MTO1_MOUSE Mass: 74855 Score: 33 Expect: 31 Matches: 23

Protein MTO1 homolog, mitochondrial OS=Mus musculus GN=Mto1 PE=2 SV=1

5. THAP2_HUMAN Mass: 26699 Score: 33 Expect: 31 Matches: 22

THAP domain-containing protein 2 OS=Homo sapiens GN=THAP2 PE=1 SV=1

6. ARF_RAT Mass: 17467 Score: 33 Expect: 36 Matches: 14

Tumor suppressor ARF OS=Rattus norvegicus GN=Cdkn2a PE=2 SV=1

7. KAPCG_HUMAN Mass: 40580 Score: 32 Expect: 46 Matches: 18

cAMP-dependent protein kinase catalytic subunit gamma OS=Homo sapiens GN=PRKACG PE=1 SV=3

8. CD4_RAT Mass: 52089 Score: 30 Expect: 71 Matches: 20

T-cell surface glycoprotein CD4 OS=Rattus norvegicus GN=Cd4 PE=1 SV=1

9. ANR46_HUMAN Mass: 26178 Score: 30 Expect: 73 Matches: 9

Ankyrin repeat domain-containing protein 46 OS=Homo sapiens GN=ANKRD46 PE=1 SV=1

10. MTO1_MACFA Mass: 77953 Score: 30 Expect: 73 Matches: 22

Protein MTO1 homolog, mitochondrial OS=Macaca fascicularis GN=MTO1 PE=2 SV=1

11. MTO1_PONAB Mass: 77936 Score: 29 Expect: 76 Matches: 23

Protein MTO1 homolog, mitochondrial OS=Pongo abelii GN=MTO1 PE=2 SV=1

12. CLC1A_BOVIN Mass: 32700 Score: 29 Expect: 84 Matches: 17

C-type lectin domain family 1 member A OS=Bos taurus GN=CLEC1A PE=2 SV=1

13. TMM27_RAT Mass: 25320 Score: 29 Expect: 86 Matches: 12

Collectrin OS=Rattus norvegicus GN=Tmem27 PE=1 SV=2

The elution fraction from the RNA affinity chromatography experiment with rbz-Dsim-45 as target was run on a 12 % SDS-PAGE. The three predominant bands (by Coomassie staining) were excised. These results are from the in-gel trypsin digest of the lowest band.

1. SYT8_RAT Mass: 44579 Score: 45 Expect: 2.2 Matches: 20

Synaptotagmin-8 OS=Rattus norvegicus GN=Syt8 PE=2 SV=1

2. NMRL1_HUMAN Mass: 33494 Score: 32 Expect: 40 Matches: 20

NmrA-like family domain-containing protein 1 OS=Homo sapiens GN=NMRAL1 PE=1 SV=1

3. RL27A_MACFA Mass: 16663 Score: 30 Expect: 71 Matches: 16

60S ribosomal protein L27a OS=Macaca fascicularis GN=RPL27A PE=2 SV=1

4. RL27A_HUMAN Mass: 16665 Score: 29 Expect: 80 Matches: 15

60S ribosomal protein L27a OS=Homo sapiens GN=RPL27A PE=1 SV=2

RL27A_PANTR Mass: 16651 Score: 29 Expect: 80 Matches: 15

60S ribosomal protein L27a OS=Pan troglodytes GN=RPL27A PE=2 SV=3
RL27A_PONAB Mass: 16665 Score: 29 Expect: 80 Matches: 15

60S ribosomal protein L27a OS=Pongo abelii GN=RPL27A PE=2 SV=3

---

5. ZN704_MOUSE Mass: 61626 Score: 29 Expect: 87 Matches: 20

Zinc finger protein 704 OS=Mus musculus GN=Znf704 PE=2 SV=1

---

6. STP2_PONPY Mass: 15918 Score: 28 Expect: 1.1e+02 Matches: 14

Nuclear transition protein 2 (Fragment) OS=Pongo pygmaeus GN=TNP2 PE=2 SV=1

---

7. RCAS1_HUMAN Mass: 24533 Score: 27 Expect: 1.2e+02 Matches: 15

Receptor-binding cancer antigen expressed on SiSo cells OS=Homo sapiens GN=EBAG9 PE=1 SV=1
RCAS1_MOUSE Mass: 24475 Score: 27 Expect: 1.4e+02 Matches: 15

Receptor-binding cancer antigen expressed on SiSo cells OS=Mus musculus GN=Ebag9 PE=1 SV=2

---

8. HRH1_CAVPO Mass: 56325 Score: 27 Expect: 1.3e+02 Matches: 17

Histamine H1 receptor OS=Cavia porcellus GN=HRH1 PE=3 SV=1

---

9. SH3L3_PONAB Mass: 10531 Score: 27 Expect: 1.3e+02 Matches: 8

SH3 domain-binding glutamic acid-rich-like protein 3 OS=Pongo abelii GN=SH3BGRL3 PE=3 SV=1

---

10. PP14C_MOUSE Mass: 17857 Score: 26 Expect: 1.5e+02 Matches: 11

Protein phosphatase 1 regulatory subunit 14C OS=Mus musculus GN=Ppp1r14c PE=1 SV=1
PP14C_HUMAN Mass: 17946 Score: 25 Expect: 2e+02 Matches: 11

Protein phosphatase 1 regulatory subunit 14C OS=Homo sapiens GN=PPP1R14C PE=1 SV=3
PP14C_RAT Mass: 18083 Score: 24 Expect: 2.9e+02 Matches: 11

Protein phosphatase 1 regulatory subunit 14C OS=Rattus norvegicus GN=Ppp1r14c PE=1 SV=1

---

11. RCAS1_RAT Mass: 24343 Score: 26 Expect: 1.6e+02 Matches: 14

Receptor-binding cancer antigen expressed on SiSo cells OS=Rattus norvegicus GN=Ebag9 PE=1 SV=1

---

12. STP2_MACMU Mass: 15690 Score: 26 Expect: 1.6e+02 Matches: 15

Nuclear transition protein 2 (Fragment) OS=Macaca mulatta GN=TNP2 PE=2 SV=1
STP2_MACFA Mass: 16109 Score: 24 Expect: 2.6e+02 Matches: 14

Nuclear transition protein 2 OS=Macaca fascicularis GN=TNP2 PE=2 SV=2

LC-MS/MS data analysis (Dr. Paul Gershon)

The following results are for PD1 from the RNA affinity chromatography experiment with rbz-Dsim-45 as the target.

hemoglobin

actin

trypsin

tropomyosin

flaggrin

transferrin

tubulin

serotransferrin

The following results are for the elution from the RNA affinity chromatography experiment with rbz-Dsim-45 as the target.

hemoglobin

trypsin

actin

serotransferrin

streptavidin

alpha 2-HS glycoprotein precursor

apolipoprotein B

bovine serum albumin

coagulation factor B

glucose 6-phosphate isomerase

alpha-1 antiproteinase precursor

tetranectin