

1-1-2010

Characterization of Novel Functions and Topologies in RNA

Aaron Steven Burton
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds

Let us know how access to this document benefits you.

Recommended Citation

Burton, Aaron Steven, "Characterization of Novel Functions and Topologies in RNA" (2010). *Dissertations and Theses*. Paper 363.

<https://doi.org/10.15760/etd.363>

This Dissertation is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

Characterization of Novel Functions and Topologies in RNA

by

Aaron Steven Burton

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Chemistry

Dissertation Committee:
Niles Lehman, Chair
Albert S. Benight
Dirk Iwata-Reuyl
Kenneth M. Stedman
Gary Brodowicz

Portland State University
2010

Abstract

The RNA World hypothesis describes a period of time during the origins of life in which RNA molecules performed all catalysis and were the only form of information storage. A great deal of evidence has been obtained in support of this hypothesis, however a few key demonstrations are lacking. The first demonstration is of a molecule capable of self-replication that could have plausibly arisen from the prebiotic soup. Previously in the Lehman Laboratory, a 198-nucleotide RNA was discovered that could be fragmented into as many as four pieces ranging from 39 – 63 nucleotides in length. When these pieces were incubated together in a test tube, they re-formed the necessary covalent bonds to regenerate the full-length 198-nucleotide RNA. Furthermore, the full-length RNAs were catalytically active and made copies of themselves from the remaining pieces in solution, providing a model system of self-replication. I was able to remove >10% of the total length of the RNA, which substantially reduced the catalytic activity of the full-length molecule. I discovered several mutations that restored catalytic activity by improved folding and increased catalytic rates using *in vitro* selection. A subset of these mutations was found to aid in the assembly of the shortened full-length RNA from smaller fragments than were possible in the original system, enhancing the prebiotic relevance of this system. A second demonstration to bolster the RNA World hypothesis would be showing that RNA is capable of harvesting energy from its environment by performing oxidation

and reduction reactions. Again using *in vitro* selection, I have completed five rounds of selection geared towards identifying a ribozyme that reduces benzoic acid to benzaldehyde using Zn^{2+} and NADH. Results to date suggest the selection is working and it should be continued for another five to ten generations. Finally, I have discovered an RNA sequence that forms knots during transcription, a phenomenon heretofore undocumented in RNA. This new topology has implications for RNA stability by rendering RNA more resistant to hydrolysis, and could impact catalysis through formation of more complex, knotted active sites. Taken together, these findings have improved our understanding of RNA folding and catalysis, and the plausibility of the RNA World.

Dedication

To my parents, Roberta and Steven Burton, who taught me to value hard work, humility, and critical thinking, all of which were necessary to persevere through graduate school.

Acknowledgments

Dr. Niles Lehman

Committee members:

Dr. Dirk Iwata-Reuyl

Dr. Albert Benight

Dr. Kenneth Stedman

Dr. Gary Brodowicz

Lehman Lab members past and present

Sinnamon Tierney for help with my NASA Fellowship

the Department of Chemistry at Portland State University

my family and friends

my beautiful and loving wife Kyra

Table of Contents

Abstract.....	i
Dedication.....	iii
Acknowledgments	iv
List of Tables	vi
List of Figures.....	vii
Chapter 1. Introduction.....	1
Chapter 2. Enhancing the Prebiotic Relevance of a Set of Covalently Self-Assembling, Autorecombining RNAs Through <i>in Vitro</i> Selection.....	13
Chapter 3. <i>In Vitro</i> Selection of a Benzoic Acid Reductase Ribozyme.	36
Chapter 4. Knotted RNAs Are Formed During <i>in Vitro</i> Transcription of the <i>Chlorella</i> Virus Pyrimidine Dimer-Specific Glycosylase Intron.	56
References	91

List of Tables

Table 1. Catalytic Parameters of <i>Azoarcus</i> Ribozyme Variants.....	30
Table 2. Sequences of molecules tested for knRNA formation.....	76

List of Figures

Figure 1. Pathways leading from the RNA World to the RNA-DNA-Protein World.	11
Figure 2. Plausible metabolic routes to DNA and RNA from abiotic precursors..	12
Figure 3. Secondary structures of the <i>Azoarcus</i> ribozyme autorecombinase..	31
Figure 4. Reverse-splicing activity of <i>Azoarcus</i> ribozyme variants..	32
Figure 5. Genotyping of the generation 10 selection population.	33
Figure 6. Mutation frequency in the G10 population.	34
Figure 7. Effects of selection-derived mutations on the self-construction activity of the <i>Azoarcus</i> Δ P6aRC ribozyme.	35
Figure 8. Balanced equation for the reduction of benzoic acid to benzaldehyde.	51
Figure 9. <i>In vitro</i> selection procedure used to select for a benzoic acid reductase ribozyme.	52
Figure 10. Reduction assay of benzoic acid reductase ability for the G ₀ and G ₅ populations..	53
Figure 11. Aldehyde-reactive molecules used for <i>in vitro</i> selection (biotin hydrazide) or the assay of benzoic acid reduction ability (FTC).	54
Figure 12. Assay of streptavidin binding by the G ₀ and G ₅ RNA populations.	55
Figure 13. Distinction between a pseudoknot (left) and an open knot (right).	77
Figure 14. <i>Chlorella</i> virus <i>pdg</i> intron location and structure.	78
Figure 15. Formation of topological isomers (knRNAs) during <i>in vitro</i> transcription of the <i>Chlorella</i> virus <i>pdg</i> intron.	79
Figure 16. Possible topological explanations for the mobility shifts observed in the knRNAs.	81
Figure 17. Gel electrophoresis of glyoxylal-treated linear and knRNAs alongside their non-glyoxylated counterparts.	82
Figure 18. Denaturing gels of knRNAs with formamide.	83
Figure 19. Debranching assay of knRNAs.	84
Figure 20. Alkaline hydrolysis of knRNAs to check for base-labile bonds.	85
Figure 21. Acid hydrolysis of knRNAs to check for the presence of acid-labile bonds.	86
Figure 22. Unknotting versus denaturation of knRNAs.	87
Figure 23. Proposed folding pathways leading to knotting by pseudoknot formation.	88

Figure 24. The 95-nucleotide intron found in the *Tetrahymena thermophila* *cnjB* locus is susceptible to knotting. 89

Figure 25. Effects of 3'-end mutations on knRNA formation. 90

Chapter 1. Introduction.

Note: This chapter is adapted from Burton and Lehman (2009).

The RNA World hypothesis postulates that there was a period during the origins of life on Earth when RNA molecules encoded all of the genetic information and performed all the catalytic functions necessary for a ribo-organism to replicate and evolve (Rich 1962; Woese 1967; Orgel 1968; Crick 1968; Kuhn 1972; Gilbert 1986; Orgel 1986; Joyce 1991, 1998, 2002). This idea is supported by several lines of evidence. The first is that many of the cofactors that modern protein enzymes use for catalysis are derived from RNA nucleotides, including but not limited to NADH, FADH₂, coenzyme A, and S-adenosylmethionine (White 1976; Jeffares *et al.* 1998). The idea is that RNA-based metabolism evolved to where RNA enzymes (**ribozymes**) depended on these cofactors, such that when a given protein evolved to replace a specific ribozyme the protein had to retain the nucleotide-based cofactor employed by each ribozyme. The second line of evidence is that the catalytic portion of the protein making machinery, the ribosome, is composed of RNA (Ban *et al.* 2000; Nissen *et al.* 2000). In the words of Thomas Cech (2000), “The ribosome is a ribozyme.” Given that ribozymes are generally much poorer catalysts than are proteins, there is no evolutionary reason for the ribosome to be made of RNA unless ribozymes preceded the use of protein as a catalyst. A third line of evidence for the RNA world pertains to

the use of RNA polymers for information storage prior to the advent of DNA. In modern biology, DNA nucleotides are made from RNA nucleotides immediately prior to DNA replication, and DNA polymerization is initiated from RNA primers. Because RNA polymers are less stable than their DNA counterparts, these observations support the conclusion that information was indeed stored in RNA before DNA, and that DNA evolved as a superior method for information storage, although some viruses still have genomes composed only of RNA.

While the RNA World is generally accepted as a period during the origins of life, our understanding of specific reactions and interactions that would have been necessary for metabolism and replication during this period is incomplete. The work contained in this thesis supports the RNA World hypothesis by enhancing the prebiotic relevance of a self-replicating set of RNAs (Chapter 2), by demonstrating that ribozymes can catalyze difficult oxidation / reduction reactions important for metabolism (Chapter 3), and by the discovery that RNA molecules can fold into knotted conformations that impact RNA stability, and may have implications for catalysis and information transfer (Chapter 4). I have also explored the nature of evolution from the RNA World to the modern system where DNA, RNA, and proteins are present (Burton & Lehman 2009). This is discussed below.

An interesting but unresolved question stemming from the RNA World hypothesis is the order in which proteins and DNA arose (Figure 1). Both transitions would be beneficial to the viability of an organism—proteins catalyze offer greater reaction rates and specificities, while the increased stability of a DNA genome (relative to one made of RNA) expands the maximum possible genome size of said organism..

A key argument against DNA preceding protein revolves around ribonucleotide reduction (Freeland *et al.* 1999; Poole *et al.* 1999). The chemistry performed by ribonucleotide reductase proteins invariably involves a free-radical mechanism (Reichard 1993; Sun *et al.* 1993), which ribozymes have not yet been shown to catalyze. Furthermore, these radicals are manipulated by sulfhydryl groups, a functionality to which naturally occurring RNAs would have no access unless sulfur-containing cofactors were employed. However, ribozymes identified in nature and through *in vitro* selection have demonstrated a diverse catalytic repertoire, including redox chemistry (Tsukiji *et al.* 2003, 2004), so it seems premature to declare RNA incapable of ribonucleotide reduction. In fact, it may prove that devising an effective *in vitro* selection scheme for a ribonucleotide reductase ribozyme is more difficult than the mechanism such a ribozyme ends up catalyzing. Moreover, it should not be assumed that a ribozyme would be required to utilize the same mechanism as protein ribonucleotide reductases; for example, alkane hydroxylation (essentially the reverse of ribonucleotide reduction), as catalyzed by methane monooxygenase protein

enzymes, proceeds via a concerted two-step mechanism without the generation of a free radical (Yoshizawa 1998). This raises the possibility that a ribozyme could perform ribonucleotide reduction through a relatively simple radical-free mechanism. More importantly, the enzymatic route to DNA nucleotides need not have begun with ribonucleotide reduction. The protein enzyme D-2-deoxyribose-5-phosphate aldolase catalyzes the formation of this product from acetaldehyde and D-glyceraldehyde-3-phosphate by the aldol condensation reaction (Heine *et al.* 2001). Because RNA can also catalyze aldol condensation chemistry (Fusz *et al.* 2005) as well as phosphate transfer (Lorsch & Szostak 1994), and acetaldehyde and D-glyceraldehyde can be formed abiotically, aldol reactions present an alternative route for the enzymatic synthesis of deoxyribose prior to the evolution of a ribonucleotide reductase. Potentially, the synthesis of deoxyribonucleotides could share many of the catalytic steps (and enzymes, perhaps) needed for the production of ribonucleotides (Figure 2). It has already been demonstrated that, within the context of double-stranded DNA, 2'-5'-phosphoester-linked ribose can be made as a consequence of the aldol condensation of phosphoglyceryl- and phosphoglycoaldehyde-terminated oligonucleotides (Oberhuber & Joyce 2005). It is likely that RNA can catalyze this reaction as well. Artificial ribozymes have been discovered that are capable of attaching purine and pyrimidine bases to activated ribose (Unrau & Bartel 1998; Lau *et al.* 2004; Lau & Unrau 2009), so all that remains is to confirm the ability of RNA to catalyze DNA-forming aldol reactions, find a ribozyme for activating ribose and deoxyribose for

nucleobase addition (which presumably had to exist to make ribonucleotides), and verify that nucleotide synthase ribozymes can act on deoxyribose. While not trivial, these feats seem well within the catalytic repertoire of RNA, given the diversity of ribozymes discovered to date. Thus, at least two discrete chemical paths exist for the RNA-catalyzed formation of DNA: ribonucleotide reduction and synthesis from small molecules.

It has been noted that deoxyribose is unlikely to have been sufficiently abiotically available to have arisen prior to its enzymatic synthesis. It should be remembered that this same argument is made for ribose. Recent research has presented myriad abiotic routes from formaldehyde and its oligomers to nucleosides (Powner *et al.* 2009). Analogous routes could potentially exist for increased production of deoxyribonucleosides relative to other molecules, which simply have not yet been found. Indeed, having one less chiral center should reduce stereochemical complexity. The question is, regardless of whether its source is abiotic or RNA-catalyzed, how much DNA does an organism need? Cells generally have only a single copy of their genome in DNA. When expression occurs, multiple copies of RNA are made, and multiple proteins can be made per mRNA template. In a putative RNA-DNA organism that consists of one DNA copy of each gene and several RNA gene products per DNA gene, the deoxyribonucleotide demand would have been much smaller than that of their ribonucleotide counterparts, which must have been sufficiently available for the

RNA World to have been a reality. Adding in the greater stability of DNA relative to RNA, the demand for deoxyribonucleotides was reduced even further. Thus, the required efficiency of DNA synthesis might have been significantly less than that needed for RNA production.

It has already been demonstrated that ribozymes are capable of acting on DNA. Group II introns have been demonstrated to be capable of DNA “polymerization,” generating longer DNA oligos via successive transesterification reactions, as well as DNA ligation (Hetzer *et al.* 1997). In addition, a minimal number of mutations allow RNA enzymes to perform catalysis on either RNA or DNA (Beaudry & Joyce 1992). It is conceivable then that existing RNA-acting ribozymes could have evolved to catalyze reactions on DNA substrates. Transitioning to an RNA-DNA World could have occurred with the development of as few as three ribozymes in the all-RNA World: (1) a DNA-producing catalyst; (2) an RNA-dependent DNA polymerase to convert the RNA genomic molecules into a more permanent storage form; and (3) a DNA-dependent RNA polymerase ribozyme to enable gene expression. The latter two, at least, would have had significant evolutionary head starts in RNA-dependent RNA polymerase ribozymes, and all three would have head starts if aldolase catalysts were responsible for ribose production. Thus, the advent of DNA was within evolutionary reach of the RNA World.

The advent of protein synthesis by translation was one of the most challenging evolutionary transitions in the history of life. In its simplest modern form, the ribosome itself has three or more rRNAs and >60 proteins, ~40 tRNAs are charged with amino acids by 18–20 amino-acyl tRNA synthetases, and 7–8 translation factors are required; thus a total of ~130 discrete genes are needed for translation (Wolf & Koonin 2007). A minimal translation system would require at least one ribosome ribozyme to catalyze peptide bond formation, and one tRNA (or equivalent structure for amino acid recognition) per amino acid, assuming all tRNAs were self-charging. Even with a primitive version of the genetic code that used 5–10 amino acids, this minimal set would have required at least 6–11 genes, without even considering the protein-coding genes themselves. It has been hypothesized that the minimum number of nucleotides a ribo-organism would need in its genome was at least 10,000–15,000 prior to the advent of translation (Jeffares *et al.* 1998). Estimations of maximum genome size are based on polymerase fidelity. Because ribozymes can tolerate some mutations as truly neutral, they may not strictly fit the Eigen (1971) model where every mutation from the master sequence is considered deleterious (Kun *et al.* 2005); (Takeuchi *et al.* 2005). The possibility remains, however, that the riboorganism genome had reached the maximum allowed by the fidelity of RNA polymerase ribozymes (Johnston *et al.* 2001; Zaher & Unrau 2007). Space constraints would further complicate the evolution of proteins, as translation-related genes would have to replace only non-essential genes. Any increase in maximum genome size would

certainly make this evolutionary step easier. One solution to genome size constraints could be achieved through the development of error correction in RNA polymerase ribozymes, but ascertaining whether RNA can perform this function will be a formidable task. Alternatively, information could be stored in a more stable genomic polymer, which would also increase the genomic length maximum. For example, Poole *et al.* (2000) proposed a model by which RNA genomes could be stabilized through methylation of certain 2'-hydroxyls. However, any such methylation would have had to occur shortly after the molecule was synthesized, as the hydroxyls most in need of protection, by definition, would be the ones that promoted the fastest self-cleavage. Polymers of DNA are not faced with this same pressure and confer additional advantages to a primitive organism. Early RNA genomes were likely double stranded, as this duplication of information allows for longer informational polymers; it is almost certain, then, that DNA-containing genomes were double stranded as well. The most advantageous way of incorporating DNA would have been in duplexes composed of one strand of RNA and one of DNA. In addition to the improved stability of DNA relative to RNA, the lack of 2'-OHs in the template strand clearly identifies it as such, which allows the information in that strand to be selectively used for gene expression. Prior to DNA, this discrimination could have been based on primitive sequence recognition or perhaps never occurred. This affords a side benefit of resource conservation, as unnecessary genome copies would not be made during the course of normal gene expression. The reduced number of hydroxyls

also limits the complexity of structures that DNA can adopt, making it easier to keep the DNA in a double-stranded state and, thus, easier from which to extract information. Another improvement in fidelity granted by DNA stems from the spontaneous deamination of cytidine to uridine. If this occurs in the context of dsRNA, it is impossible to know whether the strand containing the guanosine originally paired to the C or the U resulting from the deamination is correct. Methylation of uridines in ribozymes would add steric hindrance, which would compromise catalytic activity but not affect the structure of typical double-stranded helices. Thus, the advent of DNA allows increased genome size by exploiting multiple facets of subtle chemical differences from RNA. The extra genomic space and stability granted by this transition would allow the more difficult evolution of translation to occur.

The RNA-DNA World was, at the very least, a possible stage on the road to life, a possibility largely overlooked because of the difficulty of ribonucleotide reduction. The 2'-OH of RNA is clearly a two-edged sword—while facilitating the adoption of catalytically active tertiary structures, it also renders RNA much more susceptible to hydrolysis. Chemistries newly observed in nucleic acid catalysts have outlined a plausible path by which DNA could be made by RNA, with no requirement for amino acid cofactors. Employing DNA makes gene expression more efficient, as it is now clear which molecules are genotypes and which are phenotypes. In addition, the increased stability of DNA polymers relative to RNA improves the lifespan of a

genomic molecule and allows for larger overall genome size. The ability to store more genetic information allows for the inclusion of more genes, an important advance for an organism operating at or near its Eigen error threshold; indeed, this ability may have been essential for the development of translation and protein-coding genes. The RNA-DNA World model presented here will be bolstered by the demonstration of a few catalytic feats: showing that the aldolase ribozyme can assemble deoxyribose, confirming that nucleotide synthase ribozymes can attach bases to deoxyribose, and demonstrating that in ribozyme catalysis RNA can serve as a template for DNA synthesis or vice versa. Because all these demonstrations begin with known ribozymes, however, they do not appear to be intractable. A plausible scenario for the ancient relationships among RNA, DNA, and proteins may be within reach.

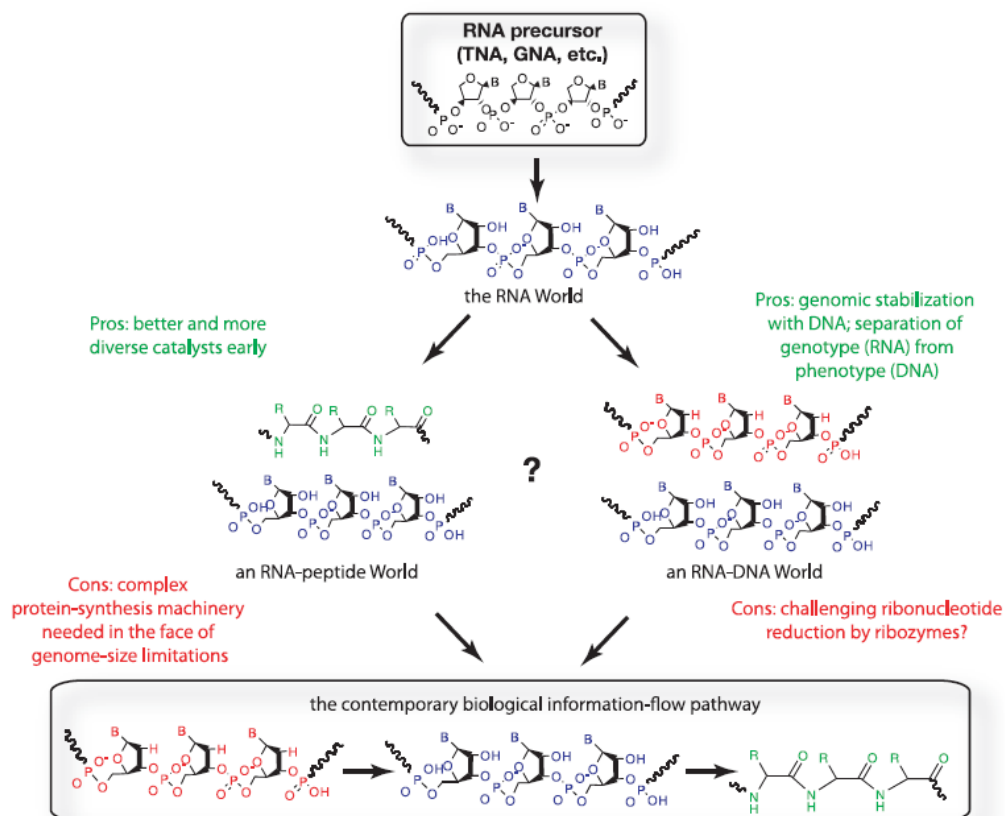


Figure 1. Pathways leading from the RNA World to the RNA-DNA-Protein World. Alternative pathways leading from the first information-rich macromolecule, such as glycerol nucleic acid (GNA) or threose nucleic acid (TNA), to present-day information flow. The order of appearance of DNA and proteins is currently unresolved.

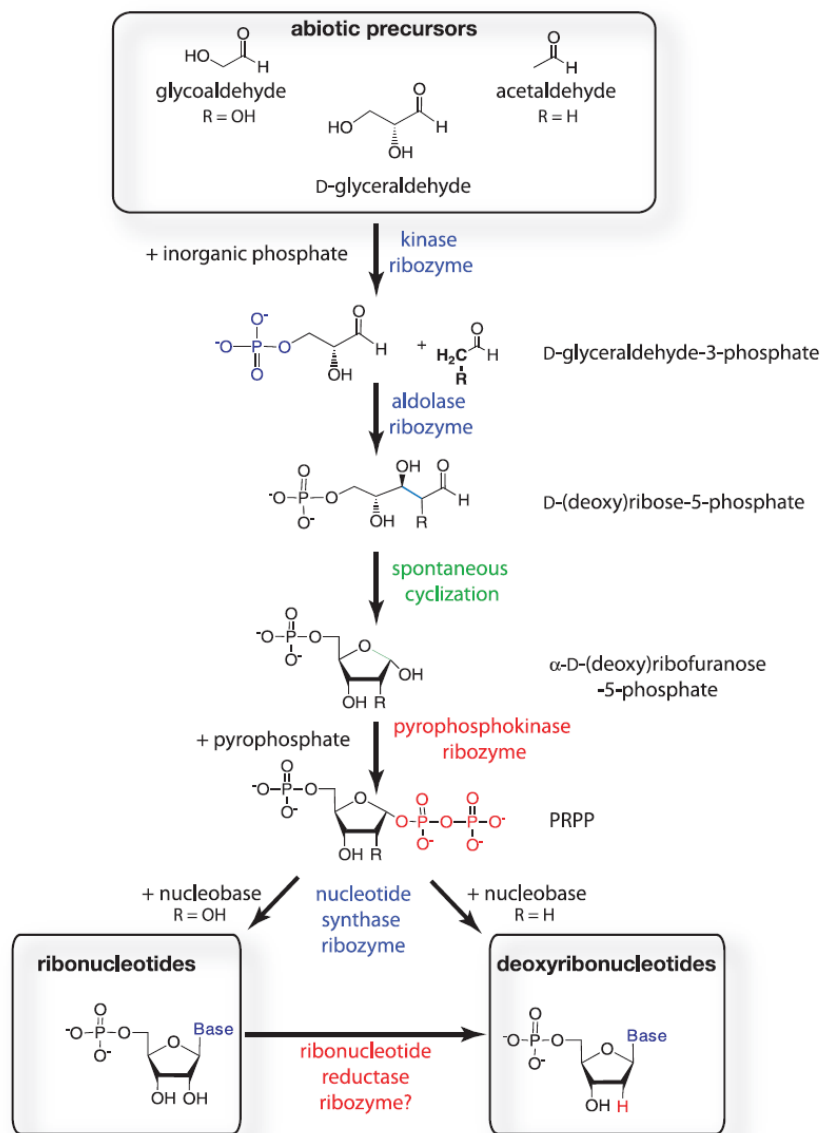


Figure 2. Plausible metabolic routes to DNA and RNA from abiotic precursors. Ribonucleotides and deoxyribonucleotides can be synthesized by the same chemistries and, possibly, the same enzymes simply by changing the initial abiotic molecules. The kinase, aldolase, and nucleotide synthase ribozymes have already been discovered; the ribose cyclization does not necessarily require a catalyst; the pyrophosphate kinase and ribonucleotide reductase ribozymes have yet to be discovered.

Chapter 2. Enhancing the Prebiotic Relevance of a Set of Covalently Self-Assembling, Autorecombining RNAs Through *in Vitro* Selection.

Note: this work is adapted from Burton & Lehman (2010).

Abstract

An *in vitro* form of the self-splicing group I intron interrupting the *Azoarcus* tRNA^{Ile} was shortened by ~10% with the removal of helix P6a. This deletion reduced the reverse-splicing activity of the ribozyme about ten-fold. Through *in vitro* selection, this activity was restored in several low-error mutants. A number of mutations were found that improved reverse-splicing activity through both increased k_{obs} and better folding. The deletion mutant could be fragmented into as many as three discrete pieces, which, when incubated together, were capable of covalent self-assembly through energy-neutral transesterification reactions, a process called autorecombination. A subset of the mutations identified through *in vitro* selection for reverse-splicing were exaptations in that they were also shown to augment the autorecombination reactions, leading to higher yields of covalently self-assembled products, making this the smallest such system yet discovered.

Introduction

The RNA World hypothesis describes a time during the origins of life, prior to the advent of DNA and proteins, when RNA served as both catalyst and informational

molecule. A great deal of evidence has been obtained in support of this hypothesis (Rich 1962; Kuhn 1972; Orgel 1986; Joyce 1991, 1998, 2002; Szostak & Ellington 1993; Gesteland *et al.* 2006; Boussau *et al.* 2008). Current research is focused on finding self-replicating RNA systems. A number of advances have been made towards an RNA replicase molecule based on RNA-dependent RNA polymerization (Johnston *et al.* 2001; McGinness & Joyce 2003; Zaher & Unrau 2007). Unfortunately, the polymerase ribozymes discovered to date are able to copy less than 1/4 of their total length. In addition, the molecules created by these polymerases are complementary copies of their targets, meaning that two complete copies have to be made for each replication event. These limitations have led to the exploration of other ribozyme chemistries for self-replicating RNA systems, including simple *trans*-esterification, which can be used to build up longer RNA molecules through recombination (Gilbert 1986; Burke & Willis 1998; Lehman 2003; Striggles *et al.* 2006).

Group I introns are a class of structurally conserved RNAs that catalyze their own excision from nascent transcripts. These self-splicing ribozymes ligate the exons they interrupt *via* consecutive *trans*-esterification reactions, the first employing exogenous guanosine as a nucleophile to free the 5'-exon, followed by attack of the 3'-hydroxyl of the now-free 5'-exon on the 3'-splice site. The two *trans*-esterification steps have been extensively studied for several group I introns. Because the *trans*-esterification reactions swap one 3'-5' phosphoester bond for another, they are essentially energy-

neutral, and therefore, readily reversible. The reverse of the second step of splicing reaction, ligation of the 3'-exon to the 3'-end of the ribozyme, has been exploited for *in vitro* selection of ribozymes with altered catalytic properties, such as the ability to use Ca^{2+} ions instead of Mg^{2+} ions for catalysis (Lehman & Joyce 1993; Burton & Lehman 2006).

The reversibility of the second step of splicing has also been used to achieve the recombination of exogenous substrates. To date, the *Tetrahymena*, *Azoarcus*, *Candida*, and *Pneumocystis* ribozymes have been shown to catalyze recombination *in vitro* (Bell *et al.* 2002; Riley & Lehman 2003; Dotson *et al.* 2008). The *Azoarcus* ribozyme specifically has demonstrated the ability to make products ranging from essentially non-structured, short RNAs to a ~200-nucleotide group I intron, from as many as four smaller substrate pieces (Riley & Lehman 2003; Hayden *et al.* 2005; Hayden & Lehman 2006). Furthermore, the ribozyme itself can be fragmented into as many as four pieces ranging from 39–63 nucleotides that are able to generate full-length, covalent ribozymes by performing recombination reactions as a non-covalent *trans* assembly. The recombinant ribozymes are active in the milieu in which they are made, and generate additional full-length ribozymes by recombining the remaining non-covalently bonded pieces of themselves (**autorecombination**). Because these autorecombinases make identical copies of themselves, they actually perform autoreplication.

The aforementioned system provides an energy-neutral mechanism for the build-up of genetic information from smaller fragments and for molecular replication (Lehman 2008). Formation of the substrates for recombination, however, almost certainly required some sort of activating agent, perhaps as nucleotide triphosphates, imidazoles, or phosphoramidates. Although it has been demonstrated that RNA polymers up to 50–60 nucleotides long can be generated by polymerization of activated monomers on clay minerals, the bulk of the products generated were 20–30 nucleotides in length (Ferris 2002). Clearly, then, smaller would have been better for any self-replicating system, both from an availability-of-materials standpoint and requirements for replication fidelity in light of Eigen's paradox (1971): no enzyme without a large genome, and no large genome without enzymes. To that effect, we have removed about 10% of the length of the autorecombinase form of the *Azoarcus* group I intron (Figure 3). This decreased the *trans*-splicing activity of the ribozyme by an order of magnitude, which we restored through *in vitro* selection. We were thus able to identify and characterize truncated *Azoarcus* ribozyme mutants capable of covalent self-assembly and autoreplication by autorecombination.

Results and Discussion

Evaluation of potential deletion sites for a shorter autorecombinase

We considered several deletion sites to reduce the size of the autorecombining form of the *Azoarcus* ribozyme (Figure 3A). Two constructs contained deletions in P9, the first

removing nucleotides 181, 182, and 197–201 ($\Delta P9_{-7}$); the second removing nucleotides 181–187 and 192–201 ($\Delta P9_{-17}$). Because the tetraloop in L9 is involved in a tertiary interaction with P4, we expected both of these deletions to reduce the reverse-splicing activity of the ribozyme, which they did (Figure 4). Unexpectedly, the $\Delta P9_{-17}$ form retained more activity than the $\Delta P9_{-7}$ form despite containing all of the deletions in $\Delta P9_{-7}$. It may be that removing the bulged region of P9 still allows the L9-P4 tetraloop-receptor interaction to take place but alters the overall conformation of the ribozyme, while the further deletions in $\Delta P9_{-17}$ spare the correct global fold at the expense of the L9-L4 interaction. We also made a construct with nucleotides 98–118 deleted ($\Delta P6aRC$, Figure 3B) that also resulted in a ribozyme with reduced activity. Because the $\Delta P9$ forms were less active than the $\Delta P6aRC$ version and formation of the *trans* recombination complex is likely to be heavily dependent on the L9-P4 tetraloop interaction, we chose the $\Delta P6aRC$ form as the starting point for the smaller autorecombinase.

In vitro selection to improve catalytic activity

The four-piece self-construction reactions demonstrated previously (Hayden & Niles Lehman 2006) require that the RNA be separated into fragments that contain the complement (5'-CAU-3') to the internal guide sequence of the *Azoarcus* ribozyme (5'-GUG-3'). When we placed the CAUs in L5, L6 and L8 of the $\Delta P6aRC$ ribozyme to make the autorecombining form $\Delta P6aRC$ (Figure 3B), the reverse-splicing activity of

the covalent ribozyme was reduced roughly seven-fold (Figure 4). As a means to restore this activity, we generated a pool of ribozymes mutagenized at 10% per position by PCR and selected for reverse splicing ability under optimal autorecombination conditions (48°C in 100 mM MgCl₂ and 30 mM EPPS (pH 7.5), keeping selection pressure constant by allowing 15 minutes for folding and reaction for every round. We started a more stringent selection line, based on the generation 6 population, that was incubated at 48°C for just 2 min, but this line was unable to accommodate the selection criteria.

Genotyping of selection population

After ten rounds of selection, we cloned and determined the sequence of 23 individuals from the G₁₀ population (Figure 5). Six genotypes were identical to that of the initial Δ P6aRC ribozyme, which was not unexpected given that it possesses some catalytic activity. Yet two mutations were observed with moderate frequency: U44C and G150A (five times each, Figure 6). We transcribed and assayed for reverse-splicing ability those genotypes that contained either of these mutations (Figure 5). We chose the two most active of these: a four-error mutant (U44C/U97A/U159G/G163C, referred to as 4EM hereafter) and a single-error mutant (G150A), for further characterization alongside the full-length wildtype *Azoarcus* ribozyme, and two less active variants, the Δ P6aRC starting molecule and a two-error mutant (U44C / G150A).

General characterization of selection mutants

To investigate how the most-active ribozymes adapted to the selection conditions, we determined k_{obs} values for each ribozyme through single-turnover kinetics and the maximum fraction of ribozymes able to adopt a catalytically active fold by incubation in the presence of excess substrate (Table 1). The decrease in reverse-splicing activity between the wildtype ribozyme and ΔP6aRC is apparently not due to a difference in k_{obs} . Instead, the reduced catalytic ability of ΔP6aRC is likely a consequence of a lessened propensity to fold into a catalytically active conformation. This phenomenon has been seen before in the behaviors of other pairs of related functional RNAs (Schmitt & Lehman 1999; Huang *et al.* 2009). Following incubation with excess substrate in 100 mM MgCl_2 and 30 mM EPPS (pH 7.5) for the time allowed in selection process, 15 minutes, only 61% of ΔP6aRC have reacted, compared to 89% of the wildtype ribozyme molecules. After 60 minutes, those percentages increase to 73% and 93%, indicating that the ΔP6aRC population folds more slowly and less completely than the wildtype ribozyme, which has nearly reached full reactivity after just 15 minutes.

The selection mutants seem to exploit both folding and kinetic improvements to accommodate the selection criteria; G150A, U44C/G150A, and 4EM adapted to the selection through improved kinetics, increasing their k_{obs} values by 3-fold, 1.4-fold, and 1.6-fold respectively, over that of the wildtype ribozyme. The mutants all fold

better than the $\Delta P6aRC$ form, although none reaches the same fraction of correctly-folded molecules as the wildtype ribozyme. These kinetic enhancements would be sufficiently large to drive a population towards fixation rapidly. Assuming, in the case of the G150A mutant, that a three-fold kinetic enhancement gives a three-fold selection advantage to that mutant, then we can approximate the number of generations it would take for that mutation to become the dominant sequence according to basic population-genetic theory of haploids:

$$\ln\left(\frac{p_t}{q_t}\right) = \ln\left(\frac{1-\mu}{\mu}\right) + t \ln w$$

where p_t and q_t are the frequencies of wildtype and mutants at generation t , respectively, μ is the mutation rate to the mutant allele, and w is the relative fitness disadvantage that the wildtype has to the mutant. If we assume a 10% mutation rate per position per generation, not unreasonable in prebiotic conditions, then the G150A mutation would occur in ~3.3% of progeny molecules. In this scenario, with minimal epistatic effects, over 99% of the species present would contain the G150A mutation after only eight generations of direct competition.

Analysis of the selection-derived mutations

Perhaps the most interesting mutant is the G150A molecule because it is the kinetically fastest ribozyme identified through the selection and folds the most quickly, reaching its reaction equilibrium by 15 minutes. This mutation is in the P8

tetraloop receptor and G150 has been identified as a monovalent metal ion binding site (Basu *et al.* 1998). Considering that we performed the selection in much higher Mg^{2+} concentrations than occur *in vivo* and in the absence of any monovalent ions, this mutation likely stabilizes formation of the L2-P8 tetraloop / receptor interaction, aiding faster and more accurate folding. Extra stabilization here may also help counter the detrimental effects of losing a portion of the P4-P6 fast folding domain. Less intuitive, however, is how this mutation provides such a dramatic improvement to k_{obs} , although presumably the new L2-P8 interaction may have long range effects that stabilize the active site in the presence of elevated Mg^{2+} .

The most active overall mutant, 4EM, contains a suite of mutations (U44C/U97A/U159G/G163C). One, U159G, may have a similar effect as G150A, as it lies directly opposite G150 and may allow access to conformations more applicable to Mg^{2+} -binding or stacking with A148 and A149. A second mutation, U97A, reverses the A97U mutation we introduced in the $\Delta P6aRC$ variant to make it autorecombination-compatible. Perhaps the resulting AAAA L6 is more stable than the original AUAA loop. Alternatively, the mutation could be essentially neutral. The remaining mutations, U44C and G163C, appear to work together to prevent misfolding. Because nucleotides 139–142 (P3) and 143–146 (P8) are both 5'-CACC-3', their corresponding base-pairing partners (40–43 and 162–165, consisting of 5'-GGUG-3'), have the potential to incorrectly base-pair. In general, formation of the P3-

P7 pseudoknot is a limiting step in group I intron folding (Zarrinkar & Williamson 1994; Pan & Woodson 1998; Sclavi *et al.* 1998; Pan *et al.* 2000; Treiber & Williamson 2001; Zhang *et al.* 2005). Native P3 formation is bolstered by the U44C mutation, which trades the U•G wobble for a C-G basepair. Both the non-native “P3” (162–165 base-pairing with 139–142) and native P8 are destabilized by G163C. Both effects combine to allow the critical P3-P7 pseudoknot to form more frequently, helping to offset the loss of part of the P4-P6 fast-folding domain and result in a better folding ribozyme population.

Covalent self-assembly is improved by incorporating selection-derived mutations

The mutations identified through *in vitro* selection improved the reverse-splicing ability of the Δ P6aRC ribozyme. However, reverse-splicing ability is only one of several properties required for covalent self-construction and autorecombination as observed in the *Azoarcus* ribozyme. The necessary fragments must also be able to fold correctly, both individually and in conjunction with their cognate pieces to form the *trans* autorecombinase, and then perform both forward- and reverse-splicing reactions. We thus turned a rational eye towards which of the selection mutations would actually benefit the self-construction process. Within 4EM, three of the four mutations were deemed to be deleterious to covalent self-assembly: U97A because it removes the L6 CAU recombination tag and both U159G and G163C because they almost certainly would destabilize the necessary *trans* complex formation. Instead, we chose to focus

on the two mutations that were the most common during the selection: G150A, which had the largest effect on reverse-splicing kinetics, and U44C, which would likely enhance *trans* complex formation.

To test this hypothesis, we first attempted simple two-piece self-construction reactions with $\mathbf{W} + h\bullet\Delta\mathbf{X}\Delta\mathbf{YZ}$ and $\mathbf{W}_{\text{U44C}} + h\bullet\Delta\mathbf{X}\Delta\mathbf{YZ}$ to verify that the shortened system was still active (Figure 7). Surprisingly, the ΔP6aRC construct was as active as the U44C version in this context, despite its drastically lower reverse-splicing ability. This parity was not maintained, however, in three-piece self-construction reactions. We tested two different connectivities, $\mathbf{W} + h\bullet\Delta\mathbf{X} + h\bullet\Delta\mathbf{YZ}$ and $\mathbf{W} + h\bullet\Delta\mathbf{X}\Delta\mathbf{Y} + h\bullet\mathbf{Z}$, with drastically different results. For the first connectivity set, the effects of the U44C and G150A mutations are roughly additive, giving a robust yield of ~8.5% after six hours, equal to or greater than the sum of either mutant alone, and about 10-fold greater than the ΔP6aRC . (The recombinant product from $\mathbf{W}_{\text{U44C}} + h\bullet\Delta\mathbf{X} + h\bullet\Delta\mathbf{YZ}$ was excised from the gel and genotyped to confirm it was the correct molecule.) The second set, however, was only capable of detectable self-construction with the G150A mutation, and was rendered essentially inactive with the ΔP6aRC , U44C and U44C/G150A double-error mutant molecules. One factor for the difference in activities observed between these two connectivity groups may be a lack of *trans* recombinase enzymes resulting from difficulty in forming the native non-covalent $\Delta\mathbf{Y}\text{-}\mathbf{Z}$ interaction, something overcome to a minor extent by the G150A mutation. Additionally, it is

possible that one or both parts of the ΔY - Z junction (the 3'-end of ΔY or the 5'-end of $h \bullet Z$) are poor substrates for recombination, and the reaction is limited by the reactivity at this junction. Both factors probably play a role, because only ~1.4% of $W \Delta X \Delta Y_{G150A}$ is obtained after 6 hours, compared to $W \Delta X$ yields of 7–9% in the $W + h \bullet \Delta X + h \bullet \Delta Y Z$ system (depending on which mutations are included). Four-piece self-construction reactions, $W + h \bullet \Delta X + h \bullet \Delta Y + h Z$ were attempted, but no full-length ribozymes—and only trace amounts of $W \Delta X \Delta Y$ (in any of the $\Delta P6aRC$ or mutant systems)—were obtained (data not shown).

Conclusion

Using strictly Darwinian *in vitro* selection techniques, it is not obvious how one could directly improve autorecombinase ability—the majority of the full-length products from a randomized pool of RNA fragments that cooperate to perform covalent self-assembly would tend to be the best recombination substrates, not the best catalysts. We thus performed *in vitro* selection to improve the reverse-splicing activity of the *Azoarcus* $\Delta P6aRC$ deletion mutant with the idea that some of the mutations enhancing catalytic ability would also augment self-construction activity. The selection-isolated mutants were able to regain some of the catalytic ability lost by the deletion of P6a by improving both their kinetic parameters and overall folding efficiency, to varying degrees. Some of the reverse-splicing mutations were in fact exaptations (Gould & Vrba 1982), helping to restore self-construction ability in three-piece reactions despite

not being evolved for that specific purpose. Our attempt to improve the autorecombination system described above highlights the unique difficulty of evolution in an RNA World context in that mutations that improve or expand the catalytic activity of a ribozyme must not too severely hamper its ability to be acted upon as a substrate and *vice versa*. Our findings underscore the requisite for multifunctionality in primordial genetic molecules; the ability of RNA or an RNA-like molecule to acquire self-replication required selection to favor catalysis, structure, and cooperation simultaneously prior to the advent of purely selfish genomes.

Materials and Methods

Construction of G_0 selection population

DNA oligonucleotides were purchased from IDT. The DNA template for the $\Delta P6aRC$ form (Figure 3B) of the ribozyme was generated by recursive gene synthesis as described (Engels & Uhlmann 1988; Rydzanicz *et al.* 2005; Hayden & Lehman 2006). The initial mutant population was generated by mutagenic PCR (Cadwell & Joyce 1992; Vartanian *et al.* 1996) to give a mutation level of 10% per position, as described previously (Burton & Lehman 2006). RNAs were synthesized by run-off transcription and gel-purified on 8% polyacrylamide / 8 M urea gels prior to selection.

In vitro selection

For selection generations 0–4, reactions contained 5 μM mutagenized *Azoarcus* ΔP6aRC RNA, 2.5 μM S9 substrate RNA (5'-AACAU•CCAAUCGCAGGCUCAGC-3'; selection is effected because the portion 3' of the “•” is appended only to the ends of active ribozymes upon reaction), 100 mM MgCl_2 and 30 mM EPPS buffer (pH 7.5) in a total volume of 20 μL . In generations 5–10, ribozyme and substrate concentrations were halved to 2.5 and 1.25 μM , respectively. Solutions containing ribozyme and substrate RNAs were heat-denatured at 80°C for 5 min, after which the reaction buffer was added and reactions were incubated at 48°C for 15 min. Reactions were quenched by adding 4 μL 0.5 M EDTA and diluting sample volumes to 500 μL with water. Substrate molecules and salts were removed by concentrating the sample to < 10 μL in a Nanosep 10K column. An additional 500 μL of water were added and the sample was concentrated again. Active ribozymes were selectively amplified in reverse transcription reactions that used all of the desalted and concentrated RNA, T9 primer DNA (5'-GCTGAGCCTGCGATTGG-3'; complementary to the S9 tail and present in a 2:1 molar ratio relative to total ribozymes), 10 mM MgCl_2 , 5 mM dithiothreitol, 50 mM Tris-HCl (pH 7.5), 0.2 mM dNTPs, and 15 U AMV reverse transcriptase (USB) in 20 μL volumes at 37°C for 1 h. Double-stranded DNA templates containing the T7 promoter sequence were generated by standard PCR using T21a (5'-CTGCAGAATTCTAATACGACTCACTATAGTGCCTTGCGCCGGGAA-3') and T20a (5'-CCGGTTTGTGTGACTTTGCC-3').

Genotyping of selection population

PCR products of the G₁₀ population were cloned into the vector pJET1.2 (Fermentas) and transformed into *E. coli* Top 10 competent cells (Invitrogen). Individual colonies were picked as templates for PCR reactions using the primers pJET1.2-F and pJET1.2-R (Fermentas), which generate products of ~325 bp (= the insert size of 205 bp plus 120 bp). Products of the correct size were genotyped using BigDye (v.3) cycle sequencing chemistry.

Assay of selection population and mutants

Colony PCR products containing frequently occurring mutations were diluted and amplified by PCR using primers T21a and T20a to regenerate dsDNA templates for transcription. Following run-off transcription and gel purification, 2.5 μM ribozymes were incubated with 1.25 μM 3'-α³²P•dATP-radiolabeled S9a (Huang & Szostak 1996; Burton & Lehman 2006) in 100 mM MgCl₂, 30 mM EPPS (pH 7.5). Reactions were quenched by adding an equal volume of an 3:1 solution of denaturing gel loading buffer and 0.5 M EDTA, respectively. Products were separated on 15% / 8 M urea or 20% polyacrylamide / 8 M urea gels and visualized by phosphorimaging. Ribozyme activities were calculated using ImageQuant software (v.5.2, GE Health Sciences).

For the most active mutants, values for k_{obs} for reverse-splicing were obtained under single-turnover conditions. Ribozymes were heat-denatured at 80°C for 5 min, then

allowed to fold for 5 min at 48°C in the reaction buffer (final concentrations of 100 nM ribozymes, 100 mM MgCl₂, and 30 mM EPPS (pH 7.5)). Reactions were initiated by the addition of 3'-radiolabeled S9a (~0.1 nM) and quenched with an equal volume of 3:1:: 2X denaturing loading dye: 0.5 M EDTA. Products were separated on 15% polyacrylamide / 8 M urea gels. Values for k_{obs} were estimated by non-linear regression according to the equation, $F = A(1 - e^{-kt})$ where F is the fraction of substrate appended to the ribozymes, and A was set as 0.204, the regression-estimated limit of reverse splicing under these conditions for the fastest mutant, G150A. Values for k_{obs} varied by less than 50% on different days.

For this same set of mutants, the fraction of ribozymes capable of catalysis (*i.e.*, correctly folded) was determined by allowing the reverse splicing reaction to approach equilibrium with 60-fold substrate excess. Reactions contained 0.5 μM ribozyme and 30 μM S9a substrate. RNAs were heated at 80°C for 5 min, and the reaction initiated by addition of the selection buffer. Aliquots of the reaction were quenched after 15 min (matching the selection conditions) and 60 min (near equilibrium) at 48°C. These aliquots of the reaction were quenched and separated on a denaturing 8% polyacrylamide / 8 M urea gel. Products were visualized by SYBR green I staining using the fluorescence mode on the Typhoon scanner and quantified using ImageQuant software.

Covalent self-assembly assays

RNA fragments as partitioned in Figure 3B were either purchased from IDT or synthesized *via* run-off transcription. Fragments other than **W** contained a five-nucleotide 'head' (h) of 5'-GGCAU-3', with the CAU serving as the recombination tag. Reactions contained 2 μ M of each fragment in 100 mM MgCl₂, 30 mM EPPS (pH 7.5), using 5'-radiolabeled **W** molecules. RNAs were heated at 80°C for 5 min prior to the addition of buffer, then immediately placed at 48°C. Aliquots were quenched with an equal volume of 3:1 :: denaturing loading buffer: 0.5 M EDTA. Products were separated on 8% polyacrylamide / 8 M urea gels and quantified by phosphorimaging.

RNA 5' -radiolabeling

To prepare for radio-labeling, RNAs were dephosphorylated in a 10 μ L reaction containing 50 mM Tris (pH 8.5) and 1 mM EDTA with 1 U calf intestine alkaline phosphatase (Roche) for 1 h at 37°C. The phosphatase protein was removed by diluting the reaction to 80 μ L and performing two phenol-chloroform extractions (pH 4.5), and one chloroform extraction. Radiolabeling reactions were performed for 1 h at 37°C. Reactions were 100 μ L and contained 3 μ L γ -³²P•ATP (MP Biomedicals), 10 U Optikinase (USB / Affymetrix) in 50 mM Tris (pH 7.5), 10 mM MgCl₂, and 5 mM DTT. The reactions were quenched by addition of an equal volume of stop solution containing urea and a stoichiometric excess of EDTA. Products were purified as described (Burton *et al.* 2009).

Table 1. Catalytic parameters of *Azoarcus* ribozyme variants.

Ribozyme genotype	k_{obs} (min^{-1})	fraction correctly folded (15 min / 60 min)
Wildtype	0.079	0.89 / 0.93
Δ P6aRC	0.080	0.61 / 0.73
G150A	0.24	0.79 / 0.79
U44C / G150A	0.11	0.77 / 0.85
4EM	0.13	0.85 / 0.9

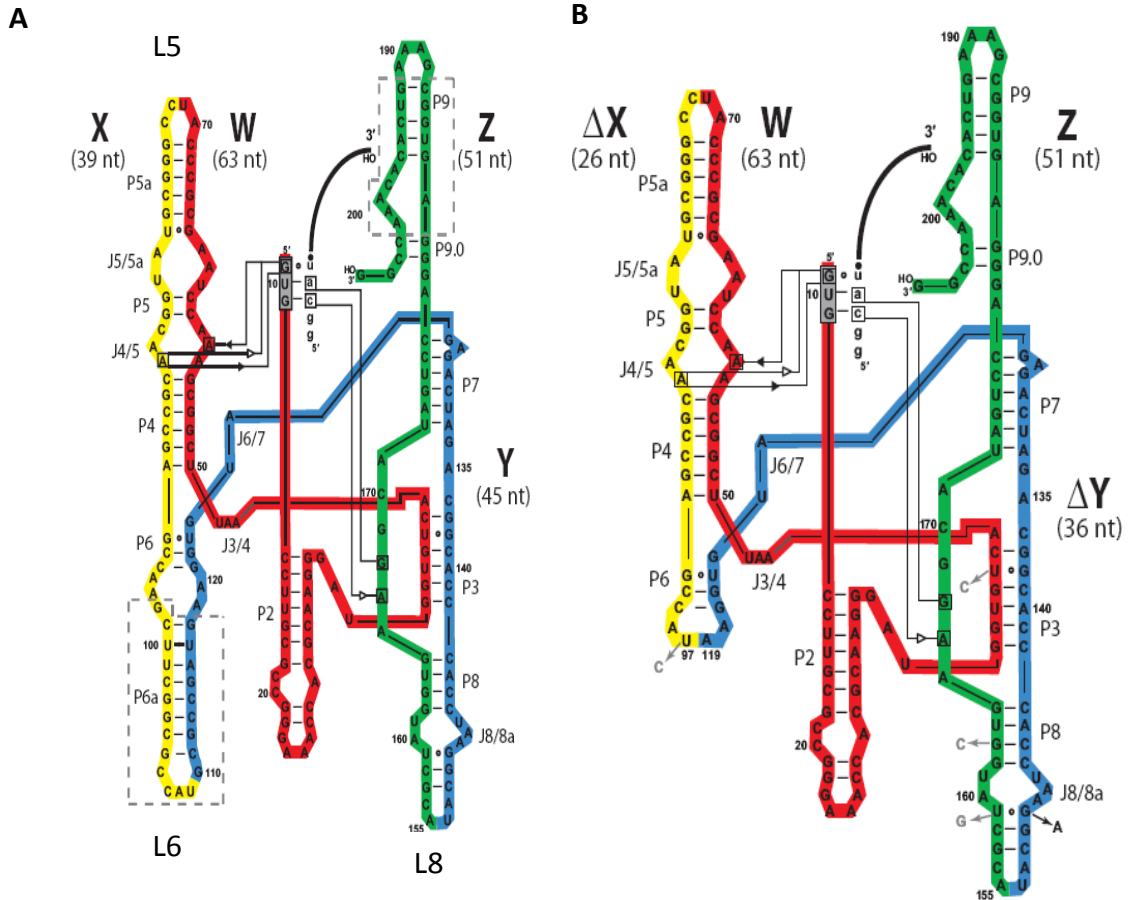


Figure 3. Secondary structures of the *Azoarcus* ribozyme autorecombinase. Fragments are separated at the end of each colored portion. Names and sizes are listed next to each fragment. The internal guide sequence of the each ribozyme is boxed; the interaction between substrate and ribozyme is indicated. A) The 198-nucleotide ribozyme contains CAUs at L5, L6, and L8 that enable it to be autorecombined. Potential deletion sites in the *Azoarcus* ribozyme are outlined by dotted lines. The resulting covalently attached ribozymes were assayed for reverse-splicing ability. B) The 176-nucleotide Δ P6aRC ribozyme. Light gray arrows pointing from nucleotides 44, 97, 159 and 163 represent mutations present in the four-error mutant, 4EM. The dark arrow shows the mutant G150A.

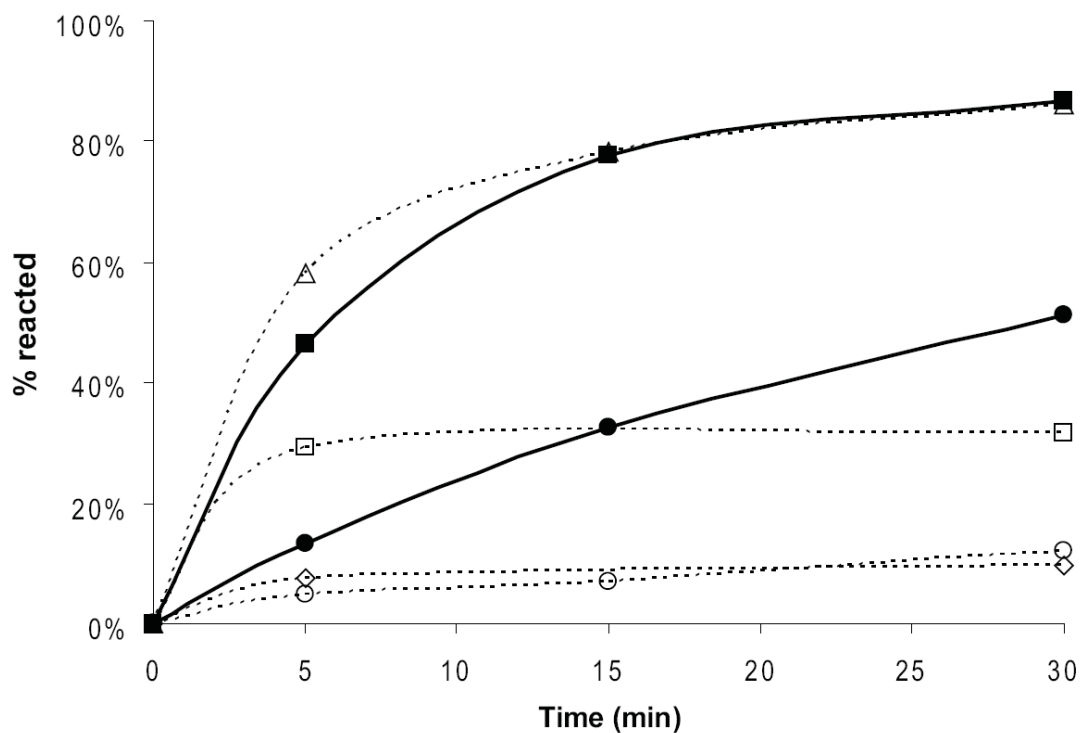


Figure 4. Reverse-splicing activity of *Azoarcus* ribozyme variants. Ribozymes (2.5 μ M) were incubated with radiolabeled substrate S9a (1.25 μ M) ratio for 0–30 min at 48°C in 100 mM MgCl₂ and 30 mM EPPS (pH 7.5). Products were separated on 15% polyacrylamide / 8 M urea gels. The fraction of substrate reverse-spliced was calculated using ImageQuant software. Dashed lines represent pre-selection ribozymes. Open triangles, open rectangles, open diamonds and open circles represent the wildtype *Azoarcus*, Δ P9-17, Δ P9-7, and Δ P6aRC ribozymes, respectively. Solid lines represent the two most active selection-derived mutants from the Δ P6aRC precursor. Closed squares and circles represent the 4EM (U44C/U97A/U159G/G163C) and G150A ribozymes, respectively.

```

ID 9 19 29 39 49 59 69 79 89 97 130 140 150 160 170 180 190 200
WT  GTCCCTTGGCCGGGAAACCGCAGAGGATGGTCTCATATTCCGGGAAACCTTAGCCCCCATCCGGGGGTATGGCAACCGGAGCCAT*AGGTTGATAGACTAGACGGCACCACCTTAGCCATACGCTATGCTGAGGCGATAGTCCAGGAGTGGGTAAGTCTACACAAACCGG
1863 -----C-----
1865 -----T-----
1867 -----T-----
1869 -----G-----
1871 -----T-----
1873 -----T-C-----
1875 -----T-----
1877 -----A-----
1879 -----T-A-----
1881 -----A-----
1883 -----C-----
1885 -----C-----A-----G-----C-----
2325 -----T-----
2327 -----T-----G-----
2329 -----C-----
2331 -----G-----A-----
2333 -----A-----
2335 -----C-----
2337 -----T-----C-----
2339 -----A-----
2341 -----A-----
2343 -----
2345 -----

```

Figure 5. Genotyping of the generation 10 selection population. Sequence alignment of 23 *Azoarcus* ribozyme Δ P6aRC-truncated isolates obtained through *in vitro* selection. The '*' in the wildtype (wt) sequence indicates the point of truncation. The units digit of numbers above the sequence indicates that position in the sequence (*i.e.* the '0' of 160 is above nucleotide 160). Underlined nucleotides represent insertions in the sequence and occur after the position in the sequence they occupy.

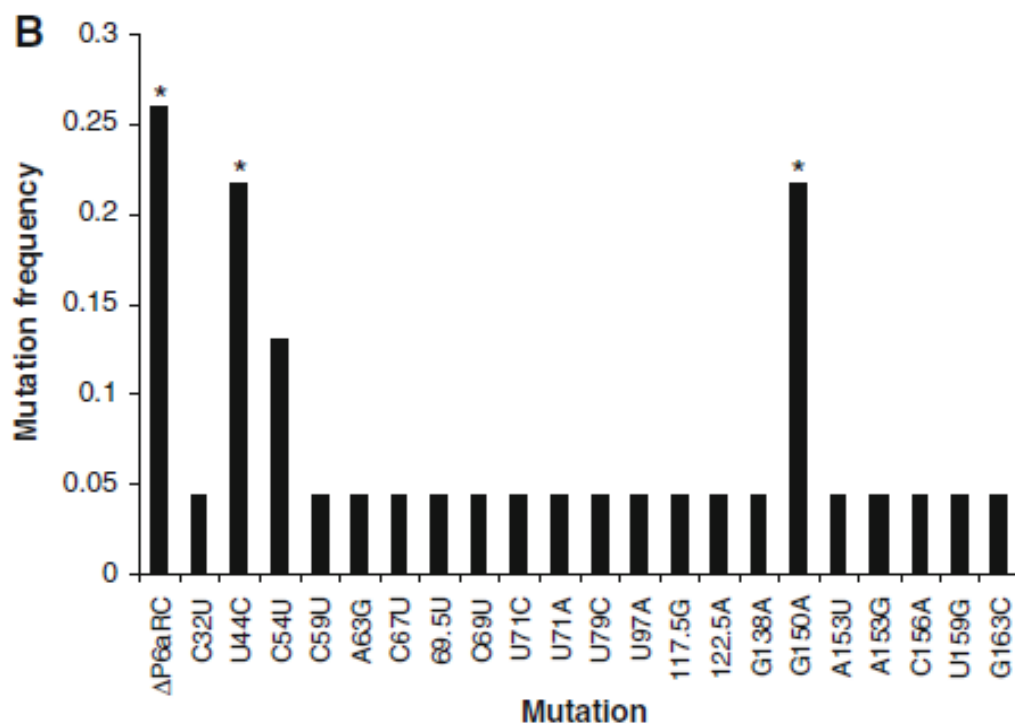


Figure 6. Mutation frequency in the G10 population. Mutations with an asterisks (*) above them were characterized further. Those listed at decimal positions represent insertions into the DP6aRC sequence, between the two flanking integers (i.e., 69.5 U is an insertion of U between nucleotides 69 and 70). The nucleotide numbering in Fig. 1 has been retained for clarity.

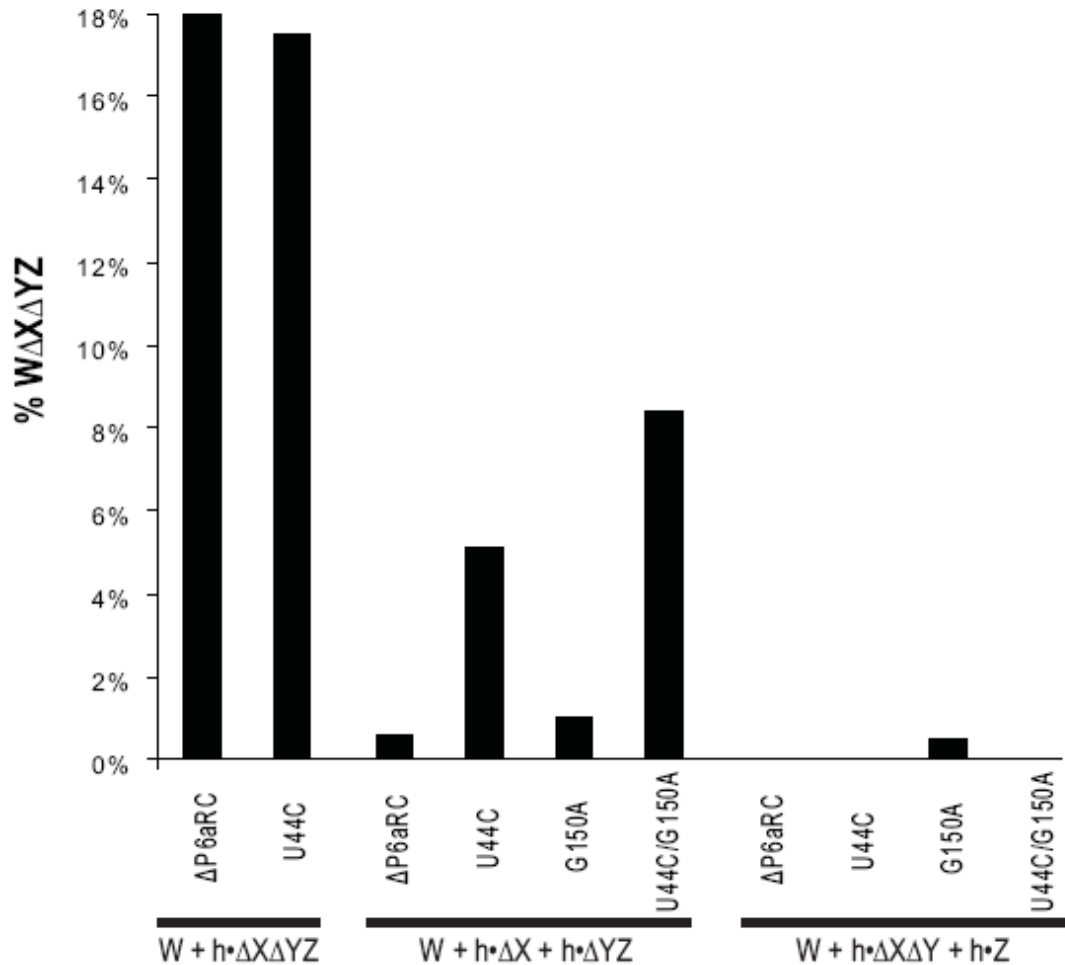


Figure 7. Effects of selection-derived mutations on the self-construction activity of the *Azoarcus* ΔP6aRC ribozyme. Self-construction reactions contained 2 μM of each fragment and were reacted at 48°C in 100 mM MgCl₂ for 6 hours. The number and breakpoints of the fragments are listed below the reactions. Products were separated on 8% polyacrylamide / 8 M urea gels. W-fragments were 5'-radiolabeled, and the percent of WΔXΔYZ was calculated as the ratio between full-length molecules to the sum of all W-containing (radiolabeled) molecules.

Chapter 3. *In Vitro* Selection of a Benzoic Acid Reductase Ribozyme.

Abstract

An RNA population composed of 45 fixed and 65 randomized nucleotides has undergone five generations of *in vitro* selection for the ability to reduce benzoic acid to benzaldehyde using NADH and Zn²⁺. The selection has been designed such that RNA molecules capable of performing the intended reduction were labeled with biotin and separated from inactive molecules based on the interaction between biotin and streptavidin. It appears that the population has responded to a selection pressure because it has evolved to migrate faster during electrophoresis. This increased mobility is likely the result of a deletion event occurring during the selection, and suggests the selection process is working. Although catalytic activity has not yet been observed, the most prominent several alternative explanations for the observed selection results have been eliminated. Consequently, the selection is well-positioned for another 5–10 generations of selection, at which point the population should be assayed again. If activity is detected at that stage, the RNA population should be genotyped and kinetically characterized. If no activity is observed at that point, the selection will be restarted, under different conditions.

Introduction

A hallmark of life is the ability of an organism to extract energy from its environment and to couple that extraction with the synthesis of necessary components for reproduction. Hence, for the RNA World to have been a reality, ribozymes must be able to perform these feats as well (Chen *et al.* 2007). To date, RNA enzymes have been shown to be capable of catalyzing a number of reactions involved in energy transfer (Lorsch & Szostak 1994; Kumar & Yarus 2001; Saito *et al.* 2001; Coleman & Huang 2002; Jadhav & Yarus 2002; Li & Huang 2005). Ribozymes have also demonstrated biosynthetic proficiency using activated leaving groups (Ekland & Bartel 1996; Lohse & Szostak 1996; Wiegand *et al.* 1997; Zhang & Cech 1997; Unrau & Bartel 1998; Zhang & Cech 1998; Johnston *et al.* 2001; Sun *et al.* 2002; Lau *et al.* 2004; Zaher & Unrau 2007; Lau & Unrau 2009). However, only one group of ribozymes capable of transferring electrons from one molecule to another is currently known (Tsukiji *et al.* 2003). Because modern organisms acquire and utilize energy by oxidation / reduction (redox) reactions, these types of reactions are our best guesses of what “riboenergetics” may have looked like in the RNA World. Thus, there is a need to more fully explore the range of redox reactions that ribozymes are capable of catalyzing.

To this end, we have performed *in vitro* selection on a randomized population of RNA molecules to identify ribozymes capable of reducing benzoic acid to benzaldehyde

(Figure 8). The provided electron source was NADH. We have also included Zn^{2+} in the reaction buffer to serve potentially as a cofactor. These conditions were chosen by analogy to the reaction catalyzed by the carboxylic acid reductase class of proteins, which are dependent on both Zn^{2+} and NADH (He *et al.* 2004; Venkitasubramanian *et al.* 2007). However, the protein enzyme also uses ATP to activate the carboxylic acid prior to forming a covalent intermediate with its associated phosphopantetheine coenzyme. This coenzyme is a functional group to which natural RNA sequences do not have access, suggesting that ribozymes capable of this performing this reaction may have to use an alternative mechanism.

Each round of selection begins with the transcription of a dsDNA template to generate RNA. The benzoic acid substrate is then covalently attached to the RNA, followed by incubation in the reaction buffer. Active molecules will reduce their benzoic acid groups to benzaldehyde, which provides a 'handle' by which the active molecules can be separated from inactive ones. The active RNAs are then converted back to DNA and amplified, completing one generation of selection. Five rounds of selection have been completed to date; the results are described below.

Results and discussion

In vitro selection

Our selection method was based on that employed in the Suga lab for their selection of a ribozyme that oxidized benzyl alcohol to benzaldehyde (Tsukiji *et al.* 2003). In their work, they covalently attached their benzyl alcohol substrate to their RNA population. A subset of these RNAs were able to convert the alcohol group to the aldehyde. The aldehyde could react with biotin hydrazide and then be fished out using streptavidin beads. In our case, we wanted to select for reduction of the carboxylic acid to the aldehyde, and then separate the active molecules using biotin and streptavidin. A number of steps were required each generation to facilitate this selection (Figure 9). First, RNA was transcribed in the presence of guanosine monophosphate, resulting in a pool of RNAs that were 5'-monophosphorylated. The monophosphate was then covalently linked to the nitrogen of the intended substrate, **4-AMBA** (4-[aminomethyl]benzoic acid), by nucleophilic substitution using **EDC** (1-ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride) as an activating agent. After purification, the derivatized RNAs were incubated in reaction buffer containing Mg^{2+} , Zn^{2+} , and NADH. The amount of reaction time allowed was progressively decreased to increase the stringency of selection (from 3 h in round 1 to 15 min in round 5). Active benzoic acid reductase molecules would have reduced their carboxylic acid moieties to aldehydes; this served as the 'tag' for the selection, identifying winners from losers. To exploit this tag, the RNA population was desalted and then incubated

with biotin hydrazide, which reacts with aldehydes but not carboxylic acids. Following biotinylation, the active ribozymes (winners) were separated from the inactive ribozymes (losers) by exploitation of the high-affinity interaction between biotin and streptavidin. In this case, magnetic beads with covalently-attached streptavidin were incubated with the RNA population, and then washed twice under denaturing conditions to remove any RNAs not attached to the streptavidin beads by biotin (the losers). Bead-bound RNAs were then washed twice with 0.1 mM EDTA to remove residual urea and rehydrated in 0.1 mM EDTA. The beads themselves were then added to a reverse transcription reaction to generate DNA copies, which were subsequently amplified by PCR.

To date, we have completed five generations of selection, after which we evaluated the ability of the population to perform benzoic acid reduction (Figure 10). RNA from the G_0 and G_5 populations was transcribed and 5'-radiolabeled, and then derivatized with 4-AMBA. The derivatized RNAs were incubated in the selection conditions for up to three hours, then desalted and incubated with fluorescein thiosemicarbazide (**FTC**), which, like biotin hydrazide, will react with an aldehyde but not a carboxylic acid. Thus, any active molecules will be able to acquire a FTC moiety, causing them to migrate more slowly through a polyacrylamide gel than their inactive counterparts. This is an indirect assay, however, because an RNA molecule has to both reduce its carboxylic acid to an aldehyde *and* react with FTC. We chose FTC for the assay rather

than the biotin hydrazide used in the selection for two reasons: 1) FTC is larger and more bulky than biotin hydrazide (Figure 11), meaning it will cause a greater mobility difference; 2) any RNAs that acquire the FTC tag almost certainly are reacting with the thiosemicarbazide group, rather than through an unintended selection mechanism involving biotin hydrazide (explained further below). We did not detect any addition of FTC in either the G₀ or G₅ populations (Figure 10).

Evidence that the selection is succeeding

While the G₅ population has not yet shown the desired catalytic activity, we do observe evidence of molecular evolution. It can be observed that the electrophoretic mobility of the RNA population has increased during the course of the five generations (Figure 10), presumably through the deletion of several nucleotides somewhere between the PCR primer binding sites, indicating that some form of molecular evolution is occurring. Likely explanations for the lack of observed catalytic activity, then, are either that the selection is working, but the population has not been sufficiently enriched for activity to be detected, or the RNA population has found a way to circumvent the selection.

RNAs do not seem to be 'cheating'

One way that non-catalytic RNAs could circumvent the selection would be by remaining bound to the streptavidin beads after the denaturing washes (Figure 9, step

4), despite not having a covalent linkage to biotin. To test this hypothesis, we incubated separately 5'-radiolabeled, 4-AMBA-derivatized G₀ and G₅ RNAs in binding buffer with streptavidin beads for 15 min. We collected the beads using a magnetic separator, and retained the supernatant. Then we tested the efficacy of the two denaturing washes by incubating the beads for 5 min at 50°C in the denaturing buffer, collecting the beads and retaining the supernatant each time. Finally, the beads were incubated for 5 min at 90°C in 8M urea, which should remove any remaining RNA (if an RNA cannot be removed under these conditions, which are more extreme than experienced prior to reverse transcription, then it is essentially impossible for that RNA to become reverse transcribed and survive the selection). Each fraction was loaded in its entirety on a polyacrylamide gel, and the amount of RNA in each band was quantified (Figure 12). If we were selecting for streptavidin-binding RNAs, it would be expected that the G₅ population would have a higher affinity for the streptavidin beads than does the G₀ population. However, it can be seen that for both populations, more than 99% of the RNAs were removed after the two denaturing washes (Figure 12). It is unlikely, then, that we are selecting for streptavidin binders. We cannot completely rule out the possibility, however, that we are simply amplifying the <1% RNAs that are separated from the streptavidin beads during the 8 M urea / 90°C wash, although the 0.1 mM EDTA washes used in the selection are much less likely to remove these from the magnetic beads.

While non-specific streptavidin binding may be the most probable means for RNAs to circumvent the designed selection, there are a few other possibilities. One is that the RNA population may catalyze the addition of biotin hydrazide to itself. An example of internal addition would be nucleophilic substitution of the hydrazide group at the phosphoramidate linkage between 4-AMBA and the RNA or at an internal phosphodiester bond. The probability of either of these reactions being catalyzed by the RNA is very low, however, because the biotinylation reaction takes place in a solution of minimal ionic strength (no divalent metals, $\sim 1 - 2 \text{ mM Na}^+$) that should minimize the ability of RNA to fold into a catalytically active structure. Furthermore, if the phosphoramidate bond was sufficiently activated for non-catalyzed nucleophilic substitution to occur, then this should have been revealed in the assay with FTC. Each reaction timepoint would be expected to produce FTC-labeled RNA, including the 0 hr reaction, if phosphoramidate activation was the mechanism by which biotin or FTC were reacting with the RNA population. We do not observe this, allowing us to discount this possibility.

Conclusion

We have generated a population of randomized RNA molecules composed of 45 fixed nucleotides and 65 nucleotides of random composition. We have also developed a method for covalently attaching the targeted selection substrate, 4-AMBA, to the RNA population using EDC activation. After five rounds of selection, the RNA appears to

have grown shorter as evidenced by its increased electrophoretic mobility. We have indirectly assayed the population for its ability to reduce the carboxylic acid group of 4-AMBA to the aldehyde using FTC, although no catalytic activity was observed. This was not entirely unexpected, however, because *in vitro* selections can often require six or more generations of selection before catalytic activity can be observed (Joyce 1998; Burton & Lehman 2006). We have discounted several possible ways by which members of the selection population could possibly circumvent the selection criteria, such as by evolving to the ability to bind streptavidin. As a result, the G₅ population is well-situated as a starting point for further selection. Another five to ten generations could be sufficient to determine if the RNAs we have selected for so far are catalytically active. If it can then be determined that the RNAs can reduce benzoic acid to benzaldehyde, they will be genotyped and characterized. However, if it is determined that the RNA is inactive after the additional selection rounds, the selection process should be restarted using different reaction conditions. In particular, it may be prudent to try different molecules to serve as the reductant, such as FADH₂ or sulfhydryl-containing molecules such as DTT or glutathione, as the reducing power of NADH may be insufficient for this task.

Materials and methods

Generation of G₀ population

A family of ssDNA oligonucleotides containing the sequence 5'-TCTCGGACGCTAACC-N₆₅-GCCTGATAACTTTTCCCTCC-3' (RONL-65; N₆₅ denotes 65 positions with an equal probability of A, G, C, or T at that site) were purchased from Operon. These molecules were made double-stranded and the T7 promoter sequence added by PCR using primers 5'-TAATACGACTCACTATAGAGCTCTTGAGGAGGGAAAAGTTATCAGGC-3' (ONL-35; the T7 promoter sequence is underlined) and 5'-TCTCGGACGCTAACC-3' (ONL-31). A randomized RNA population was made by runoff transcription of the dsDNA at 37°C for 16 h in 25 mM MgCl₂, 30 mM EPPS (pH 7.5), 5 mM dithiothreitol, 1 mM spermidine, 2 mM each of rATP, rGTP, rCTP and rUTP, 10 mM GMP (to prime transcripts for subsequent derivatization), and 400 U of T7 RNA polymerase (Ambion) in 400 µL total volume. Following transcription, 3 U of DNase I (Fermentas) were added, and the reaction incubated at 37°C an additional 2 h. A two-fold excess of EDTA was added to the reaction, and the RNA was extracted with phenol-chloroform and chloroform to remove proteins and DNA. The RNA was then ethanol precipitated and gel-purified on an 8% polyacrylamide / 8 M urea gel. Transcription products were visualized by UV-shadowing, and eluted from the gel matrix overnight with shaking in 10 mM Tris (pH 7.5) and 200 mM NaCl. The RNA was separated from polyacrylamide fragments and concentrated using Nanosep MF and 10K spin columns, and prepared for

derivatization by washing with water and then 20 mM 4-(aminomethyl)benzoic acid (**4-AMBA**; pH 6.0). The selection substrate, 4-AMBA, was covalently attached using carbodiimide chemistry, which activates the phosphate of GMP-primed RNAs for attack by the primary amine of 4-AMBA. Newly transcribed RNA (~1 nmol) was incubated in 400 μ L 20 mM 4-AMBA (pH 6.0) with 1 mg of EDC•HCl (1-ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride; Thermo Scientific) at 37°C for 2 h in the dark (this reaction time yielded the most derivatized RNA while limiting polymerization of 4-AMBA because EDC activates the carboxylic acid moiety of 4-AMBA in addition to the RNA monophosphate). The unreacted 4-AMBA and EDC were removed from the reaction by spinning through a Nanosep 10K column to a volume <10 μ L followed by addition of 500 μ L H₂O (a >50-fold dilution). This process was repeated twice more, to give a >125,000-fold dilution of the 4-AMBA.

In vitro selection for benzoic acid reduction

Derivatized RNAs (19 μ L, 400 – 500 pmol) were heat-denatured in water at 80°C for 5 min. Following heating, 8 μ L of 5X buffer (containing 30 mM HEPES (pH 7.3), 100 mM MgCl₂, and 0.5 mM ZnCl₂ at 1X) were added and the RNA was allowed to cool to room temperature for 5 min on the benchtop. The reaction was started by the addition of 13 μ L 150 mM NADH (final concentration of ~50 mM), and was incubated at 22°C. The RNAs were allowed 3 hours to react for the first round of selection; the selection stringency was increased through subsequently decreasing

reaction times to 90 minutes (generations one and two), 30 minutes (generation three) and finally 15 minutes (generation four). The reaction was quenched by adding a slight excess of EDTA relative to Mg^{2+} . RNAs were serially desalted as described above. Successful ribozymes would have reduced their benzoic acid groups to benzaldehyde. To identify these molecules, the population was incubated in 100 μ L MEB buffer (30 mM MOPS pH 6.5, 1 mM EDTA, 5 mM biotin hydrazine (Sigma)) at room temperature for 2 h. Unreacted biotin was removed by serial filtration as described above. RNAs were incubated with 100 μ L (10 mg / mL) Dynabeads M-270 Streptavidin-coated magnetic beads (Invitrogen) in binding buffer (10 mM Tris pH 7.5, 1 mM EDTA, 100 mM NaCl) for 15 min at room temperature with occasional mixing to allow the biotin-streptavidin binding to take place. Prior to the addition of RNA, the beads were twice incubated for 2 min with 0.1 M NaOH per the manufacturer's instructions to remove any potential contaminating RNAses. The suspension was placed into a magnetic particle separator (Roche) to collect the magnetic beads at the bottom of the tube and the supernatant removed. To remove RNAs that bind to streptavidin or the magnetic beads by some means other than through biotin we used denaturing washes. The tube was removed from the magnetic separator and the beads were washed (twice) by adding UTE buffer (4 M urea, 10 mM Tris (pH 7.5), 1 mM EDTA) and incubating the suspension at 50°C for 5 min, followed by magnetic separation and removal of supernatant. The beads were resuspended in 100 μ L 0.1 mM EDTA. In order to proceed with the selection, the

remaining RNA molecules (putatively active benzoic acid reductases) were reverse transcribed at 37°C for 2 h in a 50 µL reaction containing 50 mM Tris (pH 8.3), 8 mM MgCl₂, 50 mM NaCl, 1 mM DTT, 400 pmol of ONL-31a, and 7.5 U AMV reverse transcriptase (USB / Affymetrix). The reaction was stopped by the addition of 5 µL of 0.1M EDTA. The resulting DNA was amplified by PCR using ONL-35 and ONL-31a, and the next generation of RNA was formed by transcription as described above, but with reagents and total reaction volume halved because less RNA is generally needed after the first round of *in vitro* selection.

RNA 5' -radiolabeling

The G₀ and G₅ populations were transcribed as described above. To prepare for radiolabeling, 400 pmol of RNA were dephosphorylated in a 10 µL reaction containing 50 mM Tris (pH 8.5) and 1 mM EDTA with 1 U calf intestine alkaline phosphatase (Roche) for 1 h at 37°C. The phosphatase protein was removed by diluting the reaction to 80 µL and performing two phenol-chloroform extractions (pH 4.5), and one chloroform extraction. Radiolabeling reactions were performed for 1 h at 37°C. Reactions were 100 µL and contained 3 µL γ-³²P•ATP (MP Biomedicals), 10 U Optikinase (USB / Affymetrix) in 50 mM Tris (pH 7.5), 10 mM MgCl₂, and 5 mM DTT. The reactions were quenched by addition of an equal volume of stop solution containing urea and a stoichiometric excess of EDTA. Products were purified as described (Burton *et al.* 2009).

Benzoic acid reduction assay

RNAs from G₀ and G₅ were radiolabeled and derivatized with 4-AMBA as described above. Assay reactions were 40 μ L total in volume, and prepared as described above at the beginning of the *in vitro* selection section. After 0, 1, 2, and 3 h, 9 μ L aliquots were removed and transferred to Nanosep 10K spin columns containing 500 μ L H₂O and 2 μ L 0.5 M EDTA. The solutions were spun through the columns at 5,000 g until the volume was \leq 10 μ L. An additional 500 μ L of H₂O was added, and the process repeated. To the desalted RNA, 10 μ L 2 mM fluorescein thiosemicarbazide (FTC) was added. The resulting solution was allowed to incubate at room temperature for 1 h in the dark. A positive control for the labeling was performed by incubating radiolabeled, derivatized RNA with FTC in the presence of EDC (1-ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride). After 1 hr of incubation, 15 μ L of urea-containing loading dye was added (8 M urea, 10 mM EDTA, 0.1% W/V bromophenol blue), and the products were separated on an 8% polyacrylamide / 8 M urea gel. The gel was then visualized by phosphorimaging.

Streptavidin-binding assay

To assess the degree to which the RNA population is able to bind streptavidin, 20 pmol of the radiolabeled, derivatized G₀ and G₅ RNAs were incubated with 100 μ g of streptavidin-coated Dynabeads for 15 min in 20 μ L of binding buffer (10 mM Tris (pH 7.5), 100 mM NaCl, and 1 mM EDTA). After separating the magnetic beads from the

solution using the magnetic separator, the supernatant was removed and added to 20 μL of urea-containing loading dye. The beads were then incubated for 5 min with 20 μL of denaturing UTE buffer (4M urea, 10 mM Tris (pH 7.5), and 1 mM EDTA), and the supernatant again collected following bead separation. This process was repeated once more, and then 40 μL of 8 M urea-containing loading dye were added directly to the streptavidin beads and incubated for 5 min at 80°C. Each of the wash fractions was separated on an 8% polyacrylamide / 8 M urea gel, with the products visualized by phosphorimaging. The amount of radiolabeled RNA in each lane was determined, and the fraction of RNA in each lane of the four-wash series (supernatant from binding buffer, UTE 1, UTE 2, and 8 M urea) was determined.

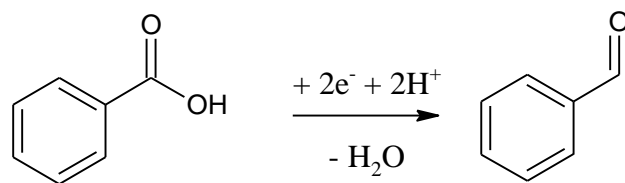


Figure 8. Balanced equation for the reduction of benzoic acid to benzaldehyde. The *in vitro* selection performed herein is designed to identify ribozymes capable of catalyzing this reaction.

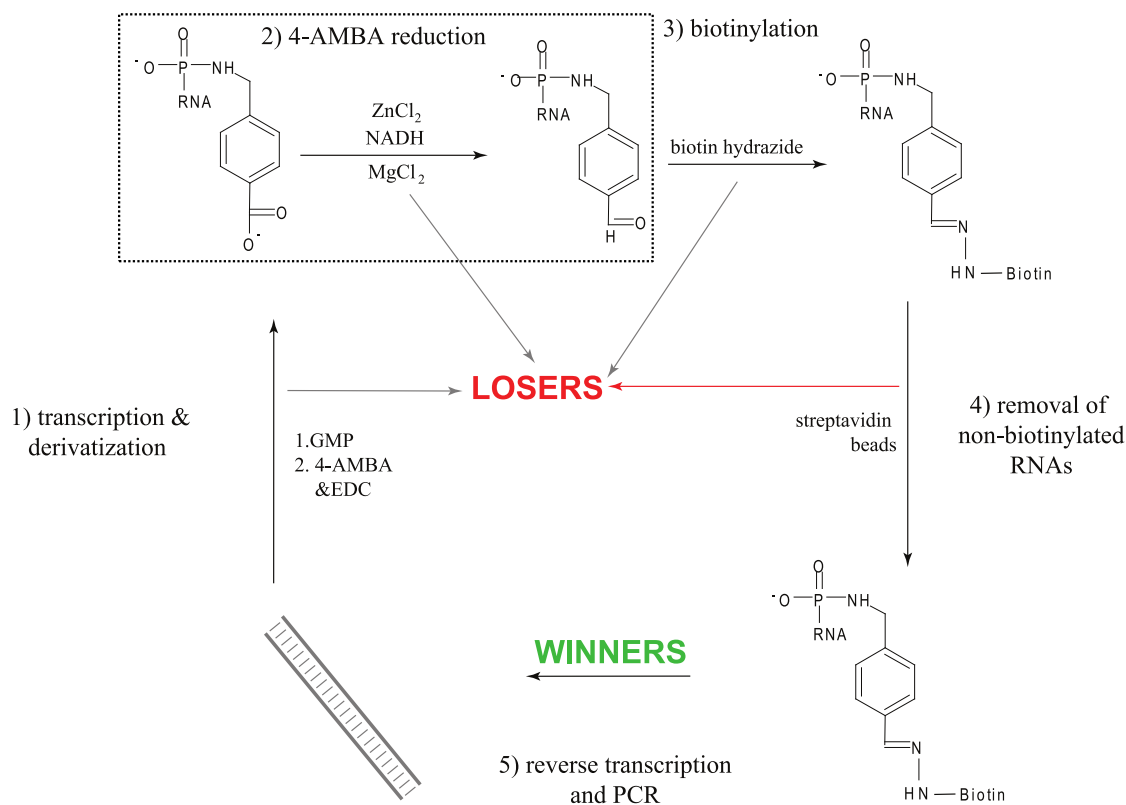


Figure 9. *In vitro* selection procedure used to select for a benzoic acid reductase ribozyme. Steps in each selection cycle are numbered 1-5 and are connected by solid black arrows. The chemical reaction to be catalyzed by the RNA population is boxed. Gray arrows indicate stages in which RNAs can fail the selection criteria, by one of the following: not acquiring the 4-AMBA molecule, failing to reduce the benzoic acid group to benzaldehyde, failing to acquire biotin tag, or not binding to the streptavidin beads. The bold arrow (step 4) indicates the stage at which RNAs that fail the selection (the losers) were removed. Only those RNAs that reach step 5 (reverse transcription/PCR) were carried forward to the next generation.

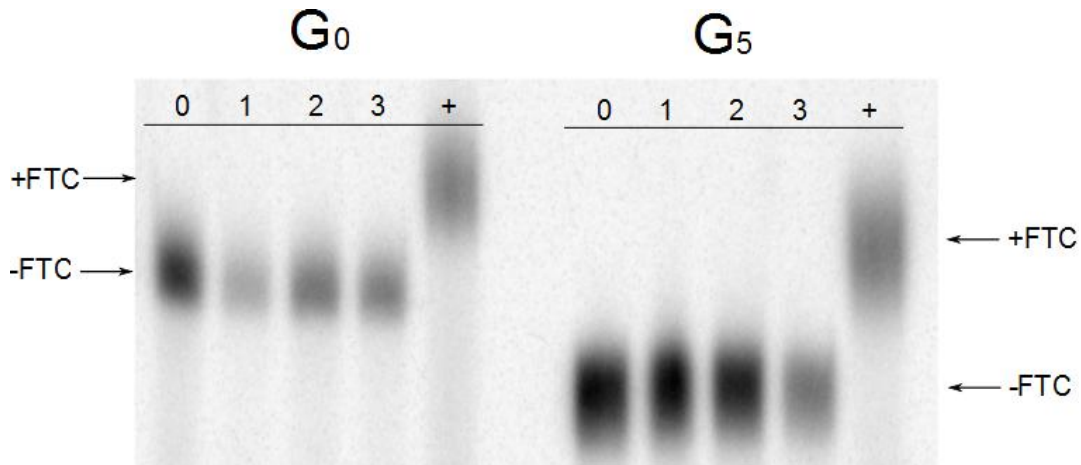
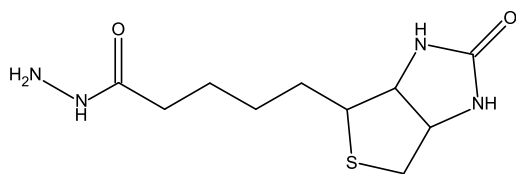
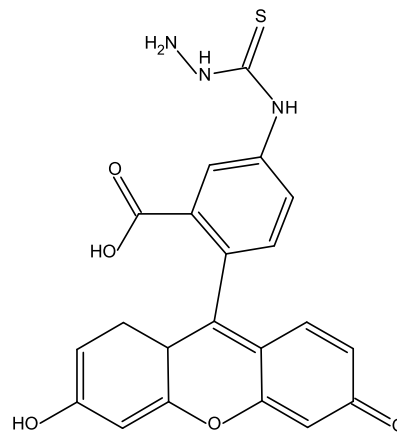


Figure 10. Assay for benzoic acid reduction in the G₀ and G₅ populations. Following incubation in conditions identical to those performed during *in vitro* selection, the radiolabeled RNAs were desalted and incubated with fluorescein thiosemicarbazide (FTC). Molecules that have reduced their carboxylic acid groups to aldehydes will be capable of acquiring the FTC molecule, which will reduce their gel mobility (as seen in the positive control lane (+) where EDC was included in the reaction mixture). The numbers above each lane indicate the incubation time in hours. While no activity is detectable in this assay, it can be seen that the G₅ population migrates noticeably faster than the G₀ population, through the polyacrylamide gel, indicating it is responding to some sort of selection pressure.



biotin hydrazide



fluorescein-5-thiosemicarbazide (FTC)

Figure 11. Aldehyde-reactive molecules used for *in vitro* selection (biotin hydrazide) or the assay of benzoic acid reduction ability (FTC). The biotin portion of biotin hydrazide interacts strongly with the streptavidin proteins coating the magnetic beads used in the selection, allowing biotinylated RNAs to be separated from their non-biotinylated counterparts. FTC reacts with aldehyde-containing RNAs (those that are catalytically active) and adds bulk to the derivatized RNA, assisting in their separation on a polyacrylamide gel.

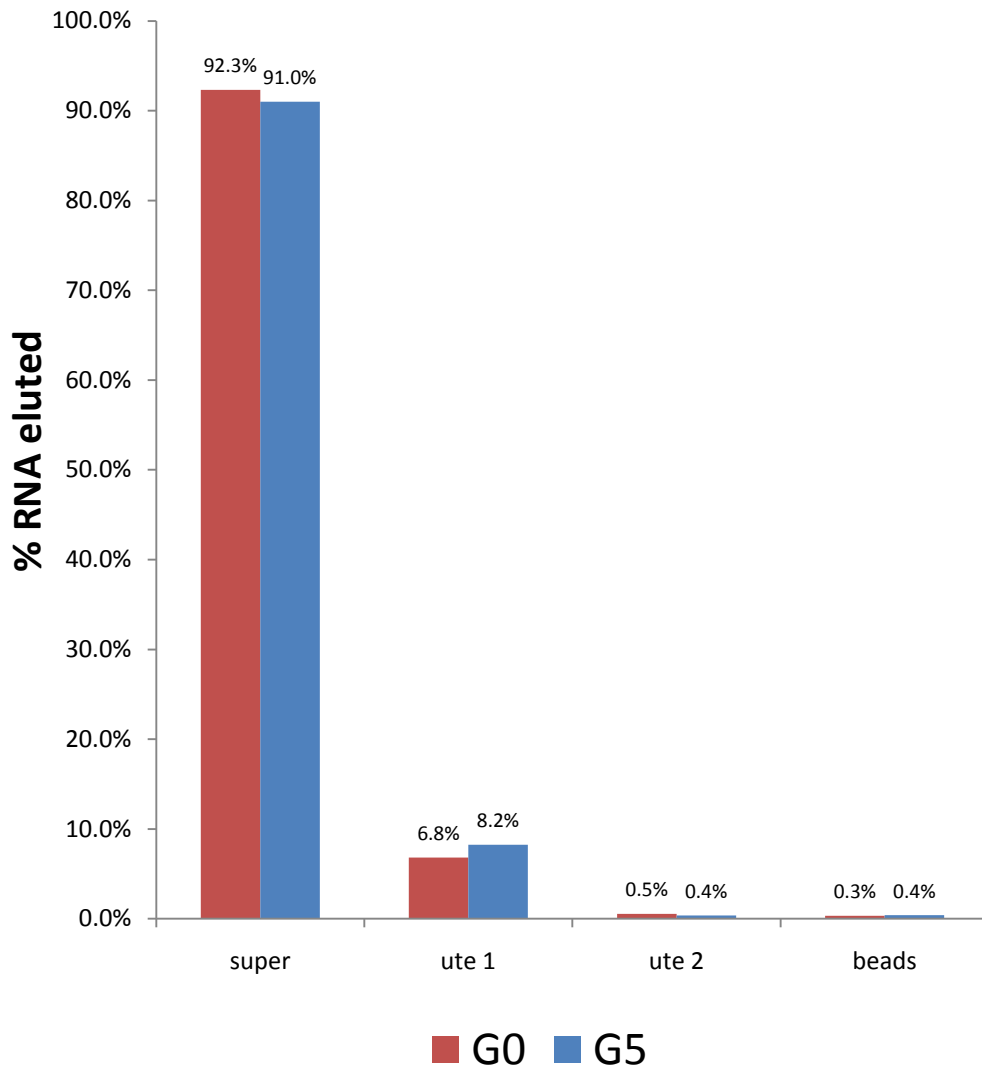


Figure 12. Assay of streptavidin-bead binding by the G₀ and G₅ RNA populations. 5'-radiolabeled RNA molecules derivatized with 4-AMBA were incubated with streptavidin beads in binding buffer for 15 min. The supernatant was removed (super). The beads were then washed twice with the washing buffer used in the selection with the supernatant retained each time (UTE 1 and UTE 2). Finally, the beads were washed with 8 M urea at 90°C to remove any remaining RNA (beads). The entire contents of the supernatant and each wash fraction were separated on a polyacrylamide gel and the amount of radioactivity in each fraction was determined.

Chapter 4. Knotted RNAs Are Formed During *in Vitro* Transcription of the *Chlorella* Virus Pyrimidine Dimer-Specific Glycosylase Intron.

Abstract

In vitro transcription of the 98-nucleotide, A+U-rich intron found in the *Chlorella* PBCV-1 virus *pdg* locus always produces, as 10–20% of the total RNA yield, at least one slower-migrating species in addition to the expected transcript. Characterization of these slower-migrating products suggests that they are knotted RNA molecules, which we call **knRNAs**. The knRNAs convert to the linear form in aqueous salt solutions but not in the presence of denaturants; the linear form cannot adopt a knRNA conformation. Knotting appears to be the result of co-transcriptional misfolding that allows the 3'-end of an intron to become threaded through an internal loop during formation of a pseudoknotted structure. Based on the parameters of high A+U content and a secondary structure predicted to contain a pseudoknot involving its 3'-end, we identified an A+U-rich intron in the *Tetrahymena* *cnjB* locus that also appears to form knots during *in vitro* transcription. Because the *Tetrahymena* intron has just 57% sequence identity with the *Chlorella* intron, knotting in RNA may be widespread.

Introduction

The structures of RNA, DNA and proteins are often approximated by the shapes of their backbone, whether they are chains of phosphoester-linked sugars or amide bonds. In all cases, the resulting shape is rope-like, where the length of the polymer is usually

far greater than its diameter. Yet higher-order topologies, particularly knots, can be induced in ropes by simply placing them in a box and then shaking (Raymer & Smith 2007). It should come as no surprise, then, that knotting has been discovered in polymers of both DNA and proteins (Liu *et al.* 1981; Ménissier *et al.* 1983; Shishido *et al.* 1987; Takusagawa & Kamitori 1996). In the case of DNA, knotting affects gene expression by making the encoded information much less accessible (Portugal & Rodríguez-Campos 1996). This inhibition is significant enough that topoisomerase enzymes exist to unravel knotted – as well as excessively supercoiled – DNA. In proteins, on the other hand, knotting can actually be a beneficial phenomenon. For example, alkaline phosphatases contain a trefoil knot, which has been shown to increase thermal stability and may provide protection from proteases (King *et al.* 2007). Proteins such as tRNA methyltransferases even have knots in their active sites (Elkins *et al.* 2003; Lim *et al.* 2003).

Given that both DNA and proteins are subject to knotting, it seems likely that RNA is as well. Pseudoknots, elements of secondary structure where a single-stranded loop region of an RNA hairpin forms base-pairs with a complementary region outside the hairpin, are well known (Rødland 2006). The knotting we refer to here, however, requires one end of the macromolecular strand to be threaded through an internal loop making an open knot, a phenomenon heretofore undocumented in RNA (Figure 13). However, this is still not a true knot by the mathematical definition, which requires

that the ends of the molecule be joined together. An RNA with an open knot RNA is a topological isomer of its linear form and should have unique physical properties that make detection and characterization possible *in vitro*.

Herein we describe two distinct RNA sequences that form topological isomers during *in vitro* transcription. One of the RNAs derives from the highly conserved intron interrupting the pyrimidine dimer-specific glycosylase (*pdg*) locus of the *Chlorella* PBCV-1 virus. The near 100% sequence similarity in all known natural isolates suggests that it serves a functional or structural purpose (Sun *et al.* 2000; Figure 14A and 14B). The second RNA sequence is found in the *Tetrahymena thermophila* conjugation-specific gene (*cnjB*; Martindale & Taylor 1988; Csank *et al.* 1990). We predicted that topological isomers will be formed during transcription of this gene based on certain similarities with the *Chlorella pdg* sequence. The present study explores the conditions that lead to the formation of the topological isomers, a phenomenon we argue is the result of knot formation, and outlines the path by which the isomers interconvert to the expected linear product. We have performed enzymatic and chemical tests to rule out other possible topologies and conclude that these topological isomers are in fact knotted RNAs or **knRNAs**. Finally, we propose a possible mechanism by which the knots form during transcription.

Results and Discussion

Observation of knotted RNA formation

In vitro transcription of the *Chlorella* virus *pdg* intron invariably produces at least two discrete products: the expected RNA transcript, and one or more products of reduced electrophoretic mobility (Figure 15). Typically, 10–20% of the total transcription yield is slower-migrating products, which we propose are knotted RNA molecules (**knRNAs**; the two predominant species have been labeled here as knRNAs I and II). Formation of knRNAs only occurs co-transcriptionally; renaturation of the full-length linear RNA does not yield detectible levels of knots (data not shown). The commercial source of viral RNA polymerase had no effect on knRNA formation. Furthermore, use of *E. coli* RNA polymerase to transcribe the *pdg* intron also gave knRNAs in addition to the linear transcript (data not shown). The knRNAs show consistent denaturing-gel electrophoretic mobilities from transcription to transcription, but migrate differently than linear standards depending on the polyacrylamide concentration. With 5% polyacrylamide, knRNAs I and II migrate slower than, and at the same rate as, a 197-nucleotide RNA, respectively (Figure 15). Yet on a 15% gel, both knRNAs migrate faster than that same 197-mer RNA. This variable mobility is a hallmark of nucleic acids with a higher order topology such as a branch, a lariat, a circle or a covalently-closed knot (Coppins & Silverman 2004; Wasserman *et al.* 1985; Wang *et al.* 1993; Wang *et al.* 1996; Vogel *et al.* 1997). To illustrate how extreme the gel mobility shifts can be as a consequence of these higher order topologies, we have included a circular

98-mer (generated by intramolecular ligation of the linear transcript) on the gels in Figure 15. We then endeavored to assign each of these topologies, as well as a few others, to the knRNAs (Figure 16).

knRNAs are not linear artifacts

T7 RNA polymerase is known to add untemplated nucleotides at the 3'-end of nascent transcripts (Chamberlin *et al.* 1979; Dunn & Studier 1983). It has also been shown that a growing RNA chain can form a hairpin that allows the transcribed RNA to act as its own template (Cazenave & Uhlenbeck 1994). The variable gel mobility described above, however, indicates that the knRNAs are not linear, allowing us to discount the hypothesis that the knRNAs are overextension products.

knRNAs are not RNAs with persistent secondary structure

A number of nucleic acid structures are known to persist even in high concentrations of urea or formamide, such as the pseudoknotted structure of the HDV ribozyme and G-quartets (Sen & Gilbert 1988; Williamson *et al.* 1989; Wadkins & Been 2002). These structures cause the molecules to have anomalous electrophoretic gel mobilities, and could potentially explain the mobility shifts of the knRNAs. To rule out this possibility, we made rigorous attempts to denature gel-purified knRNAs. Glyoxal treatment, which covalently modifies the base-pairing surface of guanine nucleobases, has been shown to denature RNAs that maintain secondary structure elements in urea

and / or formamide (Huang *et al.* 2009). Gel-purified knRNAs were treated with glyoxal according to the supplier instructions (Ambion) and electrophoresed through polyacrylamide gels both with and without urea (Figure 17). Based on the large gel-mobility shifts of the knRNAs on the non-denaturing gel, it is clear that glyoxylation is sufficient to denature the RNA, even in the absence of urea (Figure 17A). Denaturation of the knRNAs, however, does not convert them to linear molecules. We also attempted to unknot the knRNAs using formamide. Incubating the knRNAs in 95% formamide at 99°C for up to two hours, followed by electrophoresis at 50°C through an 8% polyacrylamide / 8M urea / 40% v/v formamide gel does not result in conversion to the linear product, although degradation occurs (Figure 18A). Additionally, we tried heating the knRNAs in 95% formamide at 99°C for 5 min, along with up to three iterations of incubations for 5 min at room temperature and 5 min at 99°C, followed by electrophoresis at 50°C on an 8% polyacrylamide / 8M urea / 40% v/v formamide gel. We still do not observe any conversion of knRNAs to the linear form, although some bands of slightly faster mobility than the starting knRNAs appear (Figure 18). We conclude that we are in fact denaturing the knRNAs, and that denaturation does not alleviate the gel mobility differences observed in the knRNAs. Therefore, the knRNA phenomenon is not due to persistent secondary structure.

knRNAs are not branched molecules

Branched and circular RNAs have been observed in nature and studied extensively *in vitro*. These molecules migrate much slower than linear molecules with the same number of nucleotides and show varying electrophoretic mobilities relative to linear standards depending on polyacrylamide gel composition (Semancik *et al.* 1973; Dickson *et al.* 1975; Grabowski *et al.* 1984; Padgett *et al.* 1984; Ruskin *et al.* 1984). A branched molecule, whether in a Y-shape from an intermolecular *trans*-esterification or a lariat from an intramolecular *trans*-esterification, necessarily contains a nucleotide with both 2'-5' and 3'-5' linkages. To investigate the possibility of branching, the knRNAs were incubated with yeast RNA lariat debranching enzyme (Dbr I), which cleaves the 2'-5' phosphoester linkage in branched RNA molecules (Ruskin & Green 1985; Jacquier & Rosbash 1986; Chapman & Boeke 1991). This enzyme has no effect on the mobility of the knRNAs, allowing us to discount branches and lariats as the cause of the knRNA gel mobility shifts (Figure 19).

knRNAs are not circles

Circular RNAs both migrate slower than their linear counterparts on denaturing polyacrylamide gels and experience mobility shifts during gel electrophoresis depending on the polyacrylamide content of the gel (Wang *et al.* 1996). The knRNAs we observe, however, respond differently than a 98-mer circle (Figure 15; the circle was generated by ligation of the 98-mer linear molecule). The observation that the

circle can migrate either faster or slower than the knRNAs allows us to rule out the possibility that the knRNAs are circular RNAs.

knRNAs do not contain phosphotriester bonds

It has been demonstrated that RNAs are capable of making phosphotriester bonds (Valadkhan & Manley 2001, 2003). The RNAs containing a phosphotriester bond exhibit gel mobility shifts and persist through multiple rounds of gel electrophoresis. However, this class of bonds was demonstrated to be particularly base labile, with a half-life of ~3.5 min at pH 12 at 50°C (Valadkhan & Manley 2001). When the knRNAs were incubated at pH 11.3 at 50°C, however, the degradation pattern of the RNA was a smear (Figure 20) rather than direct conversion to the linear RNA or an intermediate, as was seen with the phosphotriester bond (Valadkhan & Manley 2001). We thus ruled out base-labile covalent bonds, including RNA phosphotriesters.

knRNAs are not the result of photo-crosslinked bonds

Carbon-carbon bonds could potentially be formed between two bases as seen in photo-crosslinking experiments, although this seems unlikely given that knRNAs are formed even when transcriptions are performed in the absence of light. To test this first possibility, we subjected linear and knRNA molecules to 254 nm light under conditions that should favor reversal of any bonds resembling photo-crosslinked

products (Zhirnov & Wollenzien 2003). This treatment had no effect on the mobilities of the knRNAs, discounting this explanation (data not shown).

knRNAs do not contain phosphoramidate bonds

An amino group on a nucleobase could perform nucleophilic attack on one of the phosphoester bonds, as was proposed for the mechanism of an RNA capping ribozyme (Zaher & Unrau 2006). This would require the formation of a phosphoramidate bond, which has been shown to be labile under acidic conditions (Letsinger *et al.* 1972). To explore this possibility, 5'-radiolabeled knRNAs were incubated at room temperature for 21 hrs in solutions containing 30 mM Na⁺ ions, and various buffers at 30 mM [Tris (pH 8.0), PIPES (pH 6.0), and acetate (pH 5.2)] or 0.1 mM HCl (pH 4). Products were separated by electrophoresis on an 8% polyacrylamide / 8M urea gel (Figure 21). While conversion between knRNAs and to the linear molecule is observed, this process is not noticeably enhanced by lower pH, and may in fact be hindered. Some degradation of the knRNAs can also be observed below the linear conversion product, generally corresponding to the degradation products observed from the linear starting molecule, indicating that these molecules are of the same sequence.

knRNAs convert to linear during incubation in mild salt solutions

When the individual products were eluted from gel slices at room temperature with agitation in 300 mM sodium acetate (pH 5.2) we observed time-dependent interconversion in the order: knRNA I \rightarrow knRNA II \rightarrow Linear (Figure 22A). This interconversion pathway is also observed in a solution containing 15 mM MgCl₂, 50 mM EPPS (pH 7.5), 5 mM dithiothreitol, 2 mM spermidine (Figure 22B). We have not detected conversion from Linear \rightarrow knRNA I, Linear \rightarrow knRNA II, or knRNA II \rightarrow knRNA I, suggesting that knRNA II is a conformational intermediate on the path from knRNA I to the linear molecule (based on Figure 22B there appears to be more than one intermediate). The observation that the knRNAs are unraveled in mild salt solutions but not under severely denaturing conditions lies in stark contrast to a recent report of an RNA sequence that is able to form two distinct topological isomers during transcription (Huang *et al.* 2009). This RNA is an aptamer that inhibits the AMPA-specific receptor, and the two conformers generated during transcription cannot be interconverted; furthermore, when that RNA sequence was synthesized using solid-phase phosphoramidate chemistry, it did not adopt either of the conformations achieved by the transcribed RNAs. Unlike the knRNAs we observe in the *Chlorella* virus intron, however, the mobility differences between the two transcription-mediated folds of the AMPA aptamer were relieved by glyoxal treatment. The authors thus concluded that the topological isomers they observed were not the result of knots, but instead most likely due to persistent secondary structure (Huang *et al.* 2009).

Proposed mechanism for knot formation

To explore how knot formation could occur, it is helpful to consider tying a knot in a rope. In order to make the simplest knot, one has to make a loop in the rope, cross the free ends, and thread one of those ends through the loop so that it goes both over and under the loop, as seen in the open knot in Figure 13. To make more complex knots, one increases the number of nodes, or times the free ends wrap around each other prior to threading one of the free ends through the loop. Within the context of RNA, almost every stem-loop that is formed provides an opportunity for knot formation. Because an A-form helix completes a turn every ~11 base-pairs, the ‘free ends’ cross twice in this distance, or once every ~5.5 base-pairs. For the threading event to lead to knot formation, it must add at least two strand crossings. Pseudoknot formation between a free end and an internal loop may make this possible (Figure 23). This is easiest to envision in molecules containing pseudoknots where either the 5'- or 3'-terminus base-pairs with an internal loop of the molecule because the end of the strand must pass through the loop to lead to a knot. Complete threading becomes more and more disfavored as the length of the RNA to be passed through the internal loop increases, until only slipknot arrangements can occur, wherein the nucleotides involved in pseudoknot basepairing extrude and then return through the loop. However, it has been shown that the addition of fast-folding, non-native protein domains to the N- and C-termini of knot-containing proteins does not preclude knot-formation (Mallam *et al.* 2008). Still, many RNA molecules contain pseudoknots at or near their termini

(Rødland 2006) and none have yet been identified as forming knots, thus this cannot be the only requirement for knotting. Considering that the full-length linear molecules do not fold into their knotted counterparts, there must be one or more co-transcriptional folding intermediates that lead to knot formation. Because pseudoknots require the formation of two helical domains, knot formation by pseudoknotting could occur by one of two routes (Figure 23). In our case, the main helix could be formed first, followed by formation of the pseudoknot by the 3'-end. Conversely, the helix involving the 3'-end could form first, followed by threading of the 5'-end through the resulting loop, driven by the formation of a large number of base-pairs. Either of the routes depicted in Figure 23 might be assisted by the high A+U content of the *pdg* intron, which would presumably keep the loops more open and flexible by weakening adjacent helices.

knRNA formation in another intron

To test the premise that RNA sequences with both a high A+U content and the potential to make pseudoknots involving their 3'-end are susceptible to knotting, we made a dsDNA template of the 95-nucleotide intron found in the *Tetrahymena thermophila cnjB* locus (Figure 24A). The corresponding RNA is 87% A+U, has only 57% sequence identity with the 83% A+U *Chlorella virus pdg* intron, and the RNA-folding software we used predicted it to form a pseudoknotted structure involving its 3'-end (Figure 24B). *In vitro* transcription of this molecule did in fact produce

knRNAs (Figure 24C and Figure 25) supporting our assertion that knotting is aided by high A+U content and 3'-end pseudoknot formation.

In order to further explore regions of the *Chlorella pdg* intron critical for knotting, we introduced a series of mutations to the DNA sequence prior to transcription (Figure 25, sequences in Table 2). Mutations in the last 18 nucleotides have the greatest impact on knRNA formation. Changing 16 of these nucleotides (81–98) virtually eliminates knot formation. Mutating the last nucleotide, either individually or in conjunction with the last three (G98U or 96–98, respectively) reduces knotting significantly. Deleting the last one or four nucleotides (L-1 and L-4, respectively) also reduces knRNA formation quite effectively, in agreement with our hypothesis that an interaction involving the 3'-end is important for knotting.

How unknotting might occur

The conversions from knRNA I \rightarrow knRNA II \rightarrow Linear RNA that we describe above, apparently requiring metal ions, are consistent with the unraveling of an open knot. Higher ionic strength would provide more shielding of the phosphate backbone. This enhanced shielding makes it easier for the negatively charged phosphate groups to slide past each other, allowing the knot to unravel. In the absence of such charge minimization, the electrostatic repulsion of the phosphate backbone may trap the RNA in a knotted conformation. The denaturing treatments to which we subjected the

knRNAs may not facilitate unknotting because they all rely on low ionic strength. This highlights fundamental differences between denaturation and unknotting. For denaturation, attractive base-pairing interactions between nucleotides must be disrupted, either chemically or thermally, and phosphate backbone shielding is undesirable because it promotes secondary structure stabilization. In order to allow strand passage for unknotting, repulsive electrostatic forces between approaching phosphate groups must be minimized. It is no surprise, then, that the use of denaturants does not lead to unknotting. In the case of the glyoxal-treated RNA, the bulky adducts formed between glyoxal and the guanine moieties might actually hinder unknotting by increasing the size of the strand that must be passed through the knot.

Significance of knRNAs

Knot formation – and removal – in DNA are mediated by resolvases and topoisomerases *via* recombination events (Wasserman *et al.* 1985). The possibility of knRNA in biology is bolstered by the observation that at least one of the DNA knot-resolving enzymes, *Escherichia coli* topoisomerase III, is capable of removing covalently closed knots from synthetic, circular RNAs *in vitro* (Wang *et al.* 1996). In the present study, with the caveat that we are not transcribing under cellular conditions, the observation that the *pdg* intron appears to knot during *in vitro* transcription with both T7 and *E. coli* RNA polymerases suggest that it may also form knots *in vivo*. The knots resulting from the wildtype sequence may allow regulation of

gene expression by preventing splicing of the intron, which contains numerous in-frame translational stop codons, serving to down-regulate PDG production through early translation termination, or by preventing the pre-mRNA from binding to the spliceosome at all. Variations in knotted conformation, observed in even single-error mutants, might inhibit such regulation mechanisms, also supporting the high degree of sequence similarity observed.

The formation of knRNAs provides a striking example of the effects of different folding pathways on RNA molecules of the same sequence. This difference between co-transcriptional and post-transcriptional folding pathways has been observed previously with large RNAs (Pan *et al.* 1999; Heilman-Miller & Woodson 2003) and a small RNA (Huang *et al.* 2009). The study of knRNA folding may offer insights to how RNAs can adopt complex topologies. The large, persistent mobility shift conferred by knotting allows easy detection of knotted products by denaturing polyacrylamide gel electrophoresis. Thus, the effects of mutations can be readily observed on such gels if they change the extent or complexity of the resulting knots, making this system worth further exploration.

Materials and methods

Sources of oligonucleotides and DNA

DNA oligonucleotides were purchased from Integrated DNA Technologies. A dsDNA template of the native intron with the T7 promoter sequence was constructed by recursive gene synthesis (Engels & Uhlmann 1988; Rydzanicz *et al.* 2005). The *E. coli* promoter sequence (5'-CGAAAAACAGGTATTGACAACATGAAGTAACATGCA GTAAGATACAAATCG-3') and the various mutations were introduced during recursive gene synthesis by changing the appropriate oligonucleotide(s).

In vitro transcription

Transcription reactions generally contained ~20–30 absorbance units of PCR-amplified DNA template in 15 mM MgCl₂, 50 mM EPPS (pH 7.5), 5 mM dithiothreitol, 2 mM spermidine, 2 mM of each rNTP, and 50–200 U of T7 RNA polymerase (Fermentas or Ambion) in a volume of 10, 20, 50 or 200 µL for the specified time. Radiolabeled RNAs were generated by including 10–20 µCi of α-³²P•ATP (for internally labeled molecules), or 50–100 µCi of γ-³²P•GTP (for singly, 5'-radiolabeled RNAs; MP Biomedicals) in the transcription reaction. Following incubation, reactions were typically treated with 3 U DNase I (Fermentas) for 30–60 min at 37°C, quenched by adding EDTA to a final concentration of 20 mM, acid-phenol / chloroform extracted, and ethanol precipitated. Products were separated on 5, 8, or 15% polyacrylamide / 8 M urea gels and visualized by phosphorimaging.

Radiolabeled products were purified by phosphorimaging using Dip-N-Dot (Burton *et al.* 2009). All gel-purified products were eluted from their gel slices in 200 mM NaCl / 10 mM Tris (pH 7.5) / 0.5 mM EDTA, and desalted and concentrated using Nanosep MF and 10K columns. Quantifications were performed using ImageQuant software on images obtained from a Typhoon scanner (GE Health Sciences). *E. coli* RNA polymerase transcription reactions generally contained ~40–60 absorbance units of PCR-amplified DNA template (containing the T7 A2 promoter; Chamberlin *et al.* 1979; Dunn & Studier 1983) in 10 mM MgCl₂, 50 mM KCl, 40 mM Tris (pH 8.0), 5 mM dithiothreitol, 50 µg / mL BSA, 300 µM each NTP, and 4 U of *E. coli* RNA polymerase sigma-saturated holoenzyme (EPICENTRE) in a volume of 20 µL and were incubated for 90 minutes at 37°C. Radiolabeled RNAs were generated by including 10–20 µCi of α-³²P•ATP in the transcription reaction. Reactions were terminated, visualized, and quantified as above.

Glyoxal and formamide denaturing of knRNAs

To verify that the knRNAs were not caused by persistent secondary structure, they were chemically denatured with glyoxal. Roughly 1–5 pmol of 5'-radiolabeled knRNA or linear molecules (10 µL volume) were added to 10 µL of glyoxal loading dye (Ambion), and incubated for 60 min at 50°C. Products were separated on either an 8% polyacrylamide / 8M urea gel or a 6% polyacrylamide non-denaturing gel without urea and without EDTA in the running buffer. Gel-purified, 5'-radiolabeled knRNA I and

linear molecules were also incubated in 95% formamide solutions (Ambion) for 10, 60, and 120 min at 99°C and separated by electrophoresis through 8% polyacrylamide / 8 M urea / 40% v/v formamide gels at 50°C. A step-wise denaturation process was also performed in which 5'-radiolabeled knRNAs I and II, and the linear RNA, were incubated in 95% formamide solutions at 99°C for 5 min, or 5 min followed by one, two, or three cycles of 5 min at 99°C and 5 min at room temperature. The products were separated on 8% polyacrylamide / 8 M urea / 40% v/v formamide gels by electrophoresis at 50°C. Samples were loaded while they were at 99°C and submitted to electrophoresis between sample loadings.

Incubation of knRNAs with debranching enzyme

To verify that the topologically-altered mobility was not due to lariat formation with an internally branched nucleotide, radiolabeled topological isomer knRNA was gel-purified and incubated with yeast debranching enzyme (Dbr I, a generous gift from Jay R. Hesselberth and Stanley Fields at the University of Washington, Seattle, WA). Because this enzyme requires a “tail” of ~8 nucleotides beyond the branch site, a form of knRNA that contained the first 19 nucleotides of the 3'-exon was also incubated with this enzyme (its mobility was not altered by the enzyme; data not shown). Reactions contained 5 pmol of knRNA and ~50 ng Dbr I in 30 mM EPPS (pH 7.5), 0.5 mM MgCl₂, 4 mM MnCl₂, 100 mM KCl, and 1 mM dithiothreitol. Products were visualized by phosphorimaging. As a positive control, Dbr I successfully cleaved the

2'-5' linked portion of the branched RNA formed by the 7S11 DNAzyme (Coppins & Silverman 2004).

Alkaline hydrolysis reactions

Partial alkaline hydrolysis was performed by incubating 1–10 pmol of gel-purified, 5'-radiolabeled RNA in 50mM K₂PO₄ (pH 11.3) at 50°C for 0–10 min, alongside RNA incubated at 50°C for 10 min in 50 mM EPPS (pH 7.5). Products were separated on a denaturing 8% polyacrylamide / 8M urea gel and visualized by phosphorimaging.

UV-photocrosslinking reversal experiments

Gel-purified, 5'-radiolabeled linear and knRNA molecules in open 200 µL PCR tubes were exposed to 254 nm light from a hand-held Spectroline ENF-240C UV lamp (310 µW / cm² at a distance of 15 cm) positioned 2 cm above the samples for 10 min. Linear RNA molecules were folded for 10 min at 37°C in the T7 RNA polymerase transcription buffer and irradiated for 0–20 min using the same UV light source and set-up. Products were separated on an 8% polyacrylamide / 8 M urea gel and visualized by phosphorimaging.

Unknotting of knRNAs

In an attempt to relieve the slower migration of the knRNAs, we incubated gel-purified, 5'-radiolabeled knRNAs and the linear transcription product were incubated

in 300 mM sodium acetate (pH 5.2) or 15 mM MgCl₂ / 50 mM EPPS (pH 7.5) / 5 mM dithiothreitol / 2 mM spermidine at room temperature for up to 62 hours, with or without gentle shaking. The products were separated by electrophoresis through a 5% polyacrylamide / 8 M urea gel, then visualized and quantified by phosphorimaging.

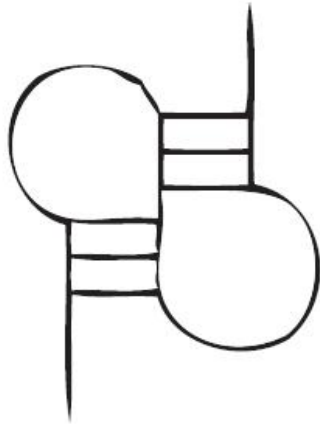
RNA structures

RNAs were folded using the ILM web server (Ruan *et al.* 2004). Predicted structures were drawn and manipulated using the XRNA software found at <http://rna.ucsc.edu/rnacenter/xrna/xrna.html>; accessed 2/01/2008).

Table 2. Sequences of molecules tested for knRNA formation. Wildtype sequences are in plain text. Mutations are listed in italics. Nucleotides within mutated regions which are not mutated are listed in bold.

Wildtype	GUAUGUAAAUAUUAAUAAUCACAAACUUAUUAAUUCAUAAUCACAAACUU AAUUAUAUUCUUUAUUUAAUAAAAUUUCGAUUUGUAAUGUUUUUGCAG
C96A	Wildtype sequence with C96 mutated to <i>A</i>
A97U	Wildtype sequence with A97 mutated to <i>U</i>
G98U	Wildtype sequence with G98 mutated to <i>U</i>
96–98	CAG mutated to <i>AUU</i>
81–98	UUUGUAAUGUUUUUGCAG mutated to <i>ACCAGUCGGACUCCCAUU</i>
L-1	Wildtype sequence minus G98
L-4	Wildtype sequence minus nucleotides 95 – 98
<i>T. thermophila</i> <i>cnjB</i> intron	GUAUUUAUUUCUGAAUAAUUUAUUUUUCUCAAAGCUUCUUUUUCAUGA AAAUAUUUUAAAUAUUAAUGAAUAAUUCAUUUUUUAAAAUUAG

Pseudoknot



Open knot

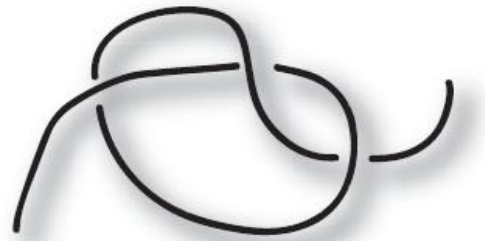


Figure 13. Distinction between a pseudoknot (left) and an open knot (right). Pseudoknots are elements of RNA secondary structure which are capable of being represented in two dimensions, and consist of RNA strands folded back on each other. An open knot cannot be represented in two dimensions, and requires one end of the molecule to be threaded through an internal loop.

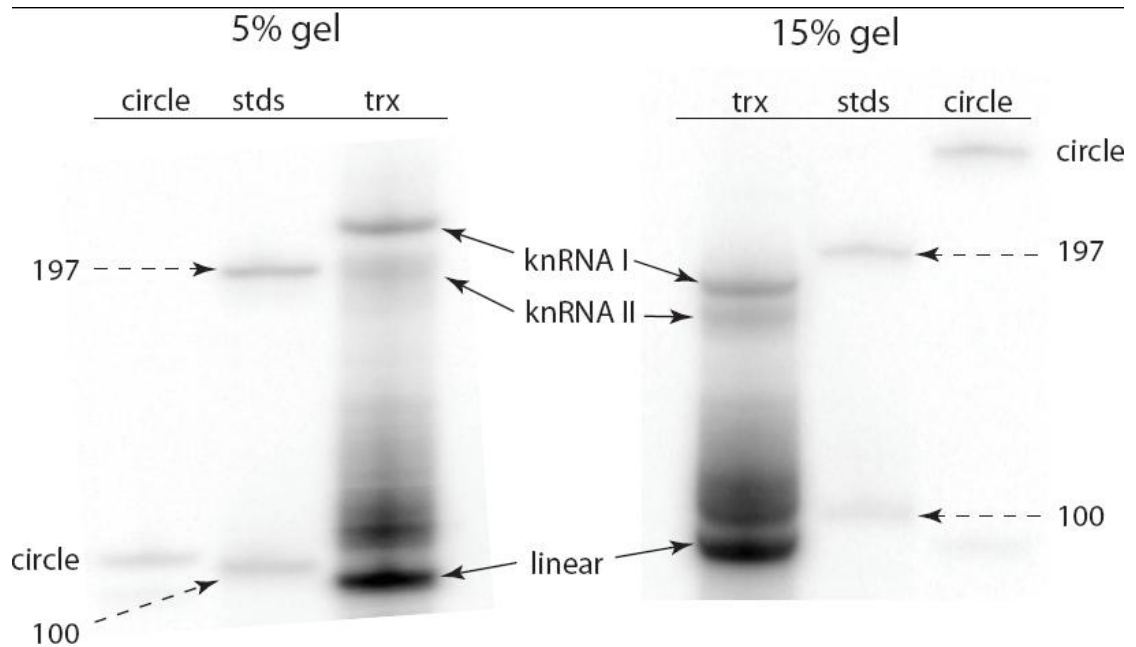

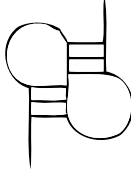
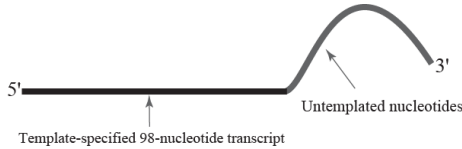
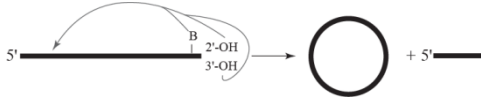
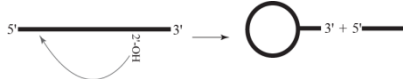
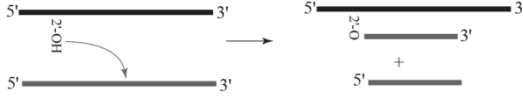
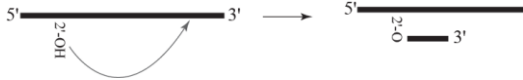
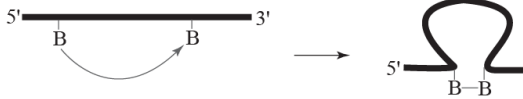
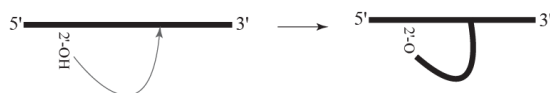


Figure 15. Formation of topological isomers (knRNAs) during *in vitro* transcription of the *Chlorella* virus *pdg* intron. A transcription reaction of the *pdg* intron DNA in the presence of α - 32 P•GTP was treated with DNase I, phenol extracted, and ethanol precipitated. Transcription products (*trx*) were electrophoresed on denaturing gels (8 M urea), with either 5% (left) or 15% (right) polyacrylamide alongside linear RNA molecules of 100 and 197 nucleotides (*stds*), and a 98-nucleotide circular molecule generated from the 98-nucleotide linear transcript (*circle*). The predominant knRNAs have been labeled as I and II.

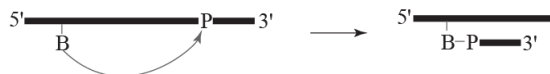
Hypothesis for anomalous gel mobility	Example	Results that refute it
Molecule with open knot		None
Persistent secondary structure or pseudoknot		knRNAs persist after glyoxal treatment, incubation in 95% formamide at 99°C or cycles in 95% formamide at 99°C and room temperature followed by electrophoresis on 8 M urea / 40% formamide polyacrylamide gels
Linear overextension artifact	 5' ————— 3' ↑ Template-specified 98-nucleotide transcript Untemplated nucleotides	Migration rate changes relative to linear standards on gels of differing acrylamide composition
Circular RNA from 3'-nucleotide attack (either 2'-hydroxyl or 3'-hydroxyl or nucleobase) on an upstream site	 5' ————— 3' 2'-OH 3'-OH B	The mobilities of the knRNAs on denaturing gels of different polyacrylamide composition does not match that of the 98-nucleotide circular control
Lariat from internal 2'-hydroxyl attacking at an upstream site	 5' ————— 3' 2'-OH	knRNAs are unaffected by yeast lariat debranching enzyme (Dbr I)
Branched RNA from an intermolecular transesterification reaction	 5' ————— 3' 2'-OH 5' ————— 3' 5' ————— 3' 2'-O 5'	knRNAs are unaffected by Dbr I
Branched RNA from an internal 2'-hydroxyl attacking at a downstream site	 5' ————— 3' 2'-OH 5' ————— 3' 2'-O	knRNAs are unaffected by Dbr I
Lariat or branched molecule from nucleobase-nucleobase bond (<i>i.e.</i> uracil or cytosine dimer)	 5' ————— 3' B B 5' ————— 3' B—B	knRNAs were not be reversed by exposure to UV light

Lariat or branched molecule from intramolecular phosphotriester



knRNAs are not base labile

Lariat or branched molecule from nucleobase-phosphate backbone bond



knRNAs are not hydrolyzed under acidic conditions

Figure 16. Possible topological explanations for the mobility shifts observed in the knRNAs. All of the listed possibilities were evaluated experimentally. All listed hypotheses other than open knotting were discounted.

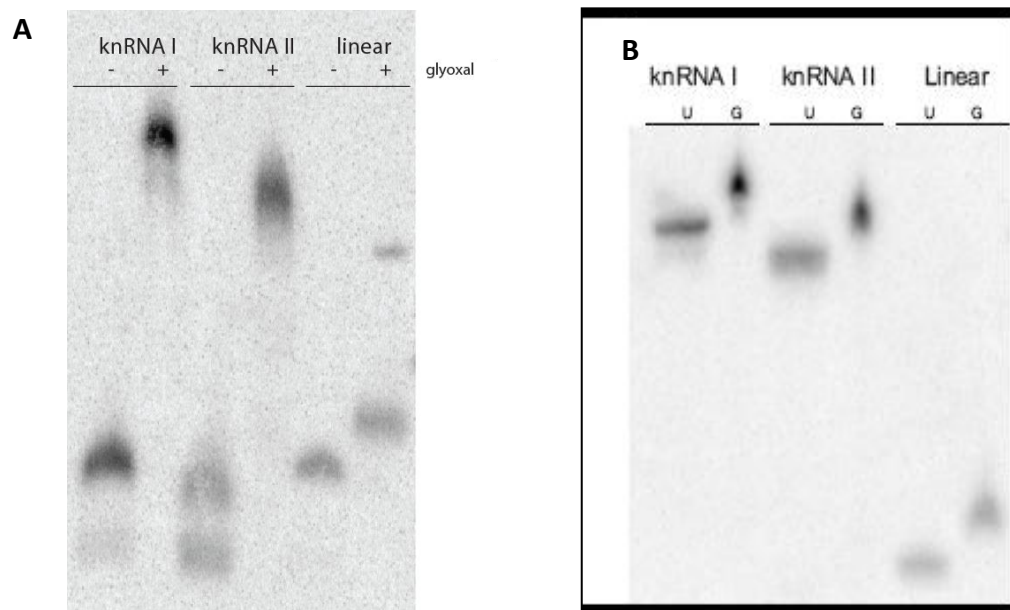


Figure 17. Gel electrophoresis of glyoxal-treated linear and knRNAs alongside their non-glyoxylated counterparts. A) A non-denaturing 6% polyacrylamide gel demonstrating the effect of glyoxal on knRNAs. The ability of glyoxal to denature the linear and knRNA molecules is demonstrated based on the extreme mobility shifts observed for each knRNA form in the absence (-) or presence (+) of glyoxal, which chemically modifies the RNA base-pairing surface. Note that no conversion between the knRNA forms and the linear is observed in the presence of glyoxal, indicating that the knRNAs are not due to persistent secondary structure in the linear form. B) Chemical denaturing of knRNAs with glyoxal. The knRNAs were either treated with glyoxal loading dye at 50°C for 60 min (G) or urea-containing loading dye at 90°C for 5 min (U), prior to loading on an 8% polyacrylamide / 8M urea gel.

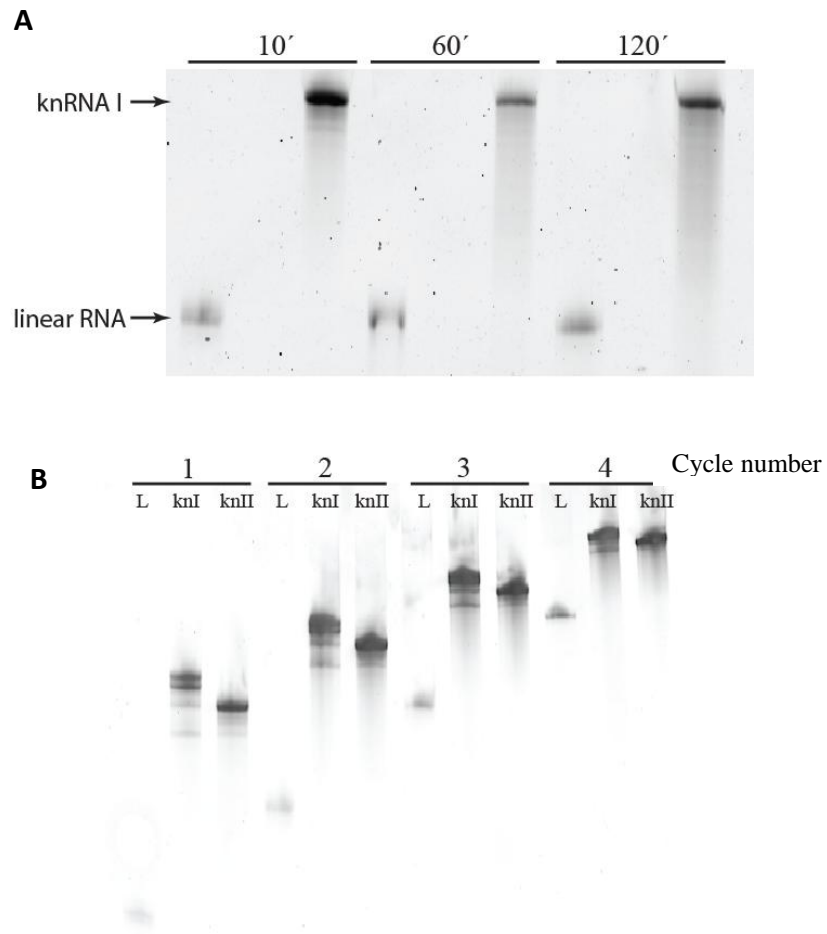


Figure 18. Denaturing gels of knRNAs with formamide. A) knRNA I was incubated at 99°C for the designated time in 95% formamide, then electrophoresed on an 8% polyacrylamide / 8M urea / 40% formamide gel at 50°C. B) knRNAs I and II and the linear molecule were subjected to 5 min at 99°C (1) followed by cycles of 5 min at room temperature and 5 min at 99 °C (2–4). The 8% polyacrylamide / 8M urea / 40% formamide gel was run continuously between cycles.



Figure 19. Debranching assay of knRNAs. Gel-purified 5'-radiolabeled knRNAs I and II were incubated with yeast debranching enzyme (Dbr I). The reaction products were separated on a 20% polyacrylamide / 8M urea gel alongside the unreacted linear molecule. To verify that the Dbr I was active, a 37-nucleotide branched RNA was also treated (three left-most lanes). The branched molecule has been denoted by the sideways 'Y' and the linear as '-'.

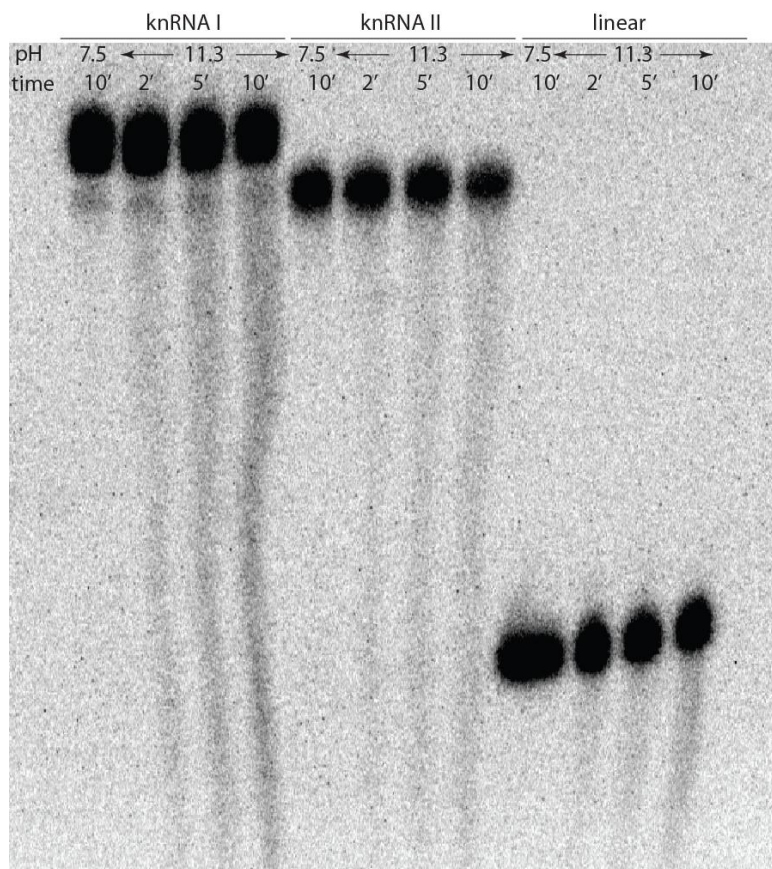


Figure 20. Alkaline hydrolysis of knRNAs to check for base-labile bonds. 5'-radiolabeled knRNAs were incubated at 50°C for up to 10 min in either 50 mM Tris (pH 7.5) or 50 mM K₂PO₄ buffer (pH 11.3). The time and pH of the reaction are indicated above each lane. Products were separated by electrophoresis through a 5% polyacrylamide / 8M urea gel.

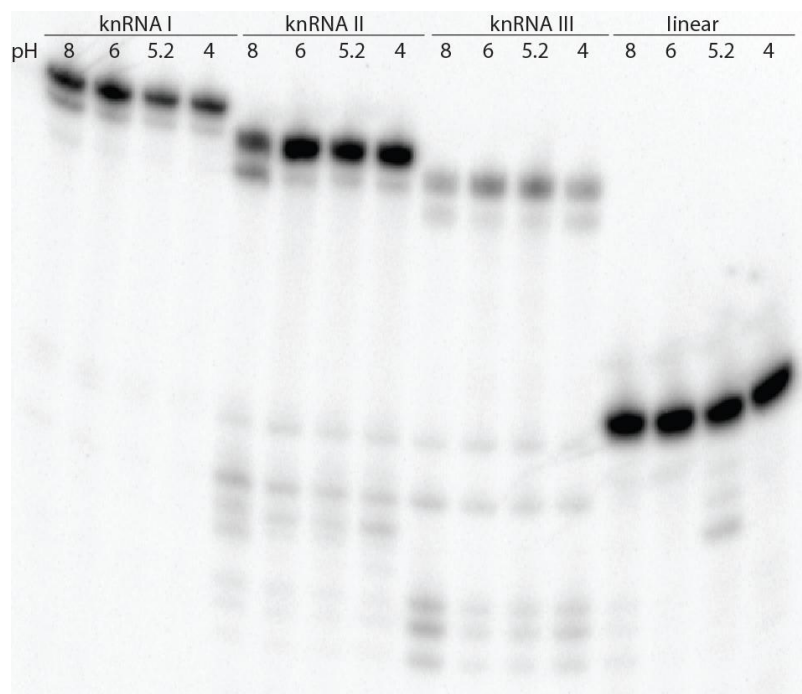


Figure 21. Acid hydrolysis of knRNAs to check for the presence of acid-labile bonds. The pH of each reaction mixture is listed above the respective lane. 5'-radiolabeled knRNAs and the linear molecule were incubated at room temperature for 21 hrs in solutions containing 30 mM Na⁺ ions, and 30 mM of buffer [Tris (pH 8.0), PIPES (pH 6.0), acetate (pH 5.2)] or 0.1 mM HCl (pH 4). Products were separated by electrophoresis on an 8% polyacrylamide / 8M urea gel.

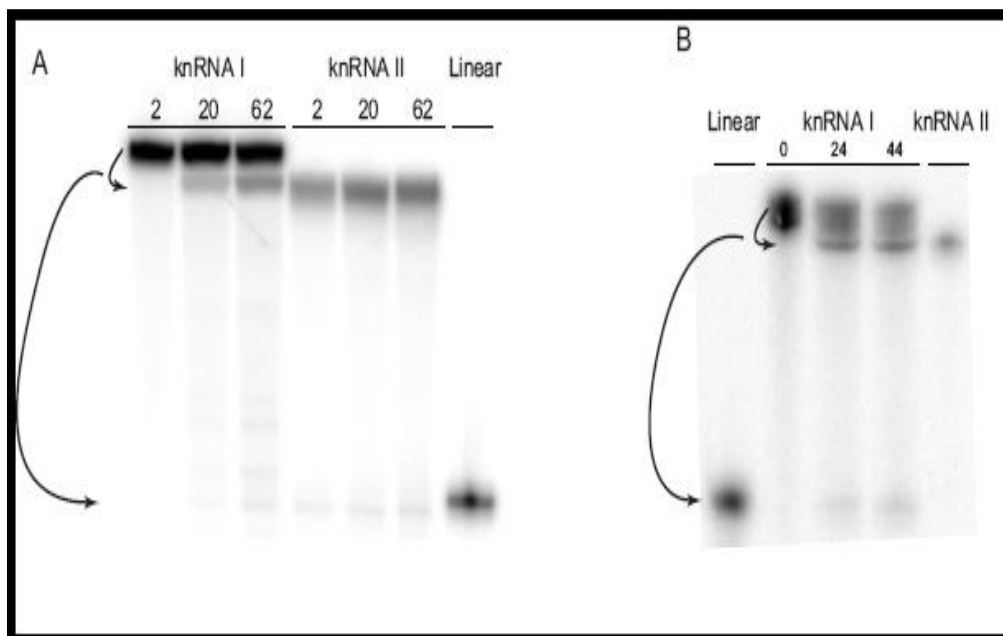


Figure 22. Unknotting versus denaturation of knRNAs. A) Incubation of the knRNAs in 300 mM sodium acetate (pH 5.2) at room temperature with agitation leads to unknotting of the knRNAs, in the order: knRNA I \rightarrow knRNA II \rightarrow Linear molecule. The time in hours is listed above each lane. B) Interconversion is also achieved when the knRNAs are incubated at 37°C in the presence of transcription buffer (pH 7.5). The time in hours is listed above each lane.

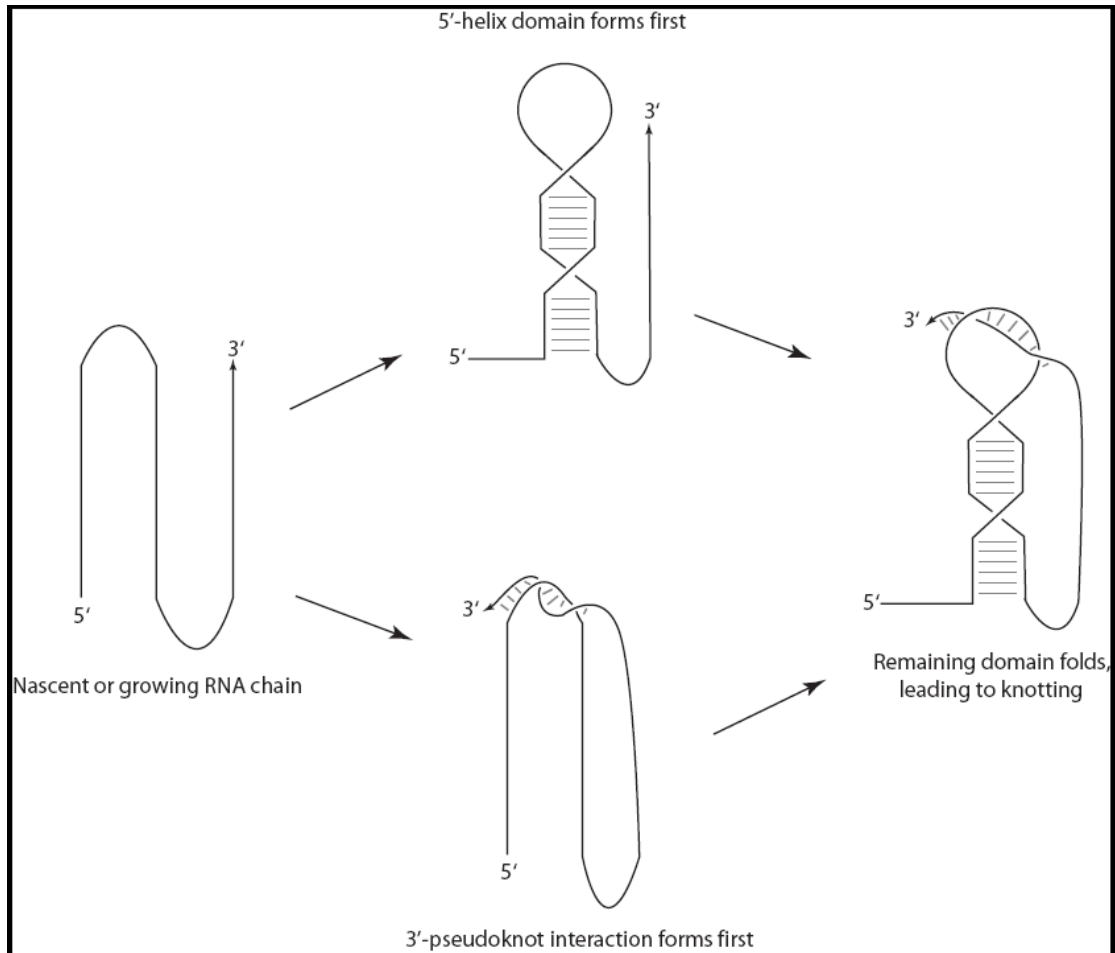


Figure 23. Proposed folding pathways leading to knotting by pseudoknot formation. Basepairing is modeled from the predicted secondary structure. From a nascent unfolded state, either of the two helical domains involved in the pseudoknot can form (top and bottom routes). In an A-form helix, every ~ 5.5 basepairs causes a crossing of the two RNA strands; hence, either depicted path leads to the same knotted structure with four strand crossings. While strand threading is likely disfavored entropically and electrostatically, the resulting basepairs provide an energetic offset, potentially to allow knotting.

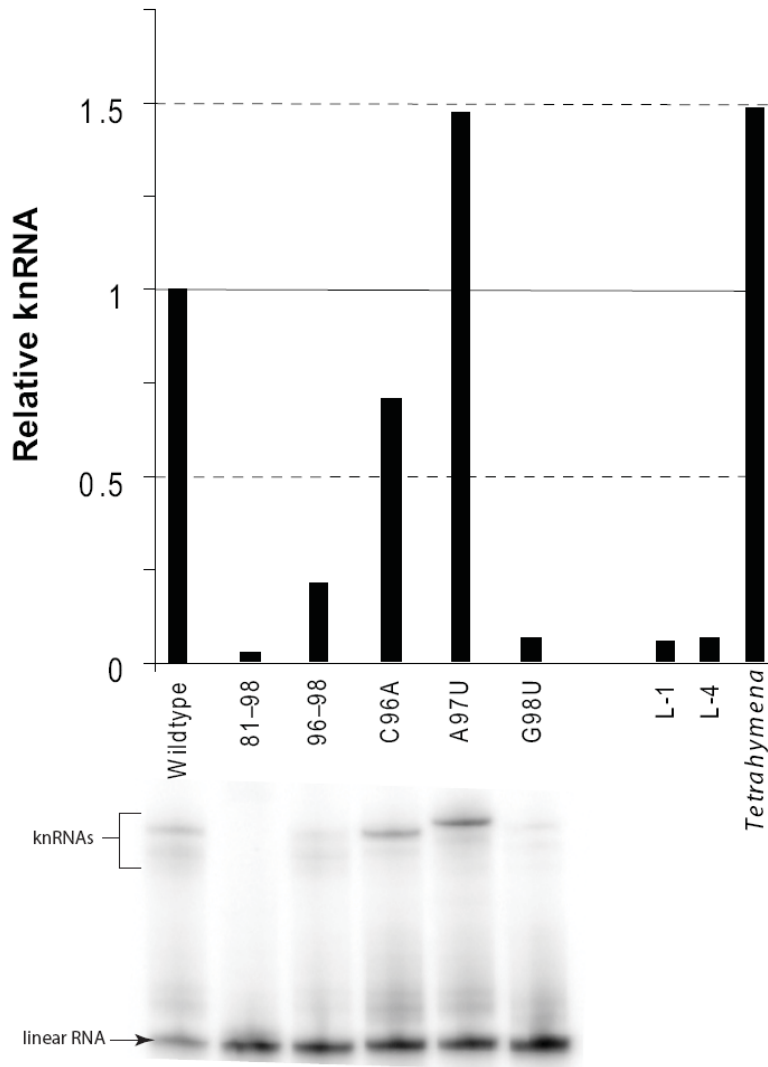


Figure 25. Effects of 3'-end mutations on knRNA formation. The wildtype and listed mutant sequences were transcribed in the presence of γ - ^{32}P •GTP and separated by electrophoresis through 8% polyacrylamide / 8M urea gels (transcription reactions of the wildtype sequence and 5 mutants are shown on the gel at the bottom). The 81-98 and 96-98 lanes contain mutations to those nucleotides in the RNA sequence; L-1 and L-4 RNAs have the last one and four nucleotides removed from their DNA templates. The fraction of knRNA for each reaction was determined; this value was then divided by the amount formed by the wildtype sequence to give relative knRNA (chart at the top); for comparison, the amount of knRNAs formed by the *Tetrahymena cnjB* intron has been included.

References

- Ban N, Nissen P, Hansen J, Moore PB and Steitz TA. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, 289(5481), 905–920.
- Basu S, Rambo RP, Strauss-Soukup J, Cate JH, Ferré-D'Amaré AR, Strobel SA and Doudna JA. 1998. A specific monovalent metal ion integral to the AA platform of the RNA tetraloop receptor. *Nature Structural Biology*, 5(11), 986–992.
- Beaudry AA and Joyce GF. 1992. Directed evolution of an RNA enzyme. *Science* 257(5070), 635–641.
- Bell MA, Johnson AK and Testa SM. 2002. Ribozyme-catalyzed excision of targeted sequences from within RNAs. *Biochemistry*, 41(51), 15327–15333.
- Benner SA, Ellington AD and Tauer A. 1989. Modern metabolism as a palimpsest of the RNA world. *Proceedings of the National Academy of Sciences of the United States of America*, 86(18), 7054–7058.
- Boussau B, Blanquart S, Nesculea A, Lartillot N and Gouy M. 2008. Parallel adaptations to high temperatures in the Archaean eon. *Nature*, 456(7224), 942–945.
- Burke DH and Willis JH. 1998. Recombination, RNA evolution, and bifunctional RNA molecules isolated through chimeric SELEX. *RNA*, 4(9), 1165–1175.
- Burton AS and Lehman N. 2006. Calcium(II)-dependent catalytic activity of the Azoarcus ribozyme: testing the limits of resolution for in vitro selection. *Biochimie*, 88(7), 819–825.
- Burton AS and Lehman N. 2009. DNA Before Proteins? Recent Discoveries in Nucleic Acid Catalysis Strengthen the Case. *Astrobiology*, 9(1), 125–130.
- Burton AS and Lehman N. 2010. Enhancing the Prebiotic Relevance of a Set of Covalently Self-Assembling, Autorecombining RNAs Through In Vitro Selection. *Journal of Molecular Evolution*, 70(3), 233–241.
- Burton AS, Madix RA, Vaidya N, Riley CA, Hayden EJ, Chepetan A, Arenas CD, Larson BC and Lehman N. 2009. Gel purification of radiolabeled nucleic acids

- via phosphorimaging: Dip-N-Dot. *Analytical Biochemistry*, 388(2), 351–352.
- Cadwell RC and Joyce GF. 1992. Randomization of genes by PCR mutagenesis. *PCR Methods and Applications*, 2(1), 28–33.
- Cazenave C and Uhlenbeck OC. 1994. RNA template-directed RNA synthesis by T7 RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America*, 91(15), 6972–6976.
- Cech TR. 2000. Structural biology. The ribosome is a ribozyme. *Science*, 289(5481), 878–879.
- Chamberlin MJ, Nierman WC, Wiggs J and Neff N. 1979. A quantitative assay for bacterial RNA polymerases. *The Journal of Biological Chemistry*, 254(20), 10061–10069.
- Chapman KB and Boeke JD. 1991. Isolation and characterization of the gene encoding yeast debranching enzyme. *Cell*, 65(3), 483–492.
- Chen X, Li N and Ellington AD. 2007. Ribozyme catalysis of metabolism in the RNA world. *Chemistry & Biodiversity*, 4(4), 633–655.
- Coleman TM and Huang F. 2002. RNA-catalyzed thioester synthesis. *Chemistry & Biology*, 9(11), 1227–1236.
- Coppins RL and Silverman SK. 2004. A DNA enzyme that mimics the first step of RNA splicing. *Nature Structural & Molecular Biology*, 11(3), 270–274.
- Crick FH. 1968. The origin of the genetic code. *Journal of Molecular Biology* 38, 367–79.
- Csank C, Taylor FM and Martindale DW. 1990. Nuclear pre-mRNA introns: analysis and comparison of intron sequences from *Tetrahymena thermophila* and other eukaryotes. *Nucleic Acids Research*, 18(17), 5133–5141.
- Dickson E, Prenskey W and Robertson HD. 1975. Comparative studies of two viroids: analysis of potato spindle tuber and citrus exocortis viroids by RNA fingerprinting and polyacrylamide-gel electrophoresis. *Virology*, 68(2), 309–316.
- Dotson PP, Johnson AK and Testa SM. 2008. *Tetrahymena thermophila* and *Candida albicans* group I intron-derived ribozymes can catalyze the trans-excision-

- splicing reaction. *Nucleic Acids Research*, 36(16), 5281–5289.
- Dunn JJ and Studier FW. 1983. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *Journal of Molecular Biology*, 166(4), 477–535.
- Dworkin JP, Lazcano A and Miller SL. 2003. The roads to and from the RNA world. *Journal of Theoretical Biology*, 222(1), 127–134.
- Eigen M. 1971. Selforganization of matter and the evolution of biological macromolecules. *Die Naturwissenschaften*, 58(10), 465–523.
- Ekland EH and Bartel, DP. 1996. RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature*, 383(6596), 192.
- Elkins PA, Watts JM, Zalacain M, van Thiel A, Vitazka PR, Redlak M, Andraos-Selim C, Rastinejad F and Holmes WM. 2003. Insights into catalysis by a knotted TrmD tRNA methyltransferase. *Journal of Molecular Biology*, 333(5), 931–949.
- Engels J and Uhlmann E. 1988. Gene synthesis. *Advances in Biochemical Engineering/Biotechnology*, 37, 73–127.
- Ferris JP. 2002. Montmorillonite catalysis of 30-50 mer oligonucleotides: laboratory demonstration of potential steps in the origin of the RNA world. *Origins of Life and Evolution of the Biosphere*, 32(4), 311–332.
- Freeland SJ, Knight RD and Landweber LF. 1999. Do Proteins Predate DNA? *Science*, 286(5440), 690–692.
- Fusz S, Eisenführ A, Srivatsan SG, Heckel A and Famulok M. 2005. A ribozyme for the aldol reaction. *Chemistry & Biology*, 12(8), 941–950.
- Gesteland R, Cech TR and Atkins J. eds. 2006. *The RNA World* 3rd ed., New York: Cold Spring Harbor Press.
- Gilbert W. 1986. Origin of life: The RNA world. *Nature*, 319(6055), 618.
- Gould SJ and Vrba ES. 1982. Exaptation—A Missing Term in the Science of Form. *Paleobiology*, 8(1), 4–15.
- Grabowski PJ, Padgett RA and Sharp PA. 1984. Messenger RNA splicing in vitro: an

- excised intervening sequence and a potential intermediate. *Cell*, 37(2), 415–427.
- Hayden EJ and Lehman N. 2006. Self-assembly of a group I intron from inactive oligonucleotide fragments. *Chemistry & Biology*, 13(8), 909–918.
- Hayden EJ, Riley CA, Burton AS and Lehman N. 2005. RNA-directed construction of structurally complex and active ligase ribozymes through recombination. *RNA*, 11(11), 1678–1687.
- He A, Li T, Daniels L, Fotheringham I and Rosazza P. 2004. Nocardia sp. carboxylic acid reductase: cloning, expression, and characterization of a new aldehyde oxidoreductase family. *Applied and Environmental Microbiology*, 70(3), 1874–1881.
- Heilman-Miller SL and Woodson SA, 2003. Effect of transcription on folding of the Tetrahymena ribozyme. *RNA*, 9(6), 722–733.
- Heine A, DeSantis G, Luz JG, Mitchell M, Wong CH and Wilson IA. 2001. Observation of covalent intermediates in an enzyme mechanism at atomic resolution. *Science*, 294(5541), 369–374.
- Hetzer M, Schweyen RJ and Mueller MW. 1997. DNA polymerization catalysed by a group II intron RNA in vitro. *Nucleic Acids Research*, 25(9), 1825–1829.
- Huang Z and Szostak JW. 1996. A simple method for 3'-labeling of RNA. *Nucleic Acids Research*, 24(21), 4360–4361.
- Huang Z, Pei W, Han Y, Jayaseelan S, Shekhtman A, Shi H and Niu L. 2009. One RNA aptamer sequence, two structures: a collaborating pair that inhibits AMPA receptors. *Nucleic Acids Research*, 37(12), 4022–4032.
- Jacquier A and Rosbash M. 1986. RNA splicing and intron turnover are greatly diminished by a mutant yeast branch point. *Proceedings of the National Academy of Sciences of the United States of America*, 83(16), 5835–5839.
- Jadhav VR and Yarus M. 2002. Acyl-CoAs from coenzyme ribozymes. *Biochemistry*, 41(3), 723–729.
- Jeffares DC, Poole AM and Penny D. 1998. Relics from the RNA world. *Journal of Molecular Evolution*, 46(1), 18–36.

- Johnston WK, Unrau PJ, Lawrence MS, Glasner ME and Bartel DP. 2001. RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science*, 292(5520), 1319–1325.
- Joyce GF. 1991. The rise and fall of the RNA world. *The New Biologist*, 3(4), 399–407.
- Joyce GF. 1998. Nucleic acid enzymes: playing with a fuller deck. *Proceedings of the National Academy of Sciences of the United States of America*, 95(11), 5845–5847.
- Joyce GF. 2002. The antiquity of RNA-based evolution. *Nature*, 418(6894), 214–221.
- King NP, Yeates EO and Yeates TO. 2007. Identification of rare slipknots in proteins and their implications for stability and folding. *Journal of Molecular Biology*, 373(1), 153–166.
- Kolesov G, Virnau P, Kardar M and Mirny LA. 2007. Protein knot server: detection of knots in protein structures. *Nucleic Acids Research*, 35, W425–428.
- Kuhn H. 1972. Self-organization of molecular systems and evolution of the genetic apparatus. *Angewandte Chemie (International Edition in English)*, 11(9), 798–820.
- Kumar RK and Yarus M. 2001. RNA-catalyzed amino acid activation. *Biochemistry*, 40(24), 6998–7004.
- Kun A, Santos M & Szathmáry E. 2005. Real ribozymes suggest a relaxed error threshold. *Nature Genetics*, 37(9), 1008–1011.
- Lai YL, Yen SC, Yu SH and Hwank JK. 2007. pKNOT: the protein KNOT web server. *Nucleic Acids Research*, 35, W420–424.
- Lau MWL, Cadieux KEC and Unrau PJ. 2004. Isolation of fast purine nucleotide synthase ribozymes. *Journal of the American Chemical Society*, 126(48), 15686–15693.
- Lau MWL and Unrau PJ. 2009. A promiscuous ribozyme promotes nucleotide synthesis in addition to ribose chemistry. *Chemistry & Biology*, 16(8), 815–825.
- Lehman N and Joyce GF. 1993. Evolution in vitro of an RNA enzyme with altered

- metal dependence. *Nature*, 361(6408), 182–185.
- Lehman N. 2003. A case for the extreme antiquity of recombination. *Journal of Molecular Evolution*, 56(6), 770–777.
- Lehman N. 2008. A recombination-based model for the origin and early evolution of genetic information. *Chemistry & Biodiversity*, 5(9), 1707–1717.
- Letsinger RL, Wilkes JS and Dumas LB. 1972. Enzymatic synthesis of polydeoxyribonucleotides possessing internucleotide phosphoramidate bonds. *Journal of the American Chemical Society*, 94(1), 292–293.
- Li N and Huang F. 2005. Ribozyme-catalyzed aminoacylation from CoA thioesters. *Biochemistry*, 44(11), 4582–4590.
- Lim K, Zhang H, Tempczyk A, Krajewski W, Bonander N, Toedt J, Howard A, Eisenstein E and Herzberg O. 2003. Structure of the YibK methyltransferase from *Haemophilus influenzae* (HI0766): a cofactor bound at a site formed by a knot. *Proteins*, 51(1), 56–67.
- Liu LF, Davis JL and Calendar R. 1981. Novel topologically knotted DNA from bacteriophage P4 capsids: studies with DNA topoisomerases. *Nucleic Acids Research*, 9(16), 3979–3989.
- Lohse PA and Szostak JW. 1996. Ribozyme-catalysed amino-acid transfer reactions. *Nature*, 381(6581), 442–444.
- Lorsch JR and Szostak JW. 1994. In vitro evolution of new ribozymes with polynucleotide kinase activity. *Nature*, 371(6492), 31–36.
- Mallam AL, Onuoha SC, Grossmann JG and Jackson SE. 2008. Knotted fusion proteins reveal unexpected possibilities in protein folding. *Molecular Cell*, 30(5), 642–648.
- Martindale DW and Taylor FM. 1988. Multiple introns in a conjugation-specific gene from *Tetrahymena thermophila*. *Nucleic Acids Research*, 16(5), 2189–2201.
- McGinness KE and Joyce GF. 2003. In search of an RNA replicase ribozyme. *Chemistry & Biology*, 10(1), 5–14.
- Ménissier J, de Murcia G, Lebeurier G and Hirth L. 1983. Electron microscopic studies of the different topological forms of the cauliflower mosaic virus DNA:

- knotted encapsidated DNA and nuclear minichromosome. *The EMBO Journal*, 2(7), 1067–1071.
- Nissen P, Hansen J, Ban N, Moore PB and Steitz TA. 2000. The structural basis of ribosome activity in peptide bond synthesis. *Science*, 289(5481), 920–930.
- Oberhuber M. and Joyce GF. 2005. A DNA-templated aldol reaction as a model for the formation of pentose sugars in the RNA world. *Angewandte Chemie (International Edition in English)*, 44(46), 7580–7583.
- Orgel LE. 1968. Evolution of the genetic apparatus. *Journal of Molecular Biology* 38, 381–93.
- Orgel LE. 1986. RNA catalysis and the origins of life. *Journal of Theoretical Biology*, 123(2), 127–149.
- Padgett RA, Konarska MM, Grabowski PJ, Hardy SF and Sharp PA. 1984. Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. *Science*, 225(4665), 898–903.
- Pan J. and Woodson SA. 1998. Folding intermediates of a self-splicing RNA: mispairing of the catalytic core. *Journal of Molecular Biology*, 280(4), 597–609.
- Pan T, Artsimovitch I, Fang XW, Landick R and Sosnick TR. 1999. Folding of a large ribozyme during transcription and the effect of the elongation factor NusA. *Proceedings of the National Academy of Sciences of the United States of America*, 96(17), 9545–9550.
- Pan J, Deras ML and Woodson SA. 2000. Fast folding of a ribozyme by stabilizing core interactions: evidence for multiple folding pathways in RNA. *Journal of Molecular Biology*, 296(1), 133–144.
- Poole A, Jeffares D and Penny D. 1999. Early evolution: prokaryotes, the new kids on the block. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 21(10), 880–889.
- Poole A, Penny D and Sjöberg, B. 2000. Methyl-RNA: an evolutionary bridge between RNA and DNA? *Chemistry & Biology*, 7(12), R207–216.
- Portugal J and Rodríguez-Campos A. 1996. T7 RNA polymerase cannot transcribe through a highly knotted DNA template. *Nucleic Acids Research*, 24(24),

4890–4894.

Powner MW, Gerland B and Sutherland JD. 2009. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* 459, 171–172.

Raymer DM and Smith DE. 2007. Spontaneous knotting of an agitated string. *Proceedings of the National Academy of Sciences of the United States of America*, 104(42), 16432–16437.

Reichard P. 1993. From RNA to DNA, why so many ribonucleotide reductases? *Science*, 260(5115), 1773–1777.

Rich A. 1962. On the problems of evolution and biochemical information transfer. In M. Kasha & B. Pullmann, eds. *Horizons in biochemistry*. New York: Academic Press, 103–126.

Riley CA and Lehman N. 2003. Generalized RNA-directed recombination of RNA. *Chemistry & Biology*, 10(12), 1233–1243.

Rødland EA. 2006. Pseudoknots in RNA secondary structures: enumeration and prevalence. *Journal of Computational Biology*, 13, 1197–213.

Ruan J, Stormo GD and Zhang W. 2004. ILM: a web server for predicting RNA secondary structures with pseudoknots. *Nucleic Acids Research*, 32, W146–149.

Ruskin, B, Krainer AR, Maniatis T and Green MR.. 1984. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell*, 38(1), 317–331.

Ruskin B and Green MR. 1985. An RNA processing activity that debranches RNA lariats. *Science*, 229(4709), 135–140.

Rydzanicz R, Zhao XS and Johnson PE. 2005. Assembly PCR oligo maker: a tool for designing oligodeoxynucleotides for constructing long DNA molecules for RNA production. *Nucleic Acids Research*, 33, W521–525.

Saito H, Kourouklis D and Suga H. 2001. An in vitro evolved precursor tRNA with aminoacylation activity. *The EMBO Journal*, 20(7), 1797–1806.

Schmitt T and Lehman N. 1999. Non-unity molecular heritability demonstrated by continuous evolution in vitro. *Chemistry & Biology*, 6(12), 857–869.

- Sclavi B, Sullivan M, Chance MR, Brenowitz M and Woodson SA. 1998. RNA Folding at Millisecond Intervals by Synchrotron Hydroxyl Radical Footprinting. *Science*, 279(5358), 1940–1943.
- Semancik JS, Morris TJ and Weathers LG. 1973. Structure and conformation of low molecular weight pathogenic RNA from exocortis disease. *Virology*, 53(2), 448–456.
- Sen D and Gilbert W. 1988. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, 334(6180), 364–366.
- Shishido K, Komiyama N and Ikawa S. 1987. Increased production of a knotted form of plasmid pBR322 DNA in Escherichia coli DNA topoisomerase mutants. *Journal of Molecular Biology*, 195(1), 215–218.
- Striggles JC, Martin MB and Schmidt FJ. 2006. Frequency of RNA-RNA interaction in a model of the RNA World. *RNA*, 12(3), 353–359.
- Sun L, Cui Z, Gottlieb RL and Zhang B. 2002. A selected ribozyme catalyzing diverse dipeptide synthesis. *Chemistry & Biology*, 9(5), 619–628.
- Sun L, Li Y, McCullough AK, Wood TG, Lloyd RS, Adams B, Gurnon JR and Van Etten JL. 2000. Intron conservation in a UV-specific DNA repair gene encoded by chlorella viruses. *Journal of Molecular Evolution*, 50(1), 82–92.
- Sun X, Harder J, Krook M, Jörnvall H, Sjöberg BM and Reichard P. 1993. A possible glycine radical in anaerobic ribonucleotide reductase from Escherichia coli: nucleotide sequence of the cloned nrdD gene. *Proceedings of the National Academy of Sciences of the United States of America*, 90(2), 577–581.
- Szostak J and Ellington A. 1993. In Vitro Selection of Functional RNA Sequences. In R. Gesteland & J. Atkins, eds. *The RNA World*. Cold Spring Harbor Press, pp. 511–533.
- Takeuchi N, Poorthuis PH and Hogeweg P. 2005. Phenotypic error threshold; additivity and epistasis in RNA evolution. *BMC Evolutionary Biology*, 5(1), 9.
- Takusagawa F and Kamitori S. 1996. A Real Knot in Protein. *Journal of the American Chemical Society*, 118(37), 8945–8946.

- Treiber DK and Williamson JR. 2001. Beyond kinetic traps in RNA folding. *Current Opinion in Structural Biology*, 11(3), 309–314.
- Tsukiji S, Pattnaik SB and Suga H. 2003. An alcohol dehydrogenase ribozyme. *Nature Structural Biology*, 10(9), 713–717.
- Tsukiji S, Pattnaik SB and Suga H. 2004. Reduction of an aldehyde by a NADH/Zn²⁺-dependent redox active ribozyme. *Journal of the American Chemical Society*, 126(16), 5044–5045.
- Unrau PJ and Bartel DP. 1998. RNA-catalysed nucleotide synthesis. *Nature*, 395(6699), 260–263.
- Valadkhan S and Manley JL. 2001. Splicing-related catalysis by protein-free snRNAs. *Nature*, 413(6857), 701–707.
- Valadkhan S and Manley JL. 2003. Characterization of the catalytic activity of U2 and U6 snRNAs. *RNA*, 9(7), 892–904.
- Van Loock MS, Harris BA and Harvey SC. 1998. To knot or not to knot? Examination of 16S ribosomal RNA models. *Journal of Biomolecular Structure & Dynamics*, 16(3), 709–713.
- Vartanian JP, Henry M and Wain-Hobson S. 1996. Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions. *Nucleic Acids Research*, 24(14), 2627–2631.
- Venkitasubramanian P, Daniels L and Rosazza JPN. 2007. Reduction of carboxylic acids by *Nocardia* aldehyde oxidoreductase requires a phosphopantetheinylated enzyme. *The Journal of Biological Chemistry*, 282(1), 478–485.
- Virnau P, Mirny LA and Kardar M. 2006. Intricate knots in proteins: Function and evolution. *PLoS Computational Biology*, 2(9), e122.
- Vogel J, Hess WR and Börner T. 1997. Precise branch point mapping and quantification of splicing intermediates. *Nucleic Acids Research*, 25(10), 2030–2031.
- Wadkins TS and Been MD. 2002. Ribozyme activity in the genomic and antigenomic RNA strands of hepatitis delta virus. *Cellular and Molecular Life Sciences: CMLS*, 59(1), 112–125.

- Wang H, Di Gate RJ and Seeman NC. 1996. An RNA topoisomerase. *Proceedings of the National Academy of Sciences of the United States of America*, 93(18), 9477–9482.
- Wang H, Du SM and Seeman NC. 1993. Tight single-stranded DNA knots. *Journal of Biomolecular Structure & Dynamics*, 10(5), 853–863.
- Wasserman SA, Dungan JM and Cozzarelli NR. 1985. Discovery of a predicted DNA knot substantiates a model for site-specific recombination. *Science*, 229(4709), 171–174.
- White HB III. 1976. Coenzymes as fossils of an earlier metabolic state. *Journal of Molecular Evolution*, 7, 101–104.
- Wiegand TW, Janssen RC and Eaton BE. 1997. Selection of RNA amide synthases. *Chemistry & Biology*, 4(9), 675–683.
- Williamson JR, Raghuraman MK and Cech TR. 1989. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell*, 59(5), 871–880.
- Woese C. 1967. The genetic code. Chapter 7, Harper and Row, New York.
- Wolf YI and Koonin EV. 2007. On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization. *Biology Direct*, 2, 14.
- Yoshizawa K. 1998. Two-step concerted mechanism for alkane hydroxylation on the ferryl active site of methane monooxygenase. *Journal of Biological Inorganic Chemistry*, 3(3), 318–324.
- Zaher HS and Unrau PJ. 2006. A general RNA-capping ribozyme retains stereochemistry during cap exchange. *Journal of the American Chemical Society*, 128(42), 13894–13900.
- Zaher HS and Unrau PJ. 2007. Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA*, 13(7), 1017–1026.
- Zarrinkar P and Williamson JR. 1994. Kinetic intermediates in RNA folding. *Science*, 265(5174), 918–924.
- Zhang B and Cech TR. 1997. Peptide bond formation by in vitro selected ribozymes. *Nature*, 390(6655), 96–100.

- Zhang B and Cech TR. 1998. Peptidyl-transferase ribozymes: trans reactions, structural characterization and ribosomal RNA-like features. *Chemistry & Biology*, 5(10), 539–553.
- Zhang L, Xiao M, Lu C and Zhang Y. 2005. Fast formation of the P3-P7 pseudoknot: a strategy for efficient folding of the catalytically active ribozyme. *RNA*, 11(1), 59–69.
- Zhirnov OV and Wollenzien P. 2003. Action spectra for UV-light induced RNA-RNA crosslinking in 16S ribosomal RNA in the ribosome. *Photochemical & Photobiological Sciences: Official Journal of the European Photochemistry Association and the European Society for Photobiology*, 2(6), 688–693.