

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Better ribozymes for trans-splicing and triphosphorylation

Permalink

<https://escholarship.org/uc/item/6rm4n2g5>

Author

Dolan, Gregory F.

Publication Date

2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Better ribozymes for *trans*-splicing and triphosphorylation

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Chemistry

by

Gregory F. Dolan

Committee in charge:

Professor Ulrich F. Müller, Chair
Professor Simpson Joseph
Professor Clifford P. Kubiak
Professor Jens Lykke-Andersen
Professor Navtej Toor

2015

Copyright
Gregory F. Dolan, 2015
All rights reserved.

The Dissertation of Gregory F. Dolan is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2015

Table of Contents

Signature page.....	iii
Table of Contents.....	iv
List of Figures	vi
List of Tables.....	viii
Acknowledgements	ix
Vita.....	xi
Abstract of the Dissertation	xii
Chapter 1 Introduction	1
1.1 Discovery of Catalytic RNA: a Brief Historical Perspective	1
1.2 Group I introns.....	2
1.3 Trans-splicing Group I Introns	7
1.4 Azoarcus Group I Intron	12
1.5 The RNA World	15
1.6 Prebiotic Triphosphorylation & the Triphosphorylation ribozyme	23
1.7 Goals of this Dissertation	27
1.8 References	28
Chapter 2 <i>Trans</i> -splicing with the group I intron ribozyme from <i>Azoarcus</i>	41
2.1 Abstract	41
2.2 Introduction.....	42
2.3 Results	46
2.4 Discussion	64
2.5 Materials and Methods	69
2.6 Acknowledgements	74
2.7 Supporting Information	76
2.8 References	82
Chapter 3 Ribozyme catalyzed triphosphorylation at low trimetaphosphate concentrations.....	88
3.1 Abstract	88
3.2 Introduction.....	89
3.3 Results	90
3.4 Discussion	99
3.5 Materials and Methods	103
3.6 Acknowledgements	108
3.7 Supporting Information	109
3.8 References	112
Chapter 4 Discussion & Future Directions	115

4.1	<i>Trans</i> -splicing with the <i>Azoarcus</i> Ribozyme.....	115
4.2	Triphosphorylation Ribozymes	121
4.3	References	126

List of Figures

Figure 1.1:	Cis-splicing mechanism of group I introns	3
Figure 1.2:	Secondary structure of <i>Tetrahymena</i> and <i>Azoarcus</i> ribozymes ...	5
Figure 1.3:	First <i>trans</i> -splicing group I intron design.....	7
Figure 1.4 :	Summary of <i>cis</i> -splicing vs. 3' <i>trans</i> -splicing	9
Figure 1.5:	Glycosidic bond formation between nucleoside and 5-phosphoryribosyl 1-pyrophosphate.....	19
Figure 1.6:	5' triphosphorylation of nucleosides with trimetaphosphate	25
Figure 1.7:	Secondary Structure of TPR1	26
Figure 2.1:	Secondary structure representations of <i>trans</i> -splicing group I intron ribozymes	44
Figure 2.2:	Experimental comparison of favored splice sites and <i>trans</i> -splicing efficiencies between <i>Tetrahymena</i> and <i>Azoarcus</i>	49
Figure 2.3:	'Classical design' for <i>trans</i> -splicing group I intron ribozymes	53
Figure 2.4:	'P1 extension design' for the <i>trans</i> -splicing <i>Azoarcus</i> ribozyme	55
Figure 2.5:	Effect of the 'anticodon stem design' on the in vitro <i>trans</i> -splicing efficiency with <i>Azoarcus</i> ribozymes.....	57
Figure 2.6:	Comparison of all three designs for the in vitro <i>trans</i> -splicing efficiency with the <i>Azoarcus</i> ribozyme on short substrates	60
Figure 2.7:	<i>Cis</i> -splicing efficiency of ribozyme constructs in <i>E. coli</i> cells.....	64
Figure S2.1:	Secondary structures of 'classical designs' for ribozyme 5'-terminal interactions.....	76

Figure S2.2:	Secondary structures of 'anticodon stem designs' for trans-splicing <i>Azoarcus</i> ribozyme	77
Figure S2.3:	Secondary structures of designed short substrate <i>Azoarcus</i> trans-splicing interactions	78
Figure S2.4:	Cis-splicing with the <i>Azoarcus</i> ribozyme testing the influence of the anticodon stem and the P10 helix interactions	79
Figure 3.1:	Proposed secondary structures for the two central triphosphorylation ribozyme variants in this study	91
Figure 3.2:	Triphosphorylation kinetics of central ribozyme variants in this study	93
Figure 3.3:	Base covariation experiments to test the formation of duplex P4.....	95
Figure 3.4:	Triphosphorylation kinetics of TPR1e under different conditions.....	97
Figure 3.5:	Determination of optimal TPR1e triphosphorylation conditions.....	99
Figure S3.1:	Diagram of doped selection performed.....	109
Figure S3.2:	Identification of mutations in isolated ribozyme clone 11 that are necessary and sufficient for full activity	110
Figure S3.3:	Position of mutations in the 58 clones isolated from the selection that had at least the same activity as the parent ribozyme TPR1	111
Figure S3.4:	Mutational load in the pool before and after selection	112
Figure 4.1:	Preferred <i>Trans</i> - and <i>Cis</i> -splicing Structural Contacts made by the <i>Azoarcus</i> ribozyme	116
Figure 4.2:	RNA and Analogue Structures.....	124
Figure 4.3:	Fluorescent Ligation Assay to measure triphosphorylation activity	126

List of Tables

Table 2.1:	Positions of all splice sites shown in Figure 2A	81
Table 2.2:	Ribozyme exon sequences at the splice sites 97, 177, and 258, for the three design principles.....	82

Acknowledgements

I would like to thank my family, friends, and co-workers who've been with me through this journey. In particular, I'd like to thank my soon to be wife, Kristin Matulich. Although there are too many things to list, I'd like to thank her for both her encouragement in my moments of self-doubt and support through my sometimes long hours of work. There is not a doubt in my mind that I could not have achieved what I have over the last five plus years without you. Thank you.

Mom & Dad. Support, unbridled love, and encouragement only begins to describe everything you've done for me. To this day, you both continue to be my role models. I love you.

To my family, thank you for being the constant in my life. I gain so much strength and self-confidence through the love you've always shown me.

A special thanks to my PI, Uli. Your dedication to science and investigation made it a pleasure to work in your lab. Thank you for your thoughtful mentorship through the years.

To all members of the Muller lab (past or present) Karen, Janina, Dario, Chengguo, Zhaleh, and Logan. Thank you for the helpful conversations, advice, and comic relief. I enjoyed working with all of you. May you all reach your personal and professional goals.

To the Toor, Joseph, and Yang Labs, thank you for the kind sharing of your lab equipment and supplies. The assistance was greatly appreciated.

Chapter 2, in full, is a reprint of the material as it appears in RNA Journal. **Dolan, GF** & Müller, UF (2014). *Trans*-splicing with the group I intron from *Azoarcus*. *RNA* 20(2):202-13. The dissertation author is the first author on this paper.

Chapter 3, in full, is currently being prepared for submission for publication of the material. **Dolan, GF**; Müller, UF. The dissertation author is the first author on this paper.

The research was funded in part by the Hellman Family Foundation and NASA. The support of the funding agencies is greatly appreciated.

Vita

University of California, Santa Barbara

2009 Bachelor of Science in Biochemistry

University of California, San Diego

2011 Master of Science in Chemistry

2015 Doctor of Philosophy in Chemistry

PUBLICATIONS

Meluzzi, D, Olson, KE, **Dolan, GF**, Arya, G & Müller, UF (2012). Computational prediction of efficient splice sites for *trans*-splicing ribozymes. *RNA* 18(3):590-602.

Olson, KE, **Dolan, GF** & Müller, UF (2014). In vivo Evolution of Catalytic RNA couples *trans*-splicing to translation. *PLoS One* 9(1):e86473.

Dolan, GF & Müller, UF (2014). *Trans*-splicing with the group I intron from *Azoarcus*. *RNA* 20(2):202-13.

Abstract of the Dissertation

Better ribozymes for *trans*-splicing and triphosphorylation

by

Gregory F. Dolan

Doctor of Philosophy in Chemistry

University of California, San Diego, 2015

Professor Ulrich F. Müller, Chair

Catalytic RNAs (ribozymes) are RNAs that can catalyze a chemical reaction without the use of protein cofactors. This dissertation centered on improving the activities for two different ribozymes: a *trans*-splicing group I intron ribozyme and a triphosphorylation ribozyme. *Trans*-splicing group I intron ribozymes can be used in gene therapy and synthetic biology applications, and this work focused on improving the *trans*-splicing activity of the group I intron ribozyme from *Azoarcus*, a small and robustly folding ribozyme located in the intervening sequence of pre-tRNA^{Ile}. These results showed that although the *trans*-splicing *Azoarcus* ribozyme could not mediate a growth phenotype in *E. coli* cells, the *Azoarcus* ribozyme is indeed capable of catalyzing the *in vitro trans*-

splicing reaction as efficiently as other *trans*-splicing ribozymes previously tested. The most efficient *trans*-splicing *Azoarcus* ribozymes contained secondary structural contacts that mimicked the contacts made by the natural *cis*-splicing ribozyme, suggesting that the natural splicing interactions are the preferred *trans*-splicing interactions for the *Azoarcus* ribozyme. The second ribozyme improved in this dissertation is the triphosphorylation ribozyme, a ribozyme that triphosphorylates the 5'-hydroxyl of RNAs using trimetaphosphate as a substrate. This ribozyme's activity is important in the context of the RNA world and in the origin of life because RNA polymerization is an energetically unfavorable process in an aqueous environment, and the 5'-triphosphate can supply the energetic driving force for RNA polymerization. A previous graduate student in the Müller laboratory performed an *in vitro* selection identifying the first triphosphorylation ribozymes. The work herein improved the activity of one of the most active original triphosphorylation ribozymes by performing a 'doped selection.' A 'doped selection' is a method to identify sequence variants with higher activity by selecting for the best ribozymes from a partially randomized pool consisting of sequence variants of the original ribozyme. This selection yielded a ribozyme with seven mutations and a catalytic rate that was 24-fold higher than the first generation ribozyme. Unlike the original ribozyme, this new ribozyme has measurable activity at prebiotically relevant trimetaphosphate concentrations and has a catalytic rate of 6.8 min^{-1} at optimal conditions.

Chapter 1

Introduction

1.1 Discovery of Catalytic RNA: a Brief Historical Perspective

The understanding of the biological functions of RNA has grown dramatically since the 1950s. In 1958, Francis Crick, while discussing his ideas about protein synthesis and the molecules involved in their production, speculated that there “should be at least two types of RNA in the cytoplasm” (1). This statement captures the state of knowledge about cellular RNA at this time: that it was involved with protein synthesis. The details of protein synthesis were just beginning to be revealed, and to this point the intimate role that RNA plays was not yet appreciated, let alone the idea of RNA catalyzing the reaction. Indeed, Crick’s prediction was correct; there are *at least* two types of RNAs, for it is now known that there are three main types of RNA molecules involved in cellular protein synthesis, mRNA, tRNA, and rRNA. The role of RNA as a cellular intermediary chiefly involved in protein synthesis remained the paradigm for much of the next decade. It wasn’t until the late 1960s that the idea of RNAs potentially possessing catalytic properties first appeared in the literature. Woese (1967), Orgel (1968), and Crick (1968) each hypothesized that an RNA world pre-dating our current protein based biology where RNAs acted as both the genome and the only genome-encoded catalyst could have existed (2-4). The

idea of catalytic RNAs remained a hypothesis until the work of Cech (1982) and Altman (1983) in the early 1980s which demonstrated that RNAs are capable of catalyzing enzyme-like reactions. Tom Cech showed that an intervening sequence in the 26S pre-rRNA of *Tetrahymena thermophila*, now known to belong to a class of catalytic RNAs (ribozymes) called group I introns, was capable of catalyzing its own removal from the primary transcript, while Sydney Altman showed that the RNA component of RNase P, a riboprotein complex that removes 5' terminal nucleotides from pre-tRNA molecules, is responsible for catalysis (5, 6). The term 'ribozyme' was coined to describe the enzyme-like activity of these catalytic RNAs. Since the discovery of these two ribozymes, a total of nine additional, natural ribozymes has been discovered: The hammerhead ribozyme, hairpin ribozyme, group II intron, the spliceosome, the HDV ribozyme, the VS ribozyme, the ribosome, the GlmS ribozyme, and the Twister ribozyme (7-17). In addition, synthetic ribozymes have also been developed using *in vitro* methods, furthering the breadth of reactions and functionalities that ribozymes cover (18-20).

1.2 Group I introns

Cis-splicing

Group I introns are catalytic RNAs (ribozymes) that can catalyze their own removal from a primary transcript and ligate the flanking sequences. This cis-splicing reaction is achieved by catalyzing two successive transesterification

reactions, and is the natural activity of group I introns. The cis-splicing reaction is summarized below (Figure 1.1). The reaction begins with an exogenous G binding into the G binding pocket of the properly folded group I intron and the 3'-hydroxyl of the exogenous G nucleophilically attacking the scissile phosphate at the 5'exon—intron junction (Figure 1.1A and 1.1B). Next, a conformational change places the omega G, a universally conserved nucleotide for all group I introns that is the final nucleotide of the intron sequence, into the G binding pocket (Figure 1.1C). This aligns the 3'exon—intron junction for nucleophilic attack from the 3'-hydroxyl of the 5' exon. This sequence of reactions separated by a conformational change results in the ligation of the two exons and the removal of the intron (Figure 1.1C and 1.1D).

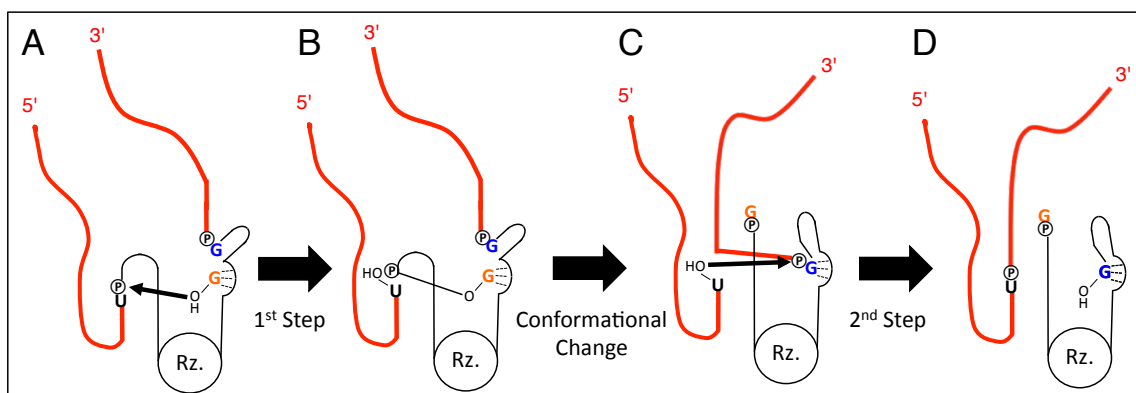


Figure 1.1: Cis-splicing mechanism of group I introns. See text for description of steps. 5' and 3' exons are in red, ribozyme is in black, exogenous G is in orange, omega G is in blue, P represents bridging phosphates at the 5' and 3' splice site, and G-binding pocket is shown as semi circle with three dashed arrows to the bound G.

Genetic Distribution & Intron Mobility

Group I introns are parasitic genetic elements that have been identified across many branches of the trees of life including prokaryotes, plants, genomes of unicellular eukaryotes, and organellar genomes (21, 22). Notably, they have

only rarely found in metazoan mitochondrial DNA and have yet to be identified in archaea (22). Propagation of group I introns predominantly occurs through vertical genetic transfer, however, mechanisms for the horizontal genetic transfer of group I introns (intron mobility) have been proposed: homing and reverse splicing. The homing mechanism allows group I introns to integrate themselves into the identical segment of intronless alleles using a homing endonuclease gene, a special type of restriction enzyme that is present as an open reading frame in some group I introns that cleaves DNA at site specific locations (23). After the homing endonuclease cleaves the DNA of the intronless allele (recipient allele), the group I intron gene (from donor allele) integrates into the recipient allele using a DNA recombination mechanism (24, 25). Indeed, the spread of group I introns via the homing pathway has been experimentally observed (26). The second mechanism for group I intron mobility, reverse splicing, has been shown to occur on the RNA level, but the complete pathway to permanent integration into the host DNA has not yet been demonstrated (27, 28). Despite the lack of experimental validation for the second mechanism, these two processes describe the current understanding of the mobility and spread of group I introns (21).

Common Structural Elements

Self-splicing group I introns have little primary sequence conservation, but share a set of common core secondary structural elements (29, 30). They are partitioned into 10 different subclasses based on their sequence homologies and common paired regions (31). The *Tetrahymena* group I intron is the most well

studied group I intron and its secondary structure is shown for reference highlighting the paired regions is shown in Figure 1.2A. This figure also identifies additional paired regions labeled as P2.1, P5a, P5b, P5c, P9.1, and P9.2. Additional paired regions differ between group I introns (see *Azoarcus* secondary structure in Figure 1.2B), but can be important for full activity (32).

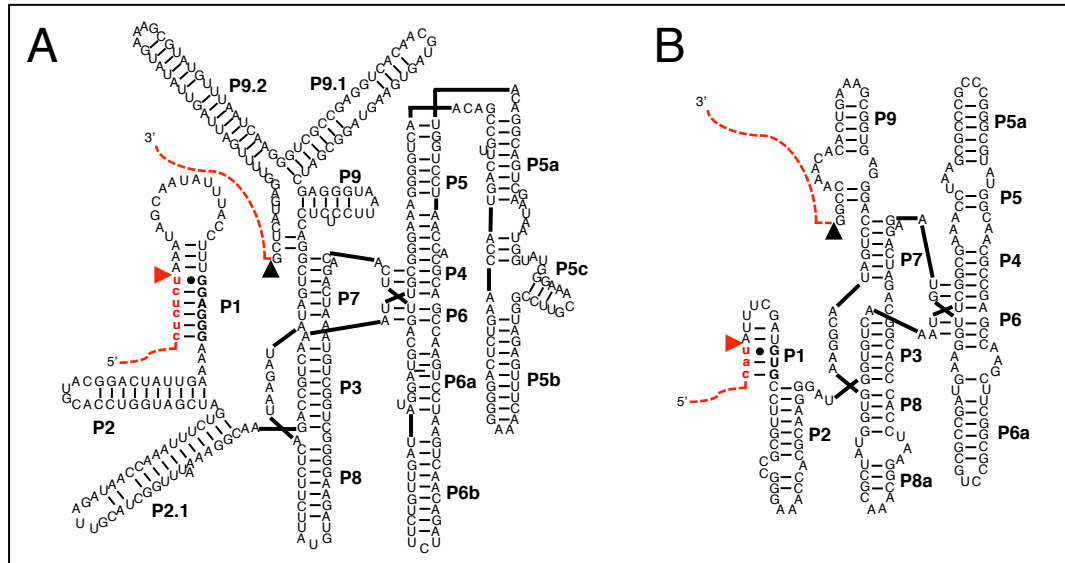


Figure 1.2: Secondary structure of *Tetrahymena* and *Azoarcus* ribozymes. The *Tetrahymena* ribozyme (A) is 414 nucleotides in length, while the *Azoarcus* ribozyme (B) has only 205 nucleotides. Paired regions are labeled P1-9 in black & bold. Note the lack of additional paired regions present on the smaller *Azoarcus* ribozyme (ex: no P9.1, P9.2, P2.1, etc.). The internal guide sequence on P1 that defines the 5' splice site is in black and bold, while corresponding paired sequences are in red. Remainder of 5' exon and 3' exon are shown in red dashed lines. The red arrowhead indicates the 5' splice site and the black arrowhead indicates the 3' splice site. The *Tetrahymena* and *Azoarcus* ribozyme secondary structures shown are modified representations from (33) and (34), respectively.

There are two main helical domains in group I introns, P4-P6 and P3,P7-P9 (referred to as P3-P9) (31). The P4-P6 domain, consisting of P4, P5, and P6 forms part of the active site at the interface with P3-P9 and is essential for group I intron folding. In the well-studied *Tetrahymena* ribozyme, the P4-P6 has been

shown to be self-folding and to act as a structural scaffold on which the rest of the ribozyme folds (35, 36). The self-folding and scaffolding function of the P4-P6 domain, however, has not held true for all group I intron ribozymes (37). The P3-P9 domain, consisting of paired regions P3, P7, P8, and P9 forms the other part of the active site at the interface with P4-P6 and also contains the G binding pocket, located in P7 of all group I introns, a region essential for catalytic function of the ribozyme (31, 38, 39). P1 is also an essential paired region in all group I introns because it contains the base pairing that delineates the 5' splice site, termed the internal guide sequence (IGS) (40). The base pairing in P1 places the splice site junction adjacent to the G binding pocket within the active site (41). The internal guide sequence is not a conserved length in all group I introns (see bold nucleotides in P1 of Figure 1.2A and Figure 1.2B), and the only sequence identity requirement in this segment is a universally conserved G at the 5' terminus of the IGS that forms a wobble base pair a U at the 5' splice site (see red arrow in Figure 1.2A and 1.2B) (42). P2, on the other hand, is a paired region that is not completely conserved in all group I introns and studies have shown that the P2 can be removed from some group I intron ribozymes without removing activity (43). Another notable universal feature of group I introns is their 3' terminal G, also called the omega G. This G binds into the G binding pocket after the first step of splicing and aligns the scissile phosphate delineating the 3' splice-site junction for nucleophilic attack from the 5' exon (see black arrow in Figure 1.2A and 1.2B) (41).

1.3 Trans-splicing Group I Introns

Initial Use

The first intermolecular, *trans*-splicing group I intron reaction reported was in 1986 with using *Tetrahymena* ribozyme when exploring the catalytic potential of group I intron ribozymes (44). In this report the 5' exon and first few nucleotides of the intron were truncated, disrupting normal base pairing in P1 and leaving the internal guide sequence free to bind with an exogenous substrate (Figure 1.3, compare to Figure 1.2A). The reaction with the 'L-19' ribozyme, as it was called, was initiated by providing a short RNA substrate in *trans*. This truncated ribozyme, or a close derivative, has been used in multiple subsequent explorations of the kinetic and functional properties of the *Tetrahymena* ribozyme (45-50).

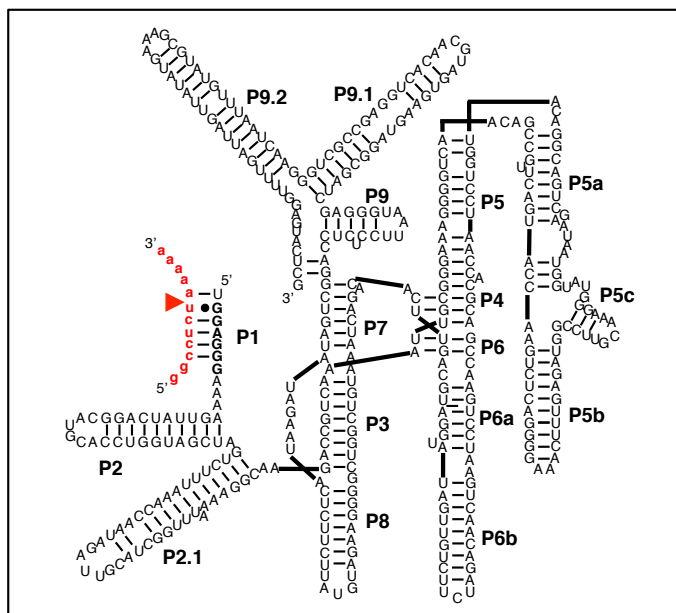


Figure 1.3: First *trans*-splicing group I intron design. Note the truncation of the 5' portion of the *Tetrahymena* ribozyme and lack of 3' exon. The *trans* substrate is shown in red with a red arrow indicating the splice site. The *Tetrahymena* ribozyme secondary structures shown is a modified representation from (33).

Gene Therapy Applications, and Reaction Improvements

In 1994, Bruce Sullenger used the *trans*-splicing *Tetrahymena* ribozyme to target and repair the 3' end of a defective mRNA molecule both *in vitro* and in *E. coli*. cells (51). This reaction was chemically analogous to the previously reported *trans*-splicing reactions, but had two major differences. First, instead of using a short RNA as the substrate, a full length mRNA was used. And second, instead of using a short or no 3' exon, as done previously, a 3' exon was used that when spliced at the correct splice site encoded a proper copy of the remainder of the targeted gene, thus repairing the targeted defective mRNA. A summary of the 3' mRNA repair *trans*-splicing reaction scheme is shown in Figure 1.4A (compare to the *cis*-splicing reaction in Figure 1.4B).

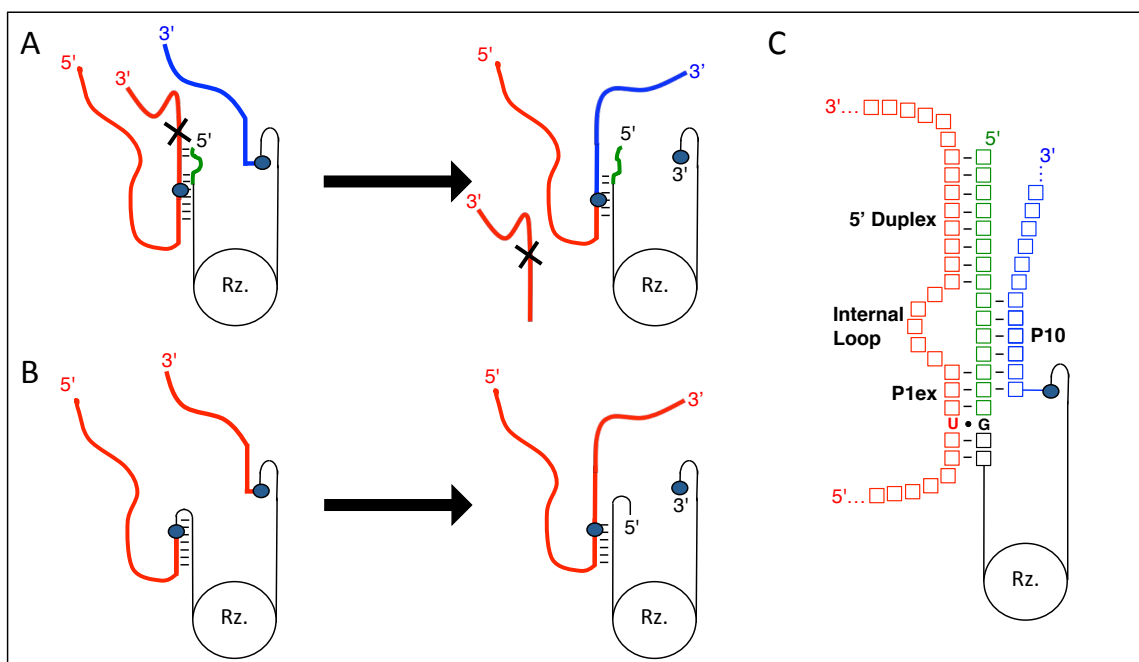


Figure 1.4 : Summary of *cis*-splicing vs. 3' *trans*-splicing. (A) *Trans*-splicing diagram showing the exogenous targeted RNA bound to the *trans*-splicing ribozyme (left), and the final reaction products showing the chimeric targeted RNA and the removed 3' portion of the targeted RNA (right). Targeted RNA is in red, 3' exon is in blue, 5' EGS of ribozyme is in green, group I intron is in black and represented with 'Rz.', and "X" on 3' portion of targeted RNA represents defective portion that can be removed with this reaction. (B) *Cis*-splicing diagram showing the primary transcript (left) and the final reaction products showing the linked flanking sequences and excised intron (right). The flanking sequences are red, while the group I intron is black and represented with 'Rz.'. (C) EGS design principles used for first *trans*-splicing ribozymes. This is a more detailed view of the *trans*-splicing summary shown in (A). The structural features of the EGS include pairing regions comprising an P1 extension (P1ex), a P10 (base pairing between EGS and 3' exon adjacent to 3' splice site), an internal loop between mRNA and EGS, and a 5' duplex to increase substrate binding and specificity. Actual lengths of P1ex, P10, loop region, and 5' duplex varied between different constructs tested. Coloring scheme is the same as in (A). Open boxes represent positions of possible nucleotide variability.

Further studies of 3' mRNA repair *trans*-splicing group I intron ribozymes tested the potential of the *Tetrahymena* group I intron ribozyme to target other mRNAs and *trans*-splice in mammalian cells (52-54). A limiting factor in these works was the low efficiency of the *trans*-splicing ribozyme in cells. Kohler *et al.*

in 1999 demonstrated that the addition of an extended guide sequence (EGS) to the 5' terminus of the *trans*-splicing *Tetrahymena* ribozyme improved the efficiency by increasing both substrate binding and target specificity (see green sequence in Figure 1.4A and 1.4C) (55). The extended guide sequences used were designed to emulate the structural contacts made in the *cis*-splicing reaction of the *Tetrahymena* ribozyme. It consisted of a P1 extension, an internal loop, a P10 interaction (base pairing between the EGS and the 3' exon, Figure 1.4C). Additionally, the EGS included a 5' terminal duplex of variable length that increased substrate binding (Figure 1.4C). These design principles, although a useful starting point, did not consistently result in highly active *trans*-splicing ribozymes with reported *trans*-splicing efficiencies between 1.2% and 10-50% (56, 57). Notably, splicing efficiencies above 10% were seen only when the ribozymes were expressed from very strong promoters. Strong promoters yield high expression levels of ribozyme that are not conducive to normal cellular function; thus, high ribozymes expression levels would not be useful in therapy or as a tool in cellular research. More recently, an alternative approach to extended guide sequence design of the *Tetrahymena* ribozyme has been developed where the extended guide sequence is randomized in a pool of *trans*-splicing ribozymes, and the pool is subjected to multiple rounds of *in vivo* selection eliminating non-productive designs and enriching for useful designs (58). This work used a weaker promoter that lowers ribozymes expression levels and was able to increase *trans*-splicing from 1% to 4%. Despite these improvements, the efficiency of group I intron 3' mRNA repair reaction remains a limiting factor, and

alternative strategies need to be developed before it can be a consistent, useful tool in gene therapy or as a tool in research.

Additional Trans-splicing Reactions

Expanding beyond the 3' mRNA repair reaction, other group I intron mediated *trans*-splicing reactions have been reported. In a different form of the 3' *trans*-splicing reaction, cytotoxic peptides are introduced using the same fundamental reaction as the 3' mRNA repair reaction to kill targeted cells. Instead of the 3' exon encoding a repaired message, the 3' exon encodes a cytotoxic peptide (59, 60). Expression of the trans-spliced RNA leads to cell death. This could be a useful tool for the treatment of cancer cells, which typically display a much higher expression level of hTert mRNA than other cell types, and to kill virally infected cells, which are the only cells in the body expressing viral RNAs (59, 61). The other types of repair reactions reported include a 5' mRNA repair reaction where instead of replacing the 3' portion of the substrate, the 5' portion of the substrate it replaced, a *trans*-insertion reaction where nucleotides are inserted within a targeted sequence, and a *trans*-excision reaction, where nucleotides are removed from a targeted sequence (62-64). Most recently, an extension of the previously reported trans-excision reaction has been reported where instead of removing only 28 nucleotides, up to 100 nucleotides can be removed (65). These 'spliceozymes' may find use in the repair of splicing diseases. Between 10% and 60% of all mutations that cause a genetic disease act via aberrant splicing, where the spliceozymes could convert incompletely spliced pre-mRNAs to translatable mRNAs (66, 67).

Other Group I Introns Tested

Since the initial report on 3' mRNA repair *trans*-splicing, many studies have sought to improve this reaction. Notably, the *Tetrahymena* ribozyme has been used for almost all of these works. In fact, a surprisingly low number of group I introns have been tested for this activity, despite there being over 1,600 known group I introns (68). Prior to this work only four different ribozymes had been tested (Didymium ribozyme (DiGIR2), two *Fuligo* ribozymes (Fse.L569, Fse.L1898), and the *Tetrahymena* ribozyme (Tth.L1925)), and besides the *Tetrahymena* ribozyme, none had been tested for *in vivo* activity (69, 70). Of the four ribozymes tested, all used the same EGS principles initially designed for the *Tetrahymena* ribozyme. Identifying design principles that could make different classes of group I introns useful for *trans*-splicing applications could be an important step in obtaining a *trans*-splicing ribozyme with improved activity.

1.4 Azoarcus Group I Intron

The *Azoarcus* group I intron ribozyme was first isolated and identified in the anticodon stem loop of pre-tRNA^{Ile} from *Azoarcus* sp. *BH72* in 1992 (71). In this initial report, the small size (205nt) and high GC base pair content (65%) were immediately identified (71). The small size of the ribozyme is due to the absence of many additional paired regions on top of the core P1-P9, and is about half the size of the *Tetrahymena* ribozyme (compare *Tetrahymena* in Figure 2A with *Azoarcus* in Figure 2B). The high GC content, along with two identified

tetraloop-receptor interactions between P9-P5 and P2-P8 gives the ribozyme high temperature tolerance, as its optimal temperature is 70-75°C *in vitro* (34). A subsequent crystal structure of the *Azoarcus* ribozyme confirmed the predicted tetraloop-receptor interactions, and also showed that it was a tightly packed structure with extensive, elongated interactions at the interface of the coaxially stacked P4-P6 and P3-P9 domains (41).

Aside from the crystal structure, extensive structural and kinetic probing experiments have been performed on the *Azoarcus* ribozyme. The compactly folded *Azoarcus* ribozyme has an internal guide sequence of only 3 nucleotides, compared to six nucleotides for the *Tetrahymena* ribozyme. As expected, this results in significantly weaker substrate binding for the *Azoarcus* ribozyme, but measured binding values are 10^6 - 10^9 stronger than expected if only simple base pairing were involved, demonstrating that tertiary contacts within the tightly folded core of the ribozyme increase stabilization and compensate for the shorter internal guide sequence (72-74). The weaker substrate binding due to the shorter internal guide sequence of the *Azoarcus* ribozyme was not expected to be a hindrance in *trans*-splicing reactions because of the extended guide sequence that will increase substrate specificity and binding. One of the most explored areas of the *Azoarcus* ribozyme is its folding kinetics and that is one the main reasons the *Azoarcus* ribozyme was selected for this work.

Azoarcus Ribozyme Folding

The kinetic folding studies of the *Azoarcus* ribozyme showed that unlike the *Tetrahymena* ribozyme, whose P4-P6 domain is independently folding and

acts as a folding scaffold, the two helical domains of the *Azoarcus* ribozyme (P4-P6 and P3-P9) formed cooperatively in a single thermodynamic transition (35-37). Careful dissection of the folding kinetics of *Azoarcus* ribozyme has shown that successful folding is dependent on the tertiary interactions formed between the tetraloop-receptors, and slight perturbations will strongly increase the amount of ribozyme that folds (75, 76). Initial folding kinetics indicated that the ribozyme completely folded within milliseconds, but more recent results coupled to activity of the ribozyme show that like many structured RNAs, a portion of *Azoarcus* ribozyme population gets stuck in kinetic folding traps (37, 77, 78). These studies revealed that ~50% the *Azoarcus* ribozyme population directly folds into the active state, while the remaining misfolded population folds into the active state within minutes (77, 78). This is compared to the *Tetrahymena* ribozyme where ~25-30% of the population directly folds into its active conformation and it takes hours for the misfolded population to resolve (77, 79). Since a higher population of properly folded ribozymes yields more opportunity for successful trans-splicing, the folding kinetics of the *Azoarcus* ribozyme make it a promising candidate for improved *trans*-splicing characteristics.

Previous Trans-Splicing Azoarcus Reactions

The *Azoarcus* ribozyme has been used in *trans*-splicing reactions previously, although none have used it towards a gene therapy or biological tool applications. The previous *trans*-splicing reactions the *Azoarcus* ribozyme has performed are *trans*-ligation reactions and include an RNA polymerization-like reaction where short oligonucleotide fragments are sequentially added to one

another, and a recombination reaction where the 3' portion of strand A replaced the 3' portion of strand B (80-82). The recombination reaction has been built upon to show that the *Azoarcus* ribozyme can catalyze its own self-assembly from four separate RNA fragments and can also catalyze the formation of an active ligase ribozyme from two separate inactive oligos (83-85). Two other notable works involving *trans*-splicing *Azoarcus* ribozymes look at (1) how self-assembled, self-replicating *Azoarcus* ribozyme fragments spontaneously formed a cooperative network from sequence fragments with randomized substrate recognition sequences (using the recombination reaction), and (2) how RNA populations with cryptic genetic mutations, giving no selective advantage to the natural reaction, can better adapt to a new environment requiring catalyzing a non-native reaction (using the reverse of second step of splicing) (86, 87).

Collectively, the smaller size, previously demonstrated catalytic plasticity, robust folding, and X-ray crystal structure elucidating high levels of structural detail make *Azoarcus* ribozyme a prime candidate for testing its 3'-mRNA *trans*-splicing potential. In chapter 2 of this dissertation I describe the results of this work.

1.5 The RNA World

The term 'RNA World' was first introduced by Walter Gilbert in 1986 (88). It describes a time pre-dating the current DNA – protein based world where RNA acted as both the genome and the only genome-encoded catalyst. This was not

the first time RNA as a catalytic, potentially primordial molecule was hypothesized. As mentioned, in 1967-68 Orgel, Crick, and Woese had each discussed the idea of the potential for RNA to precede DNA –protein based life (2-4). Several strong pieces of evidence give credence to this hypothesis. Among those are the discovery of RNAs with catalytic properties, the discovery that the ribosome is a ribozyme (the active site is completely composed of RNA), and the discovery that in the synthesis of nucleic acids in modern cells ribonucleic acids (RNA) are made first and deoxyribonucleic acids (DNA) are made by modifying the RNAs (5, 6, 14, 15, 89). The growing number of non-natural catalytic RNAs identified in the lab by *in vitro* selection also supports the hypothesis. The RNA World is an umbrella term possibly covering a vast time in history and encapsulating many different areas of research. Since this dissertation introduces improvements to a recently *in vitro* selected triphosphorylation ribozyme, a potentially important metabolic ribozyme in the RNA world, the discussion herein will briefly introduce two processes closely linked to the triphosphorylation of nucleosides: (1) the synthesis of nucleosides and nucleotides, and (2) RNA polymerization both from the prebiotic milieu and in an RNA world.

Prebiotic Synthesis of Nucleosides and in an RNA World

The synthesis of nucleosides within a prebiotic milieu has long been a question in Origin of Life research, and several different hypotheses have been proposed. One line of thought is that the ribose and nucleobase moieties were synthesized separately and then coupled together; however, most coupling

reactions have low yield (90-92). The synthesis of ribose, similar to the coupling reaction, has also proved to be a more difficult reaction. Enantiomeric purity, stability, and excessive side products hamper current known methods of prebiotic ribose synthesis (93-95). Enantiomeric purity of ribose is important because racemic mixtures of activated nucleosides (containing both D and L ribose) inhibit non-enzyme/ribozyme polymerization reactions (96). The presence of borate or lead (II) (Pb^{2+}) has been shown to increase the formation and stability of ribose in possible synthesis models, but do not address the enantiomeric purity problem (97, 98). Progress towards enantiomeric purity (as well as stability) of ribose has been shown in a reaction with cyanamide. In this reaction cyanamide reacts quickly with ribose forming a bicyclic product that spontaneously crystallizes with several regions within the crystal containing only the D- or L- enantiomer of ribose, creating enriched pockets of enantiomerically pure ribose (99). In this pathway, however, it is unclear how the sequestered ribose would be re-introduced into the synthesis of nucleosides. Despite these works, there remains no clear path to the prebiotic formation of enantiomerically pure D-ribose. The formation of nucleobases, unlike ribose and the coupling reaction, has been successful. A prebiotically plausible synthesis of each nucleobase has been reported (100-102). Adding an additional level of complexity, the differing chemistries between ribose and nucleobase formation introduce the difficulty of the two being synthesized in the same environment (103). Together, this line of work demonstrates that under the current understanding of the potential

conditions in prebiotic world, the formation of ribose and nucleobases, as well as the coupling reaction to form nucleosides would be difficult.

An alternative to the coupling of ribose and nucleobase synthesis is the idea of a direct synthesis of nucleosides that simultaneously build the ribose and nucleobase moieties. Recently, there has been great success in regioselectively synthesizing pyrimidine nucleosides using prebiotically plausible starting materials (104). Additionally, a partial synthetic route using similar aqueous chemistry has been shown that could lead to the synthesis of purines (105). Although the final steps of the purine synthesis have not been shown and there are remaining questions about the necessary purity of reactants and necessity for highly controlled conditions in some of the synthetic steps in both pyrimidine and purine pathways, these works still demonstrate a step forward in prebiotic chemistry and the thinking of how nucleosides were formed. These are two of the leading hypothesis on how nucleosides could be formed in an RNA world.

In an RNA World a faster and more efficient synthesis of nucleosides would have been required. Work towards identifying a ribozyme that can synthesize the components of nucleosides (ribose and nucleobases) has not been reported, but there has been a report of a ribozyme can that catalyze the glycosidic bond formation between nucleobase and ribose (106, 107). In this reaction, 5-phosphoryibosyl 1-pyrophosphate reacts with a nucleobase to form a nucleoside 5' monophosphate (a 5' phosphorylated version of a nucleoside)(Figure 1.5). Currently, the reaction has been shown for both a purine (G) and a pyrimidine (U), but has only been reported when the 5-phosphoryibosyl

1-pyrophosphate is tethered to the 3' end of the ribozyme, indicating that the ribozyme is unable to bind the free 5-phosphorybosyl 1-pyrophosphate and eliminating any multi-turnover potential. Regardless, the demonstration that a ribozyme is capable of catalyzing the coupling of a ribose derivative and a nucleobase is further evidence for the potential existence of an earlier RNA world. The continued discovery and improvement of additional metabolic ribozymes necessary for an RNA world organism will further delineate how an RNA world could have functioned.

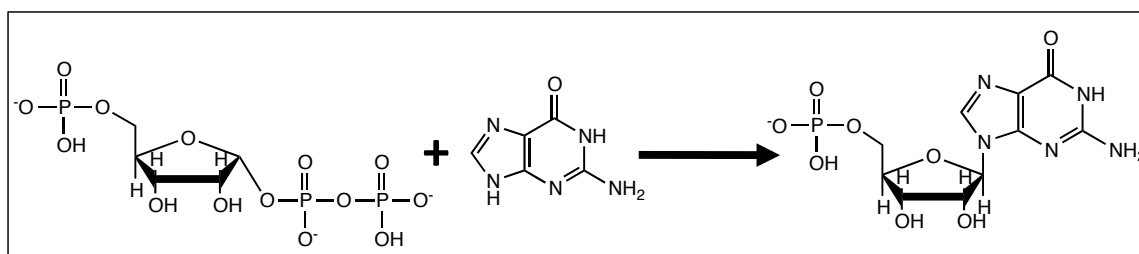


Figure 1.5: Glycosidic bond formation between nucleoside and 5-phosphorybosyl 1-pyrophosphate. The reaction diagram shown is a modified version from (107).

Prebiotic Polymerization & The Polymerase Ribozyme

Current polymerase enzymes require 5' triphosphorylated nucleosides as a starting material and use 3'-5' linkages between nucleotides. How replication was achieved before protein based enzymatic polymerization is unknown because there is no identified polymerase in the modern biological world that isn't composed of amino acids. The RNA world hypothesis predicts an RNA based RNA polymerase ribozyme that predates modern enzymes polymerases and catalyzes the RNA polymerization reaction using 3'-5' linkages and 5'

triphosphorylated nucleosides. A pre-RNA world scenario which minimizes necessary changes and conserves biologic complexity is one where RNAs are polymerized without an enzyme or ribozyme catalysis using 3'-5' linkages and 5' triphosphorylated nucleosides as well. Nucleoside 5' triphosphates, however, are relatively inert at moderate temperatures in solution (108). Nucleoside 5' triphosphates are so unreactive that most studies on non-enzymatic polymerization have used imidazolide phosphate analogs as the leaving group in place of triphosphates to perform experiments (109, 110). These analogs, while improving the timeframe of experiments for laboratory settings, are not the most likely prebiotic molecules, and generate a mixture of 3'-5' and 2'-5' phosphodiester linkages between monomers (111, 112). The most successful reaction with nucleoside 5' phosphorimidazolides monomers to form predominantly 3'-5' linkages (although some 2'-5' linkages remain) have come in the presence of mineral clay surfaces (montmorillonite), specific divalent metal ions, and eutectic phases (112-115). Minimal 2'-5' linkages in RNA polymers would not necessarily be disastrous because previous work has shown that ribozymes can tolerate a portion of 2'-5' linkages and still maintain catalytic function (116). Importantly though, none of the reactions described above can generate polymers with the necessary nucleotide distribution to form the complex structural motifs necessary for ribozyme function (117).

Previous biochemical studies have shown that desired characteristics and activities can exist and be selected for in a pool of random sequences (termed *in vitro* selections) (118-120). The non-catalyzed polymerization of nucleoside 5'-

phosphate derivatives under the appropriate conditions demonstrates there is a possible scenario where short strands of RNAs could have been formed prebiotically. *In vitro* selections, however, have many differences from the demonstrated non-catalyzed polymerization reactions. For example, *in vitro* selections generally use longer sequences for selection of activity, and desired traits are selected using carefully controlled experimental conditions. Prebiotic non-templated polymerization, on the other hand, has not been shown to not create sequences longer than 40 nucleotides and would not be selected for using carefully regulated conditions (121). These differences highlight remaining gaps in our current knowledge and demonstrate challenges of how early chemically activated nucleosides could have polymerized and how a ribozyme of a desirable catalytic activity could have emerged from the random sequences formed. Indeed, the substantial body of work on potential prebiotic polymerization of nucleosides catalogs the difficulty of a self-replicating RNA to emerge from a pool of random nucleotides, and this observation has been well documented (111, 122).

The central requirement for the RNA world hypothesis is the existence of an RNA-dependent RNA polymerase ribozyme capable of self-replication. As mentioned, there is no known non-protein polymerase ribozyme identified in nature, but by using *in vitro* selection, an RNA-dependent RNA polymerase ribozyme has been isolated in the laboratory (123). The initially isolated polymerase ribozyme could only polymerize 14 nucleotides of a single templated sequence. Subsequent improvements including the addition of an accessory

domain improved polymerization to 20 nucleotides and the addition of a template-binding region increased polymerization to 95 nucleotides of a repeating nucleotide sequence (124, 125). Using the template-binding region on a polymerase ribozyme adapted for polymerization in eutectic phases of ice, polymerization of up to 206 nucleotides of a repeating sequence has been achieved at ambient temperatures (126). Despite being able to polymerize up to 206 nucleotides of a single repeating template, the polymerase ribozyme is still not capable of self-replication due to its strong template dependence. Other work has focused on decreasing template dependence, rather than maximizing polymerized sequence length (127). More recently, a new approach selected a polymerase ribozyme that performs a “cross-chiral” reaction where the template and substrate are the opposite enantiomer of the ribozyme (128).

Enantiomerically opposite nucleotide polymers cannot form stable duplexes through Watson-Crick base pairing, forcing substrate/template recognition to operate via tertiary, non-Watson-Crick interactions (129, 130). This is the likely reason why the cross-chiral ribozyme did not show any pronounced dependence on template sequence. This ribozyme can both polymerize the addition of single nucleotides to a template and ligate oligonucleotides, and through a series of 11 ligations and polymerizations it can produce a full-length copy its sequence. It will be interesting to see if in future work the ribozyme is capable of polymerizing exclusive single nucleotide additions without sequence dependence to an excess of its length (83 nucleotides). In doing so it could be the first reported ‘self-replicating’ polymerase ribozyme (the polymerized sequence would be of the

opposite enantiomer, which also acts as a functional ribozyme). Overall, the identification of RNAs capable of catalyzing RNA-dependent RNA polymerization and the recent improvements in catalytic activity shows steady, although arduous, progress towards the identification of a self-replicating ribozyme.

1.6 Prebiotic Triphosphorylation & the Triphosphorylation ribozyme

The 5' triphosphorylation of nucleosides is an intermediate step between their synthesis and use in biological polymerization reactions. Chemical activation of the 5'-phosphate of nucleosides is required because polymerization in aqueous solution is a thermodynamically unfavorable process. Assuming the prebiotic synthesis of nucleosides, a prebiotic world would have needed a chemical capable triphosphorylating the 5' hydroxyl of nucleosides before they could participate in polymerization reactions. A possible prebiotic source of soluble phosphate that could act as a triphosphorylating agent of nucleosides is trimetaphosphate.

Prebiotic Sources of Trimetaphosphate

Trimetaphosphate could have been formed prebiotically in three different plausible mechanisms. The first is the heating of mineral calcium phosphate [$\text{Ca}_3(\text{PO}_4)_2$] in volcanic magma ($>1,200^\circ\text{C}$) to yield low percentages of trimetaphosphate (131). The second is the heating of ammonium dihydrogen phosphate [$\text{NH}_4\text{H}_2\text{PO}_4$] with urea in the range of 85°C to 100°C yielding multi-length linear polyphosphates, some of which converts to trimetaphosphate (132).

The third method, and apparently the most plausible, is the corrosion of the phosphide mineral schreibersite [$(\text{Fe,Ni})_3\text{P}$] in water (133). Phosphide minerals are found in iron meteorites and corrode completely in water between $1\text{-}10^4$ years (134). The corrosion of phosphide minerals is a multistep process that results in various soluble phosphates and a large amount of phosphites (135). Phosphites can be readily oxidized into phosphates and polyphosphates including trimetaphosphate in the presence of an oxidizing agent such as a hydroxyl radical (136). Phosphite deposits have been identified in carbonate samples from the Archean eon indicating that soluble phosphites were present in pre-biotic oceans around 3.5 Ga (137). Prebiotic estimates of soluble polyphosphates, of which trimetaphosphate would be a component, in seawater range from 10nM to 10uM, and as high as 1mM in areas of localized concentrations (tide pools) (136). Collectively, these three possible synthesis mechanisms show that trimetaphosphate was a likely prebiotic molecule.

Non-Enzymatic 5' Triphosphorylation of Nucleosides using Trimetaphosphate

The 5' triphosphorylation of nucleosides using trimetaphosphate occurs by the 5' hydroxyl group nucleophilically attacking one of the phosphates in the cyclic trimetaphosphate molecule (see Figure 1.6 for desired reaction diagram). Trimetaphosphate as a potential triphosphorylation agent has been studied in several reports in the non-enzymatic/ribozyme catalyzed reaction. Two different conditions have been tested. In the first condition the reaction was in solution, but efficient 5' triphosphorylation wasn't observed until a $\text{pH} > 12$, an inhospitable environment for RNA polymers (138). A second condition tested was under

conditions where the solutions were allowed to dry out, effectively concentrating the samples. This was done with different metal ion catalysts and produced measurable amounts of trimetaphosphate, with Ni being the best, but reaction times were on the order of weeks and yields were still low (~13%) (139). The results of the wet dry cycles, however, are not highly reliable because the pH of the reaction was not carefully controlled and no replicates were reported. Further complicating both of these non-enzymatically catalyzed reactions is the difference in pK_a between the 5' and 2'-3' diol of ribose (~15 and ~12, respectively), so as the pH increases the 2'-3' diol is phosphorylated before the 5' hydroxyl (140, 141). The pH requirements, slow reaction times, and poor specificity of the non-enzymatically catalyzed reaction make it a difficult reaction to complete, and suggests that if trimetaphosphate were used to triphosphorylate the 5' hydroxyl of nucleosides in the RNA world, it would likely be catalyzed by a ribozyme.

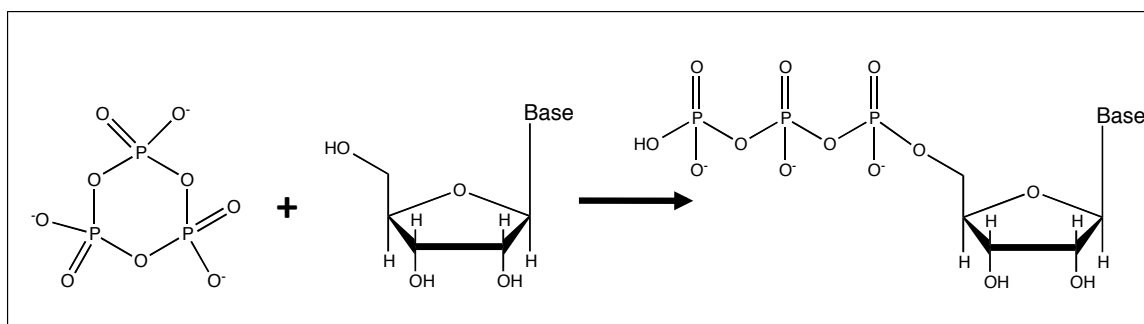


Figure 1.6: 5' triphosphorylation of nucleosides with trimetaphosphate.

Ribozyme Catalyzed 5' Triphosphorylation of Nucleosides using Trimetaphosphate

Recently in this laboratory, a ribozyme has been isolated from a randomized pool that uses trimetaphosphate to triphosphorylate the 5' hydroxyl of nucleosides using an *in vitro* selection (142). This is the first ribozyme to report this activity. The ribozyme, named TPR1, is 96 nucleotides in length (Figure 1.7), and catalyzes the reaction under near neutral pH conditions (pH 8.3) at a measure rate constant of 0.013min^{-1} converting ~90% of starting material into product within hours. TPR1, however, requires 50mM trimetaphosphate for this reaction kinetic, beyond any predicted prebiotic availability and is unable to catalyze the 5' triphosphorylation of free nucleosides. In chapter 3 of this dissertation I will describe my efforts to improve the reactivity of this ribozyme.

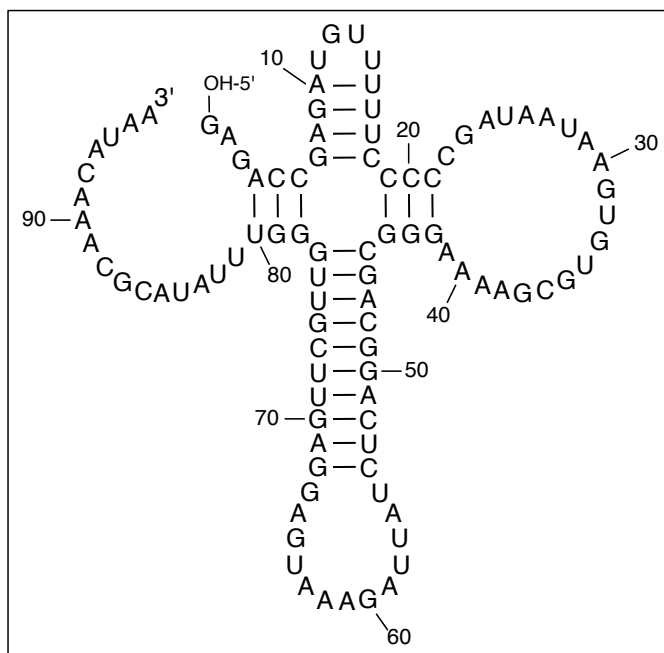


Figure 1.7: Secondary Structure of TPR1. This is the secondary structure of the 96 nucleotide long TPR1. The 5' terminal G is the site of triphosphorylation. Secondary structure representation is a modified version from (142).

1.7 Goals of this Dissertation

In this dissertation I aimed to improve two different catalytic RNAs. The first is the *trans*-splicing *Azoarcus* group I intron ribozyme. The *trans*-splicing reaction is useful in the reprogramming of targeted exogenous RNA molecules and can be useful in gene therapy applications as well as a biochemical tool. The *Azoarcus* ribozyme has never been tested for this activity and is an attractive candidate because of its small size, previously demonstrated catalytic plasticity, robust folding, and X-ray crystal structure elucidating high levels of structural detail. Very few ribozymes have been tested for *trans*-splicing activity and the testing of new ribozymes has been noted as an important contribution to the field (143). The goal of this project was to optimize the *Azoarcus trans*-splicing reaction by identifying structural contacts that maximized splicing efficiency and to make them viable in a cellular system.

The second ribozyme improved in this dissertation is the triphosphorylation ribozyme, a ribozyme that was very recently developed in the Müller lab. This ribozyme triphosphorylates the 5'-hydroxyl group of an RNA using trimetaphosphate. This or a related reaction could have been important in an RNA world, and variants of this ribozyme could have provided the chemically activated substrate required by early polymerases. The goal of this project was to increase the activity of the initially isolated triphosphorylation ribozyme as well as

to increase its prebiotic relevance by testing for activity of the improved ribozyme at more prebiotically plausible trimetaphosphate concentrations.

1.8 References

1. Crick, F. H. (1958) On protein synthesis, *Symposia of the Society for Experimental Biology* 12, 138-163.
2. Crick, F. H. (1968) The origin of the genetic code, *Journal of molecular biology* 38, 367-379.
3. Orgel, L. E. (1968) Evolution of the genetic apparatus, *Journal of molecular biology* 38, 381-393.
4. Woese, C. R. (1967) The Fundamental Nature of the Genetic Code: Prebiotic Interactions Between Polynucleotides and Polyamino Acids or their derivatives, *The Genetic Code: the Molecular basis for Genetic Expression* 59.
5. Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*, *Cell* 31, 147-157.
6. Guerriertakada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983) The Rna Moiety of Ribonuclease-P Is the Catalytic Subunit of the Enzyme, *Cell* 35, 849-857.
7. Hutchins, C. J., Rathjen, P. D., Forster, A. C., and Symons, R. H. (1986) Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid, *Nucleic acids research* 14, 3627-3640.
8. Buzayan, J. M., Gerlach, W. L., and Bruening, G. (1986) Satellite Tobacco Ringspot Virus-Rna - a Subset of the Rna Sequence Is Sufficient for Autolytic Processing, *Proceedings of the National Academy of Sciences of the United States of America* 83, 8859-8862.
9. Peebles, C. L., Perlman, P. S., Mecklenburg, K. L., Petrillo, M. L., Tabor, J. H., Jarrell, K. A., and Cheng, H. L. (1986) A self-splicing RNA excises an intron lariat, *Cell* 44, 213-223.

10. Vanderveen, R., Arnberg, A. C., Vanderhorst, G., Bonen, L., Tabak, H. F., and Grivell, L. A. (1986) Excised Group-I Introns in Yeast Mitochondria Are Lariats and Can Be Formed by Self-Splicing *In vitro*, *Cell* 44, 225-234.
11. Fica, S. M., Tuttle, N., Novak, T., Li, N. S., Lu, J., Koodathingal, P., Dai, Q., Staley, J. P., and Piccirilli, J. A. (2013) RNA catalyses nuclear pre-mRNA splicing, *Nature* 503, 229-234.
12. Sharmeen, L., Kuo, M. Y., Dinter-Gottlieb, G., and Taylor, J. (1988) Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage, *Journal of virology* 62, 2674-2679.
13. Saville, B. J., and Collins, R. A. (1990) A site-specific self-cleavage reaction performed by a novel RNA in *Neurospora* mitochondria, *Cell* 61, 685-696.
14. Nissen, P., Hansen, J., Ban, N., Moore, P. B., and Steitz, T. A. (2000) The structural basis of ribosome activity in peptide bond synthesis, *Science* 289, 920-930.
15. Noller, H. F., Hoffarth, V., and Zimniak, L. (1992) Unusual resistance of peptidyl transferase to protein extraction procedures, *Science* 256, 1416-1419.
16. Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A., and Breaker, R. R. (2004) Control of gene expression by a natural metabolite-responsive ribozyme, *Nature* 428, 281-286.
17. Roth, A., Weinberg, Z., Chen, A. G., Kim, P. B., Ames, T. D., and Breaker, R. R. (2014) A widespread self-cleaving ribozyme class is revealed by bioinformatics, *Nature chemical biology* 10, 56-60.
18. Joyce, G. F. (2004) Directed evolution of nucleic acid enzymes, *Annual review of biochemistry* 73, 791-836.
19. Fiammengo, R., and Jaschke, A. (2005) Nucleic acid enzymes, *Current opinion in biotechnology* 16, 614-621.
20. Chen, X., Li, N., and Ellington, A. D. (2007) Ribozyme catalysis of metabolism in the RNA world, *Chem Biodivers* 4, 633-655.
21. Haugen, P., Simon, D. M., and Bhattacharya, D. (2005) The natural history of group I introns, *Trends in genetics : TIG* 21, 111-119.
22. Hausner, G., Hafez, M., and Edgell, D. R. (2014) Bacterial group I introns: mobile RNA catalysts, *Mobile DNA* 5, 8.

23. Belfort, M., and Roberts, R. J. (1997) Homing endonucleases: keeping the house in order, *Nucleic acids research* 25, 3379-3388.
24. Clyman, J., and Belfort, M. (1992) Trans and cis requirements for intron mobility in a prokaryotic system, *Genes & development* 6, 1269-1279.
25. Lambowitz, A. M., Caprara, M. G., Zimmerly, S., and Perlman, P. S. (1999) Group I and Group II Ribozymes as RNPs: Clues to the Past and Guides to the Future, In *The RNA World, 2nd Ed.*, pp 451-485.
26. Dujon, B. (1989) Group I introns as mobile genetic elements: facts and mechanistic speculations--a review, *Gene* 82, 91-114.
27. Roman, J., and Woodson, S. A. (1998) Integration of the Tetrahymena group I intron into bacterial rRNA by reverse splicing in vivo, *Proceedings of the National Academy of Sciences of the United States of America* 95, 2134-2139.
28. Hedberg, A., and Johansen, S. D. (2013) Nuclear group I introns in self-splicing and beyond, *Mobile DNA* 4, 17.
29. Michel, F., Jacquier, A., and Dujon, B. (1982) Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure, *Biochimie* 64, 867-881.
30. Burke, J. M., Belfort, M., Cech, T. R., Davies, R. W., Schweyen, R. J., Shub, D. A., Szostak, J. W., and Tabak, H. F. (1987) Structural conventions for group I introns, *Nucleic acids research* 15, 7217-7221.
31. Michel, F., and Westhof, E. (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis, *Journal of molecular biology* 216, 585-610.
32. van der Horst, G., Christian, A., and Inoue, T. (1991) Reconstitution of a group I intron self-splicing reaction with an activator RNA, *Proceedings of the National Academy of Sciences of the United States of America* 88, 184-188.
33. Cech, T. R., Damberger, S. H., and Gutell, R. R. (1994) Representation of the secondary and tertiary structure of group I introns, *Nature structural biology* 1, 273-280.
34. Tanner, M., and Cech, T. (1996) Activity and thermostability of the small self-splicing group I intron in the pre-tRNA(Ile) of the purple bacterium *Azoarcus*, *RNA* 2, 74-83.

35. Murphy, F. L., and Cech, T. R. (1993) An independently folding domain of RNA tertiary structure within the Tetrahymena ribozyme, *Biochemistry* 32, 5291-5300.
36. Doherty, E. A., and Doudna, J. A. (1997) The P4-P6 domain directs higher order folding of the Tetrahymena ribozyme core, *Biochemistry* 36, 3159-3169.
37. Rangan, P., Masquida, B., Westhof, E., and Woodson, S. A. (2003) Assembly of core helices and rapid tertiary folding of a small bacterial group I ribozyme, *Proceedings of the National Academy of Sciences of the United States of America* 100, 1574-1579.
38. Michel, F., Hanna, M., Green, R., Bartel, D. P., and Szostak, J. W. (1989) The guanosine binding site of the Tetrahymena ribozyme, *Nature* 342, 391-395.
39. Been, M. D., and Perrotta, A. T. (1991) Group I intron self-splicing with adenosine: evidence for a single nucleoside-binding site, *Science* 252, 434-437.
40. Waring, R. B., Towner, P., Minter, S. J., and Davies, R. W. (1986) Splice-Site Selection by a Self-Splicing Rna of Tetrahymena, *Nature* 321, 133-139.
41. Adams, P. L., Stahley, M. R., Kosek, A. B., Wang, J., and Strobel, S. A. (2004) Crystal structure of a self-splicing group I intron with both exons, *Nature* 430, 45-50.
42. Doudna, J. A., Cormack, B. P., and Szostak, J. W. (1989) Rna Structure, Not Sequence, Determines the 5' Splice-Site Specificity of a Group-I Intron, *Proceedings of the National Academy of Sciences of the United States of America* 86, 7402-7406.
43. Beaudry, A. A., and Joyce, G. F. (1990) Minimum secondary structure requirements for catalytic activity of a self-splicing group I intron, *Biochemistry* 29, 6534-6539.
44. Zaug, A. J., Been, M. D., and Cech, T. R. (1986) The Tetrahymena ribozyme acts like an RNA restriction endonuclease, *Nature* 324, 429-433.
45. Been, M. D., and Cech, T. R. (1986) One binding site determines sequence specificity of Tetrahymena pre-rRNA self-splicing, trans-splicing, and RNA enzyme activity, *Cell* 47, 207-216.
46. Zaug, A. J., Grosshans, C. A., and Cech, T. R. (1988) Sequence-specific endoribonuclease activity of the Tetrahymena ribozyme: enhanced

cleavage of certain oligonucleotide substrates that form mismatched ribozyme-substrate complexes, *Biochemistry* 27, 8924-8931.

47. Herschlag, D., and Cech, T. R. (1990) Catalysis of RNA cleavage by the *Tetrahymena thermophila* ribozyme. 2. Kinetic description of the reaction of an RNA substrate that forms a mismatch at the active site, *Biochemistry* 29, 10172-10180.
48. Herschlag, D., and Cech, T. R. (1990) Catalysis of RNA cleavage by the *Tetrahymena thermophila* ribozyme. 1. Kinetic description of the reaction of an RNA substrate complementary to the active site, *Biochemistry* 29, 10159-10171.
49. Pyle, A. M., McSwiggen, J. A., and Cech, T. R. (1990) Direct measurement of oligonucleotide substrate binding to wild-type and mutant ribozymes from *Tetrahymena*, *Proceedings of the National Academy of Sciences of the United States of America* 87, 8187-8191.
50. Young, B., Herschlag, D., and Cech, T. R. (1991) Mutations in a nonconserved sequence of the *Tetrahymena* ribozyme increase activity and specificity, *Cell* 67, 1007-1019.
51. Sullenger, B. A., and Cech, T. R. (1994) Ribozyme-mediated repair of defective mRNA by targeted, trans-splicing, *Nature* 371, 619-622.
52. Jones, J. T., and Sullenger, B. A. (1997) Evaluating and enhancing ribozyme reaction efficiency in mammalian cells, *Nature biotechnology* 15, 902-905.
53. Lan, N., Howrey, R. P., Lee, S. W., Smith, C. A., and Sullenger, B. A. (1998) Ribozyme-mediated repair of sickle beta-globin mRNAs in erythrocyte precursors, *Science* 280, 1593-1596.
54. Zarrinkar, P. P., and Sullenger, B. A. (1999) Optimizing the substrate specificity of a group I intron ribozyme, *Biochemistry* 38, 3426-3432.
55. Kohler, U., Ayre, B. G., Goodman, H. M., and Haseloff, J. (1999) Trans-splicing ribozymes for targeted gene delivery, *Journal of molecular biology* 285, 1935-1950.
56. Rogers, C. S., Vanoye, C. G., Sullenger, B. A., and George, A. L., Jr. (2002) Functional repair of a mutant chloride channel using a trans-splicing ribozyme, *The Journal of clinical investigation* 110, 1783-1789.
57. Byun, J., Lan, N., Long, M., and Sullenger, B. A. (2003) Efficient and specific repair of sickle beta-globin RNA by trans-splicing ribozymes, *RNA* 9, 1254-1263.

58. Olson, K. E., and Muller, U. F. (2012) An in vivo selection method to optimize trans-splicing ribozymes, *RNA* 18, 581-589.
59. Ayre, B. G., Kohler, U., Goodman, H. M., and Haseloff, J. (1999) Design of highly specific cytotoxins by using trans-splicing ribozymes, *Proceedings of the National Academy of Sciences of the United States of America* 96, 3507-3512.
60. Song, M. S., and Lee, S. W. (2006) Cancer-selective induction of cytotoxicity by tissue-specific expression of targeted trans-splicing ribozyme, *FEBS letters* 580, 5033-5043.
61. Kwon, B. S., Jeong, J. S., Won, Y. S., Lee, C. H., Yoon, K. S., Hyung Jung, M., Kim, I. H., and Lee, S. W. (2011) Intracellular efficacy of tumor-targeting group I intron-based trans-splicing ribozyme, *The journal of gene medicine* 13, 89-100.
62. Bell, M. A., Johnson, A. K., and Testa, S. M. (2002) Ribozyme-catalyzed excision of targeted sequences from within RNAs, *Biochemistry* 41, 15327-15333.
63. Alexander, R. C., Baum, D. A., and Testa, S. M. (2005) 5' transcript replacement in vitro catalyzed by a group I intron-derived ribozyme, *Biochemistry* 44, 7796-7804.
64. Johnson, A. K., Sinha, J., and Testa, S. M. (2005) Trans insertion-splicing: ribozyme-catalyzed insertion of targeted sequences into RNAs, *Biochemistry* 44, 10702-10710.
65. Amini, Z. N., Olson, K. E., and Muller, U. F. (2014) Spliceozymes: ribozymes that remove introns from pre-mRNAs in trans, *PloS one* 9, e101932.
66. Hammond, S. M., and Wood, M. J. (2011) Genetic therapies for RNA mis-splicing diseases, *Trends in genetics : TIG* 27, 196-205.
67. Ward, A. J., and Cooper, T. A. (2010) The pathobiology of splicing, *The Journal of pathology* 220, 152-163.
68. Zhou, Y., Lu, C., Wu, Q. J., Wang, Y., Sun, Z. T., Deng, J. C., and Zhang, Y. (2008) GISSD: Group I Intron Sequence and Structure Database, *Nucleic acids research* 36, D31-37.
69. Lundblad, E. W., Haugen, P., and Johansen, S. D. (2004) Trans-splicing of a mutated glycosylasparaginase mRNA sequence by a group I ribozyme deficient in hydrolysis, *European journal of biochemistry / FEBS* 271, 4932-4938.

70. Fiskaa, T., Lundblad, E. W., Henriksen, J. R., Johansen, S. D., and Einvik, C. (2006) RNA reprogramming of alpha-mannosidase mRNA sequences in vitro by myxomycete group IC1 and IE ribozymes, *The FEBS journal* 273, 2789-2800.
71. Reinhold-Hurek, B., and Shub, D. A. (1992) Self-splicing introns in tRNA genes of widely divergent bacteria, *Nature* 357, 173-176.
72. Kuo, L. Y., Davidson, L. A., and Pico, S. (1999) Characterization of the Azoarcus ribozyme: tight binding to guanosine and substrate by an unusually small group I ribozyme, *Biochimica et biophysica acta* 1489, 281-292.
73. Strauss-Soukup, J. K., and Strobel, S. A. (2000) A chemical phylogeny of group I introns based upon interference mapping of a bacterial ribozyme, *Journal of molecular biology* 302, 339-358.
74. Gleitsman, K. R., and Herschlag, D. H. (2014) A kinetic and thermodynamic framework for the Azoarcus group I ribozyme reaction, *RNA* 20, 1732-1746.
75. Chauhan, S., Behrouzi, R., Rangan, P., and Woodson, S. A. (2009) Structural rearrangements linked to global folding pathways of the Azoarcus group I ribozyme, *Journal of molecular biology* 386, 1167-1178.
76. Chauhan, S., Caliskan, G., Briber, R. M., Perez-Salas, U., Rangan, P., Thirumalai, D., and Woodson, S. A. (2005) RNA tertiary interactions mediate native collapse of a bacterial group I ribozyme, *Journal of molecular biology* 353, 1199-1209.
77. Mitra, S., Laederach, A., Golden, B. L., Altman, R. B., and Brenowitz, M. (2011) RNA molecules with conserved catalytic cores but variable peripheries fold along unique energetically optimized pathways, *RNA* 17, 1589-1603.
78. Sinan, S., Yuan, X., and Russell, R. (2011) The Azoarcus group I intron ribozyme misfolds and is accelerated for refolding by ATP-dependent RNA chaperone proteins, *The Journal of biological chemistry* 286, 37304-37312.
79. Russell, R., and Herschlag, D. (1999) New pathways in folding of the Tetrahymena group I RNA enzyme, *Journal of molecular biology* 291, 1155-1167.

80. Chowrira, B. M., Berzal-Herranz, A., and Burke, J. M. (1993) Novel RNA polymerization reaction catalyzed by a group I ribozyme, *The EMBO journal* 12, 3599-3605.
81. Chowrira, B. M., Berzal-Herranz, A., and Burke, J. M. (1995) Novel system for analysis of group I 3' splice site reactions based on functional trans-interaction of the P1/P10 reaction helix with the ribozyme's catalytic core, *Nucleic acids research* 23, 849-855.
82. Riley, C. A., and Lehman, N. (2003) Generalized RNA-directed recombination of RNA, *Chemistry & biology* 10, 1233-1243.
83. Hayden, E. J., and Lehman, N. (2006) Self-assembly of a group I intron from inactive oligonucleotide fragments, *Chemistry & biology* 13, 909-918.
84. Draper, W. E., Hayden, E. J., and Lehman, N. (2008) Mechanisms of covalent self-assembly of the Azoarcus ribozyme from four fragment oligonucleotides, *Nucleic acids research* 36, 520-531.
85. Hayden, E. J., Riley, C. A., Burton, A. S., and Lehman, N. (2005) RNA-directed construction of structurally complex and active ligase ribozymes through recombination, *RNA* 11, 1678-1687.
86. Hayden, E. J., Ferrada, E., and Wagner, A. (2011) Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme, *Nature* 474, 92-95.
87. Vaidya, N., Manapat, M. L., Chen, I. A., Xulvi-Brunet, R., Hayden, E. J., and Lehman, N. (2012) Spontaneous network formation among cooperative RNA replicators, *Nature* 491, 72-77.
88. Gilbert, W. (1986) Origin of Life - the Rna World, *Nature* 319, 618-618.
89. Sprengel, G., and Follmann, H. (1981) Evidence for the Reductive Pathway of Deoxyribonucleotide Synthesis in an Archaeobacterium, *FEBS letters* 132, 207-209.
90. Fuller, W. D., Sanchez, R. A., and Orgel, L. E. (1972) Studies in prebiotic synthesis: VII. Solid-state synthesis of purine nucleosides, *Journal of molecular evolution* 1, 249-257.
91. Fuller, W. D., Sanchez, R. A., and Orgel, L. E. (1972) Studies in prebiotic synthesis. VI. Synthesis of purine nucleosides, *Journal of molecular biology* 67, 25-33.
92. Chen, M. C., Cafferty, B. J., Mamajanov, I., Gallego, I., Khanam, J., Krishnamurthy, R., and Hud, N. V. (2014) Spontaneous prebiotic formation

- of a beta-ribofuranoside that self-assembles with a complementary heterocycle, *Journal of the American Chemical Society* 136, 5640-5646.
93. Butlerow, A. (1861) Formation synthetique d'une substance sucee, *Compt. Rend. Acad. Sci.* 53, 145–147
 94. Mizuno, T., and Weiss, A. H. (1974) Synthesis and utilization of formose sugars, *Adv Carbohydr Chem Biochem* 29, 173 –227.
 95. Muller, D., Pitsch, S., Kittaka, A., Wagner, E., Wintner, C. E., and Eschenmoser, A. (1990) Chemistry of Alpha-Aminonitriles - Aldomerisation of Glycolaldehyde Phosphate to Rac-Hexose 2,4,6-Triphosphates and (in Presence of Formaldehyde) Rac-Pentose 2,4-Diphosphates - Rac-Allose 2,4,6-Triphosphate and Rac-Ribose 2,4-Diphosphate Are the Main Reaction-Products, *Helv Chim Acta* 73, 1410-1468.
 96. Joyce, G. F., Visser, G. M., van Boeckel, C. A., van Boom, J. H., Orgel, L. E., and van Westrenen, J. (1984) Chiral selection in poly(C)-directed synthesis of oligo(G), *Nature* 310, 602-604.
 97. Zubay, G. (1998) Studies on the lead-catalyzed synthesis of aldopentoses, *Origins of life and evolution of the biosphere : the journal of the International Society for the Study of the Origin of Life* 28, 13-26.
 98. Ricardo, A., Carrigan, M. A., Olcott, A. N., and Benner, S. A. (2004) Borate minerals stabilize ribose, *Science* 303, 196.
 99. Springsteen, G., and Joyce, G. F. (2004) Selective derivatization and sequestration of ribose from a prebiotic mix, *Journal of the American Chemical Society* 126, 9578-9583.
 100. Oro, J. (1961) Mechanism of synthesis of adenine from hydrogen cyanide under possible primitive earth conditions, *Nature* 191, 1193-1194.
 101. Sanchez, R. A., Ferris, J. P., and Orgel, L. E. (1967) Studies in prebiotic synthesis. II. Synthesis of purine precursors and amino acids from aqueous hydrogen cyanide, *Journal of molecular biology* 30, 223-253.
 102. Ferris, J. P., Sanchez, R. A., and Orgel, L. E. (1968) Studies in Prebiotic Synthesis III. Synthesis of Pyrimidines from Cyanoacetylene and Cyanate *Journal of molecular biology* 33, 693-704.
 103. Muller, U. F. (2006) Re-creating an RNA world, *Cellular and Molecular Life Sciences* 63, 1278-1293.

104. Powner, M. W., Gerland, B., and Sutherland, J. D. (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions, *Nature* 459, 239-242.
105. Powner, M. W., Sutherland, J. D., and Szostak, J. W. (2010) Chemoselective multicomponent one-pot assembly of purine precursors in water, *Journal of the American Chemical Society* 132, 16677-16688.
106. Unrau, P. J., and Bartel, D. P. (1998) RNA-catalysed nucleotide synthesis, *Nature* 395, 260-263.
107. Lau, M. W., Cadieux, K. E., and Unrau, P. J. (2004) Isolation of fast purine nucleotide synthase ribozymes, *Journal of the American Chemical Society* 126, 15686-15693.
108. Rohatgi, R., Bartel, D. P., and Szostak, J. W. (1996) Nonenzymatic, template-directed ligation of oligoribonucleotides is highly regioselective for the formation of 3'-5' phosphodiester bonds, *Journal of the American Chemical Society* 118, 3340-3344.
109. Sawai, H., and Orgel, L. E. (1975) Letter: Oligonucleotide synthesis catalyzed by the Zn-2+ ion, *Journal of the American Chemical Society* 97, 3532-3533.
110. Miyakawa, S., and Ferris, J. P. (2003) Sequence- and regioselectivity in the montmorillonite-catalyzed synthesis of RNA, *Journal of the American Chemical Society* 125, 8202-8208.
111. Orgel, L. E. (2004) Prebiotic chemistry and the origin of the RNA world, *Critical reviews in biochemistry and molecular biology* 39, 99-123.
112. Sleeper, H. L., and Orgel, L. E. (1979) The catalysis of nucleotide polymerization by compounds of divalent lead, *Journal of molecular evolution* 12, 357-364.
113. Ferris, J. P. (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 : the journal of the International Society for the Study of the Origin of Life* 32, 311-332.
114. Ferris, J. P., Hill, A. R., Jr., Liu, R., and Orgel, L. E. (1996) Synthesis of long prebiotic oligomers on mineral surfaces, *Nature* 381, 59-61.
115. Monnard, P. A., Kanavarioti, A., and Deamer, D. W. (2003) Eutectic phase polymerization of activated ribonucleotide mixtures yields quasi-equimolar

- incorporation of purine and pyrimidine nucleobases, *Journal of the American Chemical Society* 125, 13734-13740.
116. Engelhart, A. E., Powner, M. W., and Szostak, J. W. (2013) Functional RNAs exhibit tolerance for non-heritable 2'-5' versus 3'-5' backbone heterogeneity, *Nature chemistry* 5, 390-394.
 117. Szostak, J. W., and Ellington, A. D. (1993) *In Vitro Selection of Functional RNA Sequences*.
 118. Ellington, A. D., and Szostak, J. W. (1990) In vitro selection of RNA molecules that bind specific ligands, *Nature* 346, 818-822.
 119. Robertson, D. L., and Joyce, G. F. (1990) Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA, *Nature* 344, 467-468.
 120. Tuerk, C., and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase, *Science* 249, 505-510.
 121. Huang, W., and Ferris, J. P. (2003) Synthesis of 35-40 mers of RNA oligomers from unblocked monomers. A simple approach to the RNA world, *Chem Commun (Camb)*, 1458-1459.
 122. Joyce, G. F., and Orgel, L. E. (1999) Prospects for understanding the origin of the RNA world, *In The RNA World, 2nd ed.*, 49-77.
 123. Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E., and Bartel, D. P. (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension, *Science* 292, 1319-1325.
 124. Zaher, H. S., and Unrau, P. J. (2007) Selection of an improved RNA polymerase ribozyme with superior extension and fidelity, *RNA* 13, 1017-1026.
 125. Wochner, A., Attwater, J., Coulson, A., and Holliger, P. (2011) Ribozyme-catalyzed transcription of an active ribozyme, *Science* 332, 209-212.
 126. Attwater, J., Wochner, A., and Holliger, P. (2013) In-ice evolution of RNA polymerase ribozyme activity, *Nature chemistry* 5, 1011-1018.
 127. Yao, C. G., and Muller, U. F. (2011) Polymerase ribozyme efficiency increased by G/T-rich DNA oligonucleotides, *Rna-a Publication of the Rna Society* 17, 1274-1281.

128. Szcepanski, J. T., and Joyce, G. F. (2014) A cross-chiral RNA polymerase ribozyme, *Nature* 515, 440-442.
129. Ashley, G. W. (1992) Modeling, Synthesis, and Hybridization Properties of (L)-Ribonucleic Acid, *Journal of the American Chemical Society* 114, 9731-9736.
130. Szcepanski, J. T., and Joyce, G. F. (2013) Binding of a structured D-RNA molecule by an L-RNA aptamer, *Journal of the American Chemical Society* 135, 13290-13293.
131. Yamagata, Y., Watanabe, H., Saitoh, M., and Namba, T. (1991) Volcanic production of polyphosphates and its relevance to prebiotic evolution, *Nature* 352, 516-519.
132. Osterberg, R., and Orgel, L. E. (1972) Polyphosphate and trimetaphosphate formation under potentially prebiotic conditions, *Journal of molecular evolution* 1, 241-248.
133. Pasek, M. A., and Lauretta, D. S. (2005) Aqueous corrosion of phosphide minerals from iron meteorites: a highly reactive source of prebiotic phosphorus on the surface of the early Earth, *Astrobiology* 5, 515-535.
134. Pasek, M. A. (2008) Rethinking early Earth phosphorus geochemistry, *Proceedings of the National Academy of Sciences of the United States of America* 105, 853-858.
135. Pasek, M. A., Dworkin, J. P., and Lauretta, D. S. (2007) A radical pathway for organic phosphorylation during schreibersite corrosion with implications for the origin of life, *Geochim Cosmochim Acta* 71, 1721-1736.
136. Pasek, M. A., Kee, T. P., Bryant, D. E., Pavlov, A. A., and Lunine, J. I. (2008) Production of potentially prebiotic condensed phosphates by phosphorus redox chemistry, *Angew Chem Int Ed Engl* 47, 7918-7920.
137. Pasek, M. A., Harnmeijer, J. P., Buick, R., Gull, M., and Atlas, Z. (2013) Evidence for reactive reduced phosphorus species in the early Archean ocean, *Proceedings of the National Academy of Sciences of the United States of America* 110, 10089-10094.
138. Etaix, E., and Orgel, L. E. (1978) Phosphorylation of Nucleosides in Aqueous-Solution Using Trimetaphosphate - Formation of Nucleoside Triphosphates, *J Carb-Nucleos-Nucl* 5, 91-110.
139. Cheng, C., Fan, C., Wan, R., Tong, C., Miao, Z., Chen, J., and Zhao, Y. (2002) Phosphorylation of adenosine with trimetaphosphate under simulated prebiotic conditions, *Origins of life and evolution of the*

biosphere : the journal of the International Society for the Study of the Origin of Life 32, 219-224.

140. Ballinger, P., and Long, F. A. (1960) Acid Ionization Constants of Alcohols .2. Acidities of Some Substituted Methanols and Related Compounds, *Journal of the American Chemical Society* 82, 795-798.
141. Izatt, R. M., Rytting, J. H., Hansen, L. D., and Christensen, J. J. (1966) Thermodynamics of proton dissociation in dilute aqueous solution. V. An entropy titration study of adenosine, pentoses, hexoses, and related compounds, *Journal of the American Chemical Society* 88, 2641-2645.
142. Moretti, J. E., and Muller, U. F. (2014) A ribozyme that triphosphorylates RNA 5'-hydroxyl groups, *Nucleic acids research* 42, 4767-4778.
143. Fiskaa, T., and Birgisdottir, A. B. (2010) RNA reprogramming and repair based on trans-splicing group I ribozymes, *New biotechnology* 27, 194-203.

Chapter 2

Trans-splicing with the group I intron ribozyme from *Azoarcus*

2.1 Abstract

Group I introns are ribozymes (catalytic RNAs) that excise themselves from RNA primary transcripts by catalyzing two successive transesterification reactions. These *cis*-splicing ribozymes can be converted into *trans*-splicing ribozymes, which can modify the sequence of a separate substrate RNA, both in vitro and in vivo. Previous work on *trans*-splicing ribozymes has mostly focused on the 16S rRNA group I intron ribozyme from *Tetrahymena thermophila*. Here we test the *trans*-splicing potential of the tRNA^{Ile} group I intron ribozyme from the bacterium *Azoarcus*. This ribozyme is only half the size of the *Tetrahymena* ribozyme and folds faster into its active conformation in vitro. Our results showed that in vitro, the *Azoarcus* and *Tetrahymena* ribozymes favored the same set of splice sites on a substrate RNA. Both ribozymes showed the same *trans*-splicing efficiency when containing their individually optimized 5'-terminus. In contrast to the previously optimized 5'-terminal design of the *Tetrahymena* ribozyme, the *Azoarcus* ribozyme was most efficient with a *trans*-splicing design that resembled the secondary structure context of the natural *cis*-splicing *Azoarcus* ribozyme, which includes base pairing between the substrate 5'-portion and the ribozyme

3'-exon. These results suggested preferred *trans*-splicing interactions for the *Azoarcus* ribozyme under near-physiological in vitro conditions. Despite the high activity in vitro, however, the splicing efficiency of the *Azoarcus* ribozyme in *E. coli* cells was significantly below that of the *Tetrahymena* ribozyme.

2.2 Introduction

Group I introns are catalytic RNAs (ribozymes) that do not require the spliceosome for their removal from primary transcripts. Instead, these ribozymes fold into catalytically active structures that undergo two successive transphosphorylation reactions, facilitating their own excision and the ligation of the two flanking exons ((1); for review see (2)). These natural *cis*-splicing ribozymes can be converted into *trans*-acting ribozymes by removing their 5'-exon and modifying the ribozyme 5'-terminus. In this format, *trans*-acting group I intron ribozymes can cleave a substrate RNA (3), insert or remove one or several nucleotides (4, 5), replace the 3'-portion of the substrate RNA with the 3'-exon of the ribozyme (3, 6), or replace the 5'-portion of the substrate RNA with the 5'-exon of the ribozyme (7). These replacement reactions may be useful for therapeutic applications by either repairing the sequence of genetically mutated mRNAs (6) or by inserting the coding sequence for a toxic peptide into substrate mRNAs that indicate a disease state (8, 9). Most previous work has focused on the reaction that replaces the mRNA 3'-portion; this reaction is also the focus of the present study.

Trans-splicing group I intron ribozymes recognize the target site on the substrate RNA by base pairing, forming the P1 helix with the ribozyme's 5'-terminus (Figure 2.1). The only sequence requirement in this helix is that a uracil at the splice site of the substrate is base paired to a guanine in the ribozyme, the 5'-terminal nucleotide of the internal guide sequence (IGS) (10). Factors that affect the efficiency of the *trans*-splicing reaction include the accessibility of the target site in the substrate RNA, and the strength of the base pairing interaction between ribozyme IGS and substrate RNA (11, 12).

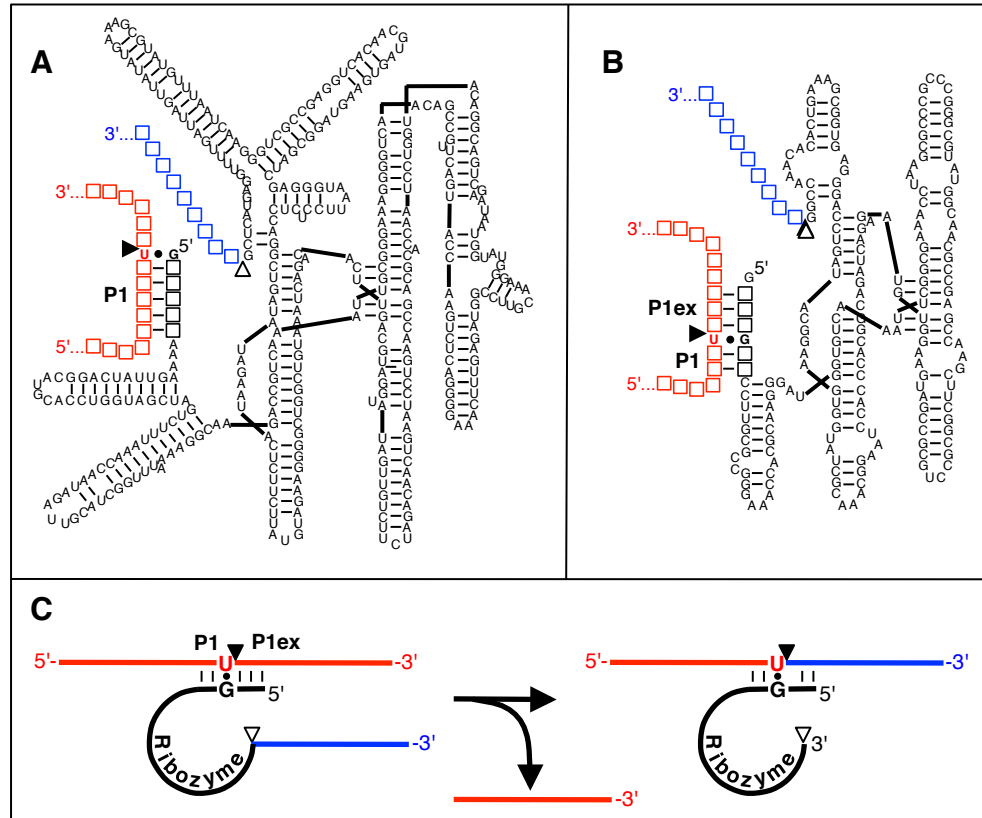


Figure 2.1: Secondary structure representations of *trans*-splicing group I intron ribozymes from (A) *Tetrahymena* and (B) *Azoarcus*. The 5'-terminus of the *trans*-splicing ribozymes (black) is truncated at or near the splice site G:U base pair (filled circle) such that the P1 helix (P1) is formed between the ribozyme 5'-terminus and the target site on the substrate (red). This helix is extended on the *Azoarcus* ribozyme construct by a P1 helix extension (P1ex) to facilitate a similar ribozyme-substrate complex stability as with the *Tetrahymena* ribozyme. A 5'-terminal guanosine was added for better in vitro transcription yields. Filled triangles indicate the 5'-splice sites, while empty triangles denote the 3'-splice sites. Empty squares denote nucleotides that differ between splice sites. (C) Schematic of a *trans*-splicing reaction. The colors for substrate and ribozyme 3'-exon are as in (A) and (B), while sequences are denoted by solid lines (with exception of the splice site G:U pair), and the ribozyme is denoted by a solid black line. During the *trans*-splicing reaction the ribozyme 3'-exon (blue) replaces the 3'-portion of the substrate.

The efficiency of the *trans*-splicing reaction can be improved by increasing the substrate-ribozyme interaction via an extended guide sequence (EGS). The EGS is an elongation of the 5' terminus of the *trans*-splicing ribozyme, and has been tested on the group I intron ribozymes from *Tetrahymena thermophila* (sub-

group IC1), *Didymium* (sub-group IE), and *Fuligo* (sub-group IC1) (13-16).

Despite the improvements mediated by the EGS, the *trans*-splicing efficiency remains a limiting factor for potential therapeutic applications of these ribozymes (17). Notably, all of these previous studies employed the same EGS design principle, which was originally developed for the *Tetrahymena* ribozyme.

Because different group I introns may prefer different EGSs, it may be possible to obtain higher *trans*-splicing efficiencies by optimizing the EGSs for ribozymes other than the *Tetrahymena* ribozyme.

To date, only four different group I introns have been tested for their *trans*-splicing potential replacing the substrate 3'-portion (6, 16, 18), compared to the more than 16,000 known group I introns (19). Of these four, the *Tetrahymena* ribozyme achieved the highest *trans*-splicing efficiencies in vitro, is understood at the most detailed biochemical level, and remains the only group I intron tested for mRNA repair / replacement *trans*-splicing in vivo (20-24). Therefore, to achieve efficient *trans*-splicing, it is promising to analyze group I intron ribozymes from different species and identify specific designs that increase their respective *trans*-splicing efficiencies.

The group I intron ribozyme from *Azoarcus* BH72 is an attractive candidate for *trans*-splicing experiments. It belongs to the IC3 group I intron family and is located in the anticodon stem-loop of tRNA^{Ile} (25). With a length of 205 nucleotides, it is about 2-fold smaller than the *Tetrahymena* ribozyme (Figure 2.1A, B). The *Azoarcus* ribozyme folds into its active conformation faster than the *Tetrahymena* ribozyme in vitro (26-29), and it can mediate *trans*-ligation

reactions at high concentrations of magnesium, ribozyme, and at elevated temperature (30-32). The structure of the *Azoarcus* ribozyme is known at the highest detail of all group I intron ribozymes, for different stages of the splicing pathway (33-35). Due to its fast folding kinetics and its ability to react in trans, the *Azoarcus* ribozyme is a promising candidate to test its potential for *trans*-splicing under near-physiological in vitro conditions and in cells.

Here we characterize the *trans*-splicing potential of the group I intron ribozyme from *Azoarcus* at near-physiological in vitro conditions and in *E. coli* cells. The results show that in vitro, the *Azoarcus* ribozyme favors the same splice sites on a model substrate mRNA as the *Tetrahymena* ribozyme. When both ribozymes are designed with their preferred secondary structure interactions they achieve similar *trans*-splicing efficiencies in vitro. Importantly, the favored secondary structure context of the *Azoarcus* ribozyme is different from that of previously established *trans*-splicing ribozymes, and reflects the natural *cis*-splicing context of the *Azoarcus* ribozyme. These results suggested how the *Azoarcus* ribozyme could be used for *trans*-splicing reactions under near-physiological in vitro conditions. In *E. coli* cells, the *trans*- and *cis*- splicing activities of the *Azoarcus* ribozyme were significantly lower than the *Tetrahymena* ribozyme activities.

2.3 Results

To analyze the potential of the *Azoarcus* ribozyme for *trans*-splicing

applications we proceeded in three steps. First, we compared the pattern of favored splice sites on a model mRNA between *trans*-splicing *Azoarcus* and *Tetrahymena* ribozymes. Second, we searched and identified an efficient secondary structure design for ribozyme-substrate interactions of *trans*-splicing *Azoarcus* ribozymes, which is different from that of the *Tetrahymena* ribozyme. Third, we tested the activity of *trans*-splicing *Azoarcus* ribozymes in *E. coli* cells.

Splice site preferences by trans-splicing Azoarcus and Tetrahymena ribozymes

To test whether the *Azoarcus* ribozyme could access the same splice sites as the *Tetrahymena* ribozyme, we determined the accessible splice sites on a model substrate mRNA, the chloramphenicol acetyl transferase (*CAT*) mRNA. Splice sites were identified using an RT-PCR assay that specifically amplifies *trans*-splicing products from a reaction with ribozymes containing a randomized substrate recognition sequence (11). The randomized substrate recognition sequence enabled the pool of ribozymes to splice at any accessible U in the substrate RNA because the phylogenetically conserved G•U pair in the P1 helix is the only requirement for ribozyme activity (10). The P1 helix of the *Azoarcus* ribozyme was elongated by 3 nucleotides beyond the splice site such that both *Azoarcus* and *Tetrahymena* ribozymes formed similarly stable six-base-pair P1 duplexes with the substrate (Figure 2.1B). After RT-PCR, cloning and sequencing of *trans*-splicing products, the identified splice sites for 56 *trans*-splicing events of the *Azoarcus* ribozyme were mapped on the *CAT* mRNA and compared to 56 *trans*-splice sites of the *Tetrahymena* ribozyme, in experiments

conducted side-by-side (12) (Figure 2.2A). Overall, the splice site patterns of the *Azoarcus* and *Tetrahymena* ribozymes were very similar. The region between position ~200 and ~400 reported few splicing events for both ribozymes. To identify additional splice sites in this region we modified the assay by using a nested 5'-PCR primer (gray symbols in Figure 2.2A). Comparison of all splice sites between both ribozymes showed no significant difference in the profile, suggesting that the *Azoarcus* and *Tetrahymena* ribozymes share the same splice site preference pattern.

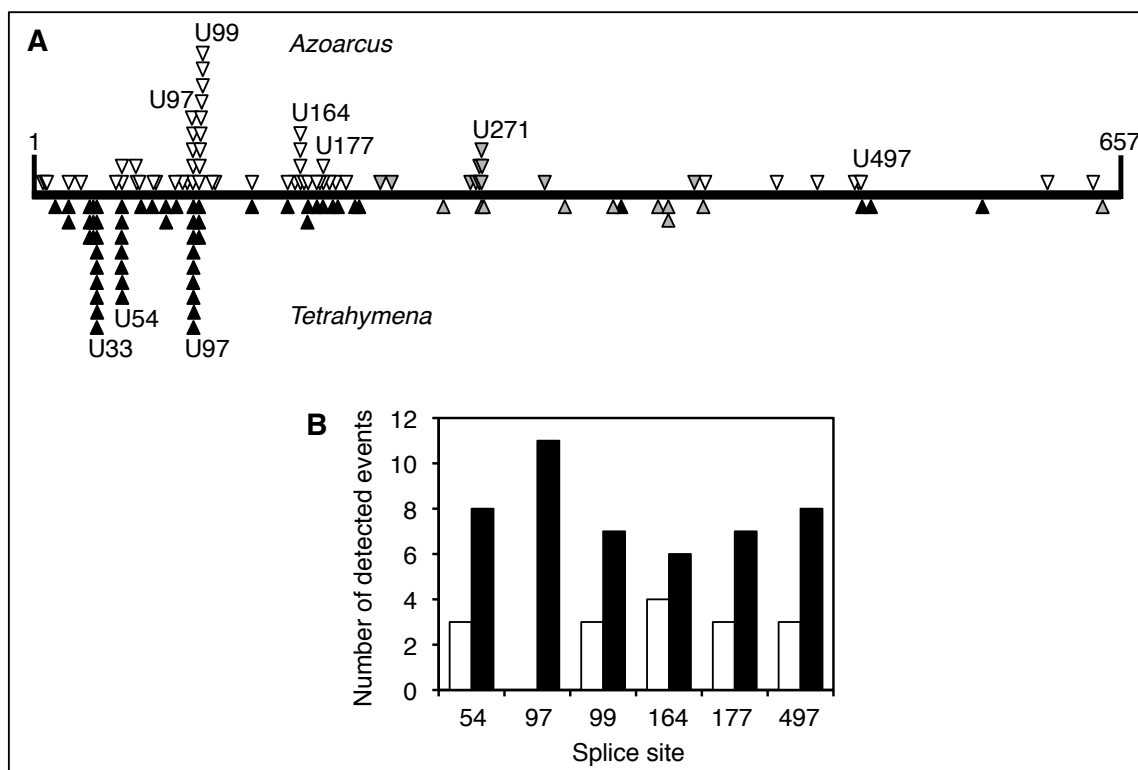


Figure 2.2: Experimental comparison of favored splice sites and *trans*-splicing efficiencies between *Tetrahymena* and *Azoarcus*. (A) Patterns of splicing events along the sequence of *CAT* mRNA substrate (horizontal line) between positions 1 and 657. The events detected for the *Azoarcus* ribozyme (above the line; open triangles) are compared to the events recorded for the *Tetrahymena* ribozyme (below the line; closed triangles; from (12)). Open and closed triangles denote splicing events that were detected with RT-PCR primers covering the entire length of the *CAT* mRNA. Gray triangles denote splicing events that were detected with a nested 5'-PCR primer to identify splice sites at least 200 nucleotides downstream of the 5'-terminus of the *CAT* mRNA. All identified splicing products are reported here, which includes ten non-U splice sites for the *Azoarcus* ribozyme and seven non-U splice sites for the *Tetrahymena* ribozyme. The positions of all individual splice sites are listed in table S1. (B) Comparison of *trans*-splicing efficiencies of *Azoarcus* ribozymes (white columns) and *Tetrahymena* ribozymes (black columns) when targeted to six specific splice sites on *CAT* mRNA.

The previous assay reported the pattern of favored splice sites but not the relative reaction efficiencies of both ribozymes. To estimate the relative efficiencies of the two *trans*-splicing ribozymes, they were reacted at equimolar concentrations in the same reaction vessel, with an excess of *CAT* mRNA.

Therefore, the number of splicing events detected for each ribozyme revealed the relative reaction efficiency of the two ribozymes (Figure 2.2B). The splicing products of the two ribozymes were discriminated by introducing a single point mutation into one of the ribozyme 3'-exons, 26 nucleotides downstream of the splice site. Because this mutation became part of the product, the sequence of the *trans*-splicing products reported which of the two ribozymes had generated the product. Six representative splice sites were chosen, some of which were favored by the *Tetrahymena* ribozyme in the previous assay, while others were favored by *Azoarcus* ribozyme (see Table S1). To target these six splice sites, the IGSs of the ribozymes had six specific sequences complementary to the six target sites on the *CAT* mRNA. After the *trans*-splicing reaction, reverse transcription, and cloning, 10-13 splicing products from each of the six reactions were sequenced. At all six splice sites, at least 10 sequences reported splicing at the targeted splice site (>80%), and at most 2 sequences (<20%) reported mis-splicing at a different splice site. For both ribozymes, a total of three mis-splicing events were detected, suggesting a similar, low error rate under these conditions (~9%). Importantly, the *Tetrahymena* ribozyme gave rise to more splicing products than the *Azoarcus* ribozyme on all six tested splice sites (Figure 2.2B). This outcome was not due to the point mutation in the ribozyme 3'-exons, because the result was the same when the point mutation in the ribozyme 3'-exon was switched between the ribozymes. These results indicated that the *Tetrahymena* ribozyme *trans*-spliced more efficiently than the *Azoarcus* ribozyme in the absence of optimized ribozyme-substrate contacts.

Design of optimal ribozyme-substrate contacts for the Azoarcus ribozyme

The extended guide sequence (EGS) is an extension of the ribozyme 5'-terminus that can improve the *trans*-splicing efficiency and specificity. This was originally demonstrated with the group I intron ribozyme from *Tetrahymena* (13), where the optimal design of the EGS includes an extension of the P1 helix by 2-4 base pairs (P1 extension), followed by an internal loop, and a duplex with at least 5 base pairs between the ribozyme 5'-terminus and the substrate RNA. The internal loop region of the EGS may also form a P10 duplex with the 3'-exon, which can benefit the *trans*-splicing efficiency (13-15) (Figure 2.3A). Subsequent studies found that this 'classical design' also increases the *trans*-splicing efficiencies of the group I introns from *Fuligo* and *Didymium* (16, 18). To test whether the 'classical design' of ribozyme-substrate contacts also increases the *trans*-splicing efficiency of the *Azoarcus* ribozyme, we generated *Azoarcus* and *Tetrahymena* ribozyme variants that formed a P1 extension, an internal loop, and a 5'-duplex (Figure S2.1). Because the formation of a P10 duplex does not always increase *trans*-splicing efficiency (15), we tested several constructs for each ribozyme and splice site with either a strong, a weak, or no predicted P10 duplex. Three splice sites were tested on the *CAT* mRNA (sites 97, 177, and 258, with position 1 being the A of the AUG start codon) to reduce a possible bias from the choice of splice sites. These variations in splice site, P10 duplex, and P1 extension length generated seven reaction contexts for each of the two ribozymes. Three of the seven *Tetrahymena* ribozyme constructs gave by far the highest *trans*-splicing efficiency, with a conversion of ~20% substrate to product,

all on splice site 258 (Figure 2.3B). In contrast, the three most efficient reactions of the *Azoarcus* ribozyme achieved efficiencies only up to 7%. The low efficiency of *trans*-splicing with all *Azoarcus* ribozyme constructs compared to the *Tetrahymena* ribozyme constructs suggested that the 'classical design' was not well suited for the *Azoarcus* ribozyme.

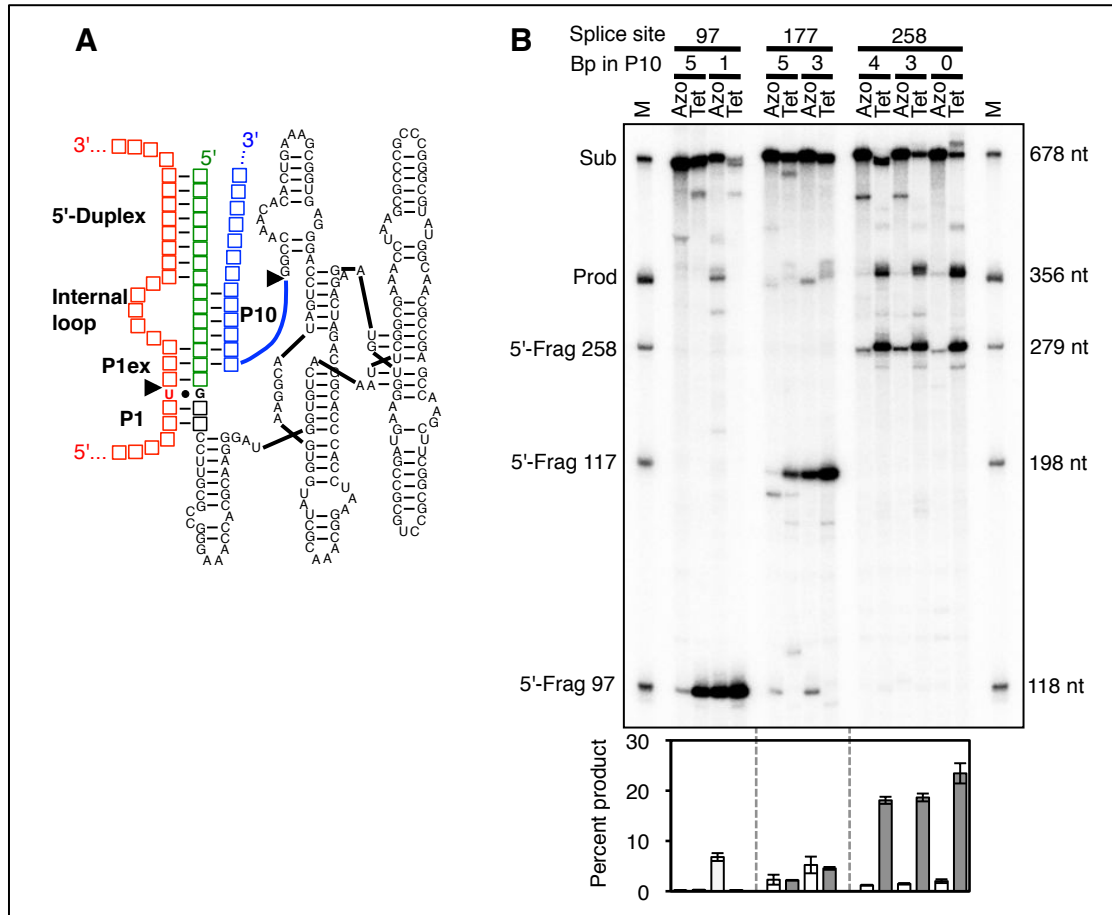


Figure 2.3: 'Classical design' for *trans*-splicing group I intron ribozymes. (A) Secondary structure of the *Azoarcus* ribozyme (black) with a 5'-terminal extended guide sequence (EGS; green), and 3'-exon (blue), formed with a pre-mRNA (red). The positions of the P1 helix (P1), P1 extension (P1ex), internal loop, and 5'-duplex are indicated. Each empty square symbolizes a position where nucleotide identities change between splice sites, while maintaining the secondary structure. Individual constructs differ by nucleotide insertions, deletions, and mutations (see figure S1). (B) Autoradiogram of 5'-radiolabeled *trans*-splicing products separated by denaturing 5% polyacrylamide gel electrophoresis. The targeted splice site (97, 177, or 258) and the designed number of base pairs (bp) in the P10 helix (bp in P10) are indicated on top of the autoradiogram. The side of the autoradiogram indicates the size of markers in the first and last lane (M), which correspond to full-length CAT mRNA substrate (Sub), the expected *trans*-splicing product (Prod), and the 5'-fragments produced by the first step of splicing (5'-Frag), at splice sites 97, 177, and 258. The length of the *trans*-splicing products is the same for all splice sites because the length of the ribozyme 3'-exons was co-varied with the splice sites. The quantitations of *trans*-splicing efficiencies are shown below the autoradiogram, as percent of substrate converted to product. Columns for the *Azoarcus* ribozyme (white) and the *Tetrahymena* ribozymes (grey) are in the same order as the lanes in the autoradiogram. Error bars denote standard deviations from three experiments.

An alternative design for *trans*-splicing *Azoarcus* ribozymes is a simple elongation of the P1 duplex past the splice site, termed the P1 extension (Figure 2.4A). We expected that the length of this P1 extension had an optimum because lengthening the P1 extensions would not only increase the strength of the ribozyme-substrate duplex but would also reduce the splice site specificity (36) and may prevent conformational changes of the ribozyme between the two catalytic steps of splicing (37). To test this design principle and identify optimal P1 extension lengths, we analyzed a total of 32 *Azoarcus* ribozyme constructs with P1 extension lengths between 3 base pairs and 63 base pairs, at the three splice sites 97, 177, or 258 (Figure 2.4B). Again, none of these constructs exceeded a *trans*-splicing efficiency of 7%. Although there was no clear optimal length of the P1 extension, the most robust *trans*-splicing appeared at a P1 length of 16-18 base pairs, where all constructs had a *trans*-splicing efficiency between 3.5% and 7%. Because these efficiencies were significantly below the *trans*-splicing efficiencies achieved with the classic design principle on the *Tetrahymena* ribozyme, these results suggested that elongated P1 helices alone (the 'P1 extension design') are not sufficient to generate efficient *trans*-splicing *Azoarcus* ribozymes.

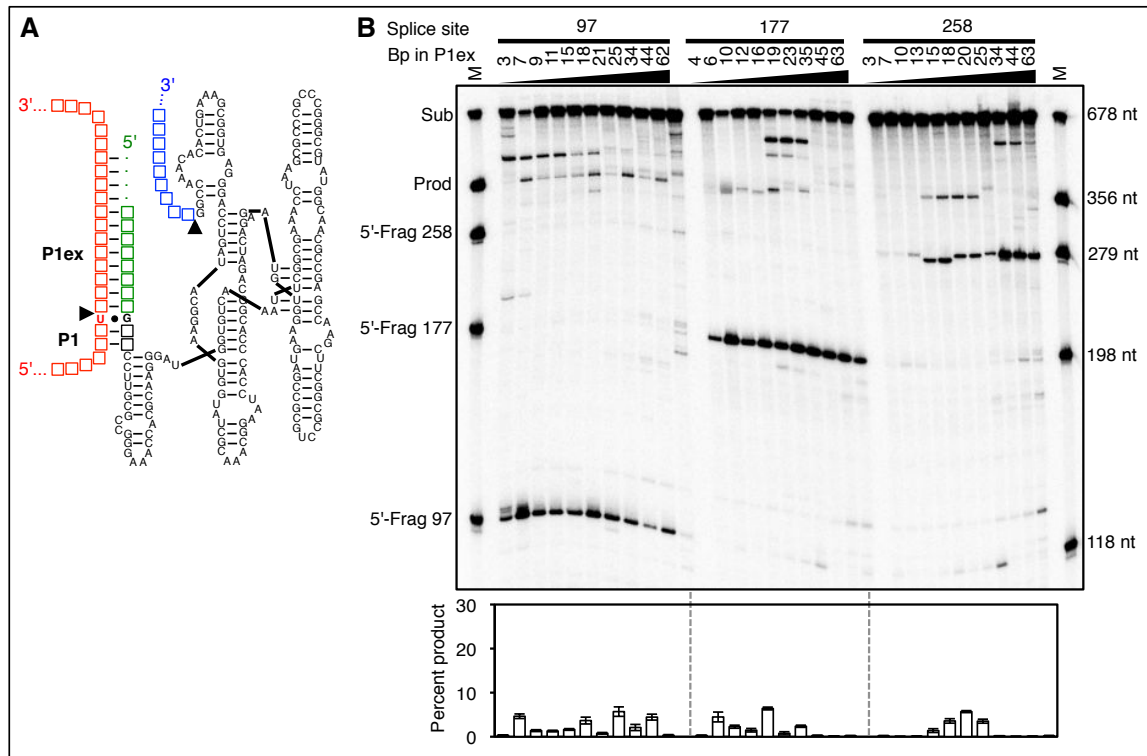


Figure 2.4: 'P1 extension design' for the *trans*-splicing *Azoarcus* ribozyme. (A) Secondary structure of the *Azoarcus* ribozyme (black), with a 5'-terminal extension of variable length (green). Depending on the length of this 5'-extension, the P1 extension helices (P1ex) are formed with corresponding lengths, stacked upon the P1 helix (P1). The ribozyme 3'-exon is in blue. (B) Autoradiogram of 5'-radiolabeled *trans*-splicing products separated by denaturing polyacrylamide gel electrophoresis. The targeted splice site (97, 177, or 258) and the number of base pairs in the P1 extension are indicated above the autoradiogram. Size markers in the first and last lane indicate the position of full-length CAT mRNA substrate (Sub), the expected *trans*-splicing product (Prod), and the 5'-fragments produced by the first step of splicing (5'-Frag), at splice sites 97, 177, and 258. The length of the *trans*-splicing products is the same for all splice sites because the length of the ribozyme 3'-exons was co-varied with the splice sites. The quantitations of *trans*-splicing efficiencies are shown below the autoradiogram, as percent of substrate converted to product. The scale is the same as in figures 3 and 5 to facilitate comparison. Error bars denote standard deviations from three experiments.

In a third design, we tested whether *trans*-splicing *Azoarcus* ribozymes would benefit from a secondary structure context similar to their native structural context near the splicing junction, the anticodon stem of tRNA^{Ile} (25, 29, 34) (Figure 2.5A). This anticodon stem contains five base pairs between the 5'-exon

and the 3'-exon, three nucleotides distant from the 5'- and 3'-splice site, respectively. In the context of *trans*-splicing *Azoarcus* ribozymes, this means that a sequence in the 3'-exon of the ribozyme needs to form base pairs with the 5'-portion of the substrate. In addition to the anticodon stem, the natural *cis*-splicing context of the *Azoarcus* ribozyme also forms a P1 extension helix before the first step of splicing (34). Because the optimal length of the P1 extension helix for the *trans*-splicing reaction was unclear, three to four different P1 extension lengths were tested for each splice site (97, 177, 258). The choice of these P1 extension lengths was based on the highest efficiency in the previous experiment, with a preference for short P1 extensions (Figure 2.4B). Satisfyingly, two of the 11 constructs resulted in *trans*-splicing efficiencies of ~20%, similar to the best *trans*-splicing efficiencies of the *Tetrahymena* ribozyme with the 'classical design' (Figure 2.3B). This suggested that the 'anticodon stem design' is a promising strategy to improve *trans*-splicing by the *Azoarcus* ribozyme. Because both the *Tetrahymena* ribozyme with the 'classical EGS' and the *Azoarcus* ribozyme with the 'anticodon stem design' showed high activity for only one splice site (see Figure 2.3B and Figure 2.5B), we suspected that the sequences flanking the splice sites might inhibit *trans*-splicing on other splice sites, which has been observed for the *Tetrahymena* ribozyme (12). This inhibition could have been masking the general trend that the *Azoarcus* ribozyme favors the 'anticodon stem design'.

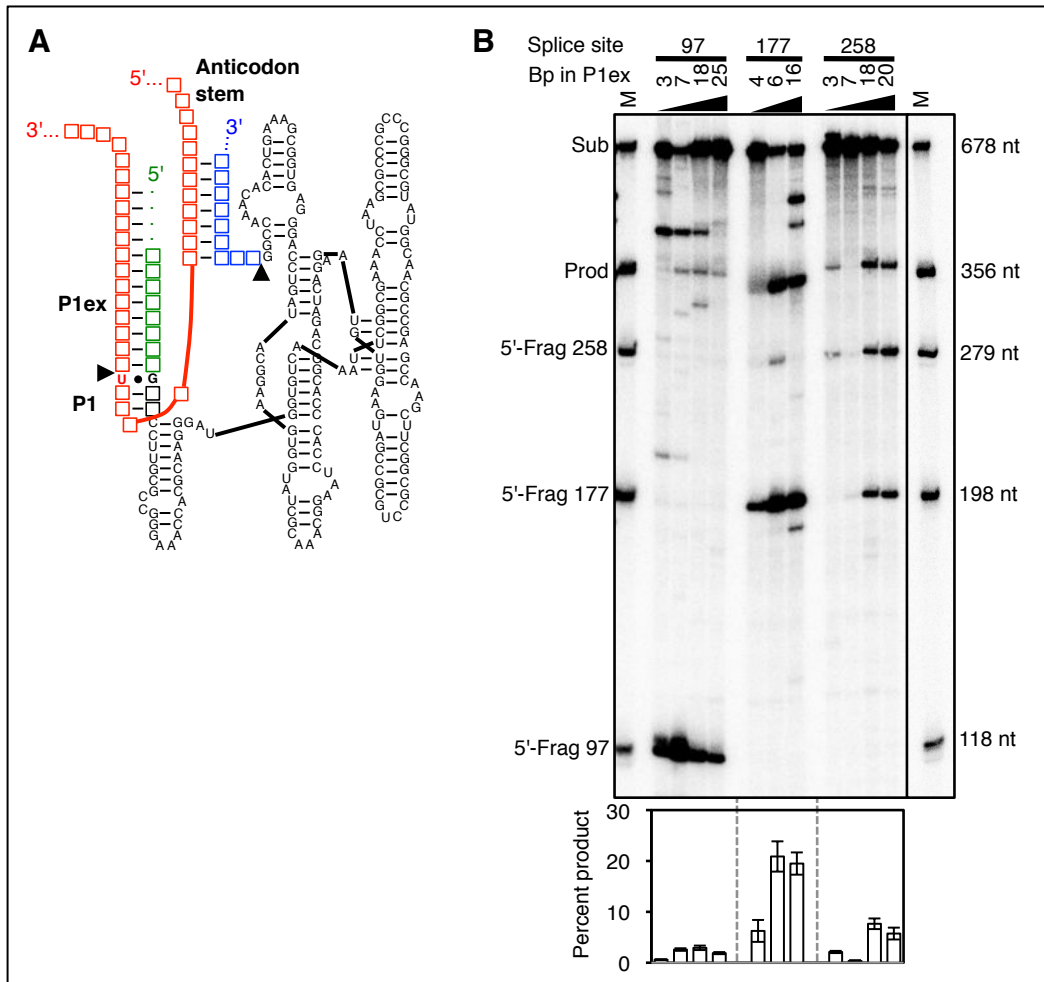


Figure 2.5: Effect of the 'anticodon stem design' on the in vitro *trans*-splicing efficiency with *Azoarcus* ribozymes. (A) Schematic for the secondary structure, with a focus on interactions between substrate mRNA (red), ribozyme 5'-extension (green), and ribozyme 3'-exon (blue). The secondary structure is based on the model describing the *Azoarcus* ribozyme crystal structure (34). The positions of the P1 helix (P1), P1 extension (P1ex), and anticodon stem are indicated. (B) Autoradiogram of 5'-radiolabeled *trans*-splicing products separated by denaturing polyacrylamide gel electrophoresis. The targeted splice site and the designed number of base pairs in the P1 extension (P1ex) are indicated for each ribozyme variant. Size markers in the first and last lane indicate the position of full-length CAT mRNA substrate (Sub), the expected *trans*-splicing product (Prod), and the 5'-fragments produced by the first step of splicing (5'-Frag) at splice sites 97, 177, and 258. The length of the *trans*-splicing products is the same for all splice sites because the length of the ribozyme 3'-exon was co-varied with the splice sites. The quantitations of *trans*-splicing efficiencies are shown below the autoradiogram, as percent of substrate converted to product. The scale is the same as in figures 3 and 4 to facilitate comparison. Error bars denote standard deviations from three experiments.

To test the 'anticodon stem design' for *trans*-splicing *Azoarcus* ribozymes without interference from flanking sequences, the *trans*-splicing reactions were repeated with short substrate RNAs (30 nt) and short ribozyme 3' exons (10 nt) (Figure S2.3). The three short substrate RNAs that were used corresponded to the three splice sites tested on the full-length RNAs (97, 177, and 258). These RNAs were long enough to establish the structural elements used by all three design principles (P1 extensions, internal loops, 5'-duplexes, P10 duplexes, anticodon stems) but short enough to minimize potentially inhibitory secondary structure formation with the flanking sequences. For each of the three substrates and for each of the three design principles, two ribozyme constructs were chosen that showed the highest *trans*-splicing efficiency with full-length *CAT* mRNA substrates, and whose P1 extension lengths were short enough to fit on the short substrates. In general, the *trans*-splicing efficiencies were higher with these short substrates than with the full-length *CAT* mRNA, confirming that the flanking sequences reduced the *trans*-splicing efficiency on the full-length constructs (Figure 2.6; compare to efficiencies in Figures 2.3, 2.4, and 2.5). Importantly, the 'anticodon stem design' generally resulted in the highest *trans*-splicing efficiencies (Figure 2.6). Specifically, all 6 constructs with the 'anticodon stem design' showed efficiencies of 19 - 82%, while only 2 / 6 constructs mediated efficiencies above 4% for both the 'classic design' and 'P1 extension design'. In only one case did the 'classical design' generate a higher *trans*-splicing efficiency than the 'anticodon stem design'. This specific construct (splice site 177, P10 length of 3 base pairs) had an advantage over all other constructs because it was

previously optimized by an in vivo selection procedure (15). Note that on each splice site, the two 'P1 extension' constructs had the same length of the P1 extension as in the two 'anticodon stem' constructs. Comparison between these six pairs of constructs confirmed the beneficial effect of the anticodon stem interaction. Together, these results confirmed that the 'anticodon stem design' is the design of choice for *trans*-splicing *Azoarcus* ribozymes.

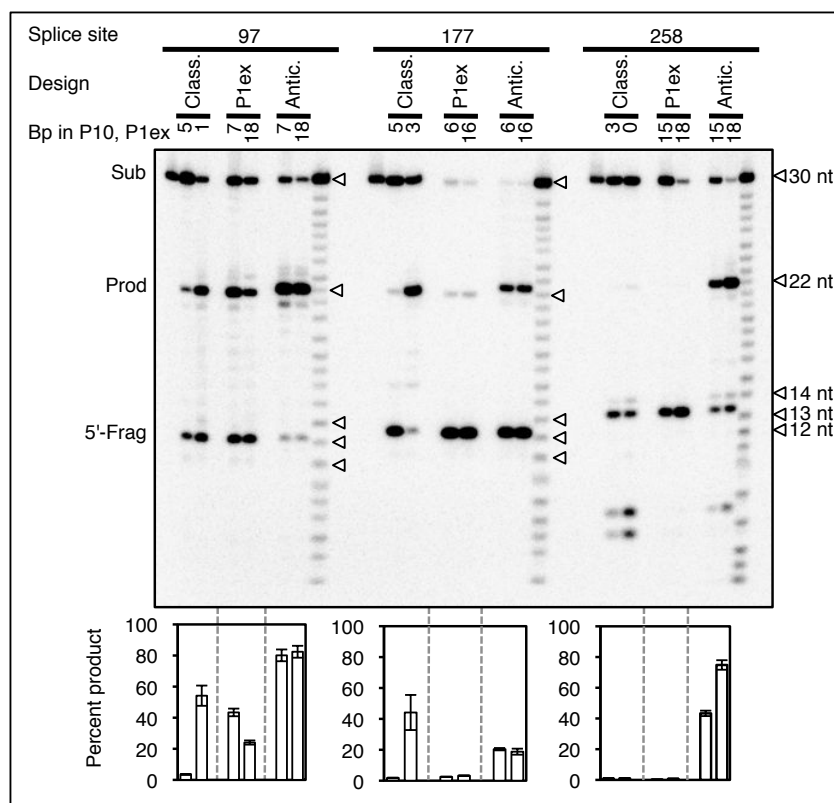


Figure 2.6: Comparison of all three designs for the in vitro *trans*-splicing efficiency with the *Azoarcus* ribozyme on short substrates (30 nt) and short ribozyme 3'-exons (10 nt). For secondary structure schematics of all individual constructs see figure S3. Shown is an autoradiogram of 5'-radiolabeled *trans*-splicing products separated by denaturing polyacrylamide gel electrophoresis. The targeted splice sites and the ribozyme designs are indicated on top of the autoradiogram. Ribozyme designs were the classical design (Class.), the P1 extension design (P1ex), and the anticodon stem design (Antic.). For each design the number of base pairs in the P10 helix (classic design) or the P1 extension (P1ex design and Antic. design) are indicated. For each splice site, the lane on the left shows unreacted substrate and the lane on the right shows an alkaline hydrolysis ladder. The position of full-length substrate (Sub), the expected *trans*-splicing product (Prod), and the 5'-fragments produced by the first step of splicing (5'-Frag), are indicated on the left. The length of the *trans*-splicing products is the same for all splice sites because the location of the splice site is the same on all short substrates. Note that the 5'-fragments (12 nucleotides) migrate corresponding to a length of 13-14 nucleotides because they contain a 3'-hydroxyl terminus while all other labeled RNAs carry a 2'-3' cyclic phosphate at their 3'-terminus, due to their synthesis method. Note that the 5'-fragments at splice site 258 are shifted up relative to those of splice sites 97 and 177, and that the *trans*-splicing products with the anticodon stem design are shifted up relative to the other two designs, due to a higher purine content in their sequence. The quantitations of *trans*-splicing efficiencies are shown below the autoradiogram, as percent of substrate converted to product. The scale is larger than in figures 3, 4, and 5 due to the higher *trans*-splicing efficiency with short substrates. Error bars denote standard deviations from three experiments.

In vivo activity of the Azoarcus ribozyme in E. coli cells

The efficiency of *trans*-splicing *Azoarcus* ribozymes in *E. coli* cells was tested by expressing them from plasmids that encoded the *trans*-splicing ribozyme construct and its substrate, the chloramphenicol acetyl transferase (*CAT*) mRNA. The *CAT* mRNA was inactivated by a frameshift mutation and the ribozyme 3'-exon was designed to repair this mutation, such that only cells expressing an efficient *trans*-splicing ribozyme were able to grow on medium containing chloramphenicol (15). Despite testing more than twenty constructs covering all design principles described in the *in vitro* portion of this study, no *trans*-splicing *Azoarcus* ribozyme facilitated growth on medium containing chloramphenicol (data not shown). In contrast, the *Tetrahymena* ribozyme with the 'classical design' mediated efficient growth. To quantify the *trans*-splicing efficiency in cells, we measured the abundance of spliced *CAT* mRNA using an RT-qPCR assay (data not shown). The results paralleled the observed chloramphenicol resistance phenotypes: In *E. coli* cells the *trans*-splicing efficiency of the *Tetrahymena* ribozyme using the 'classical' EGS was ~10% while it was <0.1% for all *Azoarcus* ribozyme constructs tested.

To identify the cause for the low *in vivo* efficiency of the *Azoarcus* ribozyme, several *cis*-splicing constructs of *Azoarcus* and *Tetrahymena* ribozymes were designed and tested for activity in *E. coli* cells (Figure 2.7). Four of these constructs expressed *CAT* pre-mRNAs with the respective ribozymes inserted at splice sites 97, 177, or 258. The *Tetrahymena* ribozyme was inserted

only at splice site 177 as a positive control. Only the *Tetrahymena* ribozyme construct mediated chloramphenicol resistance of *E. coli* cells, while none of the *Azoarcus* ribozyme constructs did. Quantitation of the *cis*-splicing efficiency by RT-qPCR confirmed the observed chloramphenicol resistance phenotype, showing that ~60% of the pre-mRNAs containing the *Tetrahymena* ribozyme were converted to product, whereas <2% of the pre-mRNAs containing the *Azoarcus* ribozyme were spliced (Figure 2.7, top four constructs). In vitro, however, the same *cis*-splicing *Azoarcus* ribozyme constructs that were inactive in cells, were similarly active as the *Tetrahymena* ribozyme (Figure S2.4). The low activity of the *Azoarcus* ribozyme in *E. coli* did not seem to be caused by a higher RNA turnover rate because the total CAT pre-mRNA levels as measured by RT-qPCR were at least 2.5-fold higher than with the *Tetrahymena* ribozyme constructs. *Cis*-splicing *Azoarcus* ribozyme constructs also failed to mediate chloramphenicol resistance in *E. coli* cells at elevated growth temperatures of 40°C and 42°C, closer to the optimal temperature of the *Azoarcus* ribozyme (38, 39). Similarly, no chloramphenicol resistance was seen when the *E. coli* growth medium was supplemented with 5 mM magnesium ions, the same concentration that led to efficient *cis*- and *trans*-splicing in vitro. These results indicated that the low *trans*-splicing activity of *Azoarcus* ribozymes in *E. coli* cells was not due to their *trans*-splicing context or increased turnover.

To test whether the native flanking sequences of the *Azoarcus* ribozyme could facilitate *Azoarcus* ribozyme activity in *E. coli* cells, the construct CisAzoCAT_177 was modified by positioning the two halves of the complete 79-

nucleotide long native tRNA^{Ile} sequence (29) at the flanks of the *Azoarcus* ribozyme, thereby replacing the corresponding *CAT* sequence (Figure 2.7, construct CisAzoCAT_tRNA). No significant *cis*-splicing activity was detected in *E. coli* cells by RT-qPCR. Because the *Azoarcus* ribozyme is naturally expressed as a non-coding RNA, we tested whether the translational apparatus in *E. coli* interfered with *cis*-splicing activity. To do that, the ribosome binding site and the start codon were mutated (construct CisAzoCAT_tRNA -RBS). Again, no significant activity was detected. The remaining *CAT* sequence was removed to test whether the exons of the *CAT* mRNA inhibited the *Azoarcus* ribozyme in *E. coli* cells (construct CisAzo_tRNA). Here, the *cis*-splicing efficiency of the *Azoarcus* ribozyme construct reached 3.8%, the highest in vivo activity of the *Azoarcus* ribozyme found in this study. When the promoter of this construct was changed from its constitutive promoter to the IPTG-inducible *trc1* promoter (construct CisAzo_tRNA_T1), the *cis*-splicing activity dropped to ~1% (the total *CAT* pre-mRNA level was increased 9 -fold with the *trc1* promoter). This raised the possibility that the transcription complex itself might influence the activity of the *cis*-splicing *Azoarcus* ribozyme. In summary, the most significant hint as to why the *Azoarcus* ribozyme was inefficient in *E. coli* cells came from the comparison of different sequence environments for the *cis*-splicing construct, where the *cis*-splicing efficiency in *E. coli* cells varied by ~200 -fold between different sequence contexts (0.018% - 3.8%), and the highest efficiency in the sequence context was closest to its natural environment in pre-tRNA^{Ile}.

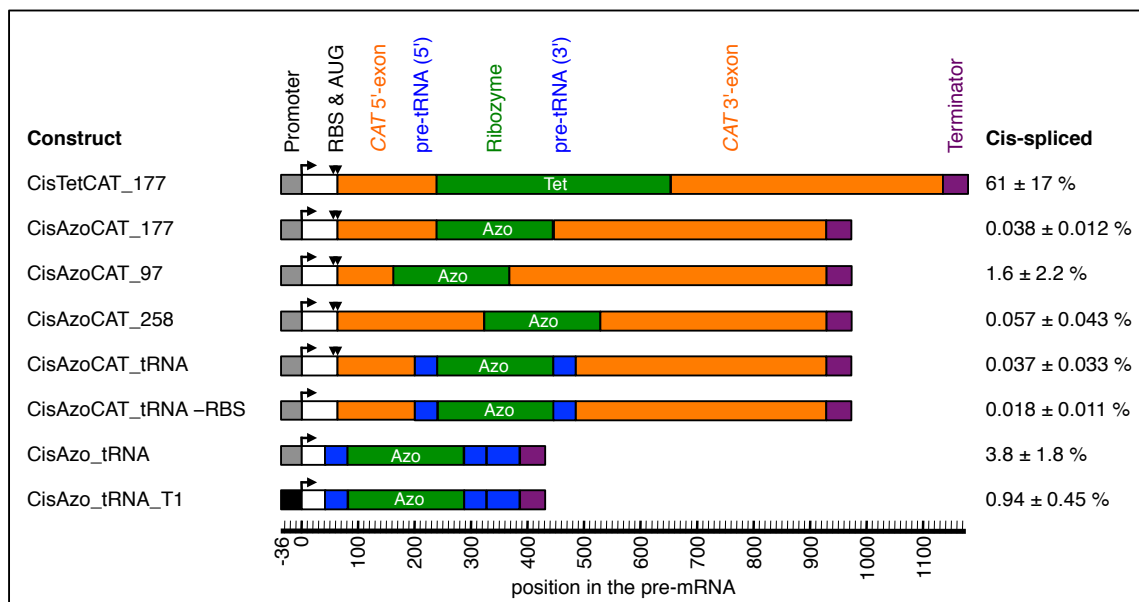


Figure 2.7: Cis-splicing efficiency of ribozyme constructs in *E. coli* cells. The name of each construct is given on the left, a schematic representation of the construct is given in the middle, and the observed *cis*-splicing efficiencies, measured by RT-qPCR, are given on the right. The names of the constructs ("Cis...") include the ribozyme ("Tet" for *Tetrahymena*, and "Azo" for *Azoarcus*), the sequence flanking the ribozyme (CAT, CAT_tRNA, and tRNA), the splice site in the CAT mRNA contexts (177, 97, 258), the mutation inactivation of the ribosomal binding site and the start codon (-RBS) and the use of a different promoter (T1). Note that the three constructs CisAzoCAT_97, CisAzoCAT_177, and CisAzoCAT_258 contained mutations in their 3'-exon immediately downstream of the 3'-splice site to facilitate anticodon stem interactions; the construct CisAzoCAT_97 additionally contained mutations in the P1 loop to generate a P10 helix. For more details on these designs see figure S4. The schematics include the promoter (gray or black boxes with angled arrow), the ribosome-binding site and start codon (RBS & AUG; black triangles), the CAT exons (orange), the ribozymes (green), the flanking tRNA^{lle} sequences (blue), and transcription terminators (purple). A second blue box indicates an extended pre-tRNA^{lle} context. The position of each element in the pre-mRNA is given on a scale below the schematics, relative to the transcription start site. Note that the numbering of splice sites (97, 177, 258) was from the AUG start codon instead. The percentage of *cis*-spliced mRNA is given on the right, with errors as standard errors from three biological replicates.

2.4 Discussion

In this study we explored the *trans*-splicing properties of the *Azoarcus* group I intron ribozyme. In the absence of a 5'-terminal extended guide sequence

(EGS) the ribozyme favored the same splice sites as the *Tetrahymena* ribozyme on a given substrate mRNA in vitro. While none of the 'classical design' constructs worked well for the *Azoarcus* ribozyme on the full-length mRNA substrate, we found that a different design, which mimicked the natural *cis*-splicing context of the *Azoarcus* ribozyme (the anticodon stem of tRNA^{Ile}), led to improved *trans*-splicing efficiencies in vitro, as high as the *Tetrahymena* ribozyme. In contrast, the efficiency in *E. coli* cells of *cis*- and *trans*-splicing *Azoarcus* ribozymes was much lower than those of the *Tetrahymena* ribozyme.

The accessibility of splice sites on our model substrate, *CAT* mRNA, was similar between *trans*-splicing *Azoarcus* and *Tetrahymena* ribozymes (Figure 2.2). This was consistent with a previous, less extensive comparison of splice site accessibility on a different mRNA substrate, between the group I intron ribozymes from *Tetrahymena*, and the myxomycetes *Fuligo* and *Didymium* (16). Here, the three ribozymes also recognized mostly the same splice sites.

When the substrate-ribozyme interactions were optimized for the *Azoarcus* ribozyme constructs, they achieved similar in vitro *trans*-splicing efficiencies as the best *Tetrahymena* ribozyme constructs on the full-length *CAT* mRNA (compare Figures 2.5B and 2.3B, respectively). The results suggested that the *Azoarcus* ribozyme preferred a different design principle than the *Tetrahymena* ribozyme. The *Tetrahymena* ribozyme achieved up to ~20% *trans*-splicing efficiency on full-length *CAT* mRNA with the 'classical design', which includes a short P1 extension, an internal loop that may form a P10 helix, and a

5' duplex with the substrate (Figure 2.3) (13-15, 18, 40). The *Azoarcus* ribozyme also achieved up to ~20% *trans*-splicing efficiency but with the 'anticodon stem design' (Figure 2.5). This design mimicked the native secondary structure environment of the *cis*-splicing *Azoarcus* ribozyme in the anticodon stem loop of tRNA^{Ile}, and outperformed two other tested design principles on the short substrate RNAs (Figure 2.6). A similar finding was made for the *Anabaena* group I intron ribozyme, where the anticodon stem also stabilized the very short P1 helix and possibly also prevented the formation of inhibitory interactions in a *cis*-splicing context (41). The failure of the P1 extension and classical designs to mediate efficient splicing indicated that the *Azoarcus* ribozyme, like the *Anabaena* ribozyme, required specific additional contacts immediately adjacent to the splice site. These contacts, comprising the anticodon stem, were not present in either the P1 extension or classical design principles (compare Figures 2.3A, 2.4A, and 2.5A). These results suggested a more general framework for designing efficient *trans*-splicing ribozymes from group I introns from different species: No single design at the splice site works well for all *trans*-splicing group I intron ribozymes. Instead, the secondary structure designs for group I intron ribozymes from different species should try to mimic the natural secondary structure context of each ribozyme.

The in vitro *trans*-splicing conditions of this study had two important differences from the *trans*-reaction conditions of previous in vitro studies with the *Azoarcus* ribozyme. Previous studies used high Mg²⁺ concentrations (10 - 100 mM) and high temperature (48°C - 60°C, with the optimum around 60°C) (28, 32,

38, 39, 42-44). Under these conditions, the *Azoarcus* ribozyme is able to catalyze an impressive variety of *trans*-reactions. By repetitively splicing a substrate the *Azoarcus* ribozyme can cause an effect like RNA polymerization (32, 42). By *trans*-ligating RNA oligomers the *Azoarcus* ribozyme can create active ribozymes from inactive ribozyme fragments (43). The latter even allows for the covalent self-assembly of ribozymes from four different fragments (30) and the spontaneous emergence of self-replicating sets from pools of ribozyme fragments with randomized substrate recognition sequences (45). In contrast, our study aimed to find designs for *Azoarcus* ribozymes that allowed for efficient *trans*-splicing at near-physiological magnesium concentrations (5 mM) and temperature (37°C). Our results identified such a design, described above. This design can now be used for *trans*-splicing reactions with the *Azoarcus* ribozyme under near-physiological in vitro conditions.

The *trans*-splicing efficiencies of *Azoarcus* ribozyme constructs in *E. coli* cells were much lower than the *trans*-splicing efficiencies of *Tetrahymena* ribozymes in *E. coli* cells of 1-10% (6, 15), and this study. Similarly, no *cis*-splicing *Azoarcus* ribozyme construct mediated chloramphenicol resistance in *E. coli* cells or achieved *cis*-splicing efficiencies above 5%, while a *cis*-splicing *Tetrahymena* ribozyme construct mediated chloramphenicol resistance and achieved ~60% *cis*-splicing efficiency in cells (Figure 2.7). For the *Azoarcus* ribozyme constructs, the 200-fold variation of *cis*-splicing efficiencies between different flanking sequences (0.018% - 3.8%) suggested that the *Azoarcus* ribozyme was quite sensitive to flanking sequences in *E. coli* cells, while the

same constructs were much less sensitive to flanking sequences in vitro (Figure S2.4).

The *Azoarcus* ribozyme's increased sensitivity to flanking sequences in *E. coli* could have been mediated by a number of different factors in the *E. coli* cellular environment, including (i) factors associated with the transcription complex, (ii) RNA chaperones, and (iii) RNA-binding proteins. First, the fourfold effect of different RNA polymerase promoters on *cis*-splicing efficiency (Figure 2.7, compare constructs CisAzo_tRNA and CisAzo_tRNA_T1) hinted that different factors interacting with the transcription complex in *E. coli* might have influenced the co-transcriptional folding such that portions of the *Azoarcus* ribozyme would be more prone to interactions with its flanking sequences. Second, one of the many RNA chaperones in *E. coli* cells (46) such as S12 (47), and StpA (48) could have unfolded portions of the *Azoarcus* ribozyme to make them accessible to inhibitory interactions with the flanking sequences. Third, proteins that bind to the group I intron ribozymes such as the tRNA-binding tyrosyl-tRNA synthetase Cyt-18 from *Neurospora crassa* (49, 50), CBP2 from *S.cerevisiae* (51), or I-Anil from *Aspergillus nidulans* (52) could have stabilized the *Azoarcus* ribozyme in a conformation that would be prone to interact with the flanking sequences. While the studies cited above investigated protein interactions that increase group I intron activity, the same proteins could have inhibited the *Azoarcus* ribozyme, because specific proteins can increase or decrease group I intron splicing efficiency depending on the identity of the group I intron ribozyme (53). It is currently unclear which, if any, of the three discussed

possibilities made the *Azoarcus* ribozyme more susceptible to interference from flanking sequences in *E. coli* cells.

2.5 Materials and Methods

Plasmids and sequences

The sequence for the *Azoarcus* ribozyme (GenBank DQ103524.1) was custom synthesized (Genscript) and used as PCR template for subsequent cloning steps. The sequence encoding the chloramphenicol acetyltransferase (*CAT*) was derived from the plasmid pLysS (Invitrogen), including its constitutive promoter. The assay for the identification of splice site preferences on the *CAT* mRNA used ribozyme 3'-exons that were the first 57 nucleotides of the same α -mannosidase sequence used previously (12, 16). Constructs were cloned into the plasmid pUC19 and sequenced. Exceptions were variants of the 5'-EGS sequence for in vitro splicing reactions and partially randomized pools for the identification of accessible splice sites on *CAT* mRNA to add the ribozyme 5'-terminus NNNGNN (for *Azoarcus*) and GNNNNN (for *Tetrahymena*). In these cases, the ribozyme 5'-terminus was modified during PCR by using 5'-PCR primers that carried the modified sequence, to generate the templates for transcription. The ribozyme 3'-exons for *CAT* mRNA repair at splice sites 97, 177, and 258 were added to the ribozyme sequences by introducing restriction sites near the 3'-splice site (BsrGI for splice site 97 and 258, and EcoRI for splice site 177) and ligating the PCR products of both fragments appended with the

same restriction sites. Introduction of the restriction sites required mutations in the 3'-exon that were silent with exception of splice site 177, where EcoRI was already present in *CAT* sequence. The resulting 3'-exon sequences are shown in supplementary table S2. To generate the *cis*-splicing constructs, the ribozyme 5'-exon sequences were linked to the ribozyme sequences by PCR of the two fragments with two 5'-phosphorylated primers, and blunt end ligation. The promoter driving the expression of *trans*-splicing ribozymes in *E. coli*, *trc1*, was a down-regulated version of the *trc* promoter as described previously (15), while the *cis*-splicing ribozymes used the promoter of *CAT* in the plasmid pLysS. The frameshift inactivation of the *CAT* gene for all in vivo experiments was a deletion of nucleotide 322 downstream of the translation start site, generated by site-directed mutagenesis.

RNAs for in vitro experiments

All RNAs for in vitro *trans*-splicing experiments were prepared essentially as described previously (12, 54), by in vitro run-off transcription with T7 RNA polymerase. Template DNA was generated by PCR and transcribed by T7 RNA polymerase at 37°C for 60 minutes in 40 mM Tris/HCl pH 7.9, 2.5 mM spermidine, 26 mM MgCl₂, 5 mM DTT, 0.01% Triton X-100, and 2 mM of each NTP. For some *trans*-splicing ribozyme constructs the incubation time was reduced to 30 minutes to reduce self-cleavage at the 3'-splice site. Transcription reactions for *cis*-splicing constructs were performed with only 6 mM MgCl₂ and 2 mM Spermidine (55), [α -³²P]UTP at 37°C for 1.5 hours to reduce self-cleavage.

Ribozymes containing truncated (10nt) 3'-exons were transcribed with hammerhead ribozymes appended to their 3'-ends to reduce 3'-end heterogeneity (see legend to Figure 2.6). All transcripts of ribozymes and substrate were purified by denaturing polyacrylamide gel electrophoresis. The concentrations of purified RNAs were calculated from their absorption at 260 nm.

Identification of accessible splice sites

The accessible splice sites on *CAT* mRNA were determined as described previously (12, 56). GTP and substrate RNA were pre-incubated separately from the ribozyme at 37°C for 10 minutes in 50 mM MOPS pH 7.0, 5 mM MgCl₂, 135 mM KCl, and 2 mM spermidine. The two solutions were mixed for a final concentration of 100 nM substrate, 10 nM ribozyme, and 200 μM GTP, and incubated at 37°C for 1 hour. Reaction products were reverse transcribed using AMV reverse transcriptase. After RNA hydrolysis (90°C for 10 minutes in 200 mM sodium hydroxide), the reverse transcription products were PCR amplified using primers that bind to the 5' end of the *CAT* mRNA and the ribozyme 3' exon, nested from the primer of reverse transcription. The PCR products were cloned into pUC19 and sequenced to identify the splice sites. All of the splice sites found for the *Tetrahymena* ribozyme (Figure 2.2A) were reported previously (12) but the current study used only the sub-set of previously reported splice sites that was determined side-by-side with the splice sites for the *Azoarcus* ribozyme in this study. These were 56 splice sites for both ribozymes on the full-length *CAT* mRNA and an additional 12 (*Azoarcus*) and 10 (*Tetrahymena*) splice sites using

a 5'-nested PCR primer, which annealed to positions 183 - 206 in the *CAT* mRNA sequence. The reactions with both *Azoarcus* and *Tetrahymena* present in the same reaction mixture were performed the same, except that each ribozyme had a final concentration of 5 nM.

In vitro splicing reactions

To generate a gel shift between substrates and *trans*-spliced products on the full-length *CAT* mRNA, the 3'-exons of *trans*-splicing ribozymes were truncated, generating a uniform product size for all splice sites (3'-exons were 239 nt for splice site 97, 159 nt for splice site 177, and 78 nt for splice site 258). The sequences for short substrates (Figure S2.3) consisted of a 28 nucleotide excerpt from the *CAT* mRNA at the splice sites 97, 177, or 258, and two 5'-terminal G's for higher in vitro transcription efficiency. The ribozyme 3'-exons for the short substrates contained seven nucleotides from the ribozyme 3'-exon used with the full-length *CAT* mRNA that were mutated for the anticodon stem designs, and a 3'-terminal AUC. Incubation conditions for the *trans*-splicing reactions were the same as for the splice site identification assay, with exception of final RNA concentrations (~60 nM 5'-[³²P]-labeled *CAT* mRNA, 1 μM ribozyme, and 200 μM GTP), and incubation times (24 hours for the full-length substrate and 8 hours for the minimal substrate). The *cis*-splicing reactions were incubated for 24 hours and contained ~30 nM internally [³²P]-radiolabeled RNA and 200 μM GTP. Splice sites of the *trans*- and *cis*- splicing constructs (splice site 97, 177, and 258 with and without anticodon stem design) were confirmed by reverse

transcription of the reaction products using AMV reverse transcriptase, PCR amplification, cloning into pUC19, and sequencing. Each *trans*-splicing construct showed the correct splicing product in one out of one sequencing reactions. Each *cis*-splicing construct, with the exception of splice site 258 without anticodon stem design were also correct in three out of three sequencing reactions. The *cis*-splicing *Azoarcus* construct targeting splice site 258 without the anticodon stem design included the omega G of the *Azoarcus* ribozyme in the splicing product of all three sequencing reactions. Radiolabeled samples from the splicing reactions were separated on 5% denaturing polyacrylamide gels for full-length *CAT* mRNAs and 12.5% denaturing polyacrylamide gels for short substrates. The product bands for splice site 177 in Figures 2.3, 2.4, and 2.5 appear to migrate slightly faster because the product co-migrates with the ribozyme of the same length (control experiments not shown). Bands were quantified on a phosphorImager (PMI, Bio-Rad) using the software Quantity One with the "lanes" method. The percentage of product formed in the in vitro *trans*-splicing reactions was calculated based on the fraction of the total radioactivity in the lane. The percentage of product formed in the in vitro *cis*-splicing reactions was calculated based on the normalized fraction of radioactivity in the lane. This normalization for *cis*-splicing reactions was necessary to account for the number of radiolabeled nucleotides in each internally labeled RNA fragment.

Quantitation of splicing in *E. coli* cells

All *E. coli* experiments used electrocompetent preparations of *E. coli* DH5 α (Invitrogen). Total RNA was extracted from *E. coli* cells logarithmically growing with 1 mM IPTG using the Nucleospin RNA II kit (Machery-Nagel). Total RNA was reverse-transcribed using Superscript III (Invitrogen). Reverse transcription products were directly used as template for quantitative PCR reactions with the Applied Biosystems qPCR master mix on the Fast 7500 RT-PCR machine (Applied Biosystems), essentially as described previously (15). The abundance r of substrate, ribozyme, and product and mRNA was calculated using the equation $r = (0.5)^N$, where N is the experimentally determined threshold cycle of the qPCR. The percentage of spliced substrate, X , was calculated for each construct by using the equation $X = 100 * (P / (S + P))$, where S and P are the abundances of substrate and product, respectively. To differentiate stringently between *CAT* mRNA substrate and spliced product, sixteen mutations were introduced into the ribozyme 3'-exon, starting 79 nucleotides downstream of the 3'-splice site. None of the used primer pairs showed significant amount of cross-amplification.

2.6 Acknowledgements

Jing Yang is thanked for sharing her qPCR machine and for reagents. Karen E. Olson and Simpson Joseph are thanked for helpful discussions. Zhaleh Amini is thanked for designing one of the ribozyme constructs. This work was supported by a grant from the Hellman family foundation to U.F.M.

Chapter 2, in full, is a reprint of the material as it appears in RNA Journal.

Dolan, GF & Müller, UF (2014). *Trans*-splicing with the group I intron from *Azoarcus*. *RNA* 20(2):202-13. The dissertation author is the first author on this paper.

2.7 Supporting Information

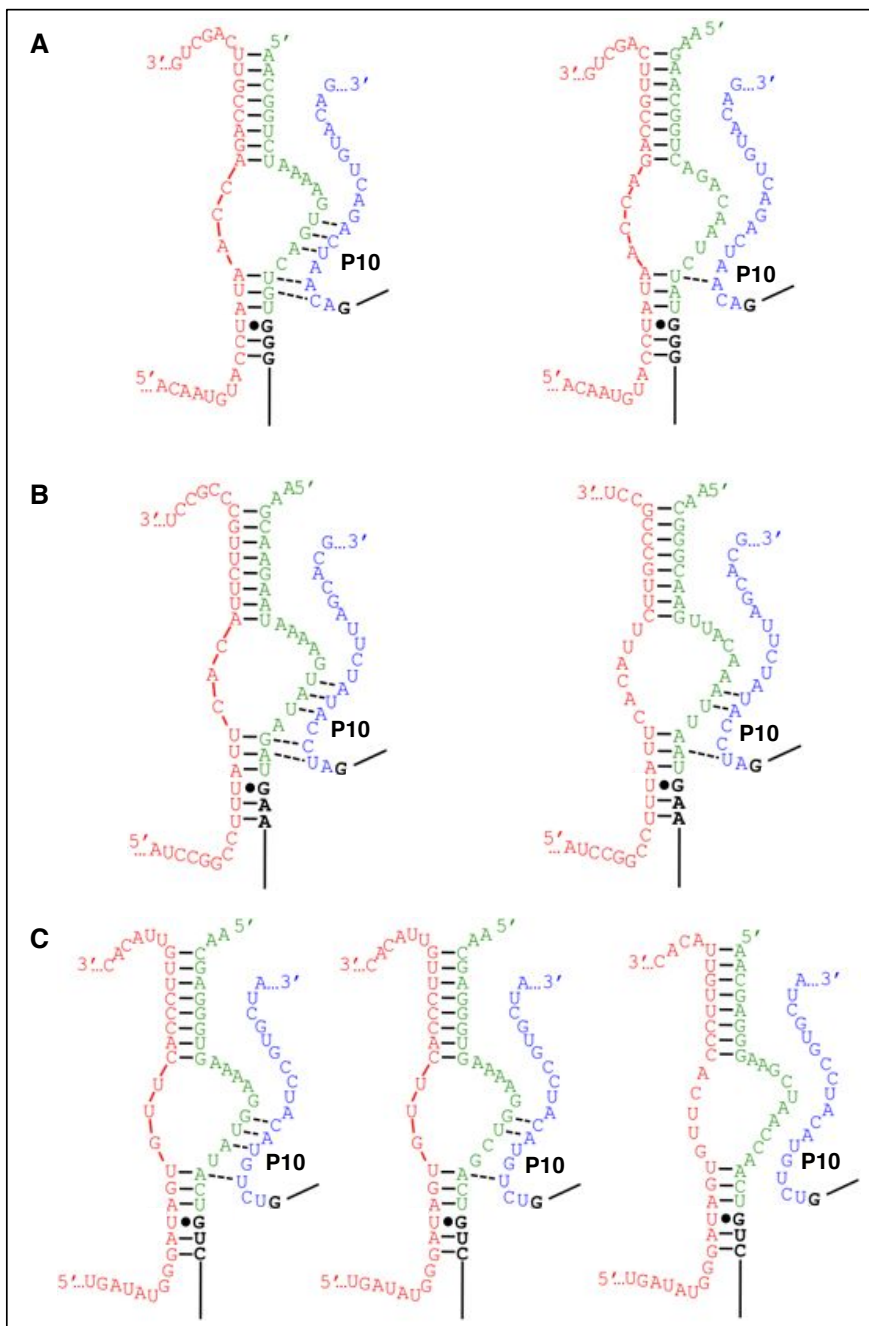


Figure S2.1: Secondary structures of 'classical designs' for ribozyme 5'-terminal interactions at splice site 97 (A), 177 (B), and 258 (C). Three designs were tested for splice site 258 due to similar base pairing between P1 extension and P10 formation. The unextended ribozyme 5'-terminus is shown in black, the ribozyme 3' exon is in blue, the ribozyme 5'-extension is in green, and the substrate is red. The predicted strength of the P10 helix is (A) 5 and 1 base pairs, (B) 5 and 3 base pairs, and (C) 4, 3, and 0 base pairs (always left to right).

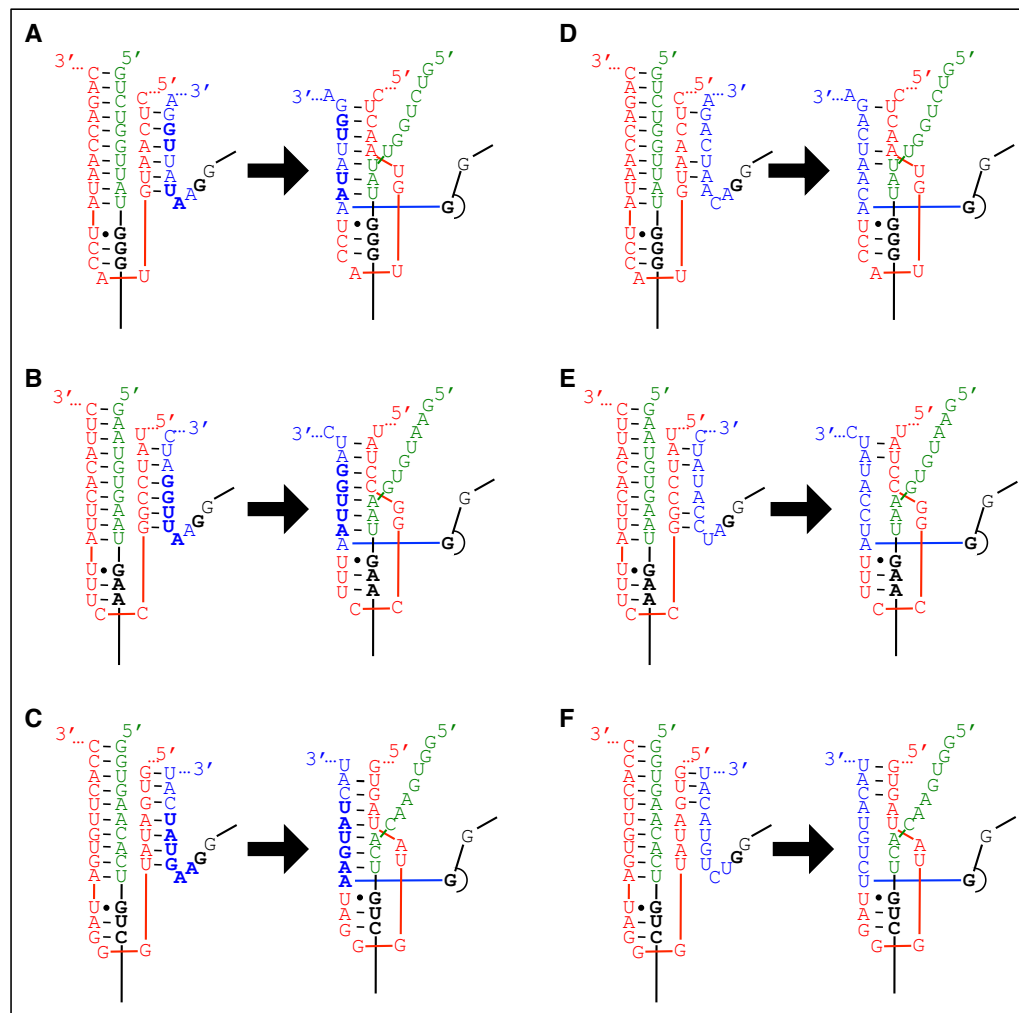


Figure S2.2: Secondary structures of 'anticodon stem designs' for trans-splicing *Azoarcus* ribozymes targeted to splice site 97 (A), 177 (B), and 258 (C). The schematic and rearrangement were adapted from Adams and Strobel, 2004. The bold blue nucleotides in the 3' exon represents the mutated nucleotides to optimize these designed interactions. The internal guide sequence (IGS) of the ribozyme is shown in black, the ribozyme 3' exon is in blue, the ribozyme 5'-extension is in green, and the substrate is red. To the left of each bold arrow, the expected structure before the first catalytic step of splicing is formed. To the right of each bold arrow, the expected structure after the conformational change and before the second step of splicing is shown. As comparison, the sequence without these mutations and their predicted base pairing is shown for splice site 97 (D), 177 (E), and 258 (F). Note that individual constructs may have different P1 extension lengths; the P1 extensions in this figure serve to illustrate the lower end of the P1 extension helix. The trans-splicing reactions shown in figure 2.7 correspond to the secondary structures in this figure S2: sub-figure (A) for splice site 97, sub-figure (B) for splice site 177, and sub-figure (C) for splice site 258.

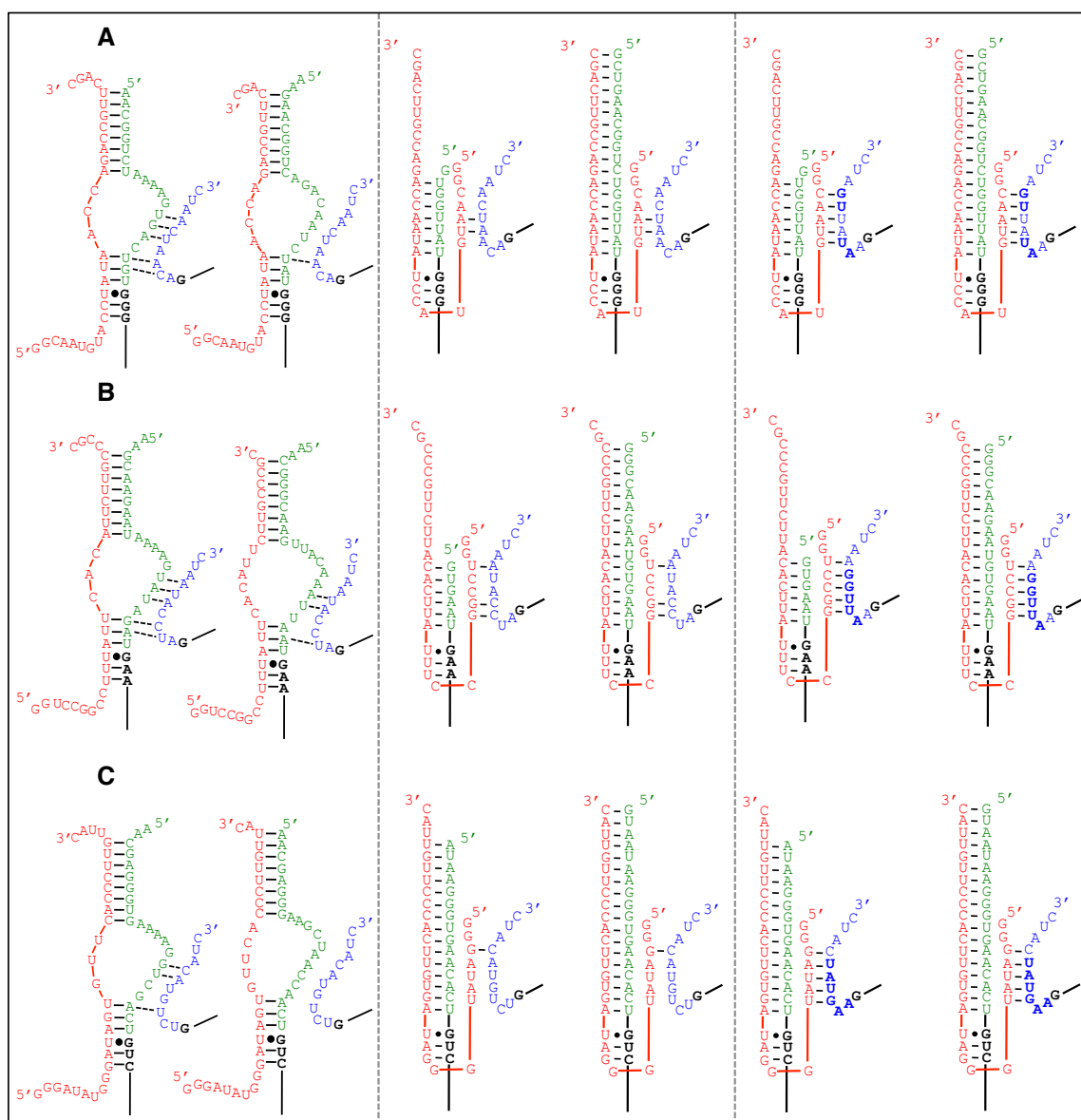


Figure S2.3: Secondary structures of designed short substrate *Azoarcus* trans-splicing interactions at splice site 97 (A), 177 (B), and 258 (C). On the left of each subfigure is the 'Classic design', the middle structure is the 'Pi extension design' and on the right is the 'anti-codon stem design'. The ribozyme 5'-terminus is shown in black, the 3' exon in blue, the ribozyme 5'-extension in green, and the substrate is in red. Bold blue characters in the 3' exons represent the mutated nucleotides to optimize the designed interactions. After the designed interactions the terminal three nucleotides of the 3' exon are the same across all structures and splice sites, to avoid potential biases due to different base pairing of the 3'-exon with the 5' terminus of ribozyme or substrate.

Figure S2.4: Cis-splicing with the *Azoarcus* ribozyme testing the influence of the anticodon stem and the P10 helix interactions predicted by Adams & Strobel, 2004. A, Secondary structure that illustrates the anticodon stem interaction between the 5'-exon (red) and the 3'-exon (blue). B, Secondary structure that illustrates the design of a perfect P10 interaction, after the first catalytic step of splicing between the 3' exon (blue) and the *Azoarcus* ribozyme (black). Note that this structure partially competes with the anticodon stem interaction shown in (A). C, Autoradiogram of a denaturing polyacrylamide gel separation of cis-splicing products with an internally radiolabeled substrate. The text above the autoradiogram indicates whether the *Azoarcus* (Azo) or the *Tetrahymena* ribozyme (Tet) was used, into which splice site of *CAT* mRNA the ribozyme was inserted (97, 177, or 258), and which structural elements were designed into the construct: No designed interactions (–/–), only the anticodon stem interaction (–/+), or the anticodon stem interaction together with the P10 interaction (mutations in the *Azoarcus* P1 ex to make all P10 bp interactions) (+/+). Markers (M) label the position of unspliced *CAT* pre-mRNA substrate (Sub), spliced *CAT* mRNA product (Prod), and 5' cleavage fragments (5'-Frag), which are different for each splice site. The column graph below the gel shows the percentage of substrate that was converted to product. Error bars are standard deviations from three experiments. D, Sequences of tested cis-splicing *Azoarcus* ribozyme constructs. In all cases, the total length of the construct was 883 nt (= 678 nt (*CAT* mRNA) + 205 nt (*Azoarcus* ribozyme)). The color scheme is the same as in (A), the IGS is underlined. Bold letters indicate mutations made to facilitate the designed interactions. Note that the three CisAzoCAT constructs tested in *E. coli* cells (figure 7) correspond to the three most efficient constructs shown above, with Azo 97 (–/–) corresponding to CisAzoCat_97, Azo 177 (–/+) corresponding to CisAzoCAT_177, and Azo 258 (–/+) corresponding to CisAzoCAT_258. The construct Azo 97 (–/–) instead of Azo 97 (+/+) because the latter contained a stop codon, and the efficiency of both constructs was the same within experimental error.

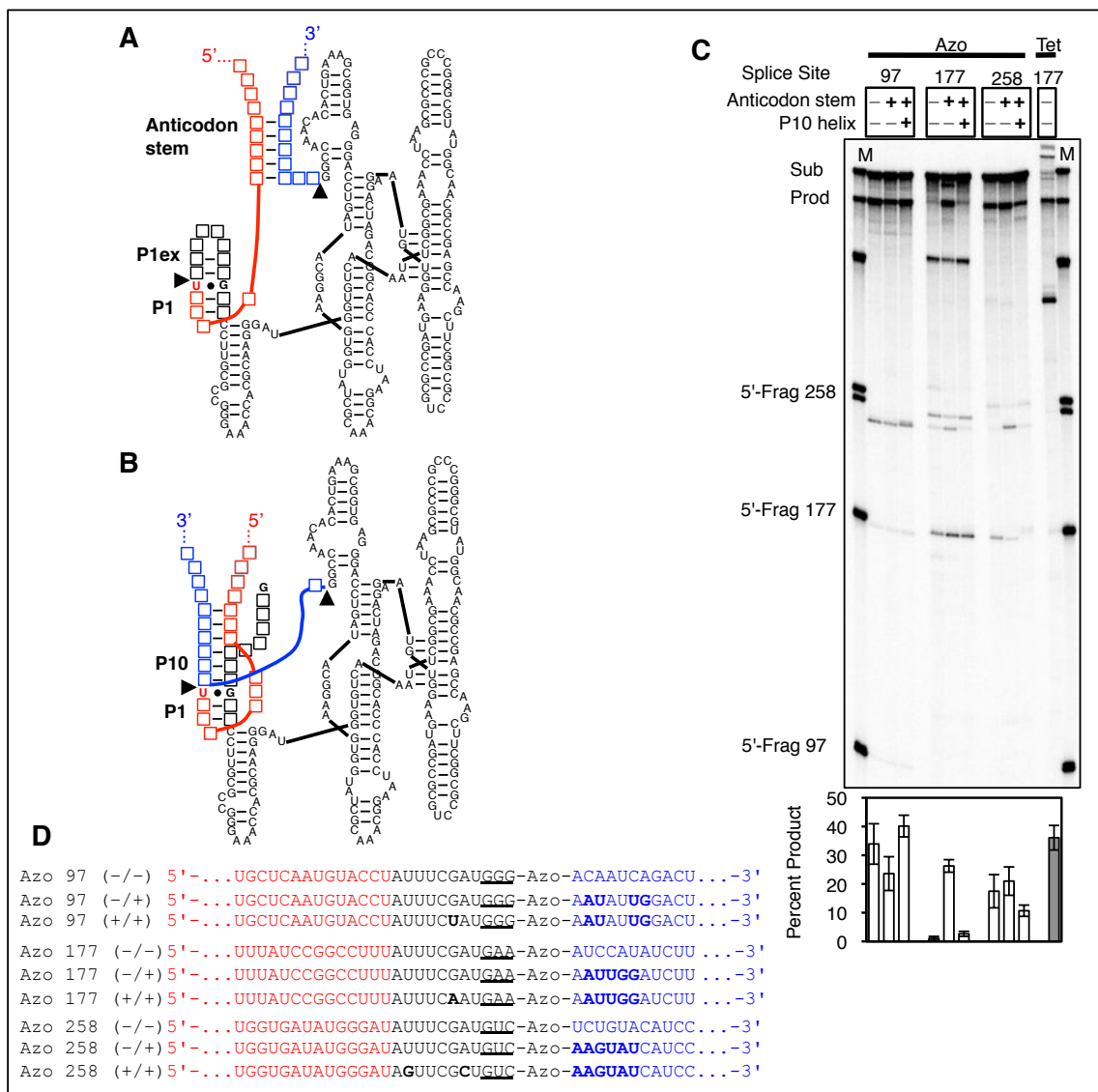


Table 2.1: Positions of all splice sites shown in Figure 2A, including the number of times they were identified. The numbering of splice sites on the *CAT* mRNA is relative to the A of the AUG start codon. Note that several splice sites tolerate a nucleotide different from U at the splice site. (A) Splice sites identified with a 5'-PCR primer annealing at the 5'-terminus of the *CAT* mRNA sequence. (B) Splice sites identified with a 5'-PCR primer annealing at positions 175-206 of the *CAT* mRNA sequence, thereby identifying splice sites only at positions >200.

A	Splice Site	Number of Occurrences	
		Tetrahymena	Azoarcus
	G6	0	1
	A8	0	1
	A9	0	1
	U14	1	0
	U18	2	1
	C29	0	1
	C30	3	0
	U32	3	0
	U33	9	0
	U51	0	1
	U54	7	2
	U63	0	2
	U64	0	1
	U65	1	0
	C71	0	1
	U73	1	1
	U83	2	0
	U87	1	1
	C88	0	1
	U93	0	1
	C95	0	1
	U97	9	5
	U99	3	9
	U110	0	1
	U111	0	1
	U131	1	0
	U133	0	1
	U153	1	1
	C159	0	1
	U164	0	4
	U165	0	1
	U166	2	1
	U175	1	1
	U176	0	1
	U177	1	2
	U179	0	1
	U180	1	0
	C183	0	1
	A184	1	0
	U189	0	1
	C196	1	0
	U197	1	0
	U356	1	0
	U405	0	1
	U448	0	1
	U473	0	1
	U496	0	1
	U497	0	1
	U499	1	0
	C505	1	0
	U573	1	0
	U611	0	1
	U639	0	1

B	Splice Site	Number of Occurrences	
		Tetrahymena	Azoarcus
	U210	0	1
	U217	0	1
	U248	1	0
	U264	0	1
	C268	0	1
	U270	0	2
	U271	1	3
	U273	1	0
	U310	0	1
	U321	1	0
	U350	1	0
	U378	1	0
	U384	2	0
	U399	0	1
	U405	1	0
	U499	0	1
	C646	1	0

Table 2.2: Ribozyme exon sequences at the splice sites 97, 177, and 258, for the three design principles. Residues that were mutated from the wild type *CAT* sequence are shown in lower case. Restriction sites are underlined. The nucleotides in bold show the last five nucleotides of the ribozyme at its 3'-splice site.

Splice site 97

Original cat sequence: ATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTTAAA

P1 and Tet design: **ACCGG**/AcAAAtcaGACtGTaCAGCTGGATATTACGGCCTTTTTTAAA

Anticodon stem design: **ACCGG**/AatAttgGACtGTaCAGCTGGATATTACGGCCTTTTTTAAA

Splice site 177

Original cat sequence: ATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTC

P1 and Tet design: **ACCGG**/ATcCAAtATctTaGCaCGcTaATGAACGCaCacCCGGAATTC

Anticodon stem design: **ACCGG**/AaTtggATctTaGCaCGcTaATGAACGCaCacCCGGAATTC

Splice site 258

Original cat sequence: AGTGTTCACCCCTTGTTACACCGTTTTCCATGAGCAAACCTG

P1 and Tet design: **ACCGG**/tcTGTaCAtCCgTGcTAtACCGTgTTtCATGAGCAAACCTG

Anticodon stem design: **ACCGG**/AagtaTCAtCCgTGcTAtACCGTgTTtCATGAGCAAACCTG

2.8 References

1. Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*, *Cell* 31, 147-157.
2. Vicens, Q., and Cech, T. R. (2006) Atomic level architecture of group I introns revealed, *Trends in biochemical sciences* 31, 41-51.
3. Inoue, T., Sullivan, F. X., and Cech, T. R. (1985) Intermolecular exon ligation of the rRNA precursor of *Tetrahymena*: oligonucleotides can function as 5' exons, *Cell* 43, 431-437.
4. Johnson, A. K., Sinha, J., and Testa, S. M. (2005) Trans insertion-splicing: ribozyme-catalyzed insertion of targeted sequences into RNAs, *Biochemistry* 44, 10702-10710.
5. Bell, M. A., Johnson, A. K., and Testa, S. M. (2002) Ribozyme-catalyzed excision of targeted sequences from within RNAs, *Biochemistry* 41, 15327-15333.
6. Sullenger, B. A., and Cech, T. R. (1994) Ribozyme-mediated repair of defective mRNA by targeted, trans-splicing, *Nature* 371, 619-622.

7. Alexander, R. C., Baum, D. A., and Testa, S. M. (2005) 5' transcript replacement in vitro catalyzed by a group I intron-derived ribozyme, *Biochemistry* 44, 7796-7804.
8. Ayre, B. G., Kohler, U., Goodman, H. M., and Haseloff, J. (1999) Design of highly specific cytotoxins by using trans-splicing ribozymes, *Proceedings of the National Academy of Sciences of the United States of America* 96, 3507-3512.
9. Song, M. S., and Lee, S. W. (2006) Cancer-selective induction of cytotoxicity by tissue-specific expression of targeted trans-splicing ribozyme, *FEBS letters* 580, 5033-5043.
10. Doudna, J. A., Cormack, B. P., and Szostak, J. W. (1989) RNA structure, not sequence, determines the 5' splice-site specificity of a group I intron, *Proc Natl Acad Sci U S A* 86, 7402-7406.
11. Jones, J. T., Lee, S. W., and Sullenger, B. A. (1996) Tagging ribozyme reaction sites to follow trans-splicing in mammalian cells, *Nature medicine* 2, 643-648.
12. Meluzzi, D., Olson, K. E., Dolan, G. F., Arya, G., and Muller, U. F. (2012) Computational prediction of efficient splice sites for trans-splicing ribozymes, *RNA* 18, 590-602.
13. Kohler, U., Ayre, B. G., Goodman, H. M., and Haseloff, J. (1999) Trans-splicing ribozymes for targeted gene delivery, *Journal of molecular biology* 285, 1935-1950.
14. Byun, J., Lan, N., Long, M., and Sullenger, B. A. (2003) Efficient and specific repair of sickle beta-globin RNA by trans-splicing ribozymes, *RNA* 9, 1254-1263.
15. Olson, K. E., and Muller, U. F. (2012) An in vivo selection method to optimize trans-splicing ribozymes, *RNA* 18, 581-589.
16. Fiskaa, T., Lundblad, E. W., Henriksen, J. R., Johansen, S. D., and Einvik, C. (2006) RNA reprogramming of alpha-mannosidase mRNA sequences in vitro by myxomycete group IC1 and IE ribozymes, *The FEBS journal* 273, 2789-2800.
17. Fiskaa, T., and Birgisdottir, A. B. (2010) RNA reprogramming and repair based on trans-splicing group I ribozymes, *New biotechnology* 27, 194-203.

18. Lundblad, E. W., Haugen, P., and Johansen, S. D. (2004) Trans-splicing of a mutated glycosylasparaginase mRNA sequence by a group I ribozyme deficient in hydrolysis, *Eur J Biochem* 271, 4932-4938.
19. Zhou, Y., Lu, C., Wu, Q. J., Wang, Y., Sun, Z. T., Deng, J. C., and Zhang, Y. (2008) GISSD: Group I Intron Sequence and Structure Database, *Nucleic acids research* 36, D31-37.
20. Michel, F., and Westhof, E. (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis, *Journal of molecular biology* 216, 585-610.
21. Lehnert, V., Jaeger, L., Michel, F., and Westhof, E. (1996) New loop-loop tertiary interactions in self-splicing introns of subgroup IC and ID: a complete 3D model of the Tetrahymena thermophila ribozyme, *Chemistry & biology* 3, 993-1009.
22. Kuo, L. Y., and Piccirilli, J. A. (2001) Leaving group stabilization by metal ion coordination and hydrogen bond donation is an evolutionarily conserved feature of group I introns, *Biochimica et biophysica acta* 1522, 158-166.
23. Koduvayur, S. P., and Woodson, S. A. (2004) Intracellular folding of the Tetrahymena group I intron depends on exon sequence and promoter choice, *RNA* 10, 1526-1532.
24. Shi, X., Solomatin, S. V., and Herschlag, D. (2012) A role for a single-stranded junction in RNA binding and specificity by the Tetrahymena group I ribozyme, *Journal of the American Chemical Society* 134, 1910-1913.
25. Reinhold-Hurek, B., and Shub, D. A. (1992) Self-splicing introns in tRNA genes of widely divergent bacteria, *Nature* 357, 173-176.
26. Mitra, S., Laederach, A., Golden, B. L., Altman, R. B., and Brenowitz, M. (2011) RNA molecules with conserved catalytic cores but variable peripheries fold along unique energetically optimized pathways, *RNA* 17, 1589-1603.
27. Russell, R., and Herschlag, D. (2001) Probing the folding landscape of the Tetrahymena ribozyme: commitment to form the native conformation is late in the folding pathway, *J Mol Biol* 308, 839-851.
28. Sinan, S., Yuan, X., and Russell, R. (2011) The Azoarcus group I intron ribozyme misfolds and is accelerated for refolding by ATP-dependent RNA

chaperone proteins, *The Journal of biological chemistry* 286, 37304-37312.

29. Rangan, P., Masquida, B., Westhof, E., and Woodson, S. A. (2004) Architecture and folding mechanism of the Azoarcus Group I Pre-tRNA, *Journal of molecular biology* 339, 41-51.
30. Hayden, E. J., and Lehman, N. (2006) Self-assembly of a group I intron from inactive oligonucleotide fragments, *Chemistry & biology* 13, 909-918.
31. Burton, A. S., 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, 233-241.
32. Chowrira, B. M., Berzal-Herranz, A., and Burke, J. M. (1993) Novel RNA polymerization reaction catalyzed by a group I ribozyme, *The EMBO journal* 12, 3599-3605.
33. Adams, P. L., Stahley, M. R., Kosek, A. B., Wang, J., and Strobel, S. A. (2004) Crystal structure of a self-splicing group I intron with both exons, *Nature* 430, 45-50.
34. Adams, P. L., Stahley, M. R., Gill, M. L., Kosek, A. B., Wang, J., and Strobel, S. A. (2004) Crystal structure of a group I intron splicing intermediate, *RNA* 10, 1867-1887.
35. Lipchock, S. V., and Strobel, S. A. (2008) A relaxed active site after exon ligation by the group I intron, *Proceedings of the National Academy of Sciences of the United States of America* 105, 5699-5704.
36. Herschlag, D. (1991) Implications of ribozyme kinetics for targeting the cleavage of specific RNA molecules in vivo: more isn't always better, *Proceedings of the National Academy of Sciences of the United States of America* 88, 6921-6925.
37. Cech, T. R. (1990) Self-splicing of group I introns, *Annual review of biochemistry* 59, 543-568.
38. Tanner, M., and Cech, T. (1996) Activity and thermostability of the small self-splicing group I intron in the pre-tRNA(Ile) of the purple bacterium Azoarcus, *RNA* 2, 74-83.
39. Kuo, L. Y., Davidson, L. A., and Pico, S. (1999) Characterization of the Azoarcus ribozyme: tight binding to guanosine and substrate by an unusually small group I ribozyme, *Biochimica et biophysica acta* 1489, 281-292.

40. Einvik, C., Fiskaa, T., Lundblad, E. W., and Johansen, S. (2004) Optimization and application of the group I ribozyme trans-splicing reaction, *Methods Mol Biol* 252, 359-371.
41. Zaug, A. J., McEvoy, M. M., and Cech, T. R. (1993) Self-splicing of the group I intron from *Anabaena* pre-tRNA: requirement for base-pairing of the exons in the anticodon stem, *Biochemistry* 32, 7946-7953.
42. Chowrira, B. M., Berzal-Herranz, A., and Burke, J. M. (1995) Novel system for analysis of group I 3' splice site reactions based on functional trans-interaction of the P1/P10 reaction helix with the ribozyme's catalytic core, *Nucleic acids research* 23, 849-855.
43. Hayden, E. J., Riley, C. A., Burton, A. S., and Lehman, N. (2005) RNA-directed construction of structurally complex and active ligase ribozymes through recombination, *RNA* 11, 1678-1687.
44. Chauhan, S., Caliskan, G., Briber, R. M., Perez-Salas, U., Rangan, P., Thirumalai, D., and Woodson, S. A. (2005) RNA tertiary interactions mediate native collapse of a bacterial group I ribozyme, *Journal of molecular biology* 353, 1199-1209.
45. Vaidya, N., Manapat, M. L., Chen, I. A., Xulvi-Brunet, R., Hayden, E. J., and Lehman, N. (2012) Spontaneous network formation among cooperative RNA replicators, *Nature* 491, 72-77.
46. Semrad, K. (2011) Proteins with RNA chaperone activity: a world of diverse proteins with a common task-impediment of RNA misfolding, *Biochemistry research international* 2011, 532908.
47. Coetzee, T., Herschlag, D., and Belfort, M. (1994) *Escherichia coli* proteins, including ribosomal protein S12, facilitate in vitro splicing of phage T4 introns by acting as RNA chaperones, *Genes & development* 8, 1575-1588.
48. Waldsich, C., Grossberger, R., and Schroeder, R. (2002) RNA chaperone StpA loosens interactions of the tertiary structure in the td group I intron in vivo, *Genes & development* 16, 2300-2312.
49. Akins, R. A., and Lambowitz, A. M. (1987) A protein required for splicing group I introns in *Neurospora* mitochondria is mitochondrial tyrosyl-tRNA synthetase or a derivative thereof, *Cell* 50, 331-345.
50. Caprara, M. G., Lehnert, V., Lambowitz, A. M., and Westhof, E. (1996) A tyrosyl-tRNA synthetase recognizes a conserved tRNA-like structural motif in the group I intron catalytic core, *Cell* 87, 1135-1145.

51. Weeks, K. M., and Cech, T. R. (1995) Protein facilitation of group I intron splicing by assembly of the catalytic core and the 5' splice site domain, *Cell* 82, 221-230.
52. Solem, A., Chatterjee, P., and Caprara, M. G. (2002) A novel mechanism for protein-assisted group I intron splicing, *RNA* 8, 412-425.
53. Vicens, Q., Paukstelis, P. J., Westhof, E., Lambowitz, A. M., and Cech, T. R. (2008) Toward predicting self-splicing and protein-facilitated splicing of group I introns, *RNA* 14, 2013-2029.
54. Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates, *Nucleic Acids Res* 15, 8783-8798.
55. Michel, F., Jaeger, L., Westhof, E., Kuras, R., Tihy, F., Xu, M. Q., and Shub, D. A. (1992) Activation of the catalytic core of a group I intron by a remote 3' splice junction, *Genes & development* 6, 1373-1385.
56. Lan, N., Howrey, R. P., Lee, S. W., Smith, C. A., and Sullenger, B. A. (1998) Ribozyme-mediated repair of sickle beta-globin mRNAs in erythrocyte precursors, *Science* 280, 1593-1596.

Chapter 3

Ribozyme catalyzed triphosphorylation at low trimetaphosphate concentrations

3.1 Abstract

In support of the RNA world hypothesis, previous studies identified trimetaphosphate (Tmp) as a plausible energy source for RNA world organisms. In one of these studies, catalytic RNAs (ribozymes) that catalyze the triphosphorylation of RNA 5'-hydroxyl groups using Tmp were obtained by *in vitro* selection. The reaction kinetics of these ribozymes, however, appeared too slow to be useful for an RNA world organism. Here, the previously characterized triphosphorylation ribozyme, TPR1, is improved by a doped selection. The resulting ribozyme, TPR1e, contains seven mutations relative to TPR1, displays a previously unidentified duplex that constrains the ribozyme's structure, and reacts at a 24-fold faster rate than the parent ribozyme. Under optimal conditions (150 mM Tmp, 650 mM MgCl₂, 40°C), the triphosphorylation rate of TPR1e reaches 6.8 min⁻¹. In synthetic seawater at a prebiotically plausible Tmp concentration of 1 mM and at 40°C the rate is 0.24 h⁻¹. These results are discussed with respect to the origin of life.

3.2 Introduction

The RNA world hypothesis states that an early stage in the evolution of life used RNA as the genome and as the only genome-encoded catalyst (1-3). This hypothesis has gained support from several directions, including the findings that RNA molecules can indeed catalyze reactions (4), that the ribosome is a catalytic RNA (5, 6), and that deoxyribonucleotides are synthesized in cells from ribonucleotide precursors (7). To find out how an RNA world organism could have functioned several laboratories are trying to generate an RNA world organism in the laboratory. An important part of producing an RNA world organism is the generation of catalytic RNAs (ribozymes) that could support a self-replicating and evolving ribozyme system. Since the invention of *in vitro* selection (8, 9) and the demonstration that novel catalytic RNAs can be isolated from randomized RNA pools (10), several types of ribozymes with possible involvement in an RNA world organism have been isolated (11), most importantly a ribozyme that catalyzes RNA-dependent RNA polymerization (12). The substrates for this ribozyme are nucleoside 5'-triphosphates, perhaps the most prebiotically plausible, chemically activated nucleotide (13). Nucleoside 5'-triphosphates can be generated from nucleosides and trimetaphosphate (Tmp) (14). Tmp likely existed prebiotically, as suggested from the finding of large amounts of its chemical precursor in 3.5 billion year old marine sediments (15, 16). However, the uncatalyzed triphosphorylation of nucleosides with Tmp occurs efficiently only at pH values above 12 (14), which would quickly hydrolyze the RNA polymers of an RNA world organism. We previously showed that the

triphosphorylation of RNA 5'-hydroxyl groups is possible at neutral pH, with ribozymes that were obtained by *in vitro* selection (17).

The only triphosphorylation ribozyme that was previously characterized in some detail, TPR1 (Figure 3.1A), displays single-exponential reaction kinetics with a rate of 0.013 min^{-1} at 50 mM trimetaphosphate and 100 mM MgCl_2 at pH 8.3 (17). Because this appears too slow to be useful in an RNA world organism, we searched here for sequence variants of TPR1 that display faster reaction kinetics and work at lower concentrations of trimetaphosphate using *in vitro* selection. The most active ribozyme identified in this study, TPR1e, contains seven activity-enhancing mutations (Figure 3.1B). Under optimal conditions, the triphosphorylation rate of TPR1e reaches 6.8 min^{-1} while the rate in synthetic seawater with 1 mM Tmp reaches 0.004 min^{-1} . The implications for an RNA world scenario are discussed.

3.3 Results

A 'doped selection' was performed to identify more active sequence variants of the previously isolated triphosphorylation ribozyme TPR1 (Figure 3.1A). A partially randomized library was generated that contained the sequence of TPR1 at each position with a frequency of 76% while containing each of the other three nucleotides at that position with a frequency of 8%. With a total of 82 nucleotides being partially randomized, this made sure that around 10^3 wild-type sequences would be contained in a library of 10^{13} sequences ($0.76^{82} * 10^{13} \sim$

1,700). The selection was performed as described previously (17) and started with an effective complexity of $7.0 \cdot 10^{13}$ sequences. After two rounds of selection the number of PCR cycles required to amplify the reverse transcription product dropped from 15 to 6. Such a drop was previously found to be a good indicator that active clones dominated the library (17). Additional rounds of selection were performed under higher selection pressure during the incubation with Tmp (shorter incubation time, lower Tmp concentration, or higher incubation temperature; Figure S3.1) in an effort to select for the most active ribozymes. After three or four rounds of selection, a total of eighty-three clones were isolated and analyzed for triphosphorylation activity (see materials and methods for details on how many clones were tested from each condition/round).

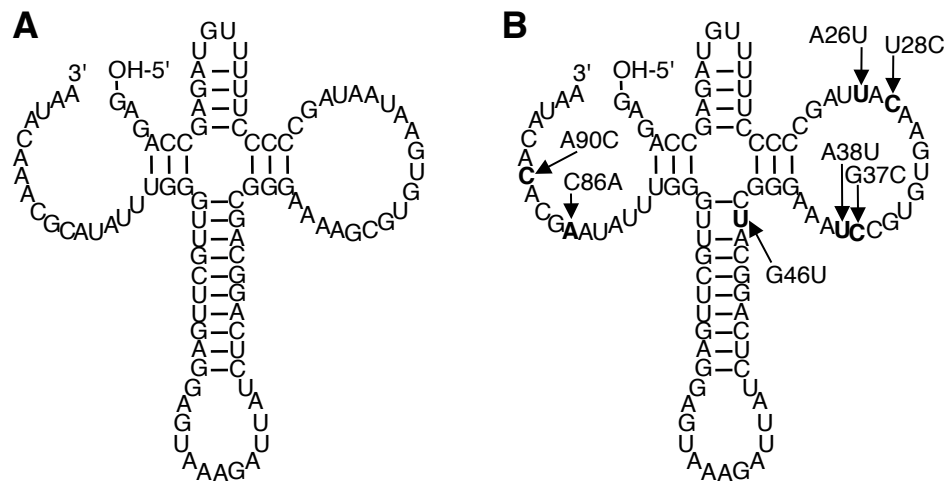


Figure 3.1: Proposed secondary structures for the two central triphosphorylation ribozyme variants in this study. (A) TPR1, the parent ribozyme that served as starting point for the present study. **(B)** TPR1e, the most active ribozyme developed in this study that contains seven beneficial mutations relative to TPR1.

Of the eighty-three sequences assayed for triphosphorylation kinetics, fifty-eight displayed triphosphorylation kinetics at least as fast as the parent

ribozyme TPR1. The six fastest ribozyme clones contained the same set of two mutations (G37C, A38U), and five of them contained the mutation C86A. The fastest ribozyme was clone 11, with 16 mutations relative to TPR1 and a rate of $0.21 \pm 0.02 \text{ min}^{-1}$ under selection conditions (50 mM Tmp, 100 mM MgCl_2 , 50 mM Tris/HCl pH 8.3) (Figure 3.2). This ribozyme was isolated from the selection line with decreased Tmp concentration. To identify the mutations in clone 11 that were necessary for improved triphosphorylation kinetics, all 16 mutations were individually reverted to the parent sequence (Figure S3.2). Eleven mutations showed decreased activity. Six of these eleven mutations were within the long central duplex of the ribozyme (P5), and four of the mutations appeared to extend the P5 duplex. Clone 11 contained destabilizing mutations in this duplex, and we hypothesized that these six mutations were only necessary in the context of additional mutations in clone 11's destabilizing P5 duplex. To test this hypothesis we separately analyzed three of the four stem-extending mutations in the context of a fully complementary P5 duplex (i.e. without the destabilizing mutations) for effects on activity. None of the mutations had an effect. With this knowledge the number of mutations necessary for full activity in ribozyme clone C11 was reduced to five mutations (U28C, G37C, A38U, C86A, A90C), leading to ribozyme TPR1_II. This ribozyme showed a rate of $0.25 \pm 0.03 \text{ min}^{-1}$ (Figure 3.2 and Figure S3.2).

In addition to the mutations in clone 11, the analysis of all 83 sequences suggested that an additional six mutations were statistically enriched in the 59 clones that were at least as active as the parent TPR1 (Figure S3.3). Each of

these mutations was individually inserted into TPR1_II to test whether TPR1_II could be improved further. Two of the six additional mutations (A26U and G46U) showed a statistically significant increase in the rate. When both mutations were combined and inserted into TPR1_II the rate was $0.309 \pm 0.017 \text{ min}^{-1}$, 24-fold faster than TPR1 (Figure 3.2). This variant was declared the 'winner' of the doped selection study. It contained seven mutations (A26U, U28C, G37C, A38U, G46U, C86A, A90C; Figure 3.1B) and was termed TPR1e (e for 'evolved').

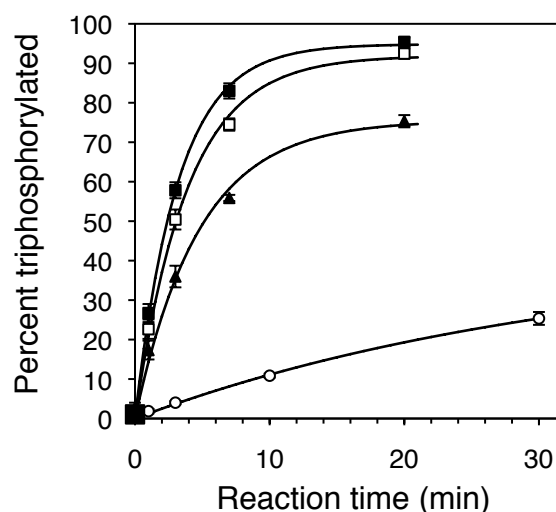


Figure 3.2: Triphosphorylation kinetics of central ribozyme variants in this study.

The starting point of the doped selection was the ribozyme TPR1 (empty circles). It has a k_{OBS} of 0.013 min^{-1} under selection conditions (100 mM MgCl_2 , 50 mM trimetaphosphate, 50 mM Tris/HCl pH 8.3). The most efficient isolate from the doped selection was a 16 mutation variant called clone 11 (filled triangles, k_{OBS} of 0.21 min^{-1}). After removal of unnecessary mutations a 5 mutation variant called TPR1-II resulted (open squares, k_{OBS} of 0.26 min^{-1}). Two separately selected mutations were inserted to yield TPR1e (filled squares), a 7 mutation variant with a k_{OBS} of 0.31 min^{-1} . Error bars denote the standard deviations from triplicate experiments.

The mutations in TPR1e suggested the existence of a previously unrecognized duplex P4 (Figure 3.3A). This duplex would be expected to form eight base pairs between the RNA segments G31 to A39 and U84 to C92. To

test whether this duplex formed, base covariation experiments were conducted at two positions in the expected duplex, testing for base pairs between G33 and C90 and between G35 and U84. No base covariation was observed with the mutations of G33C and C90G (Figure 3.3B), suggesting that these bases were not involved in the duplex or that their contribution to duplex stability was not necessary for full activity of the ribozyme. In contrast, the drop in activity by mutations G35C and C88G was rescued by the double-mutation (Figure 3.3C), suggesting that at least a portion of the P4 duplex formed in the active ribozyme. Interestingly, two of the most important activity-enhancing mutations (G37C and C86A) led to a C:A pair in the proposed P4 duplex, which may have an important role in catalysis. The newly identified P4 duplex brings the number of duplexes in the 96-nucleotide long ribozyme to five (Figure 3.3A), generating strong structural constraints and suggesting a very compact three-dimensional structure for the ribozyme.

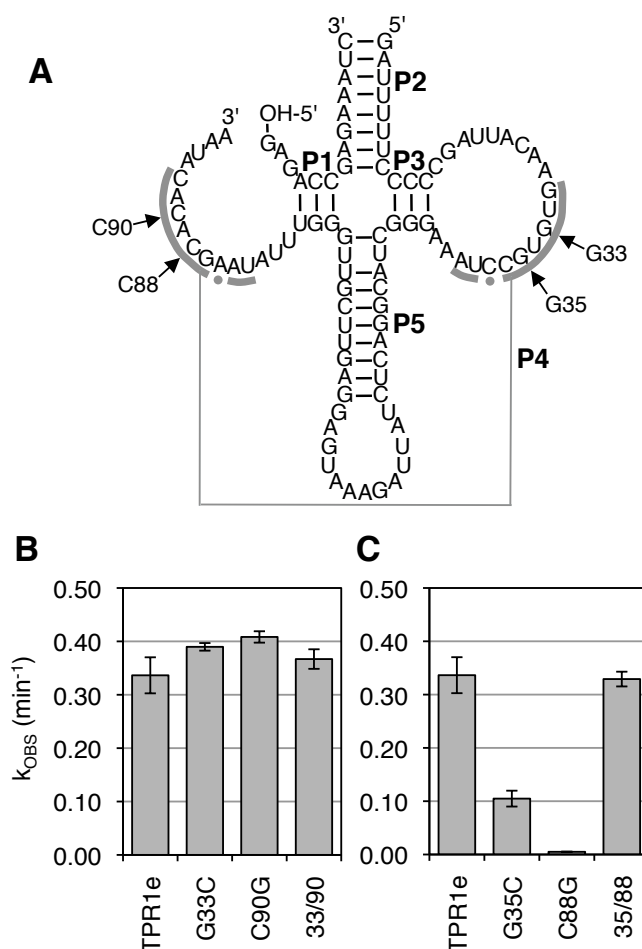


Figure 3.3: Base covariation experiments to test the formation of duplex P4. (A) Proposed secondary structure for ribozyme TPR1e with proposed paired regions labeled P1-P5. The ribozyme fragments participating in the tested duplex P4 are indicated with a thick grey line, and contain a C-A pair (grey dot) that gave rise to a large increase in activity. Nucleotides that were varied in the covariation experiments are labeled and indicated by arrows. Note that the duplexes P1 and P2 are formed *in trans* because covariation experiments were performed using the *trans* reaction. **(B)** and **(C)** Observed rates of the mutated ribozyme in the base covariation experiment of G33 and C90 (B) and G35 and C88 (C). Error bars are standard deviations of triplicate experiments.

The parent ribozyme TPR1 did not show detectable triphosphorylation of free nucleosides (17). To test whether TPR1e was able to triphosphorylate free nucleosides, ^{14}C labeled guanosine was incubated with eight different variants of TPR1e. Because the architecture of the active site and potential steric hindrance

from the 5'-terminus of the ribozyme is unknown, the eight ribozyme variants tested were truncated at different points of the 5'-terminus (full length, -1, -2, -3, -6, -18) or differed in the 5'-terminus functional group of the -1 construct (5'-hydroxyl, 5'-phosphate, or 5'-triphosphate) to allow for binding of the free guanosine. Unfortunately, no triphosphorylation of guanosine could be detected for any of the constructs (data not shown). We assume that the triphosphorylation ribozyme does not establish a sufficient number of contacts to the 5'-terminal guanosine to position it tightly at the catalytic site. This may not, however, prevent TPR1e from being a useful component of an RNA world organism because the triphosphorylation of RNA 5'-hydroxyl groups can result in RNA polymerization in 3'- to 5'-direction if RNA 5'-hydroxylation alternates with nucleoside addition (11, 17).

The experiments described above used a reaction buffer containing 50 mM Tmp. To test whether TPR1e showed RNA triphosphorylation activity under prebiotically more plausible conditions the triphosphorylation reaction was performed in synthetic seawater with 1 mM Tmp (Figure 3.4A). The concentration of 1 mM Tmp was considered to be in the upper range of prebiotically plausible concentrations (16). The synthetic seawater contained 470 mM Na^+ , 550 mM Cl^- , 28 mM SO_4^{2-} , 54 mM Mg^{2+} , 10.5 mM Ca^{2+} , 10.1 mM K^+ , 2 mM HCO_3^- , and 0.3 mM CO_3^{2-} (18, 19). The triphosphorylation rates under these conditions were 0.085 hour^{-1} at 22°C and 0.24 hour^{-1} at 40°C . The rate in synthetic seawater, which contains 54 mM MgCl_2 , can be compared to the rate with 54 mM MgCl_2 in the absence of the other components of synthetic seawater. At Tmp

concentrations between 0.1 mM to 3 mM, the rate was consistently about 15-fold lower in synthetic seawater (Figure 3.4B). The contribution of sodium chloride in this inhibition (470 mM in synthetic seawater) is about 2-fold (dashed line in Figure 3.4C). This shows (i) that TPR1e is inhibited by several components of synthetic seawater, and (ii) that the evolved ribozyme TPR1e is able to triphosphorylate with a rate of 0.24 hour^{-1} under prebiotically plausible conditions.

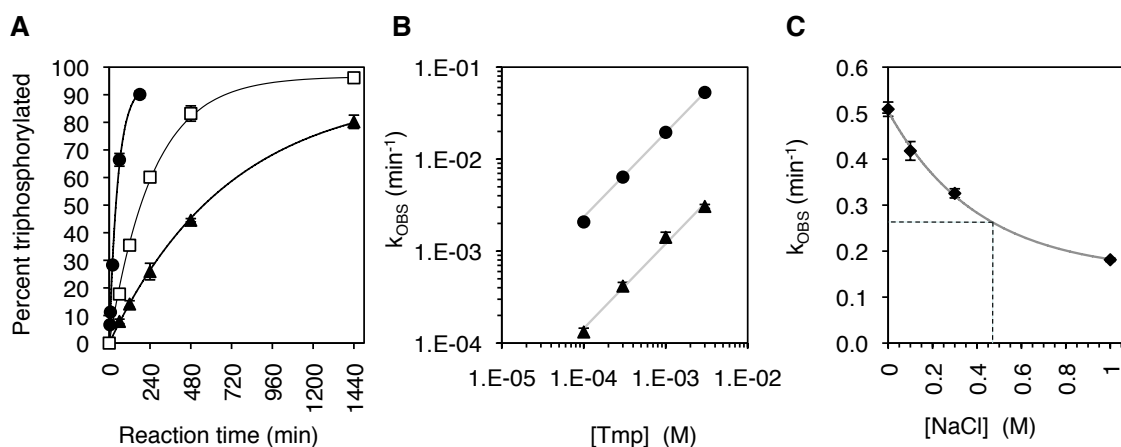


Figure 3.4: Triphosphorylation kinetics of TPR1e under different conditions. (A) At a Tmp concentration of 1 mM, the triphosphorylation kinetics are shown for synthetic seawater at 22°C (black triangles, $k_{\text{OBS}} = 0.0014 \text{ min}^{-1}$, max = 93%), and at 40°C (empty squares, $k_{\text{OBS}} = 0.0039 \text{ min}^{-1}$, max = 97%). For comparison, the reaction kinetics are shown for 54 mM MgCl₂ in Tris/HCl pH 8.3 at 22°C (black circles, $k_{\text{OBS}} = 0.020 \text{ min}^{-1}$, max = 93%). This latter condition lacks all other seawater components besides Mg. Error bars are standard deviations from triplicate experiments, and are smaller than the symbols if not visible. Curves are single-exponential fits to the data. (B) Titration of the Tmp concentration in the reaction at 22°C, in synthetic seawater (black triangles) and in 54 mM MgCl₂ with 50 mM Tris/HCl pH 8.3 (empty circles). The offset between the linear fits (grey lines) is 15-fold, on average. (C) Titration of sodium chloride into a triphosphorylation ribozyme reaction containing 50 mM Tmp and 140 mM MgCl₂. The grey line is a single-exponential fit to the data (with offset) and identifies a 1.9-fold reduction in k_{OBS} at 470 mM [NaCl], the same NaCl concentration as in synthetic seawater (dashed line). Error bars are standard deviations from triplicate experiments, and are smaller than the symbols if not visible.

To identify the rate of the evolved ribozyme, TPR1e, under optimal conditions, the influence of temperature, Tmp concentration, and magnesium

concentration were analyzed sequentially. First, the temperature was varied between 5°C and 60°C, and an optimum at 40°C was determined (Figure 3.5A). At 40°C, the optimal Tmp concentration was at 150 mM (Figure 3.5B), and the optimal concentration of free Mg^{2+} at 150mM Tmp was at 500 mM (Figure 3.5C). Interestingly, the dependence on the Tmp concentration at these high Mg conditions and high temperature showed a cooperative effect, with a significantly higher rate at 100 mM Tmp (5.7 min^{-1}) than what would results from doubling the rate at 50 mM Tmp ($2 \cdot 0.96 \text{ min}^{-1} = 1.92 \text{ min}^{-1}$). This shows that a second molecule of trimetaphosphate binds to the ribozyme and/or the first trimetaphosphate molecule and increases the reaction rate. The mechanism of the cooperative effect is unclear. Notably, the effect did not appear at more modest Mg^{2+} concentrations and 22°C, where a linear correlation between Tmp concentration and reaction rate was observed (Figure 3.4B). At the optimal conditions (40°C, 150 mM Tmp, 650 mM total MgCl_2 , 50 mM Tris/HCl pH 8.3) the cooperative effect allowed to reach a reaction rate of 6.8 min^{-1} .

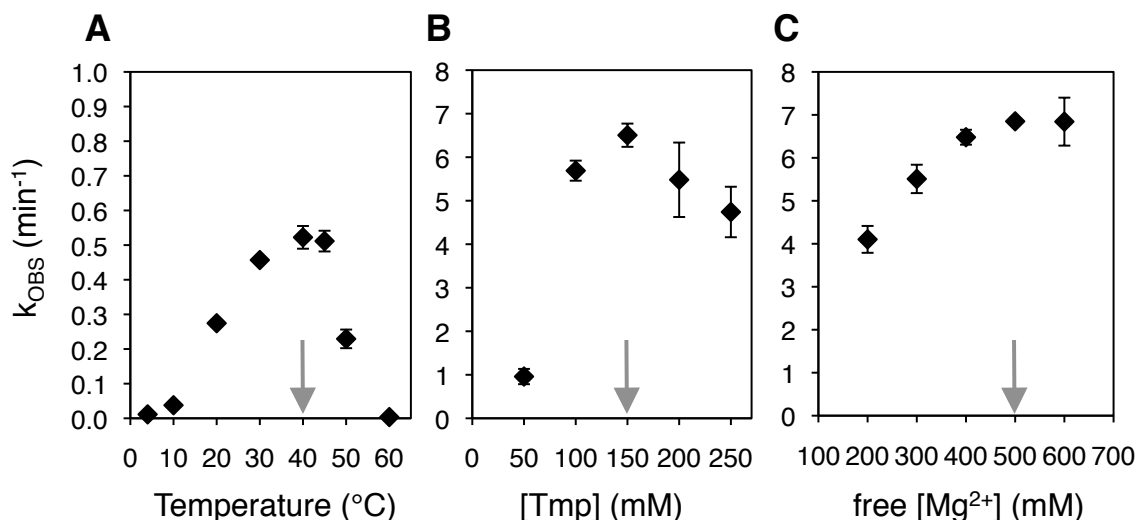


Figure 3.5: Determination of optimal TPR1e triphosphorylation conditions. (A) Observed triphosphorylation rate as function of the temperature, at 50 mM Tmp, 100 mM MgCl_2 , and 50 mM Tris/HCl pH 8.3 (B) Influence of the trimetaphosphate concentration on the observed reaction kinetics at 40 $^{\circ}\text{C}$, and with an excess of 400 mM MgCl_2 over Tmp. (C) Influence of the free Mg^{2+} concentration on the triphosphorylation rate at 40 $^{\circ}\text{C}$ and 150 mM Tmp. The free Mg^{2+} concentration is the total Mg^{2+} concentration minus the concentration of Tmp. The grey arrows indicate the optimum condition for each series of experiments. Note that the scale in (A) is different from the scale in (B) and (C). Error bars are standard deviations of triplicate experiments.

3.4 Discussion

In the present study, a ribozyme catalyzing the triphosphorylation of RNA 5'-hydroxyl groups with Tmp was subjected to a doped selection, which resulted in the improved ribozyme TPR1e. TPR1e shows triphosphorylation kinetics of 6.8 min^{-1} under optimal conditions. In synthetic seawater with 1 mM Tmp at 40 $^{\circ}\text{C}$, a triphosphorylation rate of 0.24 hour^{-1} was measured.

The usefulness of the reaction rate of 0.24 hour^{-1} in synthetic seawater at 1 mM and 40 $^{\circ}\text{C}$ depends on the specific role of the triphosphorylation ribozyme in the RNA world organism. Because TPR1e did not show detectable

triphosphorylation of free nucleosides, the role of the ribozyme could be to participate in polymerization by alternating triphosphorylation of the RNA 5'-hydroxyl group and nucleotide addition (17). In this case, the rate of only a few triphosphorylation events during the short Hadean / Archaean day of less than 15 hours (20) would require dozens of triphosphorylation ribozymes for the polymerization of one ribozyme product per day. Because this rate is well above the uncatalyzed triphosphorylation rate, this ribozyme could be useful to an RNA world organism, but there is an alternative that appears more promising: an RNA world organism existing in an area of increased Tmp. The concentration of 1 mM Tmp was used in these experiments because it seems to be in a reasonable range based on the prebiotic steady-state estimation of polyphosphate concentration of 10 nM - 10 μ M in the mixed zone of the ocean, with localized concentrations across the surface of the Earth in excess of 1 mM (16). The concentration of Tmp in evaporation pans on early Earth, however, may have been significantly higher than 1 mM. If an RNA world organism existed in an evaporation pan containing enriched levels of Tmp, the resulting increase in rate would correspond with an increase in the usefulness of TPR1e.

Also limiting the usefulness of TPR1e is its inhibition by the components of synthetic seawater (Figure 3.4). While one could speculate that the composition of prebiotic seawater may have been more favorable to TPR1e (19), it is obvious that RNA world organisms would have selected for triphosphorylation ribozymes that are not inhibited by the local environment. Within 10^{14} random RNA N₁₅₀ sequences, there exist at least dozens (and probably >100) of different

triphosphorylation ribozymes (17). This means that although TPR1e is inhibited by the components of synthetic seawater, there are likely several triphosphorylation ribozymes that are not inhibited and are capable of fast triphosphorylation within a prebiotically plausible setting.

A higher doping ratio would not have improved the probability of identifying TPR1's local fitness peak. The doping ratio used in this work (76% for the wild type sequence) together with the complexity of the pool ($7 \cdot 10^{13}$) showed that sequence space of single, double, and triple mutants of the parent ribozyme TPR1 was sampled completely, quadruple mutants were sampled to ~77% ($P=1-(1-0.76^{78} \cdot 0.08^4)^{(7 \cdot 10^{13})} = 0.77$), pentuple mutants to ~14%, and sextuple mutants to ~1.6%. Five beneficial mutations were identified in the fittest clone of the selection (Figure S3.2), and two additional beneficial mutations were identified from their enrichment in other isolated clones (Figure S3.3). The number of mutations in the fittest clone and the calculations of mutational sequence space coverage suggest that a higher doping rate might help identify ribozymes with more beneficial mutations. An increased rate of doping, however, may actually reduce the rate of active clones in the selection pool. Two observations support this idea. First, the most active ribozyme from the selection contained several mutations that destabilized stem P5, and required additional mutations elongating and re-stabilizing the stem to overcome the destabilizing effects. This suggests that a higher mutational load in a pool (i.e. a higher doping ratio) leads to more inhibitory mutations and overall, less active ribozymes. Second, the number of mutations per ribozyme was an average of 19.8 in the

starting pool (based on 10 sequences) while the selected ribozymes displayed an average of 12.2 mutations per ribozyme (based on 82 sequences; Figure S3.4). This indicates that ribozymes with fewer total mutations were preferentially selected for and more likely to maintain activity. Unfortunately, the comparison between pre- and post-selection distributions can only be done *after* a doped selection experiment. The observations in this study and in previous work (21), however, can guide the pool design for future doped selection experiments.

To further scale the TPR1 fitness landscape, selection techniques beyond a doped selection are required. A previous study found that newly selected kinase ribozymes, which were improved in their rate by about 40-fold in a doped selection (22), could be improved a further 30-fold by selection from a pool with unbiased randomization of specific, critical nucleotides in the ribozyme (23). This additional, 30-fold enhancement can be explained by the cooperative nature of nucleotides in the three-dimensional nucleotide interaction network of ribozymes (23): In a doped pool, a given position shows a majority of the wild type sequence (76% in our case) and a small fraction (8% in our case) of any mutated nucleotide. Correspondingly, when specific double mutations are necessary to identify improved sets of cooperative mutations the frequency decreases to 0.64% for double mutations, 0.05% for triple mutations, and 0.004% for quadruple mutations. Because the networks of interacting nucleotides can easily exceed seven (23), a doped pool does not appear appropriate to identify the tip of the fitness peak. The pool of choice for this purpose would contain an unbiased randomization of specific, critical nucleotides. For the case of TPR1

these critical nucleotides could be chosen from those that are mutated in the most efficient clones in this study, and additionally several nucleotides that show statistically significant enrichment (Figure S3.3). Beneficial mutations would be identified much easier after a selection from such a 'shuffled' pool compared to a doped pool because all variants occur in similar frequency in the starting pool.

3.5 Materials and Methods

Design of the doped pool

The doped pool was generated from a 113 nucleotide long DNA oligonucleotide (Sigma-Aldrich) based on the original TPR1 sequence (17). The composition of bases at each position of TPR1 was the original sequence to 76% and the other three nucleotides to 8% each. The first 14 nucleotides of TPR1 were kept constant and an additional 12 nucleotides were added to the 3' end to be used as 5' and 3' primer binding sites during the selection. Hand mixed phosphoramidites were used by Sigma at each partially randomized position to ensure the desired 76% to 8% ratio (confirmed by sequencing of 10 clones; data not shown). The hammerhead ribozyme and T7 promoter necessary for performing the selection procedure had the same sequence as previously (17), and were added as 5'-primer during PCR.

Selection

The initial pool for the selection contained an effective population size of $7 \cdot 10^{13}$ unique sequences. The protocol of the selection was identical to our previous selection (17). During the first round of the selection the incubation with Tmp lasted 3 hours, in the presence of 50 mM Tmp, 100 mM MgCl_2 , and 50 mM Tris/HCl pH 8.3, at a temperature of 22°C. Subsequent rounds of the selection were split into three lines with different selection pressures (Figure S3.1). In line one the incubation time with Tmp was decreased from 3 hours, to 2 minutes, then 20 seconds. In line two the concentration of Tmp was decreased from 50 mM, to 5 mM, then 1 mM. In line three the incubation temperature was increased from 22°C, to 42°C, then 50°C. A total of eighty-three ribozymes were then tested for activity, with 53 clones from line 1 round 3, ten clones from line 1 round 4, ten clones from line 2 round 3, and ten clones from round 3 round 3. The six fastest ribozyme clones were identified from all lines but line 3 (increased temperature). The fastest clone, clone 11, came from line 2 round 3 (lowered Tmp selection conditions).

Kinetic analysis of *cis*-triphosphorylation reactions

Initial isolates of the selection were cloned into pUC19, downstream of a hammerhead ribozyme and T7 promoter. After PCR amplification of this ribozyme cassette, ribozymes were obtained by run-off transcription from PCR products using T7 RNA polymerase. Ribozymes were internally labeled using small amounts of $\alpha[^{32}\text{P}]\text{-ATP}$ during T7 transcription. The hammerhead ribozyme cleaves co-transcriptionally and generates a 5'-hydroxyl group on the

triphosphorylation ribozymes. The ribozymes were separated by denaturing polyacrylamide gel electrophoresis (PAGE), excised by UV shadowing, eluted in 300 mM NaCl, ethanol precipitated, dissolved in 10mM Tris•HCl pH 8.3 and their concentration measured by the A_{260} . All isolates were initially tested with the additional 12 nucleotides used as a 3'-PCR primer binding site during the selection. Subsequent experiments with clone 11 to identify the functional mutations also employed this protocol with the exception that the 12 3'-terminal nucleotides were removed.

The activity assay was performed as described previously (17). In short, 8 μ M ribozyme was incubated with 50 mM Tmp, 100 mM $MgCl_2$, and 50 mM Tris•HCl pH 8.3. Small aliquots were taken at different time points and quenched with an excess of Na_2EDTA . DNAzyme reactions to separate the 8 5'-terminal nucleotides of the ribozyme were initiated by addition of DNAzyme to a final concentration of 1.6 μ M and a 24-nucleotide long DNA oligonucleotide to a final concentration of 1.6 μ M. The 24-nucleotide long DNA oligonucleotide anneals immediately downstream of the DNAzyme to the triphosphorylation ribozyme and assists in DNAzyme binding and cleavage as well as inactivating the triphosphorylation ribozyme. Note that for the faster ribozymes of this study the quenching required perfect complementarity between the 24 nucleotide oligonucleotide and the corresponding sequence of the ribozyme. After heat renaturation (2'/80°C) $MgCl_2$ was added to a final concentration of 100 mM to allow DNAzyme catalysis. Aliquots of the DNAzyme reaction were quenched in formamide and Na_2EDTA then run on a 22.5% denaturing PAGE gel to separate

the triphosphorylated and unreacted 8-nucleotide fragments resulting from the DNAzyme reaction. Bands were imaged on a PMI phosphoimager (BioRad) and quantified using the Quantity One software.

Kinetic analysis of *trans*-triphosphorylation reactions

All experiments with TPR1e used the *trans*-triphosphorylation protocol and performed essentially as shown (17). In the *trans*-triphosphorylation experiment, a radiolabeled 14-nucleotide RNA oligomer with a 5' hydroxyl group was used as substrate. The 5' portion of the ribozyme was truncated to anneal to the 14-nucleotide substrate via the P1 and P2 helices and position the substrate 5'-hydroxyl group at the active site. Ribozymes were transcribed using T7 RNA polymerase without radioactive label and purified as described above. The substrate oligonucleotide was prepared essentially as previously (17). Briefly, a PCR product was generated that encoded a T7 promoter, and the substrate flanked by hammerhead ribozymes at their 5'- and 3'-terminus. The hammerhead ribozymes were used to generate the 5'-hydroxyl group and to ensure homogeneity at the 3'-terminus of the substrate. T7 transcription was performed with α -³²P-ATP. T7 transcripts were purified as described for the ribozymes above. No concentration measurement was made; trace amounts were used in each reaction.

The triphosphorylation experiment was set up similar to the *cis*- reaction, where ribozyme and substrate were mixed with Tmp, MgCl₂, and Tris•HCl pH 8.3 (or other components as given in the text). Samples were taken by adding a

small aliquot from the reaction into formamide and EDTA. The *trans*-reaction was quenched efficiently with the addition of formamide. Reaction products were run on a 20% PAGE gel to separate the triphosphorylated from unreacted 14-nucleotide substrate.

The *trans*-reaction was used in the covariation, simulated seawater, NaCl titration, and optimal conditions experiments.

Mutagenesis for the identification of beneficial mutations, and base covariation experiments

Two strategies were used for generating the different ribozymes constructs, PCR mutagenesis and quickchange site-directed mutagenesis. Mutations reverting the sequence of clone 11 to the wild type sequence were achieved with site-directed mutagenesis for the 12 internal mutations while the four mutations near the ribozyme 3'-terminus were introduced with a 3'-PCR primer. Internal mutations in TPR1_II and its variants were introduced by overlapping PCR primers that contained the necessary mutations. PCR products were then cloned into pUC19 and sequenced. The six mutations inserted into TPR1_II were introduced using quickchange site directed mutagenesis. For the covariation experiment, positions G33C and G35C were introduced using quickchange site directed mutagenesis, positions C90G and C88G were introduced via a 3'-PCR primer.

3.6 Acknowledgements

This work was supported by a National Aeronautical Space Agency (NASA) grant awarded to U.F.M.

Chapter 3, in full, is currently being prepared for submission for publication of the material. **Dolan, GF**; Müller, UF. The dissertation author is the first author on this paper.

3.7 Supporting Information

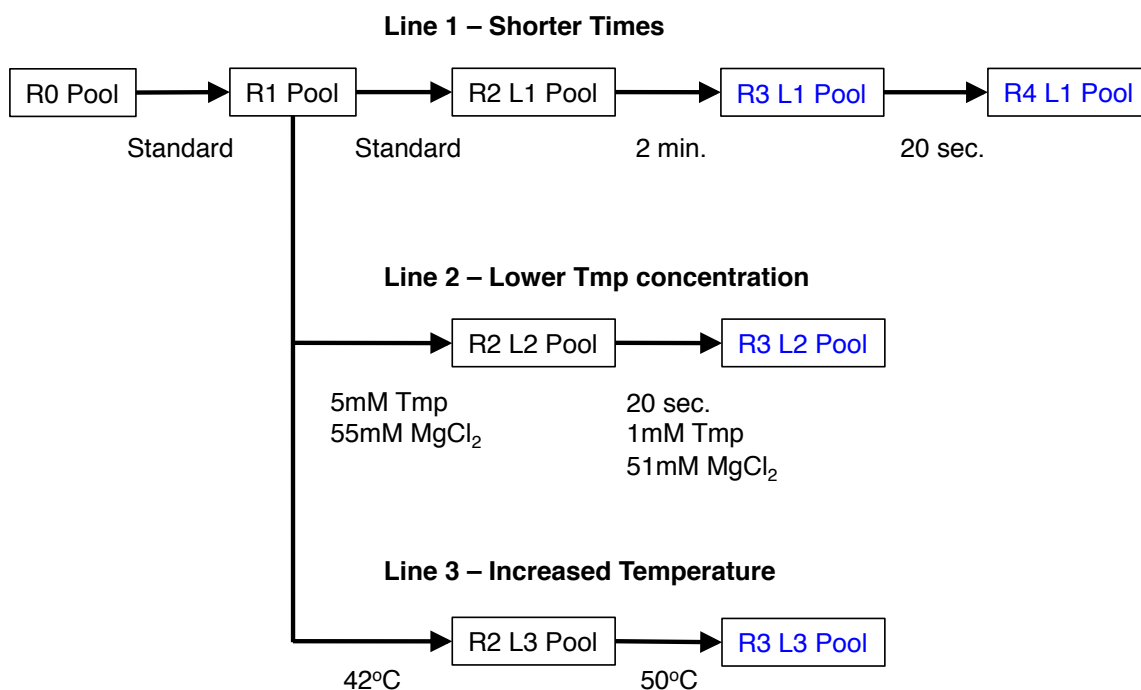


Figure S3.1: Diagram of doped selection performed. Boxes represent pools of RNA and arrows represent rounds of selection. From the initial pool (R0 pool, left), round 1 was performed using the ‘standard’ reaction conditions. ‘Standard’ describes the incubation conditions that were used in the previous selection, with an incubation time of 3 hours, an incubation temperature of 22°C, a ribozyme concentration of 100nM, a Tmp concentration of 50mM, and MgCl₂ concentration of 100mM and 50mM Tris/HCl pH 8.3. For selection steps under different conditions, only those conditions that differed from the standard condition are noted below the arrow. Round 2 was performed using three separate conditions (lines) as indicated by the branching of the R1 pool. Line 1 maintained ‘standard’ conditions in round 2, while Line 2 lowered the Tmp concentration while maintaining the same free Mg²⁺ concentration (5mM TmP / 55mM MgCl₂, as indicated under the arrow), and Line 3 used ‘standard’ conditions at an elevated temperature (42°C, as indicated under the arrow). In subsequent rounds each line was kept separate & exposed to the conditions indicated below the arrows. Blue indicates pools where individual ribozymes were cloned, isolated, and tested for activity.

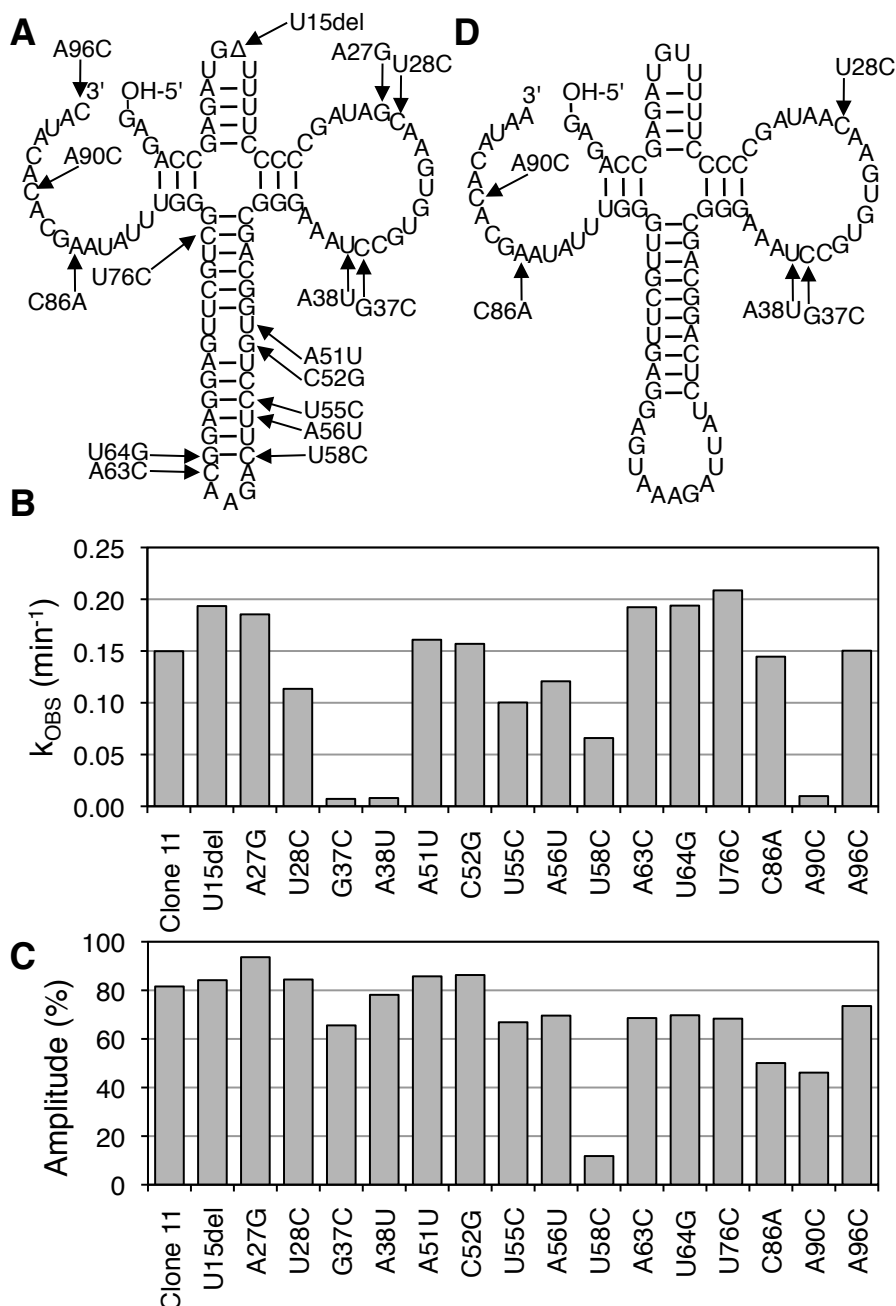


Figure S3.2: Identification of mutations in isolated ribozyme clone 11 that are necessary and sufficient for full activity. (A) Proposed secondary structure for clone 11, which contains 16 mutations relative to the parent ribozyme TPR1 (labeled with arrows). **(B)** Observed single-exponential rates of triphosphorylation kinetics with ribozyme variants of clone 11 in which the indicated mutation is reverted to the parent sequence. Error bars are standard deviations of triplicate experiments. **(C)** Observed amplitudes of the single-exponential reaction kinetics of the same clones. Error bars are standard deviations of triplicate experiments. **(D)** Proposed secondary structure of ribozyme variant TPR1_II, which contains the five necessary mutations of clone 11.

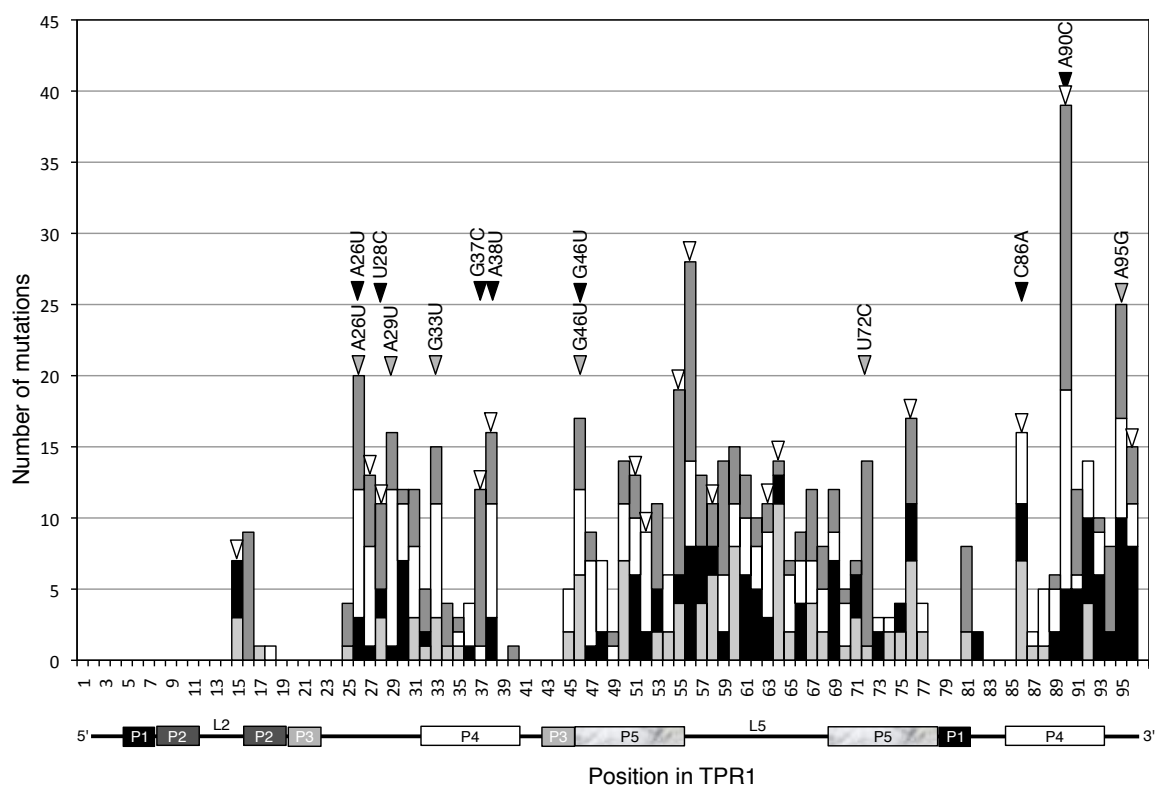


Figure S3.3: Position of mutations in the 58 clones isolated from the selection that had at least the same activity as the parent ribozyme TPR1. The number on the x-axis denotes the nucleotide position in the ribozyme. The columns are coded according to the nucleotide to which the position was mutated: dark grey: mutation to C; white: mutation to U; black: mutation to G; light grey: mutation to A. White triangles denote the 16 mutations present in clone 11. Grey triangles denote additional mutations that were tested for benefits by inserting them into ribozyme variant TPR1_II. Black triangles indicate the seven mutations of TPR1e. The illustration below the graph shows the secondary structure elements in the ribozyme.

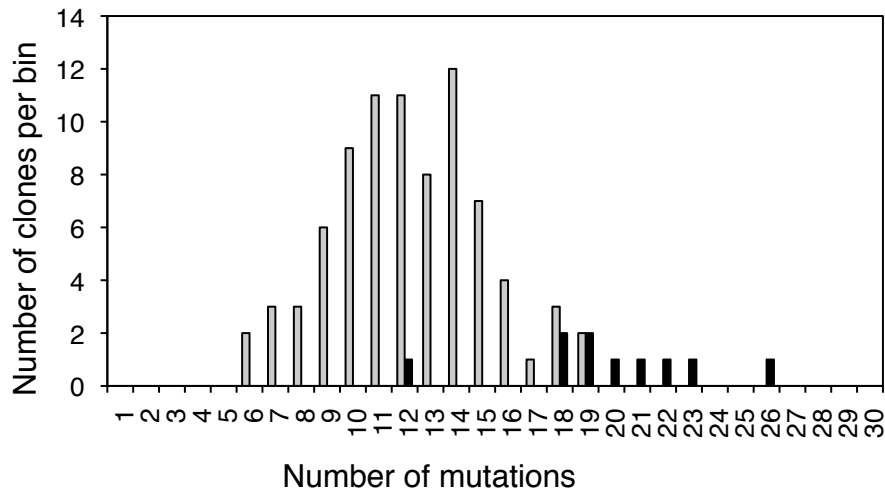


Figure S3.4: Mutational load in the pool before and after selection. The clones identified in the doped selection were binned according to their number of mutations (including deletions). The number of mutations is given on the x-axis; the number of clones per bin is given on the y-axis. Clones isolated from the pool before selection are represented in black columns. Clones isolated after selection are shown in grey columns. The average number of mutations in clones before the selection is 19.8; after the selection it is 12.2.

3.8 References

1. Woese, C. R. (1967) The genetic code. The molecular basis for genetic expression, *Harper & Row, New York*.
2. Crick, F. H. C. (1968) The origin of the genetic code, *J. Mol. Biol.* 38, 367-379.
3. Orgel, L. E. (1968) Evolution of the genetic apparatus, *J. Mol. Biol.* 38, 381-393.
4. Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*, *Cell* 31, 147-157.
5. Noller, H. F., Hoffarth, V., and Zimniak, L. (1992) Unusual resistance of peptidyl transferase to protein extraction procedures, *Science* 256, 1416-1419.

6. Nissen, P., Hansen, J., Ban, N., Moore, P. B., and Steitz, T. A. (2000) The structural basis of ribosome activity in peptide bond synthesis, *Science* 289, 920-930.
7. Sprengel, G., and Follmann, H. (1981) Evidence for the Reductive Pathway of Deoxyribonucleotide Synthesis in an Archaeobacterium, *FEBS letters* 132, 207-209.
8. Ellington, A. D., and Szostak, J. W. (1990) In vitro selection of RNA molecules that bind specific ligands, *Nature* 346, 818-822.
9. Tuerk, C., and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase, *Science* 249, 505-510.
10. Bartel, D. P., and Szostak, J. W. (1993) Isolation of new ribozymes from a large pool of random sequences, *Science* 261, 1411-1418.
11. Martin, L. L., Unrau, P. J., and Muller, U. F. (2015) RNA Synthesis by in Vitro Selected Ribozymes for Recreating an RNA World, *Life (Basel)* 5, 247-268.
12. Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E., and Bartel, D. P. (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension, *Science* 292, 1319-1325.
13. Orgel, L. E. (2004) Prebiotic chemistry and the origin of the RNA world, *Critical reviews in biochemistry and molecular biology* 39, 99-123.
14. Etaix, E., and Orgel, L. E. (1978) Phosphorylation of nucleosides in aqueous solution using trimetaphosphate: Formation of nucleoside triphosphates, *J. Carbohydrates Nucleosides Nucleotides* 5, 91-110.
15. Pasek, M. A., Harnmeijer, J. P., Buick, R., Gull, M., and Atlas, Z. (2013) Evidence for reactive reduced phosphorus species in the early Archean ocean, *Proceedings of the National Academy of Sciences of the United States of America* 110, 10089-10094.
16. Pasek, M. A., Kee, T. P., Bryant, D. E., Pavlov, A. A., and Lunine, J. I. (2008) Production of potentially prebiotic condensed phosphates by phosphorus redox chemistry, *Angew Chem Int Ed Engl* 47, 7918-7920.
17. Moretti, J. E., and Muller, U. F. (2014) A ribozyme that triphosphorylates RNA 5'-hydroxyl groups, *Nucleic acids research* 42, 4767-4778.
18. Stewart, P. B., and Munjal, P. (1970) Solubility of carbon dioxide in pure water, synthetic sea water, and synthetic seawater concentrates at -5° to

25 °C. and 10- to 45-atm. pressure, *Journal of Chemical and Engineering Data* 15, 67-71.

19. Hay, W. W., Migdisov, A., Balukhovsky, A. N., Wold, C. N., Flögel, S., and Söding, E. (2006) Evaporites and the salinity of the ocean during the Phanerozoic: Implications for climate, ocean circulation and life, *Paleogeography, Paleoclimatology, Paleoecology* 240, 3-46.
20. Zahnle, K., and Walker, J. C. (1987) A constant daylength during the Precambrian era?, *Precambrian Res* 37, 95-105.
21. Ekland, E. H., and Bartel, D. P. (1995) The secondary structure and sequence optimization of an RNA ligase ribozyme, *Nucleic Acids Res* 23, 3231-3238.
22. Curtis, E. A., and Bartel, D. P. (2005) New catalytic structures from an existing ribozyme, *Nat Struct Mol Biol* 12, 994-1000.
23. Curtis, E. A., and Bartel, D. P. (2013) Synthetic shuffling and in vitro selection reveal the rugged adaptive fitness landscape of a kinase ribozyme, *RNA* 19, 1116-1128.

Chapter 4

Discussion & Future Directions

4.1 *Trans*-splicing with the *Azoarcus* Ribozyme

The results of chapter 2 in this dissertation provide a framework for designing *trans*-splicing *Azoarcus* ribozymes for efficient *trans*-splicing *in vitro*. The *trans*-splicing *Azoarcus* ribozyme is most active when contacts are formed that closely mimic the structure of its natural *cis*-splicing interactions (Figure 4.1). In this structural context, *trans*-splicing *Azoarcus* ribozymes are as efficient as *trans*-splicing *Tetrahymena* ribozymes using their preferred structural context. This study, unlike previous works with non-*Tetrahymena* *trans*-splicing group I intron ribozyme, tested unique EGS and structural designs, rather than strictly applying previous EGS design principles. Therefore, the outcome of this study in a larger picture is that *trans*-splicing group I intron EGS designs should be individualized to take into account the structural contacts that each ribozyme forms when performing the splicing reaction.

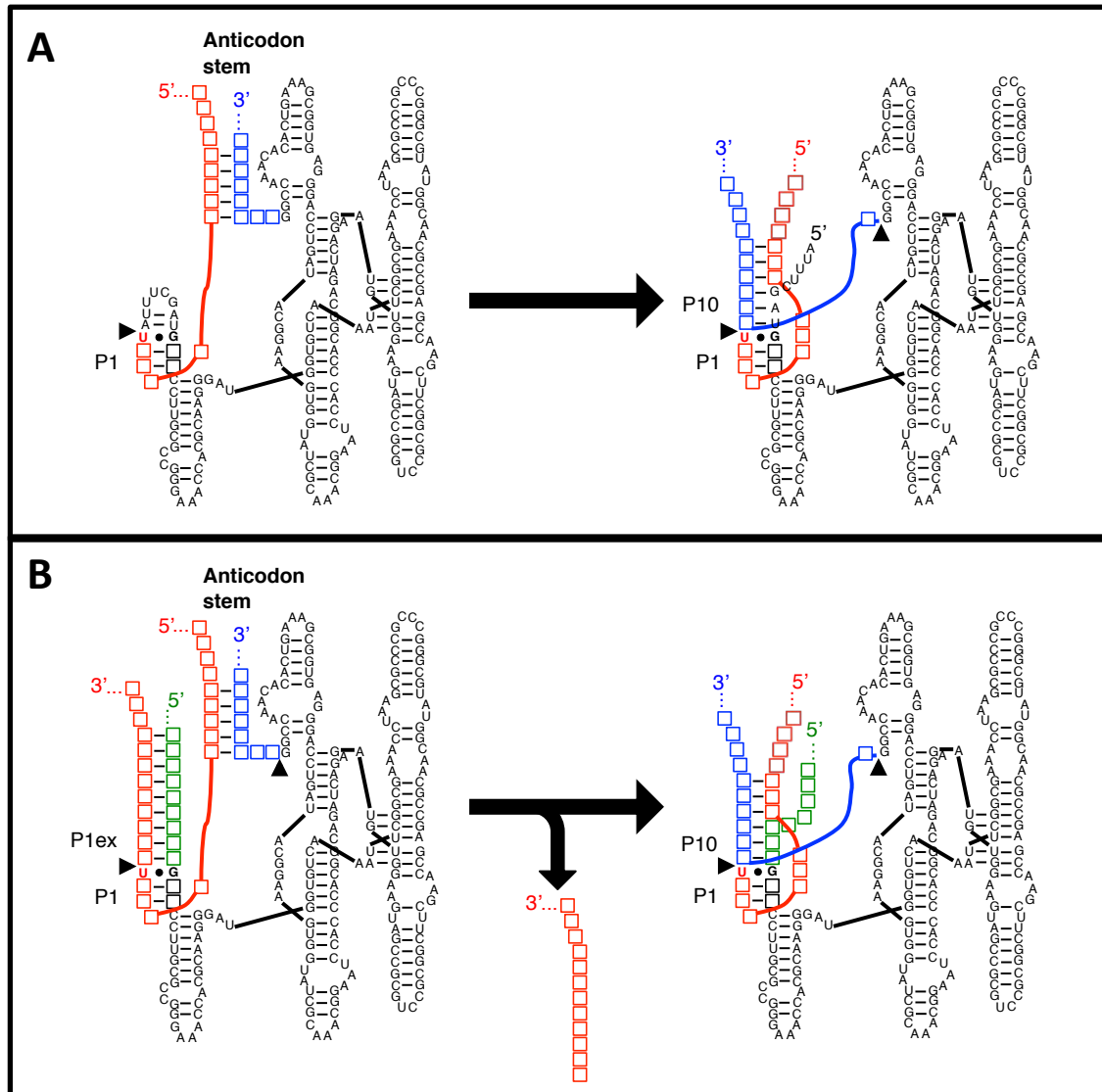


Figure 4.1: Preferred *Trans*- and *Cis*-splicing Structural Contacts made by the *Azoarcus* ribozyme. (A) *Cis*-splicing diagram of the *Azoarcus* ribozyme showing the structural contacts made before the first step (left) and second step (right) of splicing. The anti-codon stem formed between the 5' and 3' exon (left) and P10 formed between the P1 extension and the 3' exon (right) are highlighted. 5' exon is in red, ribozyme is in black, and 3' exon is in blue. (B) *Trans*-splicing diagram of the *Azoarcus* ribozyme showing the preferred structural contacts made before the first step (left) and the second step (right) of splicing. Notice the similarity of anti-codon stem formation (left) and the P10 formation (right) between *cis*- and *trans*-splicing. The color scheme is the same as in (A) except the extended guide sequence, which is in green. Open boxes represent possible variable nucleotide sequence.

The second goal of this project was to obtain a highly efficient *Azoarcus* ribozyme that was active *in vivo*. Unfortunately, *trans*-splicing activity with the *Azoarcus* ribozyme was not detectable in *E. coli* cells. Surprisingly, when switching to the *cis*-splicing reaction and inserting the ribozyme into different sites along the chloramphenicol acetyl transferase (CAT) mRNA sequence, the *Azoarcus* ribozyme was also unable to support a growth phenotype. Further experiments testing the *cis*-splicing *Azoarcus* ribozyme with only its natural flanking sequences yielded the most efficient construct measured in *E. coli* cells, but the splicing product was only ~4% as determined by RT-qPCR. Comparatively, the *Tetrahymena* ribozyme produced ~60% *cis*-splicing product even without its natural flanking sequences. The low splicing activity of the *Azoarcus* ribozyme suggests that: (1) some component of the *E. coli* cellular machinery, be it a chaperone protein or otherwise, could be interacting with and inhibiting the function of the *Azoarcus* ribozyme, or (2) there could be a specific splicing cofactor that is present in *Azoarcus* sp. BH72 that helps the ribozyme fold into its active conformation. Indeed, RNA-protein interactions have previously been found *in vitro* to alter the splicing activity of various group I intron ribozymes. Therefore, it is possible that one of the many RNA chaperone proteins in *E. coli* could have inhibited the *Azoarcus* ribozyme (1-4). (See Chapter 2, section 2.4 for a discussion of other possible components of *E. coli* that could have inhibited *Azoarcus*' activity.) Considering the other option, however, there is no known cofactor in *Azoarcus* sp. BH72 that aids in the splicing reaction of the *Azoarcus* group I intron ribozyme. Despite this, previous

studies have shown that some *in vitro* self-splicing group I introns still require protein cofactors for efficient splicing *in vivo* (5-7). Accordingly, it is possible that the *Azoarcus* group I intron may require a splicing cofactor for activity in *Azoarcus* sp. BH72. Future experiments could address both of these options.

An experiment to possibly address both of these options (whether the *Azoarcus* ribozyme is inhibited by a factor in *E. coli* cells or whether it requires a cofactor found in *Azoarcus* cells) would be to insert the *cis*-splicing *Azoarcus* ribozyme in the analogous gene (tRNA^{Ile}) of other organisms like yeast or even mammalian cells and determine if the ribozyme is capable of efficiently splicing. A similar approach has been used with the *Tetrahymena* ribozyme to probe potential *in vivo* protein facilitation of ribozyme folding and activity (8). If the *Azoarcus* ribozyme efficiently splices in other organisms, it would suggest that the observed inhibition in *E. coli* is organism specific and would provide evidence that efficient *in vivo* splicing of the *Azoarcus* ribozyme is not strictly dependent on a cofactor within *Azoarcus* sp. BH72. Once activity is established in the cells, it would be possible to test the *trans*-splicing reaction employing the defined EGS and structural principles for the *Azoarcus* ribozyme. There is reason to be hopeful that testing the *Azoarcus* ribozyme in other cellular contexts would succeed because recent *in vitro* work suggests that the *Azoarcus* ribozyme folding and activity improves in the presence of known RNA helicases from other organisms (9). In this report both CYT-19 and Mss116p, two RNA helicases from *Neurospora crassa* and *Saccharomyces cerevisiae*, increased the percentage of *Azoarcus* ribozyme that folded into its native conformation. It is unclear from this

report if the observed *in vitro* improvement would translate to increased *in vivo* activity; thus, it is necessary to perform the experiment. If the ribozyme were inactive in other cellular systems, however, then it could indicate that the cellular machinery of multiple organisms (and not just *E. coli*. cellular machinery) can inhibit the ribozyme, or alternatively, it would further suggest that a specific splicing cofactor in *Azoarcus* sp. BH72 is necessary for full activity.

In another approach, the low activity of the *Azoarcus* ribozyme could be improved in an *in vivo* evolution. This allows for successive rounds of introducing mutations and selection to enrich for the most active sequence variants. The motivation for this approach is two-fold. First, the optimal temperature for the *Azoarcus* ribozyme is 60-75°C (10, 11). By performing an *in vivo* evolution in *E. coli* cells, it would encourage the emergence of sequence and structural modifications that could shift the optimal temperature toward 37°C (growth temperature of *E. coli*) and improve the splicing activity. The Müller lab has prior experience with *in vivo* evolutions of group I intron ribozymes in *E. coli* cells, and is well suited to perform these experiments (12, 13). Second, a necessary starting point for any evolution is that the parent ribozyme (the starting point of the evolution) shows a basal level of activity that mediates a growth phenotype. This is required for the selection step in the evolution procedure. Since the best *cis*-splicing *Azoarcus* ribozymes (flanked by their native sequence) had some activity (1-4%), it is likely that such starting constructs can be generated with the *Azoarcus* ribozyme. Once the evolution system is set up it is likely that sequence variants of the *Azoarcus* ribozyme would possess improved splicing activity. With

improved *in vivo* activity of these *cis*-splicing *Azoarcus* ribozyme constructs, future *trans*-splicing experiments can be performed using the secondary structural principles found in this work.

For future *trans*-splicing studies in cells targeting previously untested group I intron ribozymes, this dissertation made another important finding. Not all *in vitro* self-splicing group I intron ribozymes are necessarily active in a foreign *in vivo* environment. Since this is the case, the first step in testing any new *trans*-splicing ribozyme should be to ensure that the *cis*-splicing reaction is functional *in vivo*. Once this is confirmed (and possibly optimized in an evolution experiment), *trans*-splicing constructs can be generated and tested for their level of *trans*-splicing activity in cells.

The disparity between known group I introns and those tested for *trans*-splicing activity is still large. *Azoarcus* is just the fifth group I intron to be tested and only the second to be tested *in vivo*. This is compared to the more than 16,000 group I intron ribozymes known (14). By continuing to test other group I intron ribozymes, more will be understood about their catalytic potential, and there is the possibility of identifying a more efficient *trans*-splicing ribozyme than the *Tetrahymena* ribozyme. Thus, the testing of more group I introns for *trans*-splicing activity remains an important contribution to the advancement of the field.

4.2 Triphosphorylation Ribozymes

The doped selection of TPR1 yielded a ribozyme with seven mutations (called TPR1e) that is ~24-fold faster under 'standard' conditions (22°C, 50mM trimetaphosphate, 100mM Mg^{2+} , and pH 8.3; $\sim 0.31 \text{ min}^{-1}$) and under optimal conditions reacts at a rate of $\sim 6.8 \text{ min}^{-1}$. At optimal conditions it was estimated that TPR1 was able to catalyze the 5' triphosphorylation rate $\sim 1 \times 10^7$ times faster than the uncatalyzed reaction (15). Now, TPR1e boasts a rate that is $\sim 2 \times 10^8$ times faster. Because of the sigmoidal nature of the trimetaphosphate titration kinetics at the conditions tested, accurate k_{cat}/K_m values cannot be reported, making comparisons with other *in vitro* selected ribozymes difficult.

Besides enhancing the catalytic rate, another goal of this project was to also obtain a ribozyme that was active at more prebiotically plausible trimetaphosphate concentrations. To test TPR1e under more prebiotically plausible conditions, the ribozyme was tested in simulated seawater with 1mM trimetaphosphate, a prebiotically plausible concentration (16-18). At 22°C the measured triphosphorylation rate was 0.0014 min^{-1} and at 40°C it was 0.004 min^{-1} . The measured rate at 22°C with the same Mg^{2+} and trimetaphosphate concentrations without the other components of the synthetic seawater was ~15-fold faster than with all seawater components. This showed that at least one of the components in synthetic seawater inhibited the ribozyme. A subsequent experiment titrating NaCl, a large component in seawater, into standard reaction conditions (Mg^{2+} , trimetaphosphate, and Tris pH 8.3) showed that NaCl inhibits the ribozyme, and at the concentration levels present in

seawater, NaCl is responsible for a 2-fold reduction in rate. Another likely inhibitory component in synthetic seawater is Ca^{2+} , which may compete with Mg^{2+} ions important for structure and/or catalysis by the ribozyme. Such an inhibition by Ca^{2+} has been observed for other ribozymes previously (19). Overall, this demonstrates the activity of TPR1e at prebiotically plausible trimetaphosphate concentrations and shows that several components of seawater including NaCl inhibit the functionality of the ribozyme.

Another step in improving the prebiotic relevance of this ribozyme would be to identify a variant that can triphosphorylate a free nucleoside. To this end, a series of experiments were set up to test TPR1e's ability to triphosphorylate a free guanosine. Because the architecture of the active site is unknown, multiple constructs of TPR1e containing different 5' truncations were tested for triphosphorylation activity of a free guanosine; the idea being that a portion of the 5' structure might interfere or sterically hinder the binding of the free guanosine. Unfortunately, none of the tested constructs showed any measurable activity.

Although TPR1e could not triphosphorylate a free nucleoside, both TPR1 and TPR1e are capable of a *trans* reaction where a short oligonucleotide with a 5' OH can be triphosphorylated. This reaction is achieved through base pairing between the short oligonucleotide and the ribozyme, but as the oligonucleotide is shortened and the base pairing decreases, activity of the ribozyme drops. It seems that the current triphosphorylation ribozyme does not bind a free guanosine nucleoside well enough to perform the reaction. Thus, substrate binding must improve to triphosphorylate free nucleosides.

To better understand the architecture and catalytic mechanism of this ribozyme, further biochemical and structural analysis is necessary. A crystal structure of this ribozyme showing 2-4 angstrom level resolution of the active site could greatly aid in experimental design of improving the *trans* reaction. Indeed, efforts are underway to crystallize this ribozyme. SHAPE structural probing, a technique that identifies flexible and paired regions of RNAs, looking at the triphosphorylation ribozyme with and without trimetaphosphate would also provide information about the dynamics of the ribozyme, specifically how the presence of trimetaphosphate alters ribozyme folding. Knowing about ribozyme dynamics and a potential binding location of trimetaphosphate can improve construct design for the *trans* reaction.

Further biochemical experiments should focus on TPR1e's means of substrate binding. If substrate recognition depends on base pairing, then triphosphorylating the other three canonical nucleosides (A, T, C) would most likely result in lowered or no activity. A known example of ribozyme base specificity required for binding of a free nucleoside/nucleotide is the group I intron's G binding pocket. It contains a G-C pair that forms a base triple with the binding G, and is unable to efficiently accommodate the binding of any other nucleotides (20). If the pair is switched to a C-G it can then only accommodate an A (21). Notably, previously *in vitro* selected ribozymes catalyzing the glycosidic bond between the ribose and the nucleobase have shown nucleoside specificity (22, 23). Potential nucleotide specificity of TPR1e would reduce its prebiotic usefulness because if this ribozyme existed in an RNA world, more than

one triphosphorylation ribozyme would have been required to triphosphorylate the various nucleosides necessary to carry meaningful genetic information.

Testing short oligonucleotides with a 5' terminal base comprising A, T, or C would be a quick way to test potential nucleotide specificity.

If substrate recognition does not depend on base pairing or nucleobase identity, then the ribose moiety would be responsible. To understand which portions of the ribose are used in recognition and binding, substituting ribose analogs could elucidate which contacts are necessary. It would be interesting to see if TPR1e can triphosphorylate a DNA moiety, which would mean the 2' hydroxyl of the ribose is not needed for binding or catalysis. Other analogues like altritol nucleic acid (ANA) could also give insights to how substrates are bound and recognized by the active site of the ribozyme (Figure 4.2).

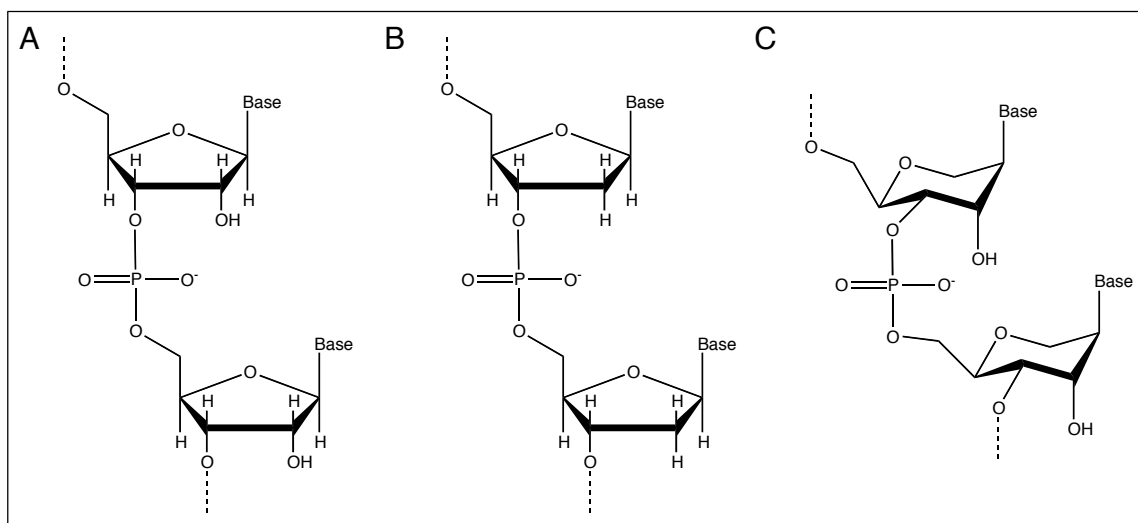


Figure 4.2: RNA and Analogue Structures. (A) The chemical structure of a ribonucleic acid (RNA) polymer. (B) The Chemical structure of a deoxyribonucleic acid (DNA) polymer. Note the absence of the 2' hydroxyl. (C) The chemical structure of an altritol nucleic acid (ANA) polymer. Note the additional carbon in the ring structure and the corresponding shift in conformation.

Other future work with the triphosphorylation ribozyme should focus on isolating, and characterizing additional ribozymes from the initial selection. Upon measuring their catalytic rate and determining their secondary structure, interesting questions can be addressed such as: determining how many different structures (or classes) of ribozymes can catalyze this reaction, ascertaining which structural motifs are possible, and establishing which structures are the most common.

One difficulty with testing many ribozymes is the effort required to measure the activity. A possible time saving step in this protocol would be using a fluorescent probe to detect activity rather than using a radioactively labeled oligonucleotide and running a gel. The proposed design of the fluorescent probe assay can be seen in Figure 4.3. It is a variation of the ligation assay used in previous works to preliminarily detect activity of isolated ribozymes. The probe is designed in a hairpin conformation with a fluorophore on one end and a quencher on the other end, such that when the probe is free in solution the 5' and 3' ends will be adjacent to one another and no signal will be detected (Figure 4.3B). The probe sequence is the reverse complement of the 5' end of the ribozyme and the short oligonucleotide β . After an active ribozyme with a 5' triphosphate is successfully ligated to a short oligonucleotide β using the ligase ribozyme (Figure 4.3A) and is heat renatured in the presence of the probe, the complete complementarity between the probe and the 5' end of the active ribozyme allows for hybridization. This separates the fluorophore and quencher resulting in increased fluorescence (Figure 4.3B). Preliminary experiments are

necessary to determine the optimal length of the hairpin stem loop and optimal length of β/β' sequence to minimize background and false positives. The use of the fluorescent probe will eliminate the necessity for radiolabeling an oligonucleotide and running a gel, saving hours of work and increasing the total number of ribozymes easily tested.

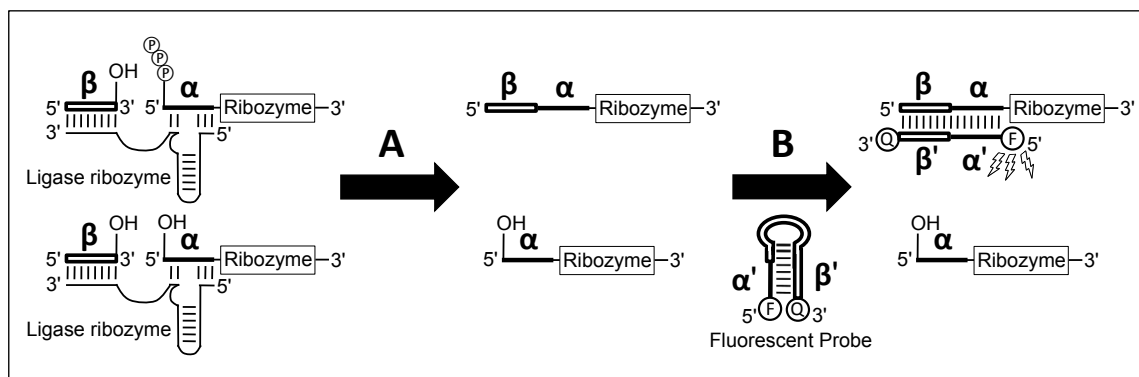


Figure 4.3: Fluorescent Ligation Assay to measure triphosphorylation activity. (A)

Active ribozymes are ligated to short oligonucleotides using the ligase ribozyme (top), while inactive ribozymes will not react, maintaining a 5' hydroxyl (bottom). The ligase ribozyme contains sequence complementarity to bind with the constant 5' region of the triphosphorylation ribozyme (α) and sequence of the short oligonucleotide (β). $\textcircled{P}$ represents a phosphate group. **(B)** Ribozymes are then heat renatured in the presence of a fluorescent probe. The fluorescent probe is designed as a hairpin and contains a fluorophore on one end ($\textcircled{F}$) and a quencher ($\textcircled{Q}$) on the other end. The probe is the reverse complement of the constant 5' region of the triphosphorylation ribozyme (α) and the short oligonucleotide (β) (α' and β'). Free in solution the hairpin probe will not emit fluorescence. After heat renaturation, the probe will hybridize with complete complementarity to active ribozymes that contain both β and α on their 5' terminus, but will not hybridize efficiently with the inactive ribozymes contain only α on the 5' terminus. Thus, active ribozymes will show an increase in fluorescence due to the spatial separation between the fluorophore and the quencher in the linear hybridized form, while inactive ribozymes will now show fluorescence, the probe preferring the hairpin conformation.

4.3 References

1. Vicens, Q., Paukstelis, P. J., Westhof, E., Lambowitz, A. M., and Cech, T. R. (2008) Toward predicting self-splicing and protein-facilitated splicing of group I introns, *RNA* 14, 2013-2029.

2. Semrad, K. (2011) Proteins with RNA chaperone activity: a world of diverse proteins with a common task-impediment of RNA misfolding, *Biochemistry research international* 2011, 532908.
3. Coetzee, T., Herschlag, D., and Belfort, M. (1994) Escherichia coli proteins, including ribosomal protein S12, facilitate in vitro splicing of phage T4 introns by acting as RNA chaperones, *Genes & development* 8, 1575-1588.
4. Waldsich, C., Grossberger, R., and Schroeder, R. (2002) RNA chaperone StpA loosens interactions of the tertiary structure in the td group I intron in vivo, *Genes & development* 16, 2300-2312.
5. Gampel, A., and Cech, T. R. (1991) Binding of the Cbp2 Protein to a Yeast Mitochondrial Group-I Intron Requires the Catalytic Core of the Rna, *Genes & development* 5, 1870-1880.
6. Lambowitz, A. M., and Perlman, P. S. (1990) Involvement of aminoacyl-tRNA synthetases and other proteins in group I and group II intron splicing, *Trends in biochemical sciences* 15, 440-444.
7. Shaw, L. C., and Lewin, A. S. (1995) Protein-induced folding of a group I intron in cytochrome b pre-mRNA, *The Journal of biological chemistry* 270, 21552-21562.
8. Zhang, F., Ramsay, E. S., and Woodson, S. A. (1995) In vivo facilitation of Tetrahymena group I intron splicing in Escherichia coli pre-ribosomal RNA, *RNA* 1, 284-292.
9. Sinan, S., Yuan, X., and Russell, R. (2011) The Azoarcus group I intron ribozyme misfolds and is accelerated for refolding by ATP-dependent RNA chaperone proteins, *The Journal of biological chemistry* 286, 37304-37312.
10. Tanner, M., and Cech, T. (1996) Activity and thermostability of the small self-splicing group I intron in the pre-tRNA(Ile) of the purple bacterium Azoarcus, *RNA* 2, 74-83.
11. Riley, C. A., and Lehman, N. (2003) Generalized RNA-directed recombination of RNA, *Chemistry & biology* 10, 1233-1243.
12. Olson, K. E., Dolan, G. F., and Muller, U. F. (2014) In vivo evolution of a catalytic RNA couples trans-splicing to translation, *PloS one* 9, e86473.
13. Amini, Z. N., and Muller, U. F. (2013) Low selection pressure aids the evolution of cooperative ribozyme mutations in cells, *The Journal of biological chemistry* 288, 33096-33106.

14. Zhou, Y., Lu, C., Wu, Q. J., Wang, Y., Sun, Z. T., Deng, J. C., and Zhang, Y. (2008) GISSD: Group I Intron Sequence and Structure Database, *Nucleic acids research* 36, D31-37.
15. Moretti, J. E., and Muller, U. F. (2014) A ribozyme that triphosphorylates RNA 5'-hydroxyl groups, *Nucleic acids research* 42, 4767-4778.
16. Stewart, P. B., and Munjal, P. (1970) Solubility of Carbon Dioxide in Pure Water, Synthetic Sea Water, and Synthetic Sea Water Concentrates at 5 Degrees to 2k Degrees C and 10-Atm to 45-Atm Pressure, *J Chem Eng Data* 15, 67-&.
17. Hay, W. W., Migdisov, A., Balukhovsky, A. N., Wold, C. N., Fogel, S., and Soding, E. (2006) Evaporites and the salinity of the ocean during the Phanerozoic: Implications for climate, ocean circulation and life, *Palaeogeogr Palaeocl* 240, 3-46.
18. Pasek, M. A., Kee, T. P., Bryant, D. E., Pavlov, A. A., and Lunine, J. I. (2008) Production of potentially prebiotic condensed phosphates by phosphorus redox chemistry, *Angew Chem Int Ed Engl* 47, 7918-7920.
19. Glasner, M. E., Bergman, N. H., and Bartel, D. P. (2002) Metal ion requirements for structure and catalysis of an RNA ligase ribozyme, *Biochemistry* 41, 8103-8112.
20. Michel, F., Hanna, M., Green, R., Bartel, D. P., and Szostak, J. W. (1989) The guanosine binding site of the Tetrahymena ribozyme, *Nature* 342, 391-395.
21. Been, M. D., and Perrotta, A. T. (1991) Group I intron self-splicing with adenosine: evidence for a single nucleoside-binding site, *Science* 252, 434-437.
22. Unrau, P. J., and Bartel, D. P. (1998) RNA-catalysed nucleotide synthesis, *Nature* 395, 260-263.
23. Lau, M. W., Cadieux, K. E., and Unrau, P. J. (2004) Isolation of fast purine nucleotide synthase ribozymes, *Journal of the American Chemical Society* 126, 15686-15693.