UNIVERSITY OF CALIFORNIA SAN DIEGO

In vitro selection of catalytic RNAs in emulsion

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

in

Chemistry

by

Arvin Akoopie

Committee in charge:

Professor Ulrich F. Müller, Chair
Professor Lin Chao
Professor Gourisankar Ghosh
Professor Navtej Toor
Professor Yitzhak Tor

2019

The dissertation of Arvin Akoopie is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California San Diego

2019

DEDICATION

To myself, for putting up with myself.

EPIGRAPH

His palms are sweaty, knees weak, arms are heavy.

There's vomit on his sweater already, mom's spaghetti.

—Eminem

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank my family for being so supportive. Alex, keep coding!

Thank you to my friends: Wesley “Gandalf the white” Chen, Regine “Penguin” Villongco, Jeremy “Chicken dinner” Hilgar, Matthew “Christian Fries” Wu, Vince “Bath salts” Harjono, Ellie “Dog” Harjono, Anna “Herbie” Guzikowski, Josh “Tacos” Arriola, Kevin “Loaded fries” Sweeney, Tim “Lamb with cumin” Wiryaman, Kayla “Iz okay” Chun, Daniel “Actor” Wine, and many others.

Special thanks to everyone who I had the pleasure of working with: Josh, Kevin, Ishani, Logan, Greg, and Zhaleh.

Significant thanks to my mentor and PI, Uli. You demonstrated to me on a daily basis, on every level, what it means to be a true scientist. You are the definition I now go by.

Credit also needs to be given to my favorite hobby, videogames. Without them I wouldn’t have gotten through this difficult period. The ever-classic Counter-Strike: Source has formed countless memories, especially the zombie, surf, and deathmatch gametypes. PUBG introduced me to a new type of FPS that I believe has tremendous potential. However, its success specifically depends on if the developers raise their standards. I’d like the thank Overwatch, but I can’t because Blizzard actively frustrates their loyal community by implementing counter-productive balance changes and fundamentally flawed designs. They over-nerfed my main, Roadhog, to the point where I haven’t picked him in over two years. I will never forgive them for this or for their lackluster matchmaking that seems intent on guaranteeing only a 50 percent win-rate, regardless of skill. Also everyone except me who plays this game is terrible and doesn’t understand that it’s a team-based effort. I had high hopes for Battlefield V, but DICE and EA seem to ignore everything but profits. Left 4 Dead 2 was a very fun game while it lasted and I enjoyed speedrunning it. Skyrim was alright, but not as good as Oblivion.

I’d also like to thank foreign policy and history for being entertaining subjects that helped keep my mind off of stressful events. The following is a randomly arranged list of interesting

topics I read during my time here: Nixon presidency (Opening to China and Watergate scandal), World War 2 (Eastern front), life in the Gulag, the Stalin regime, historical Jesus and early Christianity, Oswald's trip to Mexico City (Note: there seems to be contradictory information about his time there and the events that occurred in the Cuban embassy. Unfortunately it seems that instead of investigating all the possible details the upper echelons of the FBI and CIA were intent on obfuscating, and even destroying, evidence and possible leads. Not even they wanted to know where the rabbit hole led. As a result the final lingering questions about a potential Cuban conspiracy will probably never be answered.), Steele Dossier, Mao Zedong (Great Leap Forward and Cultural Revolution), Henry Kissinger's diplomacy, Wikileaks Cablegate, Brexit, and the Truman administration.

Certain foods helped me get through graduate school. These are: Mcdonald's 2 for 5 deal, Chik-Fil-A's spicy deluxe chicken combo, Santorini's loaded fries, Vallarta's Christian fries, Manna's KBBQ, Spicy House's entire menu, Hachi Ramen's tonton, Nijima's Ecuadorian bananas, anything dipped in bitchin' sauce, Happy Lemon's large milk tea with salted cheese foam, Taco Bell's doritos locos tacos, Ben and Jerry's cheesecake core, and hot cheetos.

Chapter 2, in full, is a reprint of the material as it appears in PLOS ONE. Dolan, Greg F.; Akoopie, Arvin; Müller, Ulrich F. 2015. The dissertation author was the second author of this paper. He contributed by characterizing the secondary structure of the ribozyme and performed base covariation experiments.

Chapter 3, in full, is a reprint of the material as it appears in Physical Chemistry Chemical Physics. Akoopie, Arvin; Müller, Ulrich F. 2016. The dissertation author was the first author of this paper.

Chapter 4, in full, is a reprint of the material as it appears in Journal of Molecular Evolution. Akoopie, Arvin; Müller, Ulrich F. 2018. The dissertation author was the first author of this paper.

Chapter 5, in full, is a reprint of the material as it appears in Nucleic Acids Research.

Akoopie, Arvin; Müller, Ulrich F. 2018. The dissertation author was the first author of this paper.

VITA

2014	B. S. in Biochemistry *Summa cum laude*, University of Nevada, Las Vegas
2016	M. S. in Chemistry, University of California San Diego
2019	Ph. D. in Chemistry, University of California San Diego

PUBLICATIONS

Akoopie A., Arriola JT., Müller UF. "A ribozyme that triphosphorylates free nucleosides [working title]", *In Preparation*, 2019.

Magde D., **Akoopie A.**, Magde M., Müller UF. "DLS measurements of a water-in-oil emulsion generated via microfluidizer [working title]", *In Preparation*, 2019.

Akoopie A., Müller UF. "The NTP binding site of the polymerase ribozyme", *Nucleic Acids Research*, 46(20), 2018.

Akoopie A., Müller UF. "Cotranscriptional 3-end processing of T7 rna polymerase transcripts by a smaller HDV ribozyme", *Journal of Molecular Evolution*, 86(7), 2018.

Akoopie A., Müller UF. "Lower temperature optimum of a smaller, fragmented triphosphorylation ribozyme", *Phys. Chem. Chem. Phys.*, 18, 20118-20125, 2016.

Dolan, GF., **Akoopie A.**, Müller UF. "A faster triphosphorylation ribozyme", *PLOS ONE*, 10(1), 2018.

ABSTRACT OF THE DISSERTATION

In vitro selection of catalytic RNAs in emulsion

by

Arvin Akoopie

Doctor of Philosophy in Chemistry

University of California San Diego, 2019

Professor Ulrich F. Müller, Chair

The RNA world hypothesis describes an early stage of life when RNA fulfilled both genomic and catalytic roles. One great effort to support this model is by generating catalytic RNAs (ribozymes) that can catalyze essential RNA world functions. One of these RNA world functions is establishing an energy metabolism in the form of chemically activated nucleotides. Trimetaphosphate has been shown to be a prebiotically plausible activating source that can generate nucleoside triphosphates from nucleosides. A main goal of the Müller lab is to generate ribozymes that can establish an energy metabolism pathway by using trimetaphosphate. Such a ribozyme could utilize trimetaphosphate and nucleosides to generated chemically activated NTPs. NTPs have an essential role in modern life (as cofactors and monomers for genetic

polymerization), and having a prebiotically plausible pathway for its generation would greatly support the RNA world hypothesis.

Previous projects from the Müller lab have established a ribozyme that can triphosphorylate its own 5' end. Although this was significant in establishing that triphosphorylation can be catalyzed by ribozymes, it did not generate freely diffusing NTPs. Here I describe a selection for a nucleoside triphosphorylation ribozyme as well as additional RNA evolutions and engineering to accomplish this goal. The main efforts taken were the evolution of a polymerase ribozyme towards a thio-modified nucleotide and characterizing the emulsion environment necessary for the selection to work.

Chapter 1

Introduction

1.1 The Origin of Life

One question that perhaps every person has pondered is how life originated. Answering this question from the scientific perspective requires two parameters to be answered: when was the Earth formed and when was the earliest organism. These two parameters would establish a time range in which natural chemical reactions would eventually lead to a genetic, reproducing system capable of open-ended evolution.

1.1.1 Age of the Earth

The Earth is approximately 4.54 billion years old [1]. As late as the early twentieth century geologists had models placing the Earth's age well below half a billion years old. Various methods were used to calculate these ages: cooling of the Earth, cooling of the Sun, Earth's tidal effects, sodium accumulation in the ocean, and sediment accumulation. Ultimately isotopic dating, calculating age from the natural radioactive decay of isotopes, enabled an accurate measurement of the age of the Earth.

There were four main ideas that enabled isotopic dating to determine the age of the Earth.

The first was Henry N. Russell's key idea that the Earth's age can be determined from the relative abundances of a "radioactive parent element and its daughter" [2]. He obtained an upper estimate of eight billion years by using then-current estimates of radium and lead in the crust and then backcalculating how much uranium and thorium was in the crust. He was then able to calculate how much time was needed for daughter lead to have been made by decaying parents uranium and thorium. Eight billion years was an upper estimate because there might have already been lead during Earth's formation.

The second main idea behind dating the Earth came from the invention of the mass spectrometer. This instrument was able to measure the masses of different isotopes to high accuracy. Most of the early pivotal work measuring the isotopes in the periodic table was done by F. W. Aston, a researcher in J. J. Thomson's lab. The important idea was that it was now possible to measure the ratio of isotopes, especially radioactive ones. Rutherford used this new data to show that ^{207}Pb was derived from ^{235}U and that, if ^{235}U and ^{238}U were in equal abundance during Earth's formation, that the Earth was 3.4 billion years old [3, 4].

The third concept establishing the age of the Earth was F. G. Houtermans' and C. C. Patterson's use of meteoritic lead isotope ratios as a "starting point" to compare current Earth lead isotopic ratios. The idea behind using meteorites is that they are isolated ancient systems that lack any sort of interfering geological activity that ancient Earth sample would contain. The main assumption made for this dating to work was that the Earth and the Canyon Diablo meteorite were formed at around the same time. Fitting the isochron curve gives Earth an age of 4.5 ± 0.3 billion years. Patterson later eliminated both assumptions. He made the "genetic" connection between Earth and meteorites (meaning they came from the same primordial composition) by predicting that modern Earth lead would fall on the same decay curve as meteoritic lead. Indeed the lead composition of deep-sea sediment in the Pacific ocean matched that of the meteorite's. Furthermore, Patterson measured five stone meteorites and added their lead composition to the previous data. The meteoritic geochron, a linear array of various lead compositions, matched

with the terrestrial one. Patterson's measurement of the age of the Earth is considered to be the first accurate one [5].

The final concept that established the age of the Earth is by F. Tera. Tera's method doesn't require the age of the lead deposits to be known, but instead assumes that the age-composition of various lead deposits have a "single point of congruency" for both their radiometric ratios and age of the reservoir they originated from. Measuring lead deposits from three different continents, Tera plotted the radiometric ratios against time from origin and found that all the lines intersect at a point 4.53 billion years ago. Tera later reanalyzed the samples using a different ratio and found the intersect at 4.54 billion years ago (Tera 1981). Using this method, it either represents the time the Earth was formed or the age of the debris which formed the Earth. Later dating of the moon and meteorites places the time-frame of which the Earth was formed to within fifty million years after the solar nebula began condensing solid matter. This condensation occurred at 4.566 billion years ago and so Tera's measurement of 4.53-4.54 billion years falls within this time-frame [6].

1.1.2 The earliest known organism

Finding evidence of the oldest organism is quite difficult [7, 8]. Although there are claims of microfossils that are 3.5 billion years old, and even though they do resemble earlier confirmed microfossils, there is debate in the community whether these are genuine signs of life or simply a mixture of human error with abiogenic processes.

Other than microfossils, stromatolites can also be viewed as evidence of early life [8]. Stromatolites are layers of sediment possibly deposited by ancient photosynthetic organisms. It's the result of calcium carbonate precipitating when the bacteria consume all the carbon dioxide in the water. Pivotal work by Allwood, Walter, Kamber, Marshall, and Burch provide the most compelling, albeit indirect, evidence for the earliest organism.

The Strelley Pool Chert (SPC) is a 3.43 billion year old rock formation in Pilbara Craton, Australia. Although the SPC had been described for over three decades, a variety of abiogenic

and mathematical models could attributed the stromatolites away from biogenic origin. Allwood et al's compelling study addressed each of these problems by doing a multi-kilomter-scale palaeontological and palaeoenviornmental study of the SPC [8]. The strongest evidence of stromatolites being of biogenic origin is that there were seven different types with distinct morphological attributes. If the SPC stromatolites were made from abiogenic origin then only one type of stromatolite would have been created and it would've been continuous. Furthermore, analysis of the "palaeovertical conical pseudocolumn" (the shape and direction of the stromatolite layers) showed a specific, vertically directed accretion. This type of accretion has no abiological explanation. Finally, there is chemical evidence of biogenesis by there being 250-fold higher rare earth element concentration compared to the surrounding areas. This would be explained by trapping of microbes in the sedimentary layers. Therefore, the earliest evidence of life is approximately 3.5 billion years ago.

If the Earth is 4.5 billion years old and the earliest living organisms were alive 3.5 billion years ago then there is a one billion year gap in which life must be originated.

1.2 The RNA World Hypothesis

The RNA World Hypothesis describes an early stage of life when RNA was responsible for both genetic and catalytic roles [9, 10, 11, 12, 13]. In modern life, RNA's purpose is primarily limited as an informational intermediary between genetic DNA and catalytic proteins. Its ancient responsibility was postulated as a solution to what was labeled as a "chicken and egg problem" [14, 15]. This paradox is the result from back-extrapolating how modern life carries out cellular function: how did the first organism develop catalytically inert double-stranded DNA for genetic information and genetically lacking catalytic proteins *at the same time*? Even when admitting the amazing nature of life itself, it was conceptually impossible for modern scientists to accept that both biomolecules would arise simultaneously and work in tandem from a prebiotic soup.

The solution was to suppose that early life consisted of only *one* biomolecule that could both genetically encode information and, from the same sequence, be catalytically active. Indeed, even though the first catalytic RNA wasn't discovered until the 1980's (two decades later), the originators of the RNA World Hypothesis realized that RNA might be able to fulfill both roles [16, 17].

There are four main lines of evidence supporting the RNA World Hypothesis: ribosomes and related translational machinery, the nature of coenzymes, introns, and the reduction of RNA to DNA. It was understood by the late 1960s that the genetic code is universal; all life uses triplet base codons to delineate one of twenty amino acids or a stop codon. One explanation for this universality is termed the "frozen accident theory", which states that any change in the genetic code would disrupt the function of enough highly-evolved proteins that it would be fatal to that mutated organism. This "careful" evolutionary pathway taken by all of life would therefore imply that all organisms evolved from a single organism. Similarly, the "Principle of Continuity", which states that "old" genetic code must still be recognizable by the next generation of organisms means that the earliest organism must have used a triplet codon-based genetic system. Since RNA is the genetic carrier of translation, this would suggest that RNA was present in the earliest organism. Furthermore, looking at the ribosome and translation machinery, one notices that it is also composed of RNA. Although it wasn't confirmed until 2003, the catalytic core of the ribosome (the peptidyl transferase) is entirely RNA [18]. tRNAs, whose job is to deliver amino acids to the nascent polypeptide-RNA-ribosome complex is also entirely RNA (although highly modified). Crick speculated that tRNA "looks like Nature's attempt to make RNA do the job of a protein". If the catalytic machinery, delivery mechanism, and information for translation is all performed by RNA, then it warrants a fair point that RNA preceded proteins during the origin of life.

Harold White published an interesting conceptual paper in 1976 named "Coenzymes as Fossils of an Earlier Metabolic State" [19]. As the second line of evidence for an RNA world, he

noted that a sizeable portion of enzymes (then over fifty percent) require coenzymes for activity. Coenzymes are small molecules located in the catalytic site of enzymes that are mechanistically essential for activity. The ”curious fact” White noticed was that many coenzymes are nucleotides. NAD, NADP, FAD, Coenzyme A, and (the most common) ATP are all nucleotides. Furthermore, other coenzymes such as thiamin pyrophosphate, THF, and pyridoxal phosphate can be derived from nucleotides. Having an RNA world preceding modern life's system would neatly explain this observation. In essence, if the earliest organisms used catalytic RNAs (ribozymes) then they might have contained these modified nucleotides as part of its machinery. Once the more efficient proteins superceded ribozymes for function only the essential catalytic core remained as a ”molecular fossil”. Another possible sign of coenzymes as fossils of an RNA world is the suggestion that the imidazole side chain of histidine could have arisen from a nucleotide coenzyme, given the fact that histidines biosynthesis starts from a nucleotide.

After the discovery of catalytic RNAs (ribozymes) in the 1980s, there was renewed interest in the RNA world. In particular, the discovery that the *Tetrahymena* ribozyme contains a self-splicing exon enabled researchers to create models where an RNA world organism could have ”sex” or genetically recombine using ancient introns [20]. Since splicing is reversible, one could envision a scenario where two introns flanking an exon could splice themselves out, including the exon, and reinsert themselves into another RNA strand. This transposon activity would help produce new combinations of genes. One could also envision a genetic system where a linear or functionally nonsensical RNA strand containing introns could then be made functional by their self-splicing. Then the reinsertion of the intron would enable genetic copying (this would explain away the problem of how to copy folded ribozymes). Indeed, White speculated that tRNAs contain an evolutionary marker of this mechanism since its secondary structure is disrupted by an intron.

The final line of evidence for an RNA world is that RNA seems to precede DNA on the biosynthetic timeline. All DNAs are derived from RNA reduction. That is, RNA is made first and

then DNA is made afterwards. The nucleotide monomer, ATP is synthesized first before being reduced to dATP. A similar pathway is taken by the three other NTPs. RNA preceding DNA on the biosynthetic pathway would suggest that RNA appeared earlier in life [21].

1.3 Role of NTPs

Modern life requires nucleoside 5' triphosphates (NTPs) for polymerization. These chemically activated monomers are used by a polymerase to be ligated onto the 3' hydroxyl of the RNA polymer. This is energetically favorable as the triphosphate contains a high-energy phosphate ester [22]. Once ligated, pyrophosphate becomes the leaving group and a monophosphate remains as the 5' OH to 3' OH backbone. The core function of an RNA world organism (as with any organism) would be replication in the form of RNA polymerization. NTPs would be the ideal substrate for polymerization for a multitude of reasons [23].

First, since modern life already uses NTPs, an RNA world organism using NTPs would remove one more evolutionary step required to explain the transition from the earliest life to modern life. Second, NTPs, although high energy, are far more stable in aqueous solution than the closest alternative, nucleoside 5' phosphoramidates. Phosphoramidates, primarily 2-methyl imidazolide, while being convenient substrates, are not prebiotically plausible for its own reasons, mainly that the intermediates for synthesis require neat formamide, they're not stable in aqueous solution, and they're inefficient substrates for nonenzymatic replication [24]. Third, NTPs can already be used by an RNA-world relevant RNA polymerase ribozyme. Fourth, nucleoside 2',3'-cyclic phosphates, although conveniently made by Sutherland syntheses, are energetically disfavored from polymerization because the equilibrium constant for dimer formation is on the order of 1.0 L/mol [25, 26]. If NTPs were the chemically activated monomers used for polymerization then a ribozyme might be able to use a prebiotically plausible pathway to catalyze its formation.

1.4 Role of trimetaphosphate

Trimetaphosphate (Tmp) is a cyclic triphosphate. It is one of the main polyphosphates generated through three prebiotically plausible routes. Firstly, Tmp can be generated by volcanic activity (Yamagata Namba 1991). Experiments by Yamagata et al demonstrate that volatization of P_4O_{10}, contained in phosphate rock, by magma followed by sudden cooling causes phosphate condensation from partial hydrolysis [27]. Subsequent analysis on the condensates revealed that trimetaphosphate and other polyphosphates survived the large temperature fluxes and hydrolysis. A second method for Tmp generation is via reaction of water with the mineral schreibersite ($(Fe,Ni)_3P$) yields phosphite and hypophosphite, which can then undergo a Fenton reaction to yield polyphosphates [28, 29]. Mechanistically, phosphate radicals are generated via OH radicals to create higher-order polyphosphates. A phosphate radical terminates with a phosphite radical to generate pyrophosphate. Subsequently, a pyroposphate radical can react with a phosphite radical to generate triphosphate. The third method to generate Tmp is via heating ammonium phosphate with urea in the presence of nucleosides [30]. A 19 percent yield was achieved when this reaction occurred at 100°C. Interestingly, less than 2 percent conversion happened in the absence of nucleosides. All three pathways are prebiotically plausible and would establish Tmp as one of the more abundant polyphosphates. An advantage of trimetaphospahte as a phosphorylating agent is

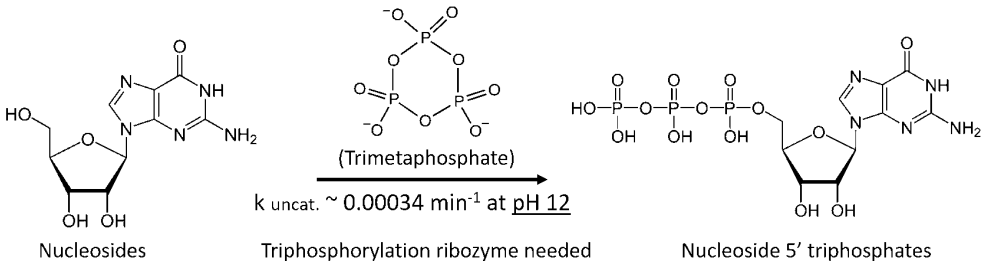


Figure 1.1: A schematic on how to generate chemically activated nucleotides. Trimetaphosphate (Tmp), a prebiotically plausible high-energy phosphorylating agent, is capable of triphosphorylating free nucleosides to generate nucleoside 5' triphosphates (NTPs). NTPs are the main energy currency of modern life and would have been the most likely substrate used in RNA polymerization during the RNA world.

that it's the most reactive polyphosphate. A landmark paper by Etaix and Orgel demonstrated that trimetaphosphate can be used as a phosphorylating source to generate NTPs directly from nucleosides. At 500 mM Tmp, 50 mM Adenosine, and 25°C 5 percent ATP was generated after 200 hours [31]. However, there are three main issues with this NTP synthetic pathway. The

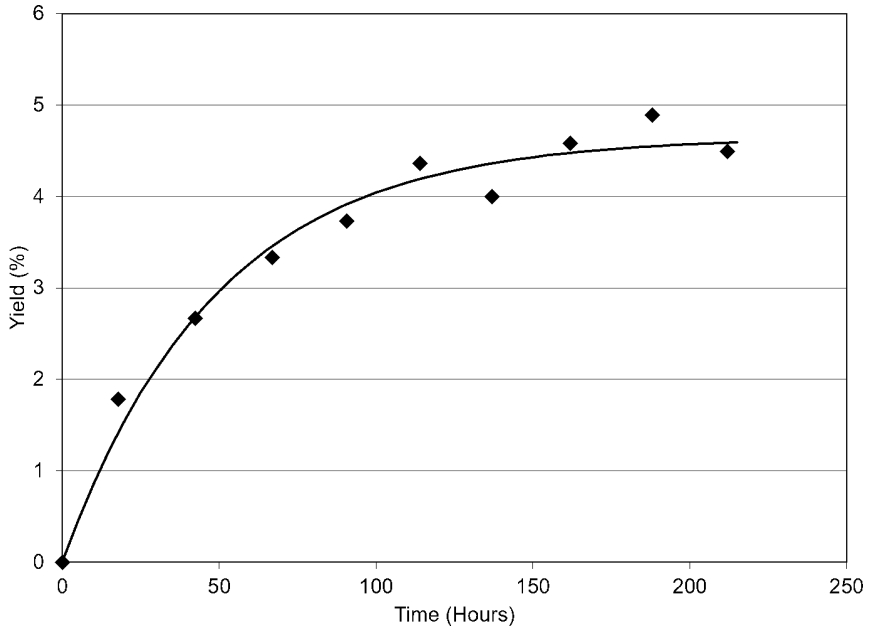


Figure 1.2: Single-order kinetics of NTP generation using the Etaix Orgel method. The yield is a maximum of five percent with a rate of 0.02 hour^{-1}. The maximum is achieved after roughly 200 hours. This figure is a recreaction from Etaix and Orgel 1978 as presented in Moretti and Muller 2014 [31, 32].

largest obstacle to this being relevant in an RNA world is the high pH required for the reaction to take place. At pH 12 any RNA polymers in the vicinity of the reaction would hydrolyze within minutes. The second fundamental issue is that it is quite slow even at pH 12, reaching a rate of 0.00034 min^{-1}. The third issue is that it is not stereochemically selective. Indeed, incubation of Tmp with Adenosine primarily yields 2' and 3' adenosine triphosphate, which readily hydrolyzes to monophosphates. Despite these rather daunting obstacles, the mere fact that this reaction can take place establishes a starting point for a RNA-world relevant pathway towards chemically activated nucleotides.

1.5 In vitro selections and TPRs

Biochemically validating the plausibility of the RNA World Hypothesis requires recapitulating an RNA organism in lab. This would mean creating an organism that uses RNA as its genetic code and also uses that same RNA for catalytic roles. Generating artificial ribozymes that perform RNA-world relevant functions has been going on since the 1990s. The largest breakthroughs have been the generation of a ligase ribozyme followed by a polymerase ribozyme [33, 34]. Subsequent evolution of the polymerase ribozyme has generated interesting variants, such as a reverse transcriptase ribozyme and a polymerase capable of RNA-PCR [35, 36]. All

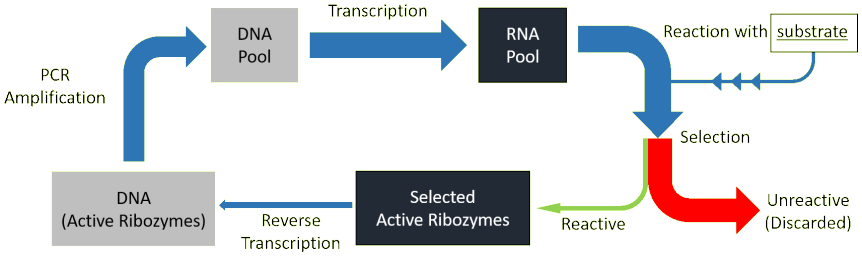


Figure 1.3: A general schematic of the steps in an in vitro selection. Starting from the top left and moving clockwise, a DNA pool containing randomized sequences is transcribed to RNA. Depending on the sequence, certain RNA molecules will fold into a structure that grants a specific activity. The RNA pool is incubated with a substrate and those active RNA molecules react with it. Some form of selection is designed that parses out active from inactive RNAs. The remaining active selected RNAs are reverse transcribed to DNA, enriching the pool for activity and starting the next round of selection.

these ribozymes were generated using a method known as *in vitro selection*. This method exploits natural selection by generating a random library of sequences and subjecting the library to a selection pressure. With careful design, a selection can be made where the sequences that survive the selection pressure are enriched for the next round. The end result is a highly-enriched library containing active sequences for the specific function selected. *In vitro selection* has been used to generate a multitude of functional sequences, such as a redox ribozyme, aptamers, and trans-splicing ribozymes [37, 38, 39]. Relevant to triphosphorylation, Moretti and Muller selected for a ribozyme that can triphosphorylate its own 5' hydroxyl using trimetaphosphate (TPRs).

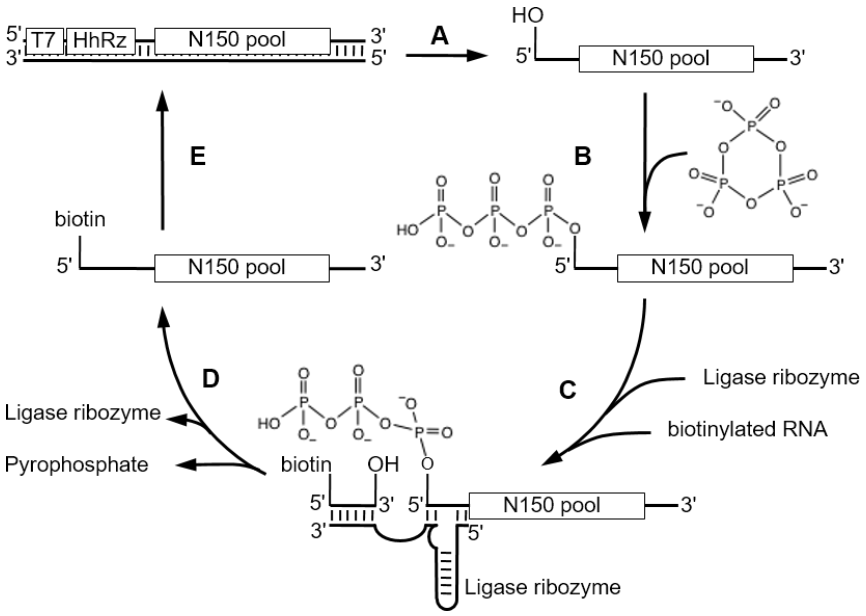


Figure 1.4: The in vitro selection of a triphosphorylation ribozyme. (A) A DNA pool is constructed containing an upstream hammerhead ribozyme sequence and a T7 promoter that enables transcription. Upon transcription, the RNA pool is automatically processed such that the 5' terminus of the randomized sequences contain a hydroxyl group. (B) This processed pool is incubated with trimetaphosphate. A few random RNAs that, by chance of their unique sequence, are active will be able to triphosphorylate their own 5' terminus, generating chemically activated RNA polymers. (C) These triphosphorylated RNA polymers are ligated to an anchor sequence containing biotin using a ligase ribozyme. Note that inactive RNAs cannot undergo this step as the ligase ribozyme requires a 5' triphosphate on the downstream RNA polymer. (D) The active RNA polymers can now be separated from inactive RNAs via streptavidin capture. (E) The recovered RNAs are reverse-transcribed to DNA and PCR amplified, starting another round of selection. Figure is a recreation from Moretti and Muller 2014.

The product is a chemically activated RNA polymer. The properties of the ribozyme were that it was 96 nucleotides, optimized at pH 8.3, a temperature of 40C, and had a rate $\tilde{1}0^7$-fold over the background rate. This selection was pivotal in that it demonstrated that triphosphorylation with trimetaphosphate could be catalyzed by RNAs. Although 5'-triphosphate RNA polymers were probably not the substrate used for polymerization it was an important foundational achievement for having a RNA-world plausible energy metabolism [32].

Going off of TPRs, the next step is to generate ribozymes, using *in vitro selection*, that can catalyze the triphosphorylation of nucleosides using Tmp. Such a ribozyme would be called a nucleoside triphosphorylation ribozyme or NTR. The motivation for such a ribozyme is that it

would clearly establish a RNA-world energy metabolism using a prebiotically plausible phosphorylating agent. Furthermore, coupling such a ribozyme with the polymerase ribozyme would generate a neat scenario where, with only two RNA polymers, one could go from unactivated nucleosides to NTPs to a copied genome.

Chapter 2

A Faster Triphosphorylation Ribozyme

2.1 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. One ribozyme (TPR1) was analyzed in more detail. TPR1 catalyzes the triphosphorylation reaction to a rate of 0.013 min^{-1} under selection conditions (50 mM Tmp, 100 mM $MgCl_2$, 22°C). To identify a triphosphorylation ribozyme that catalyzes faster triphosphorylation, and possibly learn about its secondary structure TPR1 was subjected to 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 $MgCl_2$, 40°C), the triphosphorylation rate of TRP1e reaches 6.8 min^{-1}.

2.2 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 [11, 12, 13]. This hypothesis has gained support from several directions, including the findings that RNA molecules can indeed catalyze reactions [16], that the ribosome is a catalytic RNA [40], and that deoxyribonucleotides are synthesized in cells from ribonucleotide precursors [41]. 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 inception of in vitro selection [42, 43] and the demonstration that novel catalytic RNAs can be isolated from randomized RNA pools [33], several types of ribozymes with possible involvement in an RNA world organism have been isolated [44], most importantly a ribozyme that catalyzes RNA-dependent RNA polymerization [34]. The substrates for this ribozyme are nucleoside 5'-triphosphates; these nucleotides contain the perhaps most prebiotically plausible chemical activation group [14]. Nucleoside 5'-triphosphates can be generated from nucleosides and trimetaphosphate (Tmp) [31]. Tmp likely existed prebiotically, as suggested from the finding of large amounts of its chemical precursor in 3.5 billion year old marine sediments [29, 28]. However, the uncatalyzed triphosphorylation of nucleosides with Tmp occurs efficiently only at pH values above 12 [31], 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 [32].

The only triphosphorylation ribozyme that was previously characterized in some detail, TPR1 (Fig. 1A), displays single-exponential reaction kinetics with a rate of 0.013 min^{-1} at selection conditions (50 mM trimetaphosphate and 100 mM $MgCl_2$ at pH 8.3) [32]. Here we searched for sequence variants of TPR1 that display faster reaction kinetics using in vitro selection.

The most active ribozyme identified in this study, TPR1e, contains seven activity-enhancing mutations (Fig. 1B) and reacts 24-fold faster than TPR1 under selection conditions. Under optimal conditions (150 mM Tmp, 650 mM $MgCl_2$, 40°C), the triphosphorylation rate of TPR1e reaches 6.8 min^{-1}. An additional new duplex was identified that constrains the ribozyme structure.

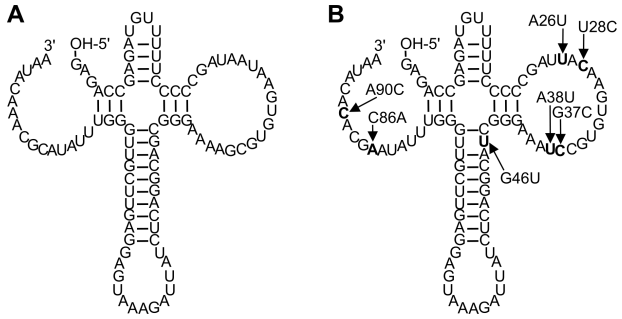


Figure 2.1: Proposed secondary structures for (A) the parent ribozyme, TPR1, and (B) the most active ribozyme identified in this study, TPR1e. TPR1e contains seven beneficial mutations relative to TPR1 (mutations are in the figure).

2.3 Results

To identify a more active sequence variant of the previously isolated triphosphorylation ribozyme, TPR1 (Fig. 1A), a 'doped selection' was performed. 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 a frequency of 8%. With a total of 82 nucleotides being partially randomized, around 103 wild-type sequences were contained in a library of 10^{13} sequences (0.76^{82} x 10^{13} ~1,700). The selection was performed as described previously [32] and started with an effective complexity of 7.0 x 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 [32]. To isolate the most efficient catalysts, higher selection pressure was used in additional

rounds of selection by changing the Tmp incubation conditions (shorter incubation time, lower Tmp concentration, or higher incubation temperature; S1 Fig.). After three or four rounds of selection, 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).

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 self-triphosphorylation rate of 0.21 ± 0.02 min^{-1} under selection conditions (50 mM Tmp, 100 mM $MgCl_2$, 50 mM Tris/HCl pH 8.3) (Fig. 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 (S2 Fig.). Only five of the sixteen mutations were necessary for increased activity (U28C, G37C, A38U, C86A, A90C). This ribozyme was called TPR1_II. The reduction to five mutations was possible by the finding (see materials and methods) that four of the eleven nonessential mutations appeared to elongate and stabilize the long central P5 duplex of the ribozyme (U55C, A56U, U58C, U64G). These elongating mutations were only necessary in the context of two destabilizing mutations within the P5 duplex of clone 11 (A51U, C52G). All clone 11 mutations within the P5 stem-loop could be deleted while maintaining (and even increasing) catalytic activity. TPR1_II showed a self-triphosphorylation rate of 0.25 ± 0.03 min^{-1} under selection conditions (Fig. 2 and S2 Fig.).

To further increase triphosphorylation activity of TPR1_II the 58 clones that were at least as active as TPR1 (S3 Fig.) were analyzed for statistically enriched additional mutations. Mutations were deemed enriched when they appeared with a frequency of more than 24% (a doping ratio of 24% was used for the pool design) and if they appeared preferably among the 20 most efficient ribozymes. Six such mutations were identified. Each of these mutations

was individually inserted into TPR1_II. 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 self-triphosphorylation rate was 0.31 $\pm$ 0.02 min^{-1} under selection conditions, 24-fold faster than TPR1 (Fig. 2). Satisfyingly, the kinetics of this ribozyme also showed the highest amplitude of 95%. This suggested that the fraction of misfolded ribozyme sub-populations was reduced from ˜15% (TPR1) to ˜8% (TPR1_II) and finally ˜5%. This variant was declared the 'winner' of the doped selection study. It contained seven mutations (A26U, U28C, G37C, A38U, G46U, C86A, A90C; Fig. 1B) and was termed TPR1e. The mutations

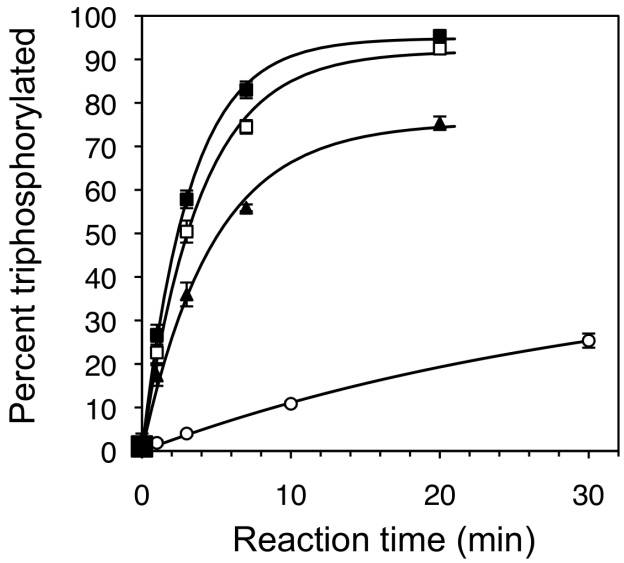


Figure 2.2: Triphosphorylation kinetics of central ribozyme variants in this study. The starting point of the doped selection was the ribozyme TPR1 (empty circles). It has a k_{obs} of 0.013 min^{-1} under selection conditions (100 mM $MgCl_2$, 50 mM trimetaphosphate, 50 mM Tris/HCl pH 8.3). The most 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 0.25 min^{-1}). Two mutations that arose independently were introduced to yield TPR1e (filled squares), a 7-mutation variant with a k_{obs} of 0.31 min^{-1}. Lines are single-exponential curve fits to the data. Error bars denote the standard deviations from triplicate experiments.

selected in TPR1e led to the complementarity between eight bases in nucleotides 31-39 and eight bases in nucleotides 84-92, interrupted only by a single C:A pair (Fig. 3A). This suggested the existence of a previously unrecognized duplex, P4. To test whether this duplex formed, base

covariation experiments were conducted at six positions in the expected duplex, testing for base pairs G33/C90, U34/A89, G35/C88, C36/G87, U38/A85, and A39/U84 (Fig. 3B). The drop in activity of individual mutations G35C and C88G was rescued by the double-mutation suggesting that this portion of the P4 duplex is formed in the active ribozyme. No changes in activity were observed with the mutations of G33C and C90G and with mutations U34A and A89U, 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. Double mutations at the positions 36/87, 38/85, and 39/84 did not rescue the inhibitory effects of the single mutations, suggesting that the identity of these bases was important for activity. In summary, the base covariation experiments of the proposed P4 duplex were inconclusive because only one out of six positions showed the characteristic deleterious effects of single mutations coupled to a rescue effect of the double mutation.

To interrogate the proposed duplex with a different method, SHAPE probing experiments with 1-methyl-7-nitroisatoic acid (1M7) were conducted [45]. This compound specifically reacts with 2'-hydroxyl groups of flexible nucleotides such that positions of low reactivity correlate with the presence of double strands. Both TPR1 and TPR1e were subjected to the same analysis. The accessibility data between nucleotides U17 and U46 were consistent with the position of the P2, P3, and P5 helix. Additionally the protected region from G31 to U38 suggested the existence of the P4 duplex (Fig. 3C). While positions U38/A85 and A39/U84 would be complementary to stack upon the P4 duplex these base pairs do not appear to form in TPR1 or TPR1e (Fig. 3C). Importantly, the protection of nucleotides A86 to C92 confirmed the existence of the duplex (Fig. 3D). The P1 duplex that appears to exist in TPR1 (Fig. 3D, position 78-80, [32]) does not appear to exist in TPR1e. Interestingly, two of the most important activity-enhancing mutations identified by the doped selection (G37C and C86A) led to a C:A pair at the end of the proposed P4 duplex. In summary, the chemical probing experiments by 1M7 confirm the existence of the P4 duplex, are consistent with previous 1M7 probing of TPR1, and suggest distinct structural

changes between TPR1 and TPR1e. The newly identified P4 duplex generates strong structural constraints and suggests a very compact three-dimensional structure for the ribozyme.

The parent ribozyme TPR1 did not show detectable triphosphorylation of free nucleosides [32]. To test whether TPR1e was able to triphosphorylate free nucleosides, 14C labeled guanosine was incubated with eight different variants of TPR1e. Unfortunately, no triphosphorylation of free nucleoside was detected with any of eight variants of TPR1e tested (data not shown). Eight ribozyme variants were tested because the ribozyme 5'-terminus could have hindered binding of the free nucleoside. In addition to the full-length TPR1e, five variants had their 5'-terminus truncated by 1, 2, 3, 6, or 18 nucleotides; all of these ribozymes contained 5'-hydroxyl groups. Additionally, the ribozyme truncated by 1 nucleotide was also tested with a 5'-phosphate and a 5'-triphosphate. So far no ribozyme has been identified that can triphosphorylate free nucleosides.

To identify the optimal reaction conditions for the evolved ribozyme (TPR1e), the influence of temperature, Tmp concentration, and Mg^{2+} concentration were analyzed sequentially. First, the temperature was varied between 5°C and 60°C, and an optimum at 40°C was determined (Fig. 4A). At 40°C, the optimal Tmp concentration was at 150 mM (Fig. 4B), and the optimal concentration of free Mg^{2+} at 150mM Tmp was at 500 mM (Fig. 4C). Interestingly, the dependence on the Tmp concentration at these high Mg^{2+} conditions and high temperature showed a cooperative effect, with a significantly higher rate at 100 mM Tmp (5.7 min^{-1}) than what would result from doubling the rate at 50 mM Tmp (2 0.96 min^{-1} = 1.92 min^{-1}). This suggested that a second molecule of trimetaphosphate binds to the ribozyme and/or the first trimetaphosphate molecule and increases the reaction rate. This effect was not observed with the parent ribozyme TPR1 [32]. The mechanism of the cooperative effect is currently unclear. Notably, the effect did not appear at more modest Mg^{2+} concentrations (~50 mM free Mg^{2+} versus 400 mM free Mg^{2+}) and lower temperature (22°C vs. 40°C), where a linear correlation between Tmp concentration and reaction rate was observed (Fig 5 and [32]). At the optimal conditions (40°C, 150 mM Tmp, 650 mM total $MgCl_2$, 50 mM Tris/HCl pH 8.3) the cooperative effect afforded a reaction rate of 6.8 min^{-1}.

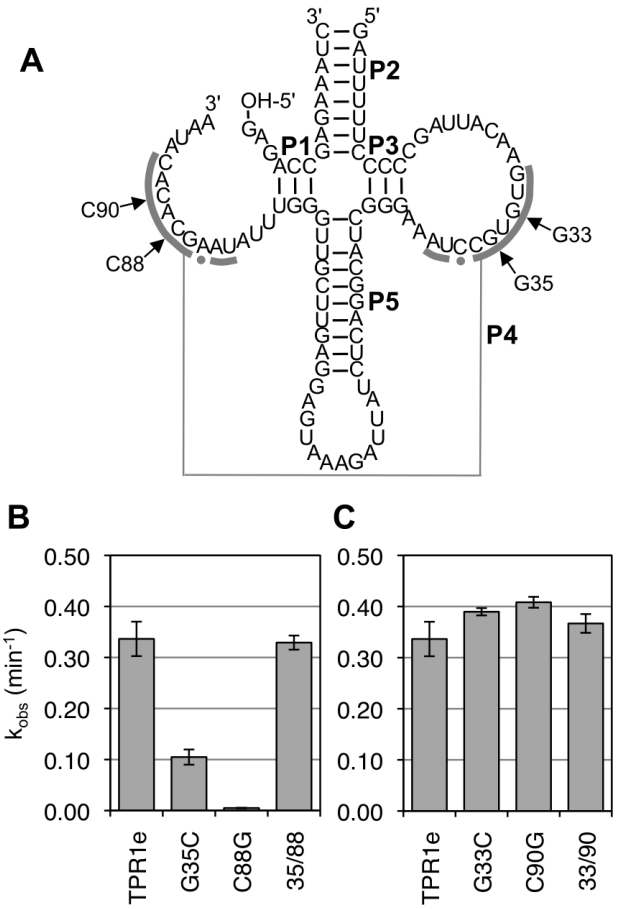


Figure 2.3: Experiments testing the formation of duplex P4. (A) Proposed secondary structure for ribozyme TPR1e showing proposed paired regions labeled P1-P5. All duplexes with exception of P4 were identified previously [32]. Nucleotides found accessible by Shape probing are colored in red, while protected nucleotides are shown in blue. C-A pairs are shown with a dot. The numbering scheme corresponds to the cis-acting ribozyme (Fig. 1). Note that the duplexes P1 and P2 are formed in trans because the experiments were performed using the trans reaction. Base pairs for duplex P1 are shown in grey because for TPR1e, nucleotides 78-80 appeared accessible by Shape probing. (B) Apparent triphosphorylation rates of the mutated ribozymes in the base covariation experiment of twelve bases in the proposed P4 duplex. The six column graphs are in the same order as the proposed based pairs in P4 of figure 3A. Each graph shows the triphosphorylation rate of TPR1e on the left, the two single mutants in the middle, and the double mutant on the right. Error bars are standard deviations of triplicate experiments. (C) Shape probing experiments for nucleotides 17 to 46 of the ribozymes TPR1 (empty circles) and TPR1e (colored circles). The normalized Shape signal is shown as a function of nucleotide position. A threshold value (grey, dashed line) is used to distinguish between low Shape reactivity (blue) and high Shape reactivity (red). The colors match the color-coding in figure 3A. The position of duplexes P1-P5 in TPR1e is indicated by thick, solid, grey, horizontal lines. (D) As in (C) but for nucleotides 57 to 94 in the ribozymes.

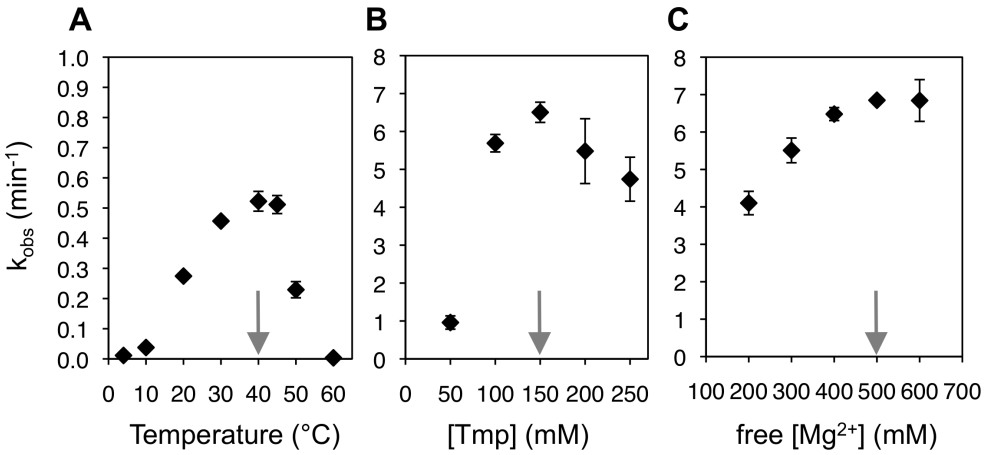


Figure 2.4: Determination of optimal TPR1e triphosphorylation conditions. (A) Observed triphosphorylation rate as function of the temperature, at 50 mM Tmp, 100 mM $MgCl_2$, and 50 mM Tris/HCl pH 8.3 (B) Influence of the trimetaphosphate concentration on the observed reaction kinetics at 40°C, and with an excess of 400 mM $MgCl_2$ over Tmp. (C) Influence of the free Mg^{2+} concentration on the triphosphorylation rate at 40°C and 150 mM Tmp. The free Mg^{2+} concentration is the total Mg^{2+} concentration minus the concentration of Tmp because each Tmp appears to be coordinated by one Mg^{2+} at these concentrations and pH 8.3 [32]. 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.

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 (Fig. 5A). The concentration of 1 mM Tmp was considered to be in the upper range of prebiotically plausible concentrations [15]. 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 CO32- [46, 47]. The triphosphorylation rates under these conditions were 0.0014 min^{-1} at 22°C and 0.004 min^{-1} at 40°C. The rate in synthetic seawater, which contains 54 mM $MgCl_2$, was 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 (Fig. 5B). The contribution of sodium chloride to this inhibition (470 mM in synthetic seawater) was about 2-fold (dashed line in Fig. 5C). This showed that TPR1e is inhibited by several components of synthetic seawater, and that the evolved ribozyme TPR1e is

able to triphosphorylate with a rate of 0.004 min^{-1} under prebiotically plausible conditions.

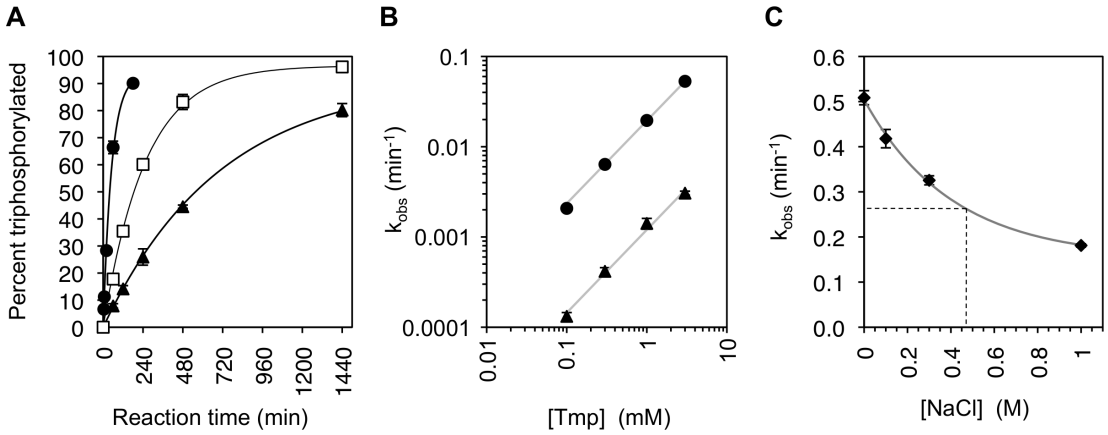


Figure 2.5: Triphosphorylation kinetics of TPR1e. (A) At a Tmp concentration of 1 mM, the triphosphorylation kinetics are shown for synthetic seawater at 22°C (black triangles, k_{obs} = 0.0014 min^{-1}, max = 93%), and at 40°C (empty squares, k_{obs} = 0.0039 min^{-1}, max = 97%). For comparison, the reaction kinetics are shown for 54 mM $MgCl_2$ in Tris/HCl pH 8.3 at 22°C (black circles, k_{obs} = 0.020 min^{-1}, max = 93%). This latter condition lacks all seawater components with exception of Mg^{2+}. 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 $MgCl_2$ with 50 mM Tris/HCl pH 8.3 (black 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 $MgCl_2$. 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.

2.4 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. A previously unidentified duplex was identified in this ribozyme, indicating a very compact structure. TPR1e shows triphosphorylation kinetics of 6.8 min^{-1} under optimal conditions. In synthetic seawater with 1 mM Tmp at 40°C, a triphosphorylation rate of 0.004 min^{-1} was measured.

The concentration of 1 mM Tmp appears to be in a prebiotically plausible range because the prebiotic steady-state concentration of polyphosphates was estimated to have been 10 nM - 10 M in the mixed zone of the ocean, with localized concentrations across the surface of the Earth in excess of 1 mM [29]. The detection of TPR1-catalyzed triphosphorylation in synthetic seawater with 1 mM Tmp is encouraging; however, the specific ribozyme presented here would most likely not have been useful to RNA world organisms for two reasons. First, TPR1e did not show detectable triphosphorylation of free nucleosides; an activity that would have generated nucleoside 5'-triphosphates that can be used in RNA polymerization [34]. Second, while TPR1e can act in trans on short RNA oligonucleotides (Fig. 3), it did not mediate multiple turnover reactions (data not shown). Multiple-turnover triphosphorylations of short RNA oligonucleotides could have been sufficient to mediate RNA polymerization if RNA polymerization would have proceeded in 3'-5' direction instead of today's biological 5'-3' direction [44, 34]. Future developments may generate ribozymes that catalyze multiple turnover triphosphorylation of RNA oligonucleotides and / or the triphosphorylation of free nucleosides.

The results of the doped selection can be helpful for the design of other doped selection experiments. The doping ratio used in this work (24% randomization of the wild type sequence) together with the complexity of the pool (7×10^{13}) showed that the sequence space of single, double, and triple mutants of the parent ribozyme (TPR1) was sampled completely, and quadruple mutants were sampled to ~77% ($P=1-(1-0.7678 \times 0.084)^{(7 \times 10^{13})} = 0.77$). This doping ratio was chosen in an attempt to maximize sequence coverage without losing the ribozyme pools catalytic activity, but two observations from this work suggest that a lower doping rate would likely have been more beneficial. First, the most active ribozyme from the selection contained several mutations that destabilized stem P5 and required additional mutations to overcome the destabilizing effects. 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; S4 Fig.). Collectively, the presence of inhibiting

mutations in even the most active clones and the preferential selection of ribozymes with fewer total mutations indicates that a lower doping ratio would likely have been more beneficial.

2.5 Materials and Methods

2.5.1 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 [32]. 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 achieve 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 [32], and were added as 5'-primer during PCR.

2.5.2 Selection

The initial pool for the selection contained an effective population size of 7 x 10^{13} unique sequences. The protocol of the selection was identical to our previous selection [32]. 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 (S1 Fig.). 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 line 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 (decreased Tmp concentration).

2.5.3 Kinetic analysis of cis-triphosphorylation reactions

The ribozyme clones were prepared as described previously [32]. PCR products from 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 alpha[32P]-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 A260. All isolates were initially tested with the additional twelve 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 twelve 3'-terminal nucleotides were removed.

The activity assay was performed as described previously [32]. In short, 8 M ribozyme was incubated at 22°C 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 Na2EDTA. DNAzyme reactions to separate the eight 5'-terminal nucleotides of the ribozyme were initiated by addition of DNAzyme and a 24-nucleotide long DNA oligonucleotide, both at a final concentration of 1.6M and two-fold above the ribozyme concentration. The 24-nucleotide long DNA oligonucleotide annealed immediately downstream of the DNAzyme to the triphosphorylation ribozyme and assisted 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 DNA 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 (1h/37°C). Aliquots of the DNAzyme reaction were quenched in formamide and Na2EDTA. Samples were 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. Triplicate experiments used one, two, or three different preparations of the ribozyme. The variability between different ribozyme preparations was smaller than the day-to-day variability of the experiments and therefore did not significantly affect the errors.

2.5.4 Kinetic analysis of trans-triphosphorylation reactions

All experiments with TPR1e used the trans-triphosphorylation protocol and performed essentially as shown [32]. In the trans-triphosphorylation experiment, an internally 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 [32]. 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 alpha-32P-ATP. T7 transcripts were purified as described for the ribozymes above. Trace amounts of substrate were used in each reaction.

The triphosphorylation experiment was set up similar to the cis- reaction, where ribozyme and substrate were mixed with Tmp, $MgCl_2$, 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.

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

Two strategies were used for generating the different ribozyme constructs, PCR mutagenesis and quickchange site-directed mutagenesis [48]. Mutations reverting the sequence of clone 11 to the wild type sequence were introduced by 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.

To reduce the number of mutations in TPR1 clone 11 to a minimum, we hypothesized that two mutations (A51U, C52G) destabilized the P5 stem, and that this effect was compensated by four additional mutations in the P5 loop (U55C, A56U, U58C, U64G) that appeared to extend the P5 stem. Indeed, when both types of mutations were removed from clone 11 (in addition to A63C and U76C), the activity of the resulting ribozyme (TPR1_II) matched and even exceeded that of clone 11.

2.5.6 SHAPE probing experiments

The Shape probing experiments were conducted essentially as described [32, 45]. Briefly, 1 M trans-acting ribozyme was incubated with 1.5 M of its 14-nucleotide substrate and 2 mM 1M7 [49] in a solution of 100 mM $MgCl_2$, 50 mM trisodium Tmp and 50 mM HEPES/NaOH pH 8.0 for 3 minutes. 1M7 was added as a 20 mM solution in DMSO, leading to a final DMSO concentration of 10% (v/v). Products were ethanol precipitated and reverse transcribed with 5'-[32P] radiolabeled primers and Superscript III reverse transcriptase (Invitrogen). The region from U17 to U46 was analyzed with a reverse transcription primer complementary to nucleotides G68 - C48. For analysis of the region U57 and U94 the ribozymes were extended at their 3'-terminus such that a 12-nucleotide reverse transcription primer could facilitate analysis of the sequence close to the ribozyme 3'-terminus. Products were separated by denaturing 15% PAGE, exposed to phosphorimager screens, scanned, and quantitated with the software Quantity One. Subtraction of the band intensities from a reaction containing DMSO without 1M7 from the intensities of the reaction containing 20 mM 1M7 in DMSO gave the SHAPE reactivities at each position. Because the reverse transcription efficiency of TPR1 was reproducibly lower than that of TPR1e the SHAPE signals were normalized to the total radioactivity in the analyzed region as shown in Figs 3C and 3D.

2.6 Supporting Information

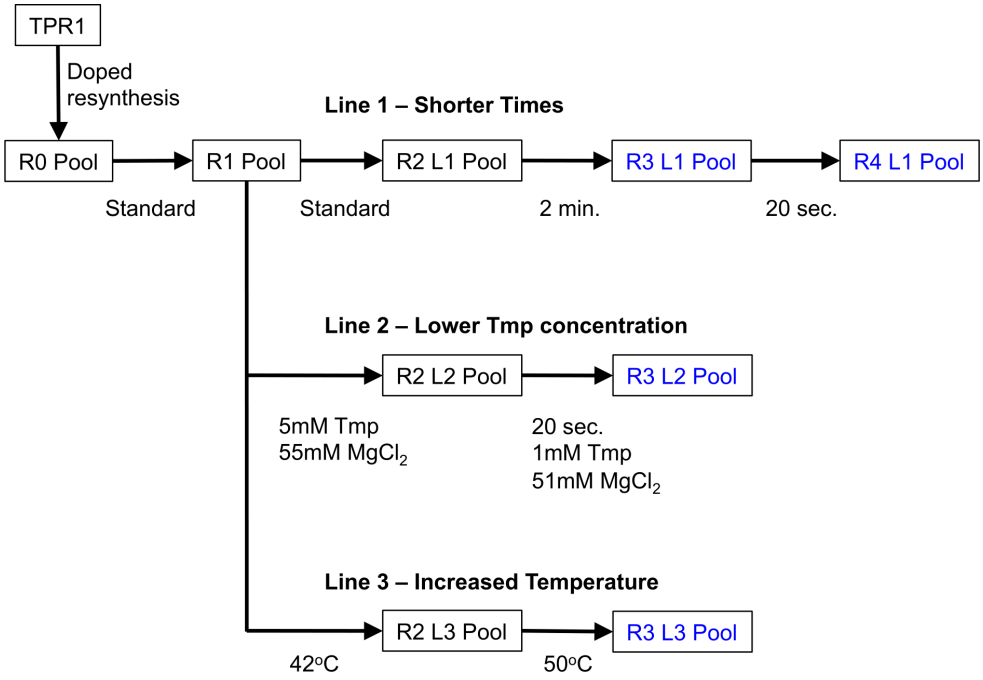


Figure 2.6: Diagram of doped selection performed. Boxes represent pools of RNA and horizontal arrows represent rounds of selection. From the starting pool of the doped selection (R0 pool, left), round 1 was performed using standard reaction conditions. 'Standard' describes the incubation conditions that were used in the previous selection, with an incubation time of 3 hours, an incubation temperature of 22°C, a ribozyme concentration of 100 nM, a Tmp concentration of 50 mM, a $MgCl_2$ concentration of 100 mM, and 50 mM 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 Mg^{2+} concentration (5mM TmP / 55mM $MgCl_2$, 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 labels indicate pools where individual ribozymes were cloned, isolated, and tested for activity.

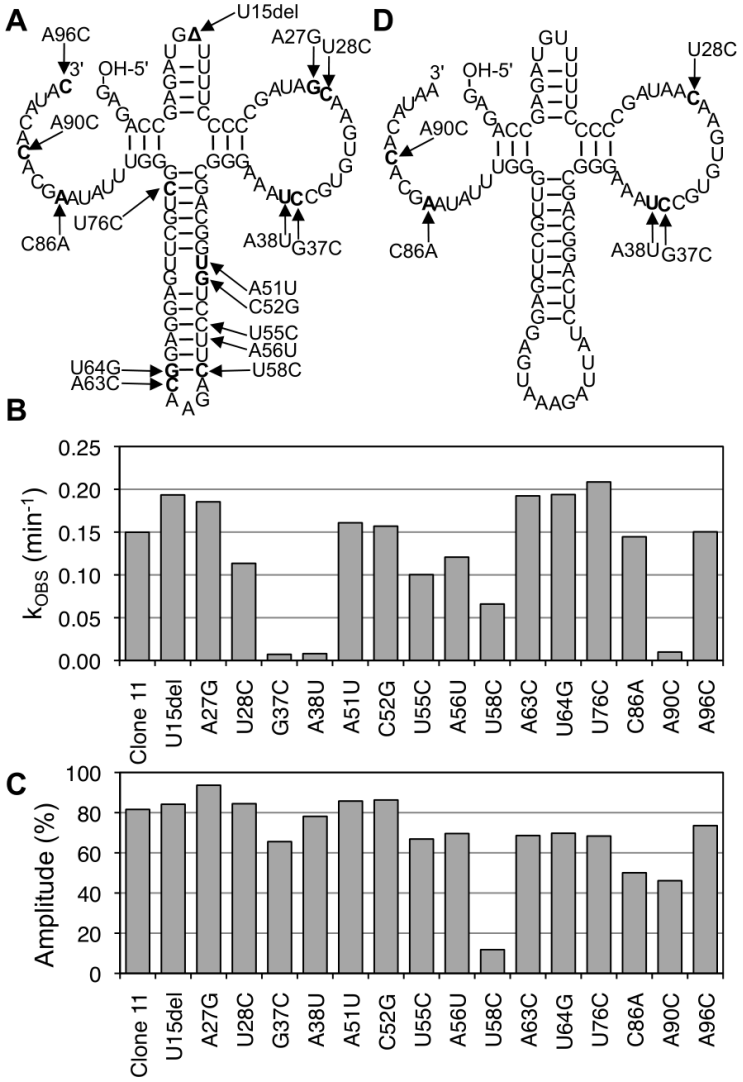


Figure 2.7: Identification of mutations in isolated ribozyme clone 11 that are necessary and sufficient for full activity. Mutated nucleotides are labeled and shown in bold. (A) Proposed secondary structure for clone 11, which contains 16 mutations relative to the parent ribozyme TPR1 (labeled with arrows). (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. The shown results are from single measurements; experimental errors were in the range of 0.02 min/textsuperscript-1. (C) Observed amplitudes of the single-exponential reaction kinetics of the same clones. The shown results are from single measurements; experimental errors were in the range of 3%. (D) Proposed secondary structure of ribozyme variant TPR1_II, which contains the five necessary mutations of clone 11.

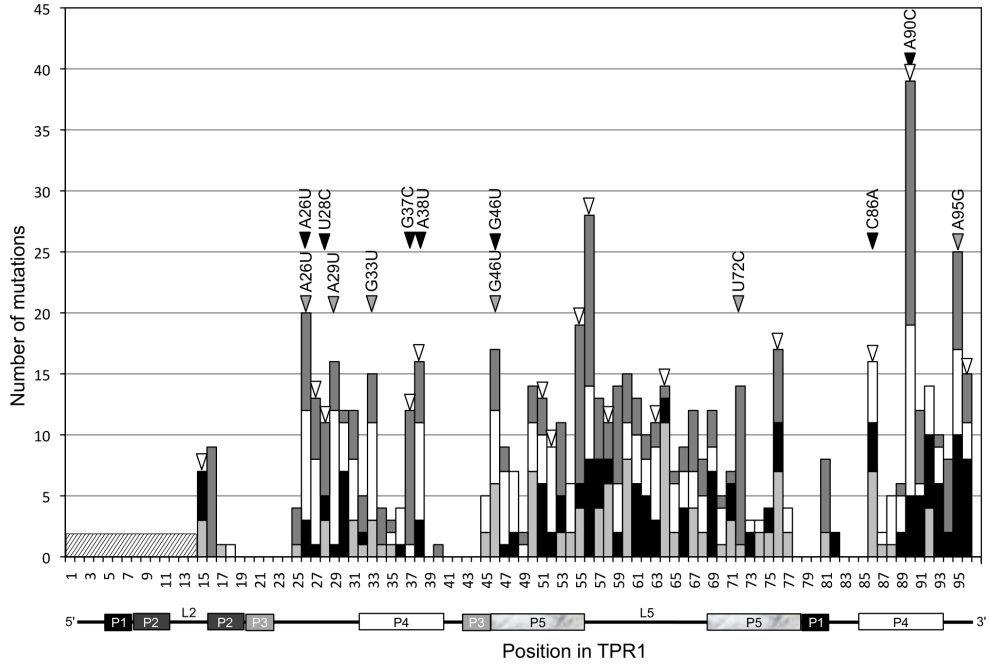


Figure 2.8: Position of mutations in the 58 clones isolated from the selection that had at least the same activity as the parent ribozyme TPR1. The number on the x-axis denotes the nucleotide position in the ribozyme. For position 1-14 (shaded box) no mutation data are available because this was the primer binding site during the selection. 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.

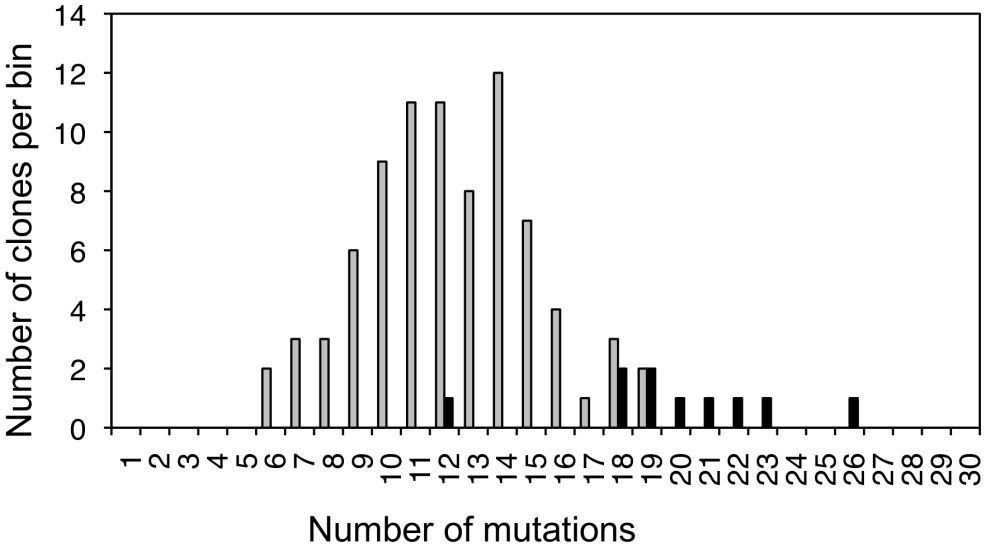


Figure 2.9: Mutational load in the pool before and after selection. The clones identified in the doped selection were binned according to their number of mutations (including deletions). The number of mutations is given on the x-axis; the number of clones per bin is given on the y-axis. Clones isolated from the pool before selection are represented in black columns. Clones isolated after selection are shown in grey columns. A total of 10 clones were analyzed from the pre-selected pool; a total of 82 clones was analyzed after the selection. The average number of mutations in clones before the selection is 19.8; after the selection it is 12.2.

2.7 Acknowledgements

We thank Gerald Joyce and David Horning for helpful discussions.

2.8 Funding

The material is based upon work supported by the National Aeronautics and Space Administration under Grant No. NNX13AJ09G issued through the Science Mission Directorate (ROSES-2011), in Astrobiology/Exobiology to U.F.M.

Chapter 2, in full, is a reprint of the material as it appears in PLOS ONE. Dolan, Greg F.; Akoopie, Arvin; Müller, Ulrich F. 2015. The dissertation author was the second author of this paper. He contributed by characterizing the secondary structure of the ribozyme and performed base covariation experiments.

Chapter 3

Lower temperature optimum of a short, fragmented triphosphorylation ribozyme

3.1 Abstract

The RNA world hypothesis describes a stage in the early evolution of life in which catalytic RNAs mediated the replication of RNA world organisms. One challenge to this hypothesis is that most existing ribozymes are much longer than what may be expected to originate from prebiotically plausible methods, or from the polymerization by currently existing polymerase ribozymes. We previously developed a 96-nucleotide long ribozyme, which generates a chemically activated 5'-phosphate (a 5'-triphosphate) from a prebiotically plausible molecule, trimetaphosphate, and an RNA 5'-hydroxyl group. Analogous ribozymes may have been important in the RNA world to access an energy source for the earliest life forms. Here we reduce the length of this ribozyme by fragmenting the ribozyme into multiple RNA strands, and by successively removing its longest double strand. The resulting ribozyme is composed of RNA fragments with none longer than 34 nucleotides. The temperature optimum was ~20°C, compared to ~40°C for the parent ribozyme. This shift in temperature dependence may be a more general phenomenon for

fragmented ribozymes, and may have helped RNA world organisms to emerge at low temperature.

3.2 Introduction

The RNA world is an early, hypothetical stage of life in which RNA polymers served as both the genome and the only genome-encoded catalyst [12, 11, 13, 20]. This RNA world hypothesis is supported by the ability to explain how an interdependent DNA-RNA-protein system could have originated from a prebiotic environment [20], by the identification of 'molecular fossils' in extant life forms [50, 40, 19, 41], and by the in vitro selection of RNA molecules that can catalyze chemical reactions required for an RNA world [51, 44]. However, direct evidence of RNA world organisms may never be found because they existed more than 3.4 billion years ago, when apparently modern life forms left traces in the fossil record [52]. Instead, it may be possible to test how RNA world organisms could have functioned by trying to generate RNA world organisms in the lab, from prebiotically plausible compounds [52, 53, 54].

The length of RNAs required to sustain an RNA world organism can be an important limitation, for two reasons. First, the emergence of the first self-replicating RNA system requires that all RNAs are synthesized prebiotically, in the absence of catalytic RNAs. It is unclear how long prebiotically generated RNAs were but it is safe to assume that shorter RNAs were more abundant than longer RNAs. Many early studies on prebiotic RNA polymerization focused on the polymerization of 5'-imidazolide- and 5'- 2-methylimidazolide-activated nucleoside 5'-phosphates [55, 56] on montmorillonite clay catalysts [57], which generated polymers with up to 40 nucleotides in length. However, these activated monomers are unlikely to have existed on prebiotic Earth because their synthesis requires high-energy intermediates. The same problem exists for 1-methyladenine as an activation group, which also mediated elongations up to 40 nucleotides [58]. The polymerization of nucleoside 5'-phosphates is possible without activation groups under temperature cycling conditions in lipid matrix or concentrated salt solutions, which

can generate polymers up to ˜100 nucleotides [59, 60]. However, the products at least from the lipid-assisted synthesis are rich in abasic sites, which casts doubt on their usefulness for an RNA world organism [61]. Nucleoside 2',3'-cyclic phosphates appear prebiotically plausible [62], polymerize to more than 13-nucleotides [63], and may be the most promising building block for prebiotically plausible RNA polymerization [64]. The thermodynamic unfavorability of their polymerization in aqueous solution [26] may be surmountable [64] but the products also contain a mixture of 2',5'- and 3',5'- phosphordiester bonds [65]. The polymerization of 3',5'-cyclic GMP, and 3',5'-cyclic AMP leads to polymers with lengths up to 25 nucleotides [66, 67]. However, the prebiotic plausibility of 3',5'-NMPs is controversial [66]. It is not yet clear which of these or other chemistries could have generated an RNA world. All of the known methods show low yields for longer polymers, suggesting that the emergence of the first RNA world organism would have been more likely if shorter RNAs were sufficient to generate an RNA world organism. Second, for early RNA world organisms it may have been a challenge to evolve efficient RNA polymerase ribozymes. This idea is suggested by studies on in vitro selected ribozymes. A catalytic RNA that polymerizes RNA in a template-dependent fashion was generated by in vitro selection [34], and further in vitro selections yielded several improved variants [68, 69, 70]. However, this ribozyme strongly favors one particular 11-nucleotide sequence such that the longest polymerization product with a functional sequence was a 24-nucleotide extension [69]. These polymerase ribozyme variants are ˜190 nucleotides in length, therefore this class of polymerase ribozymes is currently far from self-replication. Recently, a ribozyme was developed that mediates the polymerization of 5'-triphosphorylated oligonucleotides and nucleosides of the opposite stereochemistry [71]. This cross-chiral approach effectively prevented product inhibition, and mediated the polymerization of an 83-nucleotide long, catalytically active RNA from NTPs and short, triphosphorylated RNA fragments. However, RNA fragments need to be polymerized themselves, and using short RNAs as substrates (instead of nucleotides) causes problems in copying fidelity [72]. Therefore, even for an early RNA world organism the polymerization of long RNAs may have been a challenge. This

idea suggests that not only the first instance but also the early stages of RNA world organisms would be more likely to emerge if its constituent ribozymes could assemble from short RNA polymers.

The fragmentation of ribozymes into multiple strands has been observed in nature and studied in the lab on both natural and in vitro selected ribozymes (see discussion). The present study tests the effects of fragmentation and size reduction on a triphosphorylation ribozyme - a ribozyme with importance to RNA world organisms [32, 73]. This ribozyme catalyzes the synthesis of chemically activated 5'-phosphate groups by reaction of RNA 5'-hydroxyl groups with the molecule trimetaphosphate (Tmp). Tmp is prebiotically plausible because it can be generated from prebiotically available phosphides via erosion and mild oxidation, and because intermediates in this pathway are also present in 3.4-billion year old marine sediments [28, 29]. Ribozyme-catalyzed triphosphorylation activity with Tmp is important for RNA world scenarios because it makes the free energy of Tmp hydrolysis available to RNA organisms, and the second law of thermodynamics dictates that any structure-forming entity (i.e. any life form) requires the inflow of chemical energy.

Here we show the size reduction and fragmentation of a triphosphorylation ribozyme with no RNA strand being longer than 34 nucleotides. This small and fragmented ribozyme displays reaction kinetics only 2-fold below that of the full-length ribozyme at 22°C, and its temperature optimum is shifted from ~40°C to ~20°C. The results are discussed with respect to the RNA world hypothesis.

3.3 Results

The starting construct used in this study had a length of 86 nucleotides and differed from the original isolate TPR1 [32] in six mutations (Fig. 1A). These six mutations increased triphosphorylation kinetics 24-fold to 0.31 min^{-1} at 50 mM Tmp, 100 mM $MgCl_2$, and 50

mM Tris/HCl pH 8.3 [73]; a seventh, slightly beneficial mutation was not included because it destabilizes the P5 stem, which is the focus of this investigation. The ribozyme binds its 14-nucleotide substrate strand via the P1 and P2 helix, and forms a pseudoknotted structure with a sum of five helices. The helix P5 contributes a large loop and an 11-base pair long double-strand to the ribozyme. To reduce the length of required RNAs we split the ribozyme at the L5 loop and successively truncated the length of the P5 helix. The intermediates of this fragmentation-and-size reduction were studied in their ability to catalyze the triphosphorylation of the 5'-hydroxyl group at the 14-nucleotide substrate. When the ribozyme was split at the L5 loop the two individual

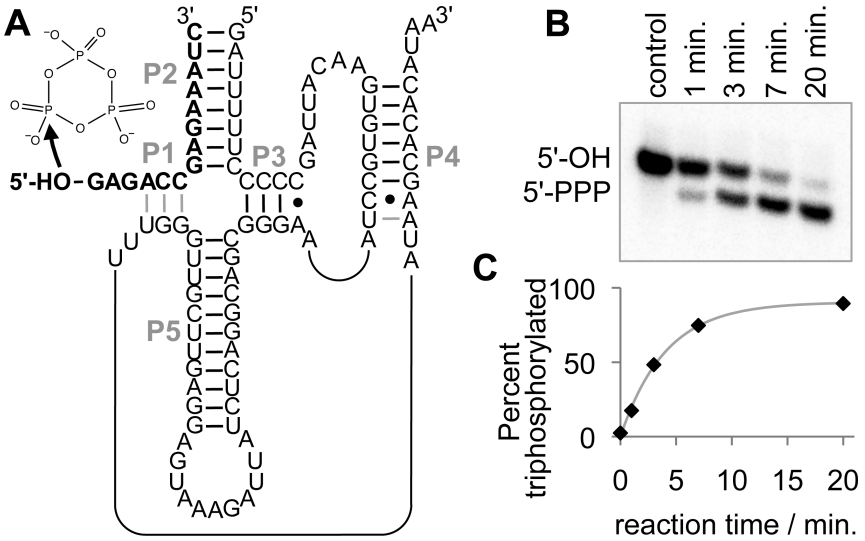


Figure 3.1: The parent triphosphorylation ribozyme, its reaction, and the assay of its activity. (A) secondary structure representation of the ribozyme. The duplexes are labeled in grey capital letters. Duplex P5 is the main focus of this study. Note that the 14-nucleotide substrate (bold) forms helices P1 and P2, and is radiolabeled internally. The nucleophilic attack of the 5'-hydroxyl oxygen on the trimetaphosphate is indicated with an arrow. (B) Phosphorimage of triphosphorylation products after PAGE separation. The triphosphorylated product (5'-PPP) migrates faster than the substrate (5'-OH) because of the increase in negative charges. (C) Kinetics from the reaction shown in (B). The grey line is a single-exponential curve fitted to the data. All triphosphorylation data analyzed in this study were described well by single-exponential fits.

fragments did not show detectable triphosphorylation activity (Fig. 2). When both fragments were combined the triphosphorylation activity of the fragmented ribozyme was within error of the full-length ribozyme. This result showed that the P5 helix can be formed in trans, leading to a

fully active triphosphorylation ribozyme.

The P5 duplex was truncated in several steps (Fig. 3A). The reduction of the P5 stem from 11 base pairs to 7, 6, and 5 base pairs did not generate a significant effect on the kinetics or the amplitude of the triphosphorylation reaction (Fig. 3B, C). In contrast, the further reduction of the P5 helix to 4, 3, 2, 1, and 0 base pairs unexpectedly caused an increase in triphosphorylation kinetics for the constructs with two and three base pairs, faster than the cis-acting ribozyme. Interestingly, the fraction of reacting ribozyme also increased from ~85% to ~95% when the P5 stem contained less than 5 base pairs. This observation suggests that the formation or separation of the P5 helix presents a rate-limiting conformational change during the triphosphorylation reaction. The shortest construct, in which the P5 helix was completely removed (Fig. 3D), showed about half of the triphosphorylation rate of the starting construct (Fig. S1). This reduction suggested that the P5 helix fulfills a function during the triphosphorylation reaction, presumably by favoring the catalytically active conformation. On the other hand, the triphosphorylation rate of 0.15 min^{-1} for the construct without P5 helix (as compared to 0.31 min^{-1} for the parental construct [73]) showed that structural elements can not only be reduced in length but also be completely omitted while maintaining a functional ribozyme. The absence of a helix that stabilizes the catalytically active conformation would be expected to shift optimal reaction kinetics to lower temperatures. To test this idea we measured triphosphorylation kinetics at different temperatures for the full-length ribozyme (Fig. 1A) and for the trans ribozyme without the P5 duplex (Fig. 3D). Indeed, the temperature optimum of the construct without the P5 duplex (15-25°C) is significantly below that of the cis ribozyme (40°C). Two additional features of this comparison are noteworthy: First, the maximal rate of the cis ribozyme is more than two-fold above the maximal rate of the fragmented ribozyme without the P5 helix. Second, at low temperatures (10°C and 15°C) the fragmented ribozyme without the P5 helix shows significantly faster triphosphorylation rates. This result has important implications for imagining how RNA world organisms could have functioned.

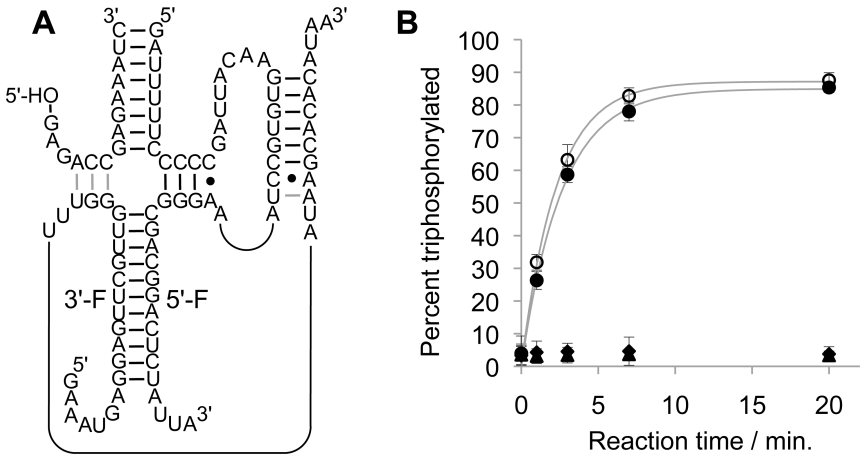


Figure 3.2: Fragmentation of the ribozyme in the loop of the P5 stem. (A) Secondary structure of the fragmented ribozyme, labeling its 5'-fragment (5'-F) and the 3'-fragment (3'-F). (B) Reaction kinetics of the full-length, unfragmented ribozyme (empty circles), the 5'-fragment alone (filled triangle), the 3'-fragment alone (filled diamonds), and the combination of 5'-fragment and 3'-fragment (filled circles). Error bars are standard deviations from three experiments; if error bars are not visible they are smaller than the symbols.

3.4 Discussion

The present study showed that an 86-nucleotide long triphosphorylation ribozyme could be fragmented in the internal duplex P5 without loss of activity, and that the P5 stem now formed in trans could be completely removed with only a 2-fold reduction in reaction kinetics. Interestingly, intermediates in the truncation series showed triphosphorylation kinetics 2-fold faster than the parent ribozyme. The fragmented and size-reduced ribozyme showed a temperature optimum of 15-25°C compared to 40°C for the parent ribozyme, with faster kinetics at low temperature.

The observed advantage of the fragmented, shortest construct over the full-length ribozyme at low temperature (Fig. 4) may reflect a more general phenomenon. The omission of stabilizing secondary structures would allow the ribozyme structure to 'breathe' more and access the conformations required for catalysis. The temperature of the Archaean ocean appears to have been mild, probably including glaciation events (especially close to the poles), because the early Sun was ˜25% less bright than today [74, 75, 76]. Indeed, early oceans may have been mostly frozen and only thawed episodically by meteorite impacts [77]. Accordingly, the behavior of

catalytic RNAs has also been explored in the lab at low temperatures, including the frozen state. A microenvironment in the eutectic spaces in ice provided the polymerase ribozyme with four key advantages [78, 79]: The hydrolysis of RNA and activated monomers is slowed, weak substrate interactions are improved (analogous to our finding in figure 4), the eutectic spaces between ice crystals afforded higher local concentrations of small molecule substrates with correspondingly faster reactions [80, 81], and an isolation of eutectic spaces from each other could have provided micro-compartments to aid evolution [70]. The current study suggests a fifth advantage: Low temperatures stabilize weak RNA structures including ribozymes assembled from multiple short fragments. This phenomenon would have allowed RNA world organisms to emerge and function by the assembly from short RNA fragments and therefore become more likely to occur. The length of required RNA polymers for ribozymes can be reduced in two ways, by ribozyme fragmentation and ribozyme minimization. First, the assembly of ribozymes from multiple fragments has been seen in nature with self-splicing group I and group II introns [82, 83]. The group I intron from Azoarcus even lent itself to disassemble into four fragments [84], the longest of which is at least 63 nucleotides in length [85]. RNase P can be assembled from four fragments, the longest of which is 109 nucleotides [86], the hairpin ribozyme can be separated into two structures interacting by tertiary interactions, with the longest strand being 50 nucleotides long [87], and the hammerhead ribozyme from two fragments, with the longest fragment 23 nucleotides in length [88]. In vitro selected ribozymes can also assemble from multiple fragments such as variants of the DSL ribozyme, where the longest fragment is 49 nucleotides long [89]. A second way to reduce the required RNA length for ribozymes is their minimization down to a catalytic core. The shortest known ribozyme is 5 nucleotides long and enhances the rate of aminoacylation by ˜25-fold above the background reaction [90]. Size-minimized ribozymes obtained by in vitro selections for other reactions require lengths between 31 nt and 119 nt in length, and typically show rate accelerations of more than a thousand-fold and often more than a million-fold [91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 37, 104, 105, 106, 107, 108, 109]. Since

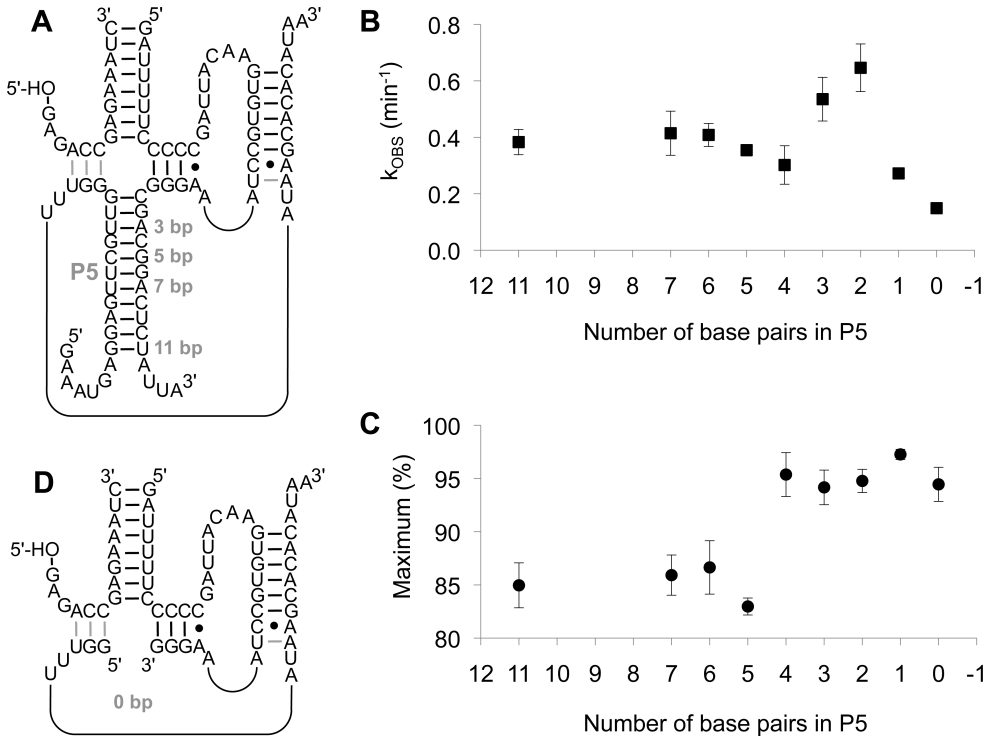


Figure 3.3: Truncation of the P5 duplex in the fragmented ribozyme. (A) Secondary structure of the fragmented ribozyme. The lengths of the P5 duplex are annotated. Note that only the 11 base pair construct carried unpaired extensions that stem from the loop in the parental construct; all shorter constructs do not contain such overhangs. (B) Observed reaction kinetics with fragmented ribozyme constructs, plotted as function of their P5 duplex length. (C) Observed extent of the triphosphorylation reaction with fragmented ribozyme constructs, plotted as function of their P5 duplex length. Error bars are standard deviations from three experiments; if error bars are not visible they are smaller than the symbols. (D) Secondary structure of the shortest, fragmented ribozyme construct. The lengths of the three RNA fragments are 14 nt (substrate), 34 nt (5'-fragment), and 19 nt (3'-fragment).

each of these selections sampled most of the sequence space up to 23 nucleotides these results also suggest that for most reactions, ribozymes shorter than 24 nucleotides don't exist. Because the first RNA world system would have to rely on prebiotically, nonenzymatically produced polymers these minimum lengths place a strong selection pressure for the first RNA world organism to access only the smallest ribozymes, or to assemble larger ribozymes from shorter fragments.

An additional advantage of ribozyme fragmentation lies in avoiding the inhibition of RNA polymerases by strong secondary structures [110, 111]. The synthesis of RNA polymers with stable secondary structures is necessary in an RNA world because most ribozymes contain strong

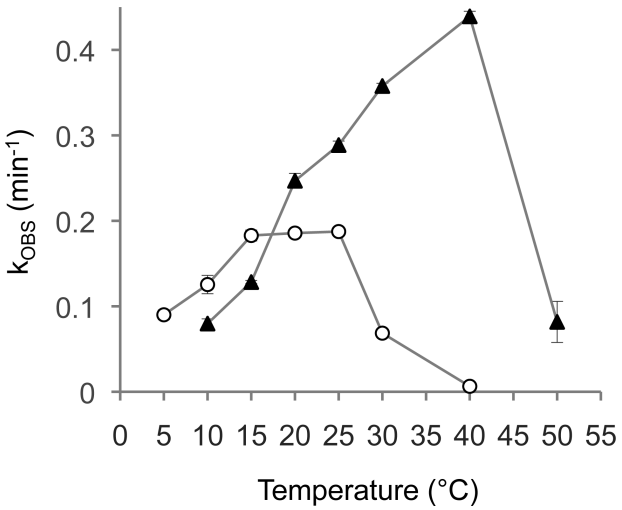


Figure 3.4: Temperature dependence of triphosphorylation kinetics. The observed reaction rates are plotted as a function of the reaction temperature for the unfragmented, full-length ribozyme (filled triangles; ribozyme sequence as shown in figure 1A) and the fragmented, shortest ribozyme (empty circles, ribozyme sequence as shown in figure 3D). Error bars are standard deviations from three experiments; if error bars are not visible they are smaller than the symbols.

secondary structures. In an RNA world this problem would be significant due to the difficulty of developing an RNA helicase [112, 113]. By separating the two partially complementary strands of a given helix onto two transcripts the helix cannot form in the transcript (or its complementary template) during synthesis, thereby avoiding the inhibition of polymerization. To our knowledge, this advantage of ribozyme fragmentation has not been published previously.

We did not expect that the complete removal of the P5 stem would lead to functional ribozymes [73]. However, the fragmented ribozyme variants with truncated P5 duplex were fully functional, and two of them - with a P5 stem length of 2 and 3 base pairs - were even faster than the ribozyme with a full P5 duplex (figure 3B). Similarly, the fraction of unreactive ribozyme was reduced from ˜15% for P5 lengths of 11 base pairs, to ˜5% for P5 lengths of 1-4 base pairs (figure 3C). This suggests that some flexibility at the base of the P5 stem helps to access the catalytically active conformation of the ribozyme. These observations are also consistent with the finding that ribozyme variants with a base mismatch at the second position of a full P5 helix accelerate reaction kinetics [73]. Future studies with many different ribozymes may be able to

determine whether this benefit of flexibility is specific for the studied ribozyme or whether it may be a general advantage of fragmented ribozymes. A general benefit for fragmented ribozymes would increase the likelihood of an RNA world emerging from a prebiotic scenario with short RNAs.

3.5 Materials and Methods

Ribozyme constructs were generated by PCR mutagenesis or the Quikchange protocol for site directed mutagenesis (Stratagene). The template was a plasmid encoding the ribozyme TPR1e as described in Dolan et al. [73]. DNA sequences were confirmed by cloning into fresh pUC19 plasmids and sequencing. However, mutations near the 5'- or 3'-termini were introduced by PCR primers during PCR amplification of the templates for transcription. RNAs were generated by run-off transcription by T7 RNA polymerase as described [32] and purified by denaturing PAGE.

The 14-nucleotide substrate was internally [32P] labeled during transcription. Two hammerhead ribozymes were encoded in the transcript, one at the 5'-terminus and one at the 3'-terminus. The 5'-terminal hammerhead ribozyme generated the 5'-hydroxyl group, while the 3'-terminal hammerhead ribozyme generated a homogeneous 3'-terminus [114]. Cleaved transcripts were purified by denaturing PAGE.

Triphosphorylation reactions were performed essentially as described [32, 73]. Triphosphorylation reactions with 5.5 M triphosphorylation ribozyme and substoichiometric concentrations of radiolabeled substrate were incubated with 50 mM Tmp, 100 mM $MgCl_2$ (corresponding to 50 mM free Mg^{2+}), and 50 mM Tris/HCl (final pH 8.1) for three hours at 22°C if no other temperature is given. At specific time points, 1.5 L of the reactions were added to 4 L of formamide PAGE loading buffer containing 50 mM Na_2EDTA, heat denatured, and separated by denaturing 20% PAGE. The gel shift is caused by covalent triphosphorylation of the RNA, which was demonstrated previously by the fact that the product RNAs served as substrate for ligase

ribozymes, and by mass spectrometric comparison of the RNA substrate and RNA product [32].

Data quantitation: After electrophoresis, PAGE gels were exposed to phosphorimager screens and scanned on a PMI phosphorimager (Bio-Rad). Signals on the phosphorimager scans were quantitated using the software Quantity One. Curve fitting was done in Microsoft excel using single-exponential equations, minimizing the sum of the squared differences to the data using the solver sub-software.

3.6 Supplemental

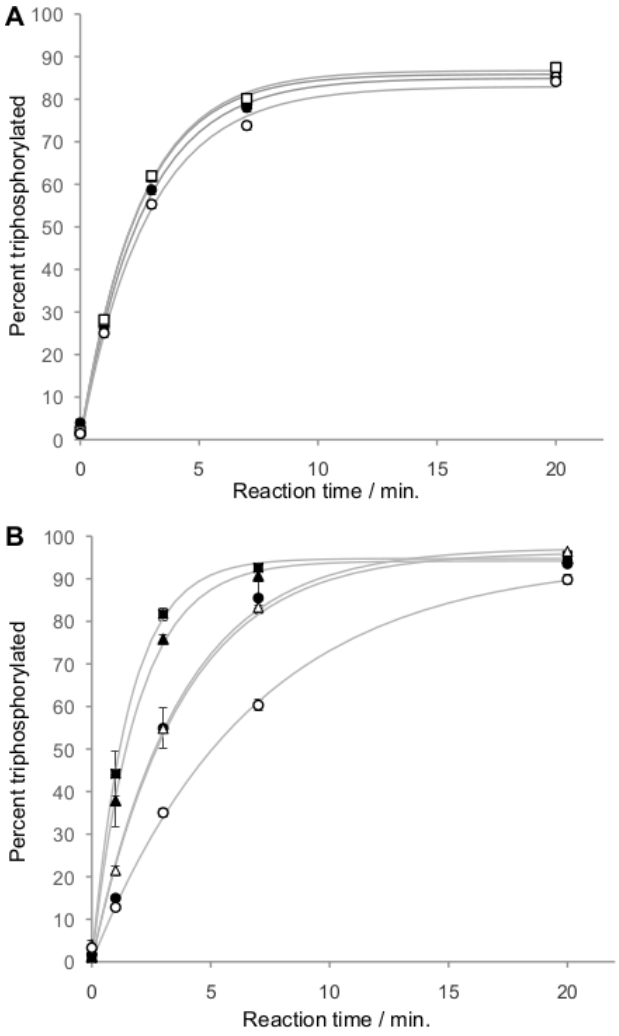


Figure 3.5: Triphosphorylation kinetics of ribozyme constructs with successive truncations int he P5 helix. The same data are displayed in figures 3C, D in different format. (A) Triphosphorylation kinetics of the ribozyme with fragmented loop L5 and full-length P5 helix (filled circles), seven (triangles), six (empty squares), and five base pairs in the P5 helix (empty circles). (B) Triphosphorylation kinetics of the ribozyme with four (filled circles), three (filled triangles), two (filled squares), one (empty triangles), and zero base pairs in the P5 helix (empty circles). Error bars are standard deviations from 3-4 experiments. Grey curves are single-exponential least-squares fits to the individual data sets.

Chapter 3, in full, is a reprint of the material as it appears in Physical Chemistry Chemical Physics. Akoopie, Arvin; Müller, Ulrich F. 2016. The dissertation author was the first author of this paper.

Chapter 4

Cotranscriptional 3'-end processing of T7 RNA polymerase transcripts by a smaller HDV ribozyme

4.1 Abstract

In vitro run-off transcription by T7 RNA polymerase generates heterogeneous 3'-ends because the enzyme tends to add untemplated adenylates. To generate homogeneous 3'-termini, HDV ribozymes have been used widely. Their sequences are added to the 3'-terminus such that co-transcriptional self-cleavage generates homogeneous 3'-ends. A shorter HDV sequence that cleaves itself efficiently would be advantageous. Here we show that a recently discovered, small HDV ribozyme is a good alternative to the previously used HDV ribozyme. The new HDV ribozyme is more efficient in some sequence contexts, and less efficient in other sequence contexts than the previously used HDV ribozyme. The smaller size makes the new HDV ribozyme a good alternative for transcript 3'-end processing.

4.2 Introduction

T7 RNA polymerase is an important tool in RNA biochemistry, by transcribing long sequences of RNA from DNA templates that contain the 17 base pair long promoter [115]. However, this enzyme tends to generate heterogeneous 3'-termini, consisting of untemplated adenylates [116] or more complicated sequences that result from self-priming of the 3'-terminus [117, 118].

Homogeneous 3'-termini are sometimes required for the pool molecules of in vitro selection and evolution experiments. If the selection procedure requires catalysis to occur at an RNA 3'-terminus then heterogeneity at this 3'-terminus can result in a large fraction of the pool being inactive, and a consequent loss of pool complexity. Ribozymes that react at their 3'-termini have been selected multiple times [96, 119, 120, 121, 122], and in one case an HDV ribozyme was used to generate homogeneous 3'-termini of the pool RNA [123]. In addition to generating homogeneous 3'-termini, these ribozymes (or deoxyribozymes) usually generate 2',3'-cyclic phosphates and 5'-hydroxyl groups. The 2',3'-cyclic phosphates can be used as chemical activation group to react with 5'-hydroxyl groups and form linear 3',5'-phosphordiester bonds. This has been used during the in vitro selection of catalytic DNAs ligating 2',3'-cyclic phosphates with 5'-hydroxyl groups [124, 125, 126]. These examples show that the efficient processing of RNA 3'-termini is an important step for a range of in vitro selection experiments.

To generate homogeneous 3'-termini in transcripts of T7 RNA polymerase, different ribozymes and deoxyribozymes have been used [127]. The cis-acting HDV ribozyme is widely used due to its convenience: The sequence of the ribozyme is simply added to the 3'-end of the desired transcript. During co-transcriptional self-cleavage, the HDV ribozyme cleaves itself off the desired sequence at one defined position [128, 129]. Other ribozymes that have been used for 3'-end processing are the hammerhead ribozyme [114], the hairpin ribozyme [114], and the VS ribozyme [128]. However, the hammerhead ribozyme and the hairpin ribozyme have a

significant sequence requirement on the desired transcript, and the hairpin ribozyme is prone to aberrant cleavage [114]. The HDV ribozyme is most general because it will cut efficiently after any nucleotide other than G [130], where cleavage kinetics are ˜10-fold slower [131]. Different upstream sequences appear to interfere with the ribozyme's formation of its P1.1 helix (Figure 1A), leading to reduced ribozyme efficiency [132, 133, 134, 135, 136].

The ideal HDV ribozyme for homogeneous 3'-termini would be small so that it does not significantly increase the size of the transcript, and its co-transcriptional self-cleavage would be efficient and sequence general. Previously used HDV ribozymes have lengths of 67 nucleotides [137], 84 nucleotides [127], and 86 nucleotides [128]. In several cases the self-cleavage efficiency is low. This can be addressed partially by temperature cycling of the transcription mixture between 25°C and 60°C for ˜10 times, with or without increase of the total magnesium ion concentration to 30-40 mM $MgCl_2$ [134]. In our hands such incubation conditions led to significant degradation of the desired transcript (data not shown).

Previously, a shorter variant of the HDV ribozyme was identified in a metagenomic data set of human sewage by a systematic, computational search for truncated HDV ribozyme variants [138]. This HDV ribozyme (drz-Mtgn-3) shows cleavage kinetics on par with many larger HDV ribozymes [139] and has a size of only 56 nucleotides. This ribozyme is the focus of this study, and was termed HDV56 for simplicity.

The current study examined the HDV56 ribozyme for its ability to process 3'-termini of transcripts by T7 RNA polymerase, and compared it to the performance of the shortest previously used, 67-nucleotide long HDV ribozyme variant [134, 137], which we termed HDV67. When compared on three different upstream sequences, HDV56 showed the highest and the lowest efficiency between the two ribozymes. To obtain a more general statement regarding the sequence generality, a random sequence with 30 nucleotides was inserted upstream of the HDV ribozyme variants. From this unbiased library, HDV56 generated the same amount of product with the correct length as the HDV67 ribozyme, within error. Together, these results show that HDV56 is

a shorter alternative for HDV ribozymes that generate homogeneous 3'-end by co-transcriptional processing.

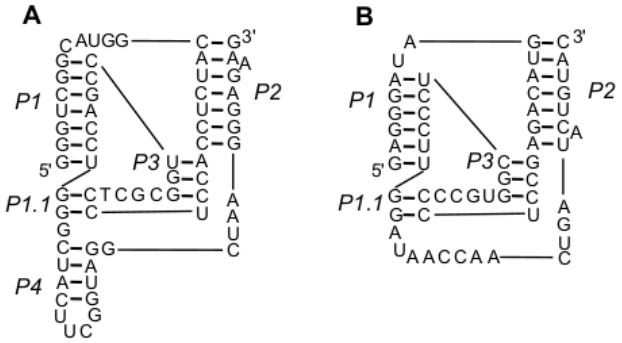


Figure 4.1: Secondary structure representations of the two HDV ribozyme variants compared in this study. (A) The previously shortest HDV ribozyme [137] with a length of 67 nt (here termed HDV67). The secondary structure was based on the sequence / secondary structure consensus model given in Riccitelli et al. [138]. (B) The small HDVdm3 ribozyme with a length of 56 nt (here termed HDV56), with the secondary structure given in Riccitelli et al. [138].

4.3 Results

To test whether the small HDV ribozyme HDV56 [138] is not only smaller than previously studied HDV ribozymes but also similarly efficient for in vitro co-transcriptional 3'-end processing, the HDV56 sequence was cloned to the 3'-terminus of three unrelated sequences. These three sequences were the highly structured RNA 5'-triphosphorylation ribozyme TPR1e (96 nt), which was developed by in vitro selection [73], and two less structured RNAs: A 115-nucleotide long portion of the mRNA of E. coli EF-Tu2, and a 191-nucleotide long portion of the mRNA for E. coli DNA polymerase III subunit alpha. Upstream of each sequence, the promoter for T7 RNA polymerase was added to facilitate in vitro transcription. The HDV67 ribozyme [134, 137] was used as comparison to the HDV56 sequence. The sequence of each construct was confirmed by cloning into the vector pUC19 and sequencing. PCR amplification of these constructs resulted in PCR product lengths of 173 - 279 base pairs. These PCR products were used as templates for the transcription reactions.

The transcription products were internally labeled through the presence of alpha[32P]-ATP in the transcription reaction. During transcription, samples were taken at different times, and separated by denaturing polyacrylamide gel electrophoresis (figure 2A). This procedure allowed following the buildup of full-length transcripts, as well as the processed transcripts. The parameter of most importance to an application of this ribozyme is the molar amount of processed RNA produced with the correct length. This amount was highest after one hour for all three upstream sequences, therefore the one-hour time point was used to compare the constructs.

The processed sequence TPR1e was produced 3.3-fold more efficiently with HDV56 than with HDV67 (Figure 2B). The processed sequence EFTu2 was produced 1.3-fold more efficiently with HDV56 than with HDV67 (figure 2C). The processed sequence DNA pol III was produced 3.3-fold more efficiently with HDV67 than with HDV56 (figure 2D). These results show that HDV56 and HDV67 have a different preference for upstream sequences. However, the number of three different upstream sequences was not sufficient to obtain a general picture about the processing efficiency of the two tested HDV ribozyme variants. To arrive at a more general statement on the 3'-processing efficiency, a random sequence with 30 nucleotides was inserted upstream of the HDV ribozyme variants. Co-transcriptional processing resulted in the same cleavage efficiency by HDV56 when compared to HDV67, within error (figure 3). These results show that the average cleavage efficiency of HDV56 is similar to that of HDV67. Due to its shorter size, HDV56 may be the first choice to test for 3'-end processing of a new transcript.

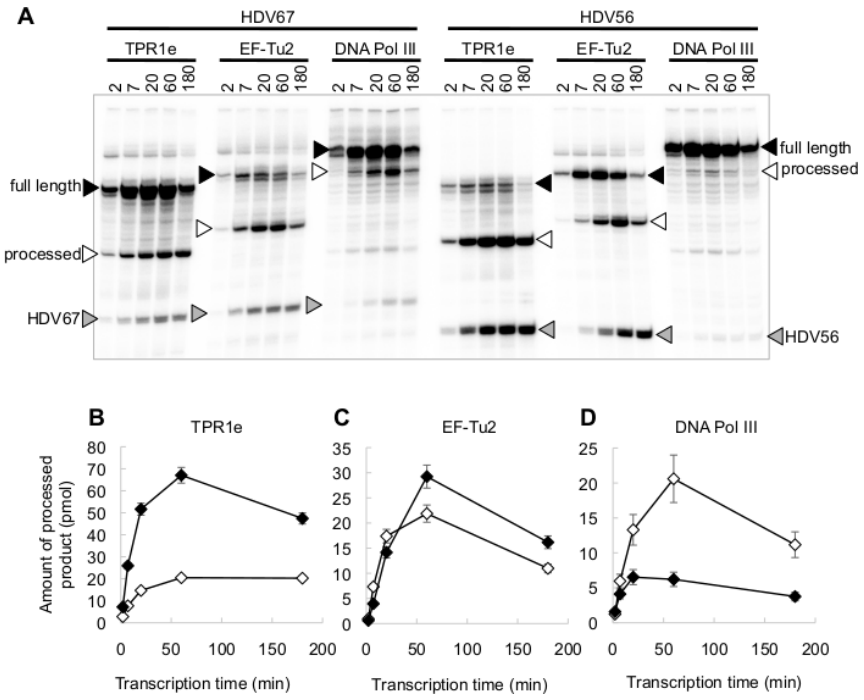


Figure 4.2: Co-transcriptional 3'-processing by HDV67 and HDV56, using three different sequences upstream of the HDV ribozyme. (A) Phosphorimage of reaction products separated by 15% polyacrylamide gel electrophoresis. Samples were taken from the transcription reactions at the indicated time points. The signals stem from internal labels caused by traces of alpha-[32P] ATP included in the transcription mixture. The three time courses on the left stem from the HDV67 ribozyme, the three time courses on the right from the (new) HDV56 ribozyme. For each ribozyme, the upstream sequence is indicated with (from left to right) the triphosphorylation ribozyme TPR1e, a 115-nucleotide fragment of the mRNA for E. coli eF-Tu2, and a 191-nucleotide fragment of E. coli DNA polymerase III. (B-D) Quantitation of processed transcript, as a function of transcription time. The previously used HDV67 ribozyme (white symbols) and the new HDV56 ribozyme (black symbols) are shown for each of the three upstream sequences. Error bars denote standard deviations from three replicate experiments. In cases where the error bars are not visible they are smaller than the symbols.

4.4 Discussion

Our study found that HDV56 is a good alternative to the HDV67 ribozyme for co-transcription processing. The availability of multiple alternative HDV ribozymes for a given upstream sequence is useful because different upstream sequences appear to lead to differences in inefficient folding of the ribozyme [134]; the different effect of upstream sequences on HDV ribozymes was confirmed in our study (see figure 2). Importantly, our study showed that HDV56

and HDV67 showed a different preference of three upstream sequences. This characteristic may be useful because if one upstrem sequence does not work well with one ribozyme then the other ribozyme may be efficient. The preference of upstream sequences is not explained

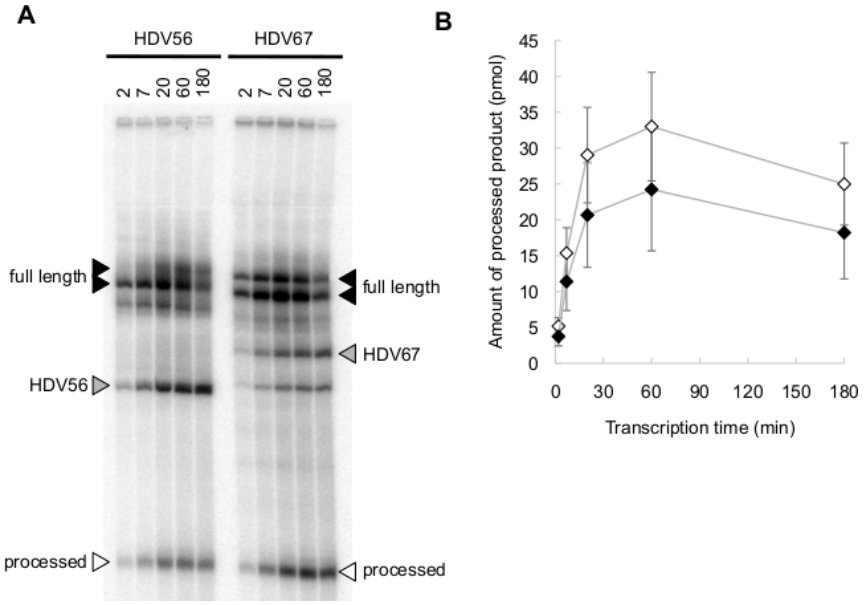


Figure 4.3: Co-transcriptional 3'-processing by HDV67 and HDV56, using a randomized N30 sequence upstream of the HDV ribozyme. (A) Phosphorimage of reaction products separated by 15% polyacrylamide gel electrophoresis. Samples were taken from the transcription reactions at the indicated time points. The signals stem from internal labels caused by traces of -[32P] ATP included in the transcription mixture. The identity of each ribozyme is indicated on the top. (B) Quantitation of processed transcript, as a function of transcription time. The molar amount of processed product was calculated from the signal strengths using a separate calibration (see materials and methods). The previously used HDV67 ribozyme is shown in white symbols and the new HDV56 ribozyme is shown in black symbols. Error bars denote standard deviations from three replicate experiments. In cases where the error bars are not visible they are smaller than the symbols.

well by a previously identified rule. This rule states that the sequence immediately upstream of the cleavage site can extend the P1 helix, and thereby interferes with the formation of the correct P1.1 helix [132, 133, 135, 136] (see figure 1). In our study, this rule explained the slight advantage of HDV56 over HDV67 on the substrate EF-Tu2 but not the stronger differences between the ribozymes on the substrates TPR1e, and DNA Pol III: For both ribozymes HDV56 and HDV67, the two nucleotides immediately upstream of the cleavage site would have to pair to

a GG sequence in the ribozymes, and therefore should not cause a difference in extending the P1 helix. Only the third upstream position the cleavage site should be able to cause a difference, which could pair with a G in HDV56 and with an A in HDV67 (see figure 1). For clarity, the third upstream nucleotide is underlined in the following three tested upstream sequences. First, the upstream sequence 5'-CCU-3' of EF-Tu2 could allow an extension of the P1 helix in HDV67 by three canonical base pairs, and only two canonical base pairs in HDV56. Therefore, HDV67 could be inhibited more by the upstream sequence, explaining the slight advantage of HDV56 over HDV67 on EF-Tu2. In contrast, the upstream sequence TPR1e (5'-TAA-3') cannot explain the much higher cleavage efficiency of HDV56 than HDV67 because both ribozymes would not be able to extend the P1 duplex by a single base pair. Similarly, the upstream sequence of DNA Pol III (5'-UGC-3') cannot explain the lower efficiency of HDV56 than HDV67 because both ribozymes could be extended by only one G:C base pair, with the second position (a G:G pair) destabilizing an elongation of the P1 duplex. These three cases show that additional rules still need to be identified to describe the influence of upstream sequences for the cleavage efficiency of HDV ribozymes. Therefore, some differences in processing efficiencies may not be due to secondary structure formation but due to the sequence itself at the processing site. While classical HDV ribozymes show no strong sequence preference except against G [131, 130] it is possible that the loss of the P4 helix renders the HDV56 ribozyme more sensitive to the upstream sequence.

The efficiency of HDV67 appears to be rather constant across the three individual, tested sequences (always around 20 pmol after 1 hour transcription) while the efficiency of HDV56 is more efficient on substrate TPR1e and less efficient on substrate DNA polymerase III. This suggests that HDV67 is more robust against different upstream sequences than HDV56, and that HDV56 is able to achieve higher cleavage efficiency than HDV67 for sub-groups of substrates (see figure 2A). The identification of molecular causes for these behaviors - or the evolutionary optimization of these ribozymes - may be able to identify HDV variants that are as small as HDV56, display the sequence generality of HDV67, and the high efficiency shown by HDV56 on

the substrate TPR1e.

To further improve the efficiency of 3-end processing by HDV ribozymes it is promising to titrate the Mg^{2+} concentration in the transcription reaction. The assays in this study used the relatively high free Mg^{2+} concentration of 10 mM, resulting from 26 mM total Mg^{2+} and 4 mM of each NTP. This is higher than the standard procedure recommended earlier [116] with 6 mM free Mg^{2+} (4 mM each NTP and total 22 mM Mg^{2+}). While the HDV ribozyme can be more efficient at Mg^{2+} concentration below 1 mM [140], the T7 RNA polymerase benefits from higher Mg^{2+} concentrations [141]. Therefore we recommend testing different Mg^{2+} concentrations for the co-transcriptional processing by HDV ribozymes.

4.5 Materials and Methods

4.5.1 Preparation of DNA constructs

The constructs for co-transcriptional 3'-processing were generated by cloning the sequence of TPR1 (96 nt), a 115-nucleotide long fragment of EF-Tu2, and a 191-nucleotide long fragment of DNA Pol III into the vector pUC19. The products were confirmed by sequencing. These plasmids were used as templates to attach the sequence for the T7 promoter at the 5'-terminus, and the sequence for HDV56 or HDV67 at the 3'-terminus. The PCR product lengths were confirmed to be of the right length by agarose gel electrophoresis, and the PCR products were purified on silica-adsorption columns (Machery-Nagel, num740609), to be used as template for the transcription assays.

To generate the DNA template for transcription of partially randomized sequences, PAGE-purified 69-nucleotide long DNA oligonucleotides (IDT) were used as template for PCR amplication of the sequences [T7 promoter]-[GGG]-[N30]-[HDV56] or [T7 promoter]-[GGG]-[N30]-[HDV67]. The sequence GGG was added upstream of the randomized sequence to allow for efficient transcription. The sequence of the HDV ribozymes was completed through the sequence

of the 56-nucleotide long or 67-nucleotide long 3'-PCR primer for HDV56 and HDV67, respectively. The PCR products were purified on silica-adsorption columns (Machery-Nagel, # 740609), to be used as template for the transcription assays.

4.5.2 Co-transcriptional 3'-end processing assay

For co-transcriptional 3'-processing, 100 nM DNA template was incubated at 37°C with 26 mM $MgCl_2$, 2.5 mM Spermidine, 0.01% (w/v) Triton X-100, 1 mM of each NTP, 5 mM DTT, and 40 mM Tris/HCl pH 7.9, with trace amounts of -[32P]-ATP for quantitation, in a total volume of 20 L. Aliquots of 3.5 L were removed after indicated time points and quenched with 3.5 L of formamide PAGE gel loading buffer (90% (v/v) formamide, 45 mM Tris(hydroxymethyl) aminomethane, 45 mM boric acid, and 30 mM Na2EDTA). Five L of each timepoint were then loaded onto a 15% denaturing polyacrylamide gel. After separation by electrophoresis, the gel was exposed to a phosphorimaging screen (Kodak screen-K) and scanned on a Personal Molecular Imager (Bio-Rad).

4.5.3 Data processing

The signals from scanning the phosphorimaging screens were quantitated with the Quantity One software using the 'rectangle' method and background subtraction. To convert the signals (cpm) to molar amounts, transcription experiments were performed for each construct in triplicate, with a 1-hour incubation time and without radiolabel. The products were separated by 15% PAGE and detected by UV shadowing. The band corresponding to processed RNA was excised, the RNA eluted, and quantified by its absorption at 260 nm. The resulting molar amount, and its ratio with the cpm of the same band after one hour transcription was used to used to calculate the molar amount of each band.

4.5.4 Sequences

The following sequences were used in this study: TPR1e: 5'-GAG ACC GAGA TGT TT TTCC CC CGA TTA CAA GT GTG CCT AAAG GGC TAC GGA CTC TAT TAG AAA TG AGGA GTT CGT TGG GTT TAT AAGC ACA CAT AA.

EF-Tu2: 5'-GG AAG TTCG TG AACT TCT GTCT CAG TACG ACTT CCC GGGC GAC GACA CTCC GAT CGTT CGT GGTT CTG CCT GAAA GCGCT GGAA GGCG ACGCA GAGTG GGAA GCGA AAAT CCT.

DNA polymerase III subunit a: 5'-GGGATGGTGGA TAACTTTATCG ACCGTAAA-CATGGTCG TGAAGAGA TCTCCTATCCGGACGT ACAGTGGCA GCATGAAAGCCTGA AACCGGTACTG GAGCCAACCTACGGC ATTATCCTGTAT CAGGAACAGGTCATG CA-GATTGCGCA GGTGCTTTCTGGTTATA CCCTCGGTGGCG CGGATATGCTGC.

Randomized sequence: 5'-GGG-N30.

HDV67: 5'-GGGTCGGCATGGCATCTCCACC TCCTCGCGGTCCGACCTGGGCTA CTTCGGTAGGCTAAGGGAGAAG

HDV56: 5'-GAGGGATAGTACAGAGCCTC CCCGTGGCTCCCTTGGATAACC AACT-GATACTGTAC.

A plasmid with the sequence of TPR1e was present in the Muller lab from a previous study [73]. A plasmid with the sequence for EF-Tu2 was a generous gift from the lab of Simpson Joseph (UCSD). The sequence for DNA polymerase III subunit alpha was amplified by PCR from E. coli genomic DNA. All other sequences were assembled by PCR using DNA oligonucleotide primers.

4.6 Acknowledgements

We thank Andrej Luptak for helpful discussions. Partial support to U.F.M came from grant NASA NNX13AJ09G and NASA grant NNX16AJ27G. Support for A.A. was provided by

a 2015/2016 Department of Education Graduate Assistance in Areas of National Need (GAANN) Training Grant P200A150251, and by a 2016 Distinguished Graduate Student Fellowship by the Department of Chemistry Biochemistry at UC San Diego.

Chapter 4, in full, is a reprint of the material as it appears in Journal of Molecular Evolution. Akoopie, Arvin; Müller, Ulrich F. 2018. The dissertation author was the first author of this paper.

Chapter 5

The NTP binding site of the Polymerase Ribozyme

5.1 Abstract

A previously developed RNA polymerase ribozyme uses nucleoside triphosphates (NTPs) to extend a primer 3'-terminus, templated by an RNA template with good fidelity, forming 3'-5'-phosphordiester bonds. Indirect evidence has suggested that the ribozyme's accessory domain binds the NTP with a highly conserved purine-rich loop. To determine the NTP binding site more precisely we evolved the ribozyme for efficient use of 6-thio guanosine triphosphate (6sGTP). 6sGTP never appeared in the evolutionary history of the ribozyme, therefore it was expected that mutations would appear at the NTP binding site, adapting to more efficient binding of 6sGTP. Indeed, the evolution identified three mutations that mediate 200-fold improved incorporation kinetics for 6sGTP. A more than 50-fold effect resulted from mutation A156U in the purine-rich loop, identifying the NTP binding site. This mutation acted weakly cooperative with two other beneficial mutations, C113U in the P2 stem near the catalytic site, and C79U on the surface of the catalytic domain. The preference pattern of the ribozyme for different NTPs changed

when position 156 was mutated, confirming a direct contact between position 156 and the NTP. The results suggest that A156 stabilizes the NTP in the active site by a hydrogen bond to the Hoogsteen face of the NTP.

5.2 Introduction

The RNA world hypothesis describes a stage in the early evolution of life in which RNA served as the genome and as the only genome-encoded catalyst [142, 12, 13, 11, 20]. To test how the RNA world could have functioned, catalytic RNAs (ribozymes) are being developed that may ultimately allow establishing an RNA world organism in the lab [143]. The central catalyst in an RNA world organism would be an RNA polymerase ribozyme that facilitates self-replication. A ribozyme with RNA polymerization activity was developed in 2001, the R18 polymerase ribozyme [34]. Several variants of this ribozyme have been developed with improved performance [68, 69, 70, 35] although so far, none are able to mediate self-replication.

The structure of the R18 polymerase ribozyme consists of two domains, the catalytic core (or ligase domain), and the accessory domain (figure 1A). The three-dimensional structure of the catalytic core has been determined by X-ray crystallography [144, 145], showing how the 3'-hydroxyl group of the extending primer is positioned for an in-line attack on the 5'-triphosphate of the incoming nucleoside triphosphate (NTP). In contrast, much less is known about the structure of the accessory domain. The accessory domain is not necessary for primer extension because three to six nucleotides can be added when a part of the catalytic core serves as template [146]. However, the accessory domain becomes necessary when a primer / template duplex is bound in trans [34]. This allows for longer polymerization to occur via a mostly non-processive mechanism of binding and releasing the primer / template duplex [147], held in place to a large extent by contacts to 2'-hydroxyl groups in the primer and template [148]. The accessory domain is draped over the vertex of the tripod structure of the catalytic core, aided by tertiary interactions between

the terminal AL4 loop of the accessory domain and the J3/4 loop within the catalytic domain [149].

The binding of NTPs appears to be mediated by a purine-rich loop in the accessory domain [149]. This purine-rich loop seems to be ideally positioned near the catalytic site of the catalytic domain, and the finding that any mutation in this loop leads to a substantial drop in polymerization activity supports the idea that the purine-rich loop is the binding site for NTPs. However, these data do not generate a causal link to NTP binding, and it is still unclear how the interactions of accessory domain and catalytic core stabilize the NTP in the catalytic site.

To identify the location of the NTP binding site in a functional assay we evolved a variant of the R18 polymerase ribozyme for efficient use of 6-thio guanosine triphosphate (6sGTP). The polymerase ribozyme has never encountered this specific NTP in its evolution/design history [34, 146, 33, 106, 150]. Therefore, evolving the polymerase ribozyme towards utilizing 6sGTP would be expected to enrich for mutations that reveal the NTP binding site. During the evolution, the mutations were allowed to occur throughout the ribozyme sequence to detect any interactions that may assist NTP binding. The most efficient isolate from the evolution had $>$200-fold improved 6sGTP ligation kinetics and relied on three mutations. The mutation A156U in the purine-rich loop was responsible for $>$50-fold improvement, suggesting a direct contact of A156 with the nucleobase of the incoming NTP. The direct interaction between A156 and the NTP was confirmed in experiments where A156 was mutated, which changed the NTP preference pattern. Mutations in position 79 and 113 had less than 3-fold effects and were weakly cooperative with A156U but not with each other. The resulting network of interactions is consistent with existing structural data and builds a picture of how the dynamic interactions in the polymerase ribozyme mediate RNA polymerization.

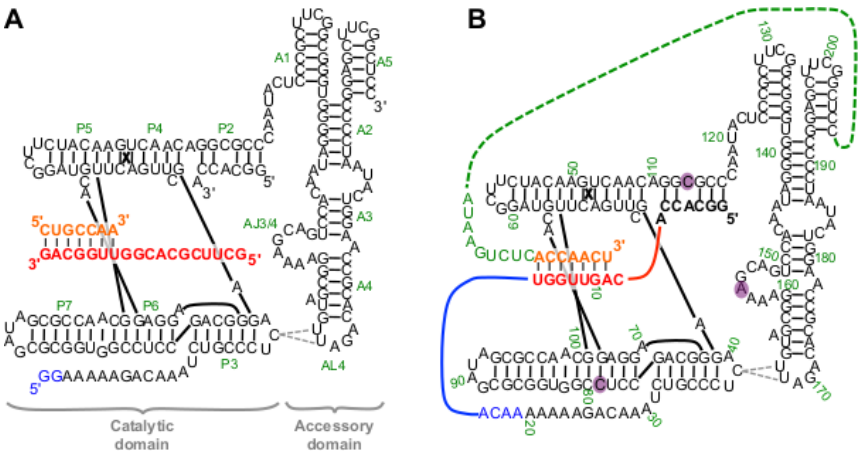


Figure 5.1: Secondary structure representations of (A) the R18 polymerase ribozyme [34] and (B) the construct used for in vitro evolution for efficient 6sGTP ligation. All nucleotides that are identical between the two constructs are shown in black. The specific secondary structure and the contact point between AL4 and C39 are based on Wang et al. [149]. Secondary structure elements are indicated in green in (A). The secondary structure elements for the ligase core (P2-P7) are labeled consistent with Ekland et al [106] and labeling for the accessory domain (A1-A5) is based on Wang et al [149]. The nucleotide numbering used in this study is shown in green in (B). The three nucleotides of central importance in this study (C79, C113, A156) are shown with a purple background.

5.3 Materials and Methods

5.3.1 Generation of the polymerase ribozyme construct

The polymerase ribozyme construct for the evolution was generated by PCR amplification from a plasmid containing the R18 polymerase ribozyme [34]. A second PCR reaction added a T7 promoter and a hammerhead ribozyme on the 5'-end similar to previous work [32], and a linker and DNAzyme recognition site (see below) on the 3 end. The resulting sequence of the DNA construct was 5 - *AAT TTA ATA CGA CTC ACT ATA* ggg tgg tgc cct gac gag cta agc gaa act gcg gaa acg cag tc GGC ACC ACA GTT GGT AC AAA AAA AGA CAA ATC TGC CCT CAG AGC TTG AGA ACA TCT TCG GAT GCA GAG GAG GCA GCC TCC GGT GGC GCG ATA GCG CCA ACG TTC TCA ACA GGC GCC CAA TAC TCC CGC TTC GGC GGG TGG GGA TAA CAC CTG ACG AAA AGG CGA TGT TAG ACA CGC CAA GGT CAT AAT C CCC GGA GCT TCG GCT CC NNN NNN NNN NNN NNN ATA AGT CTC ACC AAC T/TAT

ATG TTC TAG CGC GGA 3, where the italicized letters indicate the T7 promoter, the lowercase letters indicate the hammerhead ribozyme, the underlined regions indicate primer binding sites for PCR and reverse transcription, the N15 sequence is the linker sequence (see two paragraphs further down), and the sequence downstream of the linker sequence shows the binding site of the DNAzyme with the cleavage site marked as "/" (see next paragraph). The RNA sequence was generated by in vitro run-off transcription from this PCR product with T7 RNA polymerase under standard conditions, and purification by denaturing PAGE.

5.3.2 Processing of ribozyme 3'-ends with a DNAzyme

To generate a homogeneous 3'-terminus of the purified RNA transcript with 2'- and 3'-hydroxyl groups, a recently developed DNAzyme was employed that generates a 5'-phosphate and 2',3'-hydroxyl groups [151]. To do this the DNAzyme variant 9SK17 was modified for better complementary to the pool 3'-terminus, resulting in the sequence 5'-GCG CTA GAA CAT GCC AGC GAT CAA AGA CGG CGA GTT GTA CCC ATAG GTG TCT AGT tgg TGA GAC TT -3' where the underlined sequences were complementary to the substrate, and the three lowercase nucleotides differ from 9SK17. The pool transcript was heat renatured (3 minutes / 90°C) with a stoichiometric excess of the DNAzyme, chilled on ice for 10 minutes, and incubated at 37°C for 4 hours with the final concentrations of 70 mM HEPES/NaOH pH 7.5, 150 mM NaCl, 1 mM $ZnCl_2$, 2 mM HNO_3, 20 mM $MnCl_2$, and 40 mM $MgCl_2$. The $ZnCl_2$ stock solution was a mixture of 10 mM $ZnCl_2$ with 20 mM HNO_3 and 200 mM HEPES/NaOH pH 7.5. The reaction was stopped by adding an excess of Na2EDTA over divalent cations, and adding formamide to a final concentration of 50% (v/v). The processed RNA pool was purified by denaturing 5% polyacrylamide gel electrophoresis, resulting in a yield of about 50% for the processed pool RNA.

5.3.3 Development of linker sequences between ribozyme 3'-terminus and primer 5'-terminus

To allow self-tagging of efficient polymerase ribozymes the pool 3'-terminus was covalently linked to the primer 5'-terminus (see figure 1B). The crystal structure of the ligase ribozyme [145] suggested that a linker length of 15 nucleotides would be sufficient for this link. To identify linker sequences that would allow the ribozyme to be active, 15 randomized nucleotides were positioned between the ribozyme 3'-terminus and the primer 5'-terminus (see 'Generation of the polymerase ribozyme construct). The goal was not to identify the best linker sequence because such a sequence might have fine-tuned, and therefore modified the ribozyme conformation. Instead, three sequences were arbitrarily chosen from a large number of linker sequences that permitted ligation. To identify such linker sequences, one round of in vitro selection with 6sGTP was carried out. For this selection, a total amount of 3 pmol of the pool with the randomized linker sequence was heat-denatured in the presence of 50 mM Tris / HCl pH 8.3 at 80°C for 2 minutes followed by immediate chilling on ice for five minutes. The ligation was started by adding a premix such that the final concentration of all the components were 100 mM $MgCl_2$, 200 mM KCl, 1% PEG 20,000, 50 mM Tris / HCl pH 8.3, and 20 uM 6-thio GTP (Axxora). The mixture was incubated at 22 °C for 3 hours. The RNA was ethanol precipitated and separated by APM PAGE as described [152, 122, 153], with a total concentration of 240 uM immobilized aminophenyl-mercury in the APM layer. Note that the 10x stock solution was 2.4 mM APM in DMSO. The material at the APM interface was excised, eluted with 300 mM NaCl and 5 mM DTT, ethanol precipitated, reverse transcribed using Superscript III (NEB), PCR amplified, cloned into pUC19 plasmids, and sequenced. A total of 7 individual clones were tested for their ability to generate a signal at the APM interface when the procedure of the selection step was performed with individual ribozyme constructs that were internally labelled with a[32P]-ATP. The three best performing sequences were chosen for the following in vitro evolution (see below).

5.3.4 Evolution of ribozyme constructs with improved 6sGTP ligation efficiency

The polymerase ribozyme construct as shown under 'generation of the polymerase ribozyme construct' was subjected to 10 rounds of in vitro evolution for more efficient use of 6sGTP as substrate for ligation to its own 3'-terminus. In the first round of evolution, sequence diversity in the polymerase ribozyme sequence was generated by 30 cycles of mutagenic PCR [154]. The mutagenesis rate was about 1.0 mutations per 10 cycles of mutagenic PCR and per 100 mutagenized positions [155], consistent with the original publication of the procedure [156]. We previously found a bias among the six possible mutations with 23% AT->GC, 38% AT->TA, 18% GC->AT, 11% GC->TA, 6% AT->CG, and 5% GC->CG [155]. Of the 206 nucleotides in the polymerase ribozyme construct, 17 nucleotides on each end or the construct were fixed and not mutated during the evolution to allow primer binding for reverse transcription and PCR. In the following rounds of evolution, no mutagenic PCR was employed with exception of mutations generated by the intrinsic mutagenicity of Taq DNA polymerase.

During transcription of the pool RNA, the 5'-terminal hammerhead ribozyme removed itself, generating a 5'-hydroxyl group as described earlier [32]. Transcribed pool RNAs were purified via 5% PAGE and processed with the DNAzyme as described two sections above. During transcription, the pool RNA molecules were internally labeled by incorporating trace amounts of a[32P] ATP. This allowed locating the molecules at the APM interface for excision, as well as monitoring the fraction of pool reacting during the selective step, over the course of the evolution.

The incubation of the pool with 6sGTP was performed under the same conditions as used in the linker selection (see the previous section). The linker sequences in the ribozyme construct were alternated during the evolution to prevent the polymerase ribozyme from evolving specific sequences that interact with these linkers, with 5'-CGCCUAGACCCACGC-3' (rounds 1, 4, 7, 10), 5'-GCUCACACAAGAAAA-3' (rounds 2, 5, 8), and 5'-CAGAACUCCAAUAUA-3' (rounds

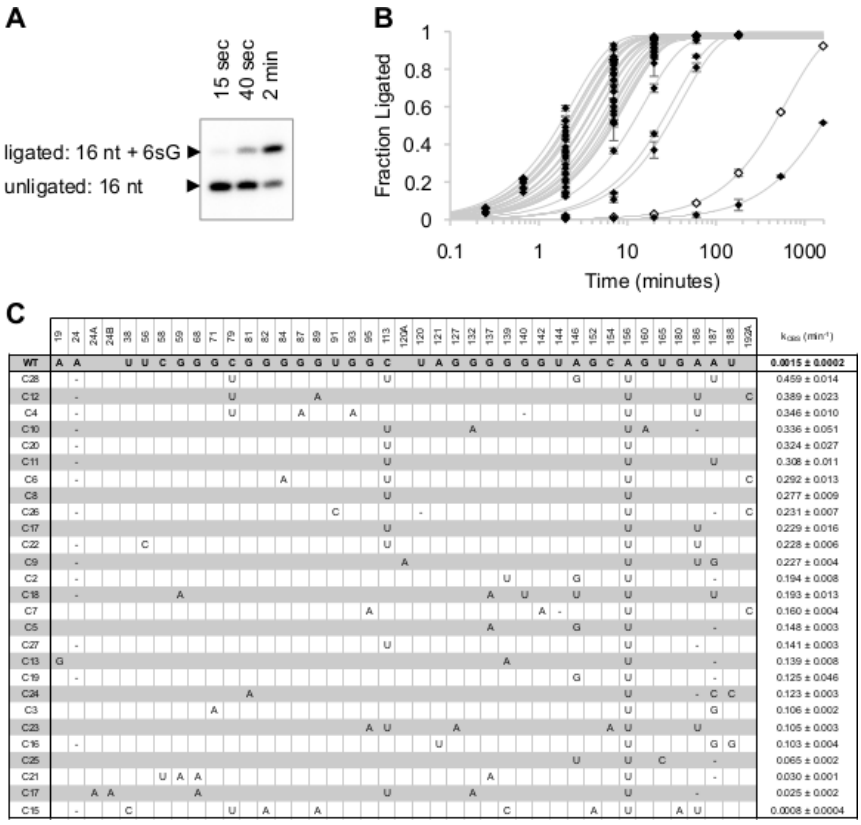
3, 6, 9). The concentration of 6-thio GTP, and the incubation time were decreased over the rounds of evolution to increase selection stringency (see figure S1). After reverse transcription of the selected RNAs, a first PCR reaction amplified the selected sequences with short PCR primers corresponding to the underlined sequence in section ”generation of the polymerase ribozyme construct”. A second PCR reaction used PCR primers with long 5’-extensions to re-generate the pool 5’-terminus with the promoter for T7 RNA polymerase, hammerhead ribozyme, and the pool 3’-terminus with linker sequence and DNAzyme target site. The resulting PCR product was the input for the next round of evolution.

After evolution round ten, the PCR products were appended with restriction sites, cloned into the plasmid pUC19, and sequenced. To test the activity of individual clones, the sequence of each clone was amplified as described in the first two paragraphs of this section.

5.3.5 Ligation assay

To measure the ligation rate for individual ribozyme sequences they were transcribed, processed at their 3’-termini, and purified as described above. The RNA substrate 5’-AUA AGU CUC ACC AAC U-3’ was transcribed with a 5’-terminal hammerhead ribozyme, which removed itself during transcription as described before [32]. After 5’-radiolabeling with polynucleotide kinase, trace amounts of the radiolabeled substrate were dissolved with the ribozymes at 1.5 uM concentration in 500 mM Tris/HCl pH 8.3. After heat renaturing for two minutes at 80°C the samples were immediately transferred to ice for five minutes. The mixture was then diluted to a final concentration of 100 nM ribozyme, 50 mM Tris/HCl pH 8.3, freshly prepared 1% (w/v) PEG 20,000, 100 mM $MgCl_2$, 200 mM KCl, and 25 uM 6sGTP. At given time points 5 uL aliquots of the reaction were quenched with 5 uL formamide gel loading buffer containing an excess of Na_2EDTA over Mg^{2+} from the reaction. Samples were separated by denaturing 20% PAGE, exposed to phosphorimager screens, then scanned and quantitated on a Typhoon phosphorimager (GE) using the Quantity One software (Bio-Rad). Rate constants were obtained

by single-exponential curve fitting to the data in Microsoft Excel, using the solver add-on. For data in figures 2-4, the amplitude for fitted curves was set to 100% because all constructs with fast ligation kinetics approached 100% in saturation. For data in figure 6, the amplitude was set to 80% because all constructs with fast kinetics approached 80% in saturation. The difference between these amplitudes appeared to be in the preparation of the substrates. Ligation rates for canonical NTPs were measured exactly the same, only by substituting 6sGTP with the respective NTP. The k_{obs} for C28 was higher in the initial screen than in the side-by-side comparisons with its variants because in the initial screen included very short timepoints (15 seconds, 45 seconds) for the fastest variants. In contrast, the analysis of the variants of C28 started at 2 minutes. Because the kinetics of clone 28 deviated somewhat from first order kinetics the inclusion or omission of the first two timepoints led to slightly different values for k_{obs}.



	19	24	24A	24B	38	56	58	59	68	71	79	81	82	84	87	89	91	93	95	113	120A	120	121	127	132	137	139	140	142	144	146	152	154	156	160	165	180	186	187	188	192A	k_{obs} (min^{-1})
WT	A	A			U	U	C	G	G	G	C	G	G	G	G	G	U	G	G	C		U	A	G	G	G	G	G	G	U	A	G	C	A	G	U	G	A	A	U		**0.0015 ± 0.0002**
C28		-									U									U									G				U						U			0.459 ± 0.014
C12		-									U				A																			U				U			C	0.389 ± 0.023
C4		-									U			A			A										-							U				U				0.346 ± 0.010
C10		-																		U				A										U	A		-					0.336 ± 0.051
C20		-																		U														U								0.324 ± 0.027
C11		-																		U														U					U			0.308 ± 0.011
C6		-												A						U														U							C	0.292 ± 0.013
C8																				U														U								0.277 ± 0.009
C26		-															C					-												U					-		C	0.231 ± 0.007
C17																				U														U				U				0.229 ± 0.016
C22		-				C														U														U				U				0.228 ± 0.006
C9		-																			A													U				U	G			0.227 ± 0.004
C2		-																									U			G				U				-				0.194 ± 0.008
C18		-						A																		A		U		U				U					U			0.193 ± 0.013
C7																			A										A	-				U							C	0.160 ± 0.004
C5																										A				G				U				-				0.148 ± 0.003
C27		-																		U														U				-				0.141 ± 0.003
C13	G																									A								U				-				0.139 ± 0.008
C19		-																												G				U				-				0.125 ± 0.046
C24												A																						U				-	C	C		0.123 ± 0.003
C3										A																								U					G			0.106 ± 0.002
C23																			A	U				A								A	U				U					0.105 ± 0.003
C16		-																					U											U					G	G		0.103 ± 0.004
C25																														U				U		C		-				0.065 ± 0.002
C21							U	A	A																	A								U				-				0.030 ± 0.001
C17			A	A					A											U				A										U				-				0.025 ± 0.002
C15		-			C						U		A		A												C				A		U			A	U				0.0008 ± 0.0004

Figure 5.2: Analysis of 27 arbitrarily chosen isolates from round ten of the in vitro evolution. (A) Representative phosphorimage of PAGE-separated ligation products of the most active clone 28 with 25 uM 6sGTP. This reaction did not use a linker between the primer 5'-terminus and the ribozyme 3'-terminus, therefore the primer was elongated from 16 nt to 17 nt. The reaction times as well as the identity of the signals are indicated. (B) Ligation kinetics of the 27 ribozymes at 25 uM 6sGTP. Time points were taken at 2, 7, 20, 60, and 180 minutes. Additional samples were taken for weakly active clones (WT and C15) at 9 hours and 27 hours, and for highly active clones at 15 and 40 seconds. Error bars are standard deviations from triplicate experiments. (C) Correlation between ribozyme mutations and their 6sGTP ligation kinetics. All positions that show at least one mutation in the 27 isolates are given on the top. The clone name is given on the left. Mutations relative to the wild type (WT) sequence of the R18 ribozyme are given in the central part. For WT the unmutated nucleotide is shown, while all other clones show only the mutated nucleotide. An empty space in the WT row with nucleotides for the isolates indicates insertion events in at least one of the selected sequences. The clones are sorted according to their ligation kinetics, with values and their standard deviations given in the right column. Note that the values of these rates are slight overestimates due to the reaction time points used in this screen.

5.4 Results

To identify the NTP binding site of the polymerase ribozyme, the ribozyme was evolved for efficient use of 6-thio guanosine triphosphate (6sGTP) as substrate instead of canonical NTPs.

The rationale was that the polymerase ribozyme had never encountered 6sGTP in its evolutionary history. Therefore, evolutionary optimization of the ribozyme for efficient use of 6sGTP was expected to lead to beneficial mutations especially at those nucleotides that were responsible for binding NTPs. The construct for the polymerase ribozyme evolution was generated such that

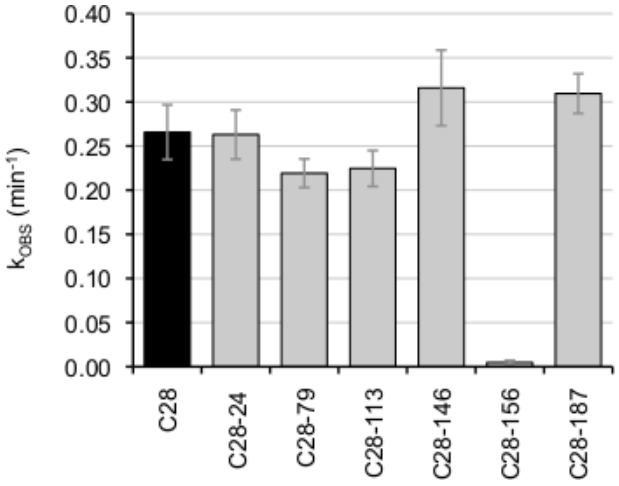


Figure 5.3: Analysis of the most efficient isolate of the evolution, clone 28. Shown are the effects on the ligation rate, by individually removing each of its six mutations. The name of each polymerase ribozyme variant is given below, where C28 indicates all six mutations and the additional number indicates which of the six mutations was removed. The observed rates were obtained by single-exponential fitting to reaction time courses. Removing mutations C79U or C113U resulted in a 1.2-fold reduction, and removing mutation A156U resulted in a 56-fold reduction. Error bars are standard deviations from three independent experiments.

successful polymerase ribozymes would tag their own 3'-terminus with 6sGTP (figure 1B). To do this, the R18 polymerase ribozyme was modified in two ways. First, the primer binding mode was converted to the format of its evolutionary ancestor, the class I ligase ribozyme [106] such that the primers 3'-terminus was paired near the ribozyme's active site, using the neighboring C nucleotide as templating base. This primer binding mode is structurally equivalent to the trans binding mode of the polymerase ribozyme, based on experiments tethering the primer-template duplex to different positions of the polymerase ribozyme [149]. Second, the 3'-terminus of the polymerase ribozyme was linked to the 5'-terminus of the primer. Three different linker sequences were identified by a single low-stringency selection step from a pool of linkers with 15 randomized

nucleotides (see materials and methods). To generate 3'-hydroxyl termini for the new, elongated ribozyme constructs, transcripts were processed using a catalytic DNA (DNAzyme) that site-specifically cleaves RNA, leaving the upstream cleavage product with a 3'-hydroxyl group [151]. In this ribozyme construct a 3'-elongation of the primer with 6sGTP would tag the 3'-terminus of the ribozyme with a sulfur-containing nucleotide. This sulfur tag was used to separate active from inactive polymerase ribozyme variants by denaturing polyacrylamide gel electrophoresis with covalently immobilized aminophenyl mercury (APM-PAGE) [152, 122, 153]l. This setup allowed the selection of active ribozymes from pools of ribozyme variants.

The starting pool for the evolution was prepared by mutagenic PCR, introducing mutations over the entire length of the polymerase ribozyme [154, 156]. Subsequent rounds of evolution did not use explicitly mutagenic conditions for PCR. However, Taq polymerase - which is not a high-fidelity polymerase - was used for PCR amplification, which introduced additional mutations via its inherent error rate. The stringency of the selection step was successively increased over ten evolution rounds, from a 6sGTP concentration of 20 micromolar and an incubation time of 3 hours to a 6sGTP concentration of 0.5 micromolar and an incubation time of 15 minutes (figure S1). The fraction of pool selected at the interface of the APM gel suggested that pool activity increased 86-fold from evolution round one to ten.

To identify the most efficient ribozyme sequence for 6sGTP ligation, 27 clones from evolution round ten were tested for ligation activity with 6sGTP (figure 2A, B). Sorting the 27 sequences according to their activity showed that a small number of mutations were highly enriched in the most efficient clones, specifically mutations A156U, C113U, and C79U (figure 2C). The enrichment of mutations in the most active clones suggested that mutation A156U had the strongest effect on 6sGTP ligation, and C79U and C113U also contributed to high activity. The most efficient isolate was clone 28, which contained 6 mutations compared to the parent ribozyme that included all three enriched mutations A156U, C79U, and C113U.

To identify the mutations that were necessary for the high activity of clone 28 the six

mutations were individually removed (figure 3). The most dramatic effect came from mutation A156U, which dropped the activity of clone 28 by 92-fold when deleted. Two additional mutations appeared to be beneficial, mutations C79U and C113U, which each dropped activity of clone 28 by ˜1.3-fold when deleted. Mutation A24-delta did not appear to benefit 6sGTP ligation activity, and mutations A146G and A187U appeared to have none or a mildly deleterious effect on ligation kinetics. This showed that in the context of the other mutations in clone 28, only the three mutations A156U, C79U, and C113U were necessary for full activity. To test whether the three mutations A156U, C113U, and C79U alone were sufficient for full activity of clone 28 they were inserted into the wild-type sequence. The triple-mutant showed the same 6sGTP ligation kinetics as clone 28, within error, showing that A156U, C113U, and C79U were sufficient for a >200-fold improvement in the kinetics of 6sGTP ligation. Because these data were recorded at a 6sGTP concentration of 25 uM, the k_{obs} = 0.30 min-1 corresponds to a first-order rate constant of 200 M^{-1} s^{-1}.

To test the contribution of each mutation in the triple mutant, all intermediates between the wild-type sequence and the triple mutant were constructed, and their 6sGTP ligation kinetics were measured (figure 4). The beneficial effect of mutation A156U was 69-fold to 199-fold, depending on the mutational background, and effects were less than 3-fold for mutations C113U and C79U. The simplest explanation for the strong effect of A156U is that the nucleobase of A156 in the purine-rich loop is directly involved in NTP recognition. The mutations A156U and C113U acted weakly cooperative, as did the mutations A156U and C79U (figure 4). In the absence of C113U, mutation A156U increased activity 69-fold and 83-fold, while in the presence of C113U, A156 increased activity 149-fold and 199-fold. This is a 2.2-fold and 2.4-fold cooperative effect, respectively. Similarly, mutation C113U increased the 6sGTP ligation rate 1.6-fold and 1.8-fold in the presence of A156U, while decreasing the rate 1.3-fold and 1.4-fold in the absence of A156U, confirming the cooperative effect between these two mutations. Mutation A156U also acted weakly cooperative with mutation C79U: The beneficial effects of mutation A156U in the

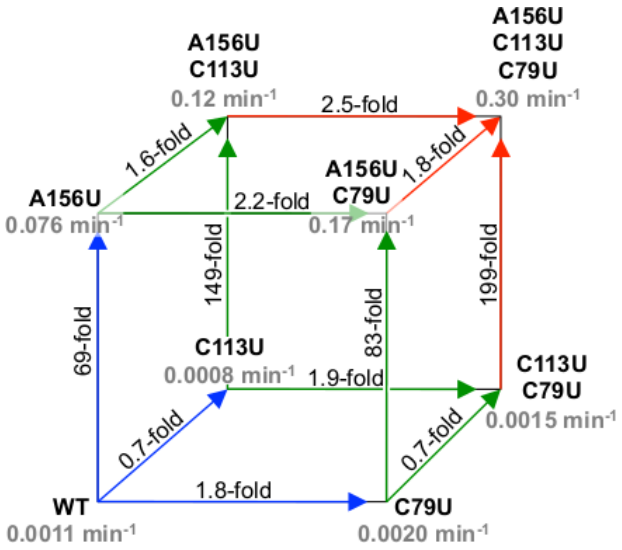


Figure 5.4: Analysis of cooperative effects between the three beneficial mutations. The 'kinetic cube' shows all intermediates from the wild-type (WT) to the triple mutant (A156U, C113U, C79U), with the 6sGTP ligation rate at 25 uM 6sGTP indicated in grey. The paths of the first mutation from the WT are labeled in blue, of the second mutation in green, and of the third mutation in red. For each path the -fold increase in ligation kinetics is indicated. The paths up correspond to A156U mutations, the paths back to the C113U mutation, and the paths right to C79U mutations. The comparison of the four different effects for each mutation showed whether mutations acted cooperatively (see text).

presence of C79U (83-fold and 199-fold) are 1.2-fold and 1.3-fold stronger than in the absence of C79U (69-fold and 149-fold). Similarly, the beneficial effects of C79U in the presence of A156U (2.2-fold and 2.5-fold) were 1.2-fold and 1.9-fold stronger than in the absence of C79U. There was no cooperative effect between mutations C113U and C79U, with influences of 1.0-fold and 1.1-fold. Together, these results suggested a network of direct or indirect interactions between the three nucleotides: A156 contacts the incoming NTP, and A156 interacts directly or indirectly with C113 in the P2 duplex, and with C79 in the body of the ligase core. The nucleobases of C79 and C113 did not seem to interact.

The simplest structural explanation for these cooperativity data is that A156 directly interacts with the incoming NTP, and that C79 as well as C113 interact with A156 but not with each other (figure 5). The direct interaction between A156 and the incoming NTP must be with the Hoogsteen face of the incoming NTP because only this side of the NTP is accessible on the

surface of the catalytic core (13). The nucleobase of A156 would have to be positioned into a narrow cleft, perhaps benefitting from stacking interactions with the P2 helix.

If there is a direct contact between position 156 and the incoming NTP then mutations in position 156 should modulate the preference for NTPs. To test this, the NTP ligation rates were measured for different polymerase ribozyme constructs (figure 6). In the background of mutations C79U and C113U, the NTP ligation rates were compared to the construct with the wild-type A in position 156 (figure 6D). While the mutant containing A156G showed a similar NTP preference as A156C, there were strong differences between A156, A156U, and A156G/C. If position 156 would act via influencing catalysis then the influence on NTP ligation kinetics would be expected to be similar between the NTPs. Therefore, the different NTP preference patterns caused by mutations in position 156 confirm that position 156 is directly involved in binding the incoming NTP.

The crystal structure of the ligase core confines the possible contacts of the purine-rich loop to the Hoogsteen face of the NTP [145] because the Watson-Crick face is involved in base pairing to the templating nucleotide, and the minor groove is inaccessible due to the position of the accessory domain (figure 5). The Hoogsteen face of 6sGTP displays the thio-modification, which largely populates the thiol form under the used reaction conditions ([157] and figure S2). The 6-thiol may act as a hydrogen donor similar to the exocyclic amino group of adenine, which would explain why the mutation A156U prefers both 6sGTP and ATP but discriminates against GTP, UTP, and CTP (figure 6D). We did not identify a similarly simple explanation for the NTP preference of mutations A156G and A156C.

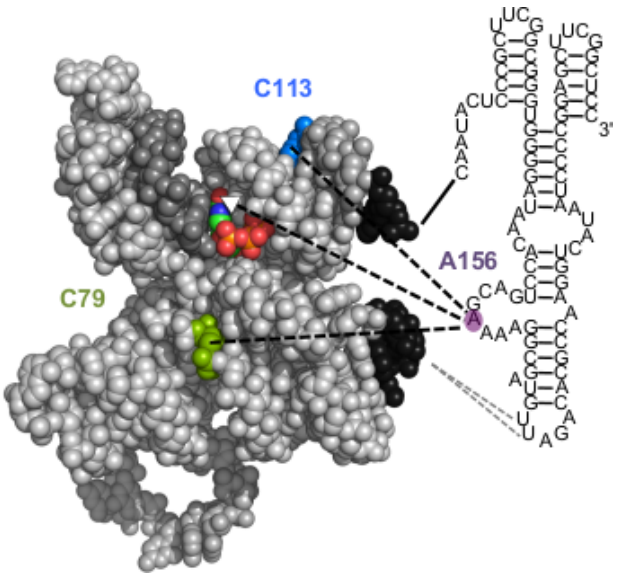


Figure 5.5: Model for the interaction between A156 (purple background) and the ligase core (grey). The NTP in the active site (atom colors), position C79 (green), and position C113 (blue) are highlighted. The results of this study suggest a direct interaction between A156 and the NTP, and indirect interactions (dashed lines) between A156 and C79, and between A156 and C113. The white arrowhead of the dashed line between A156 and the NTP points towards position 6 of the NTP nucleobase, which may serve as hydrogen donor to position156 (see text). Note that the X-ray structure corresponds to GTP in the active site therefore the red sphere (oxygen) corresponds to the 6-thio position of 6sGTP.

5.5 Discussion

The results of this study showed that mutation A156U had a >50-fold, beneficial effect on binding of 6sGTP, and that mutations C79U and C113U, with less than 3-fold beneficial effects, acted weakly cooperative with A156U but not with each other. A direct interaction between NTP and position 156 was confirmed by changes in the NTP preference patterns due to mutations in position 156. The three proposed interactions (A156-NTP, A156-C79, and A156-C113) are consistent with the crystal structure of the ligase core, crosslinking data between loop AL4 of the accessory domain and J3/4 of the ligase domain, and mutational data of the AJ3/4 loop [149]. In this bigger picture, the accessory domain is draped over the tip of the 'tripod structure' of the ligase core, with the purine-rich loop AJ3/4 of the accessory domain positioned to help stabilize an NTP in the active site [149]. The accessory domain does not form a straight connection

between its two contact points to the catalytic core (black sequences in figure 5) but rather needs to curve to the NTP binding site. This curvature is consistent with the requirement for the bulged A164, which distorts the A4 helix (figure 1A) to correctly dock with the ligase core [149].

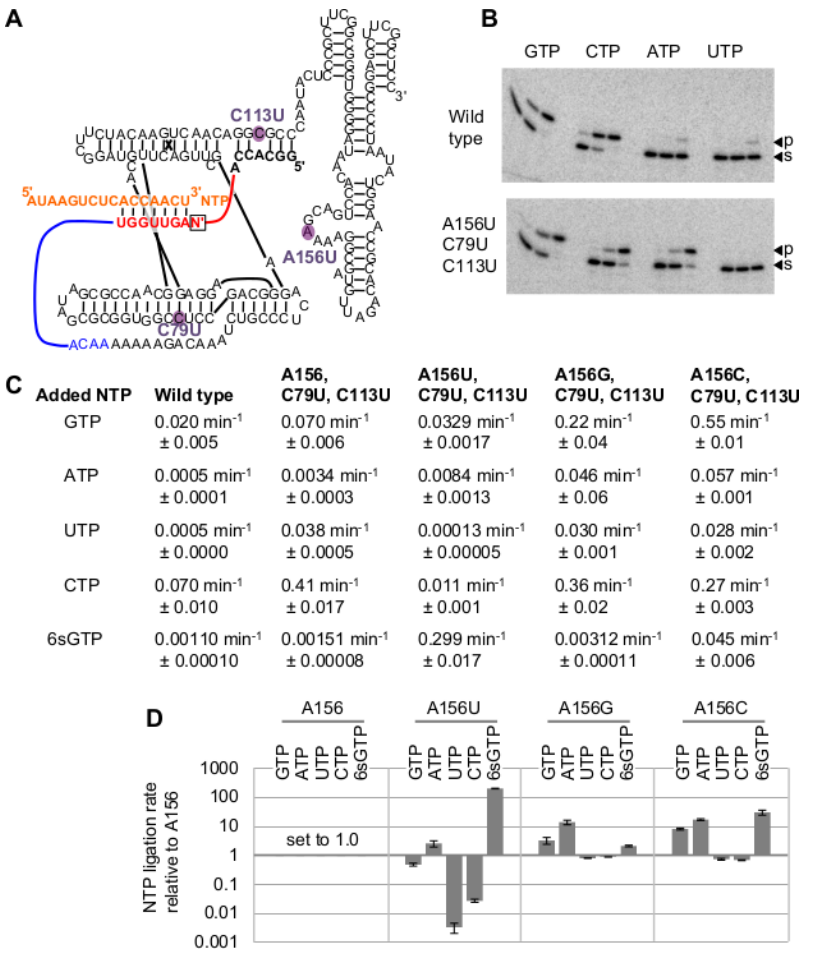
The weak cooperativity between mutations A156U and C113U is consistent with A156U being in physical contact with the incoming NTP because C113 resides in the P2 helix, adjacent to the catalytic site (figure 5). Because C113 is positioned on the opposite side of the P2 duplex as the incoming NTP (and therefore A156) we suggest that the cooperative effect between A156U and C113U results from the stabilization of the A-form duplex by the C113U mutation, by converting a C:A pair to a U:A pair. In contrast, a C:A mispair in the P2 helix makes the polymerization of multiple nucleotides more efficient [158], suggesting that a deformation in the A-form of the P2 helix is necessary for renewed binding of NTPs, or for release and re-binding of the primer/template duplex [147]. The weak cooperative effect between A156U and C79U is consistent with the purine-rich loop positioning the NTP into the active site because C79 is very close the NTP in the catalytic site: The distance of 12 between the 2'-oxygen of C79 and a non-bridging oxygen in the -phosphate of the NTP [145] could easily be spanned by one or two nucleotides of the eight-nucleotide purine-rich loop.

Two of the constructs tested in figure 6 show absolute rates of NTP ligation that are higher for all canonical NTPs when compared to the wild type ('round 18') polymerase ribozyme. Variant A156/C79U/C113U and variant A156G/C79U/C113U show higher rates for the addition of all canonical NTPs. In this case the major benefit comes from mutations C79U and C113U because no mutation is necessary for position 156. The mutation C79U was identified earlier, as one of four mutations that improved the efficiency of the polymerase ribozyme [69]. Note that mutation C113U is linked to the context of ligating single NTPs because C113U removes a mismatch in the P2 helix, which harms the function of the polymerase ribozyme [158].

To test whether the three evolved mutations were beneficial in the context of RNA polymerization in trans, the R18 polymerase ribozyme [34] was compared to the identical

construct with the three mutations C79U, C113U, and A156U in a polymerization assay. The results showed no detectable polymerization products for the triple mutant (data not shown). This was expected because the tested template selected for UTP as the first incoming NTP, which was bound much weaker by the triple mutant than by the wild-type sequence in the context of NTP ligation. Together, these data confirmed that the mutation A156U was a specific adaptation for 6sGTP binding.

This study showed that the sequence of the purine-rich loop has a dramatic effect on the efficiency of the polymerase ribozyme, via NTP binding. It may now be possible to develop more efficient variants of the polymerase ribozyme by replacing the purine-rich loop with a completely randomized sequence and re-selecting from this pool. The small size of this loop would allow complete coverage of all sequence variants so that loops with different sizes could be explored in the same experiment. Similarly, the results in this study can improve the use of the polymerase ribozyme as tool. A very recent study showed that the polymerase ribozyme can be used for 3'-tagging of RNA transcripts with chemically modified NTPs [159]. By re-selecting the most efficient purine-rich loop sequences for 3'-tagging with chemically modified NTPs, variants of the polymerase ribozyme would emerge for efficient 3'-tagging with different chemical modifications. The only difference from the selection system in this study would be that the selection step would have to be adjusted for each chemically modified NTP: While thio-modified NTPs could be selected with the same APM-PAGEs used in this study, biotinylated NTPs could be captured by streptavidine beads, and azido modified NTPs could use click chemistry for the selection step.



Added NTP	Wild type	A156, C79U, C113U	A156U, C79U, C113U	A156G, C79U, C113U	A156C, C79U, C113U
GTP	0.020 min^{-1} ± 0.005	0.070 min^{-1} ± 0.006	0.0329 min^{-1} ± 0.0017	0.22 min^{-1} ± 0.04	0.55 min^{-1} ± 0.01
ATP	0.0005 min^{-1} ± 0.0001	0.0034 min^{-1} ± 0.0003	0.0084 min^{-1} ± 0.0013	0.046 min^{-1} ± 0.06	0.057 min^{-1} ± 0.001
UTP	0.0005 min^{-1} ± 0.0000	0.038 min^{-1} ± 0.0005	0.00013 min^{-1} ± 0.00005	0.030 min^{-1} ± 0.001	0.028 min^{-1} ± 0.002
CTP	0.070 min^{-1} ± 0.010	0.41 min^{-1} ± 0.017	0.011 min^{-1} ± 0.001	0.36 min^{-1} ± 0.02	0.27 min^{-1} ± 0.003
6sGTP	0.00110 min^{-1} ± 0.00010	0.00151 min^{-1} ± 0.00008	0.299 min^{-1} ± 0.017	0.00312 min^{-1} ± 0.00011	0.045 min^{-1} ± 0.006

Figure 5.6: Influence of mutations in the polymerase ribozyme on the ligation kinetics of canonical NTPs and 6sGTP. (A) Secondary structure representation of the polymerase ribozyme constructs used for this assay. The three positions differing between wild type and triple mutant are highlighted in purple, with a purple label showing the mutation. The primer (orange) is base paired to the template strand (red) and is extended by the incoming NTP (red). In each experiment, the templating nucleotide (boxed) is adjusted to form a Watson-Crick pair with the NTP. The sequence linking the R18 polymerase ribozyme 5'-terminus with the template 3'-terminus is shown in blue, and the red line shows the connection between template 5'-terminus and the 3'-terminus of the 5'-GGCACCA sequence in the P2 stem. (B) Autoradiogram of radiolabeled ligation products that were separated by denaturing 10% PAGE. The NTPs used for each reaction are shown on the top. The three lanes for each NTP stem from samples with the incubation times of 2 minutes, 20 minutes, and 180 minutes. The lower band (labeled with 's' on the right) is the substrate, the upper band (p) is the product, which is one nucleotide longer than the substrate. The upper and lower audioradiogram show the signals from the wild type sequence and the triple mutant A156U / C79U / C113U, respectively. (C) Observed kinetic rates determined from triplicate experiments as shown in B, using single-exponential curve fitting to quantified data. (D) NTP ligation rates relative to the variant A156 / C79U / C113U. All variants contain the C79U and C113U mutation, and only the nucleotide in position 156 is varied. The NTP used for each experiment is given above each column. Error bars correspond to standard deviations from triplicate experiments.

5.6 Supplementary Data

Supplementary figures S1, and S2 are available at NAR online.

5.7 Funding

This work was supported by the National Aeronautics and Space Administration under [grant/cooperative agreement NNX16AJ27G] issued through the Science Mission Directorate, in Astrobiology / Exobiology to U.F.M.. A fellowship from the Department of Education Graduate Assistance in Areas of National Need (GAANN; award number P200A150251) provided partial support to A.A.

5.8 Acknowledgements

We thank Yitzhak Tor for helpful discussions.

Chapter 5, in full, is a reprint of the material as it appears in Nucleic Acids Research. Akoopie, Arvin; Müller, Ulrich F. 2018. The dissertation author was the first author of this paper.

Chapter 6

In emulsio selection of a nucleoside triphosphorylation ribozyme

6.1 Abstract

The RNA world hypothesis describes a stage in the early evolution of life in which RNA served as genome and as the only genome-encoded catalyst. To test how an RNA world organism could have functioned, researchers have used in vitro selection methods to generate catalytic RNAs (ribozymes) that catalyze reactions of central importance. One such reaction is the synthesis of chemically activated nucleotides. These are necessary to support RNA polymerization in aqueous medium. Here we show the development of a ribozyme that generates NTPs from nucleosides and the prebiotically plausible molecule trimetaphosphate. The development was technically challenging because this ribozyme would have to bind two small molecules (nucleosides and trimetaphosphate) and release them for multi-turnover catalysis. This challenge was addressed by selecting the ribozyme in emulsion droplets, from randomized sequence. Pool molecules generating thio-modified NTPs were tagged by a polymerase ribozyme variants attached to their 3'-terminus, and isolated on aminophenyl-mercury polyacrylamide gels. After 18 rounds of selection,

high throughput sequencing analysis identified the most successful sequences. Biochemical analysis suggested that the selected ribozymes catalyzed nucleoside triphosphorylation, with only one or a few NTPs generated by each selected ribozyme. The results are discussed in the context of the RNA world hypothesis.

6.2 Introduction

The RNA world hypothesis states that there was a stage in the early evolution of life in which RNA served both as genome and as the only genome-encoded catalyst [142, 11, 13, 12]. However, direct evidence of RNA world organisms is unlikely to be found because they likely decomposed completely over the last >3 billion years. To test how RNA world organisms could have functioned, in vitro selection [42, 43] was used to develop catalytic RNAs (ribozymes) that could have fulfilled central functions [33, 51]. The most central function in an RNA world organism would have been a ribozyme that catalyzes templated RNA polymerization to mediate the replication of the organism. An RNA-dependent RNA polymerase ribozyme has been developed by in vitro selection and design [34]. While the original ribozyme was far too inefficient for self-replication, advanced variants are getting closer to the possibility of self-replication [68, 69, 35].

The polymerization of nucleotides is thermodynamically unfavorable in aqueous environment. Different chemical activation groups for the nucleoside 5'-phophate group have been discussed and tested, with an initial focus on 2-methyl imidazolides [56], and more recently the much more active 2-amino imidazolides [160]. However, the synthesis of phosphorimidazolides requires high-energy intermediates, which puts into doubt their prebiotic plausibility [14]. Prebiotically more plausible are polyphosphates, and especially their most active member trimetaphosphate [161]. Trimetaphosphate can be generated in multiple prebiotically plausible ways, including from volcanic heating of basalt [27], by heating phosphate salts in the presence

of urea [30], by the erosion of meteoritic phosphide minerals to phosphite and polymerization under mild oxidative conditions [28], and by the reduction of phosphate to phosphite by ocean sediments containing Fe(II) [162] and subsequent polymerization [28]. Importantly, the prebiotic existence of the central phosphite intermediate has been shown by direct observation of 3.4 billion year old marine sediments [29].

Trimetaphosphate is able to directly triphosphorylate nucleoside 5'-hydroxyl groups - but this reaction proceeds efficiently only at pH >12 [31]. However, the triphosphorylation of RNA 5'-hydroxyl groups can be catalyzed by ribozymes at neutral pH [32]. Such catalysts are relatively frequent in sequence space, with >300 independent, active sequences within 10^{14} RNAs containing a random region of 150 nucleotides [163]. One of these ribozymes catalyzes rates of 6.8 min-1 under optimal conditions [73], and can be assembled from fragments no longer than 34 nucleotides [164]. However, these previously selected ribozymes triphosphorylate only their own 5'-hydroxyl group and not free nucleosides [32].

Nucleoside 5'-triphosphates (NTPs) are used in every known biological organism as central energy currency and as activated monomer for RNA polymerization, among many other roles [165]. This makes it likely that at a least at a late stage, RNA world organisms also used NTPs as chemically activated nucleotides. Correspondingly, studies on ribozyme-mediated RNA polymerization focused on NTPs as activating groups [34, 68, 69, 35, 146]. Therefore, a ribozyme that generates free NTPs from free nucleosides and trimetaphosphate would be desirable to generate an RNA world organism in the lab. However, the in vitro selection of active RNAs relies on tagging active RNA sequences so that they can be isolated from large, random pools [42, 43]. To tag RNA sequences that generated free NTPs, two conditions have to be met: First, the isolation of active sequences requires that the produced NTPs can be used to tag the pool molecules with a chemical handle. Second, because free NTPs diffuse easily between pool molecules, the selection step would have to proceed in compartments that prevent the diffusion of NTPs between pool molecules.

Here we show the in vitro selection of ribozymes that catalyze the reaction between the nucleoside 6-thio guanosine (6sGsn) and cyclic trimetaphosphate (Tmp) to form the NTP 6-thio guanosine triphosphate (6sGTP). The selective step proceeded in a water-in-oil emulsion and covered 1.6 x 10^{14} different sequences with a random region of 150 nucleotides. Active sequences were isolated via thio-tagging the pool molecules and separation on denaturing aminophenyl-mercury polyacryamide gels (APM-PAGE) [122]. After 18 rounds of selection and High Throughput Sequencing analysis the five dominant clusters were analyzed biochemically. The results are discussed in the context of the RNA world hypothesis.

6.3 Results

6.3.1 In vitro selection

An in vitro selection procedure was designed to obtain ribozymes that catalyze the triphosphorylation of 6'-thio guanosine (6sGsn) with trimetaphosphate (Tmp) to obtain 6'-thio guanosine triphosphate (6sGTP) (Figure 1). First, a double-stranded pool was generated that contained the promoter for T7 RNA polymerase, a hammerhead ribozyme, a 5'-terminal constant region, 150 base pairs of randomized region, a 3'-terminal constant region, and a recognition site for a DNAzyme. These pool molecules are transcribed using T7 RNA polymerase, during which the hammerhead ribozyme cleaves itself off co-transcriptionally and generate 5'-hydroxyl groups at the pool molecule. Because T7 RNA polymerase generates heterogeneous 3'-termini [116, 166, 118] and a precise 3'-terminus with 2',3'-diols was required for the selection procedure, transcribed RNA pool molecules were processed with a catalytic DNA that generates a 2',3'-hydroxyl terminus with a defined length [151]. The resulting pool molecule (5'-OH - constant region - N150 - constant region - 2',3'-diol) was used for the selection step.

For the selection step, the pool RNA molecules were heat renatured with an excess of a polymerase ribozyme variant that has been selected for efficient tagging of a RNA 2',3'-diol with

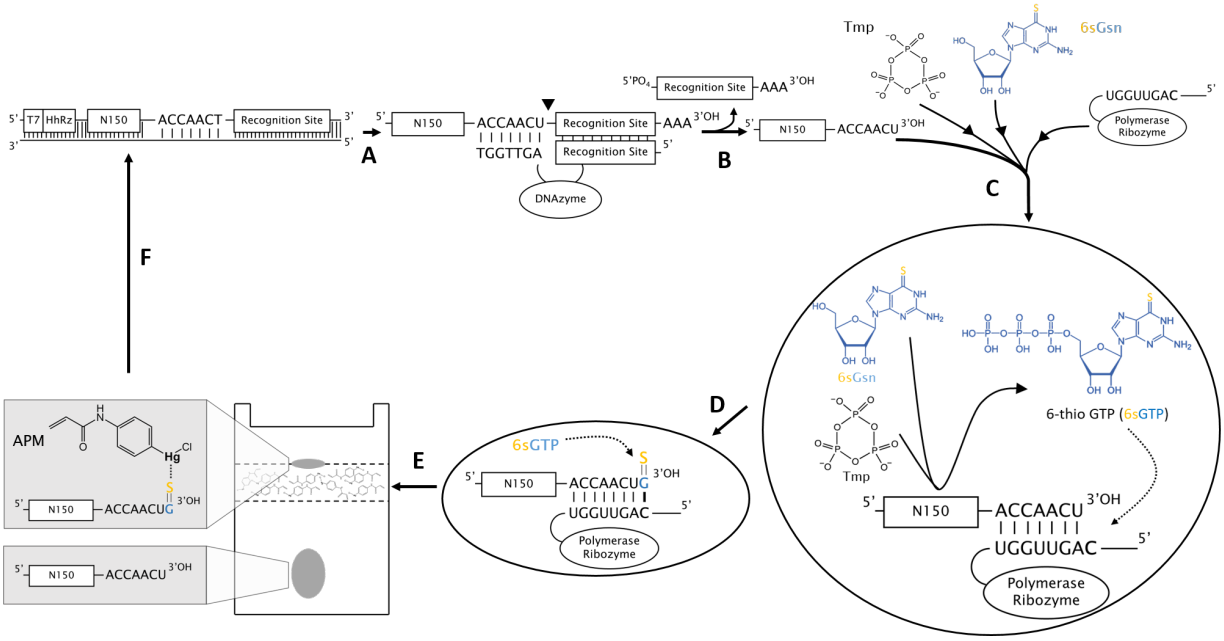


Figure 6.1: A. A pool molecule containing a T7 promoter, hammerhead ribozyme, 150 randomized resiudes, a constant primer region, and a recognition site are transcribed to RNA. The hammerhead ribozyme co-transcriptionally self-cleaves, leaving the pool molecule with a clean 5' hydroxyl end. B. The resultant transcript pool contains heterogeneous lengths of A at the 3' end due to the nature of T7 in vitro transcriptions. The pool is processed with a DNAzyme that specifically cleaves RNA in a mechanism that leaves the upstream product with a 2',3'-diol. C. The pool is mixed with trimetaphosphate, 6-thio guanosine and an evolved polymerase ribozyme. This mixture is immediately encapsulated in an emulsion such that the vast majority of droplets only contain one or no complex. If, by chance, the pool molecule is active for NTR activity it will triphosphorylate 6-thio guanosine using trimetaphosphate to generate 6-thio GTP. D. Any droplets that contain 6-thio GTP will be recognized by the polymerase ribozyme and ligated onto the 3' end of the pool molecule. Because the entire NTR-ligation reaction occurs in a water-in-oil emulsion droplet the now-thio modified pool molecule is the same one that generated the NTP. E. The emulsion is processed and the entire RNA pool is run on a modified APM-polyacrylamide (APM-PAGE). Any thio-modified RNAs are retained as soon as they encounter the APM interface. This exploits the extremely strong interaction between sulfur and mercury. Once run sufficiently long enough, the thio modified RNAs are separated enough from the unmodified ones to permit gel excision and RNA elution. F. The eluted RNAs are reverse transcribed and then PCR-amplified. This starts the next round of selection.

6sGTP [167]. Then, compounds were added to achieve final concentrations of 0.5 micromolar pool RNA, 3 micromolar polymerase ribozyme, 50 mM Tris/HCl pH 8.3, 150 mM $MgCl_2$, 200 mM KCl, 50 mM Tmp, and 1 mM 6sGsn. This aqueous phase was emulsified in an oil phase containing mineral oil and the emulsifier Abil EM90 [168]. The emulsion droplet diameter

was in the range of 150 nm (figure 5, figure 8), which resulted in an average of 3 polymerase ribozyme molecules per droplet, making sure that the large majority of RNA pool molecules was co-compartmentalized with at least one polymerase ribozyme molecule. Also, this droplet diameter determined that the generation of each single molecule of 6sGTP would raise the concentration of 6sGTP by about 1 micromolar. Since the used polymerase ribozyme displayed a 6sGTP ligation rate of 0.012 per minute and per micromolar under these ion conditions [167], even the production of a single 6sGTP molecule was expected to lead to the 3'-tagging of 50% of the co-compartmentalized pool molecules within less than 1 hour. The emulsion was incubated for 6 hours, and the pool RNA molecules extracted.

To isolate those pool RNA molecules that were tagged with 6sGTP at their 3'-terminus they were separated by a 3-layered APM PAGE [152, 153]. The thio modification of 6sGTP efficiently immobilized the pool molecules at the interface during APM-PAGE, as expected from a previous in vitro selection experiment using 6-thio guanosine [120]. After isolating the RNA pool molecules from the APM interface using DTT, they were reverse transcribed, PCR amplified, and appended with overhanging PCR primers to re-generate the pool, now enriched for active sequences.

6.3.2 Progress of the in vitro selection

Over the rounds of selection several parameters were adjusted (figure 6). In the first round of selection, the highest concentration of pool RNA molecules was used (0.5 micromolar) to be able to cover a large sequence space (1.6×10^{14} independent sequences). In subsequent rounds of selection the concentration of pool RNA was reduced to 50 nanomolar to reduce the likelihood of interactions between pool molecules. The first selection round employed 1 mM 6sGsn, close to its solubility limit, to capture even weakly active pool molecules. Upon an increase of the average pool activity the concentration of 6sGsn was reduced to 0.05 mM. For the same reason, the concentration of Tmp was reduced from 50 mM in the first round to 0.5 mM in the last round.

The average activity of the pool during the selection was followed by 5'-[^{32}P]-radiolabeling a fraction of the pool, and phosphorimaging of the APM gels. Importantly, the selection system was so sensitive that the background reaction between 6sGsn and Tmp at near-neutral pH led to detectable signals. This is because the assay employs 1,000-fold more 6sGsn (0.5 mM) than pool molecules (0.5 micromolar). This means that a conversion of 0.1% of 6sGsn to 6sGTP would have been sufficient to tag 100% of the pool molecules due to the high efficiency of the polymerase ribozyme. Because the retention of 0.1% of pool molecules at the APM interface was well within the quantifiable range, the assay allowed detecting 0.1% of 0.1% of 6sGsn molecules of 10-6 of the 6sGsn. By normalizing for the adjustments in the concentrations of 6sGsn and Tmp, an average pool ligation rate was calculated and plotted as function of the selection rounds (figure 2). After 12 cycles of selection, the pool showed a dramatic increase in activity, about 1,500-fold above the background rate of the reaction. Because activity plateaued in selection round 13 mutagenic PCR was used in round 14 to explore sequence variants of the initially selected sequences. The corresponding drop in pool activity in round 15 was followed by continued, high activity in selection rounds 16-18.

6.3.3 In emulsio selection

Over 18 rounds the randomized pool managed to enrich for activity over ˜3500-fold over the first round, after normalizing for selection pressure (Figure 2). Activity plateaued even after introducing mutations into the pool through mutagenic PCR on round 15. This is the first instance of selecting for an activity of a randomzied pool in emulsion during the entire selection. We have phrased this term, *in emulsio* selection.

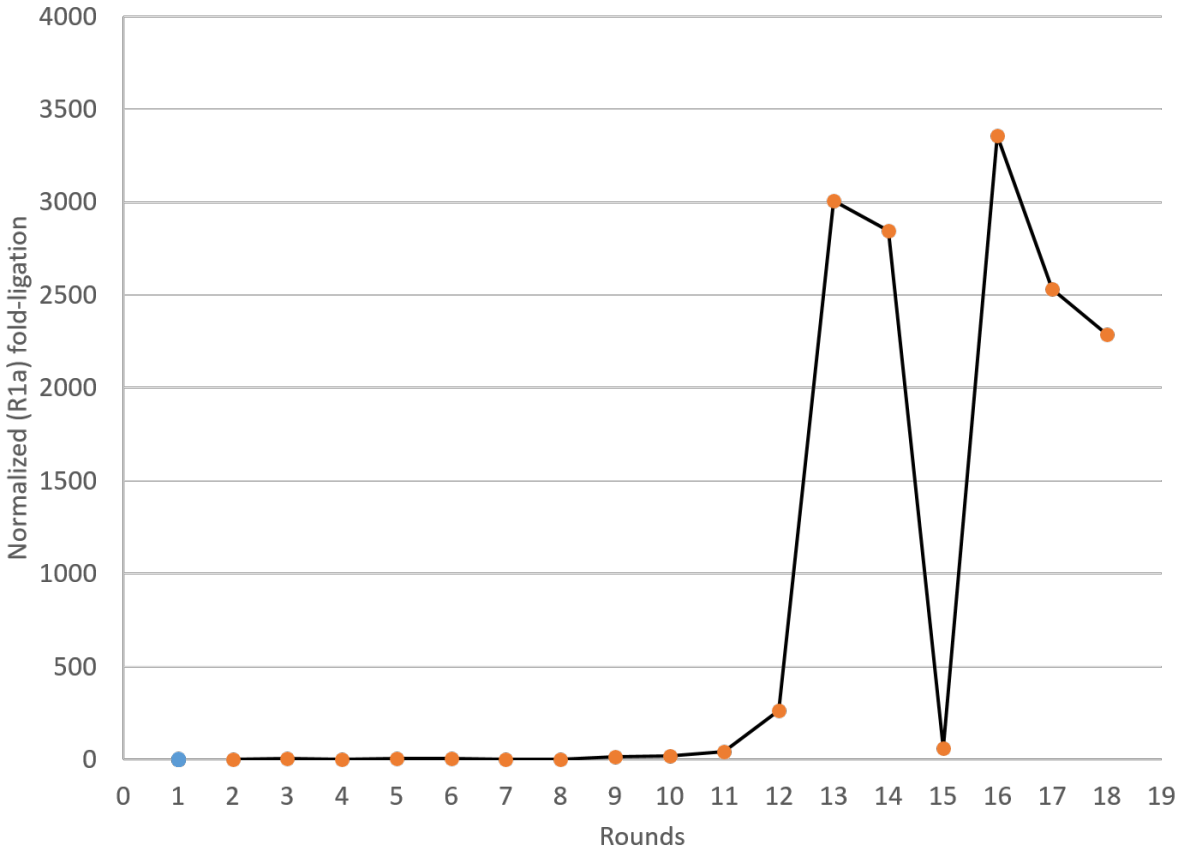


Figure 6.2: Activity of the pool on average through each round. Activity is normalized to the product of the ligation fraction with the relative concentrations of 6-thio guanosine and trimetaphosphate from round 1A. For example, if both round 1A and round 5 had the same ligation fraction but round 5 contained half the 6-thio guanosine and trimetaphosphate concentrations then round 5 would have a 4-fold improvement relative to round 1A.

6.3.4 High Throughput Sequencing analysis

To identify sequences that were highly enriched in the selection all pools were subjected to High Throughput Sequencing analysis. About 400,000 sequences were analyzed for each selection round. Sequences that showed more than 75% sequence identity were clustered, and the development of these clusters plotted as a function of the selection rounds (figure 3). In the final five rounds the pool was dominated by two clusters, c0 and c1. The HTS analysis suggested that the five clusters C0, C1, and C3 had the highest activity of all clusters (cluster C2 contained long stretches of G that likely caused the smear in the selection gels observed in the first rounds due to

G quadruplex formation). In each of the three clusters, the sequences most abundant in the last selection round were chosen for biochemical analysis (although additional sequences were tested they are not shown here because there was no significant difference in activity between the most abundant sequences in each cluster).

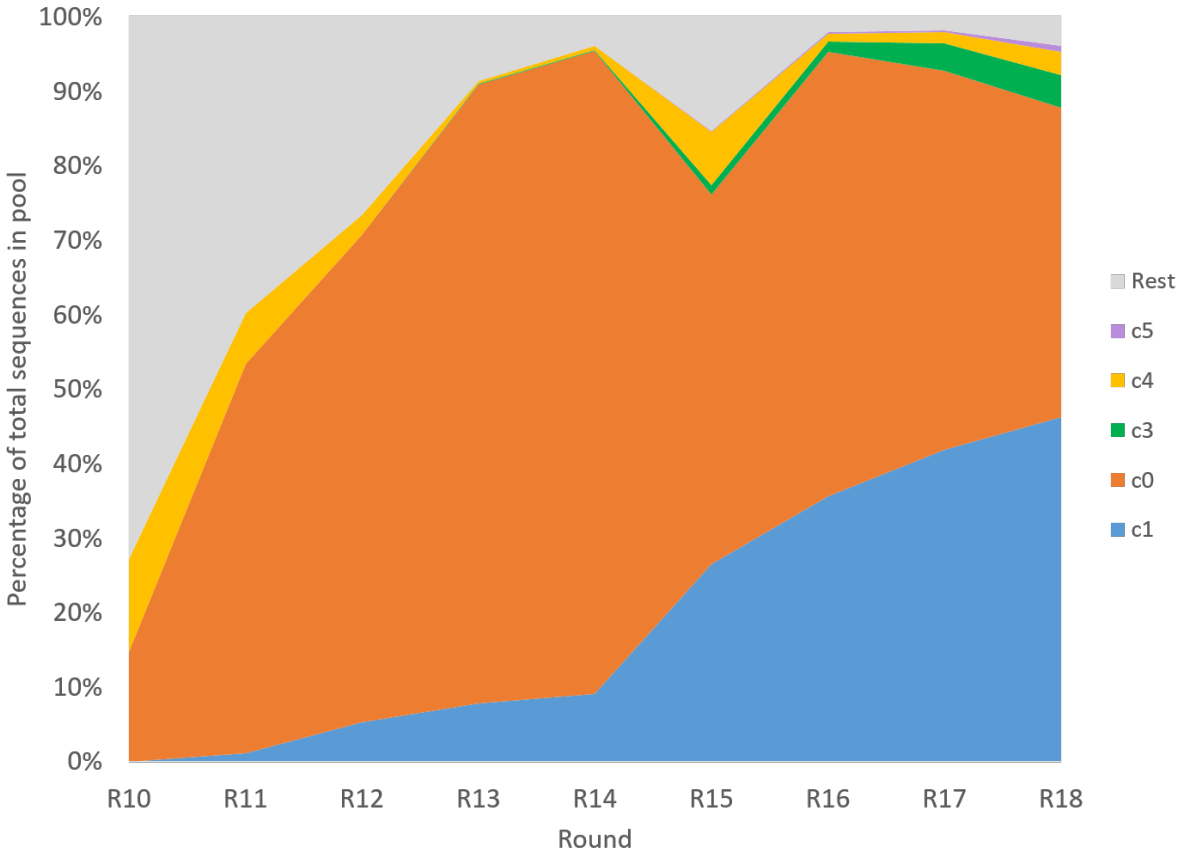


Figure 6.3: Enrichment of the top five clusters from round 18 throughout the selection. Note that this is a stacked graph showing the size (diversity) of each cluster as a percentage of the total number of unique sequences. This does not take abundance into account. Round 15 was a mutagenic PCR round, introducing a large number of unclustered sequences.

To identify the most active sequence variants within each of these clusters, the most frequent sequences were analyzed over the course of the last five rounds of the selection. The enrichment of individual sequences was measured as the slope with which their frequency in a given cluster increased over the selection cycles [163] (figure 7). Based on these data, four of the most promising sequences were generated from each of the three clusters. Additionally, one

additional sequence was generated for cluster 1 and cluster 3, which combined multiple mutations that appeared beneficial based on the HTS data.

6.3.5 Biochemical analysis of selected sequences

Two assays were used to evaluate the biochemical activity of the selected RNA clones. The first assay tried to replicate the selection system as closely as possible. Here, the RNA clone was radiolabeled, base paired to the polymerase ribozyme, and incubated with 6sGsn and Tmp. The reaction products were separated by denaturing APM-PAGE and the fraction of RNA clone at the APM interface was quantitated by phosphorimaging. In this assay, the RNA clones could affect the tagging of their 3'-terminus. Therefore, this assay was termed the cis-assay. In the second assay, a short, radiolabeled RNA substrate was base paired to a polymerase ribozyme, and the two RNAs were added to a mixture of unlabeled RNA clones base paired to a polymerase ribozyme. As in the other assay, the reaction products with 6sGsn and Tmp were separated by APM-PAGE, and the fraction of radioactivity at the APM interface was quantitated. In this assay, self-labeling of RNA clones was not visible because the RNA clones were not radiolabeled. Instead, the RNA clones needed to generate a diffusing compound that could be used to tag the 3'-terminus of the small, radiolabeled substrate. Because the reaction of this small, radiolabeled substrate was measured this assay was termed the trans-assay (figure 4).

The generation of 6sGTP from 6sGsn and Tmp was confirmed by the absence of signals when 6sGsn or Tmp were omitted (data not shown), and the 3'-tagging of pool molecules by the polymerase ribozyme with the generated 6sGTP was confirmed by a lack of signal when the RNA 3'-terminus was a 2',3'-cyclic phosphate instead of a 2',3'-diol, or when the catalytic center of the polymerase ribozyme was inactivated by the known C47U mutation [145]. *These data currently exist within different experiments and will be completed within a single experiment by the time of the defense.*

Mass Spectrometric analysis was used to test whether 6sGTP was produced by the selected

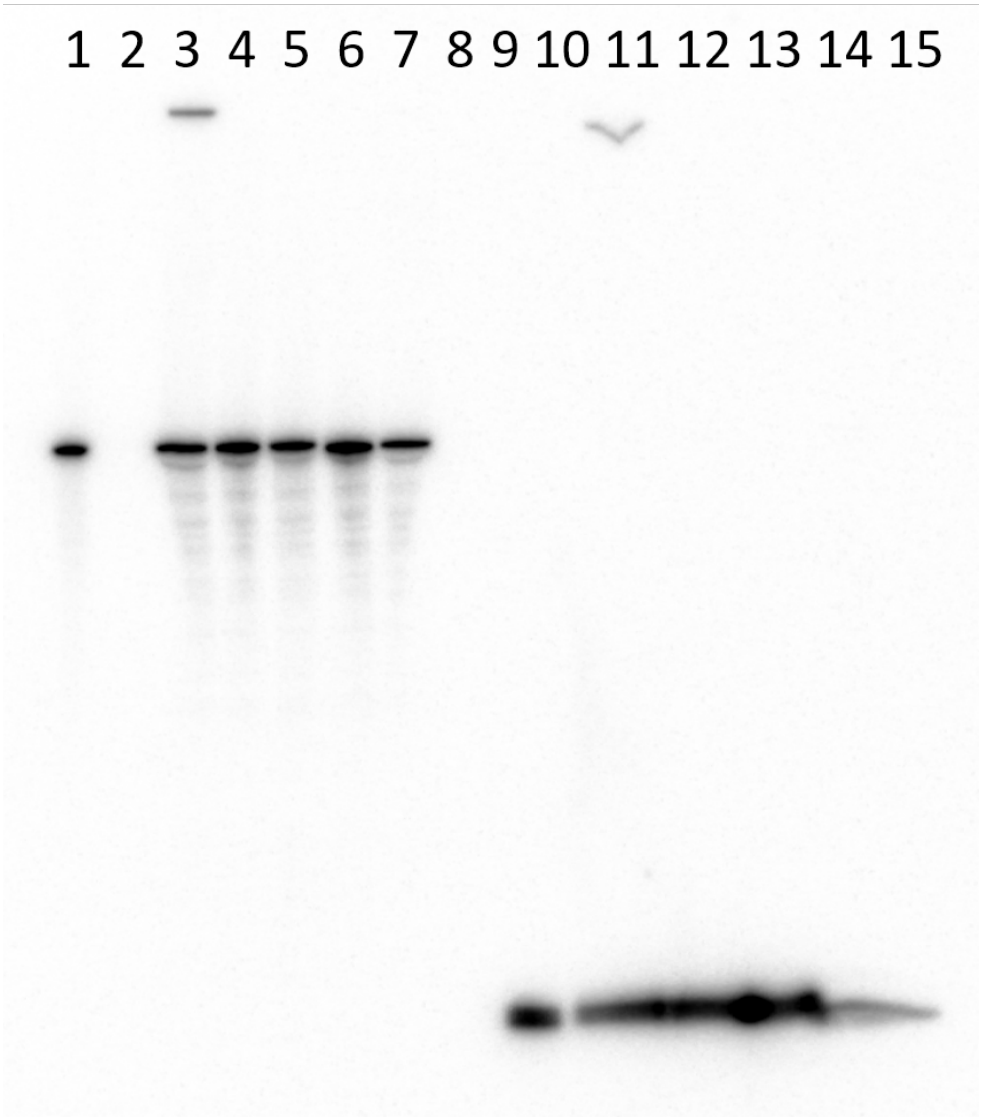


Figure 6.4: APM-PAGE of several conditions in the trans-NTR assay. The lanes are as follows: 1. NTR c0-15 marker 2. Blank 3. NTR c0-15 with a 3' hydroxyl 4. Same as 3 but no 6sGsn 5. Same as 3 but no Tmp 6. Same as 3 but an inactive polymerase ribozyme 7. NTR c0-15 with a cyclic phosphate end 8. Blank 9. Short KF substrate 10. Blank 11. IDT-prepared KF substrate with a 3' hydroxyl incubated with NTR c0-15 containing a cyclic phosphate end 12. Same as 11 but no 6sGsn 13. Same as 11 but no Tmp 14. Same as 11 but an inactive polymerase ribozyme 15. Same as 11 but the KF substrate contains a cyclic phosphate end. Only lanes 3 and 11 had detectable signals at the APM interface. Lane 3 had 10.5% ligated and lane 11 had 4.0% ligated, both after 6 hours.

ribozyme (figure 9). In short, two parallel lines of reactions were performed, one with and one without a specific NTR clone (c0-15). After six hours of incubation in 25 mM Tmp and 0.5 mM

6sGsn, both reactions were ethanol precipitated and redissolved at a lower volume. In addition, a third, positive control, sample was prepared that only contained 6sGTP. LC-ESI-MS identified where to locate the 6sGTP peak in the positive control. The sample without a NTR produced no peak. However, the sample containing NTR produced an LC peak within 0.02 min retention time. Subsequently, MS/MS on the prime 6sGTP peak (located at ˜540 m/z) showed similar "fingerprint" patterns between the positive control and the NTR sample (figure 9). Both samples contained a top peak at 168 m/z and had the same secondary peaks. Further experiments need to be done to reduce the noise introduced by the large amounts of salt and by optimizing the NTR to produce more NTPs.

6.4 Discussion

This study showed the in vitro selection of RNA catalysts from random sequence in emulsion. In vitro selection in emulsion has been used previously for selecting trans-active RNAs. However, all of these selections used the emulsion system to optimize a previously selected ribozyme. One study [169] selected variants of the previously developed class I ligase ribozyme [170], and two studies optimized the polymerase ribozyme [34] for higher polymerization efficiency [68, 69]. This sequential procedure of relying on in vitro selection experiments in bulk to generate a ribozyme and improve its performance by subsequent evolution rounds in emulsion, was combined into one procedure in a different study. Here, ribozymes were selected from random sequence during five selection rounds in bulk, then a further 6-9 rounds of selection in emulsion [171]. Therefore, the current study is the first in which a ribozyme was selected from random sequence with all selection rounds in emulsion.

Future experiments will truncate the size of the selected RNA to identify the catalytic core of the ribozyme. Future experiments will also determine the secondary structure of this truncated core, determine the optimal reaction conditions (monovalent and divalent ion concentrations, pH,

and substrate concentrations) and attempt to determine turnover numbers.

6.5 Materials and Methods

6.5.1 Generation of the pool

A 188-mer DNA oligo (150 Ns flanked by a constant region on both sides) was ordered from IDT. IDT reported that the quality of the oligomer could not be guaranteed. The solid pellet was redissolved in 270 uL water at a concentration of 44.4 uM. qPCR with a control allowed us to determine that only 2.24% of the oligo amplified to the correct size. This placed a lower limit on the pool complexity to 1.62×10^{14}. PCR with a titration of the template gave 264 nM as the optimal starting template concentration. 45 mL of PCR was performed to amplify the oligomer using short primers. This PCR1 was cleaned up and then used in PCR2 at a concentration of 30 nM. PCR2 added in the T7 transcription promoter phi 6.5 with a hammerhead ribozyme that co-transcriptionally self cleaves from the 5 end of the pool construct. Additionally, a DNAzyme recognition site was added downstream of the pool construct. This recognition site is recognized and cleaved by a DNAzyme which leaves a clean 3 OH group on the pool molecule. This 3 OH group can then be recognized by the polymerase ribozyme. The PCR2 was ethanol precipitated, redissolved in water, and then used directly in a transcription in a 1:1 volume ratio to PCR2 (200 uL PCR2 -¿ 200 uL transcription).

6.5.2 In vitro selection

The overview of one selection round was as follows: PCR, transcription, 4% PAGE, 5 end radiolabeing, 5% PAGE, DNAzyme processing, 5% PAGE, reaction in emulsion, extraction from emulsion, 5% APM-PAGE, reverse transcription. PCR and transcription were described above. The 4% PAGE following transcription was run until the bromophenol blue dye was three

quarters into the gel. This took around 2 hours. The band cut out was usually just beneath the dye. All gel elutions, except for the one following APM-PAGE, were done using 450 uL of 300 mM NaCl per tube. 5 end radiolabeling was performed using PNK and 32P on a trace amount of the pool RNA. The rest of the RNA was dissolved in 5 mM HEPES pH 7.5, 15 mM NaCl, 0.1 mM EDTA and stored in a -20 C freezer until the DNAzyme processing step. After 5% PAGE for the radiolabeling step, the radiolabeled pool was also dissolved in the same buffer. The two types of pool, trace radiolabeled and bulk unlabeled, were joined together during the subsequent DNAzyme processing step. [DNAzyme already covered].

The DNAzyme reaction was ethanol precipitated and redissolved in a lower volume so it could fit on the gel during the subsequent 5% PAGE. Running the gel for approximately two hours separated the cleaved from the uncleaved band sufficiently to cleanly cut out the lower band.

A considerable amount of polymerase ribozyme was transcribed due to its requirement in the emulsion step. The emulsion was prepared as follows: 95% heavy mineral oil, 4% ABIL EM-90 (v/v) measured by weight to give 95% total emulsion volume. This was stirred at room temperature and then degassed until no more bubbles were being generated. The oil phase was continuously stirred until the aqueous phase was prepared. The aqueous phase (2.5 mL in a 50 mL emulsion) consisted of two parts: the RNAs with buffer, and the substrate and ion components required to start the reaction. The RNAs were incubated with 100 mM Tris/HCl pH 8.3 and heat denatured using a thermocycler at 80 C for 2 minutes before being immediately chilled on ice for 5 minutes. The second aqueous mix contained $MgCl_2$, KCl, trimetaphosphate, and 6thioguanosine. Both parts were prepared such that they had equal volumes before being mixed together. The final concentrations were: 3 uM polymerase ribozyme, 0.5 uM pool, 50 mM Tris/HCl pH 8.3, 150 mM $MgCl_2$, 200 mM KCl, 50 mM trimetaphosphate, and 1 mM 6thioguanosine. As the selection progressed, the trimetaphosphate, 6thioguanosine, and pool concentrations were adjusted to increase selection pressure. The reaction was started by quickly

mixing both aqueous parts together while on ice and then pipetting the entire mixture onto the stirring oil phase. This generated a course emulsion which was mixed for 30 seconds before being loaded into the microfluidizer. Note that approximately 10 uL of aqueous phase was kept in the tube with 6sGTP added as a positive control.

The emulsion was passed by the microfluidizer 7 times (collected and loaded from the same tube during the passes) before being sealed by paraffin and left to incubate at room temperature for 6 hours. Note that with each subsequent pass, the emulsion viscosity would be lowered and the color of the emulsion stream coming out of the microfluidizer would shift from opaque white to translucent.

The incubated emulsion was processed by first adding a molar excess of EDTA to chelate the Mg ion followed by addition of 5 mL of water to increase the aqueous volume. 5 mL of diethyl ether was added followed by vortexing to break open the emulsion. This broken emulsion can be stored in the freezer overnight. The emulsion was spun at 15000 g for 30 minutes at 4 °C. The oil phase was decanted. The tube was filled with ether, vortexed, and spun again at the same settings. The oil/ether phase was decanted (note: decanting must be done very quickly as the pellet that contains the RNAs can burst). 5 mL of phenol was added to the pellet and then vortexed until the entire pellet dissolved. This was spun at 15000 g for 10 minutes at 4 °C. The aqueous phase was transferred via pipet to a new tube. Note that the aqueous phase appears as globules floating near the top of the solution. A volumetric excess of chloroform was added to the aqueous phase and vortexed. The mixture was spun at 15000 g for 5 minutes at 4 °C. The aqueous layer was transferred to a new tube. Sodium chloride was added to the solution to guarantee a concentration of at least 300 mM NaCl. This solution was then ethanol precipitated. Once redissolved, the solution was treated with 5% APM-PAGE.

APM-PAGE was performed as follows: the polyacrylamide gel was made in three parts from bottom to top. The bottom and top layers both consisted of 5% polyacrylamide. The middle layer contained 0.1 mM APM (from a stock solution of 3 mM APM in DMF). The sample was

loaded in the middle of the gel alongside a positive control consisting of the pool with spiked 6sGTP. The APM-PAGE was run for 4-6 hours, until the xylene cyanol dye was about to run off the gel. Using phosphorimaging, the radiolabeled pool molecules that had thio-modifications were captured at the polyacrylamide-APM interface, whereas the inactive population of the pool migrated through the APM layer. This interface band was excised and eluted in 300 mM NaCl and 5 mM DTT overnight at 4 °C.

Following ethanol precipitation, the pool molecules were reverse transcribed using superscript III and a short RT primer that binds to the 3 constant region.

6.5.3 High Throughput Sequencing analysis

Miseq data was given as a de-multiplexed FASTQ file format. The data were uploaded to the online Galaxy bioinformatics website where the constant regions were clipped and the files were converted to FASTA format. Each round contained its own FASTA file sorted by sequence abundance. This file was used to generate clusters. Clustering was performed using the USEARCH suite by Edgar Robert (cluster_fast function at 0.75 identity threshold). Note that clustering was done using a greedy algorithm and so the cluster centers become defined as the most abundant sequence due to the sequence order within the FASTA file. Clusters were tracked throughout the multiple rounds/FASTA files to generate a cluster progression throughout the selection. Furthermore, each cluster had an R script applied that tracked the sequence progression within each cluster through each round. Highly enriching and abundant sequences were aligned using the MUSCLE suite by Edgar Robert. This allowed the identification for highly enriching mutations within a cluster [172, 173].

6.5.4 NTR assay

The NTR assay follows the same principles as the emulsion reaction during the selection. The NTR, polymerase ribozyme, and a radiolabeled primer (KF) are heat denatured for 80 °C for 2 minutes and then slowly cooled to 4 °C. A premix containing $MgCl_2$, KCl, trimetaphosphate, and a nucleoside was added to the RNAs to start the reaction. Timepoints were made by aliquoting 5 uL reaction to 5 uL formamide loading buffer. Reaction progression was tracked by running the timepoints on a 20% PAGE and then phosphorimaging. The primer becomes ligated when there are NTPs generated, which results in a gel shift.

6.6 Supplemental

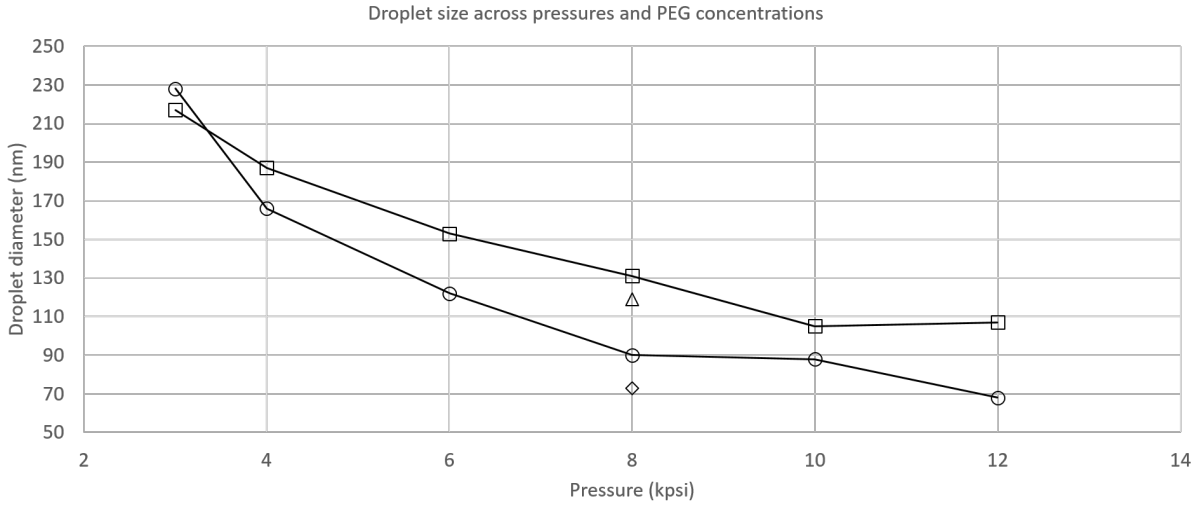


Figure 6.5: Dynamic light scattering of the emulsion revealed a well behaved emulsion system. Droplet diameter is calculated from the decay rate measured from dynamic light scattering. X-axis is the pressure used to generate the emulsion using the microfluidizer. Squares represent 0% PEG 20k. Triangles represent 5% PEG 20k. Circles represent 10% PEG 20k. Diamonds represent 20% PEG 20k. These percentages are within the aqueous phase of the emulsion, which is 5% of the total emulsion volume.

Round	% Ligated (background corrected)	Normalized	Tmp (mM)	6sGsn (mM)
1-A	0.411	1	50	1
1-B	0.942	2.29	50	1
1-C	0.688	1.67	50	1
1-D	1.04	2.53	50	1
1-E	1.33	3.24	50	1
1-F	0.498	1.21	50	1
1-G	0.754	1.83	50	1
2	0.936	2.28	50	1
3	0.339	3.30	25	0.5
4	0.0691	2.69	12.5	0.25
5	0.1022	3.98	12.5	0.25
6	0.1506	5.86	12.5	0.25
7	0.0197	0.77	12.5	0.25
8	0.0063	0.25	12.5	0.25
9	0.43	16.74	12.5	0.25
10	0.51	19.85	12.5	0.25
11	1.11	43.21	12.5	0.25
12	6.81	265.11	12.5	0.25
13	6.18	3007.30	2.5	0.1
14	1.17	2846.72	1	0.05
15	0.12	58.39	2.5	0.1
16	6.9	3357.66	2.5	0.1
17	0.52	2530.41	0.5	0.05
18	0.47	2287.10	0.5	0.05

Figure 6.6: Table of the progress of the NTR selection. Activity from all rounds is normalized to the performance of the first section of round 1. Normalization was calculated from the product of the two substrate concentrations coupled with the percent ligated at the interface.

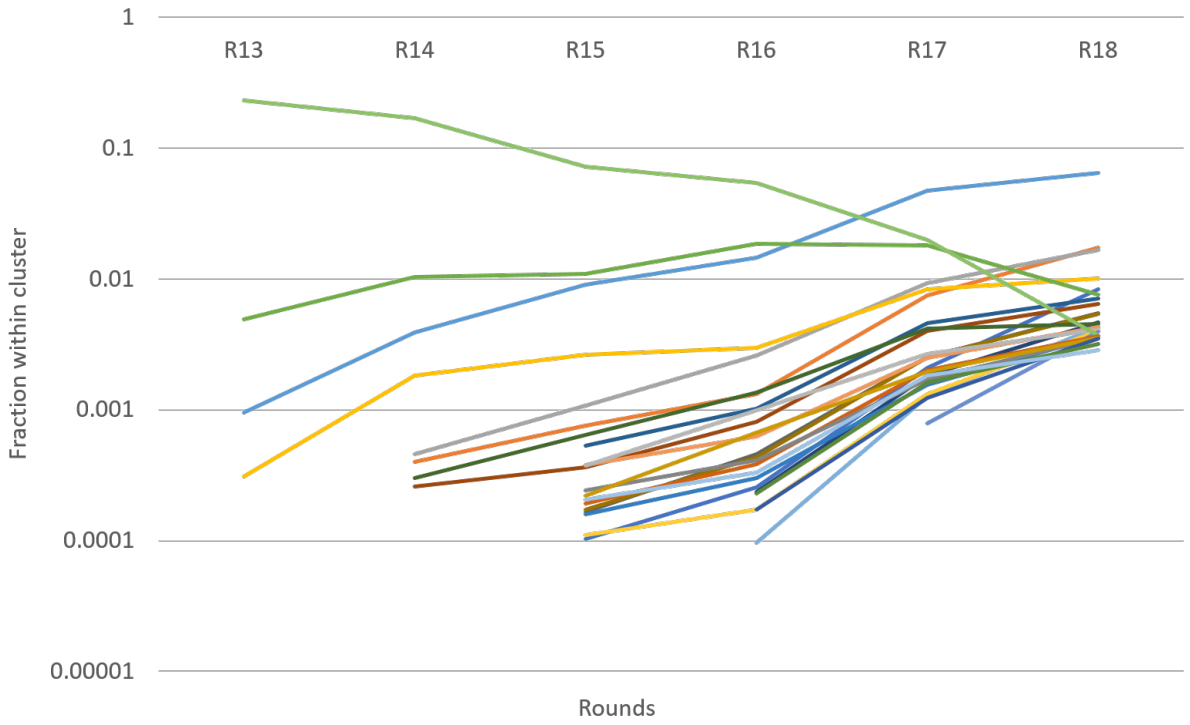


Figure 6.7: Sequence progression from cluster c1. The top 25 abundant sequences from round 18 were tracked throughout the selection as a performance of their fraction represented in the cluster at that certain round.

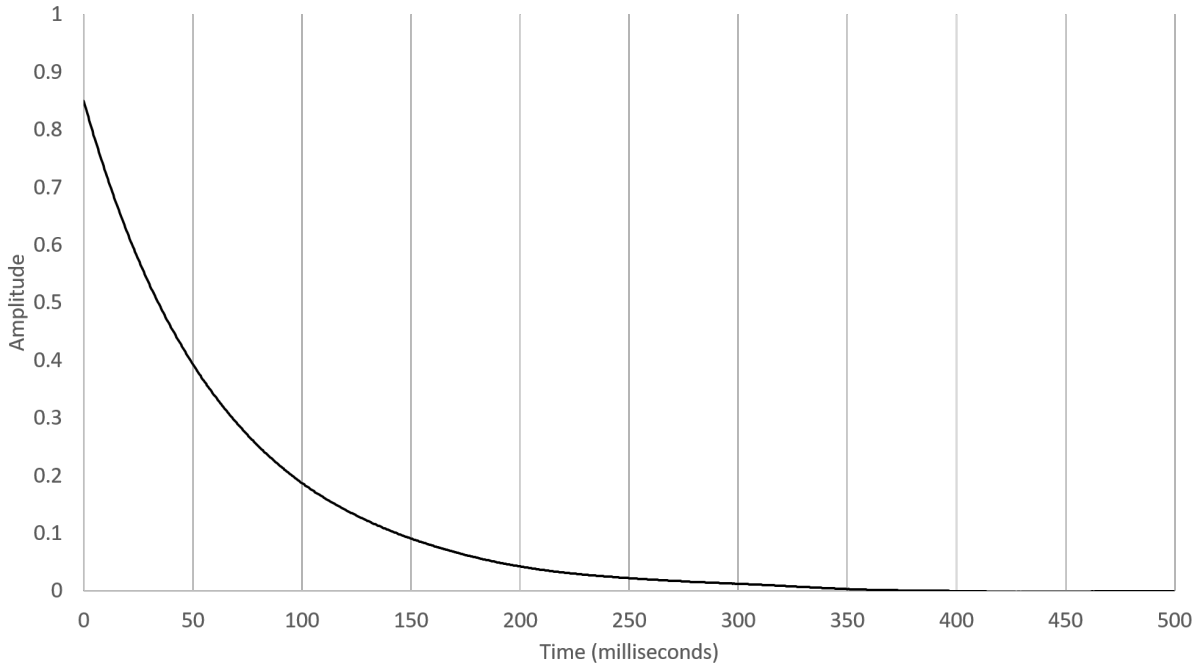


Figure 6.8: Dynamic light scattering was done on the following emulsion formulation that was used in the in vitro selection: 4% ABIL EM90, heavy mineral oil (v/v), 5% total aqueous phase.

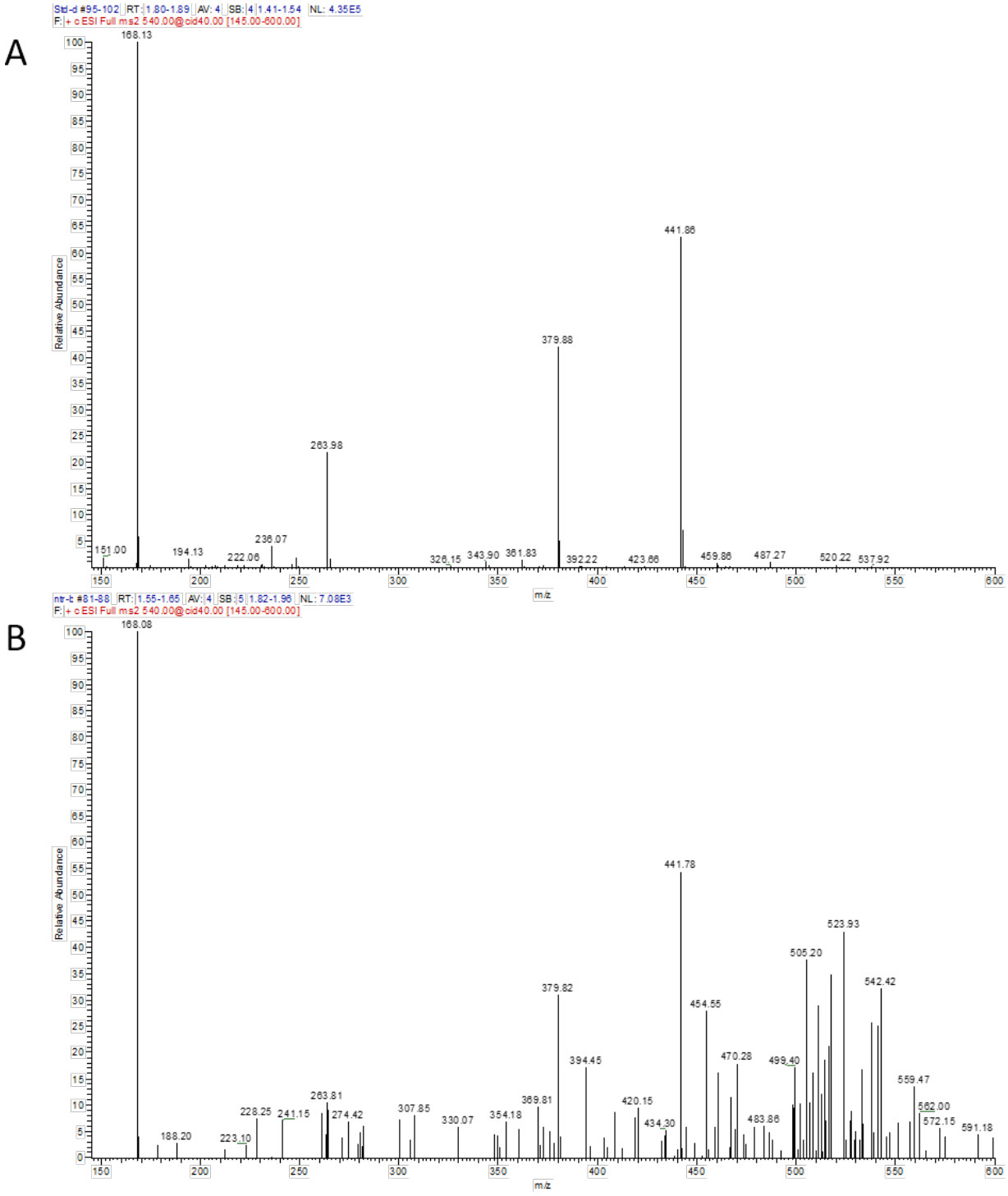


Figure 6.9: LC-ESI-MS was done on a positive control containing 6sGTP and a NTR sample that had been incubated for six hours. The following is a MS/MS "fingerprint" of the main 6sGTP peak detected during MS (540 m/z). A. MS/MS of the positive control 6sGTP sample. B. MS/MS of the NTR sample. Note that both contain peak 168 m/z as the most abundant, followed by the same pattern of abundance for secondary peaks.

6.7 Acknowledgements

Significant thanks to Doug Magde and Michael Magde for the dynamic light scattering measurements. Many thanks to Sophie Chen of Brian Zid's lab for helping develop the R script that provides sequence progression within each cluster during High Throughput Sequencing analysis. Lots of thanks to Dr. Su from the Mass Spectrometry facility in Urey Hall for providing crucial MS data on the NTR samples.

Chapter 7

Discussion and Future Directions

7.1 Experimental insights

The greatest learning opportunity presented to me during my PhD career was having a mixture of short and long-term projects. I gained insight on what RNA is capable of. Perhaps the largest misconception about RNA is that it's a fragile biomolecule. While it won't last very long at pH 12, as long as clean and proper technique is used RNA is quite stable. In my own experience there's only been a single RNAse A contamination (data not shown). Similarly, RNA does not degrade the instant it's made, and it can be dissolved and used in pure water without the need of DEPC. In my experience, catalytic RNAs can be stored in the freezer for up to one month without losing activity. A face shield was never needed when gel extracting RNAs. The only preventative technique was to hold my breath and to use new razor blades with each gel fragment. A majority of the time when there's a problem with the RNA the actual problem lies in the design and generation of the DNA template. Run-off transcription is only as good as the template it polymerizes off of.

7.2 The TPR selection system as a model

The triphosphorylation ribozyme selection established a fundamental pillar in the RNA world hypothesis. Namely, it demonstrated that triphosphorylation, which is a prebiotically plausible energy pathway, can be catalyzed by ribozymes. Although it stopped short of generating free NTPs, the selection system iteslf is still benefitting the *in vitro selection* community. The high enrichment factor of the selection system (10^4-fold per selection round), coupled with the fact that it has successfully been used five distinct times (data unpublished) shows that the TPR selection system can be used as a model to study RNA and selection behavior. Although the end result would be the same in each selection (5' cis-triphosphorylation of RNA polymers), altering the requirements and selection conditions would provide great insight into how selections in general work and elucidate how RNAs react to such selections. For example, it would be possible to do a triphosphorylation selection in the presence of lipids. This would reveal RNA-lipid interactions simply from which RNAs are enriched compared to the "naked" RNA line of selection.

There are unbounded possibilities for probing the nature of RNA complexes by, for example, doing a two-line parallel TPR selection. Using the same starting pool, one could split the selection into two parallel lines. One line would replicate the original TPR selection (only RNA), and the other line would probe for some RNA-complex interactions and characteristics by incubating the RNAs with a particular substrate during the triphosphorylation step. These lines could continue until both selected pools are enriched. Differences in enrichment would elucidate if the RNA-complex aids or harms catalysis. With high-throughput sequencing, this analysis could be taken a step further by comparing the emergent clusters and noticing if they're representation in the pool has changed and if certain mutations are favored in one line versus the other. In essence, the TPR selection system allows RNA biochemists to study fundamental properties of RNA-complexes. One caveat that must be noted from this experiment is that it might only elucidate RNA triphosphorylation, not necessarily catalysis in general.

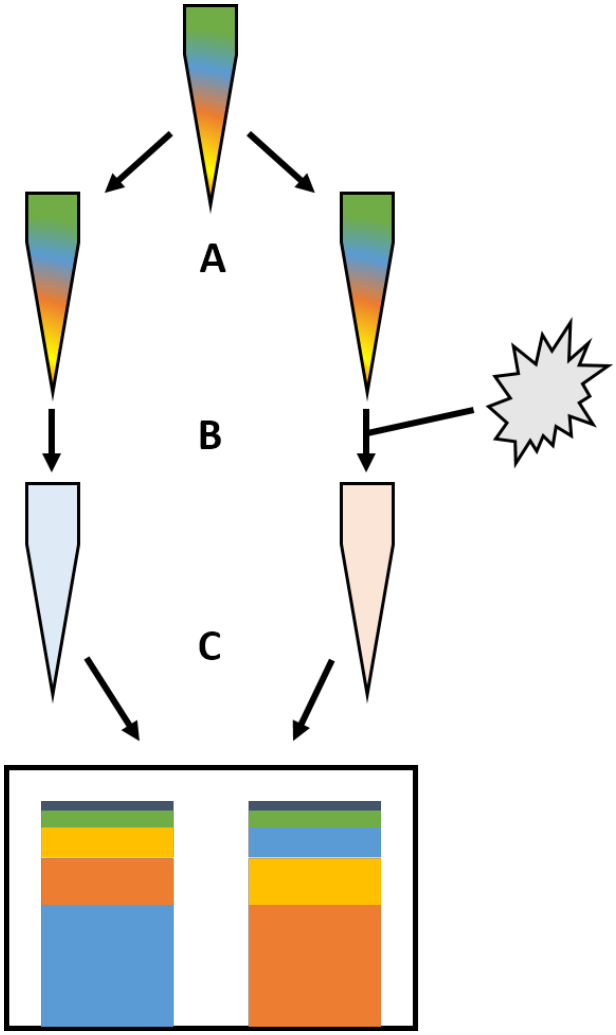


Figure 7.1: RNA-complex properties can be studied by using the following general schematic. (A) A single large randomized pool is split into two parallel lines of selection. (B) One of the lines of selection is treated exactly as in the original TPR selection. The other parallel line has an extra component (up to the researcher) that may or may not aid in sucees of the selection. Ideally this component should be added to the pool before the selective step with trimetaphosphate each round. (C) Once the selections are over, high-throughput sequencing can elucidate which clusters were enriched or hurt by the addition of the component.

7.3 A benefit of short RNAs

Selections are done with long RNAs because it's likelier to find active domains or registers than in short RNAs [174]. Despite this, it's generally agreed upon in the origin of life field that selected ribozymes would be more RNA-world plausible if they were shorter. This is

because short RNAs polymerize readily compared to longer RNAs, and also, with the fragmented ribozyme paper, perhaps short fragmented RNAs could outcompete longer ones in a particular environment [57]. That a fragmented version of TPR1e was faster at lower temperature than its linear parent may only be significant for this specific ribozyme. There is a chance, however, that this phenomena is also present for ribozymes in general. If that is the case, then shorter RNAs would be favored during a ”cold” RNA world. This isn't the first time there was evidence that would favor such an environment. Cold eutectic spaces in ice provided the polymerase ribozyme with several advantages over ambient temperature. Hydrolysis of RNA is slowed, local concentrations of various substrates increase, and shorter basepairing regions are strengthened [79]. Determining whether shorter, fragmented ribozymes generally perform better than longer ribozymes should be investigated further.

7.4 The polymerase ribozyme - the gift that keeps on giving

Although the original polymerase ribozyme is now old enough to vote, it still provides novel uses. It is able to be modified in such distinct ways as RNA-PCR, reverse transcription, and now utilizing modified nucleotide substrates [175, 35, 36, 159]. It's use in the NTR selection as a tagging tool opens up various possibilities for future selections. For example, leading off of the 6-thio GTP evolution, a recent paper utilized the polymerase ribozyme to tag the 3' end of RNAs with fluorophores, biotin, azide, or other moieties [159]. This demonstrates how a ribozyme originally meant as a proof-of-concept evovled over time to become a potential chemical biology tool that can aid in other selections or for studying biology.

There are two notes I'd like to make about the polymerase ribozyme. First is that the three-dimensional structure of the accessory domain has still not been solved. Perhaps recent tools such as XFEL or cryo-EM can aid in identifying the structure. The benefit of this is that it would confirm where the NTP binding site is and it can aid in future design and engineering of

the ribozyme.

If the structure of the accessory domain is solved it might assist in evolving a ”true” polymerase ribozyme. None of the polymerase ribozyme variants can generate sequence-independent, length-independent RNA polymers. Having such a polymerase ribozyme would be the hallmark of an RNA-world organism. Achieving such a ribozyme might be out of reach simply because perhaps RNA isn't capable of it. Although this is simply speculation.

7.5 An alternative tool to generate clean length-specific RNAs

The HDV ribozyme is a cotranscriptional tool that can efficiently generate length-specific RNA constructs. I personally would generate small RNAs with extremely low yields because it was distributed along N+1,2,3 bands. Attaching the HDV ribozyme to it would generate one clean band to purify (data not shown). The specific HDV ribozyme used wasn't efficient over general sequences. Likewise, the original HDV ribozyme also has sequence specificity. One useful future direction would be to evolve a sequence-independent HDV ribozyme. This would help the RNA community have clean efficient transcriptions without worrying about smeary products on their gel while also providing an internal size marker based on the length of the evolved ribozyme.

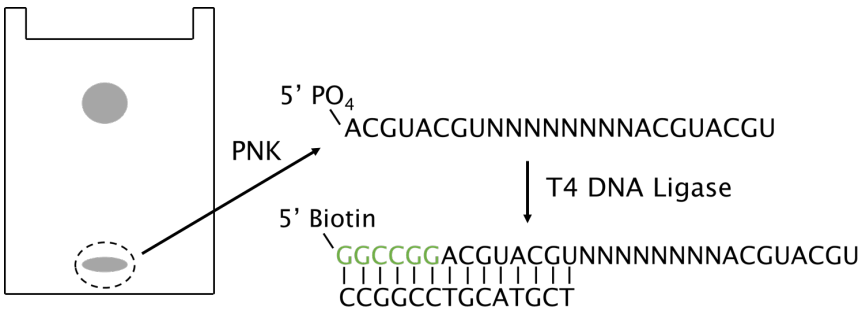


Figure 7.2: A proposed selection to generate a sequence-independent HDV ribozyme. Note that only the selective step is shown. Starting from the left and going clockwise, an RNA construct containing a randomized HDV sequence is incubated for a certain amount of time and then subjected to PAGE. Those successful HDV sequences that cleaved will migrated further down the gel. This excised gel fragment is then eluted and the RNA phosphorylated using polynucleotide kinase. This allows the cleaved HDV RNAs to be recognized as a substrate by T4 DNA ligase. Further false positives can be removed by using a streptavidin capture method. Only those RNAs that had the proper phosphate moiety and were ligated to the biotin containing oligo would be captured. PCR can then be used to amplify the pool and start a new round. Note that a splint should be used to enable proper ligation.

7.6 The NTR project four years after conception

We have a selection system in emulsion with a uniquely evolved ribozyme that consistently enriches sequences over multiple rounds. This system can be analyzed with high throughput sequencing, which greatly aids in selecting and designing active sequences. This selection system is just the beginning. Not only can the emulsion be used for other selections, but this specific NTR selection can be greatly exploited. For example, seeing if there can be a ribozyme that can generate dNTPs or canonical NTPs. Furthermore, this project established that it was possible to select from random sequence in emulsion from the very start of the selection.

Specific to the NTR selection, the next step would be to generate NTRs that are capable of multiple turnover of NTPs. This can be achieved by multiple rounds of selection using larger droplets. This adds a selection pressure in the sense that generated NTPs need to reach sufficient concentration to be recognized and utilized by the Pol Rz. In a similar vein, the unevolved Pol Rz can also be used. The step after that would be to switch to a canonical NTP. After that switching the NTP required between each round to generate a general NTR (geNTR). This would ideally be

able to generate all four canonical NTPs.

If a geNTR can be achieved, then we are tantilizingly close to an RNA world organism. This is because it can be coupled to a (hopefully by then true) polymerase ribozyme. Then, with only two ribozymes, one can go from nucleosides and trimetaphosphate to an RNA polymer. The only obstacle left would be the reproducing step. Somehow, generating the RNA polymer would then lead to a reproduction of the organism.

7.7 Other thoughts

The emulsion system currently used one batch of ABIL-EM90, a proprietary emulsifier. It would be ideal if the selection system can use a consistent, known emulsifier, perhaps Span/Tween. Similarly, processing the RNA with a DNAzyme was quite troublesome and might not have generated the 2',3'-diol we hoped for (data not shown). Therefore, in my opinion, it is worth it to spend up to a year selecting for a self-cleaving ribozyme that generates a 2',3'-diol on the upstream RNA. The design of such a system shouldn't take too long. The crucial step is to change the sequence of the upstream construct every round such that the selected ribozyme can tolerate any upstream sequence. Once such a ribozyme can be generated, it would make future NTR-like selections much simpler. It would increase RNA yield over 100 percent, save 33 percent of selection turnaround time, and be chemically defined.

One final thought about the origin of life, we will probably never know how life arose. If we do create an RNA-world organism then it will be more representative of what we thought early life was like rather than what it actually was. This doesn't mean that this research is pointless. Rather, it's questions like these, that so many people spend so much effort on, that shows the whole point of life.

Bibliography

[1] G. Brent Dalrymple. The age of the earth in the twentieth century: a problem (mostly) solved. *Geological Society, London, Special Publications*, 190(1):205–221, 2001.

[2] Henry Norris Russell and James Hopwood Jeans. A superior limit to the age of the earth's crust. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 99(696):84–86, 1921.

[3] F. W. Aston. The constitution of ordinary lead. *Nature*, 120:224, August 1927.

[4] F. W. Aston. The mass-spectrum of uranium lead and the atomic weight of protactinium. *Nature*, 123:313, March 1929.

[5] Clair Patterson. The isotopic composition of meteoritic, basaltic and oceanic leads, and the age of the earth. In *Proceedings of the Conference on Nuclear Processes in Geologic Settings, Williams Bay, Wisconsin*, 1953.

[6] Fouad Tera. Aspects of isochronism in pb isotope systematicsapplication to planetary evolution. *Geochimica et Cosmochimica Acta*, 45(9):1439–1448, 1981.

[7] J. William Schopf. Fossil evidence of archaean life. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1470):869–885, 2006.

[8] Abigail C. Allwood, Malcolm R. Walter, Balz S. Kamber, Craig P. Marshall, and Ian W. Burch. Stromatolite reef from the early archaean era of australia. *Nature*, 441:714–8, June 2006.

[9] C. R. Woese. The genetic code: The molecular basis for genetic expression. *Harper & Row New York*, pages 179–195, 1967. Cited By :248.

[10] C. R. Woese, D. H. Dugre, W. C. Saxinger, and S. A. Dugre. The molecular basis for the genetic code. *Proceedings of the National Academy of Sciences of the United States of America*, 55:966–74, April 1966.

[11] Carl R. Woese. The fundamental nature of the genetic code: prebiotic interactions between polynucleotides and polyamino acids or their derivatives. *Proceedings of the National Academy of Sciences*, 59(1):110–117, 1968.

[12] f. H. C. Crick. The origin of the genetic code. *Journal of Molecular Biology*, 38(3):367–379, 1968. Cited By :1297.

[13] l. E. Orgel. Evolution of the genetic apparatus. *Journal of Molecular Biology*, 38(3):381–393, 1968. Cited By :536.

[14] l. E. Orgel. Prebiotic chemistry and the origin of the rna world. *Critical reviews in biochemistry and molecular biology*, 39(2):99–123, March 2004. Cited By :524.

[15] Gerald F. Joyce. The antiquity of rna-based evolution. *Nature*, 418:214–21, July 2002.

[16] K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, and T. R. Cech. Self-splicing rna: autoexcision and autocyclization of the ribosomal rna intervening sequence of tetrahymena. *Cell*, 31:147–57, November 1982.

[17] C. Guerrier-takada, K. Gardiner, T. Marsh, N. Pace, and S. Altman. The rna moiety of ribonuclease p is the catalytic subunit of the enzyme. *Cell*, 35:849–57, December 1983.

[18] Peter B. Moore and Thomas A. Steitz. The structural basis of large ribosomal subunit function. *Annual review of biochemistry*, 72:813–50, 2003.

[19] H. B. White. Coenzymes as fossils of an earlier metabolic state. *Journal of Molecular Evolution*, 7(2):101–104, June 1976.

[20] Walter Gilbert. The rna world. *nature*, 319:618, 1986.

[21] Eduard Torrents, Patrick Aloy, Isidre Gibert, and Francisco Rodriguez-trelles. Ribonucleotide reductases: divergent evolution of an ancient enzyme. *Journal of molecular evolution*, 55:138–52, August 2002.

[22] Michael P. Robertson and Gerald F. Joyce. The origins of the rna world. *Cold Spring Harbor perspectives in biology*, 4, May 2012.

[23] G. F. Joyce, A. W. Schwartz, S. L. Miller, and L. E. Orgel. The case for an ancestral genetic system involving simple analogues of the nucleotides. *Proceedings of the National Academy of Sciences of the United States of America*, 84:4398–402, July 1987.

[24] A. R. Jr Hill, L. E. Orgel, and T. Wu. The limits of template-directed synthesis with nucleoside-5'-phosphoro(2-methyl)imidazolides. *Origins of life and evolution of the biosphere : the journal of the International Society for the Study of the Origin of Life*, 23:285–90, December 1993.

[25] J. E. Erman and G. G. Hammes. Relaxation spectra of ribonuclease. v. the interaction of ribonuclease with cytidylyl-3':5'-cytidine. *Journal of the American Chemical Society*, 88:5614–7, December 1966.

[26] S. C. Mohr and R. E. Thach. Application of ribonuclease t1 to the synthesis of oligoribonucleotides of defined base sequence. *The Journal of biological chemistry*, 244:6566–76, December 1969.

[27] Y. Yamagata, H. Watanabe, M. Saitoh, and T. Namba. Volcanic production of polyphosphates and its relevance to prebiotic evolution. *Nature*, 352:516–9, August 1991.

[28] Matthew A. Pasek, Terence P. Kee, David E. Bryant, Alexander A. Pavlov, and Jonathan I. Lunine. Production of potentially prebiotic condensed phosphates by phosphorus redox chemistry. *Angewandte Chemie International Edition*, 47(41):7918–7920, 2008. Cited By :46.

[29] Matthew A. Pasek, Jelte P. Harnmeijer, Roger Buick, Maheen Gull, and Zachary Atlas. 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(25):10089–10094, June 2013. Cited By :72.

[30] R. Osterberg and L. E. Orgel. Polyphosphate and trimetaphosphate formation under potentially prebiotic conditions. *Journal of molecular evolution*, 1:241–8, 1972.

[31] e. Etaix and l. E. Orgel. Phosphorylation of nucleosides in aqueous solution using trimetaphosphate: formation of nucleoside triphosphates. *Journal of Carbohydrates Nucleosides Nucleotides*, 5(2):91–110, 1978. Cited By :35.

[32] Janina E. Moretti and Ulrich F. Muller. A ribozyme that triphosphorylates rna 5'-hydroxyl groups. *Nucleic acids research*, 42:4767–78, April 2014.

[33] D. P. Bartel and J. W. Szostak. Isolation of new ribozymes from a large pool of random sequences [see comment]. *Science (New York, N.Y.)*, 261:1411–8, September 1993.

[34] W. K. Johnston, P. J. Unrau, M. S. Lawrence, M. E. Glasner, and D. P. Bartel. Rna-catalyzed rna polymerization: Accurate and general rna-templated primer extension. *Science*, 292(5520):1319–1325, May 2001. Cited By :393.

[35] David P. Horning and Gerald F. Joyce. Amplification of rna by an rna polymerase ribozyme. *Proceedings of the National Academy of Sciences of the United States of America*, 113:9786–91, August 2016.

[36] Biswajit Samanta and Gerald F. Joyce. A reverse transcriptase ribozyme. *eLife*, 6, September 2017.

[37] S. Tsukiji, S. B. Pattnaik, and H. Suga. An alcohol dehydrogenase ribozyme. *Nature Structural Biology*, 10(9):713–717, September 2003.

[38] Kevin T. Urak, Sabrina Shore, William M. Rockey, Shi-jie Chen, Anton P. Mccaffrey, and Paloma H. Giangrande. In vitro rna selex for the generation of chemically-optimized therapeutic rna drugs. *Methods (San Diego, Calif.)*, 103:167–74, July 2016.

[39] Zhaleh N. Amini, Karen E. Olson, and Ulrich F. Mller. Spliceozymes: Ribozymes that remove introns from pre-mrnas in trans. *PLOS ONE*, 9(7):1–11, July 2014.

[40] P. Nissen, J. Hansen, N. Ban, P. B. Moore, and T. A. Steitz. The structural basis of ribosome activity in peptide bond synthesis. *Science*, 289(5481):920–930, August 2000. Cited By :1522.

[41] G. Sprengel and H. Follmann. Evidence for the reductive pathway of deoxyribonucleotide synthesis in an archaebacterium. *Febs Letters*, 132(2):207–209, 1981. Cited By :12.

[42] A. D. Ellington and J. W. Szostak. In vitro selection of rna molecules that bind specific ligands. *Nature*, 346:818–22, August 1990.

[43] C. Tuerk and L. Gold. Systematic evolution of ligands by exponential enrichment: Rna ligands to bacteriophage t4 dna polymerase. *Science (New York, N.Y.)*, 249:505–10, August 1990.

[44] Lyssa L. Martin, Peter J. Unrau, and Ulrich F. Mueller. Rna synthesis by in vitro selected ribozymes for recreating an rna world. *Life-Basel*, 5(1):247–268, March 2015.

[45] S. A. Mortimer and K. M. Weeks. A fast-acting reagent for accurate analysis of rna secondary and tertiary structure by shape chemistry. *Journal of the American Chemical Society*, 129(14):4144–4145, 2007. Cited By :196.

[46] P. B. Stewart and P. Munjal. Solubility of carbon dioxide in pure water, synthetic sea water, and synthetic sea water concentrates at -5 to 25 c. and 10- to 45-atm. pressure. *Journal of Chemical and Engineering Data*, 15(1):67–71, 1970. Cited By :101.

[47] W. W. Hay, A. Migdisov, A. N. Balukhovsky, C. N. Wold, S. Flgel, and E. Sding. Evaporites and the salinity of the ocean during the phanerozoic: Implications for climate, ocean circulation and life. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 240(1-2):3–46, 2006. Cited By :99.

[48] M. P. Weiner and G. L. Costa. Rapid pcr site-directed mutagenesis. *Genome research*, 4(3):S131–S136, 1994. Cited By :34.

[49] R. Turner, K. Shefer, and M. Ares JR. Safer one-pot synthesis of the 'shape' reagent 1-methyl-7-nitroisatoic anhydride (1m7). *RNA*, 19(12):1857–1863, 2013. Cited By :15.

[50] H. F. Noller, V. Hoffarth, and L. Zimniak. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science*, 256(5062):1416–1419, June 1992.

[51] Xi Chen, Na Li, and Andrew D. Ellington. Ribozyme catalysis of metabolism in the rna world. *Chemistry & Biodiversity*, 4(4):633–655, 2007.

[52] David Wacey, Matt R. Kilburn, Martin Saunders, John Cliff, and Martin D. Brasier. Microfossils of sulphur-metabolizing cells in 3.4-billion-year-old rocks of western australia. *Nature Geoscience*, 4(10):698–702, October 2011.

[53] A. Sanchez R and E. Orgel L. Studies in prebiotic synthesis part 5 synthesis and photo anomerization of pyrimidine nucleosides. *Journal of Molecular Biology*, 47(3):531–543, 1970.

[54] Matthew W. Powner, Beatrice Gerland, and John D. Sutherland. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature*, 459(7244):239–242, May 2009.

[55] H. Sawai and L. E. Orgel. Oligonucleotide synthesis catalyzed by zn2+ ion. *Journal of the American Chemical Society*, 97(12):3532–3533, 1975.

[56] T. Inoue and L. E. Orgel. Oligomerization of (guanosine 5'-phosphor)-2-methylimidazolide on poly(c). an rna polymerase model. *Journal of molecular biology*, 162:201–17, November 1982.

[57] J. P. Ferris and G. Ertem. Oligomerization of ribonucleotides on montmorillonite - reaction of the 5'-phosphorimidazolide of adenosine. *Science*, 257(5075):1387–1389, September 1992.

[58] W. H. Huang and J. P. Ferris. Synthesis of 35-40 mers of rna oligomers from unblocked monomers. a simple approach to the rna world. *Chemical Communications*, (12):1458–1459, 2003.

[59] Sudha Rajamani, Alexander Vlassov, Seico Benner, Amy Coombs, Felix Olasagasti, and David Deamer. Lipid-assisted synthesis of rna-like polymers from mononucleotides. *Origins of Life and Evolution of Biospheres*, 38(1):57–74, February 2008.

[60] Laura Da Silva, Marie-christine Maurel, and David Deamer. Salt-promoted synthesis of rna-like molecules in simulated hydrothermal conditions. *Journal of Molecular Evolution*, 80(2):86–97, February 2015.

[61] Chaitanya V. Mungi and Sudha Rajamani. Characterization of rna-like oligomers from lipid-assisted nonenzymatic synthesis: Implications for origin of informational molecules on early earth. *Life-Basel*, 5(1):65–84, March 2015.

[62] M. Renz, R. Lohrmann, and L. E. Orgel. Catalysts for polymerization of adenosine cyclic 2',3'-phosphate on a poly (u) template. *Biochimica Et Biophysica Acta*, 240(4):463–&, 1971.

[63] M. S. Verlander and L. E. Orgel. Analysis of high molecular-weight material from polymerization of adenosine cyclic 2',3'-phosphate. *Journal of Molecular Evolution*, 3(2):115–120, 1974.

[64] William G. Scott, Abraham Szoeke, Josh Blaustein, Sara M. O'rourke, and Michael P. Robertson. Rna catalysis, thermodynamics and the origin of life. *Life-Basel*, 4(2):131–141, June 2014.

[65] M. S. Verlander, R. Lohrmann, and L. E. Orgel. Catalysts for self-polymerization of adenosine cyclic 2', 3'-phosphate. *Journal of Molecular Evolution*, 2(4):303–316, December 1973.

[66] Giovanna Costanzo, Samanta Pino, Fabiana Ciciriello, and Ernesto Di Mauro. Generation of long rna chains in water. *Journal of Biological Chemistry*, 284(48):33206–33216, November 2009.

[67] Judit E. Sponer, Jiri Sponer, Alessandra Giorgi, Ernesto Di Mauro, Samanta Pino, and Giovanna Costanzo. Untemplated nonenzymatic polymerization of 3 ',5 ' cgmp: A plausible route to 3 ',5 '-linked oligonucleotides in primordia. *Journal of Physical Chemistry B*, 119(7):2979–2989, February 2015.

[68] Hani S. Zaher and Peter J. Unrau. Selection of an improved rna polymerase ribozyme with superior extension and fidelity. *Rna*, 13(7):1017–1026, July 2007.

[69] Aniela Wochner, James Attwater, Alan Coulson, and Philipp Holliger. Ribozyme-catalyzed transcription of an active ribozyme. *Science (New York, N.Y.)*, 332:209–12, April 2011.

[70] James Attwater, Aniela Wochner, and Philipp Holliger. In-ice evolution of rna polymerase ribozyme activity. *Nature Chemistry*, 5(12):1011–1018, December 2013.

[71] Jonathan T. Sczepanski and Gerald F. Joyce. A cross-chiral rna polymerase ribozyme. *Nature*, 515(7527):440–+, November 2014.

[72] K. D. James and A. D. Ellington. The fidelity of template-directed oligonucleotide ligation and the inevitability of polymerase function. *Origins of Life and Evolution of the Biosphere*, 29(4):375–390, August 1999.

[73] Gregory F. Dolan, Arvin Akoopie, and Ulrich F. Mueller. A faster triphosphorylation ribozyme. *PLoS One*, 10(11):e0142559, November 2015.

[74] C. Sagan and G. Mullen. Earth and mars - evolution of atmospheres and surface temperatures. *Science*, 177(4043):52–&, 1972.

[75] Jacob D. Haqq-misra, Shawn D. Domagal-goldman, Patrick J. Kasting, and James F. Kasting. A revised, hazy methane greenhouse for the archean earth. *Astrobiology*, 8(6):1127–1137, December 2008.

[76] James F. Kasting. Early earth faint young sun redux. *Nature*, 464(7289):687–689, April 2010.

[77] J. L. Bada, C. Bigham, and S. L. Miller. Impact melting of frozen oceans on the early earth - implications for the origin of life. *Proceedings of the National Academy of Sciences of the United States of America*, 91(4):1248–1250, February 1994.

[78] Pierre-alain Monnard and Jack W. Szostak. Metal-ion catalyzed polymerization in the eutectic phase in water-ice: A possible approach to template-directed rna polymerization. *Journal of Inorganic Biochemistry*, 102(5-6):1104–1111, May 2008.

[79] James Attwater, Aniela Wochner, Vitor B. Pinheiro, Alan Coulson, and Philipp Holliger. Ice as a protocellular medium for rna replication. *Nature Communications*, 1:76, September 2010.

[80] Anastassia Kanavarioti, Pierre-alain Monnard, and David W. Deamer. Eutectic phases in ice facilitate nonenzymatic nucleic acid synthesis. *Astrobiology*, 1(3):271–281, September 2001.

[81] Lively Lie, Shweta Biliya, Fredrik Vannberg, and Roger M. Wartell. Ligation of rna oligomers by the schistosoma mansoni hammerhead ribozyme in frozen solution. *Journal of Molecular Evolution*, 82(2-3):81–92, March 2016.

[82] Jean-francois Pombert, Christian Otis, Monique Turmel, and Claude Lemieux. The mitochondrial genome of the prasinophyte prasinoderma coloniale reveals two trans-spliced group i introns in the large subunit rrna gene. *Plos One*, 8(12):e84325, December 2013.

[83] M. Goldschmidtclermont, Y. Choquet, J. Girardbascou, F. Michel, M. Schirmerrahire, and J. D. Rochaix. A small chloroplast rna may be required for transsplicing in chlamydomonas-reinhardtii. *Cell*, 65(1):135–143, April 1991.

[84] Eric J. Hayden and Niles Lehman. Self-assembly of a group i intron from inactive oligonucleotide fragments. *Chemistry & Biology*, 13(8):909–918, August 2006.

[85] Aaron S. Burton and Niles Lehman. Enhancing the prebiotic relevance of a set of covalently self-assembling, autorecombining rnas through in vitro selection. *Journal of Molecular Evolution*, 70(3):233–241, March 2010.

[86] C. Guerriertakada and S. Altman. Reconstitution of enzymatic-activity from fragments of m1 rna. *Proceedings of the National Academy of Sciences of the United States of America*, 89(4):1266–1270, February 1992.

[87] S. E. Butcher, J. E. Heckman, and J. M. Burke. Reconstitution of hairpin ribozyme activity following separation of functional domains. *Journal of Biological Chemistry*, 270(50):29648–29651, December 1995.

[88] Katarzyna Adamala, Aaron E. Engelhart, and Jack W. Szostak. Generation of functional rnas from inactive oligonucleotide complexes by non-enzymatic primer extension. *Journal of the American Chemical Society*, 137(1):483–489, January 2015.

[89] Nana Isomoto, Yuri Maeda, Takahiro Tanaka, Hiroyuki Furuta, and Yoshiya Ikawa. Fixation and accumulation of thermotolerant catalytic competence of a pair of ligase ribozymes

through complex formation and cross ligation. *Journal of Molecular Evolution*, 76(1-2):48–58, February 2013.

[90] Rebecca M. Turk, Nataliya V. Chumachenko, and Michael Yarus. Multiple translational products from a five-nucleotide ribozyme. *Proceedings of the National Academy of Sciences of the United States of America*, 107(10):4585–4589, March 2010.

[91] T. W. Wiegand, R. C. Janssen, and B. E. Eaton. Selection of rna amide syntheses. *Chemistry & Biology*, 4(9):675–683, September 1997.

[92] M. Wecker, D. Smith, and L. Gold. In vitro selection of a novel catalytic rna: Characterization of a sulfur alkylation reaction and interaction with a small peptide. *Rna*, 2(10):982–994, October 1996.

[93] Taek Jin Kang and Hiroaki Suga. In vitro selection of a 5 '-purine ribonucleotide transferase ribozyme. *Nucleic Acids Research*, 35(12):4186–4194, June 2007.

[94] Nobuyoshi Niwa, Yusuke Yamagishi, Hiroshi Murakami, and Hiroaki Suga. A flexizyme that selectively charges amino acids activated by a water-friendly leaving group. *Bioorganic & Medicinal Chemistry Letters*, 19(14):3892–3894, July 2009.

[95] B. Seelig and A. Jaschke. A small catalytic rna motif with diels-alderase activity. *Chemistry & Biology*, 6(3):167–176, March 1999.

[96] E. A. Curtis and D. P. Bartel. New catalytic structures from an existing ribozyme. *Nature Structural & Molecular Biology*, 12(11):994–1000, November 2005.

[97] Hani S. Zaher, R. Ammon Watkins, and Peter J. Unrau. Two independently selected capping ribozymes share similar substrate requirements. *Rna*, 12(11):1949–1958, November 2006.

[98] K. B. Chapman and J. W. Szostak. Isolation of a ribozyme with 5'-5' ligase activity. *Chemistry & Biology*, 2(5):325–333, May 1995.

[99] D. Saran, D. G. Nickens, and D. H. Burke. A trans acting ribozyme that phosphorylates exogenous rna. *Biochemistry*, 44(45):15007–15016, November 2005.

[100] N. Li and F. Q. Huang. Ribozyme-catalyzed amirloacylation from coa thioesters. *Biochemistry*, 44(11):4582–4590, March 2005.

[101] Elisa Biondi, Raghav R. Poudyal, Joshua C. Forgy, Andrew W. Sawyer, Adam W. R. Maxwell, and Donald H. Burke. Lewis acid catalysis of phosphoryl transfer from a copper(ii)-ntp complex in a kinase ribozyme. *Nucleic Acids Research*, 41(5):3327–3338, March 2013.

[102] Michael P. Robertson and William G. Scott. The structural basis of ribozyme-catalyzed rna assembly. *Science*, 315(5818):1549–1553, March 2007.

[103] J. Rogers and G. F. Joyce. The effect of cytidine on the structure and function of an rna ligase ribozyme. *Rna*, 7(3):395–404, March 2001.

[104] Tzu-pin Wang, Yu-chih Su, Yi Chen, Yi-ming Liou, Kun-liang Lin, Eng-chi Wang, Long-chih Hwang, Yun-ming Wang, and Yen-hsu Chen. In vitro selection and characterization of a novel zn(ii)-dependent phosphorothiolate thiolesterase ribozyme. *Biochemistry*, 51(1):496–510, January 2012.

[105] S. Baskerville and D. P. Bartel. A ribozyme that ligates rna to protein. *Proceedings of the National Academy of Sciences of the United States of America*, 99(14):9154–9159, July 2002.

[106] E. H. Ekland, J. W. Szostak, and D. P. Bartel. Structurally complex and highly active rna ligases derived from random rna sequences. *Science (New York, N.Y.)*, 269:364–70, July 1995.

[107] S. Fusz, A. Eisenfuhr, S. G. Srivatsan, A. Heckel, and M. Famulok. A ribozyme for the aldol reaction. *Chemistry & Biology*, 12(8):941–950, August 2005.

[108] A. Jenne and M. Famulok. A novel ribozyme with ester transferase activity. *Chemistry & Biology*, 5(1):23–34, January 1998.

[109] K. E. Chapple, D. P. Bartel, and P. J. Unrau. Combinatorial minimization and secondary structure determination of a nucleotide synthase ribozyme. *Rna*, 9(10):1208–1220, October 2003.

[110] B. Cisneros, D. Court, A. Sanchez, and C. Montanez. Point mutations in a transcription terminator, lambda ti, that affect both transcription termination and rna stability. *Gene*, 181(1-2):127–133, November 1996.

[111] I. Toulokhonov and R. Landick. The flap domain is required for pause rna hairpin inhibition of catalysis by rna polymerase and can modulate intrinsic termination. *Molecular Cell*, 12(5):1125–1136, November 2003.

[112] U. F. Muller. Re-creating an rna world. *Cellular and Molecular Life Sciences*, 63(11):1278–1293, June 2006.

[113] Nikolay Zenkin. Hypothesis: Emergence of translation as a result of rna helicase evolution. *Journal of Molecular Evolution*, 74(5-6):249–256, June 2012.

[114] S. R. Price, N. Ito, C. Oubridge, J. M. Avis, and K. Nagai. Crystallization of rna-protein complexes .1. methods for the large-scale preparation of rna suitable for crystallographic studies. *Journal of Molecular Biology*, 249(2):398–408, June 1995.

[115] J. F. Milligan, D. R. Groebe, G. W. Witherell, and O. C. Uhlenbeck. Oligoribonucleotide synthesis using t7 rna-polymerase and synthetic dna templates. *Nucleic Acids Research*, 15(21):8783–8798, November 1987.

[116] J. F. Milligan and O. C. Uhlenbeck. Synthesis of small rnas using t7 rna polymerase. *Methods in enzymology*, 180:51–62, 1989.

[117] F. J. Triana-alonso, M. Dabrowski, J. Wadzack, and K. H. Nierhaus. Self-coded 3'-extension of run-off transcripts produces aberrant products during in vitro transcription with t7 rna polymerase. *The Journal of biological chemistry*, 270:6298–307, March 1995.

[118] H. S. Zaher and P. J. Unrau. T7 rna polymerase mediates fast promoter-independent extension of unstable nucleic acid complexes. *Biochemistry*, 43(24):7873–7880, June 2004.

[119] M. Illangasekare, G. Sanchez, T. Nickles, and M. Yarus. Aminoacyl-rna synthesis catalyzed by an rna. *Science*, 267(5198):643–647, February 1995.

[120] M. W. L. Lau, K. E. C. Cadieux, and P. J. Unrau. Isolation of fast purine nucleotide synthase ribozymes. *Journal of the American Chemical Society*, 126(48):15686–15693, December 2004.

[121] H. Saito and H. Suga. A ribozyme exclusively aminoacylates the 3 '-hydroxyl group of the trna terminal adenosine. *Journal of the American Chemical Society*, 123(29):7178–7179, July 2001.

[122] P. J. Unrau and D. P. Bartel. Rna-catalysed nucleotide synthesis. *Nature*, 395(6699):260–263, September 1998.

[123] N. V. Chumachenko, Y. Novikov, and M. Yarus. Rapid and simple ribozymic aminoacylation using three conserved nucleotides. *Journal of the American Chemical Society*, 131(14):5257–5263, April 2009.

[124] K. A. Hoadley, W. E. Purtha, A. C. Wolf, A. Flynn-charlebois, and S. K. Silverman. Zn2+-dependent deoxyribozymes that form natural and unnatural rna linkages. *Biochemistry*, 44(25):9217–9231, June 2005.

[125] Diana M. Kost, Joseph P. Gerdt, P. I. Pradeepkumar, and Scott K. Silverman. Controlling the direction of site-selectivity and regioselectivity in rna ligation by zn2+-dependent deoxyribozymes that use 2 ',3 '-cyclic phosphate rna substrates. *Organic & Biomolecular Chemistry*, 6(23):4391–4398, 2008.

[126] D. R. Semlow and S. K. Silverman. Parallel selections in vitro reveal a preference for 2 '-5 ' rna ligation upon deoxyribozyme-mediated opening of a 2 ',3 '-cyclic phosphate. *Journal of Molecular Evolution*, 61(2):207–215, August 2005.

[127] Johanna M. Avis, Graeme L. Conn, and Scott C. Walker. Cis-acting ribozymes for the production of rna in vitro transcripts with defined 5' and 3' ends. *Methods in molecular biology (Clifton, N.J.)*, 941:83–98, 2012.

[128] A. R. Ferredamare and J. A. Doudna. Use of cis- and trans-ribozymes to remove 5' and 3' heterogeneities from milligrams of in vitro transcribed rna. *Nucleic Acids Research*, 24(5):977–978, March 1996.

[129] S. C. Walker, J. M. Avis, and G. L. Conn. General plasmids for producing rna in vitro transcripts with homogeneous ends. *Nucleic Acids Research*, 31(15):e82, August 2003.

[130] N. K. Tanner, S. Schaff, G. Thill, E. Petitkoskas, A. M. Craindenoyelle, and E. Westhof. A 3-dimensional model of hepatitis-delta virus ribozyme based on biochemical and mutational analyses. *Current Biology*, 4(6):488–498, June 1994.

[131] Daniel L. Kellerman, Kandice S. Simmons, Mayra Pedraza, Joseph A. Piccirilli, Darrin M. York, and Michael E. Harris. Determination of hepatitis delta virus ribozyme n(-1) nucleobase and functional group specificity using internal competition kinetics. *Analytical Biochemistry*, 483:12–20, August 2015.

[132] D. M. Chadalavada, S. M. Knudsen, S. Nakano, and P. C. Bevilacqua. A role for upstream rna structure in facilitating the catalytic fold of the genomic hepatitis delta virus ribozyme. *Journal of Molecular Biology*, 301(2):349–367, August 2000.

[133] S. Jeong, J. Sefcikova, R. A. Tinsley, D. Rueda, and N. G. Walter. Trans-acting hepatitis delta virus ribozyme: Catalytic core and global structure are dependent on the 5' substrate sequence. *Biochemistry*, 42(25):7727–7740, July 2003.

[134] Mario Moerl and Roland K. Hartmann. Production of rnas with homogeneous 5 '- and 3 '-ends. *Handbook of Rna Biochemistry, Vols 1 and 2, 2nd Edition*, pages 29–43, 2014.

[135] F. Nishikawa, M. Roy, H. Fauzi, and S. Nishikawa. Detailed analysis of stem i and its 5 ' and 3 ' neighbor regions in the trans-acting hdv ribozyme. *Nucleic Acids Research*, 27(2):403–410, January 1999.

[136] Dana J. Ruminski, Chiu-ho T. Webb, Nathan J. Riccitelli, and Andrej Luptak. Processing and translation initiation of non-long terminal repeat retrotransposons by hepatitis delta virus (hdv)-like self-cleaving ribozymes. *Journal of Biological Chemistry*, 286(48):41286–41295, December 2011.

[137] H. Schurer, K. Lang, J. Schuster, and M. Morl. A universal method to produce in vitro transcripts with homogeneous 3 ' ends. *Nucleic Acids Research*, 30(12):e56, June 2002.

[138] Nathan J. Riccitelli, Eric Delwart, and Andrej Luptak. Identification of minimal hdv-like ribozymes with unique divalent metal ion dependence in the human microbiome. *Biochemistry*, 53(10):1616–1626, March 2014.

[139] Chiu-ho T. Webb, Nathan J. Riccitelli, Dana J. Ruminski, and Andrej Luptak. Widespread occurrence of self-cleaving ribozymes. *Science*, 326(5955):953–953, November 2009.

[140] S. P. Rosenstein and M. D. Been. Self-cleavage of hepatitis delta-virus genomic strand rna is enhanced under partially denaturing conditions. *Biochemistry*, 29(35):8011–8016, September 1990.

[141] J. Lykke-andersen and J. Christiansen. The c-terminal carboxy group of t7 rna polymerase ensures efficient magnesium ion-dependent catalysis. *Nucleic Acids Research*, 26(24):5630–5635, December 1998.

[142] Alexander Rich. On the problems of evolution and biochemical information transfer. *Hoirzons in Biochemistry*, 1962.

[143] Jack W. Szostak, David P. Bartel, and P. Luigi Luisi. Synthesizing life. *Nature*, 409(6818):387, 2001.

[144] David M. Shechner, Robert A. Grant, Sarah C. Bagby, Yelena Koldobskaya, Joseph A. Piccirilli, and David P. Bartel. Crystal structure of the catalytic core of an rna-polymerase ribozyme. *Science*, 326(5957):1271–1275, 2009.

[145] David M. Shechner and David P. Bartel. The structural basis of rna-catalyzed rna polymerization. *Nature structural & molecular biology*, 18(9):1036, 2011.

[146] Eric H. Ekland and David P. Bartel. Rna-catalysed rna polymerization using nucleoside triphosphates. *Nature*, 382(6589):373, 1996.

[147] Michael S. Lawrence and David P. Bartel. Processivity of ribozyme-catalyzed rna polymerization. *Biochemistry*, 42(29):8748–8755, 2003.

[148] Ulrich F. Müller and David P. Bartel. Substrate 2-hydroxyl groups required for ribozyme-catalyzed polymerization. *Chemistry & biology*, 10(9):799–806, 2003.

[149] Qing S. Wang, Leslie Kl Cheng, and Peter J. Unrau. Characterization of the b6. 61 polymerase ribozyme accessory domain. *Rna*, 17(3):469–477, 2011.

[150] E. H. Ekland and D. P. Bartel. The secondary structure and sequence optimization of an rna ligase ribozyme. *Nucleic acids research*, 23:3231–8, August 1995.

[151] Darren J. Parker, Ying Xiao, John M. Aguilar, and Scott K. Silverman. Dna catalysis of a normally disfavored rna hydrolysis reaction. *Journal of the American Chemical Society*, 135:8472–5, June 2013.

[152] G. L. Igloi. Interaction of trnas and of phosphorothioate-substituted nucleic acids with an organomercurial. probing the chemical environment of thiolated residues by affinity electrophoresis. *Biochemistry*, 27:3842–9, May 1988.

[153] Elisa Biondi and Donald H. Burke. Separating and analyzing sulfur-containing rnas with organomercury gels. *Methods in molecular biology (Clifton, N.J.)*, 883:111–20, 2012.

[154] R. C. Cadwell and G. F. Joyce. Mutagenic pcr. *PCR methods and applications*, 3:S136–40, June 1994.

[155] Zhaleh N. Amini and Ulrich F. Muller. Low selection pressure aids the evolution of cooperative ribozyme mutations in cells. *The Journal of biological chemistry*, 288:33096–106, November 2013.

[156] R. C. Cadwell and G. F. Joyce. Randomization of genes by pcr mutagenesis. *PCR methods and applications*, 2:28–33, August 1992.

[157] Jack J. Fox, Iris Wempen, Alexander Hampton, and Peter J. Unrau. Thiation of nucleosides. i. synthesis of 2-amino-6-mercapto-9--d-ribofuranosylpurine (thioguanosine) and related purine nucleosides. *Journal of the American Chemical Society*, 80(7):1669–1675, 1958.

[158] Chengguo Yao and Ulrich F. Muller. Polymerase ribozyme efficiency increased by g/t-rich dna oligonucleotides. *RNA (New York, N.Y.)*, 17:1274–81, July 2011.

[159] Biswajit Samanta, David P. Horning, and Gerald F. Joyce. 3'-end labeling of nucleic acids by a polymerase ribozyme. *Nucleic acids research*, 46:e103, September 2018.

[160] Li Li, Noam Prywes, Chun Pong Tam, Derek K. O'flaherty, Victor S. Lelyveld, Enver Cagri Izgu, Ayan Pal, and Jack W. Szostak. Enhanced nonenzymatic rna copying with 2-aminoimidazole activated nucleotides. *Journal of the American Chemical Society*, 139:1810–1813, February 2017.

[161] Walter Feldmann. Zur chemie der kondensierten phosphate und arsenate, liii. das trimetaphosphat als triphosphorylierungsmittel fr alkohole und kohlenhydrate in wriger lsung. seine sonderstellung unter den kondensierten phosphaten. *Chemische Berichte*, 100(12):3850–3860, 1967.

[162] Barry Herschy, Sae Jung Chang, Ruth Blake, Aivo Lepland, Heather Abbott-lyon, Jacqueline Sampson, Zachary Atlas, Terence P. Kee, and Matthew A. Pasek. Archean phosphorus liberation induced by iron redox geochemistry. *Nature communications*, 9:1346, 2018.

[163] Abe Pressman, Janina E. Moretti, Gregory W. Campbell, Ulrich F. Mller, and Irene A. Chen. Analysis of in vitro evolution reveals the underlying distribution of catalytic activity among random sequences. *Nucleic acids research*, 45:8167–8179, August 2017.

[164] Arvin Akoopie and Ulrich F. Muller. Lower temperature optimum of a smaller, fragmented triphosphorylation ribozyme. *Physical chemistry chemical physics : PCCP*, 18:20118–25, July 2016.

[165] S. A. Benner, A. D. Ellington, and A. Tauer. Modern metabolism as a palimpsest of the rna world. *Proceedings of the National Academy of Sciences of the United States of America*, 86:7054–8, September 1989.

[166] F. J. Trianaalonso, M. Dabrowski, J. Wadzack, and K. H. Nierhaus. Self-coded 3'-extension of run-off transcripts produces aberrant products during in-vitro transcription with t7 rna-polymerase. *Journal of Biological Chemistry*, 270(11):6298–6307, March 1995.

[167] Arvin Akoopie and Ulrich F. Muller. The ntp binding site of the polymerase ribozyme. *Nucleic acids research*, 46:10589–10597, November 2018.

[168] Farid J. Ghadessy and Philipp Holliger. A novel emulsion mixture for in vitro compartmentalization of transcription and translation in the rabbit reticulocyte system. *Protein engineering, design & selection : PEDS*, 17:201–4, March 2004.

[169] Matthew Levy, Karl E. Griswold, and Andrew D. Ellington. Direct selection of trans-acting ligase ribozymes by in vitro compartmentalization. *RNA (New York, N.Y.)*, 11:1555–62, October 2005.

[170] N. H. Bergman, W. K. Johnston, and D. P. Bartel. Kinetic framework for ligation by an efficient rna ligase ribozyme. *Biochemistry*, 39:3115–23, March 2000.

[171] Jeremy J. Agresti, Bernard T. Kelly, Andres Jaschke, and Andrew D. Griffiths. Selection of ribozymes that catalyse multiple-turnover diels-alder cycloadditions by using in vitro compartmentalization. *Proceedings of the National Academy of Sciences of the United States of America*, 102:16170–5, November 2005.

[172] Robert C. Edgar. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19):2460–2461, 08 2010.

[173] Robert C. Edgar. Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32:1792–7, 2004.

[174] P. C. Sabeti, P. J. Unrau, and D. P. Bartel. Accessing rare activities from random rna sequences: the importance of the length of molecules in the starting pool. *Chemistry & biology*, 4:767–74, Oct 1997.

[175] David P. Horning and Gerald F. Joyce. Amplification of rna by an rna polymerase ribozyme. *Proceedings of the National Academy of Sciences*, 113(35):9786–9791, 2016.

[176] E. H. Ekland and D. P. Bartel. The secondary structure and sequence optimization of an rna ligase ribozyme. *Nucleic Acids Research*, 23(16):3231–3238, August 1995.

[177] T. Inoue and L. E. Orgel. Oligomerization of (guanosine 5'-phosphor)-2-methylimidazolide on poly(c) - an rna-polymerase model. *Journal of Molecular Biology*, 162(1):201–217, 1982.

[178] L. E. Orgel. Evolution of genetic apparatus. *Journal of Molecular Biology*, 38(3):381–&, 1968.

[179] L. E. Orgel. Evolution of the genetic apparatus. *Journal of molecular biology*, 38:381–93, December 1968.

[180] Matthew W. Powner, Beatrice Gerland, and John D. Sutherland. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature*, 459:239–42, May 2009.

[181] Xi Chen, Na Li, and Andrew D. Ellington. Ribozyme catalysis of metabolism in the rna world. *Chemistry & biodiversity*, 4:633–55, April 2007.

[182] Hani S. Zaher and Peter J. Unrau. Selection of an improved rna polymerase ribozyme with superior extension and fidelity. *RNA (New York, N.Y.)*, 13:1017–26, July 2007.

[183] E. H. Ekland and D. P. Bartel. Rna-catalysed rna polymerization using nucleoside triphosphates. *Nature*, 382:373–6, July 1996.

[184] Hani S. Zaher and Peter J. Unrau. T7 rna polymerase mediates fast promoter-independent extension of unstable nucleic acid complexes. *Biochemistry*, 43:7873–80, June 2004.

[185] P. J. Unrau and D. P. Bartel. Rna-catalysed nucleotide synthesis. *Nature*, 395:260–3, September 1998.

[186] Matthew W. L. Lau, Kelly E. C. Cadieux, and Peter J. Unrau. Isolation of fast purine nucleotide synthase ribozymes. *Journal of the American Chemical Society*, 126:15686–93, December 2004.

[187] F.w. Aston M.a. D.sc. Lxxiv. a positive ray spectrograph. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 38(228):707–714, 1919.

[188] W. Gilbert. Origin of life - the rna world. *Nature*, 319(6055):618–618, February 1986.