UNIVERSITY OF CALIFORNIA, SAN DIEGO

Catalytic RNAs for a RNA world

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

in

Chemistry

by

Janina E. Moretti

Committee in charge:
Professor Ulrich Müller, Chair
Professor Thomas Hermann
Professor Simpson Joseph
Professor Elizabeth Komives
Professor Amy Pasquinelli

2013

The Dissertation of Janina E. Moretti is approved, and is acceptable in quality and form for publication on microform and electronically:

Chair

University of California, San Diego

2013

Epigraph

SAN DIEGO – Here in a laboratory perched on the edge of the continent, researchers are trying to construct Life As We Don't Know It in a thimbleful of liquid.

Generation of scientists, children and science fiction fans have grown up presuming that humanity's first encounter with alien life will happen in a red sand dune on Mars, or in an enigmatic radio signal from some obscure star.

But it could soon happen right here on Earth, according to a handful of chemists and biologists who are using the tools of modern genetics to try to generate the Frankensteinian spark that will jump the gap separating the inanimate and the animate. The day is coming, they say, when chemicals in a test tube will come to life.

–Dennis Overbye, New York Times, July 2011

I have to say I myself do not find pipetting colorless droplets of liquid from one plastic tube to another awfully inspiring, and that's what much of molecular biology often seems to be at the bench.

–Roger Tsien, October 2008

Table of Contents

List of Figures

Acknowledgements

I would like to thank past and current members of the Müller lab—Karen Olson, Dario Meluzzi, Chengguo Yao, Greg Dolan, and Zhaleh Amini—for sharing the day to day experience of research with me and my research advisor Uli Müller for his guidance. It has been a pleasure working with you.

Thanks to the Joseph, Tezcan, Toor, and Yang labs for generously sharing reagents and equipment as well as scientific feedback.

Many teachers have supported and challenged me throughout my education. I would especially like to thank my undergraduate research advisor and mentor Karl Haushalter for giving me the opportunity to work on my first independent research project and encouraging me to apply to graduate school.

Finally, thanks to my parents, friends, and family for their love and support. I could not have done this without you.

Chapter 2, in full, has been submitted for publication, **Moretti, J.E.** and Müller, U.F. A ribozyme that triphosphorylates RNA 5'-hydroxyl groups. The dissertation author was the first author on this paper.

Chapter 3, in full, is a reprint of the material as it appears in PLoS ONE, Yao C., **Moretti J.E.**, Struss P.E., Spall J.A., Müller U.F. (2011) Arginine Cofactors on the Polymerase Ribozyme. PLoS ONE 6(9): e25030. The dissertation author was a co-author on this paper.

This research was funded in part by the National Institutes of Health Molecular Biophysics training grant.

Vita

Harvey Mudd College, Claremont

2007 Bachelors of Science Joint Degree in Biology and Chemistry

University of California, San Diego

2009 Master of Science in Chemistry

2013 Doctor of Philosophy in Chemistry

PUBLICATIONS

Yao C., **Moretti J.E.**, Struss P.E., Spall J.A., Müller U.F. (2011) Arginine Cofactors on the Polymerase Ribozyme. PLoS ONE 6(9): e25030.

Moretti J.E. and Müller U.F. (2013) A ribozyme that triphosphorylates RNA 5' hydroxyl groups

Abstract of the Dissertation

Catalytic RNAs for a RNA world

by

Janina E. Moretti

Doctor of Philosophy in Chemistry

University of California, San Diego 2013

Professor Ulrich Müller, Chair

At some point in the origin of life, life likely went through a stage where RNA acted as genome and catalyst—the RNA world. Eventually, life evolved into the form we observe today, with DNA acting as genome and proteins acting as the main catalysts. Although there is no known direct evidence of RNA world organisms, we can gain understanding of life's early ancestors by creating catalytic RNAs in the lab that could have existed in an RNA world. This dissertation aims to develop and improve two of these catalytic RNAs (ribozymes). The first is a triphosphorylation ribozyme, a catalytic RNA that

triphosphorylates the 5’ hydroxyl groups of RNA using trimetaphosphate. In a RNA world, this activity could have been useful for chemically activating RNAs for ligation or polymerization. The second is a polymerase ribozyme, a catalytic RNA that catalyzes RNA polymerization. This activity could have been important for replication in a RNA world organism.

In order to identify catalytic RNAs that can triphosphorylate their own 5’ hydroxyl groups, an in vitro selection method was established. From a library of ~10^{14} random sequences, several active sequences were identified and one was analyzed in greater detail. This triphosphorylation ribozyme was modified to act in trans and had a reaction rate of 0.16 min^{-1} under optimal conditions. Preliminary analysis of its secondary structure suggests that it forms a four helical junction motif.

A catalytic RNA that catalyzes the polymerization of RNA was previously developed, but it is too inefficient to replicate itself. Its limitation is weak substrate binding. Here efforts were made to improve the polymerase ribozyme by reducing charge repulsion between the negatively charged ribozyme and its substrate using positively charged amino acids as cofactors. However, these cofactors did not improve ribozyme polymerization. An in vitro selection method that puts selection pressure on the ribozyme to bind its substrate in trans was developed to try to find improved polymerase ribozymes. Although no active ribozymes were found, several steps of a technically challenging in vitro selection method were established.

In summary, this thesis presents work to develop two catalytic RNAs for a RNA world. An existing polymerase ribozyme was tested with amino acid cofactors and progress was made towards developing an in vitro selection that could be used to find novel polymerase ribozymes. Also, a novel catalytic RNA that can triphoshphorylate RNAs was created. This opens up new avenues that will deepen our understanding of the RNA world.

Chapter 1

Introduction

1.1 The RNA world hypothesis

How life arose and how it evolved into the life forms present today are exciting questions that have inspired researchers for generations. The discovery of microfossils in Western Australia and carbon isotope data suggest that life on Earth originated about 3.4 billion years ago (1). According to the RNA world hypothesis, a stage existed in the early origin of life where RNA acted as both genome and catalyst. Later life evolved to use DNA as genome and proteins as the main catalyst (2-4).

Multiple lines of evidence support the RNA world hypothesis. Since RNA is capable of both carrying genetic information and catalyzing reactions, it could solve the “chicken and egg” problem of how life encoded by DNA and catalyzed by proteins could have arisen. RNA could have originally filled both roles and then later been replaced by DNA, which is more stable against hydrolysis than RNA, and protein, whose 20 amino acids offer more chemical diversity than RNA. There are many examples of catalytic RNAs in modern biology, including the ribosome which is responsible for protein synthesis in all current day life forms, the splicosome which splices pre-mRNAs in eukaryotes before transcription, RNase P which processes pre-tRNAs, group I and group II introns, and five small self-cleaving RNAs (5-14). In addition to catalyzing many reactions

in current biology, RNA is capable of catalyzing a variety of other chemical reactions including ligation, redox chemistry, Michael addition, Diels-Alder reactions, and peptide bond formations (15-18). These catalytic RNAs were created in the laboratory by selecting them from large libraries of random sequences, a method that will later be discussed in detail. Furthermore, current life has many nucleotide based cofactors (such as NAD+, FAD, CoA, ATP, cAMP, and GTP) that may be “molecular fossils”: remnants of an earlier, RNA-based organism (19).

There are many important questions about the RNA world: Is a self-replicating RNA system with open-ended evolution possible? What did the RNA world look like and how did it function? How could the RNA world arise from prebiotic chemistry? This thesis helps address these questions by developing and improving catalytic RNAs (ribozymes) that could have played important roles in an RNA world—a triphosphorylation ribozyme and a polymerase ribozyme. This introduction will review current efforts to recreate a RNA world organism in the lab using in vitro selection, a method for developing novel RNA catalysts. Also, it will provide an overview of the original research presented in this thesis.

1.2 Recreating an RNA world

While direct evidence of the RNA world may never be found, one strategy to learn about life’s early ancestors is by recreating some of their relatives in the lab. Re-creating a self-replicating RNA system in the lab would help us to

understand what an RNA based world may have looked like and how it could have evolved into today's protein and DNA based world (Fig. 1.1).

A self-replicating RNA system requires a system of catalytic RNAs capable of replication and a way to exclude parasites (sequences that consume resources by being replicated without contributing a useful function) (see 20 for review). Compartmentalization of the self-replicating RNAs is needed so that successful RNA catalysts can capture the benefits of their catalysis and exclude inactive RNA sequences. Vesicles that can grow, divide, and allow small molecules to permeate have been established (21-23). These protocells are composed of a bilayer of fatty acids. Two methods for vesicle growth and division have been established. In the first, vesicles with higher osmotic pressure grow and accrue fatty acids from neighboring vesicles with lower osmotic pressure. Since higher concentrations of RNA inside the vesicle increase osmotic pressure, this system demonstrates a way to couple cell growth and division with RNA replication. Vesicles containing replicating RNAs would increase their osmotic pressure, leading to growth and division (22). Another potential vesicle system is large multilamellar vesicles that form long thin protrusions when micelles are added. These protrusions break off to form new vesicles when exposed to weak shearing forces (23). Some vesicle chemistries are also permeable to small molecules, and are more permeable to ribose than the other aldopentoses (24) which may have contributed to the development of an RNA world.

Although vesicles that can compartmentalize self-replicating RNAs have been developed, creating self-replicating RNA ribozymes has been more

challenging. Two systems of self-replicating RNAs have been established: one is based on a ligase ribozyme (25, 26) and one based on the Azoarcus group I ribozyme (27, 28). In these systems, RNA oligonucleotide fragments come together to form ribozymes that catalyze the formation of more ribozymes by connecting these fragments. These systems start from RNA fragments with relatively long specific sequences, which would be very unlikely to exist in a prebiotic environment. An additional limitation of these systems is that they cannot undergo open ended evolution because they can only recombine themselves from the existing RNA fragments but not accumulate point mutations.

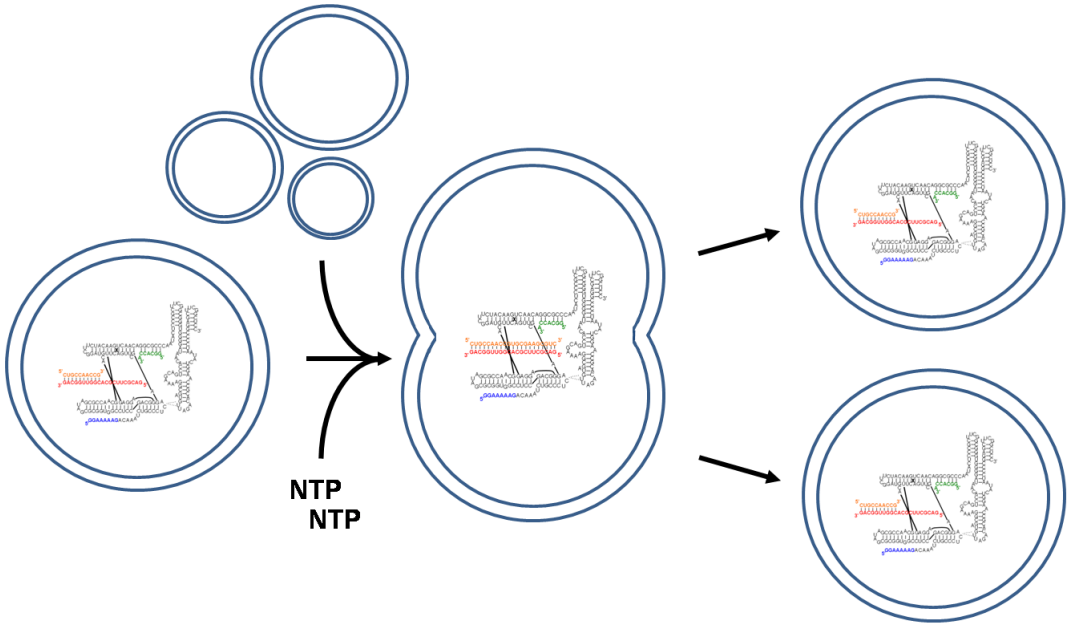

Figure 1.1: A self-replicating RNA system. This (hypothetical) self-replicating system has a lipid bilayer that is permeable to small molecules and capable of growth and division. It is shown here encapsulating ribozymes that catalyze the polymerization of RNA primers. As extended RNAs accumulate inside the vesicle, the cells grow and eventually divide. While vesicles with these properties have been developed, RNA catalysts capable of self-replication have not yet been established.

Another possible self-replicating RNA system would be one based on a replicase ribozyme—a catalytic RNA capable of catalyzing its own replication. A ribozyme that can catalyze RNA polymerization has been developed; however, it is not capable of self-replication (29). This polymerase ribozyme can catalyze the formation of 3’-5’ phosphodiester bonds between the alpha-phosphates of nucleoside triphosphates and the 3'-hydroxyl groups of an RNA primer base paired to a template strand. The polymerase ribozyme is quite accurate; with an average fidelity of 0.967 per residue. However, it is about 200 nucleotides long and its most active version can catalyze the addition of about 90 nucleotides to its primer (30). Furthermore, addition of 90 nucleotides is only possible on one specific template sequence that is preferred by the ribozyme. Much work remains to be done to develop a polymerase ribozyme that is efficient and accurate enough to be an RNA replicase. (See section 4.1 for a more in depth discussion of polymerase ribozymes and efforts to improve them.)

Another possible route for RNA replication in an RNA world is without an RNA catalyst (see 31 and 32 for review). RNA can be synthesized by combining activated nucleosides with RNA primer-template duplexes. The template aligns activated nucleosides with the primer where they react. In one of the most successful examples of template-directed RNA synthesis, a 14mer template composed of all Gs and Cs was completely copied using guanosine or cytidine 5’-phospho-2-methylimidazolides (33). However, there are still many challenges that must be addressed in order to develop syntheses that are efficient and accurate enough for RNA replication in an RNA world, including hydrolysis of

activation groups, slow reaction rates, regioselectivity (formation of a mixture of 3’-5’ and 2’- 5’ phosphodiester bonds), the requirement of high concentrations of divalent metal ions that promote hydrolysis of the product, and reaction fidelity (31).

1.3 Where did the RNA world come from?

In addition to re-creating an RNA world organism, recreating a possible path from prebiotic molecules to a self-replicating system will help us understand the origin of life (Fig 1.2). First, nucleotide monomers must be made from prebiotic starting materials. Then they must react to form RNA polymers. RNA molecules capable of catalyzing reactions beneficial to RNA world organisms would arise from these sequences and form the basis of self-replicating protocells. Eventually, this RNA-based organism would evolve into modern day life, relying mainly on protein enzymes for catalysis and DNA for information storage.

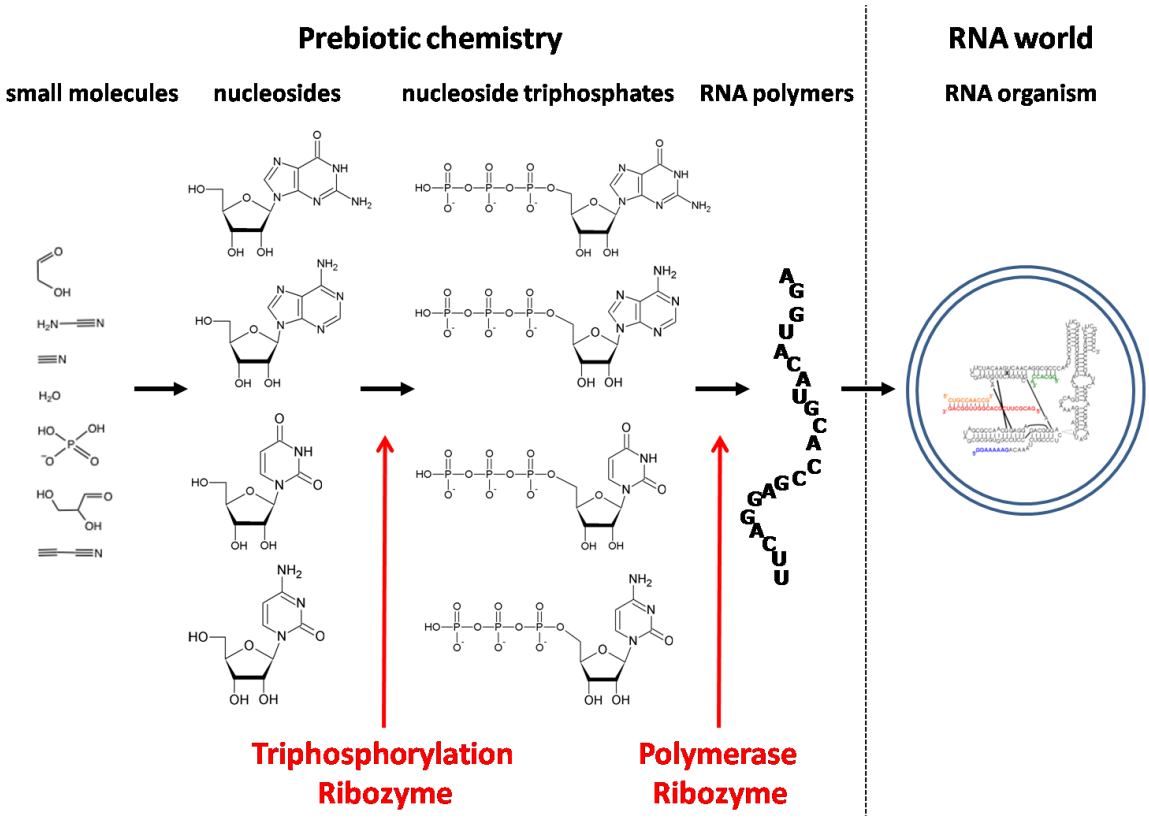

Figure 1.2: A prebiotically plausible path from small molecules to a RNA world organism. Small molecules react to form nucleosides, the building blocks of RNA. These must be activated and oligomerized to form RNA polymers. RNA molecules that are capable of catalyzing reactions that contribute to replication are encapsulated in vesicles to form an RNA world organism. Ribozymes capable of catalyzing the triphosphorylation of nucleosides or connecting activated nucleosides to form RNA polymers would have been useful in an RNA world.

The building blocks of RNA are nucleic acid monomers containing a ribose, a triphosphate, and a purine or pyrimidine base. Recently, a major advance was made towards the prebiotically plausible synthesis of the four canonical nucleosides (34, 35). Using the prebiotically plausible starting materials cyanamide, cyanoacetylene, glycolaldehyde, glyceraldehyde and inorganic phosphate, pyrimidines can be formed in a series of reactions in aqueous

solution (34) Purine precursors have been synthesized in multicomponent reactions from aminooxazoles and aminoimidazoles (35).

The next step is forming RNA polymers from nucleic acid monomers. RNA polymers can be formed from nucleoside monophosphates combined with lipid matrices and exposed to cycles of drying and wetting (36). Also, evaporation of adenosine cyclic 2',3'-phosphate in the presence of catalyst results in the formation of RNA oligonucleotides (37). A variety of other activation groups have been used to form RNA polymers such as adenine, cyanide, imidazole, and 2-methylimidazole. However, these activation groups have limited stability in water (38-42). Current biology uses triphosphates as the activation group which has the advantage of being kinetically stable in aqueous solution.

The focus of this work is to develop catalytic RNAs that could have existed and been useful in an RNA world. The first is a triphosphorylation ribozyme—an RNA that uses trimetaphosphate to triphosphorylate the 5' hydroxyl group of RNAs. This ribozyme can serve as a starting point for developing a ribozyme that triphosphorylates nucleosides. It could also be used to activate the 5' end of RNAs for 3'-5'polymerization in an early RNA world. The second is a polymerase ribozyme, an RNA capable of catalyzing RNA polymerization. Such a ribozyme has been developed, but it is not efficient enough for self-replication (43). Here efforts are made to improve it by using positively charged amino acids as cofactors and selecting a new polymerase ribozyme in the presence of these cofactors.

1.4 In vitro selection: a method for identifying novel catalytic RNA

In vitro selection is a method for developing RNA catalysts by utilizing the principles of natural selection (44, 45) (see 46 for review). Large populations of random RNA sequence are subjected to successive rounds of a selection procedure until the population is enriched with RNA sequences that perform the activity being selected for. This method has been used to generate RNA aptamers and catalysts for a variety of medical and technical applications (47-50). Here it is used to develop RNA catalysts that may have existed in an RNA world.

In vitro selections begin with large (about 10^{15} sequences) pools of randomized DNA molecules. The DNA is synthesized such that one of the four nucleotides is randomly inserted at each position. The randomized region is typically flanked by constant primer binding sites for PCR amplification, as well as a T7 RNA polymerase promoter sequence for transcription. Ribozymes are then selected from this RNA library (Fig. 1.3).

It is not clear what length of randomized sequence is best to start with in the initial pool. Usually randomized regions are no longer than ~200 nucleotides because it is technically challenging to synthesize them. However, it is unclear if longer randomized domains are better than shorter ones. The optimum length of the randomized domain depends on multiple competing influences that will vary with the specific activity being selected for. Statistically, increasing the length of the randomized domain increases the probability of finding active domains (51). However, incorrect folding is more likely in longer randomized sequences (52). At

least two experimental in vitro selections testing various randomized sequence lengths found that intermediate lengths were optimal (53, 54). In a selection for an isoleucine binding aptamer 50 to 70 nucleotides of random sequence was best; in a selection for ribozymes that synthesize coenzymes CoA, NAD, and FAD from phosphopantetheine, NMN and FMN respectively, 30 to 60 nucleotides of random sequence was best. The best length of the randomized domain also depends on how large of a motif is required to perform a specific function. Selection for activities that only require small motifs may not benefit from longer randomized domains, but if a larger, more complex structure is required, longer randomized domain are better (51). Multiple factors should be considered when choosing the length of the randomized domain in an in vitro selection, but there is no known rule dictating the optimum length of the randomized region in an in vitro selection.

Once the DNA pool is transcribed into RNA, it is then bound to (for aptamer selections) or reacted with a substrate (for catalyst selections). In in vitro selections, the substrate must be connected to a tag that can be used to separate active RNA sequences from inactive sequences. (For example, reacting RNAs with biotinylated substrates that can be pulled down on streptavidin coated magnetic particles is a common strategy.) RNA sequences that are not capable of binding or reacting with the substrate are washed away and the active sequences are reverse transcribed and PCR amplified. The amplified sequences serve as the starting material for additional rounds of selections. After successive rounds of selection, the selected pool is tested for catalytic activity. Once activity

is detected in the pool, individual sequences are cloned, tested for activity, and characterized (Fig. 1.3). In later rounds of the selection the pool may be subjected to mutagenic PCR or recombination (55-58). These techniques introduce heterogeneity into the pool sequences in order to find the most active variants of active sequences.

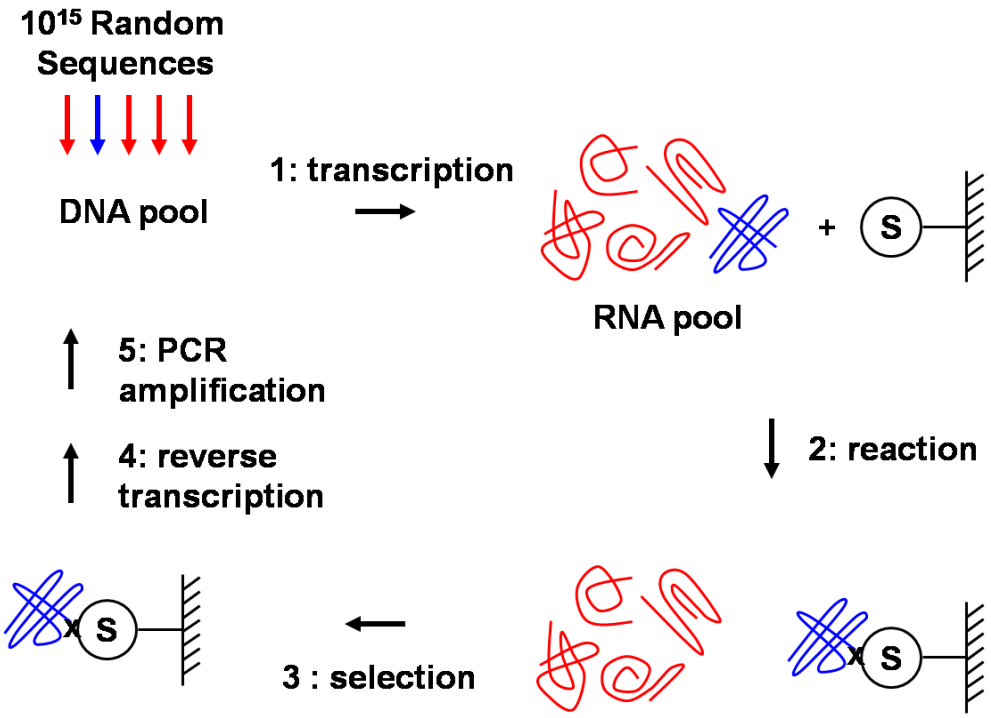

Figure 1.3: A schematic for identifying catalytic RNAs using in vitro selection. The selection begins with 10^{15} randomized DNA sequences. Most sequences encode inactive RNAs (red) while some encode active RNAs (blue). The DNA pool is transcribed and reacted with a substrate (S) that can be pulled down to separate active sequences from inactive sequences. After removal of inactive sequences, the selected sequences are reverse transcribed and PCR amplified. This pool is then used as the starting pool for additional rounds of selection.

It is usually necessary to perform multiple rounds of selection before individual, active sequences can be isolated, due to carry over of inactive RNA sequences in the selection step and the minute number of active sequences in

the starting pool. The initial pool is made as large as technically feasible (usually about 10^{15} sequences) to include as many sequences as possible and increase the likelihood of finding active RNA molecules. After each round, the pool is enriched with active sequences; the enrichment factor quantifies how many additional copies of an active sequence there are after one round of selection. For example, an enrichment factor of 100 means that the relative frequency of a specific sequence increases 100 times in each round of selection. In this case, assuming there is one active sequence in an initial pool of 10^{15} sequence, after the one round of selection there would be 100 active sequences in a background of about 10^{15} inactive sequences. At this point, there is still no detectable activity in the pool because inactive sequences still dominate the population. After an additional round of selection, each of the 100 copies of active sequence are enriched another 100 fold for 10,000 total active sequences, but this is still only 0.000000001% of the population. It is only after multiple rounds of selection that the active sequences dominate the population. At this point activity can be measured in the pool and individual clones can be sequenced and characterized.

One limitation of the in vitro selection strategy described above is that is does not permit ribozymes to perform multi-turnover reactions during the selection. Therefore, the RNA is not exposed to selection pressure for performing multi-turnover reactions. In vitro compartmentalization is a selection strategy designed to address this limitation. It has been used to select ribozymes that catalyze ligation reactions, Diels-Alder cycloadditions, and RNA polymerization (59-61). These selections are more technically challenging because the RNA

cannot covalently bind itself to its substrate during the reaction. Therefore, the sequences encoding RNAs must be connected to reacted substrates using compartmentalization, typically by encapsulating the DNA encoding the ribozymes and the ribozyme's substrate in emulsion droplets. (See section 4.1 for a more in depth discussion of selections with in vitro compartmentalization)

1.5 Goals of this thesis

The focus of this work is to develop and improve two catalytic RNAs that could have contributed important functions to RNA world organisms: triphosphorylation of RNA 5' hydroxyl groups and RNA polymerization.

Chapter 2: A ribozyme that triphosphorylates RNA 5'-hydroxyl groups

Today, RNA oligomers are made from nucleosides triphosphates. Chapter 2 describes the selection of a catalytic RNA that triphosphorylates RNA 5'-hydroxyl groups using the prebiotically plausible compound trimetaphosphate. Several active ribozymes were identified and one was studied in more detail. The triphosphorylation ribozyme was able to triphosphorylate 5'-hydroxyl groups in cis and in trans. Although eight ribozymes were tested, none were capable of catalyzing the triphosphorylation of single nucleosides. However, an RNA world organism could have used a ribozyme that triphosphorylates 5'-hydroxyl groups, if RNA polymerization was catalyzed 3'-5' (instead of 5'-3' as it is in modern biology). It is also possible that some of the untested sequences in the selection would be capable of acting on nucleosides.

Chapter 3: Arginine cofactors on the polymerase ribozyme

A ribozyme capable of catalyzing the polymerization of long, general RNA sequences would have been important for replication in an RNA world organism. Although an RNA that catalyzes template-directed RNA polymerization—a polymerase ribozyme—has been developed, it is not efficient enough for self-replication (43). It is limited by its inability to strongly bind its negatively charged substrate (62). Chapter 3 describes an attempt to improve the polymerase ribozyme by annealing short RNAs covalently linked to positively charged amino acids to it. Although the amino acid was tested with several different positions and reaction conditions, none were shown to improve RNA polymerization. This may be because a single positive charge is not sufficient to improve the reaction or because the current polymerase ribozyme is not optimized to work with the cofactors.

Chapter 4: In vitro selection for a polymerase ribozyme with improved substrate binding

Chapter 4 discusses efforts to improve the polymerase ribozyme using a selection method that puts selective pressure on ribozymes to bind their substrates in trans. To accomplish this, ribozymes must be able to act on their substrate in trans during the selection and cannot be attached to them. In order to select active ribozyme sequences, DNA sequences encoding ribozymes are attached to the ribozymes' substrates and compartmentalized in water in oil emulsion droplets. Active sequences are selected by the separation of reacted substrates (and the ribozyme genes attached to them) from unreacted

substrates. This selection did not produce active polymerase ribozyme sequences, but several steps of the selection procedure were established and optimized. Chapter 4 reports on this progress and discusses ideas for improving this selection procedure.

1.7 References

1. Wacey, D., Kilburn, M., Saunders, M., Cliff, J.B., and Brasier, M. D. (2011) Microfossils of sulphur-metabolizing cells in 3.4-billion-year-old rocks of Western Australia, *Nature Geoscience 4*, 698-702.

2. Woese, C.R. (1967) The genetic code. The molecular basis for genetic expression. Harper & Row, New York.

3. Orgel, L.E. (1968). Evolution of the genetic apparatus. *Journal of Molecular Biology 38*, 381-393.

4. Crick, F.H.C. (1968). The origin of the genetic code. *Journal of Molecular Biology* 38, 367-379.

5. Noller, H.F., Hoffarth, V., and Zimniak, L. (1992) Unusual resistance of peptidyl transferase to protein extraction procedures. *Science 256*, 1416-1419.

6. Steitz, T.A., and Moore, P.B. (2003) RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochemical Science 28*, 411-418.

7. Valadkhan, S., Mohammadia, A., Jaladat, Y., and Geisler, S. (2009) Protein-free small nuclear RNAs catalyze a two-step splicing reaction. *Proc Natl Acad Sci USA 106*, 11091-11096.

8. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell 35*, 849-857.

9. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell 31*, 147-157.

10. van der Veen, R., Arnberg, A.C., van der Horst, G., Bonen, L., Tabak, H.F. and Grivell, L.A. (1986) Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro. *Cell 44*, 225-234.

11. Hutchins, C.J., Rathjen, P.D., Forster, A.C. and Symons, R.H. (1986) Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid. *Nucleic Acids Res* 14, 3627-3640.

12. Buzayan, J.M., Gerlach, W.L. and Bruening, G. (1986) Satellite tobacco ringspot virus RNA: A subset of the RNA sequence is sufficient for autolytic processing. *Proc Natl Acad Sci USA 83*, 8859-8862.

13. Sharmeen, L., Kuo, M.Y., Dinter-Gottlieb, G. and Taylor, J. (1988) Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. *J Virol 62*, 2674-2679.

14. Saville, B.J. and Collins, R.A. (1990) A site-specific self-cleavage reaction performed by a novel RNA in Neurospora mitochondria. *Cell 61*, 685-696.

15. Ekland, E. H., Szostak, J. W., and Bartel, D. P. (1995) Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science 269*, 364-370.

16. Tsukiji, S., Pattnaik, S.B., and Suga, H. (2003) An alcohol dehydrogenase ribozyme. *Nature Structural Biology 10*, 713-717.

17. Sengle, G., Eisenfuhr, A., Arora, P.S., Nowick, J.S., and Famulok, M. (2001). Novel RNA catalysts for the Michael reaction. *Chemical Biology 8*, 459-473.

18. Wiegand, T.W., Janssen, R.C., and Eaton, B.E. (1997) Selection of RNA amide synthases. *Chemical Biology 4*, 675-683.

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

20. Muller, U. F. (2006) Re-creating an RNA world. *Cellular and Molecular Life Sciences 63*, 1278-1293.

21. Hargreaves, W. R. and Deamer, D. W. (1978) Liposomes from ionic, single chain amphiphiles. *Biochemistry 17(18)*, 3759-3768.

22. Chen, I. A., Robers, R. W., and Szostak, J. W. (2004) The empergence of competition between model protocells. *Science 305*, 1474-1476.

23. Zhu, T. F. and Szostak, J. W. (2009) Coupled growth and division of model protocell membranes. *Journal of American Chemical Society 131(15)*, 5705-5713.

24. Sacerdote, M. G. and Szostak, J. W. (2005) Semipermeable lipid bilayers exhibit diastereoselectivity favoring ribose. *Proceedings of the National Academy of Sciences 102*, 6004-6008.

25. Lincoln, T. A. and Joyce, G. F. (2009) Self-sustained replication of an RNA enzyme. *Science 323*, 1229-1232.

26. Sczepanski, J. T. and Joyce, G. F. (2012) Synthetic evolving systems that implement a user-specified genetic code of arbitrary design. *Chemistry & Biology 19*, 1324-1332.

27. Hayden, E. J. and Lehman, N. (2006) Self-assembly of a group I intron from inactive oligonucleotide fragments. Chemistry & Biology 13 (8), 909-918.

28. Vaidya, N., Manapat, M. L., Chen, I. A., Xulvi-Brunet, R., Hayden, E. J., and Lehman, N. (2012) Spontaneous network formation among cooperative RNA replicators. *Nature 491*, 72-77.

29. Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E., and Bartel, D. P. (2001) RNA-Catalyzed RNA polymerization: accurate and general RNA-templated primer extension, *Science 292*, 1319-1325.

30. Wochner, A., Attwater, J., Coulson, A., and Holliger, P. (2011) Ribozyme-catalyzed transcription of an active ribozyme, *Science 332*, 209-212.

31. Szostak, J. W. (2012) The eightfold path to non-enzymatic RNA replication. *Journal of Systems Chemistry 3*, 1-14.

32. Orgel, L. E. (2004) Prebiotic chemistry and the origin of the RNA world. *Critical Reviews in Biochemisty and Molecular Biology 39*, 99-123.

33. Acevedo, O. L., and Orgel, L. E. (1987) Non-enzymatic transcription of an oligodeoxynucleotide 14 residues long. *Journal of Molecular Biology 197*, 187-193.

34. Powner, M. W., Gerland, B., and Sutherland, J. D. (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature 459*, 239-242.

35. Powner, M. W., Sutherland, J. D., and Szostak, J. W. (2010) Chemoselective multicomponent one-pot assembly of purine precursors in water. *Journal of the American Chemical Society 132*, 16677-16688.

36. Rajamani, S., Vlassov, A., Benner, S., Coombs, A., Olasagasti, F., and Deamer, D. (2008). Lipid-assisted synthesis of RNA-like polymers from mononucleotides. *Origins of Life and Evolution of Biospheres 38*, 57-74.

37. Verlander, M. S., Lohrmann, R., and Orgel, L. E. Catalysts for the self-polymerization of adenosine cyclic 2',3'-phosphate. *Journal of Molecular Evolution 2(4)*, 303-316.

38. Prabahar, K.J., and Ferris, J.P. (1997). Adenine derivatives as phosphate-activating groups for the regioselective formation of 3',5'-linked oligoadenylates on montmorillonite: possible phosphate-activating groups for the prebiotic synthesis of RNA. *Journal of the American Chemical Society 119*, 4330-4337.

39. Lohrmann, R., and Orgel, L.E. (1968). Prebiotic synthesis:phosphorylation in aqueous solution. *Science 161*, 64-66.

40. Lohrmann, R. (1977). Formation of nucleoside 5'-phosphoramidates under potentially prebiological conditions. *Journal of Molecular Evolution 10*, 137-154.

41. Sawai, H., and Orgel, L.E. (1975). Letter: Oligonucleotide synthesis catalyzed by the Zn-2+ ion. *Journal of the American Chemical Society 97*, 3532-3533.

42. Inoue, T., and Orgel, L.E. (1982). Oligomerization of (guanosine 5'-phosphor)-2-methylimidazolide on poly(C). An RNA polymerase model. *Journal of Molecular Biology 162*, 201-217.

43. Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E., and Bartel, D. P. (2001) RNA-Catalyzed RNA polymerization: accurate and general RNA-templated primer extension, *Science 292*, 1319-1325.

44. Tuerk, C. and Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. (1990) *Science 249*, 505-510.

45. Ellington, A. D. and Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. (1990) *Nature 346*, 818-822.

46. Joyce, G. F. Directed evolution of nucleic acid enzymes. (2004) *Annual Review of Biochemistry 73,* 791-836*.*

47. Zimbres, F. M., Tarnok, A., Ulrich, H. and Wrenger, C. (2013). Aptamers: novel molecules as diagnostic markers in bacterial and viral infections? *Biomed Research International 2013*, 1-7.

48. Wang, T. and Ray, J. (2012) Aptamer-based molecular imaging. *Protein Cell 3(10)*, 739-754.

49. Hesselberth, J., Robertson, M. P., Jhaveri, S. and Ellington, A. D. (2000) In vitro selection of nucleic acids for diagnostic applications. *Reviews in Molecular Biotechnology 74(1)*, 15-25.

50. Mastroyiannopoulos, N. P. Uney, J. B., and Phylactou, L. A.(2010) The application of ribozymes and DNAzymes in muscle and brain. *Molecules 15(8)*, 5460-5472.

51. Sabeti, P. C., Unrau, P. J., and Bartel, D. P. Accessing rare activities from random RNA sequences: the importance of the length of molecules in the starting pool. (1997) *Chemistry and Biology 4*, 767-774.

52. Knight, R., Sterck, H. D., Marke, R., Smit, S., Oshmyansky, A., and Yarus, M. Abundance of correctly folded RNA motifs in sequence space, calculated on computational grids. (2005) *Nucleic Acids Research 33*, 5924-5935.

53. Lozupone C., Changayil S., Majerfeld I., and Yarus M. Selection of the simplest RNA that binds isoleucine (2003) *RNA 9*, 1315–1322.

54. Huang F., Bugg C.W., and Yarus M. RNA-catalyzed CoA, NAD, and FAD synthesis from phosphopantetheine, NMN, and FMN. (2000) *Biochemistry. 39*, 15548–15555.

55. Bartel D.P., Zapp,M.L., Green,M.R. and Szostak,J.W. (1991) HIV-1 Rev regulation involves recognition of non-Watson–Crick base pairs in viral RNA. *Cell 67*, 529–536.

56. Cadwell, R.C. and Joyce, G.F. (1992) Randomization of genes by PCR mutagenesis. *PCR Methods Appl*, 2, 28-33.

57. Zhao, H., Giver, L., Shao, Z., Affholter, J. A., and Arnold, F. H. (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nature Biotechnology 16(3)*, 258-261.

58. Lehman, N. and Unrau, P. J. (2005) Recombination during in vitro evolution. *Journal of Molecular Evolution 61*, 245-252.

59. Agresti, J. J., Kelly, B. T., Jaschke, A., and Griffiths, A. D. Selection of ribozymes that catalyze multiple-turnover Diels-Alder cycloadditions by using in vitro compartmentalization. (2005) *Proceedings of the National Academy of Sciences 120(45)*, 16170-16175.

60. Levy, M., Griswold, K., and Ellington, A. D. Direct selection of trans-acting ligase ribozymes by in vitro compartmentalization. (2005) *RNA 11*, 1555-1562.

61. Zaher, H.S., and Unrau, P. J. (2007) Selection of an improved RNA polymerase ribozyme with superior extension and fidelity, *RNA 12*, 1017-1026.

62. Lawrence, M. S., and Bartel, D. P. (2003) Processivity of ribozyme-catalyzed RNA polymerization, *Biochemistry 42*, 8748-8755.

Chapter 2

A ribozyme that triphosphorylates RNA 5’ hydroxyl groups

2.1 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 whether RNA world organisms could have used cyclic trimetaphosphate as an energy source we developed an in vitro selection strategy for isolating ribozymes that catalyze the triphosphorylation of RNA 5'-hydroxyl groups with trimetaphosphate. Several active sequences were isolated, and one ribozyme was analyzed in more detail. The ribozyme was truncated to 96 nucleotides while retaining full activity. It was converted to a trans-format and reacted with rates of 0.16 min^{-1} under optimal conditions. The secondary structure appears to contain a four-helical junction motif. This study showed that ribozymes can use trimetaphosphate to triphosphorylate RNA 5'-hydroxyl groups, and suggested that RNA world organisms could have used trimetaphosphate as their energy source.

2.2 Introduction

The early evolution of life went through a stage in which RNA molecules served as genome and as genome-encoded catalyst, according to the RNA world hypothesis (1-3). This hypothesis is widely accepted due to several lines of

evidence, including the identification of eleven catalytic RNAs (ribozymes) in biology (4-14), that protein synthesis in all extant organisms is catalyzed by a ribozyme, the ribosome (12,15), that the hypothesis can explain how today's interdependent systems of DNA and proteins could arise, and that in vitro selected RNAs can catalyze a wide range of chemical reactions (16-25). Direct fossil evidence of the RNA world, however, may never be found because the expected chemical components of these organisms would almost certainly not have survived more than three billion years until today (26).

To determine how an RNA world organism could have functioned, researchers are attempting to re-create an RNA world organism in the lab (27-32). Two components are necessary for such systems: self-replicating ribozyme systems and encapsulations. Encapsulations would be necessary to exclude molecular parasites and to contain the products of metabolism. Lipid vesicle encapsulations have been identified that can grow, divide, and are permeable to specific small molecules (33-37). In contrast, the identification of a suitable self-replicating ribozyme system appears more elusive. Self-replicating ribozyme systems have been generated but currently they do not have the potential for open-ended evolution (32,38-40). RNA may have been the first genetic polymer; this is a controversial idea (41) that is supported by the discovery of synthetic pathways with possible prebiotic relevance (42-44) and challenged by remaining problems with prebiotic synthesis of RNA and the observation that the canonical nucleobases likely underwent some form of evolution before being chosen as genetic material (45-47). Independent of this question, the replication of RNA in

RNA world organisms would have required the polymerization of nucleotides. Ribozymes that catalyze template-directed RNA polymerization have been developed (30), and variants can synthesize a short ribozyme (48).

The polymerization of RNA is entropically unfavorable. Strategies to energetically drive RNA polymerization may have been different in two phases of the RNA world. First, in the prebiotic phase and perhaps in an early phase of the RNA world, RNA polymerization may have been driven by the evaporation of water from a solution (49) and from lipid-encapsulations (50), or by the activation of the nucleotide's 5'-phosphate with a wide range of leaving groups (51) such as adenine (52), cyanide (53), imidazole (54,55), or 2-methylimidazole (56). Importantly, these activation groups lead to hydrolysis in aqueous environment within hours. Second, in later phases of the RNA world, when lipid vesicles presumably encapsulated aqueous droplets of self-replicating RNA systems (29,35,57,58), kinetically more stable activation groups would have been important. This requirement is fulfilled by nucleoside triphosphates, the universal energy currency in all known life forms.

Nucleoside triphosphates can be generated from nucleosides and cyclic trimetaphosphate (59). Trimetaphosphate is a prebiotically plausible compound because it can be generated by volcanic activity (60), by the erosion of phosphide minerals (61), and by heating of phosphate in the presence of urea (62). Additionally, trimetaphosphate results from the self-reaction of the termini of linear polyphosphates in aqueous medium (63). Although trimetaphosphate is the most reactive polyphosphorylating reagent of alcohols among all polyphosphates

(64), the reaction between nucleoside 5'-hydroxyl groups and trimetaphosphate proceeds efficiently only at pH values above 12 (59). At such high pH, RNA world organisms would hydrolyze quickly. Therefore, RNA world organisms would require a catalyst to use trimetaphosphate as energy source.

The triphosphorylation of nucleoside 5'-hydroxyl groups with trimetaphosphate appears to be well within the range of catalysis by ribozymes. Most ribozymes are acid-base catalysts and have a particular capacity for phosphoryl transfer reactions (4-7,16,17,22,24,65,66). To identify ribozymes for new reactions, the only known route is the in vitro selection from large pools of random RNA sequences (16,67,68). Therefore we aimed to develop an in vitro selection scheme for the isolation of ribozymes that catalyze the triphosphorylation of RNA 5'-hydroxyl groups using trimetaphosphate.

Here we describe the in vitro selection of a ribozyme that uses trimetaphosphate to triphophorylate RNA 5'-hydroxyl groups. After five to eight rounds of selection we obtained several active ribozymes, one of which was analyzed in more detail. This ribozyme was able to triphosphorylate RNAs in cis and in trans, with a k_{cat} of up to 0.12 min^{-1}. These results suggested that RNA world organisms could have used trimetaphosphate as their energy source.

2.3 Results

In vitro selection

An in vitro selection procedure was established to obtain ribozymes that use trimetaphosphate to triphosphorylate RNA 5'-hydroxyl groups (Fig. 2.1). First, a double-stranded DNA pool was generated that encoded the T7 RNA polymerase promoter, a hammerhead ribozyme, and 150 randomized nucleotides flanked by two constant primer binding sites. The hammerhead ribozyme served to generate 5'-hydroxyl groups at the 5'-termini of the pool molecules (69). The hammerhead ribozyme cleaved itself efficiently off the pool under transcription conditions so that > 95% of the pool was cleaved after two hours of transcription incubation (data not shown). The gel purified pool molecules were then incubated for three hours with 100 mM $MgCl_2$, 50 mM cyclic trimetaphosphate (TMP), and 50 mM Tris/HCl pH 8.3 to allow catalytically active pool molecules to triphosphorylate their own 5'-terminus. We found that TMP chelated an equimolar amount of Mg^{2+} (data not shown), therefore the triphosphorylation conditions provided 50 mM free magnesium ions. To isolate pool molecules that carried a 5'-triphosphate, the R3C ligase ribozyme was used (76). This ribozyme created 3'-5'-phosphodiester bonds between 5'-triphosphorylated pool molecules and the 3'-hydroxyl group of a 5'-biotinylated oligonucleotide. The biotin modification allowed capturing the pool on streptavidin-coated magnetic beads, and stringent washing of pool molecules that were coupled covalently to the biotin moiety. The captured pool molecules were reverse transcribed and PCR amplified. In a second PCR step, the T7 RNA polymerase promoter and the 5'-terminal hammerhead ribozyme sequence were regenerated, completing one round of the selection. In a mock selection round

we estimated that 5'-triphosphorylated RNAs were enriched at least 9,000-fold in a single selection round (Fig. S2.1).

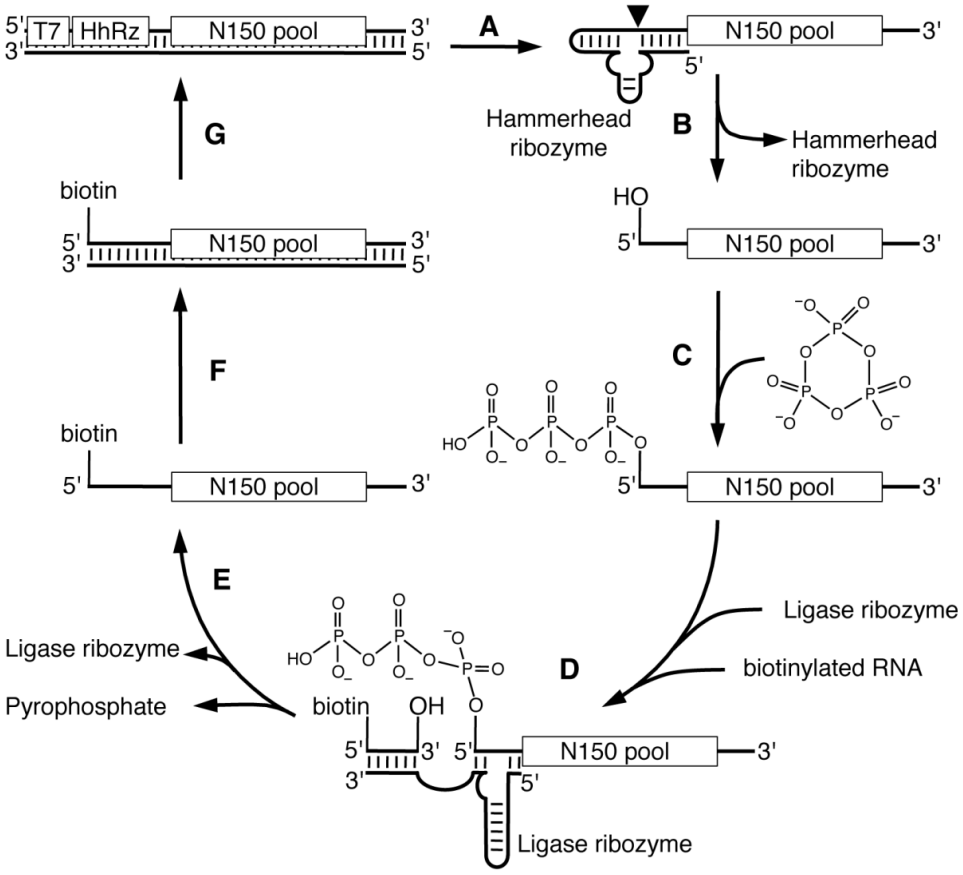

Figure 2.1: Scheme for the in vitro selection of triphosphorylation ribozymes. A) A DNA library containing the T7 RNA polymerase promoter (T7), the sequence for a hammerhead ribozyme (HhRz), and a randomized sequence (N150) was transcribed. B) The 5'-terminal hammerhead ribozyme cleaved itself off co-transcriptionally (filled triangle), generating a 5'-terminal hydroxyl group on the RNA. C) The RNA library was incubated in the presence of trimetaphosphate so active ribozymes could triphosphorylate their 5' end. D) Triphosphorylated RNA molecules were reacted with the 3'-hydroxyl group of a short, biotinylated RNA, by the R3C ligase ribozyme. E) Ligated RNAs were captured via their biotin modification on streptavidin-coated magnetic beads, and washed stringently. The RNAs were then F) reverse transcribed and G) PCR amplified to generate the DNA pool for the next round of selection.

Based on the effective complexity of the RNA pool of about $1.7 \cdot 10^{14}$ sequences (see materials and methods) and the enrichment factor of about 10^4-fold per selection round (Fig. S2.1) we anticipated that catalytically active sequences could dominate the RNA pool after four selection rounds. To follow the progress of the selection we monitored the number of PCR cycles that were required after each reverse transcription reaction to generate clearly detectable bands on agarose gels (Fig. S2.2). Satisfyingly, the number of PCR cycles dropped from at least 17 cycles in rounds one to three to 8 cycles in round four. Two modifications were used for the later selection rounds. First, mutagenic PCR was introduced in the amplification steps to allow the pool molecules to explore their sequence neighbors and perhaps find more active sequences. Second, the population was split into two branches, one exposed to the same selection pressure as before (three hours of incubation with trimetaphosphate), and one exposed to high selection pressure (only five minutes of incubation with trimetaphosphate). In round seven, a crossover branch was added to allow clones from low selection pressure to cross into the high selection pressure branch of round eight. A total of 36 sequences were obtained from the pools after five and eight selection rounds, from both low and high selection pressure branches (Fig. S2.3). The clones were named after their selection history: R5 or R8 denoting their round of isolation, a combination of the numbers 3 and 5 to denote the incubation times with trimetaphosphate (three hours and/or five minutes), and C with the clone number.

Identification of the most active ribozyme clones

To identify ribozymes with high activity we screened the 36 isolated clones for activity. This screen used the same procedure as during the selection, reacting the ribozyme with TMP followed by ligation to a short RNA using the R3C ligase ribozyme, but a 5'-radiolabeled oligonucleotide was used instead of a biotinylated oligonucleotide (Fig. S2.4). The fraction of shifted oligonucleotides in a denaturing gel reported the ligation efficiency and indirectly the triphosphorylation efficiency. The 36 ribozyme clones were sorted according to their activity, which revealed a wide distribution of activities (Fig. 2.2). Several sequences appeared multiple times but their frequency was not correlated with increased activity. The activity of several ribozyme clones in this assay was about two-fold higher than that of a positive control, made from pool RNA containing a 5'-triphosphate. We assumed that this was because the assay was not directly monitoring triphosphorylation activity, and the clones with peculiarly high activity worked especially well together with the R3C ligase ribozyme in the selection procedure. To identify those ribozyme clones that were most efficient in catalyzing the triphosphorylation reaction we chose the eight most active clones from the ligase ribozyme based assay.

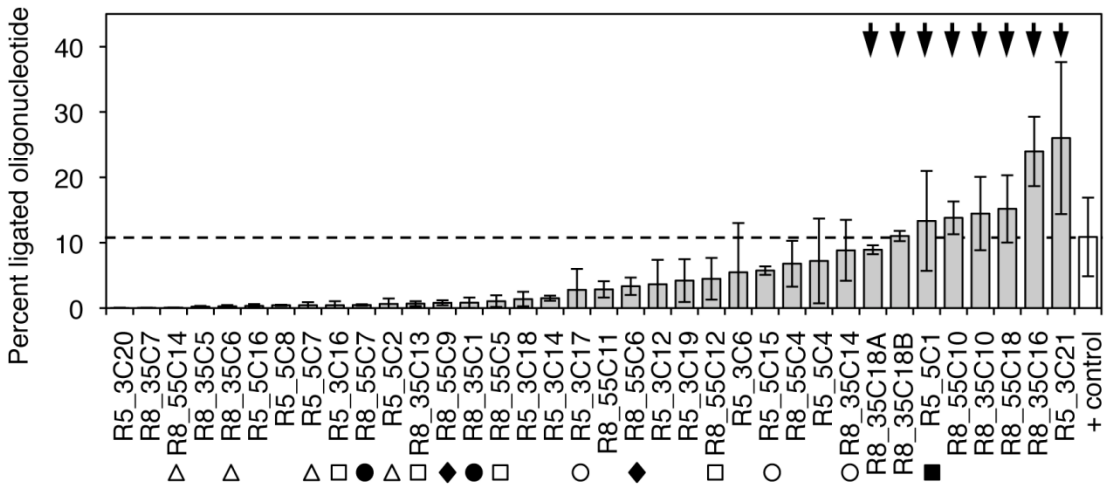

Figure 2.2: Screen of 36 in vitro selected RNAs for self-triphosphorylation activity, using an assay based on the R3C ligase ribozyme (see Fig. S2.4). The column heights show the percent of 5'-[^{32}P]-radiolabeled oligonucleotide that was ligated to the individual selected RNAs by the ligase ribozyme, after the selected RNAs were incubated with trimetaphosphate for three hours to allow for self-triphosphorylation. Therefore, the percent of ligated oligonucleotide was an indirect measure of self-triphosphorylation activity. As positive control, the unselected RNA pool was ligated to the radiolabeled oligonucleotide, facilitated by a 5'-triphosphate that was incorporated during transcription (white column and horizontal, dashed line). The labels on the x-axis denote the clone names (round of isolation_branch of the evolution_clone number). The clones were sorted according to the average percent of ligated oligonucleotide. The sequences of all 36 clones are shown in figure S2.3. Classes of related sequences are indicated by symbols, with empty squares (class 1), empty triangles (class 2), empty circles (class 3), filled circles (class 4), filled squares (class 5), and filled diamonds (class 6) each belonging to one class. Class 5 is represented by a single clone because the second clone had the identical sequence. No other clone showed related sequences among the 36 clones. The eight clones with the highest activity (arrows) were chosen for further analysis. Error bars are standard deviations from three independent experiments.

To directly monitor the triphosphorylation status at the ribozyme 5'-termini, a different gel shift assay was used (Fig. 2.3). Because the gel shift of the triphosphate group was too small to detect migration differences of the 182-

nucleotide long ribozymes we used the 8-17 DNAzyme (70) to cleave off the eight 5'-terminal nucleotides. The triphosphorylation ribozymes were internally labeled; therefore, the triphosphorylation status of the eight nucleotide fragment could be followed after separation on denaturing polyacrylamide gels and phosphoimaging. The eight nucleotide fragments of unreacted ribozymes comigrated with RNAs containing a 5'-hydroxyl terminus, and the fragment after incubation with trimetaphosphate comigrated with RNAs containing a 5'-triphosphate. This confirmed that the products were indeed triphosphorylated at their 5'-terminus. Quantitation of the signals after different triphosphorylation reaction times resulted in single-exponential time dependence, revealing pseudo-first order rate constants between 0.01 and 0.03 min^{-1} for each of the eight analyzed ribozymes. While most ribozymes reacted to a final extent of 30 - 50%, one ribozyme (clone R5_3C21) reacted to more than 80%. This clone was chosen for further analysis.

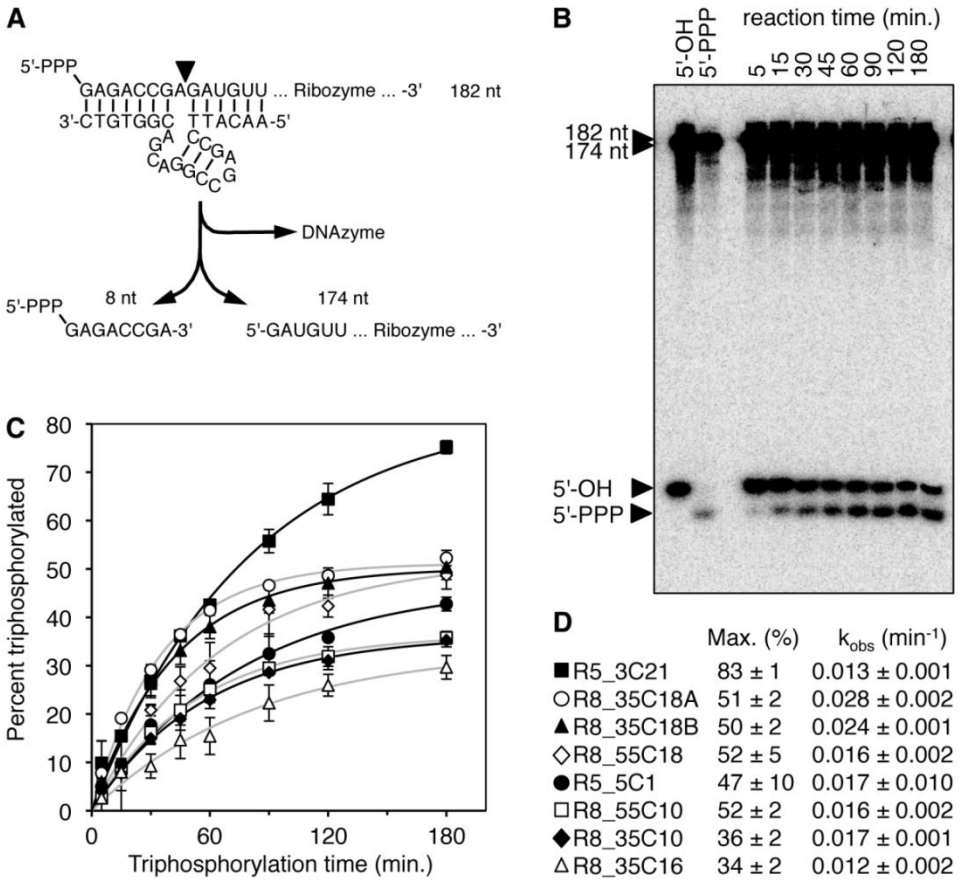

Figure 2.3: Kinetic analysis of the eight most promising ribozyme clones. A) Schematic of the assay used. After reaction with trimetaphosphate, internally radiolabeled ribozymes were cleaved by the DNAzyme to free the eight 5'-terminal nucleotides, facilitating gel separation of triphosphorylated and unreacted RNAs. B) Products after the DNAzyme reaction are separated on denaturing PAGE. An unreacted RNA (5'-OH) and a RNA that was transcribed with a 5'-terminal triphosphate (5'-PPP) were used as negative and positive controls, respectively. The incubation times with trimetaphosphate are indicated on the top. The long fragment of the cleaved ribozymes (174 nt) and uncleaved ribozymes (182 nt) migrated much slower than the eight-nucleotide fragments. The fragments were separated based on their phosphorylation status. The particular reaction shown is from clone R8_35C18A. C) Determination of triphosphorylation kinetics from signals as shown in B). The percent of triphosphorylation of the 8-mer was plotted as function of the incubation time with trimetaphosphate. Error bars are standard deviations from three experiments. Symbols are explained in D). Single-exponential curve fits are shown in black lines for filled symbols and grey lines for empty symbols. D) Symbols and clone names used in C), together with the parameters obtained by curve fits, the maximal percentage of reacted ribozyme (Max.) and the observed pseudo-first order rate constant (k_{obs}).

If the ribozymes were able to triphosphorylate single, free nucleosides they could directly generate nucleoside triphosphates. To test whether the eight most active ribozymes could triphosphorylate free nucleosides they were truncated at their 5'-terminus by one nucleotide (to accommodate the free nucleoside), and incubated with C^{14}-radiolabeled guanosine. The products were separated by PEI cellulose TLC analysis and analyzed by autoradiography (Fig. S2.5). None of the tested ribozymes generated a detectable signal for GTP. The result was the same when the ribozymes were not truncated, or when they were incubated at pH 9.5 or at different temperatures (40°C and -20°C) (data not shown).

Truncation of ribozyme R5_3C21

We removed fragments from different portions of the 182-nucleotide long R5_3C21 ribozyme to test whether a shorter, fully active ribozyme could be obtained (Fig. 2.4). The activity of the resulting truncated ribozymes was measured with the gel shift assay that directly monitored the 5'-triphosphorylation with the help of the 8-17 DNAzyme. The choice of deleted ribozyme fragments was based on the predicted secondary structures of the full-length ribozyme (71). When portions of the 5'-domain of the ribozyme were deleted, triphosphorylation activity was abolished (truncation 1 and 2). When portions of the 3'-domain were deleted, full triphosphorylation activity was retained (truncations 3 and 4). This allowed truncating the ribozyme to 96 nucleotides, while further truncation to 92 nucleotides caused a slight reduction in activity (truncation 5). Therefore the

further experiments used truncation 4 of ribzyme R5_3C21, with a length of 96 nucleotides. This ribozyme was termed TPR1.

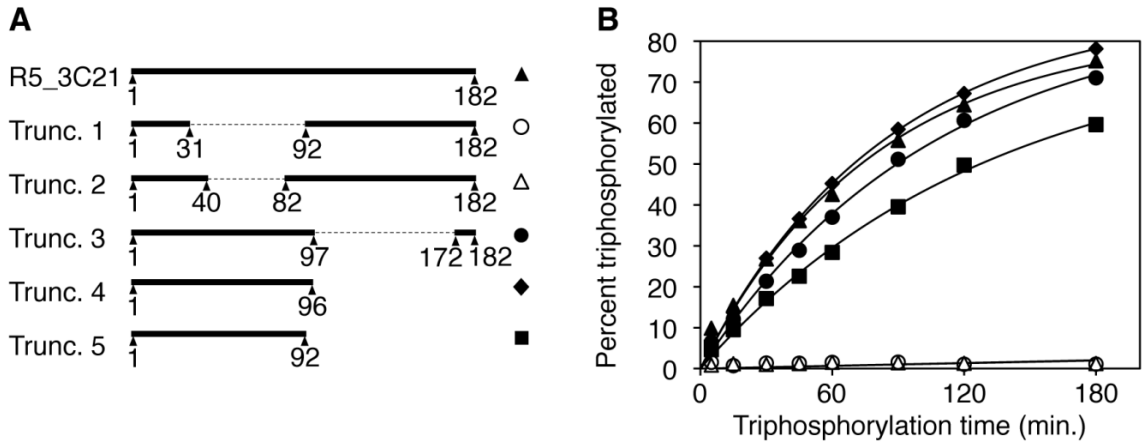

Figure 2.4: Truncation analysis of the ribozyme clone R5_3C21. A) Schematic of the tested truncations. The numbering is relative to the complete sequence of the 182-nucleotide long initial isolate (Fig. S2.3). Dotted lines indicate internal segments that were removed. Symbols to the right of each construct are consistent with the symbols in B). B) Triphosphorylation kinetics for the initial isolate and five truncated sequences. The kinetics were determined with the DNAzyme assay (Fig. 2.3). Black lines represent single-exponential curve fits to the data. Note that the symbols of truncations 1 and 2 are overlapping near the x-axis.

Triphosphorylation of short RNAs in trans

We were interested to see whether the TPR1 ribozyme would be able to triphosphorylate a short substrate RNA in trans (Fig. 2.5). For this purpose, a short hairpin structure near the ribozyme 5'-terminus was separated at its loop, and the duplex was elongated to different lengths (six to nine base pairs). This

was done using substrates with a length of 12, 13, 14, and 15 nucleotides that showed complementarity to the corresponding ribozyme 5'-termini. After incubating the radiolabeled substrates with the corresponding TPR1 ribozymes, the products were separated on denaturing polyacrylamide gels and the signals quantitated. The results showed that the TPR1 ribozyme was indeed triphosphorylating the substrates in trans. The shortest substrate, which formed a duplex of six base pairs with the ribozyme, resulted in slow reaction kinetics. In contrast, longer substrates that formed duplexes of seven to nine base pairs showed the same kinetics, and reacted to the same extent of reaction as the cis-acting TPR1 ribozyme. The duplex length of eight base pairs was chosen for further experiments, corresponding to a 14-nucleotide substrate RNA.

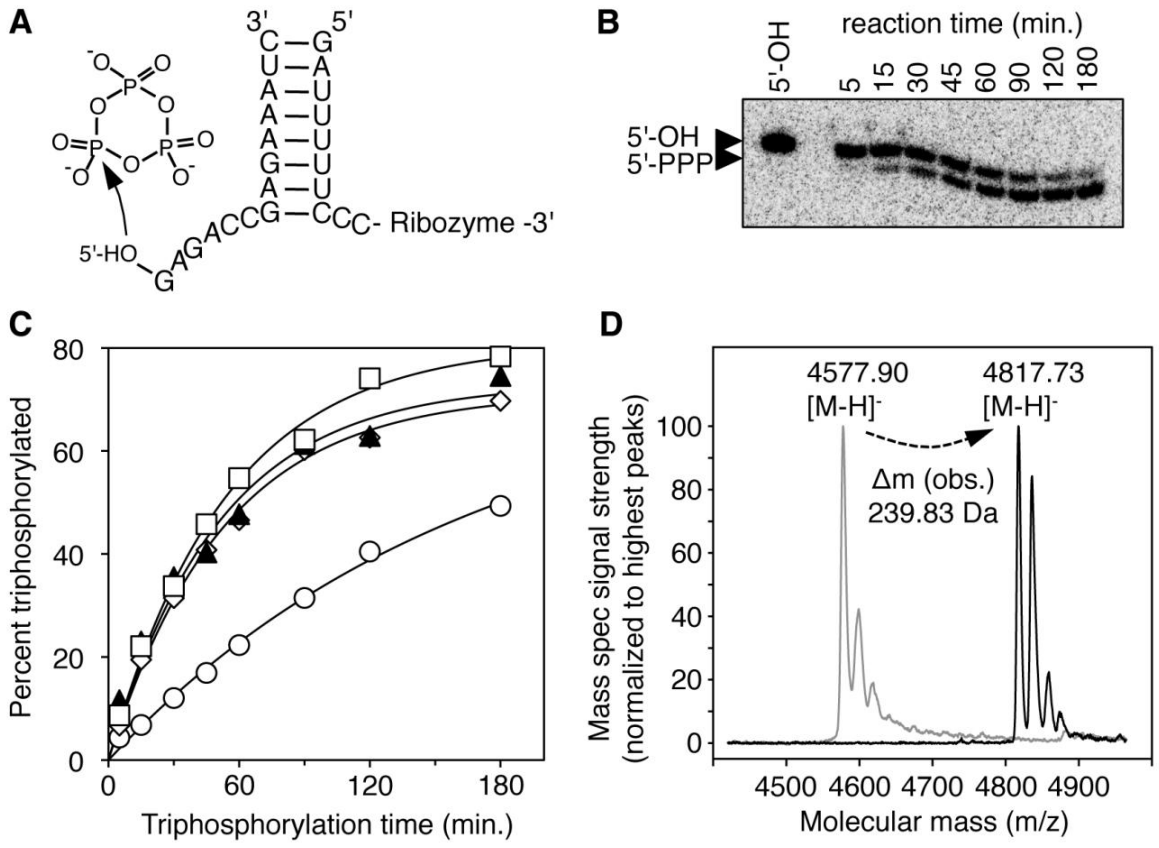

Figure 2.5: Trans-triphosphorylation of substrate RNAs, by truncated versions of the R5_3C21 ribozyme. A) Secondary structure schematic of a 14-nucleotide substrate recognized by the ribozyme with an eight base pair substrate recognition duplex. Trimetaphosphate is shown to highlight the reacting 5'-hydroxyl group. The body of the ribozyme is not shown for clarity (see Fig. 2.6). B) Autoradiogram of reaction products between substoichiometric concentrations of [^{32}P]-radiolabeled substrate and trimetaphosphate, catalyzed by 5.5 μM of the ribozyme, after separation on denaturing 22.5% PAGE. The particular image is from the reaction with a substrate recognition duplex of eight base pairs. C) Reaction kinetics of the trans-reaction with different lengths of the substrate recognition duplex. Symbols in the graph correspond to a duplex length of six base pairs (empty circles), seven base pairs (empty diamonds), eight base pairs (filled triangles), and nine base pairs (empty squares). Curved lines are single-exponential fits to the data. D) Mass spectral analysis of the substrate and the product of the trans-triphosphorylation reaction. The 14-nucleotide substrate shown in A) was analyzed before (grey lines) and after (black lines) incubation with the trans-acting ribozyme under triphosphorylation reaction conditions. The expected mass increase by 5'-triphosphate was 239.94 Da. Note that substrate and product carried a 2'-3' cyclic phosphate due to their method used to synthesize them.

Several lines of evidence suggested that the TPR1 ribozyme was generating 5'-triphosphate groups: the reaction between nucleoside 5'-hydroxyl groups and trimetaphosphate was established previously (59), the reacted ribozymes could be ligated by the R3C ligase ribozyme (Fig. S2.4), and the ribozyme reaction products had the same gel shift as a 5'-triphosphate (Fig. 2.3B). In addition, the trans-reaction presented a good opportunity for a more rigorous test. The mass of the short substrate RNA was analyzed before and after the reaction with ribozyme and trimetaphosphate (Fig. 2.5D). The expected mass difference was 239.94 Da and the observed mass difference (239.83 Da) was in excellent agreement with the prediction, confirming that the in vitro selected ribozymes used trimetaphosphate to triphosphorylate the RNA 5'-hydroxyl groups.

Secondary structure analysis

To identify single-stranded nucleotides in the ribozyme, we used SHAPE analysis with 1-methyl-7-nitroisatoic anhydride (1M7) (72). 1M7 covalently reacts with 2'-hydroxyl groups that can adopt a specific geometry, and can therefore discriminate flexible nucleotides from double-stranded nucleotides, which is used to determine RNA secondary structures (73). The 1M7 reactivities with our ribozyme identified several flexible regions and several regions that were likely base paired (Fig. 2.6A). These constraints were used to build a secondary structure model that contained a central four-helix junction (Fig. 2.6B). This structure was identified previously in RNA, such as between the spliceosomal

RNAs U2 and U6 and in the internal ribosomal entry site of HCV (74, 75).

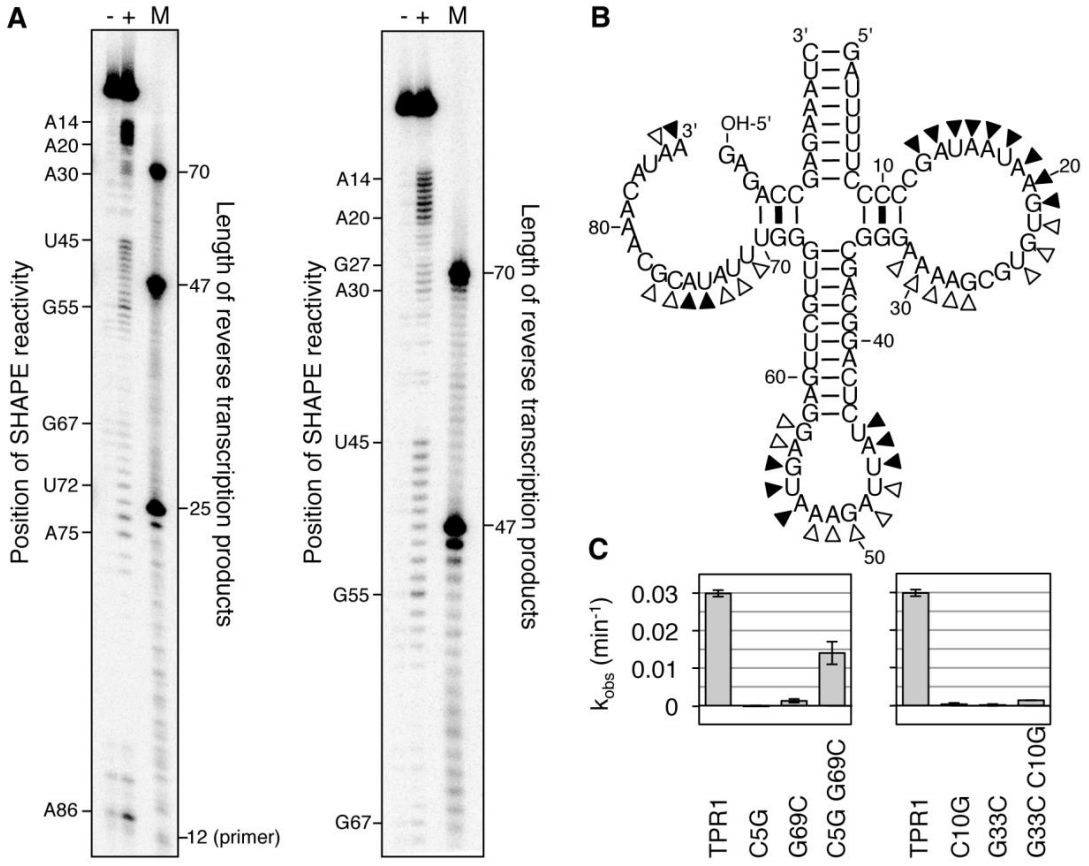

Figure 2.6: Secondary structure analysis for the trans-splicing TPR1 ribozyme. A) SHAPE analysis products, after reacting the ribozymes with 1M7 and reverse transcribing with a 5'-[^{32}P] radiolabeled primer. The left image shows a separation by 20% PAGE to separate short products. The right image shows a separation by 10% PAGE to show the longer products. Each image shows three lanes, where (-) denotes a negative control with DMSO, (+) denotes the SHAPE reaction with 1M7 dissolved in DMSO, and M denotes a marker lane with three 5'-radiolabeled DNAs that have the identical sequence as the expected reverse transcription products. The primer has a length of 12 nucleotides, base paired to 12 nucleotides that were added to the ribozyme 3'-terminus. The SHAPE signal is shifted by one nucleotide relative to the length of the reverse transcription product because the reverse transcriptase stops at the nucleotide before the SHAPE modification. B) The secondary structure was based on two types of analysis. Filled triangles indicate a strong signal from SHAPE analysis, while empty triangles indicate weaker signals. C) Two base pairs suggested by the SHAPE data were tested by single mutations of each base partner, and double mutation that should have restored activity for correct base pairs. The observed reaction rate was determined as in Fig. 5 and given for each mutant.

To test whether the two short helices formed, the predicted base pairs between C10 and G33 on the ribozyme and between C5 on the substrate and G69 on the ribozyme were chosen for analysis (Fig. 2.6C). At each base pair, three mutants were constructed and tested for activity. The single mutations removed the expected base pair and were expected to decrease activity. The double mutants were expected to restore the base pair and catalytic activity. As predicted, there was no detectable activity for any of the single mutations ribozyme variants tested. The C5G / G69C double mutation restored about 50% of the activity, but the C10G / G33C double mutation showed only slight activity. This confirmed the formation of the base pair between C5 and G69. It is possible that the C10 / G33 base pair formed but that the identity of the bases was also important for activity.

Dependence of k_{obs} on [trimetaphosphate], [Mg^{2+}], and pH.

To identify the reaction conditions at which the TPR1 ribozyme worked most efficiently, the dependence of the observed reaction rate on the concentration of trimetaphosphate, free magnesium ions, and pH were investigated (Fig. 2.7). TMP appeared to chelate an equimolar amount of Mg^{2+} (data not shown), therefore the excess of [Mg^{2+}] over [TMP] in our experiments was called the free [Mg^{2+}]. When the concentration of free magnesium ions was kept constant at 50 mM, an increase in the TMP concentration led to an increase in the observed rate to 0.03 min^{-1} at 100 mM [TMP], at pH 8.1. The [TMP] dependence was consistent with a 1:1 stoichiometry between ribozyme and TMP

(curve fit in Fig. 2.7A). Increases in the magnesium concentration at a constant TMP concentration of 100 mM led to a further increase in the reaction rate, with a k_{obs} of 0.12 min^{-1} at 400 mM of free [Mg^{2+}], at pH 8.1 (Fig. 7B). The pH dependence of the reaction was tested under these conditions (100 mM [TMP] and 400 mM free [Mg^{2+}]), as well as under the reaction conditions used during the selection (50 mM [TMP] and 50 mM free [Mg^{2+}]) (Fig. 2.7C). In these reactions three buffers were used, MES/NaOH, HEPES/NaOH, and Tris/HCl. For all reactions, the observed triphosphorylation rate increased linearly with the pH in the region between pH 5.5 and 8.5, with a slope of about one. This is consistent with a single deprotonation step being limiting for the reaction, perhaps the deprotonation of the RNA 5'-hydroxyl group in preparation for the nucleophilic attack on trimetaphosphate.

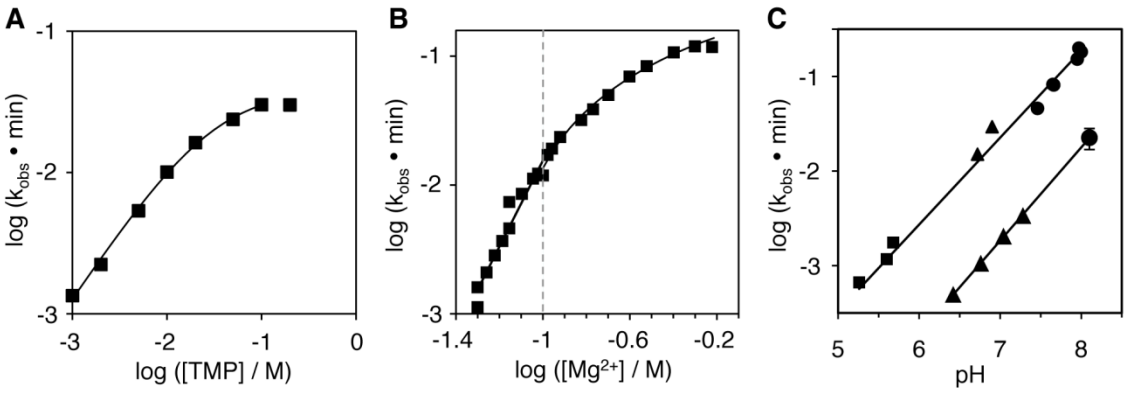

Figure 2.7: Dependence of reaction kinetics on reaction conditions, for the trans-splicing TPR1 ribozyme. A) Dependence of the observed reaction rate on the concentration of trimetaphosphate (TMP), with a free [Mg^{2+}] of 50 mM at pH 8.1. The maximal rate is reached at 100 mM [TMP]. The black line shows the fit of the Michaelis-Menten equation $k_{obs} = (k_{cat} * [TMP] / (K_M + [TMP])$ to data points from 1 to 100 mM [TMP], with $k_{cat} = 0.039$ min^{-1} and $K_M = 30$ mM. B) Dependence of the observed reaction rate on the concentration of free magnesium ions. Note that this is the excess of magnesium ions over the TMP concentration (here 100 mM), at pH 8.1. The dotted line separates reactions with [Mg^{2+}] / [TMP] < 1 (left) from reactions with [Mg^{2+}] / [TMP] > 1 (right). The fastest rate is obtained at 400 mM free [Mg^{2+}]. The half-maximal rate of 0.068 min^{-1} was observed at 150 mM free [Mg^{2+}]. C) Dependence of the observed reaction rate on the pH, at 100 mM [TMP] and 400 mM free [Mg^{2+}] (top line) and at 50 mM [TMP] and 50 mM free [Mg^{2+}] (bottom line). The slopes of the two lines were 0.92 (top line) and 0.99 (bottom line), respectively. The symbols denote the buffer systems MES/NaOH (squares), HEPES / NaOH (triangles), and Tris/HCl (circles). Note that the data point in the lower data set at pH 8.1 is the average of seven independent experiments, with error bars denoting their standard deviation.

2.4 Discussion

We established an in vitro selection strategy for ribozymes with a novel activity, the triphosphorylation of RNA 5'-hydroxyl groups. The procedure identified more than a dozen independent ribozymes. One ribozyme was characterized in more detail. This ribozyme was able to work in trans, and its product was confirmed by mass spectroscopy. Its secondary structure appears to

include a four-way helical junction. Its optimal reaction rate was at 100 mM trimetaphosphate, 400 mM free [Mg^{2+}] and increased pH.

The observed rate with the TPR1 ribozyme was 0.16 min^{-1} under optimal conditions (100 mM [TMP], 400 mM free [Mg^{2+}], pH 8.1) (Fig. 2.7). As comparison, the rate of the uncatalyzed reaction (k_{uncat}) in the absence of Mg^{2+} was estimated to be $3.4 \cdot 10^{-4}$ min^{-1}, at 500 mM [TMP] and pH 12 (Fig. 2 in (59)). Because the pH is limiting for the uncatalyzed reaction, its rate at pH 8.1 was assumed to be $10^{3.9}$ - fold slower, at $4.3 \cdot 10^{-8}$ min^{-1}. In addition, higher concentrations of trimetaphosphate lead to faster reactions for the uncatalyzed reaction (59). Because the influence of [TMP] on k_{uncat} may be nonlinear, we assumed a 3-fold lower k_{uncat} at 100 mM [TMP] compared to 500 mM [TMP]. This rate of $1.4 \cdot 10^{-8}$ min^{-1} is $1.1 \cdot 10^{7}$ -fold below the rate of 0.16 min^{-1} for the ribozyme catalyzed reaction. Mg^{2+} ions accelerate the uncatalyzed reaction (81) but no kinetic data are available for this Mg^{2+} acceleration. Therefore the value for k_{cat} / k_{uncat} of about 10^{7} describes the rate enhancement of the ribozyme-catalyzed reaction in the presence of magnesium relative to the rate of the uncatalyzed reaction without magnesium.

None of the six tested ribozymes showed triphosphorylation activity of free nucleosides (Fig. S2.5). It may be possible to overcome this in future experiments because many more ribozyme clones from our selection can now be tested for this activity. The fact that most of the isolated 36 clones appeared as single clones (Fig. S2.3) suggests that many more active sequences can be

isolated. However, RNA world organisms could have relied on the triphosphorylation of RNAs instead of free nucleosides to drive RNA polymerization: Activation groups at the nucleoside 5'-phosphate are only necessary for RNA polymerization in the 5'-3' direction. In contrast, RNA polymerization in the 3'-5' direction does not require nucleotide activation because chain elongation proceeds via the reaction of the activated RNA primer 5'-terminus with the nucleoside 3'-hydroxyl group. This principle has been shown with an evolved variant of the class I ligase ribozyme that uses NTPs for the extension of RNA primers in both the 3'-5' and 5'-3' direction (82). This suggests that instead of NTP, a nucleoside could be used to extend the growing primer 5'-terminus. In the second step, the new RNA 5'-terminus could be triphosphorylated. Therefore, RNA polymerization in an RNA world organism could have proceeded via the alternating triphosphorylation of the RNA primer, and the extension of the primer 5'-terminus with a templated nucleoside. This polymerization in the 3'-5' direction may have had an advantage for RNA world organisms: Nucleoside triphosphates are necessary for 5'-3' elongation, but their negative charge destabilizes the binding of the monomer to the template strand due to charge repulsion with the phosphodiester backbone of the template strand. This difference in affinity can be inferred from measurements of stacking interactions in nucleosides and NTPs (83). In contrast, RNA polymerization in the 3'-5' direction could use the uncharged nucleosides for the primer extension step, with a corresponding increase in monomer affinity to the template strand. Therefore, the alternating RNA 5'-triphosphorylation and extension with a

nucleoside could have been an efficient RNA polymerization strategy for RNA world organisms.

In addition to promoting RNA polymerization, RNA world organisms could have used RNA 5'-triphosphates for several other reactions. This could have included RNA-RNA ligation (16,69,84,85), the formation of phosphoamidate bonds (24), RNA capping (66,86), and amino acid activation (87). Together with the results of this study, this suggests that RNA world organisms could have used trimetaphosphate to thermodynamically drive a wide range of reactions.

2.5 Materials and Methods

In vitro selection

The DNA pool for the selection was generated from a 187-nucleotide long DNA oligomer ('ultramer', IDT) that included 150 positions with randomized sequence (phosphoramidites were hand-mixed) and primer binding sites on both termini. This 187-mer was annealed to a 79-mer, which attached the sequence of the hammerhead ribozyme and the promoter of T7 RNA polymerase to the 5'-terminus of the pool. Amplification with short primers generated double-stranded copies of the pool. The DNA pool was transcribed into the RNA pool under standard conditions, during which the 5'-terminal hammerhead ribozymes cleaved themselves from the pool, generating 5'-hydroxyl groups on the RNA pool. The sequence of the transcript was 5'-GGGCGGTCTCCTGACGAGCTAAGCGAAACTGCGGAAACGCAGTCGAGACC

GAGATGTT-N_{150}-CGCCAGTTAAGCTCCAGC-3', where the removed hammerhead ribozyme sequence is underlined. The RNA pool was purified by denaturing polyacrylamide gel electrophoresis to remove any uncleaved pool molecules.

The 5'-hydroxylated RNA pool was incubated with trimetaphosphate in a buffer containing 100 mM $MgCl_2$, 50 mM trisodium trimetaphosphate (freshly dissolved and sterile filtered), and 50 mM Tris/HCl pH 8.3. All triphosphorylation reactions during the selection were done at pH 8.3. When the solutions with TMP and $MgCl_2$ were combined the pH dropped to about 4.5. This pH was restored to the pH of the TMP solution alone (pH ~6) before the buffer (Tris/HCl pH 8.3) was added. After three hours or five minutes (depending on the selection round) of incubation at room temperature, RNAs were ethanol precipitated. The large salt pellet was extracted with a small volume of cold water to remove most of the salt. The remaining pellet was dissolved in water, desalted by size exclusion chromatography (P30 spin-columns; Bio-Rad), ethanol precipitated, and redissolved in water.

The recovered RNA pool molecules were heat renatured (2'/80°C) with a 1.25-fold molar excess of the R3C ligase ribozyme (76) (whose arms were designed complementary to the RNA pool 5'-terminus and the biotinylated capture oligonucleotide) and a 1.5-fold molar excess of biotinylated capture oligonucleotide (5'-biotin-d(GAACTGAAGTGTATG)rU-3'), in 100 mM KCl and 100 mM Tris/HCl pH 8.5. The solution was diluted to 400 nM pool RNA, 500 nM

ligase ribozyme, 600 nM capture oligonucleotide, and adjusted to 50 mM KCl, 25 mM $MgCl_2$, 2 mM spermidine, 20% (w/v) PEG 8000, and 50 mM Tris/HCl pH 8.5. After incubation for three hours at 30°C, magnesium was chelated by an excess of Na_2EDTA, and the mixture was heated (10' / 50°C) with a 10-fold excess of a DNA complementary to the ligase ribozyme to free the ligated RNAs. The biotinylated nucleic acids were captured during 30 minutes of agitation with streptavidin-coated magnetic beads (Promega) containing a 1.5-fold excess of biotin binding sites over biotinylated capture RNAs. Captured RNAs were washed first with 50 mM KCl, 20 mM HEPES/KOH pH 7.2, and 0.01% Triton X-100, then with 20 mM NaOH and 0.01% Triton X-100. RNA pool molecules were eluted from the magnetic beads by heating (3'/65°C) with 95% formamide / 1 mM Na_2EDTA. The effective complexity of the RNA pool was $1.7 \cdot 10^{14}$ sequences, based on an initial complexity of the double-stranded DNA library of $2.4 \cdot 10^{14}$ sequences, a total of 1.6 nmol of RNA pool molecules that entered the ligation step, losses of ~60% in the ligation step, losses of ~50% in the capture on streptavidin-coated beads, and losses of ~20% in further processing steps (data not shown). After ethanol precipitation, the RNAs were reverse transcribed using Superscript III reverse transcriptase (Invitrogen) and PCR amplified using Taq polymerase. Note that the 5'-PCR primer added a selective step because a part of its binding site was generated by the biotinylated capture oligonucleotide. The DNA pool for the next round of selection was regenerated by PCR amplification with the 79-nucleotide primer that was used to attach the hammerhead ribozyme

and the T7 RNA polymerase promoter. We used a previously published protocol for mutagenic PCR (77).

In pilot PCR amplifications of the reverse transcribed products, short PCR artifacts created the dominant product after about 30 PCR cycles. We assumed that this was because some pool molecules allowed the PCR primers to anneal within the randomized region, and shorter amplicons are known to outcompete longer amplicons during PCR (78). To avoid that problem we 'doped' the RNA pool molecules after the incubation with trimetaphosphate with 1/37,000 molar amount of pool 0 molecules that carried a 5'-triphosphate. Because these 5'-triphosphorylated molecules were ligated efficiently to the capture oligonucleotides, the doped pools required only 18-30 PCR amplification cycles, and suppressed short PCR artifacts. While this may have slowed the selection compared to a protocol without PCR problems, the 1/37,000-fold did not seem to harm the enrichment over multiple rounds of selection.

Triphosphorylation assays

For the self-triphosphorylation assay based on the R3C ligase ribozyme, cloned ribozymes were first transcribed in vitro (79) and purified by denaturing PAGE. Triphosphorylation reactions with 5 µM triphosphorylation ribozyme 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. The reaction mixture was diluted 10-fold with a buffer such that buffer concentrations were 500 nM ribozyme, 500 nM R3C ligase ribozyme, and 500 nM of an oligonucleotide (5'-

[^{32}P]-d(GAACTGAAGTGTATG)rU-3'), 100 mM KCl, 100 mM Tris/HCl pH 8.5, and 15 mM Na_2EDTA. After heat denaturation (2' / 65°C), the mixtures were diluted another 2-fold and adjusted to 20% (w/v) PEG 8000, 22 mM free Mg^{2+}, 50 mM KCl, 2.5 mM spermidine, and 100 mM Tris/HCl pH 8.5. This ligation reaction was incubated at 30°C for two hours, stopped by ethanol precipitation, and the products were separated by denaturing 10% PAGE. Gels were exposed to phosphorimager screens, scanned on a PMI phosphorimager (Bio-Rad), and signals were quantitated using the software quantity one.

The self-triphosphorylation assay based on the DNAzyme used triphosphorylation ribozymes that were internally [^{32}P] labeled during transcription, and purified by denaturing PAGE. Triphosphorylation reactions with 8 µM ribozyme were incubated as described above. At specific time points, 2 µL aliquots of the reactions were quenched by adding them to 8 µL of a solution containing 31.25 mM Na_2EDTA, 375 mM NaCl, 32 pmol of a DNAzyme that recognized the 14 nucleotides conserved at all triphosphorylation ribozyme 5'-termini, and 32 pmol of a DNA oligonucleotide that was complementary to the first 25 nucleotides of the "N_{150}" sequence to improve accessibility for the DNAzyme. After heat renaturation (2' / 80°C), the DNAzyme reaction was initiated by adding 10 µL of a solution containing 200 mM $MgCl_2$, 100 mM Tris/HCl pH 8.3, and incubated for one hour at 37°C. Two µL of these reactions were quenched with 8 µL of PAGE loading buffer containing 16 pmol of a DNA oligonucleotide complementary to the DNAzyme, 90% (v/v) formamide, and 37.5

mM Na_2EDTA. Products were separated by denaturing 22.5% PAGE and analyzed as described above.

The trans-triphosphorylation assay used substrate molecules that were internally [^{32}P] labeled during transcription. Two hammerhead ribozymes were contained 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 (79). Cleaved transcripts were purified by denaturing PAGE. Reaction mixtures with 5.5 µM triphosphorylation ribozyme and substoichiometric concentrations of radiolabeled substrate were incubated as described above. At specific time points, 1.5 µL of the reactions were added to 6.5 µL of formamide PAGE loading buffer containing 20 mM Na_2EDTA, heat renatured, and separated by denaturing 20% PAGE. Quantitation of the signals was as above. Note that for the titration of the TMP concentration, the chelation of Mg^{2+} with TMP generated a slight reduction in the pH for the reaction with 200 mM [TMP] such that the pH of 8.1 (all other reactions) was reduced to pH 7.6.

Generation of modified ribozyme constructs

Sequence modifications were introduced into ribozyme constructs by PCR mutagenesis or the Quickchange protocol for site directed mutagenesis (Stratagene). All DNA sequences were confirmed by cloning into pUC19 and sequencing. The only exceptions were ribozyme constructs where mutations near the 3'-terminus could be introduced during PCR amplification of the

templates for transcription. RNAs were generated by run-off transcription by T7 RNA polymerase as described (80) and purified by denaturing PAGE.

Mass spectrometric analysis

Substrate RNAs were prepared as described for the trans-triphosphorylation reaction, but without internal radiolabeling. A fraction of the substrate was incubated with an excess of the TPR1 ribozyme for three hours. The product of the triphosphorylation reaction as well as unreacted substrate RNA were purified by denaturing PAGE, and desalted on C_{18} Zip-tips (Millipore). About 20 pmol of each sample were dissolved in 4 µL containing 3-hydroxypicolinic acid and diammonium hydrogen citrate, and spotted on MALDI targets. The mass spectra were recorded by Dr. Yongxuan Su at the UCSD Molecular Mass Spectrometry Facility in negative ion mode on a Bruker Biflex IV, MALDI-TOFMS.

SHAPE analysis

For SHAPE analysis the TPR1 ribozyme was extended at its 3'-terminus by 12 nucleotides so that a 12-nucleotide reverse transcription primer could facilitate analysis of the complete ribozyme. The trans-reacting TPR1 ribozyme was incubated at a concentration of 1 µM with a 1.5-fold stoichiometric excess of its substrate, in 100 mM $MgCl_2$, 50 mM trisodium trimetaphosphate and 50 mM HEPES / KOH pH 8.0. To 9 µL of this solution, 1 µL of DMSO or a 20 mM solution of 1M7 in DMSO was added. After incubation for three minutes at room

temperature, the reaction products were ethanol precipitated. Reverse transcription with a 5'-[^{32}P] radiolabeled 12-nucleotide long reverse transcription primer was facilitated by Superscript III reverse transcriptase (Invitrogen). Reaction products were ethanol precipitated, redissolved in formamide PAGE loading buffer, heated (2' / 80°C), and separated by denaturing 10% and 20% PAGE. Signal quantitation was done as described for the triphosphorylation assays. Only primary reaction sites were observed because more than 80 % of the reverse transcription products had full length, and signals appeared with consistent kinetics in a concentration series of 1M7 (not shown). The signals were judged as strong, or weak when their intensity was at least 30%, or 8.5% of the strongest SHAPE signal intensity, respectively. Markers were synthetic DNA molecules with 5'-[^{32}P] label that had the exact sequence of expected reverse transcription products, and a length of 25, 47, and 70 nucleotides.

2.6 Acknowledgements

This work was funded by the National Institutes of Health (Molecular Biophysics Training Grant T32 GM #008326 to J.E.M.), the Hellman Family Foundation (awards 2011/12 and 2012/13) to U.F.M., and by the North American Space Agency (Astrobiology / Exobiology grant NNX13AJ09G to U.F.M).

We thank Chengguo Yao for pilot experiments with the hammerhead ribozyme, Manny Ares for the generous gift of 1M7 for SHAPE analysis, and Ram Krishnamurthy for helpful discussions.

Chapter 2, in full, has been submitted for publication, **Moretti, J.E.** and Muller, U.F. The dissertation author was the first author on this material.

2.7 Supporting Information

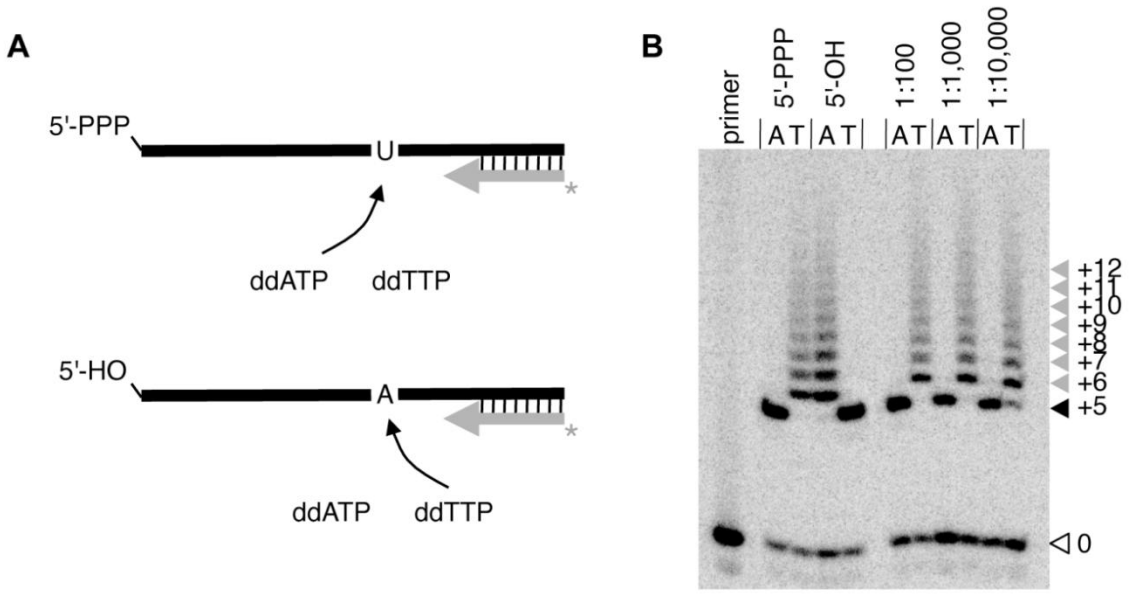

Figure S2.1: Estimation of the enrichment factor for one round of the selection. The assay was based on poisoned primer extension analysis of three mock-pools after enrichment. A) Schematic that shows 'active' (top) and 'inactive' RNA pool molecules. They are mutated with an A or a U five nucleotides upstream of the reverse transcription primer (grey arrow), which is 5'-radiolabeled (grey asterisk). During reverse transcription the primer extension on 'active' pool molecules is stopped at position +5 by ddATP but not by ddTTP. Conversely, the primer extension on the 'inactive' molecules is stopped at position +5 by ddTTP but not by ddATP. B) Autoradiogram of PAGE-separated products from the poisoned primer extension assay. The "A" or "T" indicates whether ddATP or ddTTP was used in the reaction. The length of the extension of the primer is indicated on the right by an empty triangle for unelongated primers, by a black triangle for primers that were stopped by ddNTPs at position +5, and by grey triangles for primer extension products that were not stopped at position +5. Note that the templating sequence for longer extension products is the randomized region of the pool, which causes an exponential distribution of primers that were stopped at positions higher than +5. The behavior of the 'active' RNA (5'-PPP) and the 'inactive' RNA (5'-OH) is shown in the left half of the image. The behavior of pools that were enriched from dilutions of active in inactive pools at ratios of 1:100, 1:1,000, and 1:10,000, is shown in the six lanes on the right. Note that the band at +5 in the right-most lane shows that ~10% of the primer extensions were stopped by the incorporation of ddTTP, thereby detecting the presence of 10% A at position +5. This enrichment from 1:10,000 to ~ 0.9 in one selection step suggested a lower estimate of 9,000-fold for the enrichment of active sequences in one round of selection.

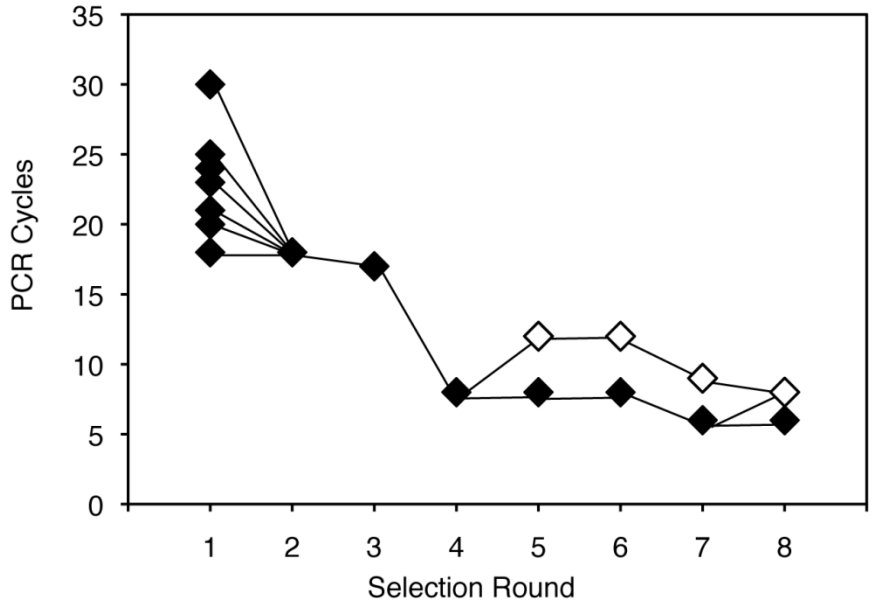

Figure S2.2: Progress of the in vitro selection. The graph shows the number of PCR cycles that was required in each round of the selection to obtain a band clearly visible by ethidium-bromide staining of agarose gels. The first round of selection was done in 11 parts, with PCR cycle numbers between 18 and 30. The drop in the PCR cycle number from >16 cycles (rounds one to three) to 8 cycles (round four) showed that a strongly increased amount of RNAs survived the selective step, suggesting that pool four contained a large fraction of active ribozymes. After round four, mutagenic PCR was employed to allow exploring the sequence neigborhood of selected clones, and the pool was split in two branches, one for low selection pressure (three hours incubation with trimetaphosphate; filled diamonds) and one for high selection pressure (five minutes incubation with trimetaphosphate; empty diamonds). After round seven, one additional branch was created, by subjecting a portion of the low-selection pressure branch to high selection pressure. Ten sequences were obtained from high-selection pressure and low selection pressure branches after five rounds and after eight rounds of selection. The name of the clone (e.g. R8_35C18) describes the round in which the clone was isolated (e.g. R8), the history of three-hour or five-minute incubations it experienced (e.g. _35), and the clone number (e.g. C18).

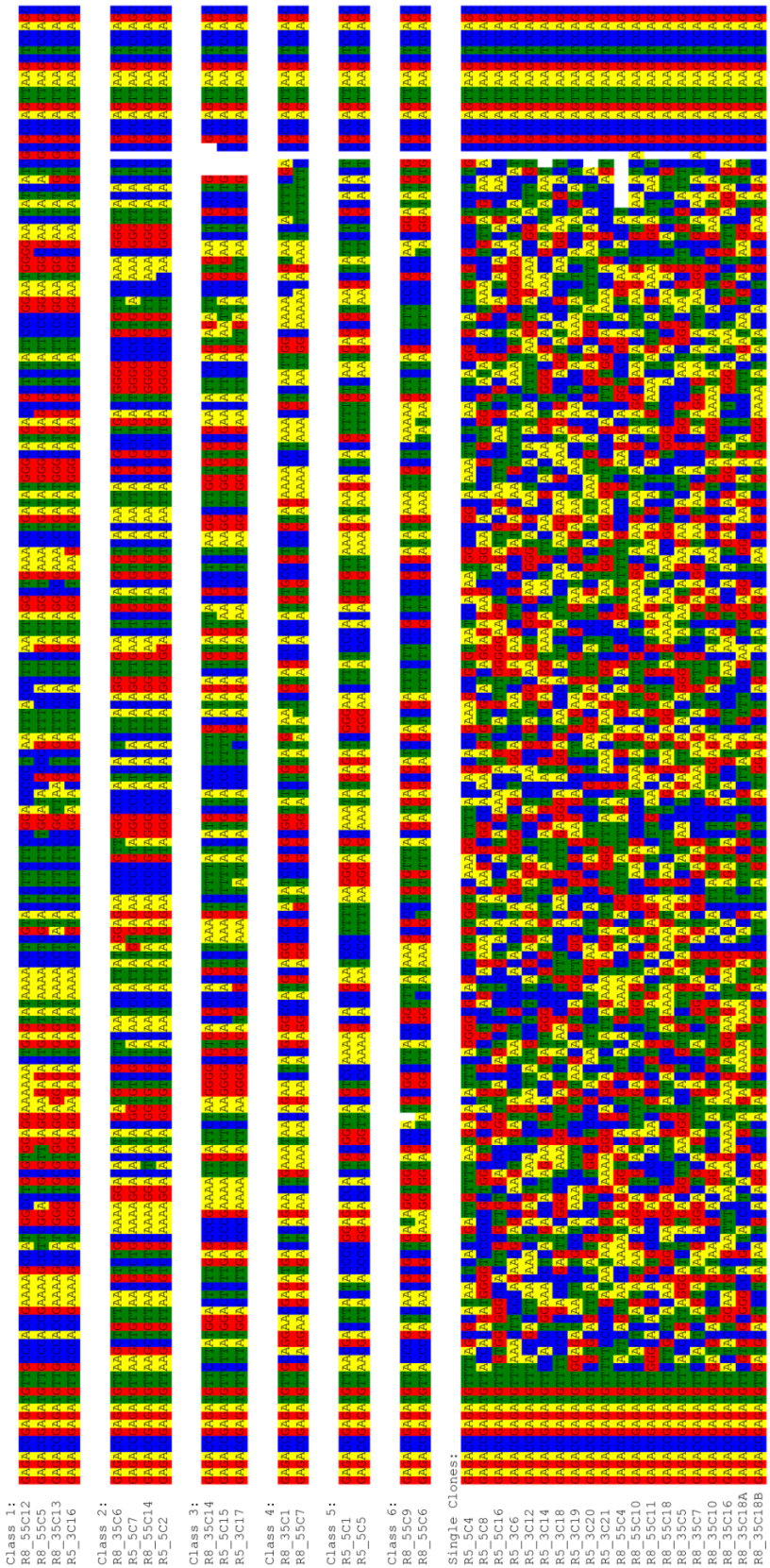

Figure S2.3: Sequences isolated after five or eight rounds of in vitro selection. Nucleotides are colored to facilitate comparison between related sequences and highlight the constant regions at the 5'- and 3'-ends. Six classes of sequences were isolated as multiple clones. Most clones were present as single clones. The eight most active clones are indicated with arrows. Note that the sequence of clone R5_5C1 is identical to the sequence of clone R5_5C5, therefore only one of the two clones was tested for activity.

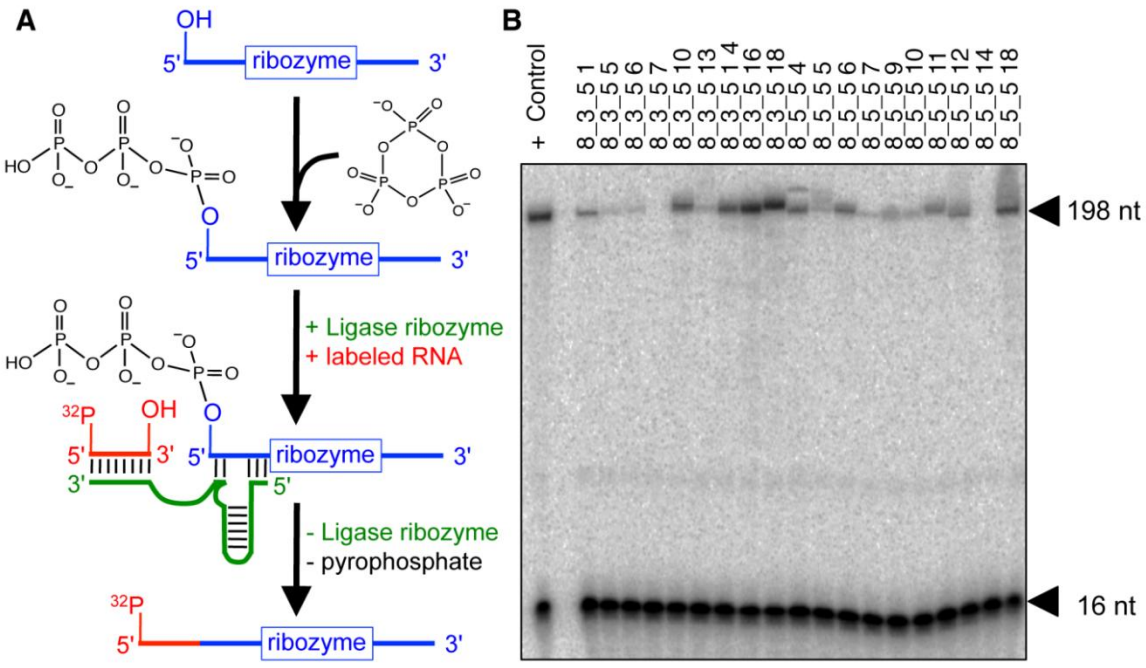

Figure S2.4: Assay for triphoshorylation activity using the R3C ligase ribozyme. A) Schematic of the assay. After the selected RNA clone (blue; 182 nucleotides) was incubated with trimetaphosphate, ligase ribozyme (green; 16 nucleotides) and 5'-[32P]-radiolabeled substrate RNA (red) were added, and incubated further. RNAs that contained a triphosphorylated 5'-terminus thereby generated a 198 nucleotide long product. B) Autoradiogram of products from the ligase ribozyme reaction, after separation by denaturing 10% polyacrylamide gel electrophoresis. The labeled substrate RNA (16 nt) is shifted if the selected RNA clone carried a 5'-triphosphate, thereby generating a large gel shift (198 nt).

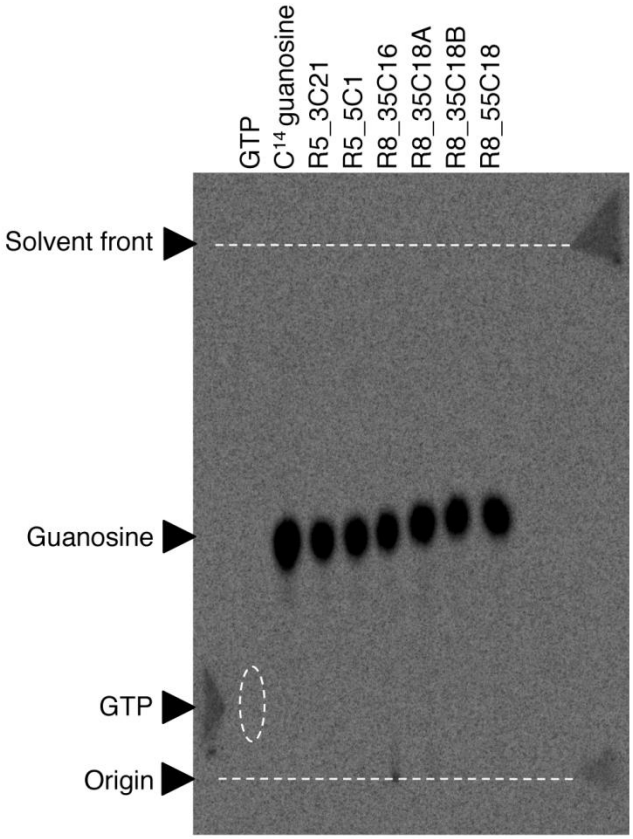

Figure S2.5: Assay testing for triphoshorylation of free nucleosides. The eight ribozymes that showed the highest activity in the self-triphosphorylation assay using the ligase ribozyme (Fig. 2.3) were tested. The 5'-terminal guanosine was removed from all ribozymes to allow for the insertion of a free guanosine. Ribozymes were incubated with C^{14} radiolabeled guanosine and trimetaphosphate. Standard reaction conditions were 8 µM ribozyme, 0.01 µCi C^{14} labeled guanosine, 100 mM $MgCl_2$, 50 mM trimetaphosphate, 50 mM Tris/HCl pH 8.3, and 16 hours incubation at room temperature. Variations used different reaction temperature (40°C, -20°C), and higher pH (9.5). Two µL aliquots of the reactions were spotted on pre-wetted PEI cellulose, together with GTP as control. Plates were developed vertically with 1 M KH_2PO_4 (pH 3.5) and dried. GTP was visualized by UV shadowing. Radioactive signals were detected by phosphoimaging. Triangles that are weakly visible in the autoradiogram were used to mark the position of the origin, the solvent front, and the GTP. Only six of the eight tested ribozymes are shown here; the remaining two ribozymes showed the same negative result.

2.8 References

1. Woese, C.R. (1967) The genetic code. The molecular basis for genetic expression. *Harper & Row, New York*.

2. Crick, F.H.C. (1968) The origin of the genetic code. *J. Mol. Biol.* 38, 367-379.

3. Orgel, L.E. (1968) Evolution of the genetic apparatus. *J. Mol. Biol. 38*, 381-393.

4. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell 31*, 147-157.

5. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell 35*, 849-857.

6. Hutchins, C.J., Rathjen, P.D., Forster, A.C. and Symons, R.H. (1986) Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid. *Nucleic Acids Res 14*, 3627-3640.

7. Buzayan, J.M., Gerlach, W.L. and Bruening, G. (1986) Satellite tobacco ringspot virus RNA: A subset of the RNA sequence is sufficient for autolytic processing. *Proc Natl Acad Sci U S A 83*, 8859-8862.

8. van der Veen, R., Arnberg, A.C., van der Horst, G., Bonen, L., Tabak, H.F. and Grivell, L.A. (1986) Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro. *Cell 44*, 225-234.

9. Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S. and Sharp, P.A. (1986) Splicing of messenger RNA precursors. *Annu Rev Biochem 55*, 1119-1150.

10. Sharmeen, L., Kuo, M.Y., Dinter-Gottlieb, G. and Taylor, J. (1988) Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. *J Virol 62*, 2674-2679.

11. Saville, B.J. and Collins, R.A. (1990) A site-specific self-cleavage reaction performed by a novel RNA in Neurospora mitochondria. *Cell 61*, 685-696.

12. Noller, H.F., Hoffarth, V. and Zimniak, L. (1992) Unusual resistance of peptidyl transferase to protein extraction procedures. *Science 256*, 1416-1419.

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

14. Roth, A., Weinberg, Z., Chen, A.G.Y., Kim, P.B., Amer, T.D.A. and Breaker, R.R. (2013) A novel class of self-cleaving ribozymes is prevalent in many species of bacteria and eukarya. *Nature Chemical Biology*, accepted.

15. Steitz, T.A. and Moore, P.B. (2003) RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem Sci 28*, 411-418.

16. Bartel, D.P. and Szostak, J.W. (1993) Isolation of new ribozymes from a large pool of random sequences. *Science 261*, 1411-1418.

17. Lorsch, J.R. and Szostak, J.W. (1994) In vitro evolution of new ribozymes with polynucleotide kinase activity. *Nature 371*, 31-36.

18. Illangasekare, M., Sanchez, G., Nickles, T. and Yarus, M. (1995) Aminoacyl-RNA synthesis catalyzed by an RNA. *Science 267*, 643-647.

19. Wecker, M., Smith, D. and Gold, L. (1996) In vitro selection of a novel catalytic RNA: characterization of a sulfur alkylation reaction and interaction with a small peptide. *RNA 2*, 982-994.

20. Zhang, B. and Cech, T.R. (1997) Peptide bond formation by in vitro selected ribozymes. *Nature 390*, 96-100.

21. Tarasow, T.M., Tarasow, S.L. and Eaton, B.E. (1997) RNA-catalysed carbon-carbon bond formation. *Nature 389*, 54-57.

22. Unrau, P.J. and Bartel, D.P. (1998) RNA-catalysed nucleotide synthesis. *Nature 395*, 260-263.

23. Sengle, G., Eisenfuhr, A., Arora, P.S., Nowick, J.S. and Famulok, M. (2001) Novel RNA catalysts for the Michael reaction. *Chem Biol 8*, 459-473.

24. Baskerville, S. and Bartel, D.P. (2002) A ribozyme that ligates RNA to protein. *Proc Natl Acad Sci U S A 99*, 9154-9159.

25. Tsukiji, S., Pattnaik, S.B. and Suga, H. (2003) An alcohol dehydrogenase ribozyme. *Nat Struct Biol 10*, 713-717.

26. Wacey, D., Kilburn, M.R., Saunders, M., Cliff, J. and Brasier, M.D. (2011) Microfossils of sulphur-metabolizing cells in 3.4-billion-year-old rocks of Western Australia. *Nature geoscience 4*, 698-702.

27. Von Kiedrowski, G. (1986) A self-replicating hexadeoxynucleotide. *Angew. Chem. Int. Ed. Engl. 25*, 932-935.

28. Orgel, L.E. (2004) Prebiotic chemistry and the origin of the RNA world. *Crit Rev Biochem Mol Biol 39*, 99-123.

29. Hargreaves, W.R. and Deamer, D.W. (1978) Liposomes from ionic, single-chain amphiphiles. *Biochemistry 17*, 3759-3768.

30. Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E. and Bartel, D.P. (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science 292*, 1319-1325.

31. Szostak, J.W. (2011) An optimal degree of physical and chemical heterogeneity for the origin of life? *Philos Trans R Soc Lond B Biol Sci 366*, 2894-2901.

32. Sczepanski, J.T. and Joyce, G.F. (2012) Synthetic evolving systems that implement a user-specified genetic code of arbitrary design. *Chem Biol 19*, 1324-1332.

33. Monnard, P.A., Apel, C.L., Kanavarioti, A. and Deamer, D.W. (2002) Influence of ionic inorganic solutes on self-assembly and polymerization processes related to early forms of life: implications for a prebiotic aqueous medium. *Astrobiology 2*, 139-152.

34. Monnard, P.A. and Deamer, D.W. (2003) Preparation of vesicles from nonphospholipid amphiphiles. *Methods Enzymol 372*, 133-151.

35. Hanczyc, M.M., Fujikawa, S.M. and Szostak, J.W. (2003) Experimental models of primitive cellular compartments: encapsulation, growth, and division. *Science 302*, 618-622.

36. Chen, I.A., Roberts, R.W. and Szostak, J.W. (2004) The emergence of competition between model protocells. *Science 305*, 1474-1476.

37. Sacerdote, M.G. and Szostak, J.W. (2005) Semipermeable lipid bilayers exhibit diastereoselectivity favoring ribose. *Proc Natl Acad Sci U S A 102*, 6004-6008.

38. Lincoln, T.A. and Joyce, G.F. (2009) Self-sustained replication of an RNA enzyme. *Science 323*, 1229-1232.

39. Hayden, E.J. and Lehman, N. (2006) Self-assembly of a group I intron from inactive oligonucleotide fragments. *Chem Biol 13*, 909-918.

40. Vaidya, N., Manapat, M.L., Chen, I.A., Xulvi-Brunet, R., Hayden, E.J. and Lehman, N. (2012) Spontaneous network formation among cooperative RNA replicators. *Nature 491*, 72-77.

41. Robertson, M.P. and Joyce, G.F. (2012) The origins of the RNA world. *Cold Spring Harb Perspect Biol 4*.

42. Sanchez, R.A. and Orgel, L.E. (1970) Studies in prebiotic synthesis. V. Synthesis and photoanomerization of pyrimidine nucleosides. *J Mol Biol 47*, 531-543.

43. Powner, M.W., Gerland, B. and Sutherland, J.D. (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature 459*, 239-242.

44. Powner, M.W., Sutherland, J.D. and Szostak, J.W. (2010) Chemoselective multicomponent one-pot assembly of purine precursors in water. *J Am Chem Soc 132*, 16677-16688.

45. Joyce, G.F., Schwartz, A.W., Miller, S.L. and Orgel, L.E. (1987) The case for an ancestral genetic system involving simple analogues of the nucleotides. *Proc Natl Acad Sci U S A 84*, 4398-4402.

46. Krishnamurthy, R. (2012) Role of pK(a) of nucleobases in the origins of chemical evolution. *Acc Chem Res 45*, 2035-2044.

47. Hud, N.V., Cafferty, B.J., Krishnamurthy, R. and Williams, L.D. (2013) The origin of RNA and "my grandfather's axe". *Chem Biol 20*, 466-474.

48. Wochner, A., Attwater, J., Coulson, A. and Holliger, P. (2011) Ribozyme-catalyzed transcription of an active ribozyme. *Science 332*, 209-212.

49. Verlander, M.S., Lohrmann, R. and Orgel, L.E. (1973) Catalysts for the self-polymerization of adenosine cyclic 2', 3'-phosphate. *J Mol Evol 2*, 303-316.

50. Rajamani, S., Vlassov, A., Benner, S., Coombs, A., Olasagasti, F. and Deamer, D. (2008) Lipid-assisted synthesis of RNA-like polymers from mononucleotides. *Orig Life Evol Biosph 38*, 57-74.

51. Prabahar, K.J., Cole, T.D. and Ferris, J.P. (1994) Effect of phosphate activating group on oligonucleotide formation on montmorillonite: the regioselective formation of 3',5'-linked oligoadenylates. *J Am Chem Soc 116*, 10914-10920.

52. Prabahar, K.J. and Ferris, J.P. (1997) Adenine derivatives as phosphate-activating groups for the regioselective formation of 3',5'-linked oligoadenylates on montmorillonite: possible phosphate-activating groups for the prebiotic synthesis of RNA. *J Am Chem Soc 119*, 4330-4337.

53. Lohrmann, R. and Orgel, L.E. (1968) Prebiotic synthesis: phosphorylation in aqueous solution. *Science 161*, 64-66.

54. Lohrmann, R. (1977) Formation of nucleoside 5'-phosphoramidates under potentially prebiological conditions. *J Mol Evol 10*, 137-154.

55. Sawai, H. and Orgel, L.E. (1975) Letter: Oligonucleotide synthesis catalyzed by the Zn-2+ ion. *J Am Chem Soc 97*, 3532-3533.

56. Inoue, T. and Orgel, L.E. (1982) Oligomerization of (guanosine 5'-phosphor)-2-methylimidazolide on poly(C). An RNA polymerase model. *J Mol Biol 162*, 201-217.

57. Chen, I.A., Salehi-Ashtiani, K. and Szostak, J.W. (2005) RNA catalysis in model protocell vesicles. *J Am Chem Soc 127*, 13213-13219.

58. Budin, I., Debnath, A. and Szostak, J.W. (2012) Concentration-driven growth of model protocell membranes. *J Am Chem Soc 134*, 20812-20819.

59. Etaix, E. and Orgel, L.E. (1978) Phosphorylation of nucleosides in aqueous solution using trimetaphosphate: Formation of nucleoside triphosphates. *J. Carbohydrates Nucleosides Nucleotides 5*, 91-110.

60. Yamagata, Y., Watanabe, H., Saitoh, M. and Namba, T. (1991) Volcanic production of polyphosphates and its relevance to prebiotic evolution. *Nature 352*, 516-519.

61. Pasek, M.A., Kee, T.P., Bryant, D.E., Pavlov, A.A. and Lunine, J.I. (2008) Production of potentially prebiotic condensed phosphates by phosphorus redox chemistry. *Angew Chem Int Ed Engl 47*, 7918-7920.

62. Osterberg, R. and Orgel, L.E. (1972) Polyphosphate and trimetaphosphate formation under potentially prebiotic conditions. *J Mol Evol 1*, 241-248.

63. Thilo, E., Schulz, G. and Wichmann, E.-M. (1953) Die Konstitution des Grahamschen und Kurrolschen Salzes. *Zeitschrift für anorganische und allgemeine Chemie 272*, 182-200.

64. Feldmann, W. (1967) Das Trimetaphosphat als Triphosphorylierungsmittel für Alkohole und Kohlenhydrate in wäßriger Lösung. Seine Sonderstellung unter den kondensierten Phosphaten. *Chem. Ber. 100*, 3850-3860.

65. Wiegand, T.W., Janssen, R.C. and Eaton, B.E. (1997) Selection of RNA amide synthases. *Chem Biol 4*, 675-683.

66. Huang, F. and Yarus, M. (1997) 5'-RNA self-capping from guanosine diphosphate. *Biochemistry 36*, 6557-6563.

67. Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science 249*, 505-510.

68. Ellington, A.D. and Szostak, J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature 346*, 818-822.

69. Uhlenbeck, O.C. (1987) A small catalytic oligoribonucleotide. *Nature 328*, 596-600.

70. Santoro, S.W. and Joyce, G.F. (1998) Mechanism and utility of an RNA-cleaving DNA enzyme. *Biochemistry 37*, 13330-13342.

71. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res 31*, 3406-3415.

72. Mortimer, S.A. and Weeks, K.M. (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J Am Chem Soc 129*, 4144-4145.

73. Hajdin, C.E., Bellaousov, S., Huggins, W., Leonard, C.W., Mathews, D.H. and Weeks, K.M. (2013) Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc Natl Acad Sci U S A 110*, 5498-5503.

74. Sashital, D.G., Cornilescu, G., McManus, C.J., Brow, D.A. and Butcher, S.E. (2004) U2-U6 RNA folding reveals a group II intron-like domain and a four-helix junction. *Nat Struct Mol Biol 11*, 1237-1242.

75. Berry, K.E., Waghray, S., Mortimer, S.A., Bai, Y. and Doudna, J.A. (2011) Crystal structure of the HCV IRES central domain reveals strategy for start-codon positioning. *Structure 19*, 1456-1466.

76. Rogers, J. and Joyce, G.F. (2001) The effect of cytidine on the structure and function of an RNA ligase ribozyme. *RNA 7*, 395-404.

77. Cadwell, R.C. and Joyce, G.F. (1992) Randomization of genes by PCR mutagenesis. *PCR Methods Appl 2*, 28-33.

78. Bustin, S.A. (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol 25*, 169-193.

79. Price, S.R., Ito, N., Oubridge, C., Avis, J.M. and Nagai, K. (1995) Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J Mol Biol 249*, 398-408.

80. Milligan, J.F., Groebe, D.R., Witherell, G.W. and Uhlenbeck, O.C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res 15*, 8783-8798.

81. Cheng, C., Fan, C., Wan, R., Tong, C., Miao, Z., Chen, J. and Zhao, Y. (2002) Phosphorylation of adenosine with trimetaphosphate under simulated prebiotic conditions. *Orig Life Evol Biosph 32*, 219-224.

82. McGinness, K.E., Wright, M.C. and Joyce, G.F. (2002) Continuous in vitro evolution of a ribozyme that catalyzes three successive nucleotidyl addition reactions. *Chem Biol 9*, 585-596.

83. Sigel, H. and Griesser, R. (2005) Nucleoside 5'-triphosphates: self-association, acid-base, and metal ion-binding properties in solution. *Chem Soc Rev 34*, 875-900.

84. Ikawa, Y., Tsuda, K., Matsumura, S. and Inoue, T. (2004) De novo synthesis and development of an RNA enzyme. *Proc Natl Acad Sci U S A 101*, 13750-13755.

85. Robertson, M.P. and Ellington, A.D. (1999) In vitro selection of an allosteric ribozyme that transduces analytes to amplicons. *Nat Biotechnol 17*, 62-66.

86. Zaher, H.S., Watkins, R.A. and Unrau, P.J. (2006) Two independently selected capping ribozymes share similar substrate requirements. *RNA 12*, 1949-1958.

87. Kumar, R.K. and Yarus, M. (2001) RNA-catalyzed amino acid activation. *Biochemistry 40*, 6998-7004.

Chapter 3

Arginine cofactors on the polymerase ribozyme

3.1 Abstract

The RNA world hypothesis states that the early evolution of life went through a stage in which RNA served both as genome and as catalyst. The central catalyst in an RNA world organism would have been a ribozyme that catalyzed RNA polymerization to facilitate self-replication. An RNA polymerase ribozyme was developed previously in the lab but it is not efficient enough for self-replication. The factor that limits its polymerization efficiency is its weak sequence-independent binding of the primer/template substrate. Here we tested whether RNA polymerization could be improved by a cationic arginine cofactor, to improve the interaction with the substrate. In an RNA world, amino acid-nucleic acid conjugates could have facilitated the emergence of the translation apparatus and the transition to an RNP world. We chose the amino acid arginine for our study because this is the amino acid most adept to interact with RNA. An arginine cofactor was positioned at ten different sites on the ribozyme, using conjugates of arginine with short DNA or RNA oligonucleotides. However, polymerization efficiency was not increased in any of the ten positions. In five of the ten positions the arginine reduced or modulated polymerization efficiency, which gives insight into the substrate-binding site on the ribozyme. These results

suggest that the existing polymerase ribozyme is not well suited to using an arginine cofactor.

3.2 Introduction

According to the RNA world hypothesis, an early stage of life used RNA both as genome and as catalyst (1-4) (for recent reviews see 5,6). The central activity in an RNA world organism would have been RNA polymerization to facilitate self-replication. To recapitulate an RNA world in the lab, RNA polymerase ribozymes were developed and improved in several laboratories (7-10).

These polymerase ribozymes have a length in the range of 200 nucleotides. Therefore, self-replication would require the polymerization of about 200 ribozyme-encoding nucleotides. However, the best existing polymerase ribozymes favor variants of a single, short template sequence with the length of less than 20 nucleotides. By concatenating multiple copies of this sequence it was possible to extend a primer by 95 nucleotides (10). However, such a template could not encode a ribozyme. On unrelated template sequences, polymerization reaches usually less than 10 nucleotides, and recent improvements made it possible to polymerize 20-30 nucleotides (9-12). However, this is still far below the level required for self-replication. The limiting factor for polymerization efficiency is the ribozyme's weak sequence-independent binding of the primer/template substrate, with a K_M in the millimolar range (13). Some of the sequence independent contacts are hydrogen bonds to template 2'-hydroxyl

groups (14). However, it may be possible to establish additional sequence independent contacts mediated by ionic interactions with the negatively charged phosphodiester groups of the primer/template substrate. To do this, the ribozyme would have to employ a positively charged cofactor.

This positive charge can be supplied by metal ions or by cationic organic molecules. The polymerase ribozyme was originally selected in the presence of 60 mM magnesium ions (7) and different versions were optimized in the presence of 36 to 184 mM free magnesium ions (9, 10). Because magnesium ions are good ligands for the phosphodiester oxygen anions of RNA (15) the continuous presence of magnesium ions during the evolutionary history of polymerase ribozymes should have found the most beneficial involvements of magnesium ions that increase polymerization efficiency. However, even at the optimal magnesium concentration of 200 mM the binding of substrate is in the millimolar range, suggesting that cations different from metal cations could play a role to improve substrate binding.

In contrast to metal ions the polymerase ribozyme did not encounter cationic organic molecules during its history. Therefore, a potential benefit from those molecules would have gone undiscovered. Specifically, the amino acid arginine carries several advantages over other cationic cofactors. Most importantly, the guanidinium group does not establish a hydration shell in aqueous solution. This helps the binding of negatively charged RNAs because it avoids the enthalpic cost of displacing a hydration shell (16). Additionally, the guanidinium group of arginine has a pK_A of 12.5 (17), maintaining a positive

charge at any pH value encountered by the ribozyme. Evidence that these factors benefit RNA binding comes from RNA binding proteins, which use arginine more than any other amino acid at the interface with RNA (18).

How could arginine cofactors compete with the high concentration of magnesium ions that are required by the polymerase ribozyme? In addition to the absence of a hydration shell our experiments carry two designs to help arginine compete with the magnesium ions. First, we decreased the free magnesium ion concentration from 184 mM to 64 mM, which facilitates near-optimal activity, and further down to 24 mM, which allows weak but quantifiable polymerization to occur (19). Second, we connected the arginine cofactor to the ribozyme via arginine-nucleic acid conjugates, which base pair to the ribozyme and thereby generate a high local concentration of arginine proximal to the binding site. We estimate that the local concentration of the arginine guanidinium group would be at least 50 mM, based on the volume accessible constrained by the length of the linker to the nucleic acid.

In an RNA world, amino acid - nucleid acid conjugates or peptide - nucleic acid conjugates could have served in the roles of cofactors and could have established the first steps in a translation system (20, 21); see also (22). The synthesis of such conjugates would have been possible in an RNA world because ribozymes can generate several different types of RNA-amino acid conjugates (23-26). One benefit of amino acid - nucleic acid conjugates for an RNA world would have been that less sequence of the ribozyme needs to evolve for pairing a conjugate compared to establishing a binding pocket for the

cofactor. This means that the 'combinatorial cost' of acquiring a cofactor is strongly reduced, and thereby the evolutionary likelihood of reaching that state is higher.

In this study, arginine was used as a positively charged cofactor for the polymerase ribozyme. An arginine - nucleic acid conjugate was positioned at ten different positions on the ribozyme located near the substrate-binding site. We tested whether the positively charged arginine could be used by the ribozyme to increase polymerization efficiency. However, the arginine did not improve polymerization in any of these ten positions, suggesting that single arginines are not sufficient to improve the existing polymerase ribozyme. This also suggests that it may be harder than previously thought to take the first step in the development of the translation apparatus, via amino acid - nucleic acid conjugates.

3.3 Results

We used arginine-RNA and arginine-DNA conjugates to position the arginine cofactor on the ribozyme. Specific sequences for the nucleic acid handle of the conjugate made it possible to base pair the conjugate to different positions on the ribozyme (Fig. 3.1). This strategy carries several advantages over the use of free amino acids or free peptides. First, a few unpaired bases on the ribozyme are sufficient to base pair to the handle of the conjugate. In comparison, free amino acids or peptides would make it necessary to establish a binding pocket for the cofactors on the ribozyme. Second, the amino acid portion of the

conjugate is accessible for interactions with the substrate. In contrast, free amino acids and short peptides require a cofactor-binding pocket that obstructs at least some of the possible interactions with the primer/template.

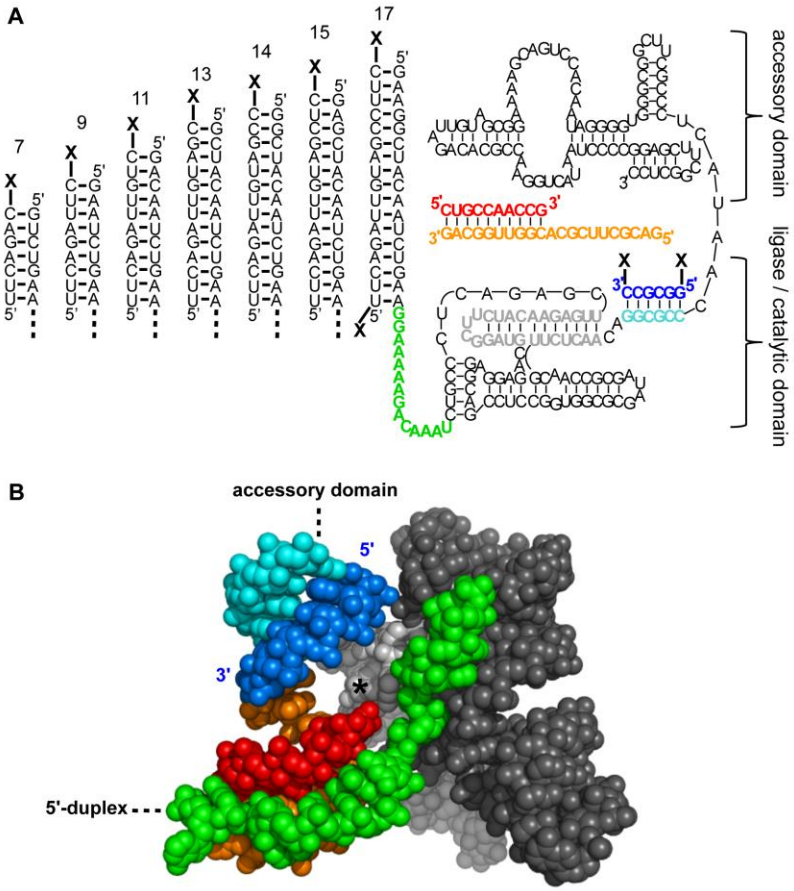

Figure 3.1: Structure of ribozyme constructs used in this study. The 5'-terminus of the ribozyme (green) is in close contact with the primer (red) and template (orange). The P2 oligo (dark blue) is base paired to a complementary region on the ribozyme (light blue), forming the P2 helix. A) Secondary structure of the polymerase ribozyme (7) with the 5'-duplexes and the P2 duplex that were used to attach arginine or amino cofactors. The length of the 5'-duplex is indicated. "X" denotes the position of the chemical modification. The P2 oligo is truncated, and the internal mismatch was removed (12). B) 3D structure of the ligase domain in the polymerase ribozyme, based on the crystal structure of the ligase (27). Atoms that do not appear in the polymerase ribozyme were deleted. The asterisk denotes the position of the catalytic site. The positions where the 5'-duplex and the accessory domain are attached to the ligase domain are indicated, as well as the 5'-terminus and 3'-terminus of the P2 oligo.

The arginine-nucleic acid conjugates were synthesized by carbodiimide peptide coupling chemistry. Fmoc-protected arginine was activated as NHS ester and reacted with amino modified DNA or RNA. The nucleic acid sequences of these conjugates were chosen to pair to one of two target sites on the polymerase ribozyme, thereby forming a 5'-terminal duplex, or the P2 duplex. The choice of these target sites was based on their vicinity to the catalytic site (Fig. 3.1;(27)) and because base pairing to these sequences did not inhibit ribozyme polymerization.

Ten polymerase ribozymes have been developed to date (7-10). Our study focuses on the first published polymerase ribozyme (7) because this was the most efficient polymerase ribozyme at the beginning of our study. Our results are relevant for at least the three most efficient variants of these ribozymes because their secondary structure is almost identical (7, 9, 10).

A 5'-duplex on the polymerase ribozyme to attach arginine conjugates

The first site for attaching the arginine conjugates is a duplex that extends the 5'-terminus of the polymerase ribozyme (19). Choosing this 5'-duplex for attaching the arginine has the benefit that the length of the duplex can be varied, thereby tethering an arginine to the distal end of the duplex places it at different positions along a 'helical ruler' on the ribozyme (28). Additionally, the proximal end of the duplex accesses another position. Therefore, the 5'-duplex allowed us to place an arginine cofactor at eight positions on the polymerase ribozyme (Fig.

3.1). In the absence of a complementary RNA or DNA, all single-stranded 5'-terminal ribozyme sequences inhibited polymerization (data not shown), confirming that the conjugates annealed to their intended position at the ribozyme 5'-terminus.

To measure the effect of each positioned arginine on ribozyme function we quantified the polymerization efficiency with and without the arginine modification. Additionally, we measured the influence of an amino group, which was used to couple the arginine with the nucleic acid handle. The polymerization efficiency was measured as the average number of nucleotides added to each primer molecule. This readout is sensitive enough to allow the detection of single hydrogen bonds between the ribozyme and the primer/template substrate (14). However, the arginine modification did not show any effect on ribozyme polymerization when it was placed at the proximal end of the 5'-duplex or at the distal end of the 5'-duplex, with duplex lengths of 9, 14, and 17 base pairs (Fig. 3.2 and 3.3).

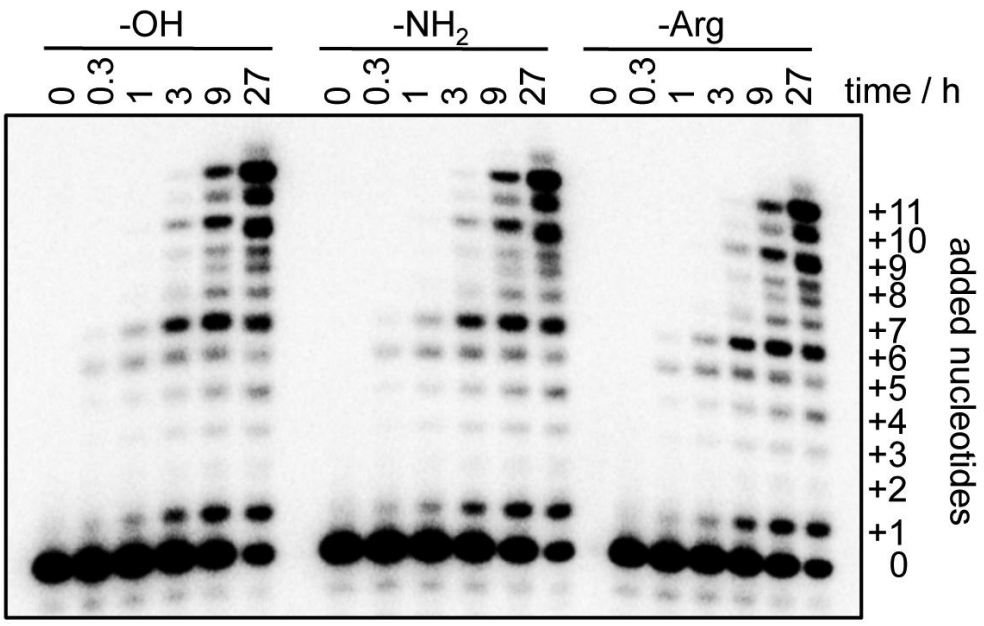

Figure 3.2: Influence of arginine and amino modifications at the proximal end of the 5'-duplex, on polymerization. Shown is an autoradiogram of PAGE separated polymerization products. For each sample, the polymerization products at six incubation times are shown. The number of nucleotides added to the primer during polymerization is indicated. The eighth nucleotide addition results in two bands due to nucleotide misincorporation. The length of the 5'-duplex was 17 base pairs. No difference in polymerization efficiency was found between unmodified and modified ribozymes, within the errors of three replications of the experiments.

When the 5'-duplex had a length of 7 base pairs, the DNA derivatives showed lower polymerization efficiency than the RNA derivatives (Fig. 3.3). This can be explained by the lower thermodynamic stability of DNA/RNA duplexes relative to RNA/RNA duplexes (29). We assume that the 7-base pair DNA/RNA duplex was partially dissociated so that the single-stranded 5'-sequence of the ribozyme could inhibit polymerization. This interpretation is supported by our observations that the optimal reaction temperature with the 7-base pair

DNA/RNA duplex was slightly lower than with longer duplexes, and that the single-stranded 5'-sequence inhibits polymerization (data not shown). All 5'-duplexes longer than 7 base pairs appeared to be stable under the used reaction conditions.

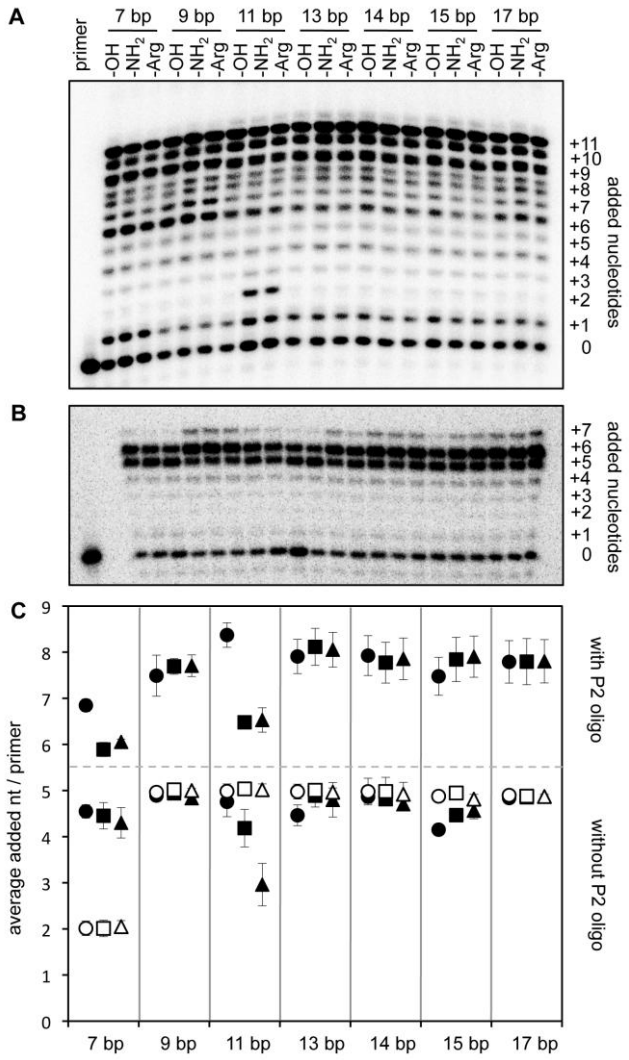

Figure 3.3: Influence of arginine and amino modifications at the distal end of the 5'-duplex on polymerization. A) PAGE separated polymerization products, with RNA/RNA duplexes at the 5'-terminus of the ribozyme, in the presence of our P2 oligo. The length of the 5'-duplexes and the chemical modification, are indicated. B) PAGE separated polymerization products, with RNA/RNA duplexes at the 5'-terminus of the ribozyme, without P2 oligo. The length of the 5'-duplexes and the chemical modification, are indicated. C) Quantitation of polymerization efficiencies for RNA/RNA (filled symbols) and DNA/RNA (open symbols) duplexes at the ribozyme 5'-terminus. The polymerization efficiency is described as the average number of nucleotides added per primer. For each length of the 5'-duplex, three variants were tested: Unmodified (circles), amino modified (squares), and arginine modified (triangles) duplexes. Symbols above the grey dashed line show the results of reactions in the presence of the P2 oligo; symbols below the grey dashed line show the results in the absence of a P2 oligo. Errors are standard deviations from three experiments.

Inhibitory effects of arginine at the 5'-duplex

When the length of the 5'-duplex was 11 base pairs the arginine and amino modification showed an inhibitory effect on polymerization (Fig. 3.3C). While this effect occurred both in the absence and the presence of a P2 oligo, an inhibitory effect at the 7 base pair duplex appeared only in the presence of the P2 oligo. The P2 oligo was introduced as a heptanucleotide that complements the P2 duplex on the polymerase ribozyme (Fig. 3.1A) and improved some aspects of polymerization (7). However, later studies showed that polymerization also proceeds well in the absence of the P2 oligo (19) and that a truncated P2 oligo is more efficient than the heptanucleotide (12). Because the P2 oligo binds adjacent to the catalytic site we tested most effects found in this study with and without the optimized, truncated P2 oligo. Because the P2 duplex affected the positioning of the 5'-duplex with 7 base pairs on the ribozyme but not that of longer 5'-duplexes we assume that the ribozyme forms fundamentally different interactions with 5'-terminal duplexes of 7-base pairs and longer duplexes.

Arginine at the 5'-duplex can rescue inhibitory effects

When the 5'-duplex had a length of 15-base pairs the RNA/RNA duplex slightly decreased polymerization, in the absence of a P2 oligo (Fig. 3.3C, lower panel). This was concluded from comparing the polymerization efficiency between RNA/RNA duplexes and DNA/RNA duplexes, as well as between the unmodified, amino modified, and arginine modified RNA/RNA duplex. The inhibitory effect of the 2'-hydroxyl group was rescued by the 2'-deoxy modification

as well as by the 3'-terminal amino or arginine modification of the RNA. The rescue by 3'-terminal modifications showed that the 3'-terminal 2'-hydroxyl group caused the inhibitory effect and not internal 2'-hydroxyl groups in the RNA/RNA duplex. These effects probably also existed for a duplex length of 13 base pairs and in the presence of the P2 oligo but were too small to have strong statistical significance (Fig. 3.3C).

5'-terminal duplexes can enter the catalytic site

To explain the inhibitory effect of the 3'-terminal RNA 2'-hydroxyl group at the 5'-duplex we hypothesized that the distal terminus of the RNA/RNA duplex entered the active site and interfered with binding of the primer/template. To test whether the inhibitory effect of the 3'-terminal 2'-hydroxyl group could be due to insertion into the catalytic site we monitored whether the 5'-duplex could be used as a primer/template duplex and extended by the polymerase ribozyme. To obtain a templating sequence the 5'-terminus of the polymerase ribozyme was elongated by four nucleotides. Polymerization assays showed that the radiolabeled RNAs at the 5'-duplex were indeed extended by the polymerase ribozyme, with a strong dependence on the length of the 5'-duplex (Fig. 3.4). The same length dependence was visible in the absence and the presence of the P2 oligo. The dependence followed a pattern that coincided with the periodicity of an A-form helix (11 base pairs), with the exception of the 7 base pair duplex. These results showed that the distal terminus of the 5'-duplex entered the catalytic site of the ribozyme, confirming the hypothesis that the 3'-terminal 2'-hydroxyl group

of 5'-terminal RNA/RNA duplexes could inhibit polymerization by insertion into the catalytic site.

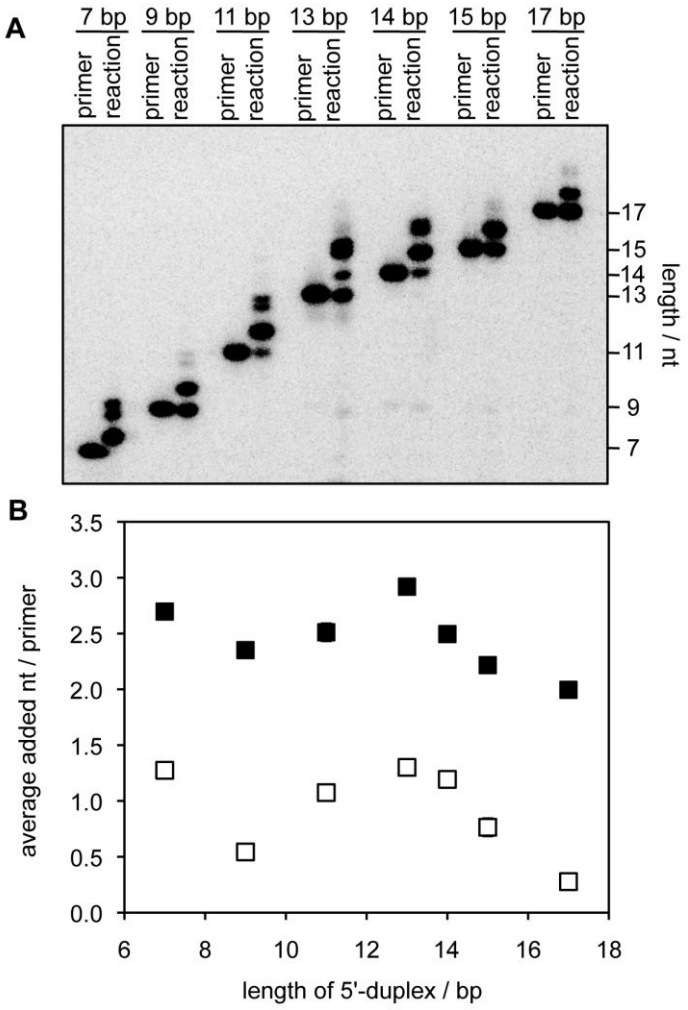

Figure 3.4: Extension of primers that were base paired to the 5'-terminus of the polymerase ribozyme. The extension efficiency of these primers was measured as a function of the length of the 5'-duplex. A) Autoradiogram of PAGE separated polymerization products, in the absence of the P2 oligo. For each length of the 5'-duplex (indicated) the unreacted primer and the reaction products are shown. B) Quantification of polymerization efficiencies. The polymerization efficiency was measured as the average number of nucleotides added per primer and plotted as a function of the length of the 5'-duplex. The polymerization efficiencies in the absence (open squares) and in the presence of the P2 oligo (filled squares) are shown. Errors are standard deviations from triplicate experiments and were usually smaller than the symbols.

Arginine at the P2 duplex can mediate an initial burst of polymerization

The second site on the ribozyme that was used for the attachment of conjugates is the P2 duplex, which was formed by the polymerase ribozyme base pairing to the RNA hexanucleotide 5'-GGCGCC-3' (7, 12) (Fig. 3.1). The P2 oligo is positioned adjacent to the catalytic site as judged by the crystal structure of the catalytic core of the ribozyme (Fig. 3.1B, (27)). To test whether the positive charge next to the catalytic site could improve polymerization further we modified both the 5'-terminus and the 3'-terminus of this P2 oligo with arginine (Fig. 3.1A).

An arginine or amino modification at the 5'-terminus of the P2 oligo resulted in an initial burst of polymerization but caused a stalling of polymerization after five or six nucleotides were added (Fig. 3.5). This mirrors the behavior when the P2 oligo is a heptanucleotide, differing from our hexanucleotide by a 3'-terminal adenosine (12). This 3'-terminal adenosine interacts with the single-stranded portion of the same template as used in this study (T21) but not with other templates (T50a, T50b). Consistent with that we did not find an influence of modifications at the 5'-terminus of the P2 oligo when other templates were used (T50a or T50c from reference (12); data not shown). When the arginine or amino modification was placed at the 3'-terminus of the P2 oligo it did not affect polymerization efficiency. This is consistent with a previous study, which found that nucleotide extensions at the 3'-terminus of the P2 oligo are tolerated (12).

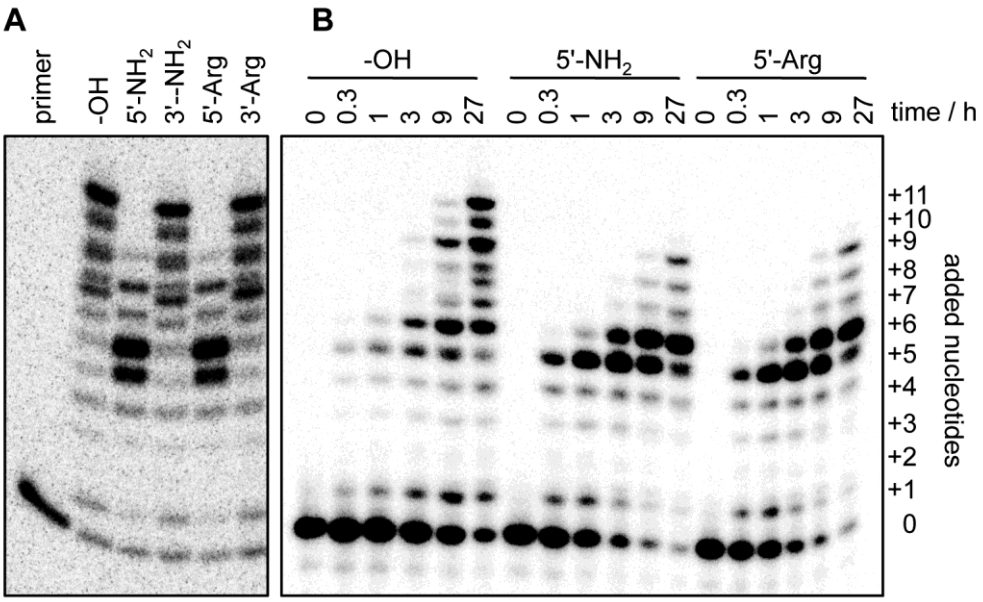

Figure 3.5: Influence of arginine and amino modifications at the P2 oligo. The position and the modification of the DNA P2 oligo 5'-GGCGCC-3' is shown for each reaction, as well as the number of nucleotides added to the primer. The images are representative for three experiments. A) Autoradiogram of PAGE separated polymerization products, after 24 hours of polymerization. B) Autoradiogram of PAGE separated polymerization products, after increasing times for polymerization.

Influence of arginine conjugates at low magnesium concentrations

All experiments above were conducted at magnesium ion concentrations of 80 mM Mg^{2+} (64 mM free Mg^{2+}), which may be too high for arginine to compete with, to bind to phosphodiester oxygens. Therefore, we reduced the concentration of Mg^{2+} to 40 mM (24 mM free Mg^{2+}), which is high enough to obtain quantifiable data from polymerization but perhaps low enough to see a positive effect of arginine cofactors (19). A positive effect by arginine cofactors at this concentration would not mean that the polymerase ribozyme efficiency is improved over its optimal activity (which requires 200 mM Mg^{2+}) but that single

arginines could have a role in nucleic acid interactions in an RNA world, at these lower Mg^{2+} concentrations. However, even at this low concentration we did not detect increased polymerization efficiencies due to arginine (Fig. 3.6). On the contrary, the arginine modification was inhibitory when the 5'-duplex had a length of 11 or 13 base pairs. The inhibitory effect at a duplex length of 11 base pairs was consistent with the effect at 80 mM Mg^{2+} whereas the inhibitory effect at a 5'-duplex length of 13 base pairs was not seen at 80 mM Mg^{2+} and may therefore reflect a minor structural change of the ribozyme between 40 mM and 80 mM Mg^{2+}.

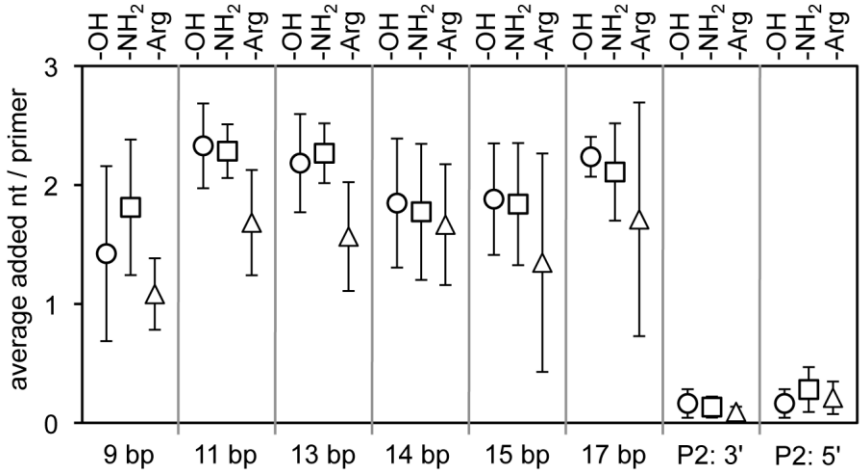

Figure 3.6: Influence of arginine and amino modifications on polymerization at low magnesium concentration. Quantitation of polymerization efficiencies for DNA conjugates at the ribozyme 5'-terminus (in the absence of a separate P2 oligo) and at the P2 site. The polymerization efficiency is described as the average number of nucleotides added per primer. For each experiment, three DNAs were tested: Unmodified (circles), amino modified (squares), and arginine modified (triangles) DNAs. Errors are standard deviations from three experiments.

3.4 Discussion

In an effort to increase the polymerization efficiency of the polymerase ribozyme we tested whether arginine conjugates could improve polymerization. However, we found that arginine did not improve polymerization when placed at ten different positions on the polymerase ribozyme.

Why did the use of arginine-nucleic acid conjugates not improve the efficiency of the polymerase ribozyme? One possibility is that a single arginine is not sufficient to show a strong effect on primer/template binding. However, our assay is sensitive enough to detect even single hydrogen bonds that affect substrate binding (14). Second, although we tested ten different positions for arginine on the ribozyme the best location may not have been among them. Third, the current polymerase ribozymes may not benefit from the conjugates because the ribozymes were optimized in the absence of these conjugates. A partial randomization and re-selection in the presence of these conjugates may find polymerase ribozymes that efficiently use the conjugates. Lastly, it is possible that the magnesium concentration that is necessary for activity of the polymerase ribozyme shielded the phosphodiester groups sufficiently that the effect of an arginine was too low to detect. Although we decreased the concentration of free magnesium ions to 24 mM (Fig. 3.6) we did not find a beneficial effect of arginine on polymerization. This suggests that at the magnesium concentration necessary for activity of the ribozyme single arginines cannot improve polymerization of the existing polymerase ribozymes.

The potential benefit of the amino acid histidine for acid-base catalysis in ribozymes and deoxyribozymes was explored previously. Histidine promised to be useful for a catalytic function because it has a pK_A close to the neutral pH, whereas nucleic acids do not (30). Indeed, an in vitro selection found deoxyribozymes that use free histidine as cofactor, probably with a catalytic role (31). However, the rate enhancements of histidine-using deoxyribozymes are not higher than those that use divalent metal ion cofactors or no cofactors at all (32, 33), and it was found that ribozymes can perturb the pK_As of nucleobases close to the neutral pH (34-36). Additionally, it appears easier for nucleic acids to use divalent cations rather than histidine as cofactor (37). Therefore, histidine (and perhaps any other amino acid) does not seem to be important for general acid-base catalysis in ribozymes or deoxyribozymes.

Peptides and proteins fulfill several non-catalytic roles in natural ribozymes. The bacterial RNase P ribozyme requires the C5 protein for recognition of the pre-tRNA substrate (38) and ribosomal proteins fulfill a very diverse set of functions (39). Although natural hammerhead ribozymes do not require a protein cofactor, a trans-acting variant of the hammerhead ribozyme benefits from the nonspecific binding of the HIV p7 nucleocapsid protein, for the annealing of substrates and the dissociation of products (40). Therefore, trans-acting ribozymes can benefit from peptides or proteins for the function of substrate interactions. Although our study did not identify how single arginines can assist ribozyme polymerization we assume that a different setup with short

peptides can help the polymerase ribozyme to bind the primer/template substrate.

Ribozymes have been selected previously to require the presence of a peptide or protein to be active (41-43). Here, the peptides / proteins appear to stabilize the catalytically active structure of the ribozyme. However, in none of these cases was the activity of the RNP complex higher than that of the parent ribozyme. Therefore, these RNPs show how the activity of a ribozyme can be regulated by a peptide but not how the activity can be increased. In contrast, our study aimed solely to obtain ribozymes with higher efficiency.

In an RNA world, amino acid - nucleid acid conjugates could have been crucial intermediates for establishing a translation system (20, 21, 24). In the first step amino acid-nucleic acid conjugates would have been synthesized. Although we did not find a functional benefit of single amino acid conjugates for the polymerase ribozyme they could have carried different immediate evolutionary advantages (20). With respect to the evolution of the translational apparatus these conjugates would have served as the ancestors of aminoacyl-tRNAs (44). The next step in the evolution would have been the attachment of multiple amino acids to a single conjugate. This formation of peptide bonds can be catalyzed by ribozymes (45) and could have carried immediate benefits, for example by tighter interactions of diarginine with RNA than of arginine. If the source of this second amino acid would have been another conjugate then the ribozyme that catalyzed this peptidyl transfer would have been a primitive ribosome: the conjugates (precursors of tRNAs) would be aligned by base pairing to an mRNA (either a

sequence in the ribozyme or a separate RNA). The nucleic acid portion of the conjugate would then have served as precursor to the tRNA anticodon and facilitated the first encoded peptide synthesis. Further evolutionary steps would have improved the efficiency and accuracy of this machinery to the present-day translation apparatus. One strength of this model is that each of these evolutionary steps has been shown to be accessible to ribozymes, and that each evolutionary step carried an evolutionary advantage for the RNA world organism (20, 46).

3.5 Materials and Methods

Ribozymes and substrates

Ribozymes were synthesized by in vitro transcription from PCR products, using bacteriophage T7 RNA polymerase as described (19). Transcribed ribozymes were purified by 7 M urea 5% polyacrylamide gel electrophoresis (PAGE). RNAs were purchased from Dharmacon, and DNAs were purchased from IDT. All RNAs and DNAs were PAGE purified. Primers were radiolabeled using T4 polynucleotide kinase (NEB) and [γ-^{32}P] ATP (Perkin-Elmer). All chemicals were Molecular Biology grade or higher.

Synthesis of conjugates

Arginine conjugates were synthesized from amino-modified RNAs or DNAs via NHS-activated arginine. The amino modifications contained a tether to the nucleic acid by six methylene groups (C6-linker). In dry DMF, 143 mM a-

amino-Fmoc-arginine were reacted with 143 mM N-hydroxy succinimide (NHS) and 143 mM N,N'-dicyclohexyl carbodiimide (DCC) for 1 hour at 50°C. After cooling to room temperature the reaction mixture was mixed with the 2.5-fold volume of an aqueous solution with 100 mM amino-modified nucleic acid and 200 mM MES/NaOH pH 6.5. After incubation for 2 hours at room temperature the reaction mixture was dried in vacuum, then deprotected in an excess of 100 mM NaOH for 1 hour at 40°C. RNA coupling products were deprotected in an excess of 50 mM NaOH and 50 mM Na_2CO_3 for 2 hours at 10°C. Deprotected products were neutralized, ethanol precipitated, and purified by 7 M urea 20% PAGE. The overall yield was 5-10 % for RNA-arginine conjugates and 10-20% for DNA-arginine conjugates, as calculated based on the nucleic acid. The identity of the DNA conjugates was confirmed by MALDI Mass Spectroscopy, using hydroxypicolinic acid and ammonium citrate as matrix and a DNA 12mer and 21mer as internal standard for calibration. Expected mass for our test amino modified DNA: 5330.6; found: 5330.7. Expected mass for the corresponding arginine modified DNA with Fmoc protection: 5709.0; found: 5708.7. Expected mass for deprotected arginine modified DNA: 5486.8. Found: 5486.9. Additionally, the identity of both DNA and RNA conjugates was confirmed by the migration pattern of 5'-radiolabeled samples in denaturing PAGE.

Ribozyme Reactions

Ribozyme reactions were performed as described (19). All RNAs were dissolved in water at the appropriate concentration (final reaction concentration:

2 μM Ribozyme, 1 μM template, less than 50 nM 5'-radiolabeled primer, 2.5 μM P2 oligo, 2.5 μM 5' terminus oligo), heat denatured (2 min/80°C) and cooled to the reaction temperature (17°C) at 0.1°C/sec. Reactions were started by adding 2.5x reaction buffer containing magnesium chloride, buffer (Tris/HCl, pH 8.5), and NTP (an equimolar mix of the four nucleoside triphosphates) so that the final concentrations were 50 mM Tris/HCl and 4 mM of each NTP. Magnesium chloride was 80 mM with the exception of primer extensions at the 5'-duplex (200 mM $MgCl_2$), or reactions annotated to contain 40 mM $MgCl_2$. Reaction times were 24 hours for reactions with P2 oligo and 3 hours for reactions without P2 oligo if not indicated otherwise. Reaction times for reactions with 40 mM $MgCl_2$ were 22 hours. The reason for the different incubation time is that in the absence of the P2 oligo polymerization is fast during the first hours and then stalls, whereas in the presence of the P2 oligo polymerization is slower during the first hours but extends further (12). The reactions were stopped by the addition of a 1.5 fold volume of stop buffer (80% (v/v) formamide, 200 mM Na_2EDTA at pH 8.4) and a template-complementary RNA added in 20-fold excess over the template. The mixtures were heat denatured (2 min/80°C) and cooled to room temperature at 0.1°C/sec before loading and separating on 7 M urea 0.5x TBE 20% PAGE.

Data analysis

Autoradiographs of the PAGE separations were recorded by a PMI phosphorimager (Bio-Rad) and quantitated using the software Quantity One.

Shifts higher than 11 nt above the primer were counted as full-length extension. The values for “average nucleotides per primer” were obtained by multiplying the fraction of intensity for each band (minus background signal) with the number of added nucleotides corresponding to that band. For quantifying the effect of 2'-deoxy substitutions, the method was described previously (14). All experiments were repeated at least in triplicate.

3.6 Acknowledgements

This work was funded by NSF grant 0743985 to U.F.M. and NIH MBTG grant GM #08326 to J.E.M. We thank Jerry Yang for advice in the synthesis and coupling of NHS esters, and Russ Doolittle for helpful discussions on arginine chemistry. Jing Yang is thanked for helpful comments on the manuscript.

Chapter 3, in full, is a reprint of the material as it appears in PLoS ONE, Yao C, **Moretti JE**, Struss PE, Spall JA, Müller UF (2011) Arginine Cofactors on the Polymerase Ribozyme. PLoS ONE 6(9): e25030.The dissertation author was the second author on this paper.

3.7 References

1. Woese CR (1968) The fundamental nature of the genetic code: prebiotic interactions between polynucleotides and polyamino acids or their derivatives. *Proc Natl Acad Sci U S A 59*: 110-117.

2. Crick FHC (1968) The origin of the genetic code. *J Mol Biol 38*: 367-379.

3. Orgel LE (1968) Evolution of the genetic apparatus. *J Mol Biol 38*: 381-393.

4. Gilbert W (1986) The RNA world. *Nature 319*: 618.

5. Chen X, Li N, Ellington AD (2007) Ribozyme catalysis of metabolism in the RNA world. *Chem Biodivers 4*: 633-655.

6. Cheng LK, Unrau PJ (2010) Closing the circle: replicating RNA with RNA. *Cold Spring Harb Perspect Biol 2*: a002204.

7. Johnston WK, Unrau PJ, Lawrence MS, Glasner ME, Bartel DP (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science 292*: 1319-1325.

8. Lawrence MS, Bartel DP (2005) New ligase-derived RNA polymerase ribozymes. *RNA 11*: 1173-1180.

9. Zaher HS, Unrau PJ (2007) Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA 13*: 1017-1026.

10. Wochner A, Attwater J, Coulson A, Holliger P (2011) Ribozyme-catalyzed transcription of an active ribozyme. *Science 332*: 209-212.

11. Attwater J, Wochner A, Pinheiro VB, Coulson A, Holliger P (2010) Ice as a protocellular medium for RNA replication. *Nat Commun 1*: doi:10 1038/ncomms1076.

12. Yao C, Muller UF (2011) Polymerase ribozyme efficiency increased by G/T-rich DNA oligonucleotides. *RNA 17*: 1274-1281.

13. Lawrence MS, Bartel DP (2003) Processivity of ribozyme-catalyzed RNA polymerization. *Biochemistry 42*: 8748-8755.

14. Müller UF, Bartel DP (2003) Substrate 2'-hydroxyl groups required for ribozyme-catalyzed polymerization. *Chemistry and Biology 10*: 799-806.

15. Misra VK, Shiman R, Draper DE (2003) A thermodynamic framework for the magnesium-dependent folding of RNA. *Biopolymers 69*: 118-136.

16. Mason PE, Neilson GW, Dempsey CE, Barnes AC, Cruickshank JM (2003) The hydration structure of guanidinium and thiocyanate ions: implications for protein stability in aqueous solution. *Proc Natl Acad Sci U S A 100*: 4557-4561.

17. Schmidt CLA, Kirk PL, Appleman WK (1930) The apparent dissociation constants of arginine and of lysine and the apparent heats of ionization of certain amino acids. *J Biol Chem 88*: 285-293.

18. Hoffman MM, Khrapov MA, Cox JC, Yao J, Tong L, et al. (2004) AANT: the Amino Acid-Nucleotide Interaction Database. *Nucleic Acids Res 32*: D174-181.

19. Müller UF, Bartel DP (2008) Improved polymerase ribozyme efficiency on hydrophobic assemblies. *RNA 14*: 552-562.

20. Orgel LE (1989) The origin of polynucleotide-directed protein synthesis. *J Mol Evol 29*: 465-474.

21. Szathmary E, Maynard Smith J (1997) From replicators to reproducers: the first major transitions leading to life. *J Theor Biol 187*: 555-571.

22. Maizels N, Weiner AM (1994) Phylogeny from function: evidence from the molecular fossil record that tRNA originated in replication, not translation. *Proc Natl Acad Sci U S A 91*: 6729-6734.

23. Illangasekare M, Yarus M (1999) Specific, rapid synthesis of Phe-RNA by RNA. *Proc Natl Acad Sci U S A 96*: 5470-5475.

24. Lee N, Bessho Y, Wei K, Szostak JW, Suga H (2000) Ribozyme-catalyzed tRNA aminoacylation. *Nat Struct Biol 7*: 28-33.

25. Kumar RK, Yarus M (2001) RNA-catalyzed amino acid activation. *Biochemistry 40*: 6998-7004.

26. Baskerville S, Bartel DP (2002) A ribozyme that ligates RNA to protein. *Proc Natl Acad Sci U S A 99*: 9154-9159.

27. Shechner DM, Grant RA, Bagby SC, Koldobskaya Y, Piccirilli JA, et al. (2009) Crystal structure of the catalytic core of an RNA-polymerase ribozyme. *Science 326*: 1271-1275.

28. Joseph S, Weiser B, Noller HF (1997) Mapping the inside of the ribosome with an RNA helical ruler. *Science 278*: 1093-1098.

29. Lesnik EA, Freier SM (1995) Relative thermodynamic stability of DNA, RNA, and DNA:RNA hybrid duplexes: relationship with base composition and structure. *Biochemistry 34*: 10807-10815.

30. Joyce GF (1998) Nucleic acid enzymes: playing with a fuller deck. *Proc Natl Acad Sci U S A 95*: 5845-5847.

31. Roth A, Breaker RR (1998) An amino acid as a cofactor for a catalytic polynucleotide. *Proc Natl Acad Sci U S A 95*: 6027-6031.

32. Emilsson GM, Breaker RR (2002) Deoxyribozymes: new activities and new applications. *Cell Mol Life Sci 59*: 596-607.

33. Geyer CR, Sen D (1997) Evidence for the metal-cofactor independence of an RNA phosphodiester-cleaving DNA enzyme. *Chem Biol 4*: 579-593.

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

35. Wilson TJ, McLeod AC, Lilley DM (2007) A guanine nucleobase important for catalysis by the VS ribozyme. *EMBO J 26*: 2489-2500.

36. Cottrell JW, Scott LG, Fedor MJ (2011) The pH dependence of hairpin ribozyme catalysis reflects ionization of an active site adenine. *J Biol Chem*.

37. Faulhammer D, Famulok M (1996) The Ca^{2+} ion as a cofactor for a novel RNA-cleaving deoxyribozyme. *Angew Chem Int Ed Engl 35*: 2837-2841.

38. Koutmou KS, Zahler NH, Kurz JC, Campbell FE, Harris ME, et al. (2010) Protein-precursor tRNA contact leads to sequence-specific recognition of 5' leaders by bacterial ribonuclease P. *J Mol Biol 396*: 195-208.

39. Zimmermann RA (2003) The double life of ribosomal proteins. *Cell 115*: 130-132.

40. Herschlag D, Khosla M, Tsuchihashi Z, Karpel RL (1994) An RNA chaperone activity of non-specific RNA binding proteins in hammerhead ribozyme catalysis. *EMBO J 13*: 2913-2924.

41. Robertson MP, Ellington AD (2001) In vitro selection of nucleoprotein enzymes. *Nat Biotechnol 19*: 650-655.

42. Robertson MP, Knudsen SM, Ellington AD (2004) In vitro selection of ribozymes dependent on peptides for activity. *RNA 10*: 114-127.

43. Atsumi S, Ikawa Y, Shiraishi H, Inoue T (2003) Selections for constituting new RNA-protein interactions in catalytic RNP. *Nucleic Acids Res 31*: 661-669.

44. Nagel GM, Doolittle RF (1995) Phylogenetic analysis of the aminoacyl-tRNA synthetases. *J Mol Evol 40*: 487-498.

45. Zhang B, Cech TR (1997) Peptide bond formation by in vitro selected ribozymes. *Nature 390*: 96-100.

46. Noller HF (2004) The driving force for molecular evolution of translation. *RNA 10*: 1833-1837.

Chapter 4

Developing an in vitro selection for a polymerase ribozyme with improved substrate binding

4.1 Introduction

Creating a RNA catalyst capable of replicating itself would bolster the RNA world hypothesis and also support efforts to make a self-replicating RNA system. This would further our understanding of life's early ancestors. Although a self-replicating ribozyme has not yet been developed, significant progress has been made towards this goal.

In the early nineties, Dave Bartel and Jack Szostak used in vitro selection to develop a ligase ribozyme—a catalytic RNA that catalyzes the formation of a 3',5'- phosphodiester between two RNA molecules aligned on a template (1). This ligase ribozyme was then used as a starting point to develop a ribozyme that catalyzes RNA polymerization (2). The ligase ribozyme was partially randomized and 76 random nucleotides were added to its 3' terminus to form a pool of variants that served as the starting point of an in vitro selection. This selection produced the R18 polymerase ribozyme, a 189 nucleotide long ribozyme capable of adding 14 nucleotides to an RNA primer annealed to a template (Fig. 4.1). It catalyzes 3'-5' phosphodiester bond formation using ribonucleoside triphosphates and a primer-template duplex as its substrate. A second in vitro selection that used the same starting pool, but made some

modifications to the selection procedure identified eight additional polymerase ribozymes that were also derived from the ligase ribozyme (3).

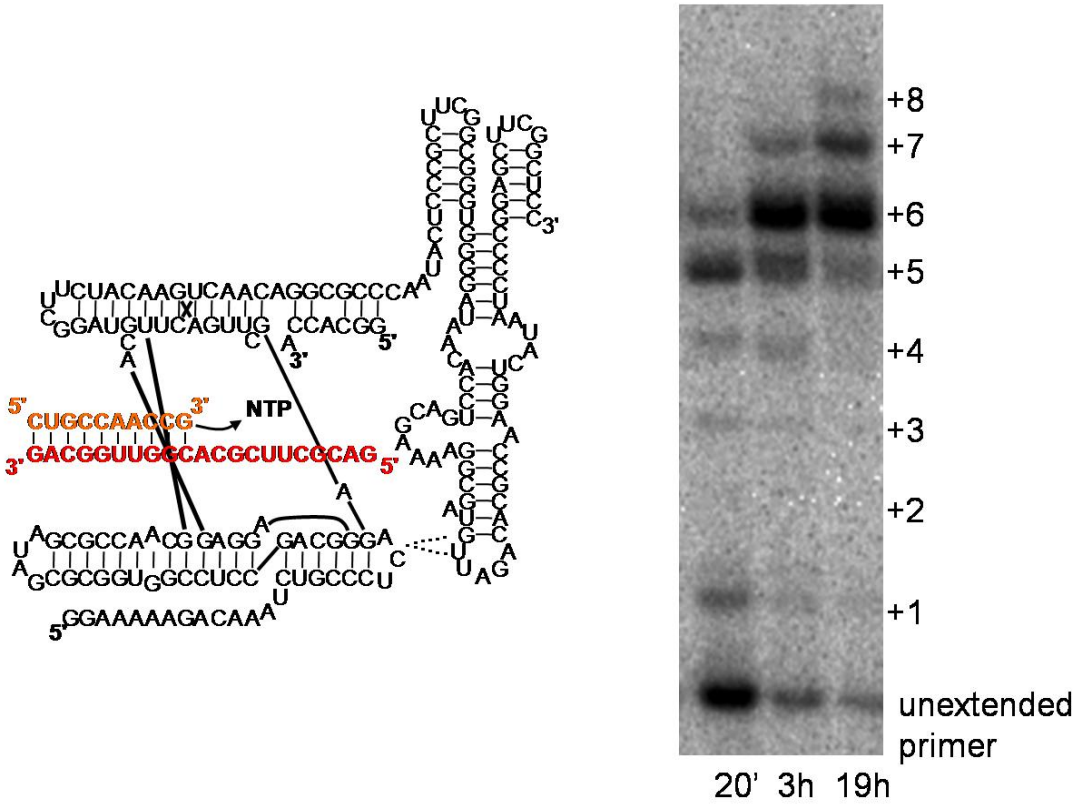

Figure 4.1: The polymerase ribozyme is an RNA that catalyzes RNA polymerization. A) The secondary structure of the R18 polymerase ribozyme is represented in black. Its substrate is an RNA duplex with a primer (orange) annealed to a template (red). B) After 19 hours the polymerase ribozyme extends this ten nucleotide long primer by an additional eight nucleotides. This reaction can be monitored by separating radiolabeled primers on a polyacrylamide gel after the polymerization reaction.

The R18 polymerase ribozyme can catalyze general RNA polymerization, but it is not efficient enough for self-replication. Specifically, it is limited by a low substrate affinity, with a K_M of about 3mM (4). Therefore, efforts to improve RNA

catalyzed RNA polymerization have utilized different strategies that aim to improve substrate binding. Using templates whose single stranded region can form base pairs with the polymerase ribozyme can improve polymerization efficiency (5, 6). However, this strategy constrains the template sequence and is insufficient for polymerizing sequences long enough to enable self-replication. Polymerization efficiency has also been shown to improve when the ribozyme is co-localized with its substrate by attaching hydrophobic anchors to both (7) or by concentrating them in eutectic ice phases (5). Annealing G/T-rich DNA oligonucleotides to the polymerase ribozyme also improves polymerization. The DNA makes a series of non-specific interactions with the non-base paired section of the template that improve substrate binding (8).

Another strategy for improving polymerase ribozymes is to use in vitro selection methods that enable multi-turnover catalysis and put selective pressure on ribozymes to bind their substrates during the selection. The R18 polymerase ribozyme was connected to its substrate during the selection and therefore was not exposed to selection pressure for substrate binding. In vitro selections for trans-binding ribozymes rely on compartmentalization of the DNA encoding the ribozyme with the ribozyme's substrate. This enables active sequences to be selected without attaching them to their substrates during the reaction. Zaher and Unrau developed an in vitro selection strategy that uses water in oil emulsions to compartmentalize DNAs encoding polymerase ribozyme variants with their substrates. They found an improved version of the R18 polymerase ribozyme, termed B6.61, that can catalyze 20 nucleotides of primer extension (9) (Fig. 4.2).

Wochner et al. developed a selection strategy—compartmentalized bead-tagging—that uses both water in oil emulsion and binding of ribozyme genes and substrates to streptavidin-coated magnetic beads. This method yielded polymerase ribozyme tC19Z which synthesized 95 nucleotides of RNA and was also used to synthesize an active ribozyme from an RNA template (Fig. 4.2). However, this ribozyme performs best on a template that is a series of 11 nucleotide repeats of a specific sequence that is favored by the polymerase ribozyme (6).

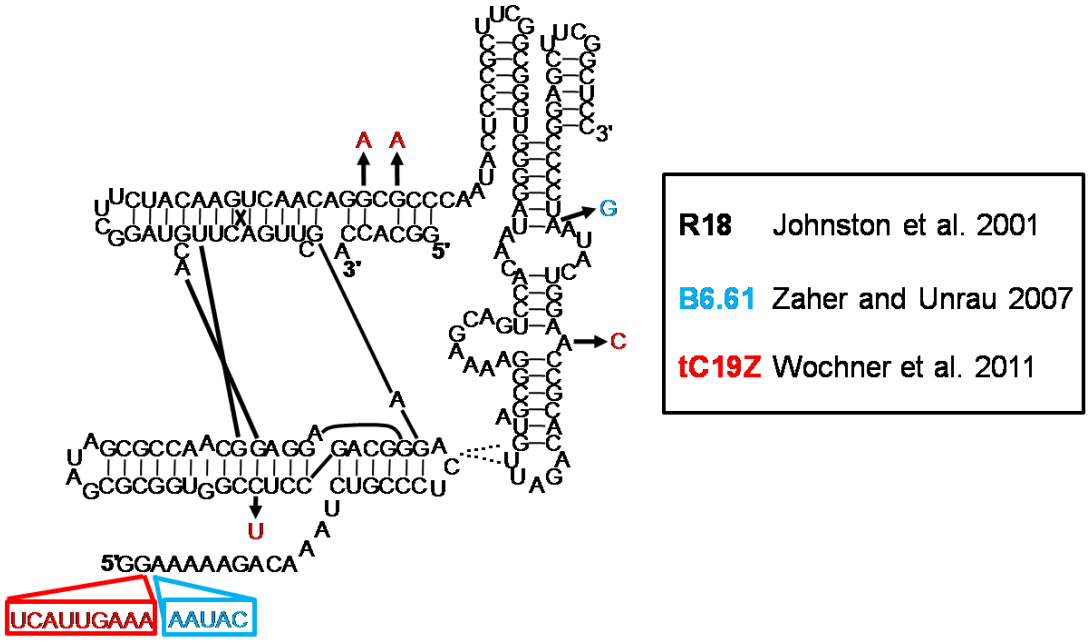

Figure 4.2: Variants of the polymerase ribozyme. The original R18 polymerase ribozyme is shown in black. The B6.61 polymerase ribozyme has five nucleotides inserted at the 5' end and an A to G mutation (shown in blue). The tC19Z variant has nine nucleotides inserted at the 5' end in addition to four mutations (shown in red).

In this project we aimed to develop an in vitro selection for finding polymerase ribozymes with improved substrate binding and polymerization efficiency by utilizing three strategies: 1) We used an in vitro selection method that put selective pressure on ribozymes to bind their substrate in trans. 2) We did the selection in the presence of positively charged cofactors that could aid the ribozyme by reducing charge repulsion between the negatively charged ribozyme and its negatively charged substrate. 3) We added 75 randomized nucleotides to the 5' end of previously selected polymerase ribozyme sequences to enable the selection of a new substrate binding domain.

We used an in vitro selection strategy based on the one developed by Zaher and Unrau (9). During the selection, the ribozyme is not covalently attached to its substrate so that the ribozyme has to bind its substrate in trans. To select for a trans-reacting ribozyme, the ribozyme or its gene must be physically linked to its reaction product. This is accomplished by covalently linking the ribozyme's substrate to a DNA sequence encoding the ribozyme. In the selection, the DNA is transcribed into ribozymes by T7 RNA polymerase. Functional ribozymes will catalyze RNA polymerization of the substrate covalently bound to their gene. To ensure that freely diffusing ribozymes act only on the substrate attached to the DNA encoding them, the reaction is performed in a water in oil emulsion. The emulsion droplets compartmentalize each ribozyme together with the DNA encoding it. Isolation of elongated primers will therefore isolate the genes of functional polymerase ribozymes (Fig. 4.3).

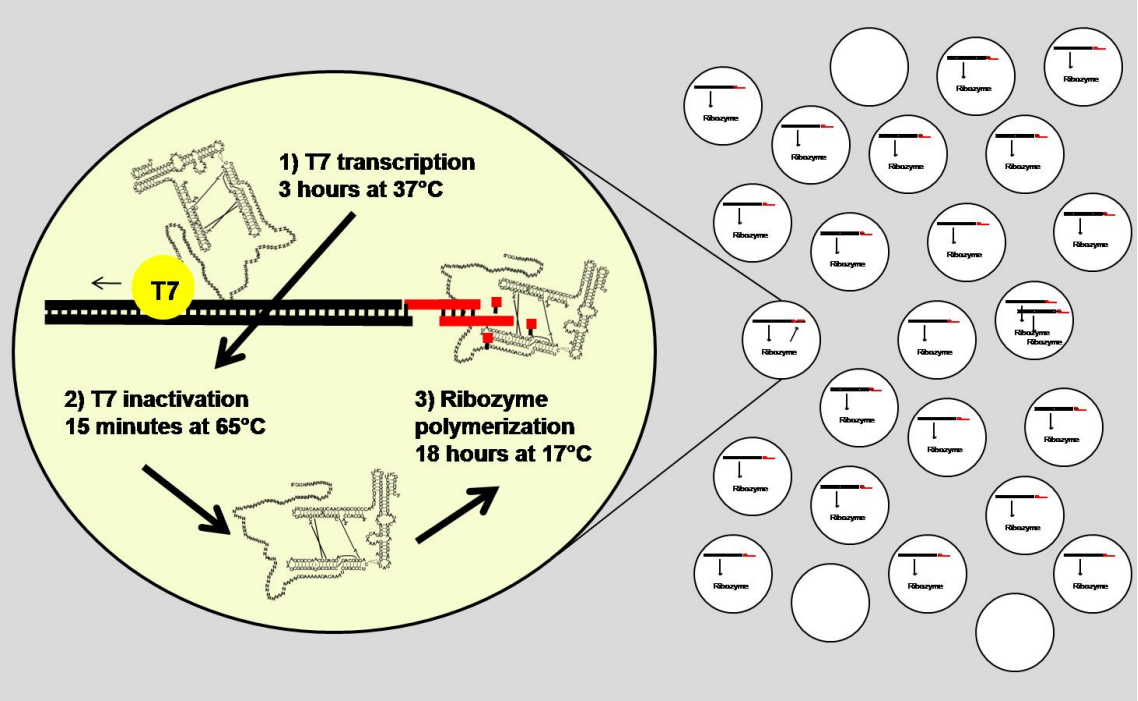

Figure 4.3: Compartmentalization of ribozyme variants with their substrates. Each ribozyme gene is ligated to the ribozyme's substrate. The genes are transcribed by T7 RNA polymerase and allowed to react with their substrates. Each gene is compartmentalized in a water in oil emulsion droplet, so active ribozymes will only act on substrates connected to their own gene. During the selection, there are ~10^{14} droplets with an average of one gene per droplet. In the first round of selection, the vast majority of sequences will be inactive.

During the selection, the ribozyme variants were exposed to amino acids covalently linked to short DNAs, which can form Watson-Crick base pairs with the polymerase ribozyme. This method enables the ribozyme to bind the DNA-amino acid conjugate through base-pairing interactions while leaving the amino acid free to interact with the substrate. DNA-amino acid conjugates containing lysine, arginine, and di-arginine were used in the selection. These amino acids are positively charged and could therefore be utilized by the ribozyme to help bind its negatively charged substrate.

As the starting point for the selection, we used a pool of polymerase ribozyme variants. Three previously selected polymerase ribozymes—R18 from the Bartel lab, B6.61 from the Urau lab, and tC19Z from the Holliger lab—were used in the first round of the selection, with each representing a third of the pool (2, 9, 6). Seventy-five randomized nucleotides were added to the 5' end of all ribozyme sequences. About 10^{14} sequences were present for the first round of selection.

After doing 10 rounds of this selection, no RNA polymerization was detected in any of the pools. In this section I will summarize the progress that was made establishing this selection. This includes developing methods for forming small, uniform water in emulsion droplets, doing T7 RNA polymerization of ribozymes and ribozyme polymerization simultaneously in emulsion droplets, and discriminating between RNA primers that were extended by the polymerase ribozyme. I will also discuss ways that the selection methods could be improved to select an improved polymerase ribozyme.

4.2 Results

There are several steps in the selection procedure for trans-acting polymerase ribozymes (Fig. 4.4). First a DNA pool of 10^{14} polymerase ribozyme variants are synthesized. The pool sequences include a BtgI restriction site, a T7 promoter, 75 randomized nucleotides, and the sequence of an existing polymerase ribozyme. This pool is digested and ligated to an RNA that serves as a substrate for the polymerase ribozyme. The substrate is annealed to an RNA

template containing a thiol at its 5' terminus. These pool molecules are then compartmentalized in a water in oil emulsion along with T7 RNA polymerase and other reaction components including nucleoside triphosphates and buffer. The pool is transcribed and active RNA sequences are able to extend the primer. The pool is then purified from the emulsions by a series of extractions and extended primers are separated from unextended primers. The selected sequences are PCR amplified and used as the pool for the next round of selection.

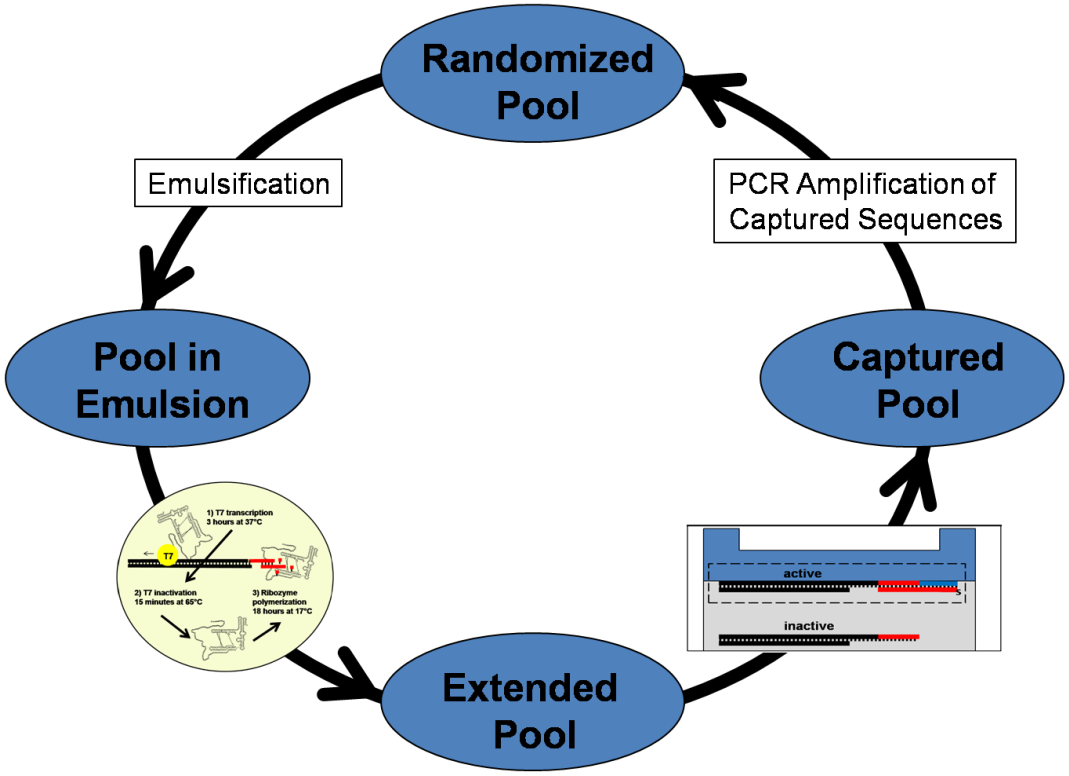

Figure 4.4: Overview of one round of in vitro selection in emulsion. The selection begins with a pool of 10^{14} polymerase ribozyme genes connected to RNA substrates. They are then compartmentalized in water in oil emulsion droplets, transcribed into RNA, and allowed to react with their substrates. The pool is then purified from the emulsion and substrates that have been extended are separated from those that have not. The selected sequences are PCR amplified and prepared for the starting pool of the subsequent round of selection.

In order to perform this selection, several experimental techniques were first optimized independently, including making small, homogeneous emulsion droplets, optimizing transcription by T7 RNA polymerase and ribozyme polymerization in the droplets, and selectively capturing functional ribozyme genes.

In the first round of selection there are 10^{14} sequences and the vast majority are inactive. When these sequences are compartmentalized in emulsion droplets, the size of the droplets determines how many sequences are in each droplet on average. If the droplets are too large, droplets would contain multiple sequences and allow inactive sequences encapsulated with an active sequence to be carried over to the next round of selection. For this selection, droplets about 1µm in diameter are needed for an average of one sequence per droplet. To generate these, a mixture of 5% surfactant in mineral oil was combined with the aqueous phase. These were mixed by magnetic stirring and then passed through a 100µm channel in a microfluidizer at 4,000psi. This generated a homogeneous water in oil emulsion with droplets of the desired size (Fig 4.5).

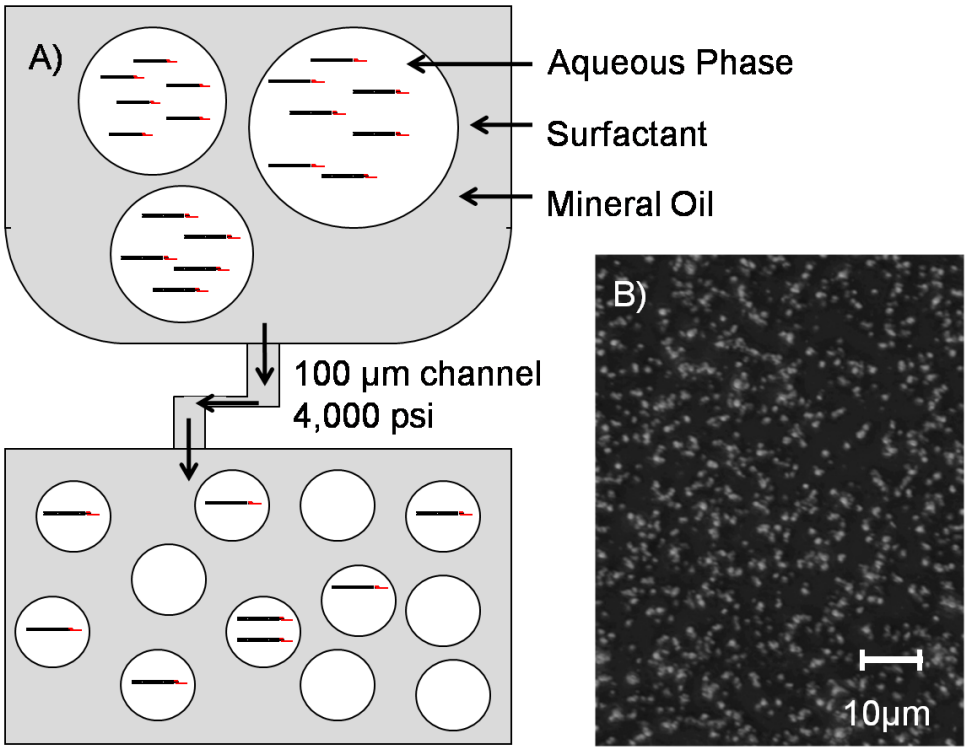

Figure 4.5: Formation of small, homogeneous water in oil emulsion droplets. A) The emulsion droplets are formed by forcing a mixture of aqueous phase, surfactant, and mineral oil through the narrow (100µm) channel at 4,000psi. B) A microscopy image of the emulsion. The droplets are roughly 1µm in diameter.

Inside each emulsion droplet, T7 transcription and RNA catalysis must occur in the same reaction solution. These reactions were optimized inside emulsion droplets using radiolabeled primers to monitor the reaction. Adding sodium chloride to the reaction reduced non-specific interactions between the T7 RNA polymerase and adding glycerol improved T7 polymerase activity. Under optimized conditions, we observed primer extension to five nucleotides by the transcribed polymerase ribozyme (Fig. 4.6).

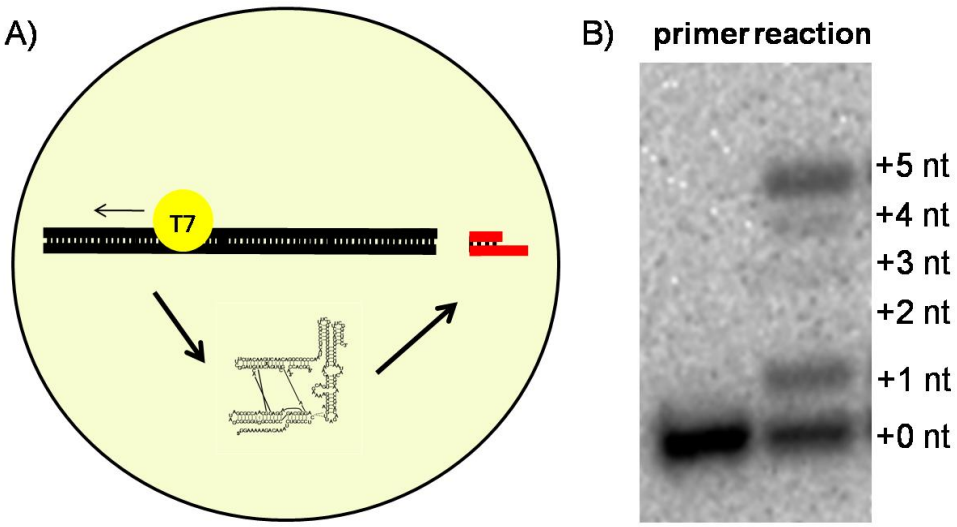

Figure 4.6: Concerted T7 transcription and ribozyme polymerization reactions in emulsion droplets. A) Inside the emulsion droplet, the T7 RNA polymerase transcribes the polymerase ribozyme. In the same reaction volume, the polymerase ribozyme can then catalyze RNA polymerization on a RNA primer-template substrate. B) A reaction mix containing 60mM $MgCl_2$, 80mM NaCl, 40mM TrisHCl pH 7.9, 4mM each NTPs, 2.5U/µl T7 RNA polymerase, and 25% glycerol was combined with 5% ABIL in mineral oil. After passing this mix through a microfluidizer at 4,000psi, the reaction proceeded at 37°C 3 hours, 65°C 15 minutes, and 17°C overnight. The primers were extracted from the emulsion mixture and separated on a 20% polyacrylamide gel. Fifty percent of the primers are extended by five nucleotides.

During the ribozyme polymerization reaction, all primers were base-paired with complementary RNA oligonucleotides containing a thiol modification. After the reaction in emulsion, primers that were extended by active polymerase ribozymes must be separated from unextended primers. To do this, the primer-template duplexes are heated with a competitor RNA oligo that is complementary to the template RNA under specific annealing conditions. Extended primers remain annealed to the template and unextended primers are displaced. The highly specific and strong interaction between mercury and sulfur is used to

capture all the genes that remain annealed to their template on a native polyacrylamide gel containing mercury (Fig. 4.7). These selected sequences are then gel purified and PCR amplified for the next round of selection.

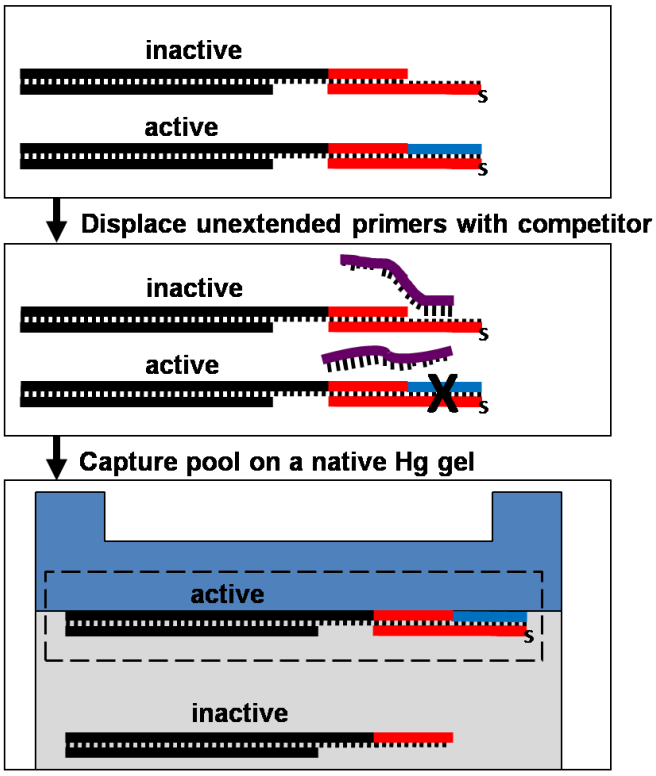

Figure 4.7: Strategy for the selection of extended primers. After the reaction, the genes of active polymerase ribozymes will have their primers extended (blue). These must be separated from inactive genes with unextended primers. All sequences are mixed with a competitor RNA oligo (purple) that is complementary to the thiol containing template (red). Under specific annealing conditions, the competitor will displace the unextended primer, while leaving the extended primer annealed to the template. The active sequences are then gel purified on a native polyacrylamide gel containing mercury.

To test the capture procedure before the actual selection, short radiolabeled RNAs representing unextended (10 nucleotides long) and extended

(15 nucleotides long) primers were annealed to thiol containing templates. They were then incubated with a 16 nucleotide long competitor RNA in an annealing buffer containing 1M Urea, 100mM TrisHCl pH 8.5, and 200mM KCl, 100mM DTT for 10 minutes at 50°C. Under these conditions, unextended primers are displaced and migrate uninhibited through a native mercury gel, while extended primers are retained at the interface between the gel's stacking layer and mercury containing layer (Fig. 4.8).

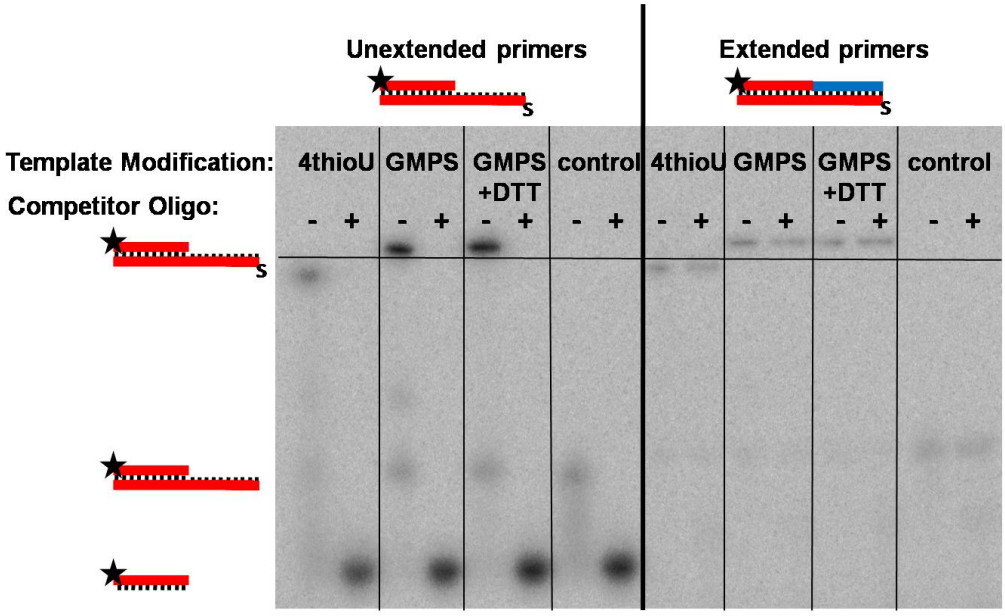

Figure 4.8: Capture of extended primers annealed to thiol containing templates. Radiolabeled primers representing the extended (15 nucleotides long) and unextended (10 nucleotides long) were annealed to thiol containing templates, incubated with a competitor RNA, and run on a native polyacrylamide gel. The black horizontal line shows the gel interface between a stacking layer with no mercury and a layer with mercury. Templates with two different types of thiol modification—4thioU and guanosine monophosphorothioate (GMPS)—were tested.

After establishing methods for generating the emulsion, doing a concerted transcription and polymerase ribozyme reaction in emulsion, and isolating

extended sequences, the actual selection was performed. For the initial pool DNA sequences containing a BtgI restriction site, T7 promoter, seventy-five randomized nucleotides, and the sequence of previously selected polymerase ribozyme were synthesized. The pool was digested with the restriction enzyme BtgI and ligated to a DNA-RNA hybrid oligo to serve as substrate for the polymerase ribozyme. The primer was annealed to an RNA template containing a guanosine monophosphorothioate at its 5' end.

After six rounds of selection, pools for each round of the selection were transcribed and tested for RNA polymerization activity. No activity was observed in any round. To test the selection procedure, one round of selection was performed on a pool that contained a 1:500 dilution of active polymerase ribozyme sequences (the B6.61 polymerase ribozyme) in inactive sequences (B6.61 C47U, an inactive variant). After one round, the pool should be enriched with active sequences and there should be a detectable increase in the ribozyme polymerization activity of the pool. However, no increase in ribozyme polymerization activity was observed after one round (Fig. 4.9).

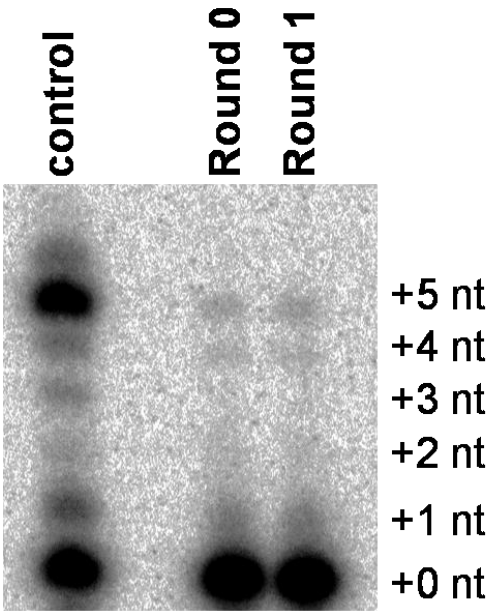

Figure 4.9: Testing the enrichment after one round of selection. A pool containing 500 inactive polymerase ribozyme sequences for every active sequence was subjected to one round of the selection procedure. Both pools were transcribed and tested in ribozyme polymerization reactions. No increased activity was observed after one round of selection.

4.3 Discussion

After six rounds of in vitro selection, there was no detectable activity in the pools. An experiment designed to quantify how much active sequences are enriched after a round of selection showed that pool activity did not increase after one round of selection (Fig. 4.9). Before proceeding with the selection, the procedure should be tested with this type of experiment to demonstrate that the pool is enriched with active sequences after each round. This confirms that the protocol is working and also provides an estimate for the enrichment factor—a number quantifies how many additional copies of an active sequence are in the pool after each round of selection. In this section, I will discuss strategies that

could be used to improve the selection method for a polymerase ribozyme with improved general substrate binding.

Although we showed that annealing a competitor RNA with extended and unextended substrates can be used to separate them in certain contexts, (Fig. 4.8) this experiment was not completely identical to the conditions during the selection. During the selection, the short RNA primers are ligated to longer DNA sequences that include a randomized domain. These randomized sequences could interfere with the competitor annealing. Also, this method was optimized for discriminating unextended sequences from sequence extended five nucleotides. This puts selection pressure on the polymerase ribozymes to catalyze at least five nucleotides of polymerization. Active polymerase ribozymes capable of catalyzing the addition of less than five nucleotides could be lost.

Our capture method relies on non-covalent, base-pairing interactions to discriminate active from inactive sequences. Using covalent interactions to capture active sequences during the selective step would be a more robust strategy. One way to do this is to use thiol modified nucleoside triphosphates, such as 4thioU, during the selection. If the polymerase ribozyme incorporates them into the primer, extended primers can be immediately isolated on a mercury gel. Since the transcription by T7 RNA polymerase and RNA polymerization by the polymerase ribozyme occur in the same reaction volume during the selection, thiol modified nucleosides used during the selection will be incorporated into the transcribed ribozymes. However, Zaher and Unrau successfully used this

strategy in a selection for a polymerase ribozyme, so it is possible to develop active ribozymes with these modifications (9).

Another way to use nucleoside triphosphates with thiol modifications during the selection is to do successive emulsions: one for the transcription of the ribozymes and a second for the RNA polymerization reactions. Wochner et al. used this strategy by anchoring biotinlyated polymerase ribozyme genes to streptavidin coated magnetic beads, transcribing them in emulsion, and then ligating the transcripts to biotinlyated hairpin RNAs anchored to the same beads. These beads were extracted from the emulsion, decorated with substrates for the polymerase ribozyme, and re-emulsified (6). An advantage of this method is that separate reaction mixes are used for the T7 transcription and RNA polymerization reaction. Therefore, thiol modified nucleoside triphosphates could be used in the RNA polymerization reaction without incorporating them into the ribozymes. Furthermore, this method allows for optimizing each reaction mix for the appropriate reaction and eliminates any possibility of the T7 RNA polymerase acting on the polymerase ribozyme's substrate. However, a major drawback of this method is that using one streptavidin magnetic bead per sequence in the selection significantly limits the initial pool compelxity. Wochner et al. had $5x10^7$ sequences in their initial pool, while typical in vitro selections have 10^{13} to 10^{15} sequences.

Ultimately, the advantages of this method probably do not out weigh the sacrifice of several orders of magnitude of pool complexity. However, if another technique could be developed that would allow for the transcription and ribozyme

polymerization reactions to occur in separate reactions volumes, it would improve the method. One possibility is using disulfide bonds to link the transcribed ribozymes and their genes in the first emulsion. This could be accomplished by annealing a short thiol modified RNA to the 5' end of the transcribed ribozymes and attaching a thiol linker to the DNA encoding the pool. Forming a disulfide bond between the DNA and the short RNA annealed to the transcribed ribozymes would connect each ribozyme to the gene encoding it. The bond could then be reduced in a second emulsion to free the ribozymes to act on their substrates. Since this strategy does not rely on magnetic particles, larger pool complexities could be achieved.

In addition to optimizing the selection technique, changing the sequence of the initial pool could also yield improved polymerase ribozymes. Both of the previous selections for polymerase ribozymes with improved processivity and substrate binding used pools based on the R18 polymerase ribozyme. Zaher and Unrau generated their pool by making R18 sequences with 10% mutagenesis in the accessory domains and 0% or 3% mutagenesis the ligase domain. They also added five randomized nucleotides in the 5' end of the ribozyme (9). Wochner et al. also used a mutagenized R18 polymerase ribozyme with thirty-six randomized nucleotides inserted into the 5' end of the ribozyme (6). Since the R18 polymerase ribozyme was selected under conditions that did not provide selective pressure for substrate binding, it may not be the optimum starting point in these selections. An in vitro selection for improved substrate binding that

started with a completely randomized accessory domain could produce polymerase ribozymes with better processivity.

Using in vitro selection strategies that utilize compartmentalization to select for better processivity and general substrate binding are a technically challenging, but promising route to developing a catalytic RNA capable of self replication. Here, some methods were established to set up an in vitro selection in emulsion for the selection of a polymerase ribozyme with improved general substrate binding. However, ultimately the selection was not successful in finding new ribozymes. In the future, developing a selective step that relies on covalent (instead of base-pairing) interactions and improving the pool design could improve the selection.

4.4 Acknowledgements

This work was funded by NSF grant 0743985 to U.F.M. and NIH MBTG grant GM #08326 to J.E.M. The dissertation author is the first author on this material.

4.5 References

1. Bartel, D. P., and Szostak, J. W. (1993) Isolation of new ribozymes from a large pool of random sequences, *Science 261*, 1411-1417.

2. Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E., and Bartel, D. P. (2001) RNA-Catalyzed RNA polymerization: accurate and general RNA-templated primer extension, *Science 292*, 1319-1325.

3. Lawrence, M. S., and Bartel, D. P. (2005) New ligase-derived RNA polymerase ribozymes, *RNA 11*, 1173-1180.

4. Lawrence, M. S., and Bartel, D. P. (2003) Processivity of ribozyme-catalyzed RNA polymerization, *Biochemistry 42*, 8748-8755.

5. Attwater, J., Wochner, A., Pinheiro, V. B., Coulson, A., and Holliger, P. (2010) Ice as a protocellular medium for RNA replication, *Nature Communications 1*, 76.

6. Wochner, A., Attwater, J., Coulson, A., and Holliger, P. (2011) Ribozyme-catalyzed transcription of an active ribozyme, *Science 332*, 209-212.

7. Muller, U. F. and Bartel, D. P. (2008) Improved polymerase ribozyme efficiency on hydrophobic assemblies, *RNA 14*, 552-562.

8. Yao, C. and Muller, U. F. (2011) Polymerase ribozyme efficiency increased by G/T-rich DNA oligonucleotides, *RNA 15*, 1274-1281.

9. Zaher, H.S., and Unrau, P. J. (2007) Selection of an improved RNA polymerase ribozyme with superior extension and fidelity, *RNA 12*, 1017-1026.

Chapter 5

Future directions

5.1 Introduction

This thesis sought to create new, and improve existing catalytic RNAs that could contribute to our understanding of an RNA world. The first is a triphosphorylation ribozyme, an RNA that uses trimetaphosphate to triphosphorylate the 5’ hydroxyl groups of RNAs. Triphosphorylation ribozymes could contribute to an RNA world by activating nucleic acids in preparation for ligation or polymerization. The second is a polymerase ribozyme, an RNA that catalyzes RNA polymerization. Although a polymerase ribozyme has already been developed, it is not capable of self replication (1, 2). We attempted to improve the polymerase ribozyme, first by adding positively charged amino acid cofactors and then by performing an in vitro selection in the presence of these cofactors.

5.2 Triphosphorylation ribozyme

An in vitro selection for ribozymes that use trimetaphsophate to triphosphorylate RNA 5’ hydroxyl groups was established and used to identify multiple sequences with this activity. One of the more promising sequences was truncated from 196 nucleotides to 96 nucleotides without loss of activity. It was capable of triphosphorylating its own 5’ end, as well as the 5’ end of short RNA

substrates. At optimum reaction conditions (100mM trimetaphosphate, 400mM $MgCl_2$, and pH 8.1) the reaction rate was 0.16 min^{-1}.

Catalytic RNAs that can triphosphorylate nucleosides could have played a role in activating nucleic acid monomers in the RNA world. A recently established synthesis of nucleosides from small molecules has brought us much closer to a prebiotically plausible synthesis of RNA monomers (3, 4). Ribozymes capable of chemically activating these nucleosides could have provided activated monomers for RNA polymerization. Although eight different triphosphorylation ribozymes were screened for the ability to triphosphorylate nucleosides, no activity was detected (Fig. S2.5). Since twenty-three active sequences were identified during the selection from testing thirty-six clones, more sequences could be screened for this activity (Fig. 2.2). Also, determining the tertiary structure of identified triphosphorylation ribozymes could provide more information about the active site and generate ideas about how the ribozymes could be modified to enable nucleoside triphosphorylation.

However, activating the 5’ hydroxyl groups of short RNAs could also be useful for the RNA world. It is possible that RNA polymerization occurred in the 3’-5’ direction in an earlier form of life. In this scenario, the RNA being extended would be activated with a triphosphate on its 5’ hydroxyl group and react with the 3’ hydroxyl groups of unactivated nucleosides. Polymerization in this direction has an advantage—having the negatively charged triphosphate on the RNA primer instead of the incoming monomer would reduce charge repulsion between the template and monomer. This could improve polymerization efficiency.

Preliminary investigations of the triphosphorylation ribozyme's secondary structure suggested that it forms a four way helical junction motif (Fig 2.6). By partially randomizing the ribozyme sequence of this ribozyme and performing more rounds of selection, the secondary structure could be elucidated by base covariation analysis. This experiment could also generate the more active variants of the starting triphosphorylation ribozyme. In the initial selection, about 10^{14} sequences were screened. Since 150 nucleotides were randomized in the selection, there are 4^{150} (~10^{90}) possible sequences and only a tiny fraction are sampled. By taking sequences that already show activity and partially randomizing them, the most active variants can be found (5). Selections starting with partial randomizations of two of the most active triphosphorylation ribozymes identified in the selection are already underway.

During the selection for the triphosphorylation ribozyme, the first evidence of activity in the pool (decreased number of PCR cycles required to amplify the pool after reverse transcription of the selected RNAs, see Fig. S2.2) was observed in the fourth round of selection. At this point, mutagenic PCR was used to introduce variation into the sequences to find the most active variants. Also, the selection was split into two branches: one with higher selection pressure (five minute reaction time) and one with the same selection pressure as in previous rounds (three hour reaction time). Clones were sequenced after round five and eight from both high and low selection pressure branches. When the clones were screened for activity, many clones were active, but there was no observable correlation between clone activity and round of selection or selection pressure

(Fig. 2.2). Indeed one of the most active clones, the one chosen for further analysis, came from round five and low selection pressure. There is no evidence that increased selection pressure, additional rounds of selection, and mutagenic PCR generated more active sequences in this selection. The mutagenic PCR conditions were adopted from a previously established protocol that generated about one mutation per 100 nucleotides (6, data not shown). It is possible that this was not enough mutations for ribozymes to explore enough alternative sequences. If mutagenic PCR conditions were varied to introduce more mutations on average per sequence, the sequence space of each variant could be better explored and it may result in better triphosphorylation ribozymes.

5.3 Polymerase ribozyme

An RNA world organism would require a way to replicate its genome. A catalytic RNA capable of self-replication could fill this role. Although a ribozyme capable of catalyzing RNA polymerization has been developed, it is not capable of self-replication and it is limited by weak substrate binding (1, 2). Since the ribozyme and its substrate are both negatively charged RNA molecules, reducing charge repulsion could improve substrate binding. The ribozyme can form base pairs with the single stranded portion of its template to improve polymerization efficiency. However, this severely limits how far the primer can be extended and which sequences can be extended. For self replication, the ribozyme must be processive and bind general template sequences.

We attempted to address this issue by annealing RNA-arginine conjugates to the ribozyme. The RNA acted as a handle, annealing the polymerase ribozyme so that the positively charged amino acid cofactor could interact with the substrate. However, this did not result in improved ribozyme polymerization (Fig. 3.2 and 3.3). It is possible that the arginine wasn't positioned correctly in active site, since the current polymerase ribozyme was not selected in the presence of RNA-amino acid conjugates. To address this, we attempted to perform an in vitro selection for a polymerase ribozyme with RNA-amino acid conjugates present. In addition to potentially improving the polymerase ribozyme, RNA-amino acid could have been precursors of modern day tRNAs. If short RNA-amino acid conjugates could improve catalytic RNA function in an RNA world, they could later evolve into part of today's translation system.

Because we were interested in finding polymerase ribozymes with improved substrate binding, we used an in vitro selection method that would allow the ribozyme and its substrate to not be covalently attached during the selection. This method uses compartmentalization of ribozyme and substrate in water in oil emulsion droplets. However, we were not able to get active sequences from the selection.

Several parts of the technically challenging in vitro selection in emulsion were established (Fig. 4.5, 4.6, and 4.8). In the future, the selection method should be tested with a pool of active sequences diluted with inactive sequences to establish that the selection is working and to determine the enrichment factor (and hence the expected number of rounds of selection needed to saturate the

pool with active sequences). The selection method could be improved by developing a capture method that relies on covalent modifications of the ribozyme's substrate (see section 4.3 for discussion).

Although technically challenging, an in vitro selection that allows for processivity during the reaction could yield improved polymerase ribozymes. There have been two in vitro selections for polymerase ribozymes that utilize this method (7, 8). Both selections found variants of the polymerase ribozyme that improved polymerization, but not enough for self-replication. However, in these selections the initial pool was based on the original polymerase ribozyme. The starting pools had the original polymerase ribozyme mutagenized with short randomized domains added to the 5' end of the ribozyme. The original polymerase ribozyme was developed by attaching a randomized domain to a ligase ribozyme and performing an in vitro selection. This selection yielded the first polymerase ribozyme and later a second selection from the same starting pool found multiple, diverse domains capable of catalyzing RNA polymerization (1, 9). An in vitro selection for polymerase ribozymes that provides selection pressure for substrate binding in trans could be performed with a completely randomized accessory domain. This could yield novel polymerase ribozymes with improved substrate binding that may be able to efficiently polymerize general substrate sequences.

5.4 Acknowledgments

Thanks to Greg Dolan who is working on doped selections of triphosphorylation ribozymes and Karen Olson and Zhaleh Amini for establishing the mutagenic PCR protocol in our lab and determining the average number of mutations per sequence.

5.5 References

1. Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E., and Bartel, D. P. (2001) RNA-Catalyzed RNA polymerization: accurate and general RNA-templated primer extension, *Science 292*, 1319-1325.

2. Lawrence, M. S., and Bartel, D. P. (2003) Processivity of ribozyme-catalyzed RNA polymerization, *Biochemistry 42*, 8748-8755.

3. Powner, M. W., Gerland, B., and Sutherland, J. D. (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature 459*, 239-242.

4. Powner, M. W., Sutherland, J. D., and Szostak, J. W. (2010) Chemoselective multicomponent one-pot assembly of purine precursors in water. *Journal of the American Chemical Society 132*, 16677-16688.

5. Bartel D.P., Zapp,M.L., Green,M.R. and Szostak,J.W. (1991) HIV-1 Rev regulation involves recognition of non-Watson–Crick base pairs in viral RNA. *Cell 67*, 529–536.

6. Cadwell, R.C. and Joyce, G.F. (1992) Randomization of genes by PCR mutagenesis. *PCR Methods Appl*, 2, 28-33.

7. Zaher, H.S., and Unrau, P. J. (2007) Selection of an improved RNA polymerase ribozyme with superior extension and fidelity, *RNA 12*, 1017-1026.

8. Wochner, A., Attwater, J., Coulson, A., and Holliger, P. (2011) Ribozyme-catalyzed transcription of an active ribozyme, *Science 332*, 209-212.

9. Lawrence, M. S., and Bartel, D. P. (2005) New ligase-derived RNA polymerase ribozymes, *RNA 11*, 1173-1180.