UNIVERSITY OF CALIFORNIA SAN DIEGO

Influence of Peptide and Lanthanide Cofactors on Ribozymes and on the Origin of Life

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

in

Chemistry

by

Kevin Josef Sweeney

Committee in charge:

Professor Ulrich Müller, Chair
Professor Thomas Hermann
Professor Simpson Joseph
Professor Jens Lykke-Andersen
Professor Wei Wang

2021

The dissertation of Kevin Josef Sweeney is approved,
and it is acceptable in quality and form
for publication on microfilm and electronically.

University of California San Diego

2021

EPIGRAPH

In the beginner's mind there are many possibilities,
but in the expert's there are few.

Shunryu Suzuki

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Tmp	Trimetaphosphate, a cyclic triphosphate species.
SHAPE	Selective 2’-Hydroxyl Acylation analyzed by Primer Extension
Tf	Trifluoromethanesulfonate
HTS	High-Throughput Sequencing
PTC	the Peptidyl-Transferase Center of the ribosome

LIST OF FIGURES AND TABLES

ACKNOWLEDGEMENTS

Chapter 2, in full, is currently being prepared for submission for publication of the material. Sweeney, Kevin J.; Jorge, Micaella Z.; Schellinger, Joan; Muller, Ulrich F. The dissertation author was the primary investigator of this material.

Chapter 3, in full, is currently being prepared for submission for publication of the material. Sweeney, Kevin J.; Muller, Ulrich F. The dissertation author was the primary investigator of this material.

VITA

2015	Bachelor of Science, University of California, Santa Cruz
2015-2016	Teaching Assistant, University of California San Diego
2016-2020	Research Assistant, University of California San Diego
2017	Master of Science, University of California San Diego
2020-2021	Teaching Assistant, University of California San Diego
2021	Doctor of Philosophy, University of California San Diego

FIELDS OF STUDY

Major Field: Chemistry

Studies in RNA Biochemistry, Origin of Life

Professor Ulrich Müller

ABSTRACT OF THE DISSERTATION

Influence of Peptide and Lanthanide Cofactors
on Ribozymes and on the Origin of Life

by

Kevin Josef Sweeney

Doctor of Philosophy in Chemistry

University of California San Diego, 2021

Professor Ulrich Müller, Chair

Throughout human history, the nature of life's origin has been a prominent question seen with much philosophical importance. Our best scientific model for the origin of life is the RNA world hypothesis, which describes the earliest forms of life as having RNA as the only genome and the only genome-encoded catalysts, as opposed to the DNA-RNA-protein world of modern biology. Research into geology and prebiotic chemistry, however, describe a complex chemical environment in which the ribozymes would have had many catalytic possibilities to explore. In this dissertation, I present the results of two in vitro selections which examined the utility of

peptides and lanthanides as cofactors for triphosphorylation ribozymes. These ribozymes utilize trimetaphosphate, a prebiotically plausible polyphosphate source, to triphosphorylate their own 5'-ends as a simpler analog for studying nucleoside activation. The first selection was done in the presence of ten octapeptides of varying prebiotic plausibility, and two ribozymes were examined in detail. Ribozyme 20 displayed a strongly beneficial relationship with one specific peptide, which increased the RNA's activity by 900 (+/- 300)-fold, providing an example of a ribozyme that could benefit from a primitive translation system. In contrast, the mild (2-10 fold) benefit for ribozyme 23 interacting with six of the ten peptides provides an example of a ribozyme that could possibly benefit from random, abiotically assembled peptides. The other selection, using ytterbium as a metal ion cofactor instead of the standard magnesium, yielded many active sequences. One sequence, ribozyme 51, was studied in detail and was very sensitive to the ionic radius of the lanthanide. It required a lanthanide for activity and seemed to only be able to bind trimetaphosphate as a complex with the lanthanide, but still seemed to prefer other metal cations for folding. These results provide insight into the relationships of peptides and lanthanides with the development of catalytic RNAs, as well as opening doors for future study for understanding both the origin of life and the general principles of their interactions.

Chapter 1: Introduction to the RNA World

The RNA World

The question of how life originated is one of the most fundamental, yet persistent in all of humankind. Every culture, from ancient to modern and in all parts of the world, has an origin myth that explains how we as humans, and our world, came to be. In my doctoral research, I hoped to advance our modern scientific understanding of the origin of life, both to better understand the history of life on Earth and to help the search for life in other parts of the universe. In order to do so, it is first necessary to define life.

The definition of life I am using here includes the ability to self-replicate, a process by which living systems are able to assemble replicate by assembling non-living components. This does not necessarily mean that a single molecule is self-replicating, but for example in modern biology, the DNA-RNA-protein system of individual organisms is self-replicating. The second criterion is that the replication must be imperfect, meaning that genetic information is transmitted with fidelity, but including some errors. If both of these criteria are met, open-ended Darwinian evolution is a natural and unavoidable consequence, and both are necessary for the simplest living systems to evolve into multicellular, intelligent life such as humans[1].

In the wake of Darwin's seminal book, much research was conducted on the evolution of various organisms, and later how genetic information is stored and copied. However, direct chemical study of the origin of life (the transition from abiotic chemistry to living systems) began with Stanley Miller's research[2]. Miller's work, where simple compounds (CH_4, NH_3, H_2O, and H_2 in the original experiment) were mixed as gases and exposed to an electrical discharge before being condensed and collected, showed that amino acids can be synthesized without biological influence under conditions similar to those under which life may have originated.

The work of Stanley Miller coincided with the growing understanding of what is now called the "central dogma" of molecular biology, which is that in today's biology, DNA carries genetic information and is transcribed into RNA, which is then translated into proteins, which catalyze most of the different reactions happening in living cells. The molecular understanding of how information is encoded and used to create catalysts in biology lead to much discussion about how these systems could have originated and developed. Notably, in the 1960s, several prominent biochemists began to think that nucleic acids could have been the primary basis for life's earliest forms rather than an interdependent system of DNAs and proteins, and narrowed this down to RNA specifically[3,4,5,6]. This idea gained support when the catalytic potential of RNA was confirmed by later studies by Sidney Altman and Thomas Cech, for which they split the 1989 Nobel Prize in Chemistry[7,8]. About a decade later, it was shown that the catalytic center of the ribosome, which synthesizes all genetically-encoded proteins, consists entirely of RNA with no peptides within 18 angstroms of the site of peptide bond synthesis[9,10]. This means that ribozyme is responsible for synthesizing genetically-encoded proteins, and therefore most self-replication machinery, despite the greater versatility of proteins as catalysts due to the more chemically diverse amino acids. Combined with the fact that the ribosome is translating an mRNA sequence and the amino acids are delivered to the ribosome by tRNAs, this provides more support for an RNA-first model, since a protein-based translation system would likely be more efficient and therefore would not have evolved into a less efficient RNA-based system. Additionally, many proteins utilize nucleotide-based cofactors even though there is no need to do so, which may be relics from a time when those reactions were catalyzed by RNAs whose role was eventually taken over by proteins[11]. All of these discoveries provide support for the idea that the origin of life began with RNA serving both as the genome and the only genome-encoded catalyst, which is now often called the "RNA World" hypothesis after the term was coined by Walter Gilbert in 1986[12].

In discussing how an RNA world may have begun, it is necessary to understand the chemical environment on Earth at the time. Isotopic analysis of lead deposits, among other sources, allows us to confidently estimate the Earth's age at approximately 4.54 billion years[13]. Life as we imagine it could not

have existed on Earth at this time, however, as the planet would have been covered in magma at the time and the moon forming impact happened 4.5 billion years ago, sterilizing the earth. However, analysis of Australian Zircons indicates that the Earth may have at least started cooling enough to have liquid water on the surface as long as 4.4 billion years ago[14]. This offers a maximum age for life on Earth, as life here as we know it requires the presence of liquid water.

The earliest chemical systems that could be called alive are unlikely to have been preserved in geological record, especially if not contained within membranous barriers, as the separation between the organism and its environment would not have been substantial enough to form recognizable sediments surviving billions of years in the Earth's crust. There is substantial evidence, however, that cellular life has existed on Earth for at least 3.4 billion years. In Shark Bay, along the western coast of Australia, fossils have been identified that were left behind by microbial mats, or layers of single-celled autotrophic organisms[15]. There are some formations that may indicate cellular life at least 3.77 billion years ago, but it is possible that these formations may have been abiotic in origin[16,17]. This means we can confidently say that life on Earth originated between 3.4 and 4.4 billion years ago, and studies of rock formations of this age have yielded important information about the chemical environment at the time and how an RNA world could have developed.

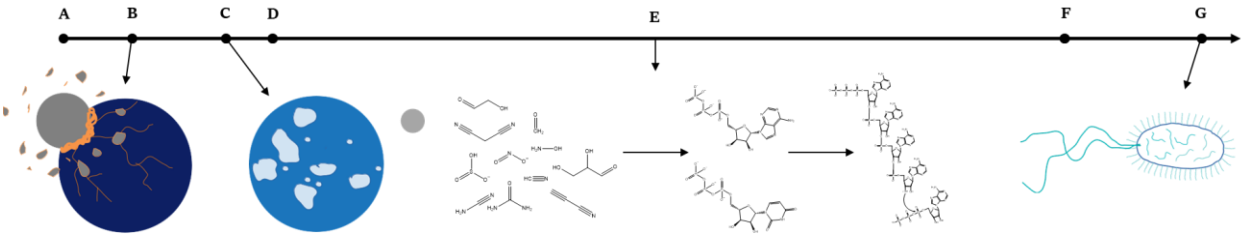

Figure 1.1: Timeline for the origin of life. **(A)** Approximate age of our solar system = 4.56 Ga[13]. **(B)** Moon-forming impact occurs on the still cooling Earth at approximately 4.51 Ga, sterilizing the Earth and re-mixing the mantle[13]. **(C)** Liquid water present in enough abundance to leave evidence in zircons at approximately 4.4 Ga[14]. **(D)** Possible second large impact occurs, delivering iron and creating a reducing atmosphere at approximately 4.36 Ga[21]. **(E)** Time period for a developing RNA world. Prebiotic nucleotide synthesis from Becker and coworkers (2019)[19]. **(F)** Microfossils indicate possible evidence of cellular life at 3.77Ga[16]. **(G)** Cellular life existed at 3.5 Ga, supported by fossilized evidence of microbial mats[15].

Peptides at the Origin of Life

It has been suggested by some that the RNA world would have originated on its own and only created or utilized DNA and proteins later in development. There are issues with this hypothesis though, because in contrast to amino acids, nucleosides have not been found in experiments like those by Stanley Miller. Recent publications, especially from the labs of John Sutherland and Thomas Carell, have made great strides in demonstrating increasingly prebiotically plausible syntheses for nucleosides / nucleotides[19,20,20]. Some precursors for these syntheses would require a reducing atmosphere to be present in abundance, which may have been created by a large impact on Earth by an iron-laden meteorite about the size of the moon[21]. Even so, given the ease with which amino acids form under prebiotic Earth simulation experiments compared to nucleic acids, it seems almost certain that amino acids would have been present during an RNA world scenario.

Not only did Stanley Miller's original experiment show synthesis of the amino acids aspartic acid, glycine, alpha-amino-butyric acid, alanine, and beta-alanine, but later analysis of samples collected from a similar apparatus showed that many more amino acids and similar compounds could be synthesized under such conditions, including more than half of those used by modern organisms[22]. In addition, it seems likely that short peptides would also have been present. In a 2015 study, Forsythe and coworkers demonstrated that the inclusion of alpha-hydroxy acids along with amino acids going through wet/dry cycles can increase both the rate of formation and the length of the products (mixed depsipeptides) formed[23]. This occurs because the alpha-hydroxy acids are more likely to form ester bonds during the dry phase than the amino acids because the alcohol is a better nucleophile, while the amino acids are also more likely to replace an alpha-hydroxy acid than react with a carboxy-terminus because the alpha-hydroxy acid is a better leaving group than a hydroxide ion. After successive cycles, the length and amino acid content of the depsipeptides increase, producing detectable depsipeptides up to 10 units in length after four cycles. This demonstration of prebiotically plausible peptide synthesis further supports the idea that an RNA world would have occurred in the presence of peptides. Combined with the ancient and ubiquitous nature of certain modern RNA-protein interactions, such as the ribosome, it's possible that

peptides were assisting catalytic RNAs from the beginning. With this in mind, the goal of my main thesis project was to investigate if the presence of short peptides could aid the emergence of catalytic RNAs.

Table 1.1: Table of amino acids comparing their abundance seen in prebiotic simulation experiments and their likelihood to interact with nucleic acids. The amino acids seen in Miller-type experiments are listed in order of abundance based on Parker and coworkers (2011)[22], with the exception of proline. Proline and all amino acids not seen in Miller-type experiments are listed in order of likely development according to Trifonov (2004)[33]. Protein-RNA interactions are taken from the Amino Acid Nucleotide Interaction database[34], and refer to the number of predicted interactions between an amino acid in a protein and an RNA strand, coming from a total of 930 solved Protein-RNA complex structures.

Amino Acids	Seen in Miller-type experiments?	Modern RNA-protein interactions
Glycine	Yes	458
Alanine	Yes	147
Valine	Yes	144
Proline	Yes	92
Serine	Yes	879
Threonine	Yes	753
Aspartic Acid	Yes	361
Glutamic Acid	Yes	446
Leucine	Yes	118
Isoleucine	Yes	95
Methionine	Yes	70
Arginine		2716
Asparagine		657
Glutamine		501
Histidine		269
Lysine		1813
Cysteine		46
Phenylalanine		64
Tyrosine		444
Tryptophan		73

In vitro Selection of Triphosphorylation Ribozymes

In the years since the articulation of the RNA world hypothesis, much work has been done using selection and evolution techniques to identify RNA sequences capable of catalyzing reactions that would be necessary for an RNA world organism, with great success[24]. One reaction critically necessary for such an organism would be RNA polymerization. While no fully self-replicating ribozyme has yet been

discovered, David Bartel's lab produced the first RNA polymerase ribozyme, an important step towards self-replicating RNA[26,26]. Another critical reaction would be chemical activation of nucleosides so that they would be usable for polymerization. My work centers around a model system designed to approximate this reaction.

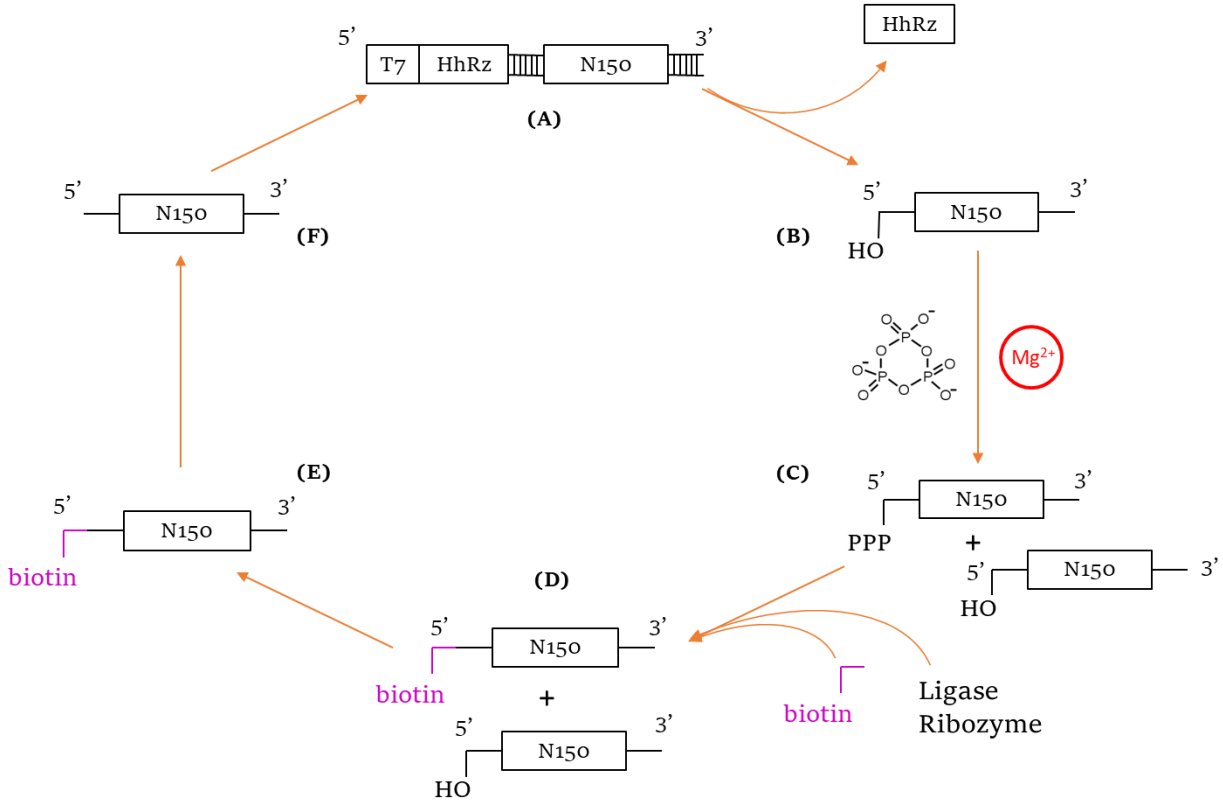

Figure 1.2: In vitro selection scheme established by Moretti and Muller in 2014 which is the basis for both projects discussed in this dissertation. **(A)** Double-stranded pool DNA, which contains a T7 promoter, a hammerhead ribozyme (HhRz), and a 150 nucleotide randomized region (N150) flanked by 5' and 3' constant regions. The DNA is transcribed by T7 polymerase and the hammerhead ribozyme cleaves itself off cotranscriptionally, leaving behind **(B)** the pool RNA molecules with 5' hydroxyl groups. The RNAs consist of the randomized region and the constant regions. They are then incubated with trimetaphosphate and Mg^{2+} in a buffered solution, and **(C)** some RNAs will be able to coordinate trimetaphosphate and perform a nucleophilic attack with the 5'-hydroxyl group. The whole pool is then incubated with a ligase ribozyme[29] which will covalently connect 5'-triphosphorylated RNAs with a biotinylated oligonucleotide to produce **(D)** biotinylated active pool RNA molecules. Streptavidin-coated magnetic beads are then used to **(E)** separate the active, biotinylated sequences from the inactive ones by stringent washing. The pool is then **(F)** reverse transcribed into cDNA, which is then PCR amplified to regenerate a double-stranded DNA pool that is now enriched for catalytic sequences. Adapted from Moretti and Muller (2014)[27].

In order to examine how peptides may have affected the development of RNA world ribozymes, we wanted to use a model system that allows sampling a large RNA sequence space while requiring ribozyme activity similar to a reaction that would be important to an RNA world. Previously, our lab

developed an in vitro selection system that began with more than 10^{14} different sequences and found ~300 sequence families of ribozymes that triphosphorylate their own 5'-ends[28,28] (Figure 1.2). The triphosphorylation reaction is critical because a 5'-triphosphate is still necessary for polymerization in modern biology, and is likely to have been an important activating process in an RNA world. The ribozymes activate themselves instead of free nucleosides in this selection procedure because this makes it much simpler to isolate the active sequences from the inactive ones. If activating free nucleosides, the nucleoside triphosphate could then immediately diffuse away, leaving no way to know which RNA activated it. The phosphorylating molecule used by the ribozymes is trimetaphosphate, which is a cyclic triphosphate molecule that could be produced under the conditions of the early Earth[30]. The cyclic structure makes the phosphorous atoms more susceptible to nucleophilic attack from RNA than a linear triphosphate, making it a better phosphorylating agent[31]. It has been demonstrated that free adenosine can react freely with trimetaphosphate to produce a variety of phosphorylated species, but the reaction required a pH of 12 to yield detectable adenosine 5'-triphosphate[32]. My project centered on using the triphosphorylation ribozyme selection system to study the effects of peptides on the ribozymes that could be identified by the selection.

In experiments designed to simulate prebiotic amino acid synthesis, some of the 20 modern encoded amino acids are produced more abundantly than others, and some have not been seen at all. Generally, the negatively charged or small, uncharged amino acids are often seen as products in these experiments, while the bulkier, aromatic, and positively charged amino acids are not[33]. This is in contrast with the amino acids that are seen directly interacting with RNA in modern biology; positively charged aromatic amino acids are most often seen directly interacting with RNA, with arginine at the top of the list[34,35]. The residues involved are not surprising, given that so many RNA-protein interfaces involve ionic interactions with the phosphate backbone or aromatic stacking with the nucleotide bases[36]. This provides a challenge for designing peptides to investigate these interactions in an RNA world context, as the peptides that seem most likely to have a strong influence on the selection will also be the least likely to exist in abundance in a system before biological amino acid synthesis. In my project, we covered the

range of prebiotic plausibility and likelihood of interacting with RNA by using a mix of 10 peptides, each at a length of 8 amino acids to stay at a plausible length based on the 2015 study from Forsythe and coworkers[23]. All 20 modern encoded amino acids are present, with some peptides being composed entirely of abundant products from Stanley Miller's experiments, while others include multiple positively charged and aromatic amino acids.

Lanthanides as Ribozyme Cofactors

Multivalent cations are extremely important for RNA function, with magnesium being the most common and best-studied example. Multivalent cations can perform multiple important roles, including assisting in folding by interacting with the phosphate backbone, facilitating catalysis, and helping bridge interactions with other molecules (especially negatively charged ones)[37]. Recently, several studies investigated whether Fe^{2+}, which likely existed at millimolar concentrations in the prebiotic ocean due to the absence of oxygen and the much better solubility of Fe^{2+} as compared to Fe^{3+}, could have been an important ribozyme cofactor[38,39]. The goal of my second thesis project was to examine the possibility of RNA using lanthanides as cofactors, hoping to learn about the potential for RNA to utilize previously untested ions for catalysis.

Protein enzymes have recently been discovered that use lanthanides as cofactors[40,41,42]. These bacterial enzymes are involved in metabolism of small organic compounds, especially methanol, and are closely related to similar enzymes that use calcium instead of lanthanides. Lanthanides can be very useful for catalysis due to their strong Lewis acid character, which may be why at least one bacteria preferentially uses the lanthanide-dependent enzyme when both are available[43]. It was also an indication that they may be able to participate in the same triphosphorylation reaction used in my other project. A more direct indication came from Huskens and Peters in 1995, who demonstrated that lanthanides could hydrolyze trimetaphosphate, especially when coordinated by EDTA or NTA, organic molecules that chelate multivalent cations[44].

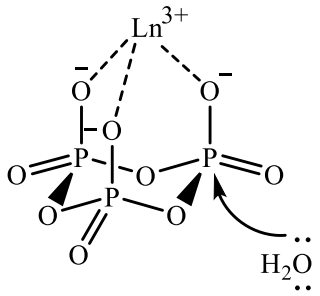

Figure 1.3: A lanthanide ion coordinates the oxygen atoms of trimetaphosphate. This activates the Tmp for hydrolysis via nucleophilic attack of a water at a phosphorus atom[44]. Additional coordination of the lanthanide by chelators can strongly increase the lanthanide's catalysis of the reaction.

Additionally, despite comprising most of the elements referred to as "Rare Earth Elements," they are not among the least abundant elements in the Earth's crust. Lanthanides are on average ~6 orders of magnitude less abundant than the most common elements, Oxygen and Silicon, but ~5 orders of magnitude more abundant than the least common, Iridium and Osmium[45]. Beyond that, their presence in the Earth's crust is not uniform; they are unable to integrate into the most common minerals where other metals can, and thus are pushed out to higher local concentrations where they can form different minerals. Lanthanides are often seen concentrated (mixed with each other) in pegmatite minerals that would have been among the last to crystallize from magma during the cooling of the Earth[46]. Given their potential to catalyze trimetaphosphate hydrolysis and their not insignificant abundance in the Earth's crust, we decided to use our lab's triphosphorylation ribozyme selection system to identify if ribozymes could use a lanthanide as a cationic cofactor for this reaction. In this selection process, after transcription and gel purification, the RNA pool was incubated with trimetaphosphate, ytterbium trifluoromethanesulfonate, and sodium chloride in a buffered solution, meaning that ytterbium would be the only multivalent cation added for the ribozymes to use for the reaction.

Summary

Due to the unique usage of RNAs in modern biology as both information carriers and catalysts, in addition to the ribosome's status as a ribozyme and the common usage of nucleotide-based cofactors by

protein enzymes, it seems likely that all life on Earth today descended from living systems that utilized RNAs the only genome and the only genome-encoded catalysts. However, our understanding of the chemical environment and possible synthetic pathways at the time when an RNA world would have existed indicates that other polymers, including peptides, would have been present. Combined with the critical role of protein-RNA interactions in modern biology, including in the ribosome, this suggests that peptides could have assisted in the evolution of those RNA world organisms. Additionally, many other cofactors may have been important to these earliest forms of life, and thus it is useful to explore the chemical space of RNA catalysis using our selection system to see how they can influence ribozyme development. In the following chapters, I will present my research investigating cofactors for in vitro-selected triphosphorylation ribozymes and discuss the implications of my research on our current understanding of how life may have originated on Earth.

References

1. Darwin, Charles. "On the origin of species" (1859).
2. Miller, Stanley L. "A production of amino acids under possible primitive earth conditions." *Science* 117.3046 (1953): 528-529.
3. Rich, Alexander. "On the problems of evolution and biochemical information transfer." *Horizons in biochemistry* (1962): 103-126.
4. Woese, Carl R. "The fundamental nature of the genetic code: prebiotic interactions between polynucleotides and polyamino acids or their derivatives." *Proceedings of the National Academy of Sciences of the United States of America* 59.1 (1968): 110.
5. Orgel, Leslie E. "Evolution of the genetic apparatus." *Journal of molecular biology* 38.3 (1968): 381-393.
6. Crick, Francis HC. "The origin of the genetic code." *Journal of molecular biology* 38.3 (1968): 367-379.
7. Stark, Benjamin C., Kole, Ryszard, Bowman, Emma J., and Sidney Altman. "Ribonuclease P: an enzyme with an essential RNA component." *Proceedings of the National Academy of Sciences* 75.8 (1978): 3717-3721.
8. Kruger, Kelly, Grabowski, Paula J., Zaug, Arthur J., Sands, Julie, Gottschling, Daniel E., and Thomas R. Cech. "Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena." *Cell* 31.1 (1982): 147-157.
9. Nissen, Poul, Hansen, Jeffrey, Ban, Nenad, Moore, Peter B., and Thomas A. Steitz. "The structural basis of ribosome activity in peptide bond synthesis." *Science* 289.5481 (2000): 920-930.

10. Ban, Nenad, Nissen, Poul, Hansen, Jeffrey, Moore, Peter B., and Thomas A. Steitz. "The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution." *Science* 289.5481 (2000): 905-920.

11. White, Harold B. "Coenzymes as fossils of an earlier metabolic state." *Journal of molecular evolution* 7.2 (1976): 101-104.

12. Gilbert, Walter. "Origin of life: The RNA world." *Nature* 319.6055 (1986): 618-618.

13. Dalrymple, G. Brent. "The age of the Earth in the twentieth century: a problem (mostly) solved." *Geological Society*, London, Special Publications 190.1 (2001): 205-221.

14. Hoskin, Paul WO. "Trace-element composition of hydrothermal zircon and the alteration of Hadean zircon from the Jack Hills, Australia." *Geochimica et cosmochimica acta* 69.3 (2005): 637-648.

15. Duda, Jan-Peter, Van Kranendonk, Martin J., Thiel, Volker, Ionescu, Danny, Strauss, Harald, Schäfer, Nadine, and Joachim Reitner. "A rare glimpse of Paleoarchean life: Geobiology of an exceptionally preserved microbial mat facies from the 3.4 Ga Strelley Pool Formation, Western Australia." *PloS one* 11.1 (2016): e0147629.

16. Dodd, Matthew S., Papineau, Dominic, Greene, Tor, Slack, John F., Rittner, Martin, Pirajno, Franco, O'Neil, Johnathan, and Crispin T. S. Little. "Evidence for early life in Earth's oldest hydrothermal vent precipitates." *Nature* 543.7643 (2017): 60-64.

17. Brasier, Martin D., Owen R. Green, and Nicola McLoughlin. "Characterization and critical testing of potential microfossils from the early Earth: the Apex 'microfossil debate'and its lessons for Mars sample return." *International Journal of Astrobiology* 3.2 (2004): 139-150.

18. Powner, Matthew W., Béatrice Gerland, and John D. Sutherland. "Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions." *Nature* 459.7244 (2009): 239-242.

19. Becker, Sidney, Feldmann, Jonas, Wiedermann, Stefan, Okamura, Hidenori, Schneider, Christina, Iwan, Katharina, Crisp, Antony, Rossa, Martin, Amatov, Tynchtyk, and Thomas Carell. "Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides." *Science* 366.6461 (2019): 76-82.

20. Okamura, Hidenori, Becker, Sidney, Tiede, Niklas, Wiedermann, Stefan, Feldmann, Jonas, and Thomas Carell. "A one-pot, water compatible synthesis of pyrimidine nucleobases under plausible prebiotic conditions." *Chemical Communications* 55.13 (2019): 1939-1942.

21. Benner, Steven A., Bell, Elizabeth A., Biondi, Elisa, Brasser, Ramon, Carell, Thomas, Kim, Hyo-Joong, Mojzsis, Stephen J., Omran, Arthur, Pasek, Matthew A., and Dustin Trail. "When did life likely emerge on Earth in an RNA-first process?." *arXiv preprint* arXiv:1908.11327 (2019).

22. Parker, Eric T., Cleaves, Henderson J., Dworkin, Jason P., Glavin, Daniel P., Callahan, Michael, Aubrey, Andrew, Lazcano, Antonio, and Jeffrey L. Bada. "Primordial synthesis of amines and amino acids in a 1958 Miller H2S-rich spark discharge experiment." *Proceedings of the National Academy of Sciences* 108.14 (2011): 5526-5531.

23. Forsythe, Jay G., Yu, Sheng-Sheng, Mamajanov, Irena, Grover, Martha A., Krishnamurthy, Ramanarayanan, Fernández, Facundo M., and Nicholas V. Hud. "Ester-mediated amide bond formation driven by wet–dry cycles: A possible path to polypeptides on the prebiotic Earth." *Angewandte Chemie International Edition* 54.34 (2015): 9871-9875.

24. Chen, Xi, Na Li, and Andrew D. Ellington. "Ribozyme catalysis of metabolism in the RNA world." *Chemistry & biodiversity* 4.4 (2007): 633-655.

25. Ekland, Eric H., and David P. Bartel. "RNA-catalysed RNA polymerization using nucleoside triphosphates." *Nature* 382.6589 (1996): 373-376.

26. Johnston, Wendy K., Unrau, Peter J., Lawrence, Michael S., Glasner, Margaret E., and David P. Bartel. "RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension." *Science* 292.5520 (2001): 1319-1325.

27. Moretti, Janina E., and Ulrich F. Müller. "A ribozyme that triphosphorylates RNA 5′-hydroxyl groups." *Nucleic acids research* 42.7 (2014): 4767-4778.

28. Pressman, Abe, Moretti, Janina E., Campbell, Gregory W., Müller, Ulrich F., and Irene A. Chen. "Analysis of in vitro evolution reveals the underlying distribution of catalytic activity among random sequences." *Nucleic Acids Research* 45.14 (2017): 8167-8179.

29. Rogers, Jeff, and Gerald F. Joyce. "The effect of cytidine on the structure and function of an RNA ligase ribozyme." *RNA* 7.3 (2001): 395-404.

30. Pasek, Matthew A., Kee, Terrance P., Bryant, David E., Pavlov, Alexander A., and Jonathan I. Lunine. "Production of potentially prebiotic condensed phosphates by phosphorus redox chemistry." *Angewandte Chemie International Edition* 47.41 (2008): 7918-7920.

31. Feldmann, W. "Trimetaphosphate as a phosphorylation agent for alcohols and carbohydrates in aqueous solution. Its special position under the condensed phosphate." *Chem. Ber* 100 (1967): 3850-3860.

32. Etaix E, Orgel LE. "Phosphorylation of nucleosides in aqueous solution using trimetaphosphate: Formation of nucleoside triphosphates." *J Carbohydr Nucleosides Nucleotides* (1978): 5:91–110

33. Trifonov, Edward N. "The triplet code from first principles." *Journal of Biomolecular structure and dynamics* 22.1 (2004): 1-11.

34. Hoffman, Michael M., Khrapov, Maksim A., Cox, J. Colin, Yao, Jianchao, Tong, Lingnan, and Andrew D. Ellington. "AANT: The amino acid–nucleotide interaction database." *Nucleic Acids Research* 32.suppl_1 (2004): D174-D181.

35. Das, Chandreyee, and Alan D. Frankel. "Sequence and structure space of RNA-binding peptides." *Biopolymers: Original Research on Biomolecules* 70.1 (2003): 80-85.

36. Corley, Meredith, Margaret C. Burns, and Gene W. Yeo. "How RNA-binding proteins interact with RNA: molecules and mechanisms." *Molecular cell* 78.1 (2020): 9-29.

37. Bowman, Jessica C., Lenz, Timothy K., Hud, Nicholas V., and Loren Dean Williams. "Cations in charge: magnesium ions in RNA folding and catalysis." *Current opinion in structural biology* 22.3 (2012): 262.

38. Athavale, Shreyas S., Petrov, Anton S., Hsiao, Chiaolong, Watkins, Derrick, Prickett, Caitlin D., Gossett, J. Jared, Lie, Lively, Bowman, Jessica C., O'Neill, Eric, Bernier, Chad R., Hud, Nicholas V., Wartell, Roger M., Harvey, Stephen C., and Loren Dean Williams. "RNA folding and catalysis mediated by iron (II)." *PLoS One* 7.5 (2012): e38024.

39. Popović, Milena, Palmer S. Fliss, and Mark A. Ditzler. "In vitro evolution of distinct self-cleaving ribozymes in diverse environments." *Nucleic Acids Research* 43.14 (2015): 7070-7082.

40. Nakagawa, Tomoyuki, Mitsui, Ryoji, Tani, Akio, Sasa, Kentaro, Tashiro, Shinya, Iwama, Tomonori, Hayakawa, Takashi, and Keiichi Kawai. "A catalytic role of XoxF1 as La3+-dependent methanol dehydrogenase in Methylobacterium extorquens strain AM1." *PloS one* 7.11 (2012): e50480.

41. Keltjens, Jan T., Pol, Arjan, Reimann, Joachim, and Huub J.M. Op den Camp. "PQQ-dependent methanol dehydrogenases: rare-earth elements make a difference." *Applied microbiology and biotechnology* 98.14 (2014): 6163-6183.

42. Daumann, Lena J. "Essential and ubiquitous: the emergence of lanthanide metallobiochemistry." *Angewandte Chemie International Edition* 58.37 (2019): 12795-12802.

43. Chu, Frances, and Mary E. Lidstrom. "XoxF acts as the predominant methanol dehydrogenase in the type I methanotroph Methylomicrobium buryatense." *Journal of bacteriology* 198.8 (2016): 1317-1325.

44. Huskens, Jurriaan, Kennedy, Anna D., van Bekkum, Herman, and Joop A. Peters. "The Hydrolysis of Trimetaphosphate Catalyzed by Lanthanide (III) Aminopolycarboxylate Complexes: Coordination, Stability, and Reactivity of Intermediate Complexes." *Journal of the American Chemical Society* 117.1 (1995): 375-382.

45. Haxel, Gordon. Rare earth elements: critical resources for high technology. Vol. 87. No. 2. US Department of the Interior, US Geological Survey, 2002.

46. Sigel, Helmut, ed. Metal Ions in Biological Systems: Volume 40: The Lanthanides and Their Interrelations with Biosystems. CRC Press, 2003.

Chapter 2: Peptide Cofactors at the Origin of Life

Kevin J. Sweeney, Micaella Z. Jorge, Joan Schellinger, Ulrich F. Muller

Abstract

Early stages of life likely employed catalytic RNAs (ribozymes) in many functions that are today filled by proteins. However, the earliest life forms must have emerged from heterogenous chemical mixtures, which included amino acids and short peptides. Here we explored whether the presence of short peptides can help the emergence of catalytic RNAs. To do this, we conducted an in vitro selection for catalytic RNAs in the presence of ten different peptides with a length of eight amino acids. This length is consistent with experimental models for prebiotically plausible polymerization of amino acids. The in vitro selection generated dozens of ribozymes, several of them benefitting from the presence of peptides. Two ribozymes were characterized in more detail because they represent two different benefits from peptides: Ribozyme 20 showed at least 500-fold higher activity in the presence of one specific peptide but no benefit from the nine other peptides. The requirement for one specific peptide means that this ribozyme would have benefitted from the invention of encoded translation. In contrast, ribozyme 23 benefitted from most of the ten peptides, mostly so under suboptimal incubation conditions. This non-specific effect could have been important for RNA function before encoded translation emerged. Together, these results establish two model systems that allow us to study how peptides could have aided ribozymes during the emergence of life.

Introduction:

The RNA world hypothesis states that an early stage in the evolution of life used RNA as genome, and as the only genome-encoded catalyst[1,2,3,4,5]. However, the emergence of a self-replicating RNA system directly from a prebiotic environment would be unlikely if RNA oligomers containing only canonical nucleotides would have to stem directly from a chemically heterogeneous mixture[6]. The concepts of

chemical evolution and systems chemistry address this challenge: First, chemical evolution describes a process that shifts the spectrum of different chemical species over time, based on chemical characteristics such as stability, structure formation, as well as reactions with other compounds[7,8]. For example, canonical nucleosides seem to have been selected for high photochemical stability[9], and their N-glycosidic bond is hydrolytically more stable in canonical nucleosides than in alternative nucleosides[10]. Other features must have been selected on the level of duplex formation for advantages that are only apparent at a higher level of organization. For example, the pK_A and pK_B values at the Watson-Crick face of canonical nucleosides are just right for efficient base pairing[11], and the conformation of ribose is optimal for positioning of the bases in base pairing position[12]. This selection advantage could have been mediated in part by the higher hydrolytic stability of double stranded RNA as compared to single stranded RNA[13]. Second, systems chemistry studies the molecular interactions in a heterogeneous local environment, with a special focus on synergistic interactions[14]. For example, the presence of alpha-hydroxy acids benefits the oligomerization of amino acids under wet-dry cycling conditions[15], and pyrimidine deoxynucleotides benefit the amidophosphorylation of purine deoxynucleotides[16]. Finally, in a combination of chemical evolution and systems chemistry, pairs of compounds may be able to stabilize each other, and thereby enrich in the environment[17].

The mutual interactions of peptides and RNAs are promising to study regarding the origin of life because peptides likely existed in prebiotic environments that could give rise to RNAs: Amino acids naturally arise under conditions that generate nucleotides[18,19,20], and amino acids are able to form peptides under wet-dry cycling conditions[21], especially in the presence of alpha-hydroxy acids[22], which likely existed together with amino acids[23]. Prebiotic model reactions for peptide formation give decent yields up to about 8-mers[22]. Additionally, mutual RNA / peptide interactions could have been beneficial because the two polymers complement each other, as seen in today's RNA/protein complexes: RNA has an advantage for the sequence specific recognition of nucleic acids, and polypeptides can use their larger chemical diversity to establish more powerful catalytic sites. A central question is whether and how peptides could have aided the emergence of catalytic RNAs (ribozymes). Previous studies found that cationic proto-

peptides increase the thermodynamic stability of duplex RNA[24], and that coacervates consisting of polyanionic and polycationic peptides can modulate the secondary structure of tRNA[25]. The presence of an RNA-binding peptide can serve as switch for ribozyme activity[26], and cationic peptides can enhance the function of existing ribozymes[27,28]. However, to our knowledge, no study has analyzed how short peptides could aid the emergence of catalytic RNAs from a random sequence pool. In vitro selection procedures can be used as powerful model systems to explore ligands for potential benefits[29,30,31]. An in vitro selection of new ribozymes in the presence of peptides would address the question address the question if peptides can help the emergence of catalytic RNAs.

As model system for a prebiotically relevant in vitro selection, we focused on the reaction with the prebiotically plausible energy source cyclic trimetaphosphate (Tmp)[32,33,34,35]. The resulting ribozymes catalyze the nucleophilic attack of their 5'-hydroxyl group on Tmp, thereby generating an RNA 5'-triphosphate. 5'-triphosphates are used as universal energy currency in all known forms of life in the form of nucleoside 5'-triphosphates (NTPs). Therefore, catalysts that convert a prebiotically plausible energy source into a biologically useful energy source is important to models for origins of life. Importantly, the described in vitro selection generates a >1,000-fold enrichment factor per selection cycle, can identify dozens to hundreds of active sequences in a single experiment, and identified few or no inactive sequences[35,36,37,38]. Therefore, this in vitro selection procedure is an excellent model system to test whether peptides could benefit the emergence of ribozymes from random sequence.

Here we show the in vitro selection of self-triphosphorylating ribozymes in the presence of ten different octapeptides. After five rounds of selection the RNA library was dominated by active sequences. Two ribozymes that benefitted from peptides (ribozyme 20 and 23) were chosen for further analysis. Ribozyme 20 showed 900 (+/- 300)-fold higher activity with one specific peptide, but not with any of the other peptides. Ribozyme 23 benefitted from most of the ten peptides, especially under suboptimal reaction conditions. These examples establish two model systems to study how early life forms may have benefitted from peptides, before and after the invention of a translation apparatus.

Results:

To determine whether the presence of peptides can aid the emergence of catalytic RNAs we performed an in vitro selection of ribozymes in the presence of ten different octapeptides. The ten peptides were chosen to represent sets of amino acids that reflect different stages in the early evolution of life (Figure 2.1A): Peptides P1 to P6 are dominated by amino acids that are abundant in Miller-Type reactions[39] and were likely present in the earliest stages of life[40]. Several of these peptides explored the addition of arginine and lysine: Arginine precursors are generated in cyanosulfidic models for prebiotic reactions[18,19], and lysine has been identified in meteorites [41,42]. These two amino acids are most frequently found in modern proteins at RNA contacts[43] and were therefore promising to explore for their benefit to ribozyme function. Peptides P8, P9, and P10 incorporated amino acids that emerged later in evolution, including the aromatic amino acids whose synthesis may have required sophisticated enzymes. The length of the peptides was chosen so that they could have originated prebiotically due to wet-dry cycling in the presence of alpha-hydroxy carboxylic acids[22]. The sequence of amino acids within each of the peptides was more or less arbitrary, to reflect some of the many possibilities. Additionally, all amino acids are the (biological) L-enantiomers, whereas prebiotic amino acids likely contained racemic mixtures. Therefore, these ten peptides are no more than a starting point for future studies: If a specific peptide were found to be beneficial to ribozymes, it could be used as a foothold to explore the chemical space of related peptides that could have been useful in a prebiotic setting.

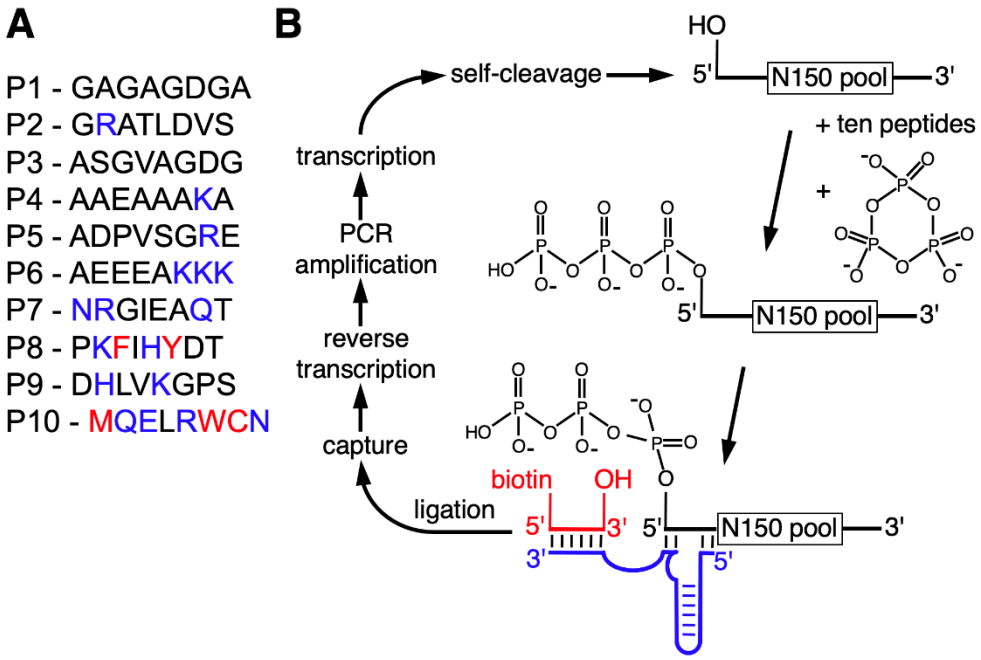

Figure 2.1: In vitro selection of ribozymes in the presence of ten octapeptides. **(A)** Ten octapeptides used for the in vitro selection. Peptides differ in how early their amino acids likely appeared: Black amino acids are generated in Miller-Type prebiotic model reactions. Blue amino acids likely appeared in intermediate stages. Red amino acids, which includes the three aromatic amino acids, likely appeared latest. **(B)** Schematic for the in vitro selection procedure for self-triphosphorylating ribozymes in the presence of peptides. The RNA pool (top) is prepared with a 5'-terminal hydroxyl group through the use of a self-cleaving hammerhead ribozyme. Upon incubation with the ten octapeptides and trimetaphosphate, catalytically active molecules convert their 5'-terminus to a 5'-triphosphate. Those molecules are covalently coupled to a biotinylated oligonucleotide (red) by using a ligase ribozyme (blue). The biotin allows the capture of active molecules on streptavidin beads. Reverse transcription, PCR amplification, and transcription complete one selection round.

The in vitro selection procedure for self-triphophorylation ribozymes was described earlier[35], and was modified for the current study to incorporate the presence of peptides during ribozyme reactions (Figure 2.1B). The selection procedure started from a DNA library that contained a randomized sequence of 150 nucleotides, flanked by constant regions that served as primer binding sites, and preceded by the sequence of a hammerhead ribozyme that cleaved itself co-transcriptionally and thereby generated a 5'-hydroxyl at the library's 5'-terminus. The purified RNA library with 5'-hydroxyl termini was incubated with Tmp and the ten peptides. After purification of the RNA library, those library molecules that catalyzed their own 5'-triphosphorylation were ligated to a short biotinylated DNA/RNA oligonucleotide, captured on magnetic streptavidin beads, reverse transcribed, and PCR amplified. The resulting DNA library was used for the next cycle of the selection procedure. The effective complexity of the library in

the first round of selection was at least 7.8 x 10^{14}. The library showed an increased fraction of self-triphosphorylated pool molecules after five rounds of selection, as judged by the required cycles of PCR after reverse transcription (Figure 2.S1), and later High-Throughput Sequencing (HTS) showed the pool enriched for active sequences in four rounds. To isolate the most active ribozymes, the selection pressure was increased starting in round six of the selection, by reducing the incubation time with Tmp and peptides, and reducing the Tmp concentration. The peptide concentration was kept constant at 0.5 mM for each peptide (for a total of 5mM together) in all selection rounds. It was not reduced over the rounds because this would have penalized ribozymes that relied on peptides. A total of ten selection rounds were performed.

Identification of peptide-using ribozymes

To identify self-triphosphorylation ribozymes that utilized peptide cofactors, two different methods were used: Biochemical analysis of arbitrarily selected clones, and high throughput sequencing (HTS) analysis. The biochemical assay had the advantage of detecting small changes in activity, while the HTS analysis had the advantage of covering millions of sequences.

The biochemical assay detected 5'-self-triphosphorylated RNAs by ligating them to a short, radiolabeled oligonucleotide and separating the products by denaturing polyacrylamide gel electrophoresis (PAGE) (Figure 2.S2). It was performed under the same buffer conditions as the first selection rounds (100 mM $MgCl_2$, 50 mM Tmp, 50 mM Tris/HCl pH 8.3 for 3 hours at 22°C). All of 32 arbitrarily selected clones from selection round five showed self-triphosphorylation activity. Two of the 32 ribozymes showed significantly increased activity in the presence of peptides, and the ribozyme with the largest average increase (a 40% increase over the reaction without peptide; clone 19) was chosen for further analysis. Ribozyme clone 19 was tested for activity in the presence of individual peptides (Figure 2.2). Six of the ten peptides increased its activity between 2-fold and 5-fold.

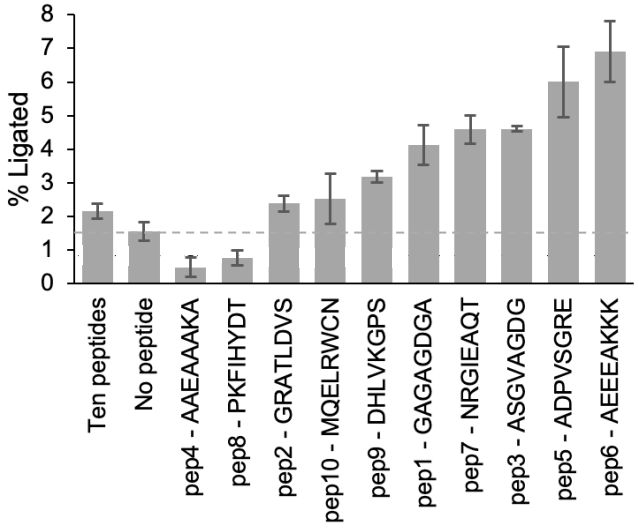

Figure 2.2: Effect of peptides on the self-triphosphorylation activity of ribozyme cluster 23 clone 19, under selection conditions. The activity was measured as the percent of ribozyme ligated to a short, radiolabeled oligonucleotide due to its 5'-triphosphate. In the sample 'ten peptides', each of the ten peptides had a concentration of 0.5 mM, with a total concentration of 5 mM. Samples 1-10 contained each one peptide at a concentration of 5 mM. The peptide sequence is given in each label. The dashed, horizontal line represents the ribozyme activity without peptides. Error bars are standard deviations from triplicate experiments.

HTS analysis detected sequence clusters after four rounds of selection, and found that the pool was dominated by active sequences after five rounds of selection (Figure 2.3A). While many clusters were detectable after five rounds of selection, one cluster (C13) dominated the population after 8 rounds, and another cluster (C74) dominated after ten rounds of selection.

To identify ribozyme clusters dependent on peptides for activity, the library after selection round five was subjected to 12 parallel sub-rounds of selection: One sub-round contained no peptides, one sub-round contained all peptides, and ten sub-rounds contained each one individual peptide. Comparison of the HTS results from these selection sub-rounds identified individual ribozyme clusters that benefitted from specific peptides in their activity (Figure 2.3B). Among the 55 sequence clusters that were most enriched in selection round 5, most show a positive or negative response to several peptides. Peptides 4 and 8 led to the strongest effects on ribozyme enrichment, both for activity enhancement (clusters 20, 71, 2, 4) and activity inhibition (cluster 41). Even though beneficial effects were often weak, they were at least sometimes relevant, as demonstrated by the peptide response of cluster 23, which is the cluster containing the ribozyme chosen from the biochemical screen (clone 19, see above). Based on the two

types of analysis (biochemical screen and HTS analysis), the two ribozyme clusters 20 and 23 were chosen for further analysis because they appeared to represent two different types of responses to the peptides.

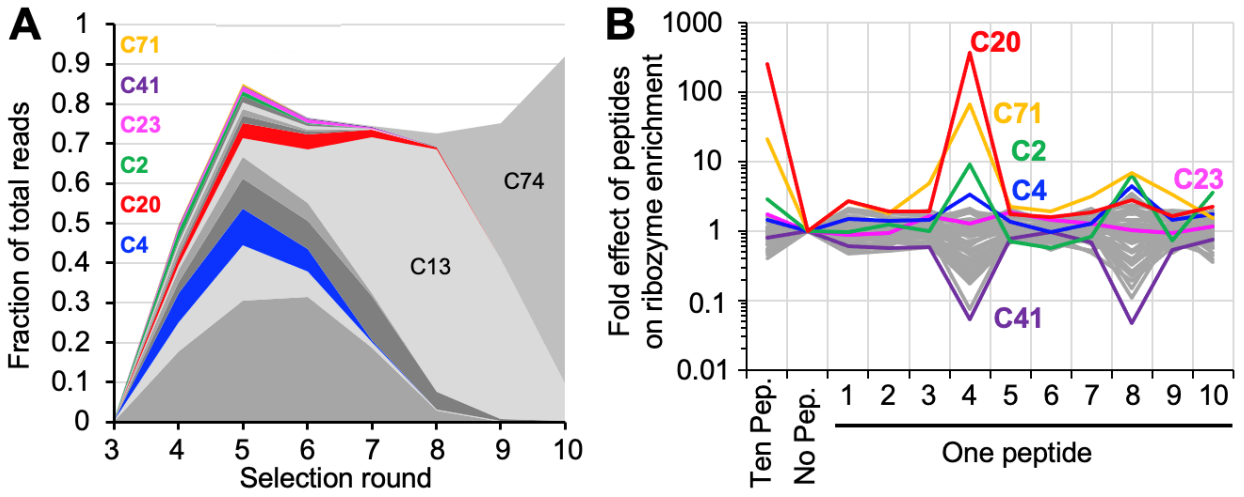

Figure 2.3. High Throughput Sequencing analysis of the selected sequences. **(A)** Enrichment of sequence clusters over the ten rounds of selection. Each color represents one sequence cluster. The height of each cluster section corresponds to the fraction of sequences in that selection round. Clusters of interest are labeled in color (note that clusters C41 and C71 are only tiny slivers at the top), and the two clusters C13 and C74 that dominate the population late are labeled in black. **(B)** Effects of peptides on the activity of the top 50 ribozyme clusters in selection round 5, based on HTS data. The activity is estimated from the frequency of a given cluster's sequences in the presence of a given peptide, as compared to the absence of peptides. By definition, the frequency is 1 for all clusters in the selection without peptide ('No Pep.'). The response to the presence of only one specific peptide is shown to the right. A ratio of 1 denotes no peptide effect, a ratio larger than one suggests an increase in ribozyme activity by the peptide, and a ratio smaller than one a decrease in activity by the peptide. Data shown in color were analyzed biochemically, with their cluster names indicated; the data for all other clusters are shown in gray. Labels below the graph show the number of each peptide (see Figure 2.1A).

The two ribozymes representing clusters 20 and 23 were truncated at their 3'-termini while maintaining (or increasing) activity (Figure 2.S3). Ribozyme cluster 20 was truncated to 120 nucleotides, and ribozyme cluster 23 to 171 nucleotides. In addition, cluster 20 was analyzed for the effect of five mutations that showed strong enrichment over selection rounds during HTS analysis (Figure 2.S4A). Interestingly, all of these mutations were close to each other, at positions 91 to 95 in the ribozyme (U91C, C92U, G93A, G93U, G95A). Despite this, none of the abundant sequences from cluster 20 had a mutation at A94, suggesting that it may have an important role. After testing variants of ribozyme 20 with these individual mutations and their combinations, mutation G93U was chosen as the winner because its

average activity was highest among all single, double, and triple mutants of the five mutations, and 3.8-fold higher than the peak sequence of cluster 20 (Figure 2.S4B). Both the peak sequence and the mutant G93U benefitted 4.3-fold from the combined presence of the ten peptides under selection conditions (each peptide at 0.5 mM). In summary, the two ribozymes chosen for further analysis were (i) the peak sequence of ribozyme cluster 20 after truncation to 120 nucleotides and insertion of the mutation G93U, and (ii) the clone 19 sequence of ribozyme cluster 23 after truncation to 171 nucleotides. For simplicity, these two ribozymes were called 'ribozyme 20' and 'ribozyme 23', respectively.

Ribozyme 20 and its peptide interactions

To identify the optimal reaction conditions for ribozyme 20, the concentrations of $MgCl_2$, Tmp, as well as the reaction pH and temperature were systematically varied. All assays were done in parallel in the absence and presence of peptides to learn where ribozyme activity may depend more or less on the peptides. The ribozyme showed highest activity at 200 mM $MgCl_2$ and 100 mM Tmp, when Tmp and $MgCl_2$ were co-varied (Figure 2.4A). Activity in the absence of peptide 4 was barely detectable. The activity was highest at pH 8.3 (Figure 2.4B), and at 25°C (Figure 2.4C).

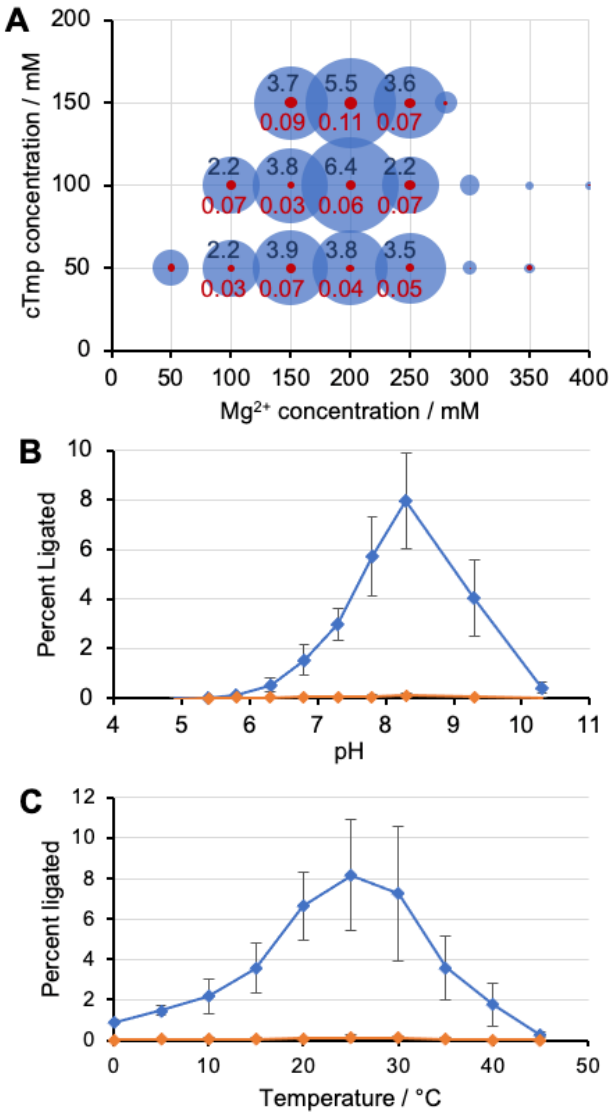

Figure 2.4. Optimization of reaction conditions for ribozyme 20. **(A)** Co-optimization of Tmp and Mg^{2+}. The size of the blue circles represents the activity with peptides; the size of the red circles represents the activity without peptides. The numbers in blue and red are the percent of reacted ribozymes with and without peptides, respectively. Peptides were used at 0.5 mM each, and a total of 5 mM. **(B)** Ribozyme activity as a function of the pH, in the presence of 200 mM Mg^{2+} and 100 mM Tmp. Blue symbols represent activity with peptides, orange symbols represent activity without peptides. **(C)** Ribozyme activity as function of temperature at 200 mM Mg^{2+}, 100 mM Tmp, and pH 8.3. Blue symbols represent activity with peptides, orange symbols represent activity without peptides. Error bars are standard deviations from at least three experiments.

When ribozyme 20 was tested under these optimized conditions with individual peptides (Figure 2.5A), peptide 4 showed a 900 (+/-300)-fold increase in activity, whereas no other peptide showed any effect above background. This confirms the conclusion from HTS analysis that peptide 4 shows a specific, and strong benefit to ribozyme 20. The concentration dependence on peptide 4 gave a sigmoidal behavior

with a half-maximal stimulation slightly above 1 mM peptide 4, and saturation at 1.75 mM (Figure 2.5B). The sigmoidal shape of the dependence suggested that two or more peptide 4 molecules may interact with one ribozyme 20 molecule to achieve the active conformation. Because the ribozyme showed some activity without peptide 4, the role of the peptide was unlikely a contribution to catalysis and more likely the stabilization of the active ribozyme conformation.

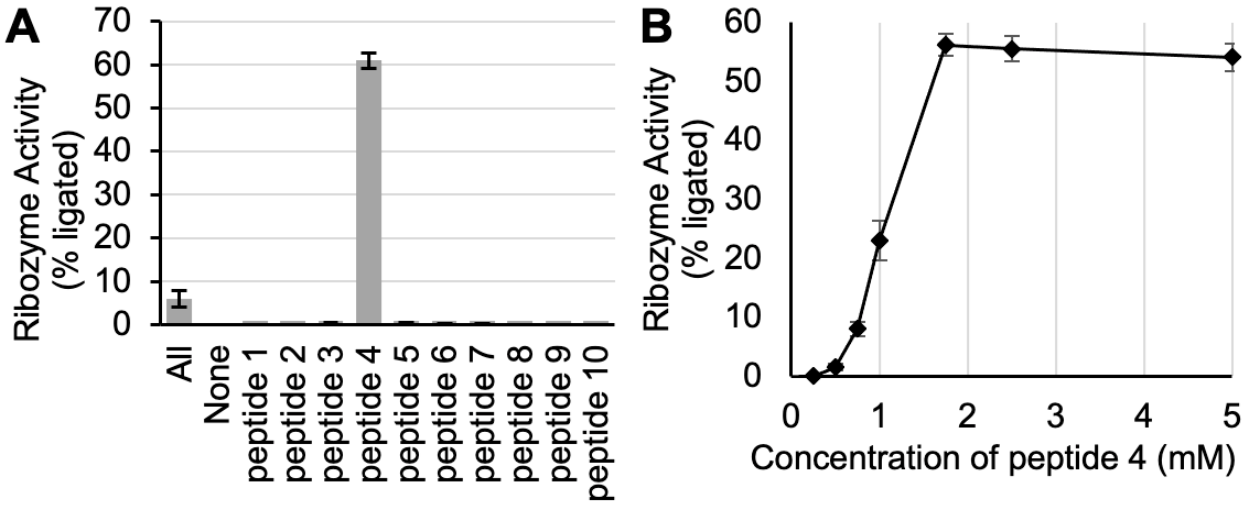

Figure 2.5. Interactions of ribozyme 20 with individual peptides. **(A)** Activity of ribozyme 20 in the presence of ten individual peptides, each at 5 mM, The reaction with all peptides has each at 0.5 mM concentration. **(B)** Titration of peptide 4 into the activity assay of ribozyme 20. Error bars are standard deviations from three experiments.

To elucidate how peptide 4 may modify the structure of ribozyme 20, the ribozyme secondary structure was analyzed by SHAPE probing[44] (Figure 2.6A). Differences in SHAPE reactivity with and without peptide suggested four distinct changes in the ribozyme structure (Figure 2.6B). First, the decreased accessibility in positions 7-9 suggested that the loop of helix 1, which is close to the catalytic site at position 1, becomes more rigid in the presence of peptide 4. Second, peptide 4 modulated helix 3 by increasing flexibility of the upper part of the stem (58-61 and 68-72) and constraining the loop positions 63 and 64. Third, the nucleotides 83-85 and 102-106 flanking helix 4 become more accessible. Fourth, position 91 in the loop of the helix 4 becomes less accessible, which is important because the loop of helix 4 contains all five mutations that improved ribozyme activity (Figure 2.S4). In total, 25 out of 105 analyzed positions in the ribozyme changed SHAPE accessibility outside experimental error. These

results show how peptide 4 causes a significant, but very defined modification in the secondary structure of ribozyme 20 by destabilizing helix 3, causing a protection if the 3'-flanking region of helix 4, and possibly as downstream effects the higher accessibility in the 5'-flanking region of helix 4, and rigidification of the loops on helix 1 and helix 4.

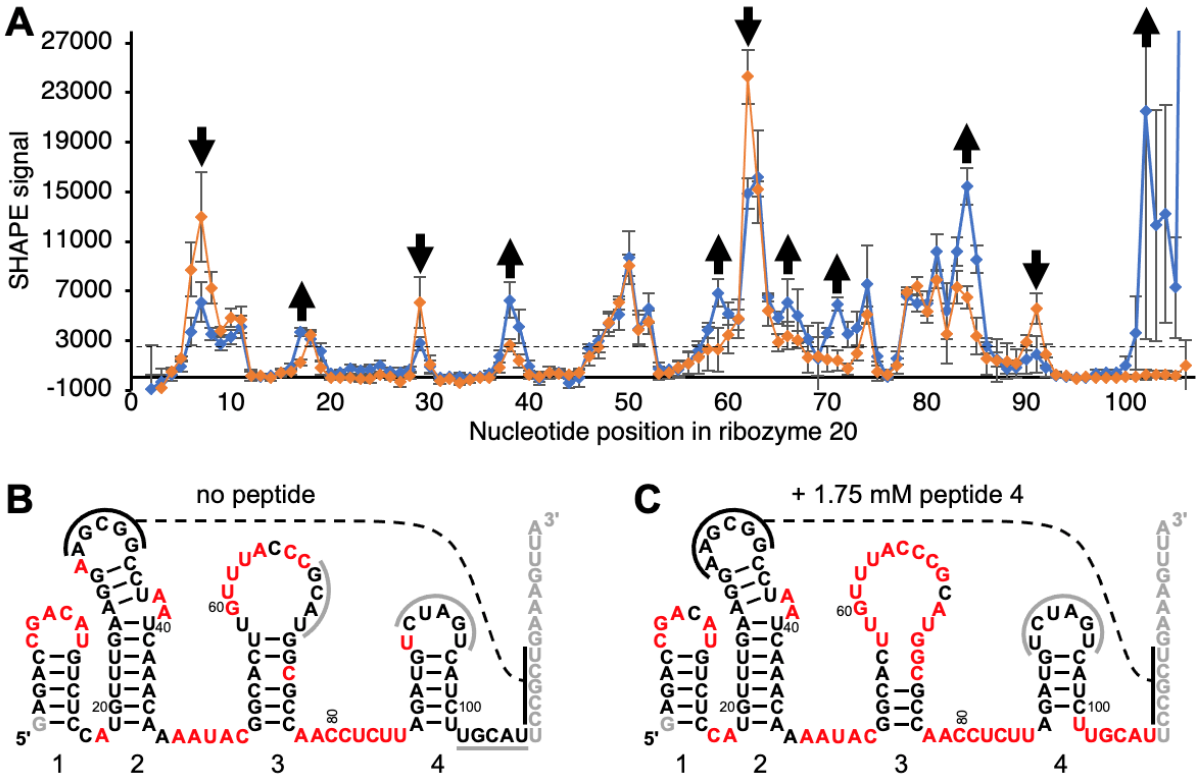

Figure 2.6. Secondary structure analysis of ribozyme 20 by SHAPE chemical probing in the absence and presence of peptide 4. **(A)** The SHAPE accessibility is plotted as a function of nucleotide position, for the ribozyme without peptide (orange) and with 1.75 mM peptide 4 (blue). The accessibility to the SHAPE reagent 1M7 was determined by blocking reverse transcription with 5'-[^{32}P] radiolabeled primers, and PAGE analysis. Nucleotides past position 106 are not analyzed because they were obscured by the reverse transcription primer. Error bars are standard deviations of triplicate experiments. For secondary structure prediction, the average signal intensity of 3,000 was used as cutoff to discriminate between 'accessible' and 'not accessible' (dashed line). Positions where the presence of peptide 4 increased or decreased SHAPE accessibility are marked with an up arrow or down arrow, respectively. **(B)** Predicted secondary structure without peptide, using constraints from SHAPE analysis. Positions in red denote SHAPE accessibility above 3,000 as shown in (A). The helices are numbered 1-4 below the image. Black solid lines indicate a possible tertiary contact with five base pairs, and the dashed line denotes the interaction. Note that positions 92-95, which are important for function (Figure 2.S4) are within a 'loop' without SHAPE accessibility. Grey solid lines indicate protected loop regions and unknown interaction partners. Grey positions are unconstrained by SHAPE because they were covered by the reverse transcription primer, or too close to the 5'-terminus. **(C)** Predicted secondary structure with 1.75 mM peptide 4, using constraints from SHAPE analysis. See (B) for details.

Ribozyme 23 and its peptide interactions

Ribozyme 23 showed highest activity at a combination of 150 mM Mg^{2+} and 100 mM Tmp, and a larger benefit of peptides under sub-optimal reaction conditions, at 50 mM Mg^{2+} and 50 mM Tmp (Figure 2.7A). This pattern of maximum peptide benefit under suboptimal reaction conditions was a general theme for ribozyme 23: At 150 mM Mg^{2+} / 100 mM Tmp, the highest ribozyme activity was at pH 9.8, while the peptide benefit was undetectable at pH 9.8, 1.3-fold at pH 8.3, and above 2-fold below pH 7 (Figure 2.7B). At 50 mM Mg^{2+} / 50 mM Tmp, highest ribozyme activity was also at pH 9.8, while the peptide benefit was lowest at pH 9.8, 1.7-fold at pH 8.3, and 2-fold at pH 7.3 (Figure 2.7C). At 150 mM Mg^{2+} / 100 mM Tmp and pH 8.3, ribozyme activity was highest at 35°C but the peptide benefit was highest at 20°C (Figure 2.7D). At 50 mM 50 mM Mg^{2+} / Tmp and pH 8.3, ribozyme activity and peptide benefit were both highest at 25°C (Figure 2.7E). One possible explanation for the different conditions of highest ribozyme activity and conditions with largest peptide benefit is that peptides may help ribozyme 23 to escape kinetic folding traps.

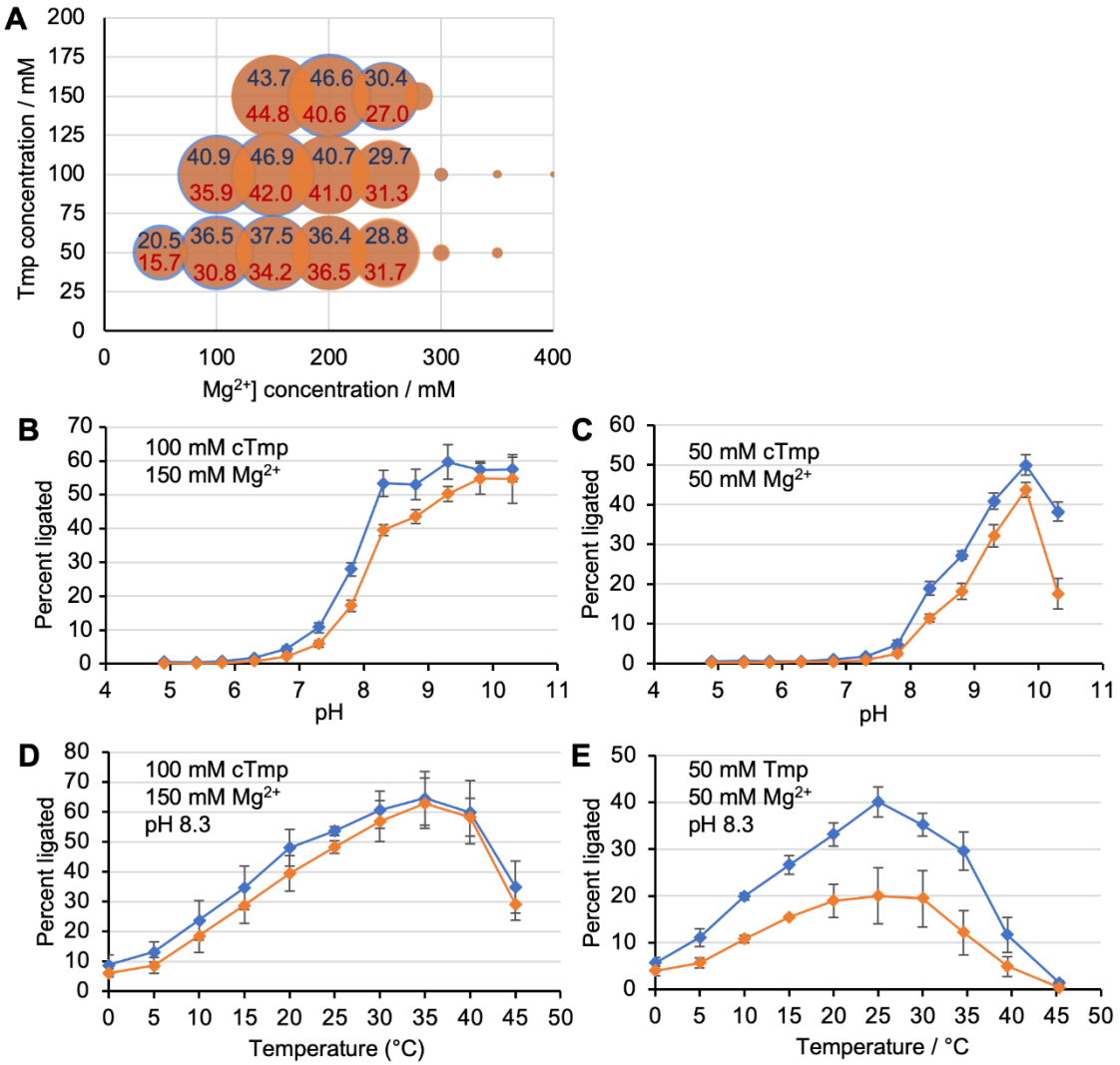

Figure 2.7. Optimization of reaction conditions for ribozyme 23. **(A)** Co-optimization of Tmp and Mg^{2+}. The size of the blue circles represents the activity with peptides; the size of the orange circles represents the activity without peptides. The numbers in blue and red are the percent of reacted ribozymes with and without peptides, respectively. Peptides were used at 0.5 mM each, and a total of 5 mM. **(B)** Ribozyme activity as a function of the pH, in the presence of 150 mM Mg^{2+} and 100 mM Tmp. **(C)** Ribozyme activity as a function of the pH, in the presence of 50 mM Mg^{2+} and 50 mM Tmp. **(D)** Ribozyme activity as function of temperature at 150 mM Mg^{2+}, 100 mM Tmp, and pH 8.3. **(E)** Ribozyme activity as function of temperature at 50 mM Mg^{2+}, 50 mM Tmp, and pH 8.3. Blue symbols represent activity with peptides, orange symbols represent activity without peptides. Error bars are standard deviations from at least three experiments.

To test which of the ten octapeptides generated the largest benefit for ribozyme 23 activity under optimized conditions, each peptide was incubated at 5 mM concentration with the ribozyme at 150 mM Mg^{2+}, 100 mM Tmp, pH 8 and 20°C, and compared to a reaction without peptides, and to a reaction containing all ten peptides with each 0.5 mM concentration (Figure 2.8A). The strongest benefit came

from peptide 6 and peptide 4. After 20 minutes of reaction time, all peptides except peptide 8 benefitted ribozyme activity, with the strongest benefits (>2-fold) from peptides 4 (AAEAAAKA) and 6 (AEEEAKKK). Even the most prebiotically plausible peptide 1 (GAGAGDGA) showed a 24% improvement on ribozyme activity under these conditions. However, almost all peptide benefits disappeared for reaction times of 3 hours (Figure 2.S5), showing the kinetic benefits to ribozyme activity. When the three peptides were titrated into the reaction incubated for 20 minutes (Figure 2.8B), peptide 6 showed a significant effect starting at 0.5 mM, while peptide 4 required 2.5 mM, and peptide 1 showed only a barely significant effect at 5 mM concentration. For all three peptides, their benefit did not appear to saturate at 5 mM concentration, suggesting that higher benefits can be reached at even higher peptide concentrations. Again, peptide benefits were less after 3 hours, with only 5mM peptide 6 showing a significant benefit (Figure 2.S5).

To further test the idea that the peptides were helping to resolve kinetic folding traps, the reaction kinetics for ribozyme 23 were recorded in the presence of 5 mM of peptides 1, 4, 6, and compared to the reaction without peptides (Figure 2.8C). These kinetics were recorded under optimal conditions (150 mM Mg^{2+}, 100 mM Tmp, 20°C, and pH 8.3) as well as sub-optimal conditions (the same conditions but at pH 6.8). Under optimal conditions, the peptides 4 and 6 benefitted ribozyme activity between 3-fold and 2-fold from the first time point to about one hour reaction time, whereas the benefit by peptide 1was barely detectable. Under suboptimal conditions, the reaction rates were slower but the peptide benefits more pronounced. Even peptide 1 showed a clear benefit between 3 hours and 27 hours, and the benefit of peptides 4 and 6 were about 3-fold for the first 9 hours of the time course. Together, these results describe the kinetic benefits of ribozyme 23 activity from peptides 6, 4, and 1.

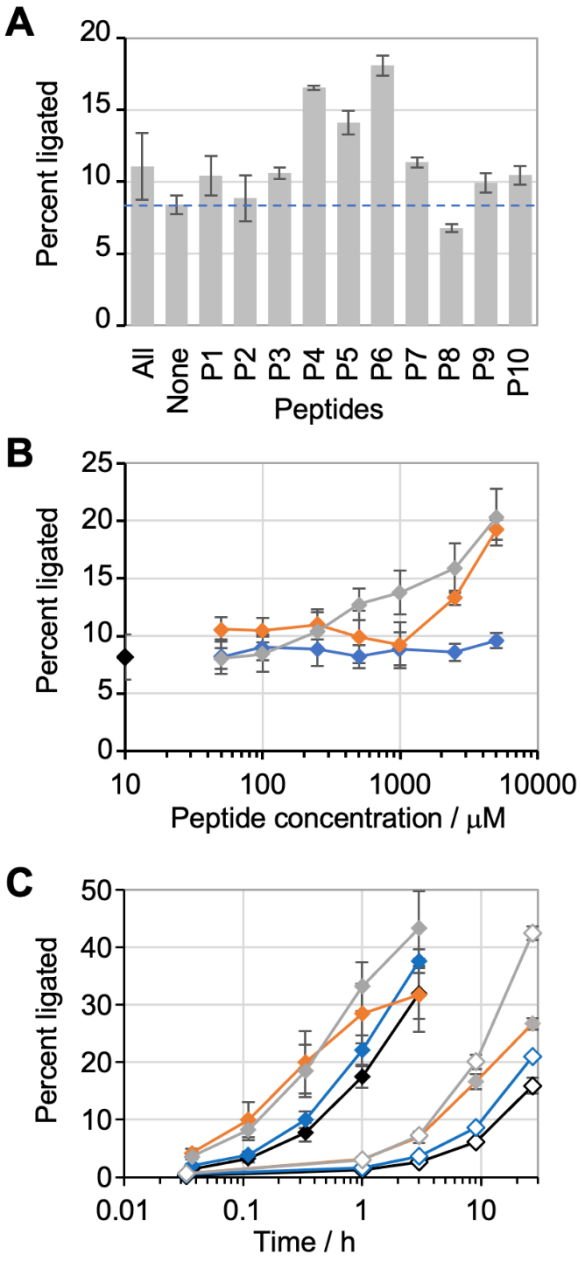

Figure 2.8. Effect of individual peptides on ribozyme 23 activity. **(A)** Activity of ribozyme 23 in the presence of ten individual peptides. All peptides were used at 5 mM. **(B)** Titration of peptides 1, 4, and 6 into the activity assay of ribozyme 23, with an incubation time of 20 minutes. The activity of the ribozyme without peptides (black diamond) is shown on the y-axis. The activity in the presence of peptide 6 (grey), peptide 4 (orange), and peptide 1 (blue) are shown as function of the peptide concentrations. **(C)** Reaction kinetics of ribozyme 23 in the absence (black) and presence of peptides 1 (blue), 4 (orange), and 6 (grey). Filled symbols denote optimal conditions; open symbols denote the same conditions but at pH 6.8. Error bars are standard deviations from three experiments.

To test whether the peptides benefitted ribozyme 23 by helping to escape kinetic folding traps, we probed the ribozyme secondary structure in the absence and presence of peptide 6 using SHAPE [44]

(Figure 2.9). A total of 56 positions out of the ribozyme's 156 tested positions showed a change in SHAPE accessibility outside error of triplicate experiments, corresponding to a strong effect of peptide 6 on the secondary structure of ribozyme 23. Most of these changes (50) were due to decreased SHAPE accessibility, which strengthens the idea that peptide 6 helps ribozyme 23 to overcome kinetic folding traps and maximize base pairing interactions in a more compact structure.

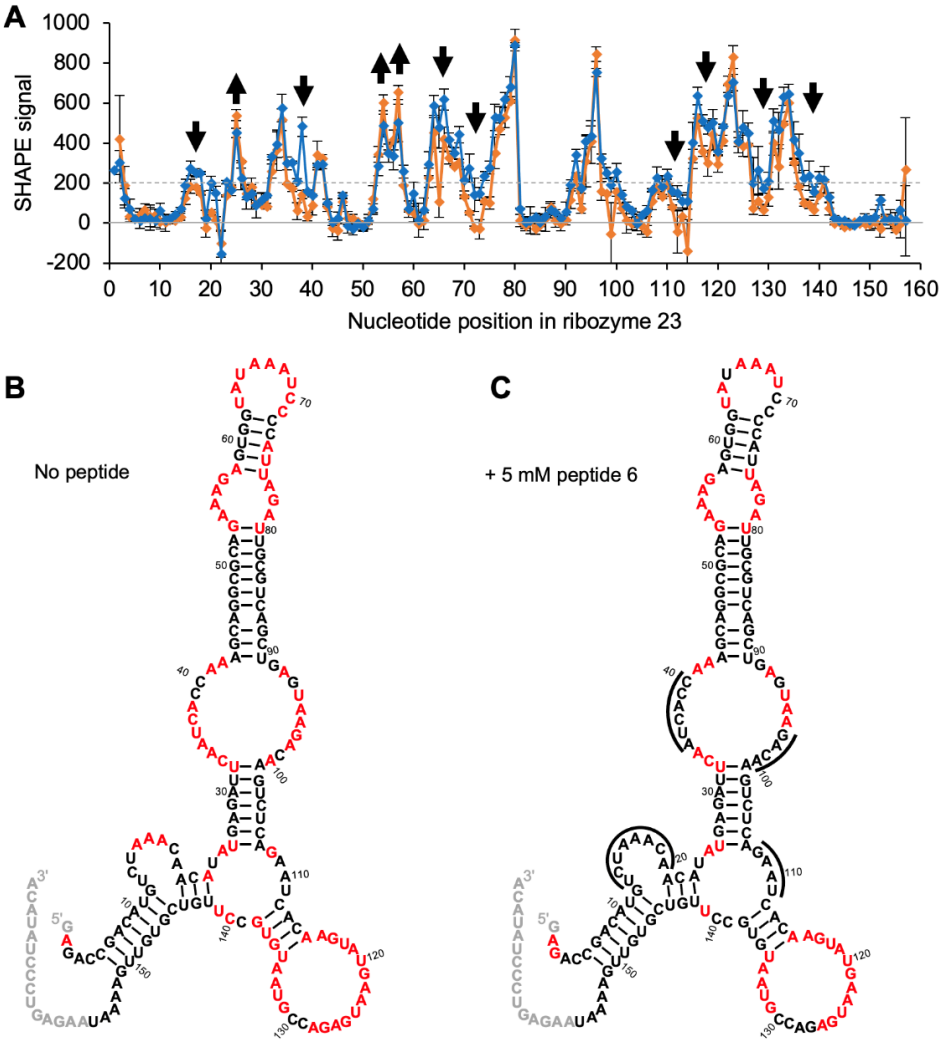

Figure 2.9. Secondary structure analysis by chemical probing of ribozyme 23, with and without peptide 6. **(A)** The SHAPE accessibility is plotted as a function of nucleotide position, for the ribozyme without peptide (blue) and with 5 mM peptide 6 (orange). Nucleotides past position 156 were not analyzed because they were obscured by the reverse transcription primer. Error bars are the standard deviations of triplicate experiments. For secondary structure prediction, the average signal intensity of 200 was used as cutoff to discriminate between 'accessible' and 'not accessible' (dashed line). Positions where the presence of peptide 6 increased, or decreased SHAPE accessibility are marked with an up arrow, or down arrow, respectively. **(B)** Predicted secondary structure of ribozyme 23 in the absence (left) and the presence of peptide 6 (right). The accessibility from SHAPE analysis in indicated in red, using the SHAPE cutoff in (A). Regions with at least four consecutive, protected positions not explained by the secondary structure are marked with a curved, black line that could correspond to pseudoknots or tertiary interactions.

Discussion:

This study explored how peptides can aid the emergence of catalytic RNAs from random sequence by using an in vitro selection system for self-triphosphorylating ribozymes in the presence of ten different octapeptides. Several ribozymes were identified, and two were studied in more detail. Ribozyme 20 increased 900 (+/-300)-fold in activity in the presence of peptide 4. While the specific requirement of one peptide would have benefitted most from the emergence of encoded translation, the specific recruitment may also be a way of selecting a useful peptide from a prebiotic mixture of peptides. In contrast, ribozyme 23 benefitted from six to eight of ten peptides. Peptide benefit was most pronounced under suboptimal reaction conditions and early reaction time points, suggesting that ribozyme 23 is helped by peptides to escape kinetic folding traps. In summary, the two ribozymes demonstrate different principles how peptides could have aided ribozymes in early stages of life.

RNA molecules are prone to get trapped into inactive folding intermediates; this has been termed the RNA folding problem[45]. Modern proteins that aid RNA folding without the use of energy can be divided in two classes: proteins that form sequence specific, stable complexes with RNA, and proteins that bind non-specifically, and transiently to RNA[46]. These two classes are mirrored in the two ribozymes analyzed in this study: Ribozyme 20 benefits 900 (+/-300)-fold from specific interaction with peptide 4, and ribozyme 23 benefits to a lesser extent, from multiple different peptides (Figures 2.5 and 2.8).

The peptide specificity of ribozyme 20 illustrates the evolutionary benefits for a primitive ribosome. A primordial ribosome may not have had the ability to synthesize a protein with sophisticated structure - but it would have been able to synthesize short peptides with specific sequences[47,48,49]. Such peptides could play a chaperone-like role by binding to and stabilizing RNA structures[48], and could have enlarged the structural repertoire of RNA through peptide-induced rearrangements of RNA structure[49]. Because ribozyme 20 shows detectable activity even in the absence of peptide 4 (Figure 2.5B), it appears more likely that peptide 4 simply stabilizes ribozyme 20 in the active form, and that the active form is not well populated without peptide. The identification of the ribozyme 20 / peptide 4 model system now

allows addressing the detailed mechanism how short, encoded peptides can benefit the function of catalytic RNAs.

Peptide 4 (AAEAAAKA) is dominated by alanine residues, and its glutamate and lysine are positioned ideally to form a salt bridge that stabilizes alpha-helices[50]. While the length of eight amino acids is likely too short to form an alpha-helix in free form, it is possible that the RNA-bound form (and therefore the activity mediating form) stabilizes an alpha-helix. Previous studies on an N-terminal, arginine-rich peptide of the HTLV-Rex protein showed that binding to an RNA aptamer stabilized an alpha-helical conformation using only two hydrogen bonds between carbonyl oxygen and amide nitrogen of the i to 1+3 form[51]. In general, peptide alpha-helices can bind into the major groove of RNA if the narrow major groove is widened by non-Watson-Crick interactions[52]. Similar widening of the major groove can be accomplished by fraying near the end of a duplex, or disruptions by internal loops, especially if the two sides of the loop have different length[53]. Because the narrow major groove is much richer in hydrogen-bonding partners than the openly accessible minor groove,[54] peptide 4 may be able to insert itself into helix 3 of ribozyme 20 to distort the helix, thereby destabilizing some secondary structure elements while stabilizing others. Future studies will reveal the conformation of peptide 4 when bound to ribozyme 20.

The non-specific benefit of peptides to ribozyme 23 may rely on ionic bonds and hydrogen bonds with effects similar to the soft ion bonds that allow the polyamines spermine and spermidine to promote or inhibit ribozyme function, which reduce friction between helices and help form the active conformation of the hammerhead ribozyme[55], hairpin ribozyme[56], and the HDV ribozyme[57]. Spermidine helps group I intron ribozymes fold into the active conformation by destabilizing long-range kinetic folding traps[58], it rescues the activity of group I intron ribozymes that lost part of their structural scaffold[59,60], and the larger group II intron ribozymes depend on spermidine in the absence of proteins[61]. In further support of the analogy between polyamines and peptides, spermidine appears to bind group II intron ribozymes at the same sites as peptides[62], and for RNase P the absence of the protein component can be rescued by spermidine[63].

Could the presence of prebiotic peptides have helped the emergence of ribozymes in early stages of life? The identification of ribozymes 20 and 23, which benefit from peptides in different ways, suggest that the answer is yes. Based on the HTS data (Figure 2.3B), it appears that out of the top 50 ribozyme clusters, four show a similar peptide benefit profile as ribozyme 20 (ribozyme clusters 20, 71, 2, 4). It is less clear which other ribozymes clusters show a mild benefit from peptides - similar to ribozyme 23 / peptide interactions - because these weaker effects are hard to detect by HTS analysis. A caveat of the current study for origin-of-life models is that the ten peptides are based exclusively on the twenty biological amino acids in their L-conformation, and that only an initial set of ten peptides was incorporated. Additionally, this study is confined to ribozymes that bind one negatively charged substrate (Tmp) and catalyze the nucleophilic attack of the ribozyme's 5'-hydroxyl group to a Tmp phosphorus, under a specific set of selection conditions. Future studies on different peptides, and different ribozymes will broaden our understanding how peptides could have helped the emergence of catalytic RNAs.

The ribozyme 23 / peptide model system is of particular importance to the beginnings of an RNA world because it shows that even before the advent of encoded translation, peptides could have benefitted RNA catalysis in a non-specific way. Ribosomal peptides appear to reflect how peptides benefitted the function of RNAs through early stages of life[65]. This history can be recapitulated by following the signals left by the evolutionary steps of the ribosome, from the peptidyl transferase center (PTC) to the periphery[66]. Peptides and proteins close to the center are almost devoid even of secondary structure, while proteins in layers successively closer to the periphery contain more and more sophisticated tertiary structures[65]. The non-specific benefits of peptides to ribozyme 23 could reflect how prebiotically generated peptides with varying sequence could have helped the proto-ribosome core - the PTC - fold into its active conformation at low temperature and / or high salt concentration. Therefore, future studies on the ribozyme 23 / peptide complex may help understand how the PTC core of the ribosome started its long evolutionary journey as an RNA interacting with a small, unencoded peptide.

Materials and Methods:

Synthesis and purification of peptides

Peptides were synthesized on rink amide PS resin (0.77 mmol/g, Novabiochem) and Rink Amide ChemMatrix (0.45 mmol/g, Biotage) following the Fmoc/tBu strategy using a microwave-assisted peptide synthesizer (Alstra, Biotage). Side chain protecting groups for amino acids were as follows: Ser(tBu), His(Trt), Lys(Boc), Asp(OtBu), Glu(OtBu), Trp(Boc), Arg(Pbf), AsN(Trt), Thr(tBu), Gln(Trt), Tyr(tBu), Cys (Trt). Typical coupling reactions were performed at 0.200 - 0.300 mmol scale with 3.0 equiv. of Fmoc-protected amino acids, 2.94 equiv. of HBTU and 6.0 equiv. of NMM for 5 min at 75 ºC. All reagents were pre-dissolved in DMF at 0.5 M. Fmoc deprotections were performed with 20% piperidine in DMF for 10 min at rt. Washing was performed after every deprotection and coupling step using DMF. For the syntheses of some sequences, the deprotection and coupling steps described above were either performed twice or with increased reagent equivalence to ensure reaction completion. Cleavage of the peptide sequences from the solid support with concomitant side chain deprotection was accomplished by placing the resin in fritted SPE tube and treating with TFA cleavage cocktail (~20mL/g) containing 90:5:2.5:2.5 TFA:dimethoxybenzene:H_2O:TIS for 2 hr. For peptide sequences containing cysteine and methionine, EDT was added at 90:4:2.5:2.5:1.0 TFA:dimethoxybenzene:H_2O:EDT:TIS. Cleaved peptides were then precipitated in cold ether, centrifuged, dissolved in methanol and reprecipitated in cold ether (3x).

HPLC characterization and purification of peptides were carried out at rt on analytical (Jupiter C18 5 μm, 300 Å, 150 x 4.6 mm by Phenomenex, Torrance, CA) and semi-preparative columns (Aquasil C18 5 μm, 100 Å, 150 × 10 mm by Keystone Scientific Inc., Waltham, MA) with Prostar 325 Dual Wavelength UV-Vis Detector from Agilent Technologies with Varian pumps (Santa Clara, CA) with detection set at 225 and 406 nm. Peptides were eluted from column following a gradient using mobile phases A: 0.1% TFA in H20 and B: 0.1% TFA in CH_3CN. MS analyses were obtained on a LTQ ESI-MS spectrometer (San Jose, CA). Solutions were prepared in either methanol or methanol/water(formic acid

1%) with flow rate of 10 μL/min, spray voltage at 4.50 kV, capillary temperature at 300°C, capillarity voltage at 7.00 V, tube lens at 135.00 V. Purified peptides were characterized by analytical HPLC (with purity typically greater than 90%) and MS (either by direct injection or LC-MS).

To remove a possible carryover of TFA, peptides were dissolved in a total volume of 2 to 5 mL 100 mM $(NH_4)HCO_3$, and frozen as a thin film on the inside of a glass bulb cooled in a dry ice / isopropanol bath. After desiccating the frozen solution to dryness in oil vacuum (~2 mbar), the process was repeated once with 100 mM $(NH_4)HCO_3$, and once with water. The desiccated peptide was weighed, and dissolved in water to a stock concentration of 10 mM.

In vitro selection

The in vitro selection was performed essentially as described[35], with two important modifications: First, the RNA library was incubated with Tmp in the presence of a total of 5 mM peptides. The peptides had to be removed after this incubation to facilitate the ligation, and pull-down of successful library molecules. Second, the library 5'-terminus, ligase ribozyme, capture oligo, and library 3'-terminus were re-designed to avoid the selection of sequences from the previous study in our lab.

The DNA library consisted of a T7 RNA polymerase promoter, a hammerhead ribozyme, a 5'-constant region, 150 randomized positions, and a 3'-constant region. This double-stranded library was generated by PCR amplification from a custom-synthesized ultramer^R DNA (IDTDNA) with the sequence 5'-TGCGATTACGTGTATA-N150-AGACATGTCGGTCTCGACTG-3' (lower strand), the 5'-PCR primer 5'-

*AATTTAATACGACTCACTATA*gggcggtctcctgacgagctaagcgaaactgcggaaacgcagtcGAGACCGACATGTCT-3' and the 3'-PCR primer 5'-TGCGATTACGTGTATA-3' where the italicized portion constitutes the T7 RNA polymerase promoter with 5'-terminal enhancer sequence[67], the lower case portion denotes the sequence of the 5'-terminal hammerhead ribozyme that is required to generate a 5'-terminal hydroxyl group, and the underlined portion is complementary to the ultramer^R DNA. The minimum effective

complexity of the pool was determined by using the number of viable pool DNA molecules and accounting for the likelihood of duplicate sequences (complexity = 1-(1-(1/(8.3*10^14))^(2.49*10^15))). The percent viability of the pool molecules was determined by qPCR relative to a gel-purified standard. The number of pool DNA molecules is likely a low estimate since the qPCR used more stringent PCR conditions than the Taq amplification in the actual selection. After transcription and PAGE purification, the library RNA was incubated at a concentration of 200 nM with 50 mM Tris/HCl pH 8.3, 100 mM $MgCl_2$, 50 mM Na_3-Tmp, and 3.3 mM NaOH (to compensate for the pH drop due to Mg^{2+} chelation by Tmp), and 0.5 mM of each of the ten peptides, in volumes of 10 mL for three hours at room temperature. After ethanol precipitation the large pellet was eluted with a total of 200 uL water, and the small, remaining pellet was dissolved in 0.5 mL water and extracted with 0.5 mL phenol equilibrated to 10 mM Tris/HCl pH 8.3, then with phenol/chloroform, and finally with chloroform. After ethanol precipitation the recovery was consistently about 80% of the inserted RNA.

The recovered RNA was ligated to the 3'-terminus of the biotinylated oligonucleotide 5'-bio-GTAGTGCTTCAArU-3' using the R3C ligase ribozyme[68]. The construct based on the R3C ligase ribozyme was designed to minimize interactions between the two arms of the ribozyme after testing several constructs. The final construct showed a ligation efficiency of ~40% on the N150 pool. The ligation products were captured on streptavidin coated magnetic beads (Promega), eluted with 96% formamide at 65°C / 5 minutes, and collected by ethanol precipitation. Recovered RNA was reverse transcribed with the 3'-PCR primer and Superscript III Reverse Transcriptase (Invitrogen). RT products were first PCR amplified with 5'-PCR primer 5'-TGCGATTACGTGTATA-3' and 3'-PCR primer 5'-TGCGATTACGTGTATA-3', then with the long 5'-PCR primer containing T7 promoter and hammerhead ribozyme to complete one round of selection. The number of PCR cycles in the first PCR that was necessary to see a clear band on an agarose gel was used to monitor the progress of the selection.

HTS analysis

DNA pools from rounds 1-10 of the selection, in addition to any subrounds with different peptides included, were prepared for high-throughput sequencing by PCR addition of barcode (each pool with a different barcode), adapter, and index sequences. DNA pools were then mixed together for the sequencing reaction, which was performed at the UCSD Institute for Genomic Medicine (IGM) Genomics Center. An Illumina MiSeq platform was used for paired-end 200 base pair reads.

High-throughput sequencing data was received from the UCSD sequencing core already de-multiplexed (separated/sorted from the same pool into subpools by barcode) into fastq files. The reads from opposite ends were merged using merge function of usearch[70]. A bash script converted the fastq to readable fasta files, and then a python script removed any reads that did not contain the full 5' and 3' constant regions or a long enough variable region (any with less than 135nt of the expected 150nt were removed) to remove partial sequences. This was considered the full set of sequences for each subpool. Then a python script counted how many reads were present for each unique sequence and generated a fasta formatted file for each subpool with a total read count for each unique sequence. Using the subpool from round 5, the clustering function of usearch was used to create clusters of sequences with each sequence sharing at least 75% identity. For each cluster with over 30 total reads in round 5, a python script found all sequences from that cluster in every other subpool so that the abundance over the selection/under different peptide conditions could be followed.

Self-triphosphorylation activity assays

To measure the self-triphosphorylation activity of different ribozymes and their variants, an assay was used in which short, radiolabeled oligonucleotides are ligated to triphosphorylated ribozymes, and where this ligation is detected in a gel shift assay[35]. Because the radiolabeled oligonucleotide and the ribozymes were used at equimolar concentration the results report also the degree to which the ribozyme was triphosphorylated. The ribozymes were prepared by run-off transcription from PCR products

containing the sequence of the T7 RNA polymerase promoter, the hammerhead ribozyme, and the triphosphorylation ribozyme. After transcription and co-transcriptional self-cleavage of the hammerhead ribozyme, the self-triphosphorylation ribozyme carrying a 5'-hydroxyl group was purified by denaturing PAGE. The 14-mer 5'-d(AGTAGTGCTTCAA)rU-3' (identical to the biotinylated oligonucleotide as in the selection but without biotin) was 5' radiolabeled using polynucleotide kinase (NEB) and γ-[^{32}P] ATP (Perkin-Elmer), followed by PAGE purification. Under standard conditions, the self-triphophorylation ribozyme was incubated at 5 uM concentration with 50 mM Tris-HCl pH 8.3, 100 mM $MgCl_2$, 50mM trisodium trimetaphosphate, 3.33mM NaOH, and peptides at 0 to 5 mM final concentration. The Tmp solution was made fresh. This mixture was incubated at ~22°C for 3 hours, after which time 2μL were removed and added to 8μL of a ligation pre-mix on ice. The final concentrations were 1μM triphosphorylation ribozyme, 1μM R3C ligase ribozyme, 1μM unlabeled 14-mer, less than 10 nM of 5'-[^{32}P] labeled 14-mer with 10,000 cpm, 100 mM Tris-HCl pH~8.5, 25 mM sodium-EDTA pH~8.0 (a 5mM excess over $MgCl_2$), and 100 mM KCl. The mixture was heated to 65°C and cooled at ~0.1°C/sec to 30°C to anneal the arms of the ligase ribozyme to the triphosphorylation ribozyme and the 14-mer. Upon reaching 30°C, 10μL of this mixture were added to 10μL of 50 mM $MgCl_2$, 4 mM spermidine, and 40%(m/v) PEG 8000. After incubation for 2 hours at 30°C the reaction was stopped by ethanol precipitation. The products were separated on 7M urea 10% PAGE 45 minutes. Overnight exposures to a phosphorimaging screen were scanned on a Typhoon phosphorimager (GE) imager, and the bands were quantified with the software Quantity One (Bio-Rad) using the rectangle method. The 'fraction ligated' was calculated as the [volume of ligated band] / [volume of ligated and unligated bands].

Because the fraction of ligated RNA depends not only on the self-triphosphorylation activity but also on the ability of the ribozyme to serve as substrate to the ligase ribozyme[35], a calibration was performed (Figure 2.S6). Specifically, the final constructs of ribozyme 20 and ribozyme 23 with a 5'-triphosphate due to their transcription without a hammerhead ribozyme were tested in the assay. Ribozyme 20 resulted in an average 'fraction ligated' of 68.6 ± 2.0%, and ribozyme 23 resulted in an average of 53.7% ± 2.6%. For both ribozymes, the starting conditions was identical to their incubation

under optimized conditions. In both cases there was no significant difference when this pre-incubation was performed in the absence, or presence of peptides. Therefore, the values of 'fraction ligated' under optimized conditions with the final ribozyme constructs can be converted to 'fraction triphosphorylated' when multiplying the 'fraction ligated' by 1.46 for ribozyme 20, and with the factor of 1.86 for ribozyme 23.

SHAPE probing

At the buffer conditions optimized for each ribozyme (Tris/HCl replaced by HEPES/KOH or AMPSO/KOH), 49 uL of a solution with 200 nM ribozyme were mixed with 1 uL of a solution with 100 mM 1M7[44] in dry DMSO or a control of DMSO, and incubated at the given temperature for 3 minutes. After ethanol precipitation, the product was reverse transcribed using Superscript III Reverse Transcriptase (Invitrogen) and a 5'-[^{32}P] radiolabeled primer complementary to the ribozyme 3'-terminus. RT products were heated in 750 mM NaOH for 5' / 80°C, ethanol precipitated, and separated on 7M urea 10% polyacrylamide gels for 2-4 hours. Exposures of phosphorimager screens were scanned on a Typhoon Phosphorimager (GE). Each of the bands were quantified using the rectangle method in Quantity One (Bio-Rad). The volume of the bands in the 'minus 1M7' land were subtracted from the volume of each band in the 'plus 1M7' lane. Based on the RT primer and the full-length ribozyme, the bands were assigned to positions in the ribozyme, generating a profile for each experiment. Each profile was normalized, and the averages and standard deviations from three experiments were reported. The cutoffs chosen between 'weak signals' and 'strong signals' are indicated for each ribozyme. The secondary structures are based on secondary structure predictions by mfold[69], constrained by the position of flexible / single-stranded nucleotides derived from the SHAPE experiments.

SHAPE probing was done with ribozymes carrying a 5’-hydroxyl group in the absence of Tmp. For ribozyme 20, SHAPE probing was done at 50mM MgCl2, 50mM HEPES / KOH pH 8.0. For ribozyme 23, SHAPE probing was done at 50mM MgCl2, 50mM MOPS/HCl pH 6.8.

Acknowledgements:

We thank Russell Doolittle for help with the final purification of peptides. We thank Ram Krishnamurthy for helpful discussions. We thank Logan Norrell for pilot experiments. NASA is thanked for funding of KJS, JS, MJ, and UFM.

Chapter 2, in full, is currently being prepared for submission for publication of the material. Sweeney, Kevin J.; Jorge, Micaella Z.; Schellinger, Joan; Muller, Ulrich F. The dissertation author was the primary investigator of this material.

Supplementary Figures:

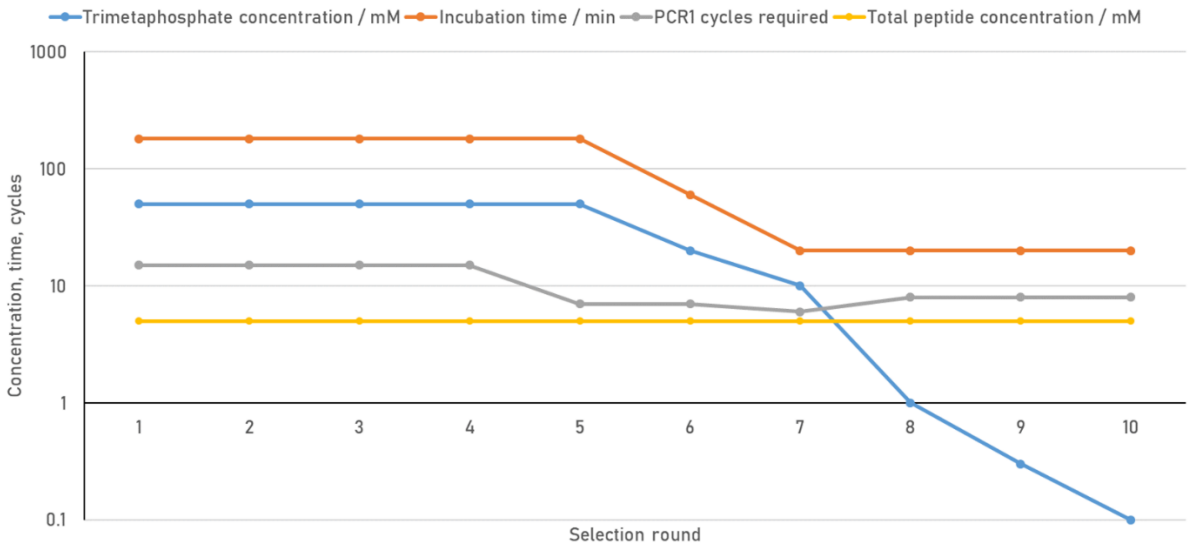

Figure 2.S1: Conditions and progress of the in vitro selection rounds 1-10. Shown are the trimetaphosphate concentration in mM, the total mixed peptide concentration in mM, the trimetaphosphate incubation time in minutes, and the number of cycles required to PCR amplify the pool at the end of each round. The drop in PCR1 cycles required in round 5 indicates that active sequences dominated the pool, as it meant that more RNAs had survived the selection step than in previous rounds. Total peptide concentration remained constant throughout the selection, while trimetaphosphate concentration was lowered towards the end of the selection to increase the selection pressure for binding trimetaphosphate. Incubation time was lowered from 180 minutes to 60 minutes for round 6 and from 60 minutes to 20 minutes for round 7, but was not lowered further to increase the odds that ribozymes could participate in multiple separate binding events (peptide(s) and trimetaphosphate).

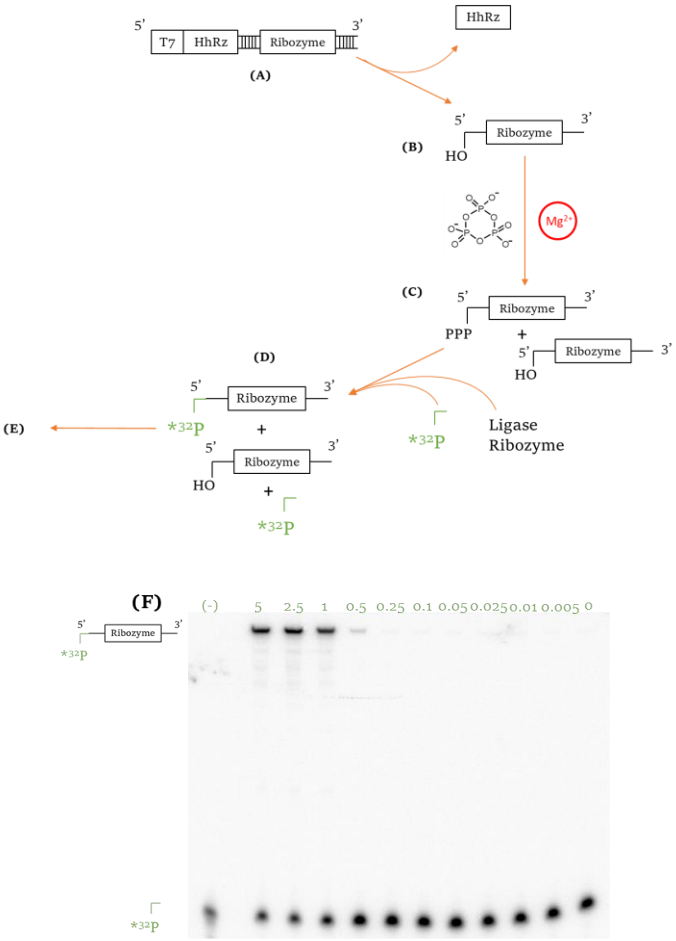

Figure 2.S2: Activity assay of triphosphorylation ribozymes. **(A)** The double-stranded DNA, a PCR product of a single clone's sequence, contains a T7 promoter, the hammerhead ribozyme, and the ribozyme's sequence, which includes the 5' and 3' constant regions along with the sequence selected from the original N150 randomized region. The DNA is transcribed by T7 polymerase and the hammerhead ribozyme cleaves itself off cotranscriptionally, leaving behind **(B)** the ribozyme with a 5'-hydroxyl group. The ribozyme is then incubated with trimetaphosphate, Mg^{2+}, and peptide(s) in a buffered solution, and **(C)** some RNAs will be able to coordinate trimetaphosphate and perform nucleophilic attack with the 5'-hydroxyl group. The reaction is then quenched, and the ribozyme is incubated with a ligase ribozyme, which will react 5'-triphosphorylated RNAs with a radiolabeled oligonucleotide to produce **(D)** radiolabeled active pool RNA molecules. **(E)** The products are then separated by denaturing PAGE, and the resulting separation of radiolabeled oligonucleotides being ligated or not ligated to the ribozyme is visualized using a phosphorimaging screen and phosphorimager. **(F)** An example gel scan showing ligated and unligated radiolabeled oligonucleotides. The experiment was a reaction of ribozyme 20 in the presence of different concentrations of peptide 4 (data shown in Figure 2.5).

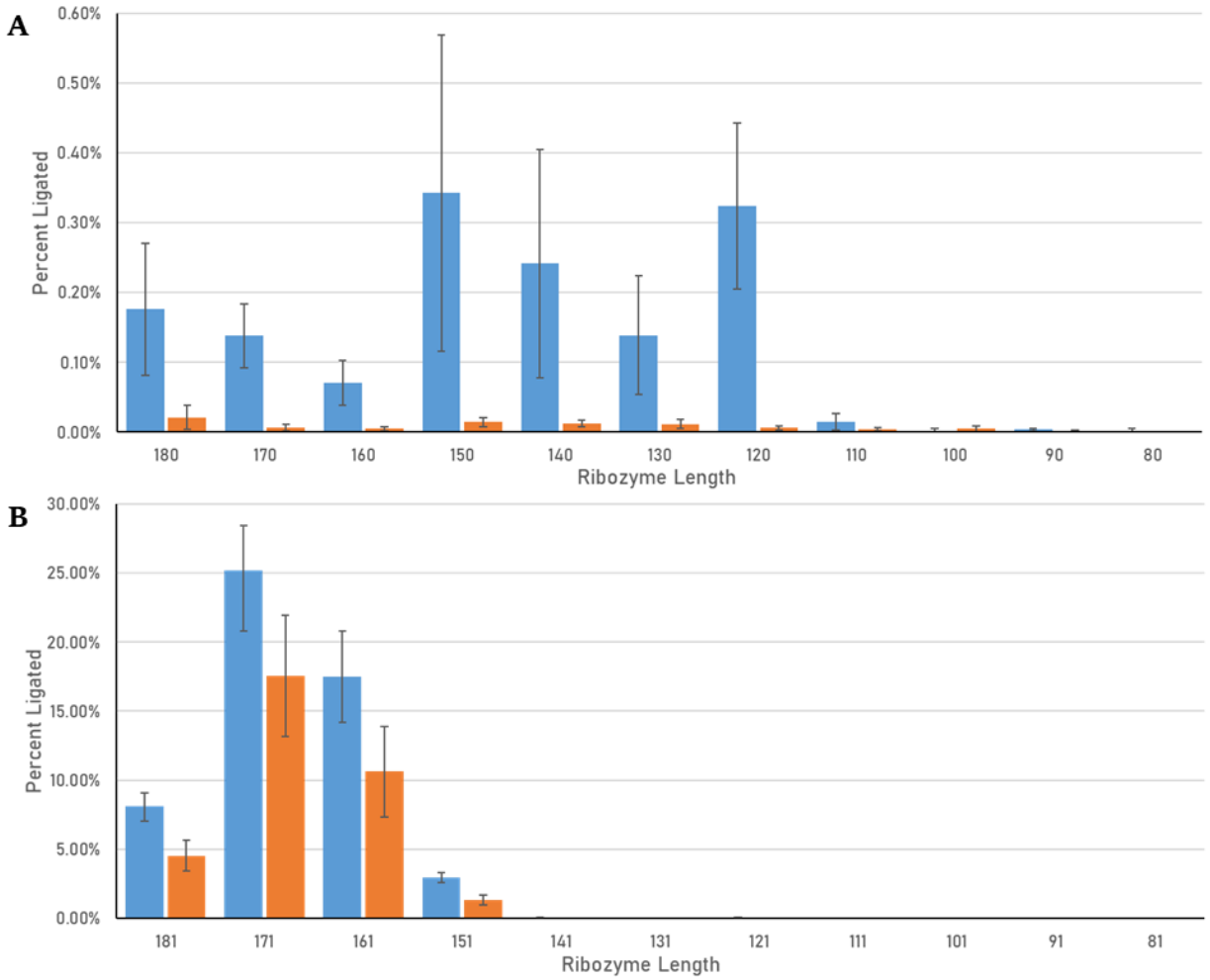

Figure 2.S3: Activity assay of 3'-truncated variants of **(A)** the cluster 20 peak sequence and **(B)** clone 19. Truncations were created by PCR amplifying the construct using a 3' primer that annealed partway into the ribozyme, not at the 3' end. Blue columns represent activity with peptides, orange columns represent activity without peptides. Error bars represent standard deviations from triplicate experiments.

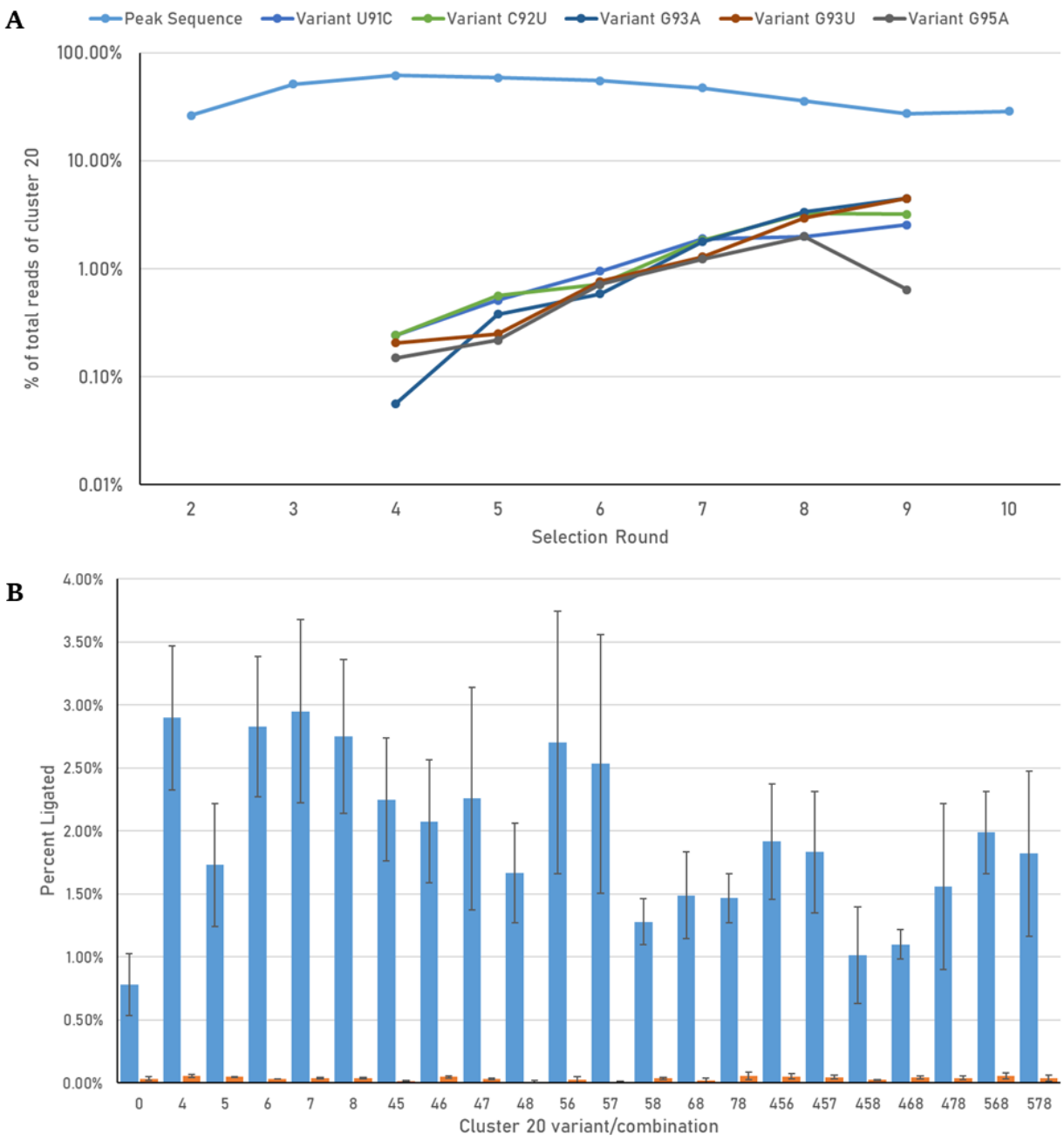

Figure 2.S4: Mutations of cluster 20 variants discovered using High-Throughput Sequencing improve the ribozyme's activity. **(A)** High-Throughput Sequencing data showing the relative abundance of each individual sequence within the total sequence reads from cluster 20. Shown are the peak sequence (the most abundant from the cluster and the first one tested) and five variants which were becoming relatively more abundant as the selection progressed, indicating possible activity benefits. All five mutations are within a four nucleotide stretch (91-95). **(B)** Biochemical testing of single, double, and triple mutants using the mutations found in the cluster 20 variants. All single mutants showed increased activity, as did most multi-mutants. Blue columns represent reactions with peptides, orange columns represent data without peptides. Error bars represent standard deviations from triplicate experiments.

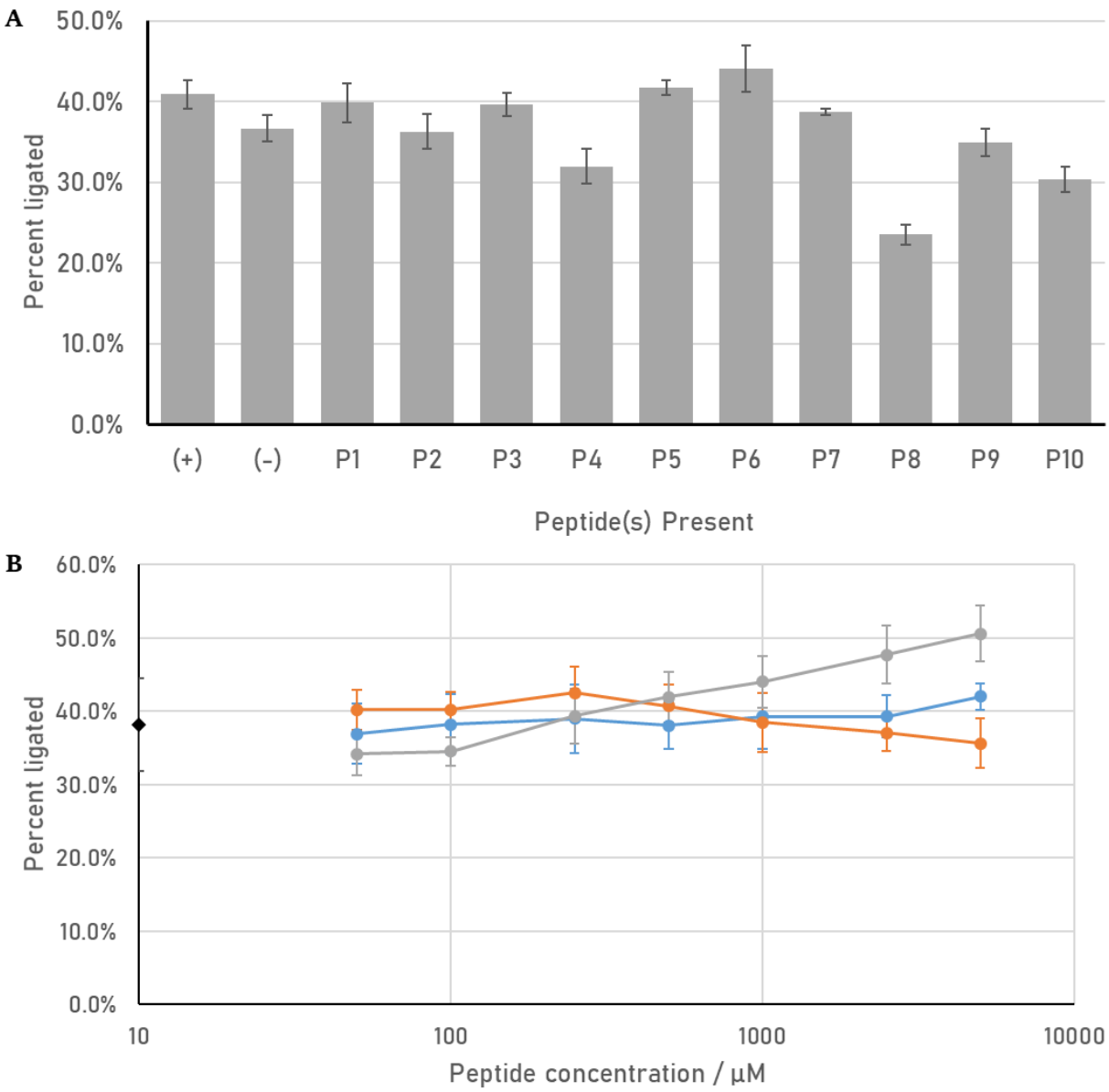

Figure 2.S5: Diminished peptide benefits for ribozyme 23 after 3 hours. **(A)** Screen of 10 individual peptides at 5mM in each reaction, with incubation time being 3 hours (20 minute reaction data shown in Figure 2.8A). (+) represents all peptides included at 0.5mM each (5mM total), (-) represents no peptides included. **(B)** Titration of peptides 1, 4, and 6 into the activity assay of ribozyme 23, with an incubation time of 3 hours (20 minute reaction data shown in Figure 2.8B). The activity of the ribozyme without peptides (black diamond) is shown on the y-axis. The activity in the presence of peptide 6 (grey), peptide 4 (orange), and peptide 1 (blue) are shown as function of the peptide concentrations. Error bars represent standard deviations from triplicate experiments.

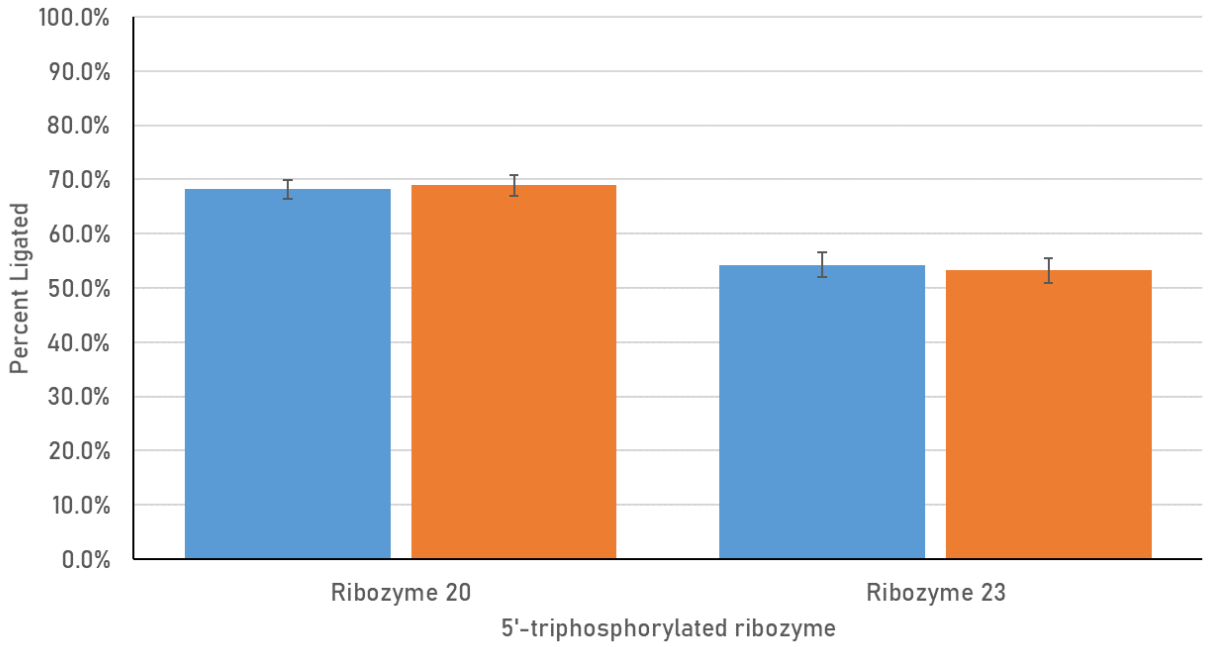

Figure 2.S6: Calibration of the ligation assay. The RNAs for this experiment were transcribed without a hammerhead ribozyme, meaning that the RNA was 100% triphosphorylated. Blue columns represent data in the presence of peptides, orange columns represent data in the absence of peptides. Error bars represent standard deviations from triplicate experiments.

References:

1. Rich, A. On the problems of evolution and biochemical information transfer. *Horiz. Biochem. New York, NY*, 103–126 (1962).

2. Woese, C. R. The genetic code. The molecular basis for genetic expression. *Harper Row N. Y.* (1967).

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

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

5. Gilbert, W. The RNA world. *Nature* **319**, 618 (1986).

6. Joyce, G. F. The antiquity of RNA-based evolution. *Nature* **418**, 214–21 (2002).

7. Higgs, P. G. Chemical Evolution and the Evolutionary Definition of Life. *J. Mol. Evol.* **84**, 225–235 (2017).

8. Krishnamurthy, R. & Hud, N. V. Introduction: Chemical Evolution and the Origins of Life. *Chem. Rev.* **120**, 4613–4615 (2020).

9. Rios, A. C. & Tor, Y. On the Origin of the Canonical Nucleobases: An Assessment of Selection Pressures across Chemical and Early Biological Evolution. *Isr. J. Chem.* **53**, 469–483 (2013).

10. Rios, A. C., Yu, H. T. & Tor, Y. Hydrolytic fitness of N -glycosyl bonds: comparing the deglycosylation kinetics of modified, alternative, and native nucleosides: HYDROLYTIC FITNESS OF N -GLYCOSYL BONDS. *J. Phys. Org. Chem.* **28**, 173–180 (2015).

11. Krishnamurthy, R. Role of pK(a) of nucleobases in the origins of chemical evolution. *Acc Chem Res* **45**, 2035–44 (2012).

12. Krishnamurthy, R. RNA as an Emergent Entity: An Understanding Gained Through Studying its Nonfunctional Alternatives. *Synlett* **25**, 1511–1517 (2014).

13. Soukup, G. A. & Breaker, R. R. Relationship between internucleotide linkage geometry and the stability of RNA. *Rna* **5**, 1308–25 (1999).

14. Islam, S. & Powner, M. W. Prebiotic systems chemistry: Complexity overcoming clutter. *Chem* **2**, 470–501 (2017).

15. Gavette, J. V., Stoop, M., Hud, N. V. & Krishnamurthy, R. RNA-DNA Chimeras in the Context of an RNA World Transition to an RNA/DNA World. *Angew Chem Int Ed Engl* **55**, 13204–13209 (2016).

16. Jiménez, E. I., Gibard, C. & Krishnamurthy, R. Prebiotic Phosphorylation and Concomitant Oligomerization of Deoxynucleosides to form DNA. *Angew. Chem. Int. Ed.* **60**, 10775–10783 (2021).

17. Guttenberg, N., Virgo, N., Butch, C. & Packard, N. Selection first path to the origin of life. *ArXiv170605831 Q-Bio* (2017).

18. Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D. & Sutherland, J. D. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat Chem* **7**, 301–7 (2015).

19. Islam, S., Bučar, D.-K. & Powner, M. W. Prebiotic selection and assembly of proteinogenic amino acids and natural nucleotides from complex mixtures. *Nat. Chem.* **9**, 584–589 (2017).

20. Xu, J., Green, N. J., Gibard, C., Krishnamurthy, R. & Sutherland, J. D. Prebiotic phosphorylation of 2-thiouridine provides either nucleotides or DNA building blocks via photoreduction. *Nat Chem* **11**, 457–462 (2019).

21. Lahav, N., White, D. & Chang, S. Peptide formation in the prebiotic era: thermal condensation of glycine in fluctuating clay environments. *Science* **201**, 67–9 (1978).

22. Forsythe, Jay G., Yu, Sheng-Sheng, Mamajanov, Irena, Grover, Martha A., Krishnamurthy, Ramanarayanan, Fernández, Facundo M., and Nicholas V. Hud. Ester-mediated amide bond formation driven by wet-dry cycles: A possible path to polypeptides on the prebiotic Earth. *Angew Chem Int Ed Engl* **54**, 9871–9875 (2015).

23. Parker, E. T., Cleaves, H. J., Bada, J. L. & Fernández, F. M. Quantitation of α-hydroxy acids in complex prebiotic mixtures via liquid chromatography/tandem mass spectrometry. *Rapid Commun. Mass Spectrom. RCM* **30**, 2043–2051 (2016).

24. Frenkel-Pinter, Moran, Haynes, Jay W., Mohyeldin, Ahmad M., C., Martin, Sargon, Alyssa B., Petrov, Anton S., Krishnamurthy, Ramanarayanan, Hud, Nicholas V., Williams, Loren Dean, and Luke J. Leman. Mutually stabilizing interactions between proto-peptides and RNA. *Nat. Commun.* **11**, 3137 (2020).

25. Cakmak, F. P., Choi, S., Meyer, M. O., Bevilacqua, P. C. & Keating, C. D. Prebiotically-relevant low polyion multivalency can improve functionality of membraneless compartments. *biorxiv* (2020) doi:10.1101/2020.02.23.961920v1.

26. Robertson, M. P. In vitro selection of ribozymes dependent on peptides for activity. *RNA* **10**, 114–127 (2004).

27. Tagami, S., Attwater, J. & Holliger, P. Simple peptides derived from the ribosomal core potentiate RNA polymerase ribozyme function. *Nat Chem* **9**, 325–332 (2017).

28. Poudyal, Raghav R., Guth-Metzler, Rebecca M., Veenis, Andrew J., Frankel, Erica A., Keating, Christine D., and Philip C. Bevilacqua. Template-directed RNA polymerization and enhanced ribozyme catalysis inside membraneless compartments formed by coacervates. *Nat Commun* **10**, 490 (2019).

29. Faulhammer, D. & Famulok, M. The Ca2+ ion as a cofactor for a novel RNA-cleaving deoxyribozyme. *Angew Chem Int Ed Engl* **35**, 2837–2841 (1996).

30. Stephenson, J. D., Popović, M., Bristow, T. F. & Ditzler, M. A. Evolution of ribozymes in the presence of a mineral surface. *RNA* **22**, 1893–1901 (2016).

31. Popovic, M., Fliss, P. S. & Ditzler, M. A. In vitro evolution of distinct self-cleaving ribozymes in diverse environments. *Nucleic Acids Res* (2015) doi:10.1093/nar/gkv648.

32. Pasek, M. A., Kee, T. P., Bryant, D. E., Pavlov, A. A. & Lunine, J. I. Production of potentially prebiotic condensed phosphates by phosphorus redox chemistry. *Angew Chem Int Ed Engl* **47**, 7918–20 (2008).

33. Pasek, M. A., Harnmeijer, J. P., Buick, R., Gull, M. & Atlas, Z. Evidence for reactive reduced phosphorus species in the early Archean ocean. *Proc Natl Acad Sci U A* **110**, 10089–94 (2013).

34. Herschy, B., Chang, S. J., Blake, R., Lepland, A., Abbott-Lyon, H., Sampson, J., Atlas, Z., Kee, T. P., Pasek, M. A. Archean phosphorus liberation induced by iron redox geochemistry. *Nat Commun* **9**, 1346 (2018).

35. Moretti, J. E. & Muller, U. F. A ribozyme that triphosphorylates RNA 5'-hydroxyl groups. *Nucleic Acids Res* **42**, 4767–78 (2014).

36. Dolan, G. F., Akoopie, A. & Muller, U. F. A Faster Triphosphorylation Ribozyme. *PLoS One* **10**, e0142559 (2015).

37. Pressman, A., Moretti, J. E., Campbell, G. W., Muller, U. F. & Chen, I. A. Analysis of in vitro evolution reveals the underlying distribution of catalytic activity among random sequences. *Nucleic Acids Res* **45**, 8167–8179 (2017).

38. Arriola, J. T. & Müller, U. F. A combinatorial method to isolate short ribozymes from complex ribozyme libraries. *Nucleic Acids Res.* **48**, e116–e116 (2020).

39. Parker, Eric T., Cleaves, Henderson J., Dworkin, Jason P., Glavin, Daniel P., Callahan, Michael, Aubrey, Andrew, Lazcano, Antonio, and Jeffrey L. Bada. Primordial synthesis of amines and amino acids in a 1958 Miller H2S-rich spark discharge experiment. *Proc Natl Acad Sci U A* **108**, 5526–31 (2011).

40. Trifonov, E. N. The triplet code from first principles. *J Biomol Struct Dyn* **22**, 1–11 (2004).

41. Georgiou, C. D. Functional properties of amino acid side chains as biomarkers of extraterrestrial life. *Astrobiology* **18**, 1479–1496 (2018).

42. Cobb, A. K. & Pudritz, R. E. Nature's starships. I. Observed abundances and relative frequencies of amino acids in meteorites. *Astrophys J* **783**, 12 (2014).

43. Hoffman, Michael M., Khrapov, Maksim A., Cox, J. Colin, Yao, Jianchao, Tong, Lingnan, and Andrew D. Ellington. AANT: the Amino Acid-Nucleotide Interaction Database. *Nucleic Acids Res* **32**, D174-81 (2004).

44. Mortimer, S. A. & Weeks, K. M. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J Am Chem Soc* **129**, 4144–5 (2007).

45. Herschlag, D. RNA chaperones and the RNA folding problem. *J Biol Chem* **270**, 20871–4 (1995).

46. Doetsch, M., Schroeder, R. & Fürtig, B. Transient RNA-protein interactions in RNA folding: Transient RNA-protein interactions in RNA folding. *FEBS J.* **278**, 1634–1642 (2011).

47. Szathmary, E. & Maynard Smith, J. From replicators to reproducers: the first major transitions leading to life. *J Theor Biol* **187**, 555–71 (1997).

48. Poole, A. M., Jeffares, D. C. & Penny, D. The Path from the RNA World. *J. Mol. Evol.* **46**, 1–17 (1998).

49. Noller, H. F. The driving force for molecular evolution of translation. *RNA* **10**, 1833–7 (2004).

50. Marqusee, S. & Baldwin, R. L. Helix stabilization by Glu-...Lys+ salt bridges in short peptides of de novo design. *Proc. Natl. Acad. Sci.* **84**, 8898–8902 (1987).

51. Jiang, F., Gorin A., Hu, W., Majumdar, A., Baskerville, S., Xu, W., Ellington, A., Patel, D. J. Anchoring an extended HTLV-1 Rex peptide within an RNA major groove containing junctional base triples. *Structure* **7**, 1461-S12 (1999).

52. Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L., Frankel, A. D., Williamson, J. R. Alpha Helix-RNA Major Groove Recognition in an HIV-1 Rev Peptide-RRE RNA Complex. *Science* **273**, 1547–1551 (1996).

53. Weeks, K. & Crothers, D. Major groove accessibility of RNA. *Science* **261**, 1574–1577 (1993).

54. Seeman, N. C., Rosenberg, J. M. & Rich, A. Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci.* **73**, 804–808 (1976).

55. Hammann, C. A spermidine-induced conformational change of long-armed hammerhead ribozymes: ionic requirements for fast cleavage kinetics. *Nucleic Acids Res.* **25**, 4715–4722 (1997).

56. Chowrira, B. M., Berzal-Herranz, A. & Burke, J. M. Ionic requirements for RNA binding, cleavage, and ligation by the hairpin ribozyme. *Biochemistry* **32**, 1088–1095 (1993).

57. Suh, Y.-A., Kumar, P. K. R., Taira, K. & Nishikawa, S. Self-cleavage activity of the genomic HDV ribozyme in the presence of various divalent metal ions. *Nucleic Acids Res.* **21**, 3277–3280 (1993).

58. Heilman-Miller, S. L., Pan, J., Thirumalai, D. & Woodson, S. A. Role of counterion condensation in folding of the Tetrahymena ribozyme II. Counterion-dependence of folding kinetics. *J. Mol. Biol.* **309**, 57–68 (2001).

59. Pace, U. & Szostak, J. W. Mutations in a semiconserved region of the *Tetrahymena* intron. *FEBS Lett.* **280**, 171–174 (1991).

60. Joyce, G. F., Horst, G. van der & Inoue, T. Catalytic activity is retained in the *Tetrahymena* group I intron despite removal of the large extension of element P5. *Nucleic Acids Res.* **17**, 7879–7889 (1989).

61. Peebles, C. L., Perlman, P. S., Mecklenburg, K. L., Petrillo, M. L., Tabor, J. H., Jarrell, K. A., Cheng, H.-L. A self-splicing RNA excises an intron lariat. *Cell* **44**, 213–23 (1986).

62. Marcia, M. & Pyle, A. M. Principles of ion recognition in RNA: insights from the group II intron structures. *RNA* **20**, 516–527 (2014).

63. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–57 (1983).

64. Olive, J. E. & Collins, R. A. Spermine Switches a *Neurospora* VS Ribozyme from Slow Cis Cleavage to Fast Trans Cleavage †. *Biochemistry* **37**, 6476–6484 (1998).

65. Hsiao, C., Mohan, S., Kalahar, B. K. & Williams, L. D. Peeling the onion: ribosomes are ancient molecular fossils. *Mol Biol Evol* **26**, 2415–25 (2009).

66. Bokov, K. & Steinberg, S. V. A hierarchical model for evolution of 23S ribosomal RNA. *Nature* **457**, 977–980 (2009).

67. Tang, G.-Q., Bandwar, R. P. & Patel, S. S. Extended Upstream A-T Sequence Increases T7 Promoter Strength. *J. Biol. Chem.* **280**, 40707–40713 (2005).

68. Jaeger, L., M.C. Wright, and G.F. Joyce, A complex ligase ribozyme evolved in vitro from a group I ribozyme domain. Proc Natl Acad Sci U S A, 1999. 96(26): p. 14712-7.

69. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406–15 (2003).

70. Edgar, Robert. *Usearch*. Lawrence Berkeley National Lab (LBNL), Berkeley, CA (United States), 2010.

Chapter 3: Lanthanide Cofactors for Triphosphorylation Ribozymes

Kevin Sweeney and Ulrich F. Müller

Abstract

The RNA world hypothesis describes a stage in the early evolution of life in which RNA served as the only genome and genome-encoded catalyst. Central catalytic activities in an RNA world organism would have included ribozymes that generate chemically activated nucleotides, which would then have served as monomers for RNA polymerization and self-replication. We previously identified hundreds of ribozymes that use the prebiotically plausible compound trimetaphosphate to generate chemically activated RNA 5'-phosphates, which showed that ribozymes are able to generate chemically activated nucleotides. The previous in vitro selection was performed in the presence of Mg^{2+} as the only multivalent cation. To test whether lanthanides are able to serve as cofactor for this reaction instead of Mg^{2+}, we performed an in vitro selection in the absence of Mg^{2+} and in the presence of Yb^{3+}. The best isolated ribozyme was able to catalyze the reaction with a maximum rate of 0.67 min^{-1}, and used Yb^{3+} specifically to interact with trimetaphosphate. The ribozyme was also able to use four other heavy lanthanides for the reaction, and it required much lower concentration of trimetaphosphate than the Mg^{2+}-using ribozymes. The findings are discussed in the context of the RNA world hypothesis.

Introduction

The RNA world hypothesis describes an early stage in the evolution of life in which RNA served as the genome and as the only genome-encoded catalyst[1,2,3,4]. To study how an RNA world organism could have functioned, researchers have developed catalytic RNAs (ribozymes) in the lab, using in vitro selection[5,6,7]. The central metabolites in an RNA world organism would have been chemically activated nucleotides to facilitate self-replication by RNA polymerization[8]. Such chemically activated nucleotides could have been nucleoside triphosphates (NTPs) because NTPs are the central energy currency in all known living organisms, because their involvement in diverse, fundamental metabolic and signaling

processes point to an ancient origin[9,10] and because there are chemical pathways that could have led to their prebiotic synthesis[11,12].

A prebiotic synthesis of NTPs could have been the reaction of nucleosides with the prebiotically plausible triphosphorylation reagent trimetaphosphate (Tmp)[11,12,13,14]. To test whether this chemistry could be catalyzed by ribozymes, we previously developed an in vitro selection to identify hundreds of ribozymes that catalyze the reaction between the 5'-hydroxyl group of their 5'-terminal nucleoside with trimetaphosphate, generating 5'-triphosphates[15,16]. The previous selection was performed with Mg^{2+} cofactors because Mg^{2+} is known to aid ribozyme catalysis for a wide range of activities[7,17,18,19,20,21], and because Mg^{2+} likely existed at concentrations around 10-20 mM in the prebiotic ocean[22,23].

Our interest to explore lanthanides as cofactors for triphosphorylation ribozymes was sparked by a publication from the Peters lab[24], where lanthanides were used to catalyze the hydrolysis of Tmp. The lanthanide (III) ion appeared to coordinate the three negatively charged oxygens of Tmp, and thereby activate the phosphorus for nucleophilic attack. Importantly, the coordination of the lanthanide with chelating ligands had a strong, enhancing effect on the catalysis rate, and different ligands (ethylenediaminetetraacetate; EDTA, and nitrilotriacetate; NTA) led to different rate enhancements. This situation is analogous to ribozyme-catalyzed triphosphorylation reactions: Triphosphorylation ribozymes catalyze the nucleophilic attack of an RNA 5'-hydroxyl group onto a phosphorus of Tmp, in part by the coordination of Mg^{2+} ions by the Tmp oxygens[15]. The ligands chelating the lanthanides for Tmp hydrolysis can be seen as analogs of ribozymes, which often chelate, and activate multivalent cations at the catalytic site[25,26,27,28,29]. However, lanthanides display excellent catalytic properties not only for mediating nucleophilic attacks on Tmp. Their optimal hydrolysis constants and their high water exchange rate constants make them excellent catalysts for a wide range of reactions[30], suggesting that they could be beneficial for multiple different reactions in RNA world organisms. Together, these arguments convinced us to explore the potential of lanthanides as cofactors for RNA triphosphorylation ribozymes.

The idea that lanthanide cofactors could have been used in an RNA world may not be far-fetched for two reasons. First, lanthanides are not as rare as the name 'rare earth elements' suggests: Their

abundance in the Earth's crust is in the range of several mass ppm, similar to cobalt and molybdenum[31,32,33]. Due to their unique chemical behavior, they are highly enriched in specific ores and pegmatites[34], leading to higher local concentrations. The low mobility of lanthanide minerals is increased by specific organic compounds in the environment[35], which could have made the lanthanides more accessible. Second, it has been recently discovered that some methylotrophic bacteria use lanthanides as cofactors for metabolic enzymes, especially methanol dehydrogenases, which are closely related to similar calcium-using enzymes[36,37,38]. Some organisms possessing both enzymes have been shown to preferentially use the lanthanide-using enzymes if lanthanides and calcium are both available, likely due to their stronger Lewis acid character and resulting greater catalytic potential[39]. The gene encoding the lanthanide dependent enzyme can be found in a variety of coastal areas[40], and other lanthanide-binding proteins have been identified as well[41].

To test whether triphosphorylation ribozymes could use a lanthanide as a cofactor, we performed an in vitro selection as described previously[15], but conducting the triphosphorylation reaction in the absence of Mg^{2+} and in the presence of the lanthanide Yb^{3+}. After five rounds of selection, about half of the isolated clones showed detectable activity with Yb^{3+}. The most active ribozyme, ribozyme 51, was characterized in detail. Ribozyme 51 used Yb^{3+} for catalysis but preferred other cations to assist in folding, and was very sensitive to the ionic radius of the lanthanide. Together, these results showed that ribozymes can use Yb^{3+} as a cofactor and demonstrated the unique relationship of a ribozyme with such a strong Lewis acid.

Results

In our first experiment, we tried to identify lanthanide-using ribozymes from libraries that were previously selected to use Mg^{2+} as cofactor[15]. These libraries contained hundreds of ribozyme clusters with unrelated sequences[16]. To identify lanthanide-using ribozymes from these libraries, they were subjected to four rounds of selection in the absence of Mg^{2+}, but in the presence of 3 mM Yb^{3+}. We chose

Ytterbium (Yb^{3+}) as lanthanide because Yb^{3+} has one of the smallest ion radius of the lanthanides, which gives it one of the highest charge densities and one of the strongest Lewis acidities[42]. Four selection rounds were previously sufficient to enrich active ribozymes from a pool with more than 10^{14} different sequences to dominate the pools[15,16]. Because the library size was in this case only a few hundred ribozyme clusters and the enrichment factor per selection round is at least 10^4, we expected that lanthanide-using ribozymes would dominate the pool within one or at most two selection rounds. However, even after four selection rounds there was no drop in the number of PCR cycles necessary to amplify the selected RNAs, which we use as a proxy for pool activity during the selection. When 18 individual clones were chosen arbitrarily from this selection round and tested biochemically, none of the clones showed any activity above background (~0.1% reacted per hour, data not shown). These results suggested that the previously selected library with hundreds of Mg^{2+}-using ribozymes did not contain a single clone that was able to use Yb^{3+} as cofactor.

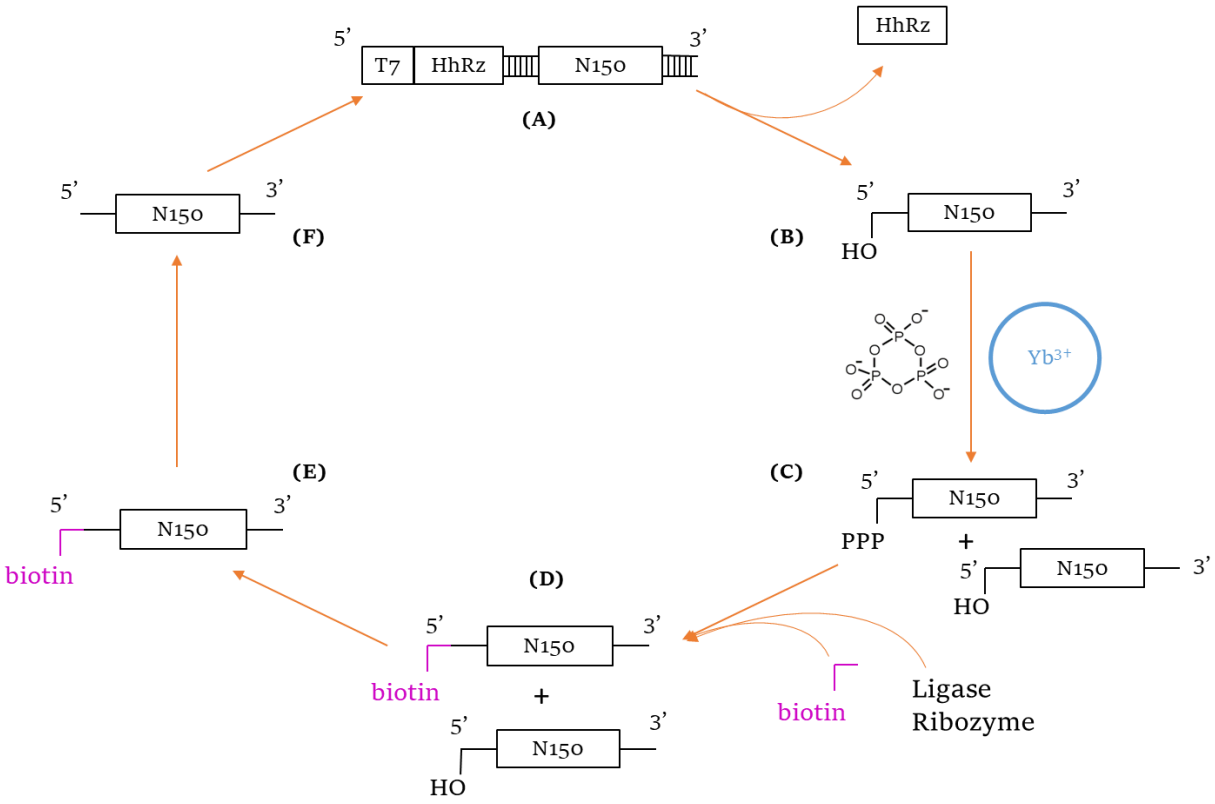

Figure 3.1: Scheme for the in vitro selection, based on the original of Moretti and Muller (2014)[15]. **(A)** Double-stranded pool DNA, which contains a T7 promoter and enhancer, a hammerhead ribozyme, and a 150 nucleotide randomized region flanked by 5' and 3' constant regions. The DNA is transcribed by T7 polymerase and the hammerhead ribozyme cleaves itself off cotranscriptionally, leaving behind **(B)** the pool RNA molecules with 5' hydroxyl groups. The RNAs consist of the randomized region and the constant regions. They are then incubated with trimetaphosphate and Yb^{3+} in a buffered solution, and **(C)** some RNAs will be able to coordinate trimetaphosphate and perform nucleophilic attack with the 5'-hydroxyl group. The whole pool is then incubated with a ligase ribozyme, which will react 5'-triphosphorylated RNAs with a biotinylated oligonucleotide to produce **(D)** biotinylated active pool RNA molecules. Streptavidin-coated magnetic beads are then used to separate the **(E)** active, biotinylated sequences from the inactive ones after stringent washing. The pool is then **(F)** reverse transcribed into cDNA, which is then PCR amplified to regenerate the starting double-stranded DNA pool.

To test more generally whether Yb^{3+} could be used by ribozymes as cofactor for triphosphorylation we conducted a new in vitro selection, starting from random sequence (Figure 3.1). The starting pool was the same for the Yb^{3+} cofactor selection as for the previously published Mg^{2+} cofactor selection. The effective library size was slightly larger (2.0×10^{14} as opposed to previously 1.7×10^{14}), because from the initial library with a complexity of 2.4×10^{14} sequences we effectively sampled 83 % during the first round of selection. After four rounds of selection, the pool appeared to be dominated by active sequences because the number of PCR cycles necessary to amplify the pool was 16 in the first

round without any 5'-triphosphorylated RNA added in the ligation step (round 4, see materials and methods) before decreasing to 13 for round 5 and 9 for round 6 (Figure 3.S1). After eight cycles of the selection, a ninth selection round in the absence of Yb^{3+}, half in the presence of Mg^{2+} and half in the absence of any multivalent cation, required 20 and 17 PCR cycles, respectively, suggesting that the vast majority of clones in the selection specifically required Yb^{3+} for activity. To identify clones with high activity, 29 clones were analyzed biochemically for triphosphorylation (Figure 3.2). When the 29 clones were tested for triphosphorylation under selection conditions, activity was barely detectable. To test whether this unexpectedly low activity was due to unfavorable reaction conditions, we tested the most active clone (clone 26) with and without added NaCl, and at pH values from 6.3 to 9.3 (Figure 3.S2). Surprisingly, the highest activity was seen with NaCl at pH 7.3, while the highest activity without NaCl was at pH 8.3. Therefore the reaction conditions to screen for activity were altered to 6 mM YbCl3, 6 mM Tmp, 50 mM HEPES/KOH pH 7.3, and 150 mM NaCl. When the 29 clones were tested again under these conditions, 19 showed detectable activity (Figure 3.2). One clone (clone 15) showed more than 10-fold higher activity than the second-best clone. Clone 15 was chosen for further analysis.

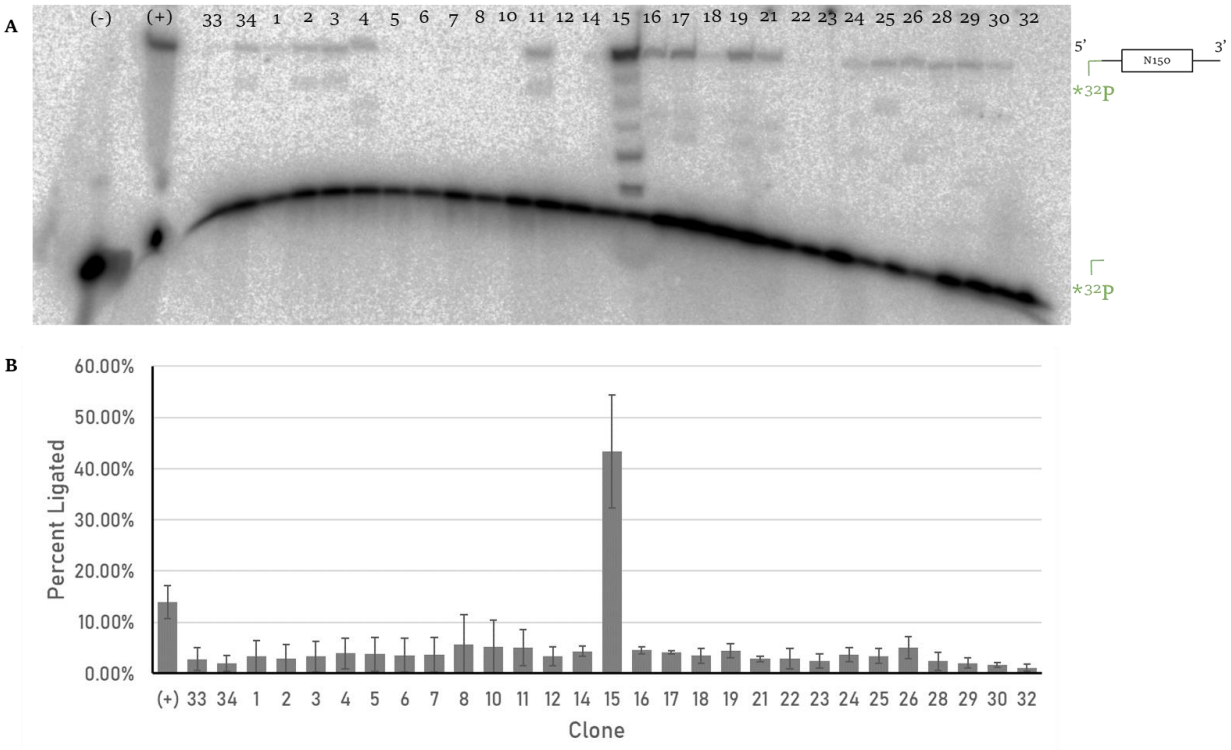

Figure 3.2: Screen of 29 clones from the eighth round of the selection. The ribozymes were transcribed from PCR products of individual clones. Each RNA was transcribed with a hammerhead ribozyme as in the selection, and then gel purified and incubated with trimetaphosphate and Yb^{3+} in a buffered solution. The ligation is to the same oligonucleotide as the selection process, but with a 5'-^{32}P instead of a biotin. (+) represents random pool RNA transcribed without a hammerhead ribozyme, thus having a 5'-triphosphate. Each number represents a different clone. **(A)** One of three polyacrylamide gels showing separation of the unligated radiolabeled oligonucleotides from those which were ligated to active ribozymes. (-) represents a lane loaded with unreacted radiolabeled oligonucleotide. **(B)** Graph of the data from triplicate experiments. Error bars represent standard deviation.

The initial isolates have a size of 181 nucleotides. To identify the minimal sequence of the ribozyme necessary for full activity, we truncated 80 nucleotides from the ribozyme 3'-terminus in 10-nucleotide increments (Figure 3.S3). Even the truncation of 10 nucleotides abolished activity, suggesting that the 3'-terminus of the 181 nt long RNA is required for activity. This finding matches the computationally predicted secondary structure (Figure 3.3A), which structure positions the ribozyme 3'-terminus close to the 5'-terminus, where catalysis takes place.

To test whether internal regions of the ribozyme could be deleted, we used the computationally predicted secondary structure from unafold[43] as guide. Indeed, the truncated ribozymes removing three different sections of the region between position 49 and 163 retained most of the activity of the full-length

ribozyme (Figure 3.3). However, the deletion of all three fragments in this region reduced activity by 75% relative to the full length ribozyme. This showed that the catalytic activity resided in the 5'-terminal and 3'-terminal ribozyme fragment, and that the role of the central region of the ribozyme could be fulfilled by multiple sequences. To determine whether this central region could be replaced by a shorter fragment, we inserted a sequence of 20 randomized nucleotides between position 53 and 159 of the ribozyme, and subjected this pool to four rounds of selection, with each successive round decreasing the concentration of Yb^{3+} and Tmp (see Materials and Methods). Each round required only 7-8 PCR cycles to amplify the selected sequences, suggesting that the majority of the pool was active. Twenty-three clones were arbitrarily chosen from the fourth round of this selection, and clone 51 (henceforth known as ribozyme 51) was determined to be most active (Figure 3.S4).

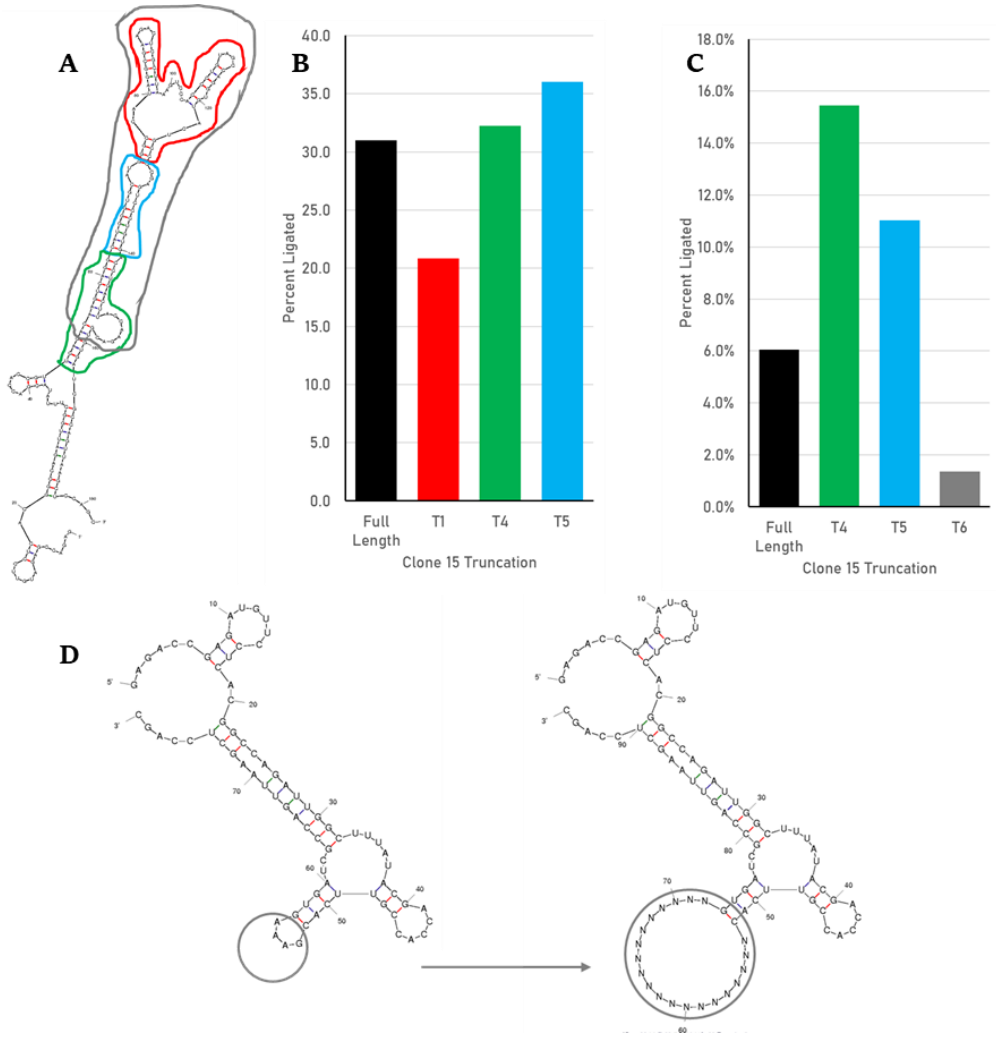

Figure 3.3: Truncation analysis of clone 15 leads to a re-selection. **(A)** Predicted secondary structure of clone 15 using unafold[43]. Circled in different colors are the different regions of the ribozyme removed for truncation activity analysis. Red corresponds to T1, green to T4, blue to T5, and gray to T6. **(B)** Data from the first activity assay of truncated variants of clone 15. Colors match (A). **(C)** Data from the second activity assay of truncated variants of clone 15, this time including T6 with most of the middle of the ribozyme removed. Colors match (A). **(D)** Predicted secondary structure of clone 15-T6 using mfold, and the same structure but with the circled GAAA tetraloop replaced by a 20 nucleotide randomized region (N20). This was used as the starting pool for the re-selection.

To identify the optimum reaction conditions for ribozyme 51, we began by co-varying the concentration of Yb^{3+} and Tmp (Figure 3.4). The highest maximum activity (58.4%) was seen at 3 mM Tmp and 0.3 mM Yb^{3+}, while the fastest rate (0.67 min^{-1}) was seen at 10mM Tmp and 1mM Yb^{3+} (full data shown in Figure 3.S5). Detectable activity was seen with concentrations of Yb^{3+} as low as 33µM. The ratio of Tmp:Yb^{3+} seemed to have an important influence; reactions with a ratio of 10:1 showed both the greatest maximum and the greatest rate. This suggests that the ribozyme binds the two simultaneously

as a complex instead of with two separate binding events. The sharp decrease in activity with greater than 10:1 Tmp:Yb^{3+} is likely because multiple Tmp molecules may then chelate one Yb^{3+}, which would make it inaccessible to the ribozyme.

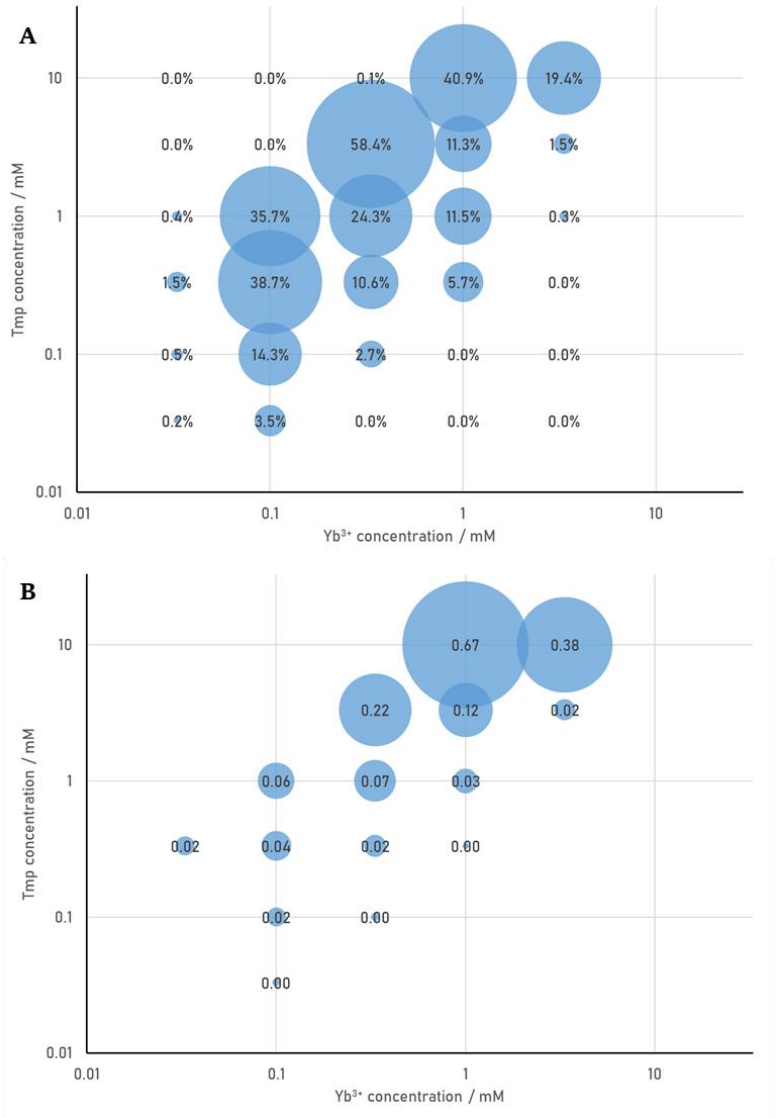

Figure 3.4: Ytterbium and trimetaphosphate dependence of ribozyme 51. All data is averaged from triplicate experiments. **(A)** Maximum activity of ribozyme 51 seen after 9 hour reaction time. **(B)** Rate of ribozyme 51 self-triphosphorylation. Rates are in min^{-1}, and were only calculated for reactions showing >1% maximum percent ligated. Full data for all reaction conditions in this experiment are shown in Figure 3.S5.

To test if ribozyme 51 requires Yb^{3+} specifically or can use any lanthanide, activity was tested with trivalent cationic salts of all lanthanides that could be readily obtained (Figure 3.5). Activity remained high with Lu^{3+}, but decreasing atomic mass correlated with decreasing activity, until no activity

was detected with Dy^{3+} or any lighter lanthanide (Figure 3.5A). The lanthanide contraction means that this correlates as well with the ionic radius of each lanthanide (Figure 3.5B). Together with the similar chemical behavior of the lanthanide series, this suggests that ribozyme 51 has a binding site that is very sensitive to the size of the lanthanide cofactor.

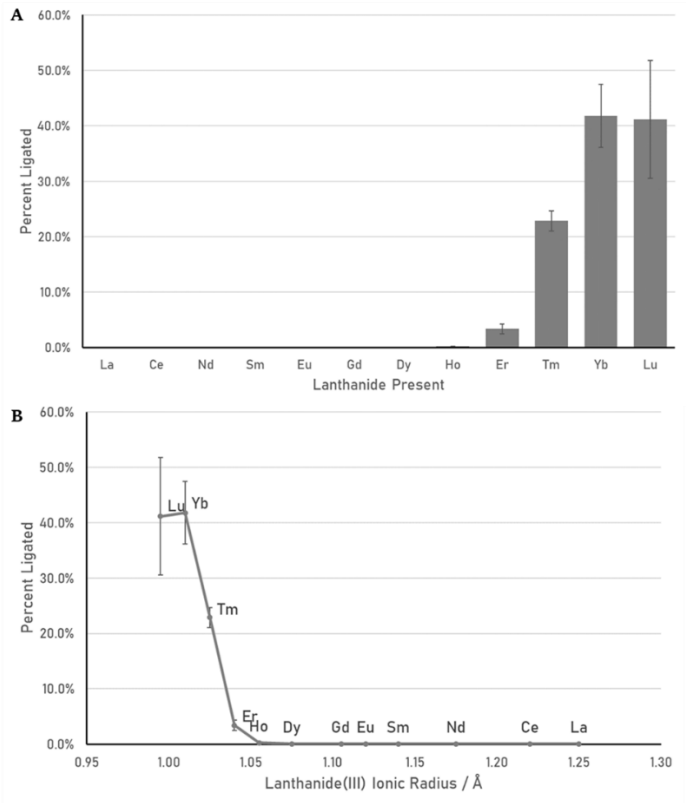

Figure 3.5: Ribozyme 51 activity using other lanthanide(III) cations. **(A)** Graph of ribozyme 51 activity with different lanthanide(III) cations, shown in order of increasing atomic mass. **(B)** Graph of ribozyme 51 activity against the ionic radius of the lanthanide(III) cations. Error bars represent standard deviations from triplicate experiments.

The other conditions for the ribozyme were then optimized by testing different temperatures, pH values, and concentrations of other cations (Figure 3.6). First, temperature was tested, showing an optimum of 25°C (Figure 3.6A). Activity was seen at a broad range of temperatures, with detectable activity even seen when the reaction mix and all components were kept on ice for the entire 20 minutes. The influence of pH on the activity of ribozyme 51 was examined next, and the optimal pH for activity was shown to be 6.8 (Figure 3.6B). There was no activity seen with this ribozyme at pH 7.8 or higher,

indicating that the ribozyme needs a neutral or acidic environment for activity. This pH dependence is very different from other self-triphosphorylating ribozymes, which were increasingly active with higher pH until the point of RNA hydrolysis (~pH 10). Future studies will analyze the cause of the decreased activity of ribozyme 51 at higher pH.

Ribozyme 51 was then tested for activity in the presence of various concentrations of Na^+, K^+, and Mg^{2+} individually (Figure 3.6C and 3.6D). In reactions with none of these ions added, the only cations present were the 9.9 mM sodium from the Na_3Tmp and the 330 μM Yb^{3+}, and activity was low (~5%). The selection had been done in the presence of 155 mM Na^+, but ribozyme 51 operated optimally at higher concentrations of monovalent cations, with the optimum being in the range of 400 mM. K^+ may have a slightly stronger effect than Na^+ because the ribozyme's optimum is broader with K^+. Mg^{2+} was also able to aid in the reaction, although at much lower concentrations (the optimum was at 5 mM) and with lower maximal activity. In order to test if the Mg^{2+} and the monovalent cations were aiding ribozyme 51 in the same way, we tested its kinetics in the presence of different combinations of 500 mM K^+, 5 mM Mg^{2+}, and 0.33 mM Yb^{3+} (Figure 3.6E). The ribozyme was inactive without Yb^{3+}, indicating that the other cations were not replacing Yb^{3+} in catalysis. Ribozyme 51 showed a >100-fold rate increase with either K^+ or Mg^{2+} as opposed to neither, but another ~2-fold rate increase and ~1.2-fold higher maximum activity with both together. This indicates that while K^+ and Mg^{2+} are likely fulfilling similar roles for the ribozyme, they are both necessary for optimal activity. We hypothesized that Yb^{3+} was bound at the catalytic site as a complex with trimetaphosphate, while K^+ and Mg^{2+} were important for the folding of the ribozyme.

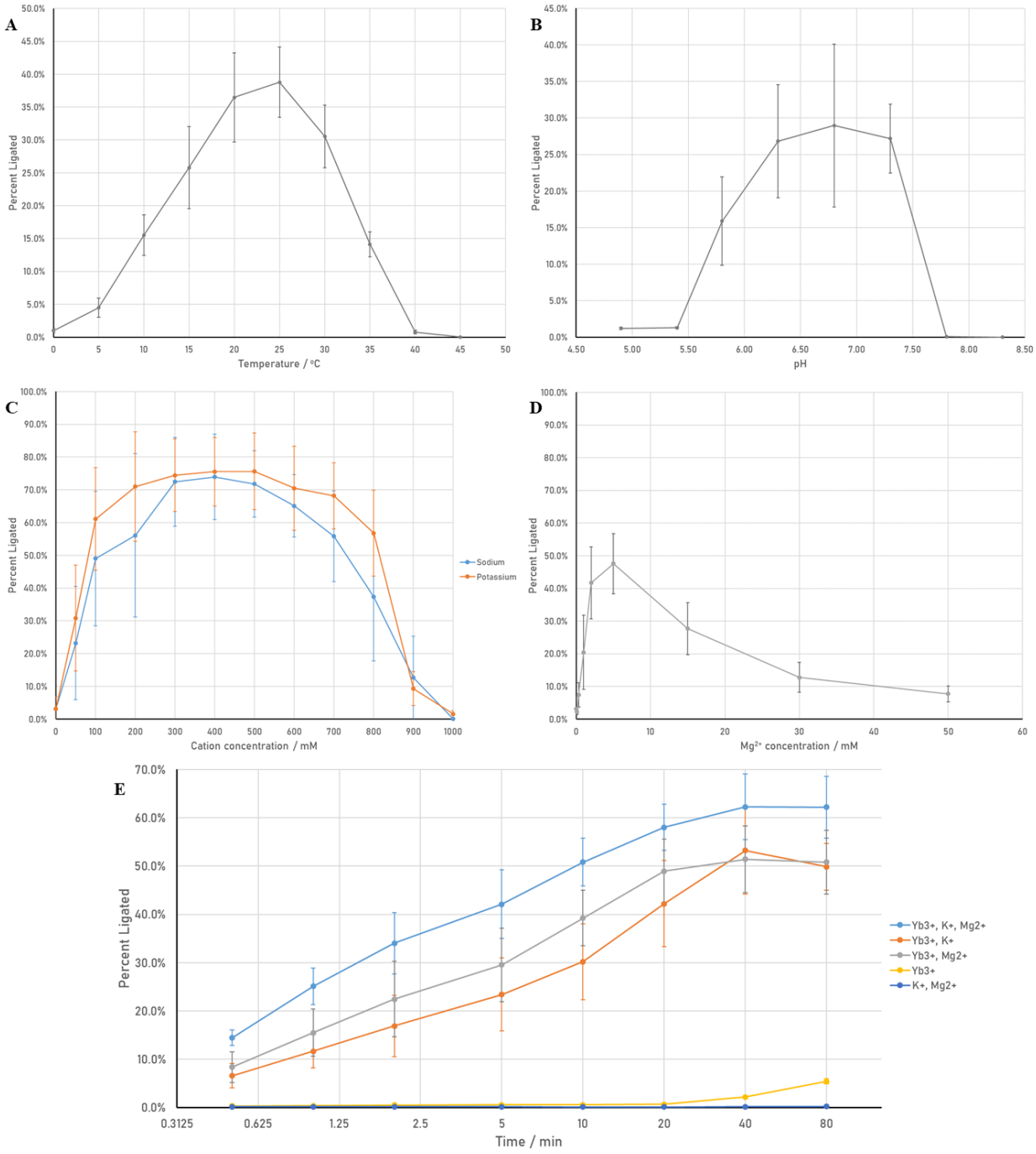

Figure 3.6: Optimization of reaction conditions for ribozyme 51. All data shown is from triplicate experiments, and error bars represent standard deviations. **(A)** Activity of ribozyme 51 tested in 5ºC increments from 0ºC (reacted on ice) to 50ºC. **(B)** Activity of ribozyme 51 at pH 4.9 and 5.3 (NaOAc/HOAc), 5.8 and 6.3 (MES/NaOH), 6.8 (MOPS/NaOH), 7.3 and 7.8 (HEPES/NaOH), and 8.3 (Tris/HCl). **(C)** Activity of ribozyme 51 with varying concentrations of KCl or NaCl added to the triphosphorylation reaction. Colors are indicated with the legend to the right. **(D)** Activity of ribozyme 51 with varying concentrations of $MgCl_2$ added to the triphosphorylation reaction. **(E)** Activity of ribozyme 51 over time with different combinations of salts (500mM KCl, 5mM $MgCl_2$, and $Yb(Tf)_3$). Colors are indicated with the legend to the right.

In order to test how the ions affected the folding of ribozyme 51, we used SHAPE[44] to examine the ribozyme's secondary structure (Figure 3.7). The first SHAPE experiments we conducted were to identify the influence on the secondary structure of Yb^{3+}/Tmp binding (Figure 3.7A), and predicted structural changes are shown (Figure 3.7C). These reactions contained 400mM K^{+} and no Mg^{2+}, and varied in containing Yb^{3+} and Tmp, just Yb^{3+}, just Tmp, or neither. Despite minor differences, the accessibility of nucleotides in both conditions without Yb^{3+} appear largely the same, whether Tmp was present or not (data with just Yb^{3+} was inconsistent and is shown in Figure 3.S6A). This suggests that the ribozyme 51 only interacts with trimetaphosphate via coordination with Yb^{3+}. This matches the earlier activity data titrating Tmp and Yb^{3+}. When comparing the SHAPE data from reactions without Yb^{3+} to the reaction with Yb^{3+} and Tmp, multiple regions seem to exhibit structural changes. Loop 2, from nucleotides 25-29, appears to be less accessible, suggesting that this loop is important for binding the complex. Additionally, the reaction with Yb^{3+} and Tmp showed increased accessibility of nucleotides 34, 35, and especially 37, but lower accessibility of nucleotide 36. Therefore, this region (34-37) seems to be involved in binding Yb^{3+} and Tmp as well. Additional major changes are seen throughout the helix pairing nucleotides 40-59 with 64-85. When Yb^{3+} is present, the otherwise very accessible nucleotides 58-60 become more protected, leaving only a mild peak at nucleotides 59-61. Instead, nucleotides on both sides of that long helix (47-55 and 63-70) become more accessible. These regions do not base pair with each other in our predicted structure so we cannot say if the whole helix is becoming more flexible, but this is part of the region that was re-selected (53-72), so it seems that the re-selection improved Yb^{3+}/Tmp binding. The comprehensive set of SHAPE data can be found in Figure 3.S6.

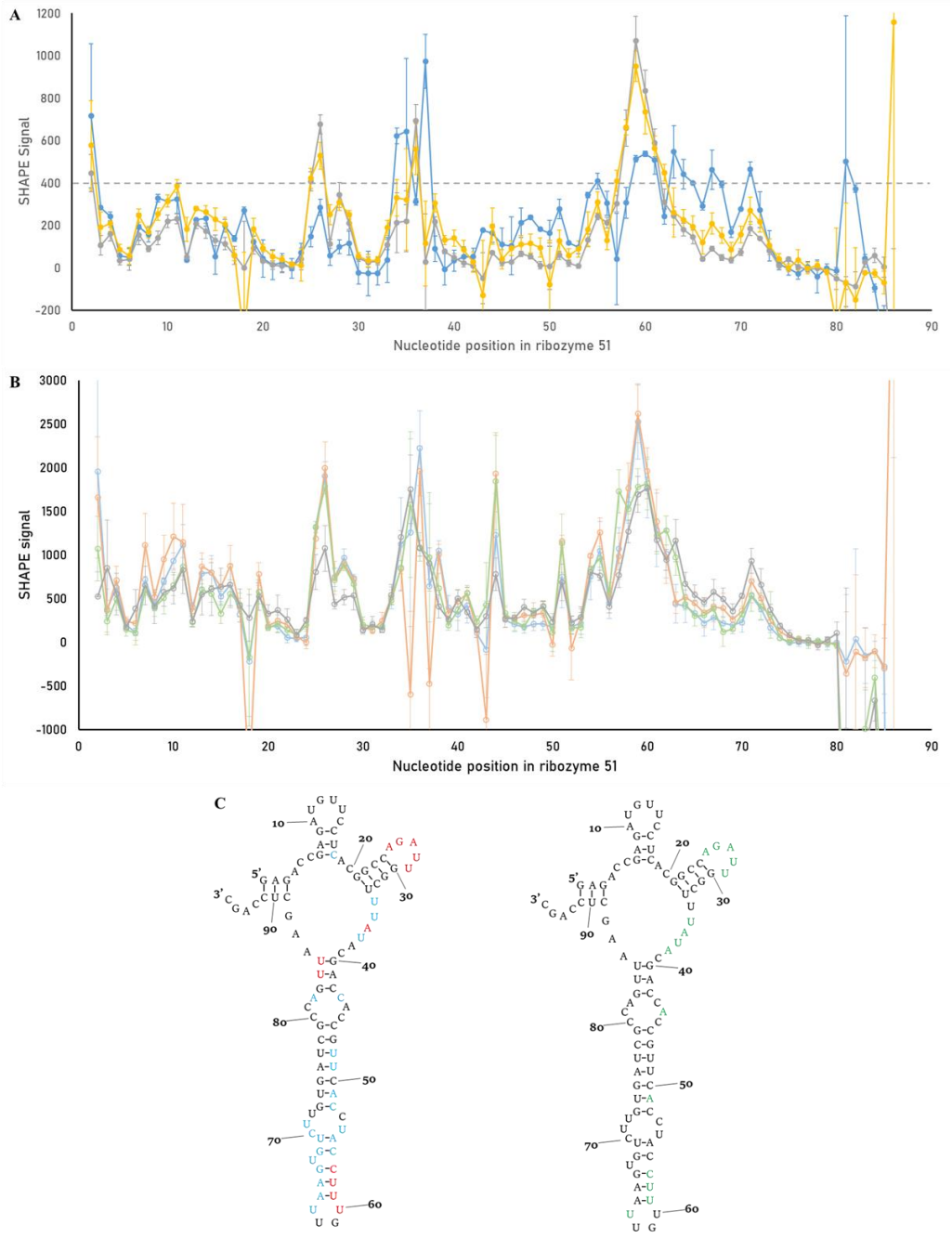

Figure 3.7: Secondary structure analysis of ribozyme 51 using SHAPE probing[44]. Error bars represent standard deviations from triplicate experiments. **(A)** Experiments testing the effect of Yb^{3+} and Tmp on ribozyme secondary structure. The blue line represents data with both present, the gray line represents data with just Tmp present, and the yellow line represents data with neither present. **(B)** Experiments testing the effect of K^{+} and Mg^{2+} on ribozyme secondary structure with no Yb^{3+} or Tmp present. The blue line represents data with both present, the orange line represents data with just K^{+}, the green line represents data with just Mg^{2+}, and the gray line represents data with neither. **(C)** Secondary structure prediction showing the influence of different reaction components. The structure on the left shows the influence of Yb^{3+} and Tmp, with blue nucleotides being more accessible with those components and red nucleotides being more accessible without them. The right structure shows the effects of K^{+} and Mg^{2+}, with green nucleotides representing regions with a strong influence. Full SHAPE data shown in Figure 3.S6.

To test the influence of K^+ and Mg^{2+} on the structure of ribozyme 51, we conducted further SHAPE experiments (Figure 3.7B). Shown are the data from the four different conditions (with 500mM K^+ and 5mM Mg^{2+}, with just K^+, just Mg^{2+}, or with neither) not including Yb^{3+} or Tmp, which clearly indicate some structural changes (the data with Yb^{3+} and Tmp is shown in Figure 3.S6B). Most notably, loop 2 (25-29) was less accessible when neither K^+ nor Mg^{2+} were present. This suggests that without folding assistance from other multivalent cations, this stem (21-24, 30-33) and loop were less structured, and probably less able to bind the Yb^{3+}/Tmp complex. Additionally, the peak in accessibility at nucleotide 59 was only present when K^+ is included regardless of Mg^{2+} presence, allowing insight into a benefit that may be specific to K^+. These findings matches the activity data showing that the ribozyme can be active with Yb^{3+} and no other cations, but that K^+ and Mg^{2+} have an important influence in forming the active structure of the ribozyme.

To better understand the importance for Yb^{3+}/Tmp binding and activity of the randomized region that had been added for the re-selection (nucleotides 53-72 Figure 3.3), activity of ribozyme 51 and clone 15-T6 were tested under optimized conditions (50 mM MOPS-NaOH pH 6.8, 500 mM KCl, and 5 mM $MgCl_2$, and a 10:1 molar ratio of Yb^{3+}:Tmp) across a range of Yb^{3+}/Tmp concentrations (Figure 3.S7). The difference was minor, indicating that the change in conditions rescued activity for this shorter sequence as well, and that the role of the re-selected region was relatively minor. However, this benefit was greater when the concentration of Yb^{3+} and Tmp were further removed from the optimum (now 0.1mM and 1mM, respectively), suggesting that the re-selected region does improve the ribozyme's ability to bind the Yb^{3+}/Tmp complex. Additionally, this experiment shows that the ribozyme can show detectable activity with as little as 18 µM Yb^{3+} and 180 µM Tmp, showing that it is very sensitive to those species.

Discussion

In this study, we have described the first ribozyme that uses a lanthanide as a catalytic cofactor, which was identified using an in vitro selection for self-triphosphorylating ribozymes. There are three

points of evidence for the lanthanide's participation in catalysis: First, it was shown previously that lanthanides strongly coordinate the three negatively charged oxygens of Tmp and thereby help the nucleophilic attack on the phosphorus atoms. Second, the ribozyme is unable to catalyze the self-triphosphorylation in the absence of a lanthanide(III) cation, even in the presence of other cations, including Mg^{2+}. Thirdly, Tmp/Yb^{3+} concentration dependence shows that there is a co-dependence of Tmp concentration and Yb^{3+} concentration, which would be expected if Tmp and Yb^{3+} form a complex at the catalytic site but not if the Yb^{3+} is necessary only for structural roles in ribozyme folding.

The biochemical relevance of lanthanides is already being investigated: the recent investigations of lanthanide-using methanol dehydrogenases (MDHs) have demonstrated the biological importance of lanthanides for metabolism in methanotrophs[36,37,38]. These enzymes are not the only ones using lanthanides[40,41], however they are the best-studied. They are closely related to similar enzymes that use calcium as the metal ion cofactor, but one species has been shown to preferentially use the lanthanide-dependent enzymes if the ions are available in the media[39]. They also require much lower concentration of lanthanide than the calcium-dependent enzymes, and these two observations are likely both related to the exceptional Lewis-acidity of lanthanides. This is also reflected in ribozyme 51 described in this work, as the highest maximum activity is observed in the range of 100-330 μM of Yb^{3+}, and increasing beyond that decreases the ribozyme's maximum activity despite increasing the rate. It is possible that this is due to misfolding caused by the lanthanide's strong electrophilicity, as the SHAPE experiments show large error bars both when Yb^{3+} is present without Tmp or without any other added cations (Figure 3.S6A, 3.S6G). Another similarity between ribozyme 51 and the lanthanide-dependent MDHs is the size-selectivity of the lanthanide binding. The lanthanides are very similar in coordination geometry and chemistry, leaving ionic radius as the biggest difference across the series[42]. Ribozyme 51 effectively uses only the heaviest lanthanides, showing catalytic activity with Ho-Lu, while most lanthanide-dependent MDHs seem to utilize only the lighter lanthanides[45].

A similar size-selectivity is seen in minerals where lanthanides can crystallize together[34], but also in lanthanide-dependent DNAzymes that have been generated by in vitro selections[46]. In 2013 it was

demonstrated that a previously generated RNA-cleaving DNAzyme could use lanthanides as cofactors[47], after which multiple in vitro selections have been done identifying DNAzymes that depend specifically on lanthanides[48,49,50]. Ours is the first ribozyme known to use lanthanides as cofactors.

Given that lanthanides are such strong Lewis acids compared to other metal ions, it is not surprising that a modern organism can preferentially use lanthanide cofactors when they are available, but switch to other ions like calcium when lanthanides are not available[39]. If an RNA world organism could do the same, that could enhance its evolutionary fitness. However, given that ytterbium-using ribozymes did not emerge more quickly from the selection than magnesium-using ribozymes in the original triphosphorylation ribozyme selection using the same random pool, we speculate that possible lanthanide-using ribozymes in an RNA world would not have expanded outside small niches. There are also other cations that early ribozymes may have used for catalysis that are not still used by modern ribozymes, including Fe^{2+}, although many metal ions can be used by ribozymes as cofactors[51,52]. It is possible that a ribozyme from the selection in this study could also use other cofactors, and future evolution experiments could be conducted to explore that idea. Given that at the end of the selection, the pool seemed slightly more active with Mg^{2+} than without any multivalent cation (Figure S1), it is possible that a sequence from the pool other than ribozyme 51 is able to catalyze the reaction with both Yb^{3+} and Mg^{2+}. Alternatively, the sequence may not use any multivalent cation for catalysis, instead just using Mg^{2+} to aid in folding into the active structure. This study contributes to the growing body of literature supporting the benefits of cations for ribozymes.

Materials and Methods

In Vitro Selection

The in vitro selection was performed essentially as described[15], with one important modification: the triphosphorylation reaction step was performed with 3mM Yb^{3+} instead of Mg^{2+} and the trimetaphosphate concentration was brought down to 10mM. Yb^{3+} concentrations only up to 3 mM were

used in this study because up to this concentration, no significant RNA degradation was seen at neutral pH within 3 hours of incubation; some degradation was visible at 10 mM Yb^{3+} (data not shown).

The starting DNA pool molecules consisted of a T7 RNA polymerase enhancer and promoter, a hammerhead ribozyme, a 5'-constant region, 150 randomized nucleotides, and a 3'-constant region. The library was initially synthesized in part as custom ultramer^{R} single-stranded DNA (Integrated DNA Technologies, IDT) with the sequence 5'-**GCTGGAGCTTAACTGGCG**-(N150)-**AACATCTCGGTCTC**GACTG-3' (lower strand) using hand-mixing for the randomized region. The double-stranded pool for the start of the selection used this as the template for PCR, with 5' primer 5'-*AATTTAATACGACTCACTATA*GGGCGGTCTCCTGACGAGCTAAGCGAAACTGCGGAAACGCAGT CGAGACCGAGATGTT-3' and 3' primer 5'-**GCTGGAGCTTAACT**-3'. In all sequences, the italicized portion is the T7 promoter and enhancer, the underlined portion is the hammerhead ribozyme, and the bold regions are the constant regions. After transcription and PAGE purification, the pool RNA was incubated at 100nM with 50mM Tris/Hcl pH8.3, 3mM $Yb(Tf)_3$, 10mM Na_3Tmp, 150mM NaCl (to compensate for the decreased sodium relative to the original selection due to severely decreased Na_3Tmp concentration), and 5mM NaOH (to account for the pH drop due to Tmp chelating metal ions). The incubation was performed in 100mL volume for the first round. After ethanol precipitation, water was added in 50μL increments to remove salt until RNA was detected by A260. Then the pellet was dissolved in water and desalted by size exclusion chromatography (P30 spin-columns, Bio-Rad) and ethanol precipitated.

The recovered RNA was ligated to a biotinylated oligonucleotide (5'-biotin-d(GAACTGAAGTGTATG)rU-3') using R3C ligase ribozyme[53]. The annealing step of this reaction contained 800nM pool RNA, 1000nM ligase ribozyme, and 1200nM biotinylated oligonucleotide in addition to 100mM KCl, 100mM Tris/Hcl pH8.5, and 2pM of random pool RNA transcribed without a hammerhead ribozyme, meaning that it had a 5'-triphosphate. This extra RNA was included for only the first 4 rounds because before the active sequences dominated the pool, the high number of PCR cycles required to amplify at the end of those rounds would have generated many unwanted side-products. The

ligation products were bound to streptavidin-coated magnetic beads (Promega) while other RNAs were washed away. The bound RNA was eluted with 96% formamide at 65ºC for 5 minutes, followed by ethanol precipitation. The RNA was then reverse transcribed with the 3' PCR primer using Superscript III Reverse Transcriptase (Invitrogen). RT products were then PCR amplified twice. The first PCR used 5' primer 5'-GAACTGAAGTGTATGTGAGACCGAGA-3' and 3'-primer 5'-GCTGGAGCTTAACT-3', then the second PCR used the same 3' primer with the 5' primer from the beginning of the selection that adds the T7 promoter and hammerhead ribozyme. This completes one round of selection, and the number of PCR cycles required to see a clear band on an agarose gel in the first PCR for each round was used to track the progress of the selection.

The 4 rounds of re-selection were performed similarly, but with a few minor differences. First, the pH of the incubation with trimetaphosphate was lowered to 7.3 and the Tmp and Yb^{3+} concentrations were adjusted to 6mM each after an experiment suggested that this improved activity (Figure 3.S2). For each successive following round, the concentrations of both were lowered to 3mM each, then to 1mM each, and finally to 0.3mM each.

Self-triphosphorylation activity assays

To measure the self-triphosphorylation activity of ribozymes from this selection, an assay was used which resembled the first part of an in vitro selection round. Instead of a random pool, a specific DNA sequence from the PCR product of an individual clone was transcribed using the same T7 promoter for runoff transcription. The sequence also included the hammerhead ribozyme to leave behind a 5'-hydroxyl group on the ribozyme RNA, and the RNA was purified by denaturing PAGE after transcription. The standard triphosphorylation reaction, after changing conditions after the selection but before optimizing for ribozyme 51 specifically, included 50 mM Tris pH 7.3, 150 mM NaCl, 3 mM $Yb(Tf)_3$, 10 mM Na_2Tmp, and 5 µM ribozyme RNA. These solutions would be incubated at room temperature (~22 ºC) for 3 hours, after which 2 µL would be removed and added to 8 µL of another solution such that the 10 µL mix would contain 100 mM Tris pH 8.0, 100 mM KCl, 5.6mM Na_2EDTA (a 5 mM excess over

Yb^{3+}), 1 μM of R3C ligase ribozyme, 1 μM of the ribozyme RNA, 1 μM of the oligonucleotide 5'-d(GAACTGAAGTGTATG)rU-3', and 10,000 counts per minute of the same oligonucleotide with a 5'-^{32}P radiolabel (this was prepared by using polynucleotide kinase (NEB) and γ-^{32}P ATP (Perkin-Elmer), followed by denaturing PAGE purification). This mix was heated to 65ºC, then cooled at ~0.1ºC/second down to 30ºC in order to allow the arms of the ligase ribozyme to anneal to the triphosphorylation ribozyme and the short oligonucleotide. Then equal volume of a solution containing 50mM $MgCl_2$, 4mM Spermidine, and 40%(m/v) PEG 8000 was added, resulting in the concentrations of all components being ½ of that listed. This solution was incubated for 3 hours, then stopped by ethanol precipitation. The products were separated on 10% denaturing PAGE for 45 minutes, then exposed overnight (~10 hours) to phosphorimaging screens. The screens were then scanned on a Typhoon (GE) phosphorimager, and the bands were quantified using the rectangle tool in Quantity One (Bio-Rad). The fraction ligated was calculated as the volume of the ligated band divided by the volume of the sum of the ligated and unligated bands. Because the short radiolabeled oligonucleotide, ligase ribozyme, and triphosphorylation ribozyme were all at equimolar concentrations during the ligation reaction, the results of this assay can report the degree of triphosphorylation.

SHAPE probing

First, a 49 μL solution was made of 50 mM MOPS/KOH pH 6.8, the specified amounts of KCl and $MgCl_2$, and 330 μM $Yb(Tf)_3$ and 3.3mM trimetaphosphate (if included) along with 200nM ribozyme 51. This was then mixed with 1 μL of a solution with either 100mM 1M7[44] in dry DMSO or a control of just DMSO. This was incubated at 25ºC for three minutes, followed by ethanol precipitation. The RNA was then reverse transcribed using Superscript III Reverse Transcriptase (Invitrogen) and an 8 nucleotide 5'-^{32}P radiolabeled DNA primer complementary to the ribozyme's 3' terminus. The length of 8 nucleotides was chosen after a pilot experiment showed that this was the shortest primer that offered full extension (data not shown). RT products were heated to hydrolyze RNA at 80ºC in 750mM NaOH for five minutes, followed by ethanol precipitation. They were then separated on 7M urea, 10%

polyacrylamide gels for 3 or 5 hours. Exposures of phosphorimager screens were scanned on a Typhoon Phosphorimager (GE). Each of the bands was quantified using the rectangle tool in Quantity One (Bio-Rad). For each band, the volume in the control DMSO lane was subtracted from the volume in the lane with 1M7. Based on the RT primer length and ribozyme length, the bands were assigned to positions in the ribozyme, generating a profile for each experiment. Each profile was normalized, and the averages and standard deviations from three experiments are reported. The cutoffs chosen between weak and strong signals are indicated for each condition. The secondary structures are based on computational predictions by unafold[43], including the constraints of the nucleotides with strong signals in the SHAPE experiments.

Acknowledgements:

Zhaleh N. Amini and Gregory F. Dolan are thanked for conducting pilot experiments.

Chapter 3, in full, is currently being prepared for submission for publication of the material. Sweeney, Kevin J.; Muller, Ulrich F. The dissertation author was the primary investigator of this material.

Supplementary Figures:

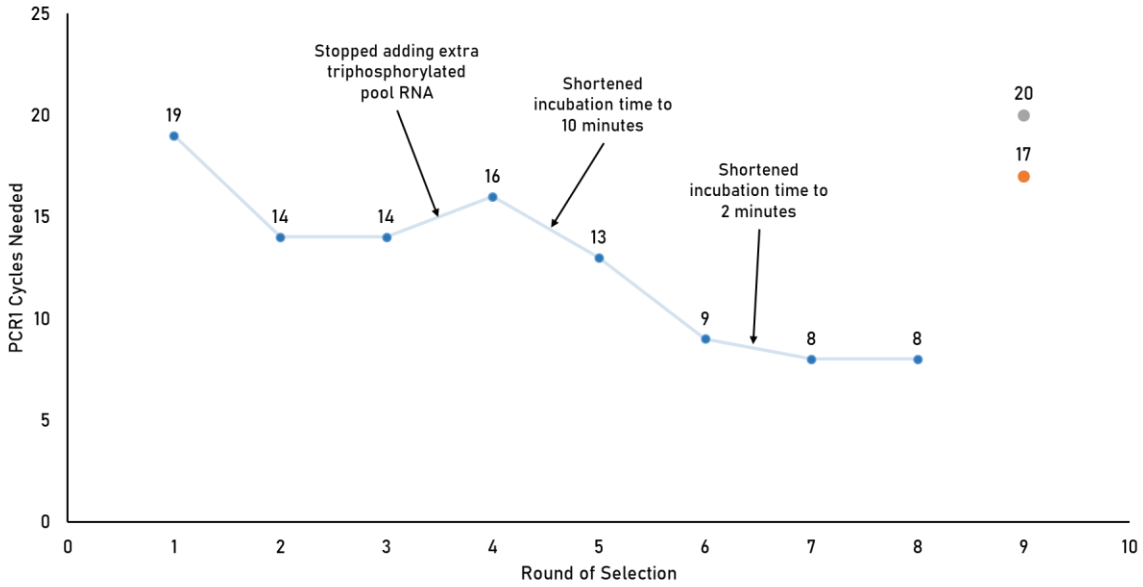

Figure 3.S1: Progress of the lanthanide selection beginning with a random pool, measured by PCR cycles required after each round. The round 4 ligation was the first of the selection not to include ~2pM 5'-PPP pool RNA, so the fact that only 16 PCR cycles were required indicated that the pool was then dominated by active sequences. Starting in round 5, the incubation time was shortened to 10 minutes, and for round 7 it was shortened to 2 minutes, both times to encourage the selection of ribozymes with faster rates. Two separate round 9s were performed – one with Mg^{2+} replacing Yb^{3+} (PCR cycles required represented in orange), and one with no multivalent cation in the triphosphorylation reaction (PCR cycles required represented in gray).

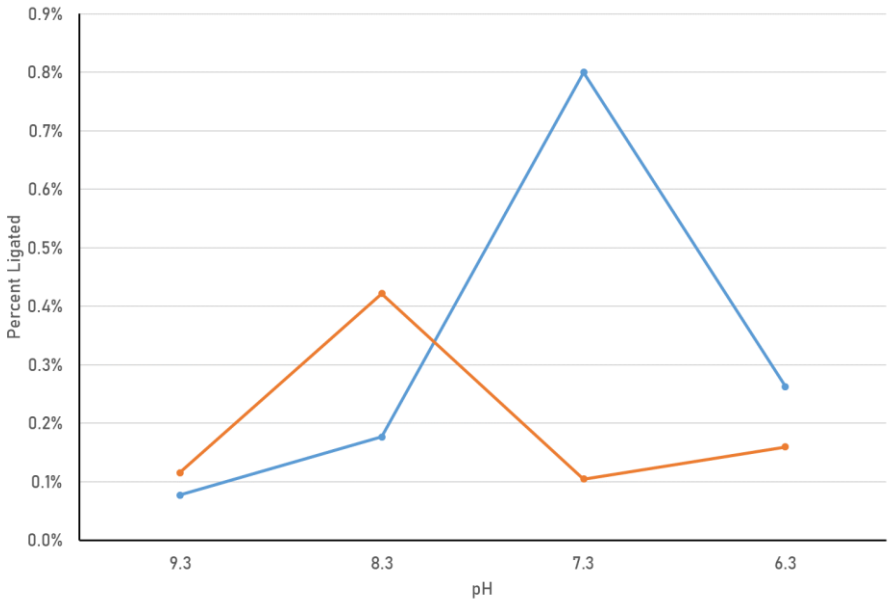

Figure 3.S2: Optimizing the pH and NaCl concentration for activity of clones after 8 rounds of in vitro selection with Yb^{3+}. The results are shown from a single experiment. The blue line represents data from reactions with 150mM NaCl, and the orange line represents data from reactions with no added NaCl (the only sodium present was from the Na_3Tmp). The original selection conditions were 150 mM NaCl and pH 8.3.

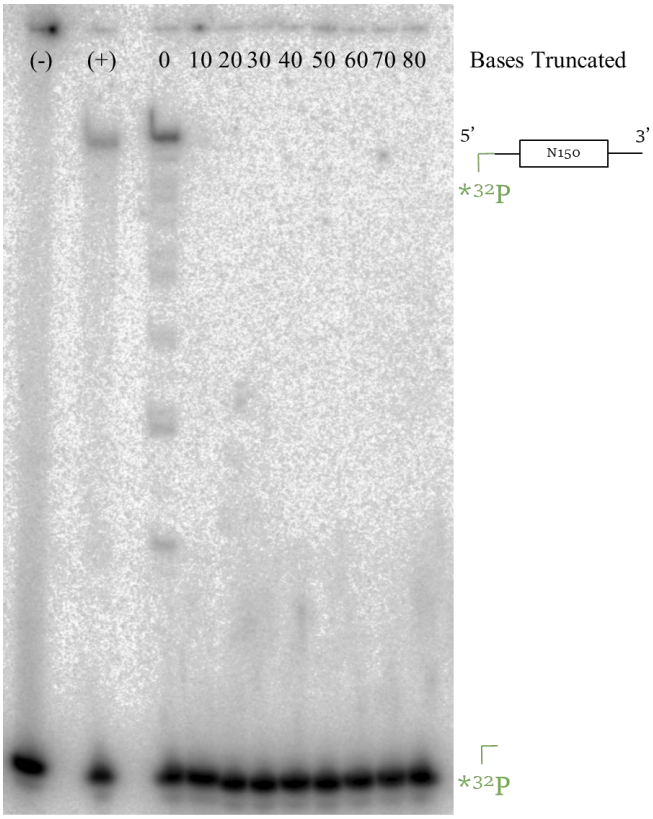

Figure 3.S3: 3'-Truncations of clone 15 are completely inactive. Shown is a PAGE separation of activity assays with ribozyme 51 full length (181 nucleotides, shown as 0 bases truncated), along with constructs truncated at the 3' end in 10-nucleotide increments down to 80 nucleotides removed. (-) represents a lane containing unreacted, radiolabeled oligo. (+) represents random pool RNA transcribed without a hammerhead ribozyme, meaning the 5' ends are 100% triphosphorylated.

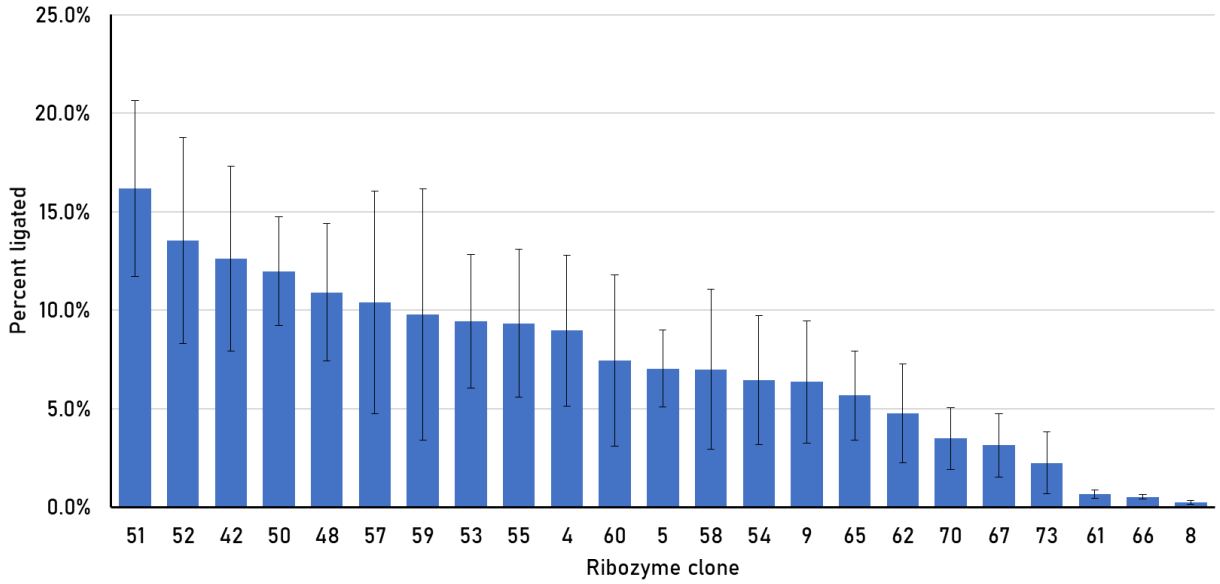

Figure 3.S4: Screen of 23 clones for activity after the N20 re-selection based on clone 15-T6 from the original selection (see Figure 3.3). All data are results from triplicate experiments, and error bars represent standard deviations.

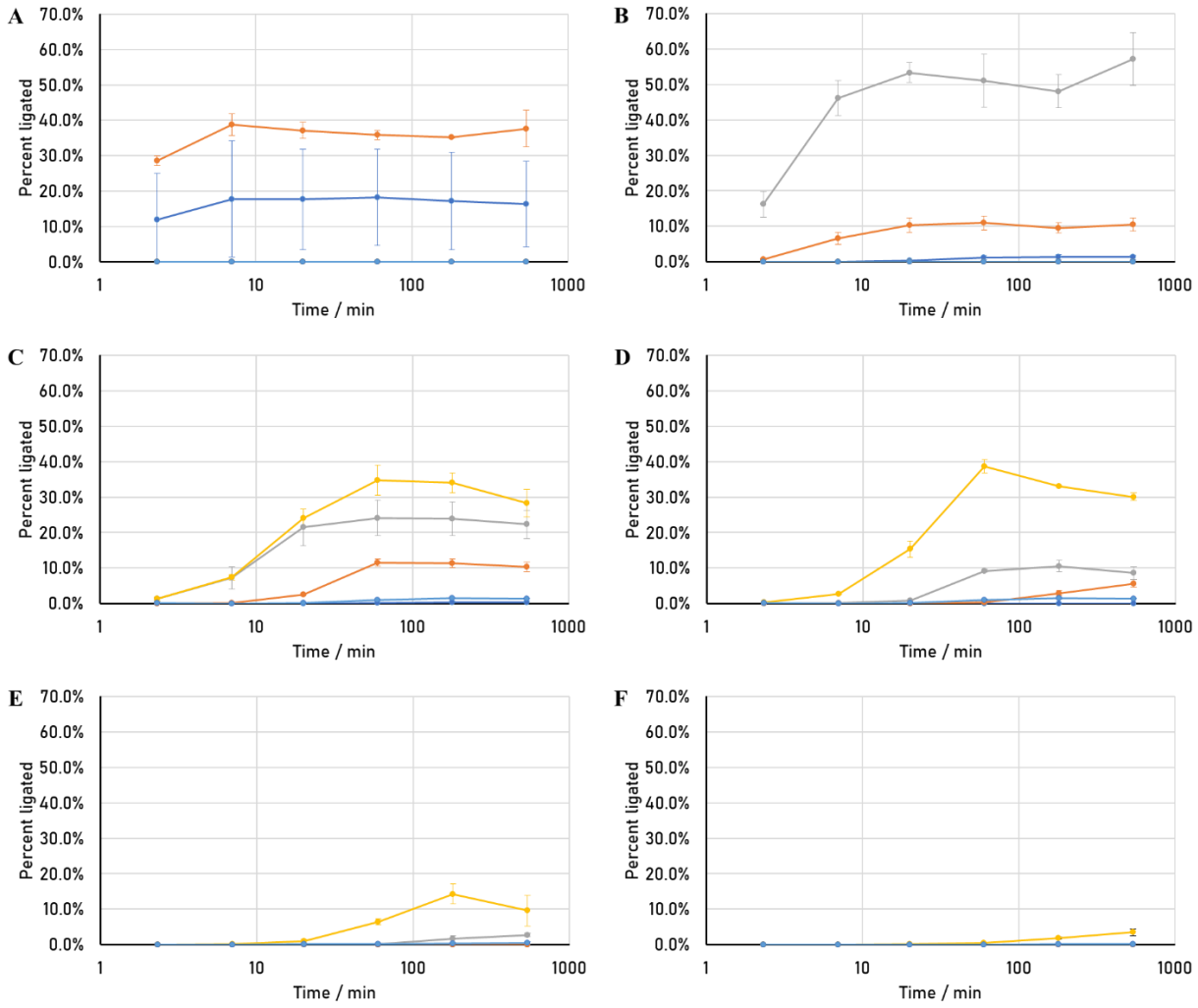

Figure 3.S5: Full kinetics data of ribozyme 51 at various Yb^{3+} and Tmp concentrations, corresponding data shown in Figure 3.4. Data shown is from triplicate experiments, and error bars represent standard deviation. Each graph displays data with one trimetaphosphate concentration: **(A):** 10mM Tmp. **(B):** 3mM Tmp. **(C):** 1mM Tmp. **(D):** 0.3mM Tmp. **(E):** 0.1mM Tmp. **(F):** 0.03mM Tmp. In each graph, the dark blue line represents 3mM Yb^{3+}, the orange line represents 1mM Yb^{3+}, the gray line represents 0.3mM Yb^{3+}, the yellow line represents 0.1mM Yb^{3+}, and the light blue line represents 0.03mM Yb^{3+}. These experiments were conducted at pH 7.3, with 150mM NaCl, no KCl, and no $MgCl_2$.

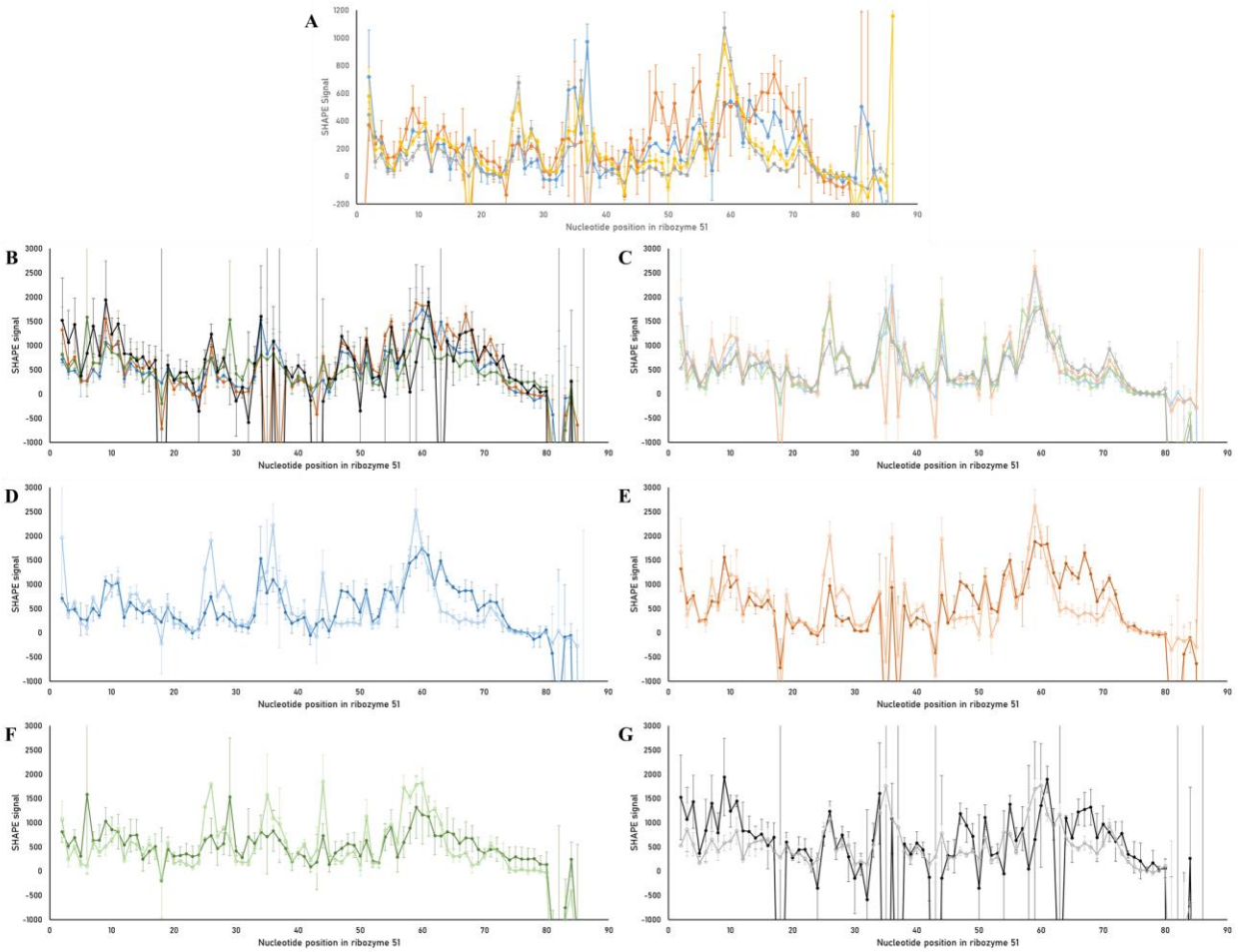

Figure 3.S6: Full SHAPE data from all experiments testing the influence of different reaction components on secondary structure (corresponding data shown in Figure 3.7). All data are from triplicate experiments, and error bars represent standard deviations. **(A)** Ribozyme 51 SHAPE data with 400 mM KCl and 50 mM MOPS-NaOH pH 6.8. Varying presence of 0.33 mM $Yb(Tf)_3$ and 3.3 mM Na_3Tmp presence, including $Yb(Tf)_3$ and Na_3Tmp (blue), just $Yb(Tf)_3$ (orange), just Na_3Tmp (gray), and neither (yellow). **(B)-(G):** Ribozyme 51 SHAPE data with 50mM MOPS-NaOH pH 6.8 and varying presence of 500 mM KCl, 5 mM $MgCl_2$, 0.33 mM $Yb(Tf)_3$, and 3.3 mM Na_3Tmp. **(B)** Data with $Yb(Tf)_3$ and Na_3Tmp, showing with KCl and $MgCl_2$ (blue), just KCl (orange), just $MgCl_2$ (green), and neither (black). **(C)** Data without $Yb(Tf)_3$ and Na_3Tmp, with the colors the same as (B) but lighter. **(D)-(G):** Same data as (B) and (C), but showing each pair with and without $Yb(Tf)_3$ (darker color) and Na_3Tmp (lighter color). **(D)** Data with KCl and $MgCl_2$. **(E)** Data with just KCl. **(F)** Data with just $MgCl_2$. **(G)** Data with neither.

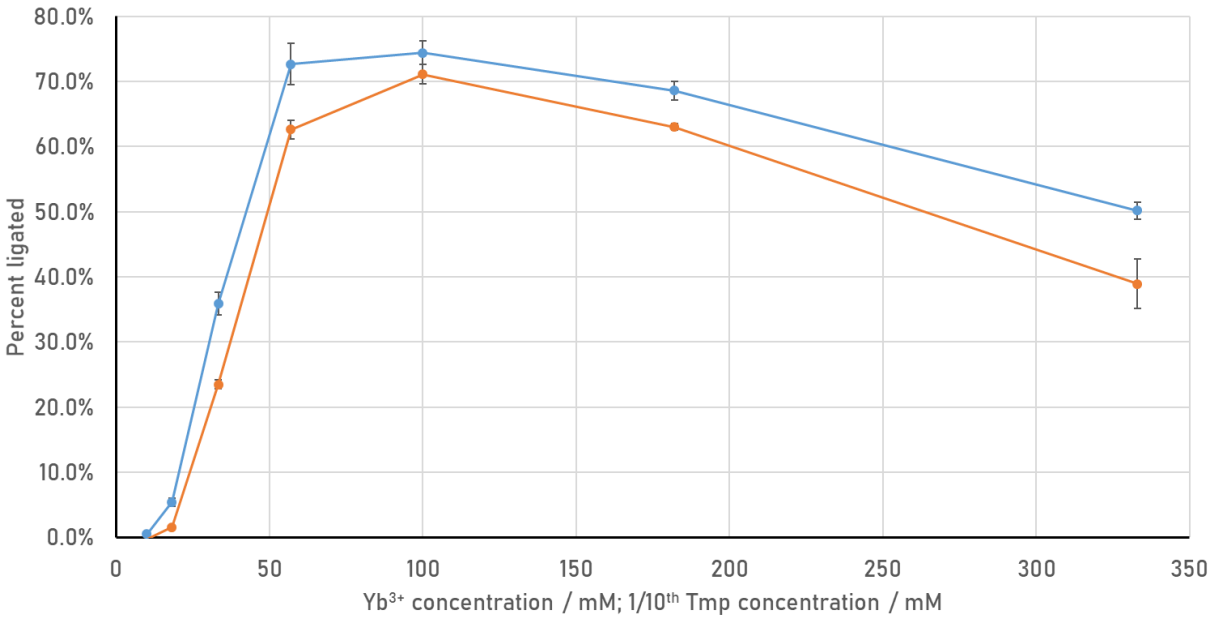

Figure 3.S7: Activity data at various Yb^{3+}/Tmp concentrations, comparing ribozyme 51 with the sequence from before the N20 re-selection, clone 15-T6 (which has just a GAAA tetraloop instead of the N20 region, which is nucleotides 53-72 in ribozyme 51). Activity of ribozyme 51 is shown in blue, while activity of clone 15-T6 is shown in orange. Data is from triplicate experiments containing 50 mM MOPS-NaOH pH 6.8, 500 mM KCl, 5mM $MgCl_2$, and always a 10:1 molar ratio of Na_3Tmp:$Yb(Tf)_3$. Error bars represent standard deviations.

References:

1. Rich, A., On the problems of evolution and biochemical information transfer. *Horizons in biochemistry*, 1962. New York, NY(Academic Press Inc.): p. 103-126
2. Crick, F.H.C., The origin of the genetic code. *J. Mol. Biol.*, 1968. 38: p. 367-379.
3. Woese, C.R., The fundamental nature of the genetic code: prebiotic interactions between polynucleotides and polyamino acids or their derivatives. *Proc Natl Acad Sci U S A*, 1968. 59(1): p. 110-7.
4. Orgel, L.E., Evolution of the genetic apparatus. *J. Mol. Biol.*, 1968. 38: p. 381-393.
5. Ellington, A.D. and J.W. Szostak, In vitro selection of RNA molecules that bind specific ligands. *Nature*, 1990. 346(6287): p. 818-22.
6. Tuerk, C. and L. Gold, Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, 1990. 249(4968): p. 505-10.
7. Bartel, D.P. and J.W. Szostak, Isolation of new ribozymes from a large pool of random sequences. *Science*, 1993. 261(5127): p. 1411-8.
8. Johnston, Wendy K., Unrau, Peter J., Lawrence, Michael S., Glasner, Margaret E., and David P. Bartel. RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science*, 2001. 292(5520): p. 1319-25.
9. White, H.B., 3rd, Coenzymes as fossils of an earlier metabolic state. *J Mol Evol*, 1976. 7(2): p. 101-4.
10. Nelson, J.W. and R.R. Breaker, The lost language of the RNA World. *Sci Signal*, 2017. 10(483).
11. Etaix, E. and L.E. Orgel, Phosphorylation of nucleosides in aqueous solution using trimetaphosphate: Formation of nucleoside triphosphates. *J. Carbohydrates Nucleosides Nucleotides*, 1978. 5(2): p. 91-110.
12. Cheng, C., Fan, C., Wan, R., Tong, C., Miao, Z., Chen, J., Zhao, Y. Phosphorylation of adenosine with trimetaphosphate under simulated prebiotic conditions. *Orig Life Evol Biosph*, 2002. 32(3): p. 219-24.
13. Pasek, Matthew A., Kee, Terrance P., Bryant, David E., Pavlov, Alexander A., and Jonathan I. Lunine. Production of potentially prebiotic condensed phosphates by phosphorus redox chemistry. *Angew Chem Int Ed Engl*, 2008. 47(41): p. 7918-20.
14. Pasek, M.A., Harnmeijer, J. P., Buick, R., Gull, M., Atlas, Z. Evidence for reactive reduced phosphorus species in the early Archean ocean. *Proc Natl Acad Sci U S A*, 2013. 110(25): p. 10089-94.
15. Moretti, J.E. and U.F. Muller, A ribozyme that triphosphorylates RNA 5'-hydroxyl groups. *Nucleic Acids Res*, 2014. 42(7): p. 4767-78.
16. Pressman, Abe, Moretti, Janina E., Campbell, Gregory W., Müller, Ulrich F., and Irene A. Chen. Analysis of in vitro evolution reveals the underlying distribution of catalytic activity among random sequences. *Nucleic Acids Res*, 2017. 45(14): p. 8167-8179.
17. Seelig, B. and A. Jaschke, A small catalytic RNA motif with Diels-Alderase activity. *Chem Biol*, 1999. 6(3): p. 167-76.

18. Unrau, P.J. and D.P. Bartel, RNA-catalysed nucleotide synthesis. *Nature*, 1998. 395(6699): p. 260-3.

19. Sengle, G., Eisenführ, A., Arora, P. S., Nowick, J. S. Novel RNA catalysts for the Michael reaction. *Chem Biol*, 2001. 8(5): p. 459-73.

20. Jaeger, L., M.C. Wright, and G.F. Joyce, A complex ligase ribozyme evolved in vitro from a group I ribozyme domain. *Proc Natl Acad Sci U S A*, 1999. 96(26): p. 14712-7.

21. Zhang, B. and T.R. Cech, Peptide bond formation by in vitro selected ribozymes. *Nature*, 1997. 390(6655): p. 96-100.

22. Jones, C., Nomosatryo, S., Crowe, S.A., Bjerrum, C.J., Canfield, D.E. Iron oxides, divalent cations, silica, and the early earth phosphorus crisis. *Geology*, 2015. 43(2): p. 135-138.

23. Halevy, I. and A. Bachan, The geologic history of seawater pH. *Science*, 2017. 355(6329): p. 1069-1071.

24. Huskens, Jurriaan, Kennedy, Anna D., van Bekkum, Herman, and Joop A. Peters. The Hydrolysis of Trimetaphosphate Catalyzed by Lanthanide(III) Aminopolycarboxylate Complexes: Coordination, Stability, and Reactivity of Intermediate Complexes. *J. Am. Chem. Soc.*, 1995. 117: p. 375-382.

25. Reiter, N.J., Osterman, A., Torres-Larios, A., Swinger, K.K., Pan, T., Mondragón, A. Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature*, 2010. 468(7325): p. 784-9.

26. Stahley, M.R. and S.A. Strobel, RNA splicing: group I intron crystal structures reveal the basis of splice site selection and metal ion catalysis. *Curr Opin Struct Biol*, 2006. 16(3): p. 319-26.

27. Toor, N., Keating, K.S., Taylor, S.D., Pyle, A.M. Crystal structure of a self-spliced group II intron. *Science*, 2008. 320(5872): p. 77-82.

28. Thaplyal, P., Ganguly, A., Hammes-Schiffer, S., Bevilacqua, P.C. Inverse thio effects in the hepatitis delta virus ribozyme reveal that the reaction pathway is controlled by metal ion charge density. *Biochemistry*, 2015. 54(12): p. 2160-75.

29. Shechner, D.M. and D.P. Bartel, The structural basis of RNA-catalyzed RNA polymerization. *Nat Struct Mol Biol*, 2011. 18(9): p. 1036-42.

30. Kobayashi, Shū, Satoshi Nagayama, and Tsuyoshi Busujima. "Lewis acid catalysts stable in water. Correlation between catalytic activity in water and hydrolysis constants and exchange rate constants for substitution of inner-sphere water ligands." *Journal of the American Chemical Society* 120.32 (1998): 8287-8288.

31. Yaroshevsky, A.A., Abundances of chemical elements in the Earth's crust. *Geochemistry International*, 2006. 44(1): p. 48-55.

32. Balaram, V., Rare earth elements: A review of applications, occurrence, exploration, analysis, recycling, and environmental impact. *Geoscience Frontiers*, 2019. 10(4): p. 1285-1303.

33. Haxel, Gordon. Rare earth elements: critical resources for high technology. Vol. 87. No. 2. US Department of the Interior, US Geological Survey, 2002.

34. Sigel, Helmut, ed. Metal Ions in Biological Systems: Volume 40: The Lanthanides and Their Interrelations with Biosystems. CRC Press, 2003.

35. Aide, Michael. "Lanthanide soil chemistry and its importance in understanding soil pathways: Mobility, plant uptake and soil health." *Lanthanides. IntechOpen*, 2018.

36. Nakagawa, Tomoyuki, Mitsui, Ryoji, Tani, Akio, Sasa, Kentaro, Tashiro, Shinya, Iwama, Tomonori, Hayakawa, Takashi, and Keiichi Kawai. "A catalytic role of XoxF1 as La3+-dependent methanol dehydrogenase in Methylobacterium extorquens strain AM1." *PloS one* 7.11 (2012): e50480.

37. Keltjens, Jan T., Pol, Arjan, Reimann, Joachim, and Huub J.M. Op den Camp. "PQQ-dependent methanol dehydrogenases: rare-earth elements make a difference." *Applied microbiology and biotechnology* 98.14 (2014): 6163-6183.

38. Daumann, Lena J. "Essential and ubiquitous: the emergence of lanthanide metallobiochemistry." *Angewandte Chemie International Edition* 58.37 (2019): 12795-12802.

39. Chu, Frances, and Mary E. Lidstrom. "XoxF acts as the predominant methanol dehydrogenase in the type I methanotroph Methylomicrobium buryatense." *Journal of bacteriology* 198.8 (2016): 1317-1325.

40. Taubert, Martin, Grob, Carolina, Howat, Alexandra M., Burns, Oliver J., Dixon, Joanna L., Chen, Yin, and J. Colin Murrell. "XoxF encoding an alternative methanol dehydrogenase is widespread in coastal marine environments." *Environmental microbiology* 17.10 (2015): 3937-3948.

41. Cotruvo Jr, Joseph A., Featherston, Emily R., Mattocks, Joseph A., Ho, Jackson V., and Tatiana N. Laremore. "Lanmodulin: a highly selective lanthanide-binding protein from a lanthanide-utilizing bacterium." *Journal of the American Chemical Society* 140.44 (2018): 15056-15061.

42. Sinha, Shyama P., ed. Systematics and the Properties of the Lanthanides. Vol. 109. *Springer Science & Business Media*, 2012.

43. Markham, Nicholas R., and Michael Zuker. "UNAFold." *Bioinformatics*. Humana Press, 2008. 3-31.

44. Mortimer, Stefanie A., and Kevin M. Weeks. "A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry." *Journal of the American Chemical Society* 129.14 (2007): 4144-4145.

45. Lumpe, Henning, Pol, Arjan, Op den Camp, Huub J. M., and Lena J. Daumann. "Impact of the lanthanide contraction on the activity of a lanthanide-dependent methanol dehydrogenase–a kinetic and DFT study." *Dalton Transactions* 47.31 (2018): 10463-10472.

46. Huang, P. J., and Juewen Liu. "In vitro selection and application of lanthanide-dependent DNAzymes." *Methods in Enzymology* 651 (2021): 373-396.

47. Javadi-Zarnaghi, Fatemeh, and Claudia Höbartner. "Lanthanide cofactors accelerate DNA-catalyzed synthesis of branched RNA." *Journal of the American Chemical Society* 135.34 (2013): 12839-12848.

48. Huang, Po-Jung Jimmy, Mahsa Vazin, and Juewen Liu. "In vitro selection of a new lanthanide-dependent DNAzyme for ratiometric sensing lanthanides." *Analytical chemistry* 86.19 (2014): 9993-9999.

49. Huang, Po-Jung Jimmy, Mahsa Vazin, and Juewen Liu. "In vitro selection of a DNAzyme cooperatively binding two lanthanide ions for RNA cleavage." *Biochemistry* 55.17 (2016): 2518-2525.

50. Zhou, Wenhu, Ding, Jinsong, and Juewen Liu. "A DNAzyme requiring two different metal ions at two distinct sites." *Nucleic acids research* 44.1 (2016): 354-363.

51. Glasner, Margaret E., Nicholas H. Bergman, and David P. Bartel. "Metal ion requirements for structure and catalysis of an RNA ligase ribozyme." *Biochemistry* 41.25 (2002): 8103-8112.

52. Schnabl, Joachim, and Roland KO Sigel. "Controlling ribozyme activity by metal ions." *Current opinion in chemical biology* 14.2 (2010): 269-275.

53. Rogers, Jeff, and Gerald F. Joyce. "The effect of cytidine on the structure and function of an RNA ligase ribozyme." *RNA* 7.3 (2001): 395-404.

Chapter 4: Conclusion

Summarized Work

In the work described in this dissertation, I have conducted two in vitro selections for catalytic RNAs (ribozymes) and analyzed three ribozymes that utilize the cofactors of interest. My first project, the selection of triphosphorylation ribozymes in the presence of peptides, resulted in more than 50 sequence clusters of ribozymes with varying responses to the peptides, as seen in the high-throughput sequencing data. Two clusters displaying very different behavior were investigated, each resulting in one ribozyme sequence that was studied biochemically in more depth. Ribozyme 20 showed a benefit to its triphosphorylation activity of 900 (+/-300)-fold from only one specific peptide sequence (N-AAEAAAKA-C). It is thus an example of a ribozyme that would benefit from an RNA world organism developing the ability to produce specific peptide sequences via a primitive translation system. The effect of peptide concentration on ribozyme activity showed a sigmoid dependence, suggesting that the peptide is bound as a dimer or multimer (Figure 4.1). Effects of the peptide on the ribozyme's secondary structure were identified via SHAPE probing, describing a defined and limited effect of the peptide on the ribozyme's secondary structure. Overall, the results are consistent with the ribozyme forming a specific complex with a multimer of the peptide.

Ribozyme 23, from the other cluster studied, showed a weaker, but more general benefit from multiple peptides, sometimes in the range of 2-10 fold depending on the conditions. Because the ribozyme benefitted from up to six of the ten octapeptides tested, it is an example of a sequence that might benefit from peptides before the advent of a primitive translation system, where random or semi-random peptide sequences would have been assembled without RNA catalysis in the prebiotic environment. The ribozyme had a bigger increase in activity when the conditions were suboptimal, indicating that the presence of the peptides could aid this ribozyme in staying active while conditions change. In addition, the benefit conferred was mostly kinetic, as the ribozyme's rate increased more than

its amplitude. These data, and the strong effect of peptides on the ribozyme's secondary structure, suggest that the peptides may help the ribozyme folding by escaping kinetic folding traps (Figure 4.2). Both ribozymes can serve as models for future study of peptide-RNA interactions, both in the RNA world context and more generally.

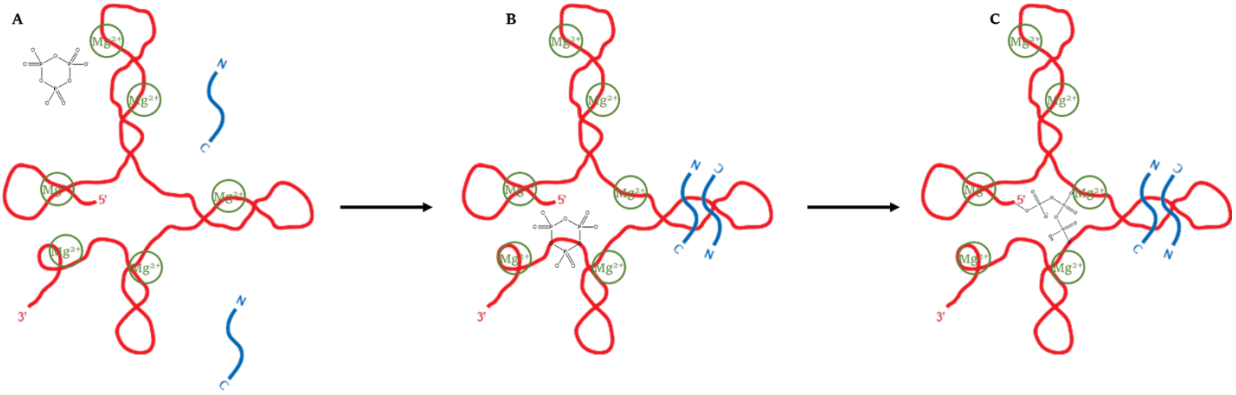

Figure 4.1: Proposed mechanism for peptide 4 assemblies aiding the triphosphorylation activity of ribozyme 20. **(A)** Ribozyme 20, peptide 4, Mg^{2+}, and Tmp mixed together in solution. Ribozyme 20 is in a structure closely resembling the active structure. **(B)** Multiple peptide 4 molecules come together to form one or more assemblies, which helps ribozyme 20 find the active structure and bind trimetaphosphate. **(C)** Now in the active structure interacting with more than one peptide 4 molecule, ribozyme 20 performs nucleophilic attack on the trimetaphosphate, ending with a 5'-triphosphate.

In addition, I performed a similar selection using ytterbium as the multivalent cation for the triphosphorylation reaction instead of magnesium. The selection began with a fully random pool, as triphosphorylation ribozymes previously selected in the presence of Mg^{2+} were unable to transition to using Yb^{3+}. The most active resulting ribozyme, ribozyme 51, was dependent on ytterbium for catalysis and seemed to bind trimetaphosphate as a complex with ytterbium with an optimal molar ratio of 10:1 Tmp:Yb^{3+} (Figure 4.3). Ribozyme 51 was able to use lutetium to a similar level of activity as ytterbium, but successively lighter lanthanides had strongly decreased ability to support the reaction. This is likely due to selectivity of the binding site for the ion radius of the lanthanide, given the similar character of lanthanides across the series. The ribozyme also benefitted greatly from monovalent cations (K^+ or Na^+) and from Mg^{2+}: despite these three cations not catalyzing the reaction, the presence of at least one of the three could increase the ribozyme's rate by >100-fold compared to only Yb^{3+} and just 10mM Na^+ being

the only cations present. SHAPE experiments identified specific secondary structure elements that were likely involved in binding the ytterbium/trimetaphosphate complex. The ribozyme is able to utilize quite low concentrations of Yb^{3+}, with activity seen at concentrations as low as 18μM, which demonstrates the ribozyme's high affinity for Yb^{3+} and/or the Yb^{3+}/Tmp complex. This selection expands the useful cofactors for catalytic RNAs to include lanthanides, and finds that lanthanides are possible, but probably not important cofactors in an RNA world scenario.

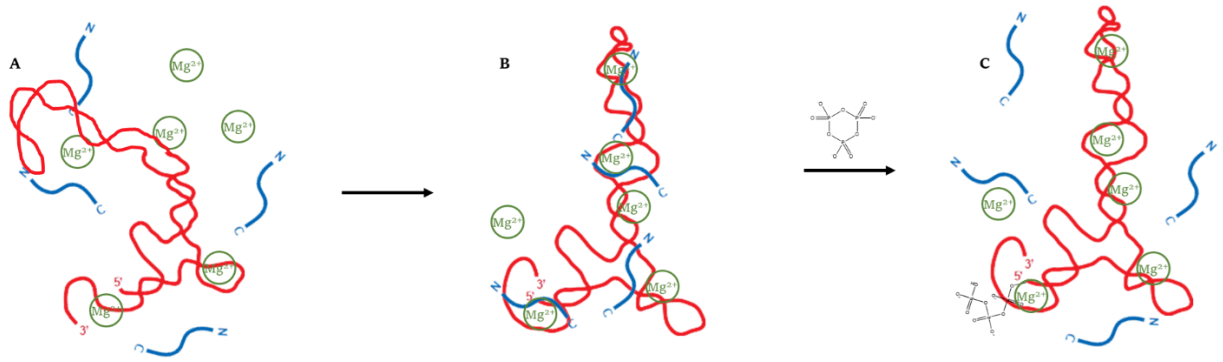

Figure 4.2: Proposed mechanism for peptides helping ribozyme 23 to escape a kinetic folding trap, allowing it to find the active structure. **(A)** Ribozyme 23 folds into a stable, but inactive/less active structure, which is stabilized by Mg^{2+}. **(B)** Peptides help ribozyme 23 fold into the optimal structure for activity, promoting binding of trimetaphosphate. **(C)** Peptides dissociate. With ribozyme 23 now in the most active structure, it performs a nucleophilic attack on the trimetaphosphate, ending with a 5'-triphosphate.

Current Impact

Our selection of triphosphorylation ribozymes using peptides as cofactors was the first to directly address the question of how peptides may affect the development of ribozymes from random sequences. The first in vitro selection/evolution of an RNA-peptide interaction was published in 2003 by identifying RNAs that could bind to the HIV Rev Response Element, but this selection was only for binding and did not involve catalysis[1]. The HIV Rev Response Element is also a peptide evolutionarily optimized for interactions with RNA. Other experiments have successfully identified novel ribozyme-peptide interactions. In one such study, Andrew Ellington's lab evolved the L1 ligase ribozyme to only be active in the presence of the HIV Arginine Rich Motif, illustrating a novel interaction between a catalytic RNA and a peptide[2], and showing the protein binding to an RNA can modulate the RNA's catalytic activity.

However the evolution began with a randomized region added to an already existing ribozyme, and the HIV-ARM was also a peptide evolutionarily optimized to interact with RNA. Philip Holliger's lab tested the activity of a previously evolved RNA polymerase ribozyme with and without peptides[3]. However, in this study, the ribozyme was not evolved in the presence of peptides, and the peptides used were either already evolutionarily optimized for interactions with ribosomal RNA or strictly polycationic peptides, which are prebiotically implausible.

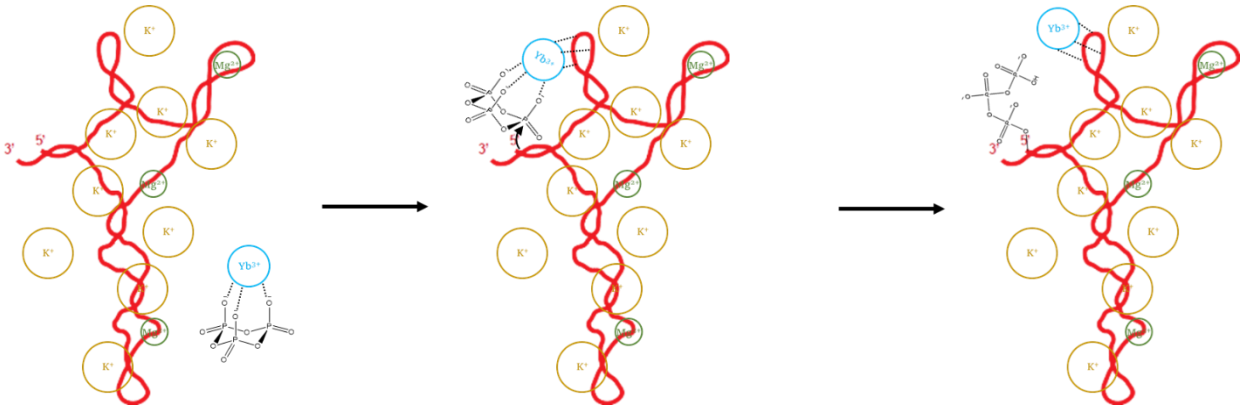

Figure 4.3: Proposed mechanism for ytterbium acting as a metal ion cofactor for ribozyme 51's self-triphosphorylation activity. **(A)** K^+ and Mg^{2+} ions interact with the backbone of the ribozyme, helping it fold into the most active structure. Meanwhile, Yb^{3+} forms a complex with Tmp. **(B)** The ribozyme binds the Yb^{3+}/Tmp complex, coordinating it with the loop at nucleotides 25-29 and positioning for the nucleophilic attack using the 5'-hydroxyl group. **(C)**: Ribozyme 23 has reacted with the trimetaphosphate, ending with a 5'-triphosphate.

Recent studies examined the formation of small, membraneless compartments involving RNA (and often peptides) and their effect on RNA folding and catalysis. These compartments, known as coacervates, form upon association of high concentrations of anions and cations, with the components at a local effective concentration of 1000-fold or more higher than in bulk solution[4]. In these coacervates, RNA acts as a polyanion that coordinates with the polycations, including peptides, provided that small ions are also present at millimolar concentrations. One study demonstrated that hammerhead ribozyme catalysis can be accelerated by coacervate formation with polylysine and polyarginine, and that other, non-peptide polycations can improve template-directed RNA polymerization when Mg^{2+} concentration is suboptimal (<1mM)[5]. The most important mechanism appears to be the co-localization of RNA strands in the coacervate. The formation of coacervates can also be controlled by phosphorylation of their peptides,

as that disrupts the ionic interactions[6]. These studies have advanced our understanding of RNA-peptide interactions, while our in vitro selection system experiments appear to have identified two completely different mechanisms by which peptides could aid the emergence of ribozymes.

Despite being the most commonly seen amino acids interacting with RNA in modern biology, the cationic amino acids are not found in abundance in Miller-type experiments, and analysis of the genetic codes across different kingdoms of life suggests that they were likely only produced in abundance by biocatalysts once life had begun[7,8,9,10,11]. When designing the peptides for the selection, we wanted to cover the range of prebiotic plausibility for the 20 encoded amino acids. Therefore, some peptides were enriched in more prebiotically plausible amino acids, while others were enriched in more positively charged and aromatic ones. For example, while peptides such as P6 (N-AEEEAKKK-C) and P10 (N-MQELRWCN-C) include multiple amino acids unlikely to be abundant prebiotically, peptides such as P1 (N-GAGAGDGA-C) and P3 (N-ASGVAGDG-C) are composed only of amino acids seen in Miller-type experiments. The resulting ribozymes 20 and 23 are the first characterized ribozyme sequences selected from a completely random pool in the presence of peptides, and thus our project is the first to directly address the question of how peptides may affect the development of ribozymes from random sequences. Based on the HTS data, the selection generated several additional ribozymes that appear to benefit from peptide cofactors, providing additional model systems for RNA-peptide interactions and the benefit of peptides for the emergence of RNA catalysis.

One possibility for how ribozyme 23 may benefit from peptides is that the peptides generally may help the ribozyme fold into its most active structure. One of the main issues faced by ribozymes is that RNAs often get stuck in kinetic folding traps, especially in the presence of hard cations such as Mg^{2+}, which stabilize RNA structures[12]. However, softer cations, such as ammonium and spermidine, can help RNAs escape these folding traps and find their optimal structure. Additionally, recent work has shown that mixed cationic proto-peptides of amino acids and alpha-hydroxy acids can increase the thermal stability of RNA secondary structures while being protected somewhat against hydrolysis by that RNA[13]. Taken together, these may help explain why peptide 6 (which has a C-terminal tri-lysine stretch) provides

the most benefit for ribozyme 23, in addition to why the benefit to activity from peptides is generally greater under reaction conditions that are further away from the ribozyme's optimum (pH, temperature, Mg^{2+}/trimetaphosphate concentration).

The near-total reliance of ribozyme 20 on peptide 4 for activity (900 +/- 300-fold) likely follows a different mechanism, as it is a specific interaction that confers a much greater benefit to the ribozyme. The sequence of peptide 4 (N-AAEAAAKA-C) is such that it seems to be the octapeptide sequence most likely to form an alpha-helix, given the high helix-propensity of alanine and the possible salt-bridge formed between the glutamate and lysine[14,15]. However, a peptide of only 8 amino acids is unlikely to form an alpha-helix on its own without a disulfide bond, because only four hydrogen bonds between backbone amide groups would stabilize the alpha-helical structure[16]. Previous studies have shown that longer peptides with repeats of a similar sequence to peptide 4 have shown the ability to form alpha-helices, either in a concentration-dependent manner or with multiple repeats separated by a loop, implying that multiple peptides forming a larger complex together can help stabilize the individual alpha-helices[17]. Further study of the same peptide with transmission electron microscopy (TEM) showed that they form long fibril-like structures at high concentrations[18]. The sigmoidal concentration dependence for alpha-helix formation in these studies seems to mirror the sigmoidal concentration dependence of ribozyme 20's activity with peptide 4, indicating that peptide 4 may also be forming alpha-helices when this peptide assembles into a larger complex. The sigmoidal change does not necessarily mean alpha-helices are being formed, but it is one possible structure. The peptides could instead be coming together in unorganized, mostly hydrophobic aggregates. In the high throughput sequencing data, peptide 4 caused the strongest increase in sequence abundance for any cluster compared to the 'no peptide' reaction, and multiple other sequence clusters received strong (>10-fold) benefits to their fitness from peptide 4. Given that peptide 4 caused major changes to the composition of the ribozyme sequence pool, it seems that the presence and abundance of peptides forming the same structures or aggregates as peptide 4 could be an important factor in the development of ribozymes. Overall, the selection with peptides has provided several model systems

for how peptides can influence the emergence of ribozymes from a random RNA pool, which can inform us how peptides could have benefitted catalytic RNAs at early stages of life's emergence.

The selection for triphosphorylation ribozymes with ytterbium as a cofactor provided evidence that ribozymes can be active with lanthanides as metal-ion cofactors. While our study is the first to develop an RNA that uses a lanthanide cofactor, protein enzymes have been identified that use lanthanide cofactors. Lanthanide-using methanol dehydrogenases are one example, and are closely related to calcium-using enzymes that perform the same function[19,20,21]. At least one organism with both enzymes (it is common for methylotrophs to have both) will preferentially use the lanthanide-dependent enzyme rather than the calcium-dependent one, likely due to the lanthanides' stronger Lewis acidity[22]. A previously selected DNAzyme has also been demonstrated using lanthanides as cofactors[23], which lead to a series of in vitro selection studies resulting in DNAzymes dependent on one or more lanthanides for activity[24,25,26]. Our study is the first to demonstrate a ribozyme using lanthanides for catalysis. Ribozyme 51, the most active ribozyme, was able to use only the heavier lanthanides, which is likely due to the change in ionic radius across the series since the lanthanides behave chemically very similarly. This trend of decreasing ion radius going from lighter to heavier lanthanides is known as lanthanide contraction, and the size-selective incorporation is commonly seen for lanthanides in minerals[27], protein enzymes[28], and DNAzymes[29]. Based on activity dependence with both components and secondary structure analysis, ribozyme 51 seems to bind to trimetaphosphate only as a complex with lanthanides. The molar ratio of 10:1 Tmp:Yb^{3+} was optimal, suggesting that each lanthanide needs to have more than one Tmp molecule coordinated for catalysis. This ideal 10:1 molar ratio was not caused by Tmp protecting the ribozyme against hydrolysis by Yb^{3+}, since we did not detect significant RNA hydrolysis even at 1:1 molar ratios and at higher Yb^{3+} concentrations. As seen in the 1995 study by Huskens and Peters, lanthanides can coordinate trimetaphosphate and accelerate its hydrolysis, especially if the lanthanide is coordinated with EDTA or NTA[30]. This study expands our knowledge of the ability for ribozymes to utilize various ions other than the most common Mg^{2+} (figure 4.4)[31,32,33,34]. The importance for lanthanides for origin-of-life scenarios is likely limited because, even though lanthanides are found at not insignificant concentrations

in the Earth's crust (see introduction), they do not seem to be superior to the much more abundant Mg^{2+} in supporting ribozyme catalysis. While unlikely, the biological examples of lanthanide cofactors suggest that it would remain a possibility for lanthanides to have been used as a cofactor by RNA-based organisms.

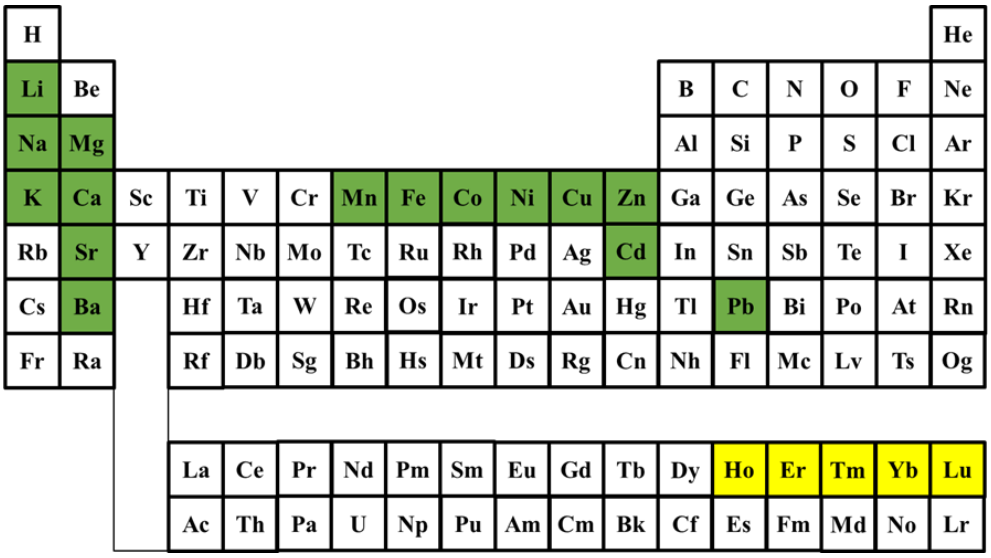

Figure 4.4: Periodic table of the elements, with all ions filled in that are known to be able to participate in ribozyme catalysis. Those known from previous studies are highlighted in green[31,32,33,34], those shown in this work are highlighted in yellow.

Future Directions

The project identifying peptide-using triphosphorylation ribozymes opens many doors for future study, given the unique new interactions found and the importance of RNA-peptide interactions from the origin of life all the way through to modern biology. The interaction of ribozyme 20 with peptide 4, for example, could yield interesting results with deeper exploration. Circular Dichroism (CD) experiments using peptide 4 could identify if it forms an alpha-helix. Doing CD at various concentrations of the peptide could also test if the peptide's alpha-helical structure is dependent on the assembly of multiple peptides, as was the case in the study from Kojima and coworkers[17]. If it is observed, it would indicate that peptide 4 alpha-helix formation, and/or formation of the larger complex of alpha helices, confers a 900 (+/- 300)-fold benefit of activity to ribozyme 20, and possibly a >10-fold benefit to multiple other

ribozymes as well. If alpha-helices are not observed with peptide 4, it should be tested with the ribozyme present as well since it is possible that the ribozyme could help stabilize the alpha helix, as has been seen for a portion of the HIV Rev protein[35]. However, whether or not alpha-helices are observed, the strong influence of peptide 4 on several clusters in the selection indicates that its sequence or structure beneficial to include in future in vitro selection experiments.

Further study of ribozyme 23 and its interaction with peptides could yield very useful information about the influence of peptides on ribozyme folding and kinetics. Activity experiments could be done testing this ribozyme with other cations to see if that is the primary benefit, including polycationic peptides (such as polylysine or polyarginine) or smaller cationic compounds like spermine. The ribozyme's kinetics could also be tested at other suboptimal conditions than pH, including temperature, magnesium concentration, and trimetaphosphate concentration, which would provide more information about how the ribozyme benefits from peptides. To further investigate if the benefit of the peptides is to folding, thermal denaturation experiments could be conducted comparing ribozyme 23 with and without peptides, especially peptide 6. Finally, since the high-throughput sequencing results do not show as much benefit for ribozyme 23 with peptide 6 as the biochemical testing does, it is possible that many other ribozymes would also benefit to a similar degree from peptide 6, or other polycationic peptides. This could be investigated by looking at other ribozymes from the high-throughput sequencing data.

In addition to the two ribozymes tested in more depth, many other interesting new peptide-RNA interactions could be identified by investigating other sequences from this selection. For example, multiple sequences ordered for the initial biochemical testing also showed benefits in the HTS analysis from peptide 4, including those from clusters 2 and 71. Cluster 2 showed similar benefits from peptides 8 and 10, and cluster 71 also benefitted from peptide 8. Given these different behaviors from ribozyme 20, which was strongly enriched only with peptide 4, they may utilize the peptides differently. These sequences showed very low activity and thus undetectable peptide influence in the initial screen, but cluster 20 was barely better before optimizing conditions. If the reaction conditions for ribozymes from clusters 2 and 71 were optimized and activity was tested with individual peptides, a dramatic effect like

that seen with ribozyme 20 could possibly be identified. Beyond that, many clusters (and individual sequences/mutations within clusters) can be identified in the high-throughput sequencing data as having a strong benefit from peptides, so there are many possibilities for other sequences from this selection to investigate. Finally, similar investigations could be done with differently designed peptides depending on the desired goal. For example, for a more direct focus on prebiotically plausible peptides, a selection could be done that includes only prebiotically plausible amino acids in racemic mixtures, mixed lengths, and proto-peptides/depsipeptides synthesized via wet/dry cycling with alpha-hydroxy acids. A project could also be done focusing on polycationic peptides and their implications for RNA folding.

The implications of the selection with ytterbium on origin of life research are unknown. Given that ytterbium-using ribozymes did not emerge more quickly from the selection than magnesium-using ribozymes in the original triphosphorylation ribozyme selection using the same random pool, ytterbium may not be more useful than Mg^{2+} for ribozymes in an RNA world. However, this work demonstrated the utility of this selection procedure in identifying possible new metal ion cofactors for ribozymes, which could be done with other new cofactors. In addition, evolution experiments could be conducted starting with lanthanide-using ribozymes from this selection such that the ribozyme evolves to use successively lighter lanthanides. High-throughput sequencing analysis of this selection could also yield interesting results about the number of active sequence clusters, and thus the general ability of Yb^{3+} to act as a cofactor for self-triphosphorylation using trimetaphosphate. Rather than concluding that lanthanide-using ribozymes would have been important to an RNA world, this project illustrates the importance of exploring the chemical space of catalytic RNAs and the utility of this selection for doing so, as any number of other ions and prebiotically plausible compounds could have been useful cofactors at the origin of life. Lanthanide-using ribozymes could also be used as tools, such as for detecting lanthanides in solution, as DNAzymes can be used[29].

As our understanding of the origin of life deepens, we are now able to learn more about how the RNA world could have been aided in its emergence and development by the other available chemicals. My work has contributed to a greater understanding of how the emergence of catalytic RNAs can be

influenced by the available cofactors. In addition, the peptide-RNA pairs described (and others not yet investigated in detail from the pool) can serve as simple model systems to study the mechanisms for how peptides can help RNAs, both in catalysis and in folding. A better understanding of how life on Earth could have originated can influence our search for life on other celestial bodies, and it can also bring into context our lives now and our place in the universe.

References

1. Peled-Zehavi, Hadas, Horiya, Satoru, Das, Chandreyee, Harada, Kazuo, and Alan D. Frankel. "Selection of RRE RNA binding peptides using a kanamycin antitermination assay." *RNA* 9.2 (2003): 252-261.
2. Robertson, Michael P., Scott M. Knudsen, and Andrew D. Ellington. "In vitro selection of ribozymes dependent on peptides for activity." *RNA* 10.1 (2004): 114-127.
3. Tagami, Shunsuke, James Attwater, and Philipp Holliger. "Simple peptides derived from the ribosomal core potentiate RNA polymerase ribozyme function." *Nature chemistry* 9.4 (2017): 325-332.
4. Frankel, Erica A., Philip C. Bevilacqua, and Christine D. Keating. "Polyamine/nucleotide coacervates provide strong compartmentalization of Mg2+, nucleotides, and RNA." *Langmuir* 32.8 (2016): 2041-2049.
5. Poudyal, Raghav R., Guth-Metzler, Rebecca M., Veenis, Andrew J., Frankel, Erica A., Keating, Christine D., and Philip C. Bevilacqua. "Template-directed RNA polymerization and enhanced ribozyme catalysis inside membraneless compartments formed by coacervates." *Nature communications* 10.1 (2019): 1-13.
6. Aumiller, William M., and Christine D. Keating. "Phosphorylation-mediated RNA/peptide complex coacervation as a model for intracellular liquid organelles." *Nature chemistry* 8.2 (2016): 129-137.
7. Miller, Stanley L. "A production of amino acids under possible primitive earth conditions." *Science* 117.3046 (1953): 528-529.
8. Parker, Eric T., Cleaves, Henderson J., Dworkin, Jason P., Glavin, Daniel P., Callahan, Michael, Aubrey, Andrew, Lazcano, Antonio, and Jeffrey L. Bada. "Primordial synthesis of amines and amino acids in a 1958 Miller H2S-rich spark discharge experiment." *Proceedings of the National Academy of Sciences* 108.14 (2011): 5526-5531.
9. Trifonov, Edward N. "The triplet code from first principles." *Journal of Biomolecular structure and dynamics* 22.1 (2004): 1-11.
10. Hoffman, Michael M., Khrapov, Maksim A., Cox, J. Colin, Yao, Jianchao, Tong, Lingnan, and Andrew D. Ellington. "AANT: The amino acid–nucleotide interaction database." *Nucleic acids research* 32.suppl_1 (2004): D174-D181.
11. Das, Chandreyee, and Alan D. Frankel. "Sequence and structure space of RNA-binding peptides." *Biopolymers: Original Research on Biomolecules* 70.1 (2003): 80-85.

12. Herschlag, Daniel. "RNA Chaperones and the RNA Folding Problem (∗)." *Journal of Biological Chemistry* 270.36 (1995): 20871-20874.

13. Frenkel-Pinter, Moran, Haynes, Jay W., Mohyeldin, Ahmad M., C., Martin, Sargon, Alyssa B., Petrov, Anton S., Krishnamurthy, Ramanarayanan, Hud, Nicholas V., Williams, Loren Dean, and Luke J. Leman. "Mutually stabilizing interactions between proto-peptides and RNA." *Nature communications* 11.1 (2020): 1-14.

14. Chakrabartty, Avijit, Tanja Kortemme, and Robert L. Baldwin. "Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions." *Protein Science* 3.5 (1994): 843-852.

15. Marqusee, Susan, and Robert L. Baldwin. "Helix stabilization by Glu-... Lys+ salt bridges in short peptides of de novo design." *Proceedings of the National Academy of Sciences* 84.24 (1987): 8898-8902.

16. Scholtz, J. Martin, and Robert L. Baldwin. "The mechanism of alpha-helix formation by peptides." *Annual review of biophysics and biomolecular structure* 21.1 (1992): 95-118.

17. Kojima, Shuichi, Kuriki, Yukino, Sato, Yoshihiro, Arisaka, Fumio, Kumagai, Izumi, Takahashi, Sho, and Kin-ichiro Miura. "Synthesis of α-helix-forming peptides by gene engineering methods and their characterization by circular dichroism spectra measurements." *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* 1294.2 (1996): 129-137.

18. Kojima, Shuichi, Kuriki, Yukino, Yoshida, Takao, Yazaki, Kazumori, and Kin-ichiro Miura. "Fibril formation by an amphipathic α-helix-forming polypeptide produced by gene engineering." *Proceedings of the Japan Academy, Series B* 73.1 (1997): 7-11.

19. Nakagawa, Tomoyuki, Mitsui, Ryoji, Tani, Akio, Sasa, Kentaro, Tashiro, Shinya, Iwama, Tomonori, Hayakawa, Takashi, and Keiichi Kawai. "A catalytic role of XoxF1 as La3+-dependent methanol dehydrogenase in Methylobacterium extorquens strain AM1." *PloS one* 7.11 (2012): e50480.

20. Keltjens, Jan T., Pol, Arjan, Reimann, Joachim, and Huub J.M. Op den Camp. "PQQ-dependent methanol dehydrogenases: rare-earth elements make a difference." *Applied microbiology and biotechnology* 98.14 (2014): 6163-6183.

21. Daumann, Lena J. "Essential and ubiquitous: the emergence of lanthanide metallobiochemistry." *Angewandte Chemie International Edition* 58.37 (2019): 12795-12802.

22. Chu, Frances, and Mary E. Lidstrom. "XoxF acts as the predominant methanol dehydrogenase in the type I methanotroph Methylomicrobium buryatense." *Journal of bacteriology* 198.8 (2016): 1317-1325.

23. Javadi-Zarnaghi, Fatemeh, and Claudia Höbartner. "Lanthanide cofactors accelerate DNA-catalyzed synthesis of branched RNA." *Journal of the American Chemical Society* 135.34 (2013): 12839-12848.

24. Huang, Po-Jung Jimmy, Mahsa Vazin, and Juewen Liu. "In vitro selection of a new lanthanide-dependent DNAzyme for ratiometric sensing lanthanides." *Analytical chemistry* 86.19 (2014): 9993-9999.

25. Zhou, Wenhu, Ding, Jinsong, and Juewen Liu. "A DNAzyme requiring two different metal ions at two distinct sites." *Nucleic acids research* 44.1 (2016): 354-363.

26. Huang, Po-Jung Jimmy, Mahsa Vazin, and Juewen Liu. "In vitro selection of a DNAzyme cooperatively binding two lanthanide ions for RNA cleavage." *Biochemistry* 55.17 (2016): 2518-2525.

27. Sigel, Helmut, ed. Metal Ions in Biological Systems: Volume 40: The Lanthanides and Their Interrelations with Biosystems. *CRC Press*, 2003.

28. Lumpe, Henning, Pol, Arjan, Op den Camp, Huub J. M., and Lena J. Daumann. "Impact of the lanthanide contraction on the activity of a lanthanide-dependent methanol dehydrogenase–a kinetic and DFT study." *Dalton Transactions* 47.31 (2018): 10463-10472.

29. Huang, P. J., and Juewen Liu. "In vitro selection and application of lanthanide-dependent DNAzymes." *Methods in Enzymology* 651 (2021): 373-396.

30. Huskens, Jurriaan, Kennedy, Anna D., van Bekkum, Herman, and Joop A. Peters. "The Hydrolysis of Trimetaphosphate Catalyzed by Lanthanide (III) Aminopolycarboxylate Complexes: Coordination, Stability, and Reactivity of Intermediate Complexes." *Journal of the American Chemical Society* 117.1 (1995): 375-382.

31. Glasner, Margaret E., Nicholas H. Bergman, and David P. Bartel. "Metal ion requirements for structure and catalysis of an RNA ligase ribozyme." *Biochemistry* 41.25 (2002): 8103-8112.

32. Schnabl, Joachim, and Roland KO Sigel. "Controlling ribozyme activity by metal ions." *Current opinion in chemical biology* 14.2 (2010): 269-275.

33. Pan, Tao, and Olke C. Uhlenbeck. "A small metalloribozyme with a two-step mechanism." *Nature* 358.6387 (1992): 560-563.

34. Murray, James B., Seyhan, Atilla A., Walter, Nils G., Burke, John M., and William G. Scott. "The hammerhead, hairpin and VS ribozymes are catalytically proficient in monovalent cations alone." *Chemistry & biology* 5.10 (1998): 587-595.

35. Tan, Ruoying, Chen, Lily, Buettner, Joseph A., Hudson, Derek, and Alan D. Frankel. "RNA recognition by an isolated α helix." *Cell* 73.5 (1993): 1031-1040.