UC Irvine
UC Irvine Electronic Theses and Dissertations

Title
Investigation of Phosphate-Based Reactions in the Context of an Evolving RNA World

Permalink
https://escholarship.org/uc/item/2258h4bg

Author
Kent, Alexandra

Publication Date
2021

Copyright Information
This work is made available under the terms of a Creative Commons Attribution-NoDerivatives License, available at https://creativecommons.org/licenses/by-nd/4.0/

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA,
IRVINE

Investigation of Phosphate-Based Reactions in the Context of an Evolving RNA World

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Chemistry

By

Alexandra D. Kent

Dissertation Committee:
Professor Andrej Lupták, Chair
Professor Gregory A. Weiss
Professor Rachel W. Martin

2021

Chapter 1 © 2021, John Wiley & Sons, Ltd.
Chapter 2 © 2021, Wiley-VCH GmbH
All other materials © 2021 Alexandra D. Kent

DEDICATION

To

Andrej Lupták and Jennifer Heemstra

TABLE OF CONTENTS

	Page
LIST OF FIGURES	vi
LIST OF TABLES	ix
ACKNOWLEDGMENTS	x
CURRICULUM VITAE	xii
ABSTRACT OF THE DISSERTATION	xv
CHAPTER 1: Catalytic RNA	1
1.1 Introduction	1
1.2 Catalytic Mechanisms	3
1.3 Classes of Catalytic RNAs	4
1.4 Final Remarks	30
1.5 References	31
CHAPTER 2: Enzymatic RNA Production from NTPs Synthesized from Nucleosides and Trimetaphosphate	37
2.1 Abstract	37
2.2 Introduction	38
2.3 Results	39
2.4 Conclusion	44
2.5 Acknowledgments	45
2.6 Materials and Methods	45
2.7 Supplementary Data	50
2.8 References	54
CHAPTER 3: Ribozyme Mediated Repair of Damaged and Mismatched RNA Sequences	58
3.1 Abstract	58
3.2 Introduction	59
3.3 Results	60
3.4 Discussion	71
3.5 Materials and Methods	74
3.6 Supplementary Data	84
3.7 References	100
APPENDIX A: Template-Directed, Primer-Free Polymerization of Nucleoside Triphosphates by a Polymerase Ribozyme	104
A.1 Introduction	104
A.2 Results and Discussion	104
A.3 Materials and Methods	107
A.4 References	108

APPENDIX B: Using Nucleoside Diphosphates as Substrates for Polymerization 111
B.1 Introduction 111
B.2 Results and Discussion 111
B.3 Materials and Methods 115
B.4 References 116

APPENDIX C: Effect of Substrate Energetics on Ribozymes Selected from a Random Pool 118
C.1 Introduction 118
C.2 Results and Discussion 119
C.3 Materials and Methods 122
C.4 References 124

APPENDIX D: Bioluminescent Probes for Imaging Cell-Cell Contact *In Vivo* 127
D.1 Abstract 127
D.2 Introduction 127
D.3 Results and Discussion 130
D.4 Summary and Future Directions 138
D.5 Supporting Information 139
D.6 Materials and Methods 144
D.7 References 154

LIST OF FIGURES

Page

Figure 1.1 Structure and proposed transition states of the ribosome and RNase P. 7

Figure 1.2 Structures and proposed transition states of the group I and group II self-splicing introns, and the spliceosome. 12

Figure 1.3 Structures and proposed transition states of the hepatitis delta virus and hammerhead ribozymes. 17

Figure 1.4 Structures and proposed transition states of the hairpin and *glmS* ribozymes. 23

Figure 1.5 Structures and proposed reaction mechanisms of the *in vitro* selected Diels Alderase, flexizyme, and RNA polymerase ribozyme. 29

Figure 2.1 Triphosphorylation of adenosine with cTmp. 38

Figure 2.2 Luminescence-based detection of ATP production. 39

Figure 2.3 Reaction analysis of nucleoside triphosphorylation by trimetaphosphate. 41

Figure 2.4 *In vitro* RNA production coupled to cTmp derived rNTPs. 43

Figure 2.S1 ATP synthesis does not depend on boric acid. 50

Figure 2.S2 Temperature dependence of ATP synthesis is independent of buffer identity. 51

Figure 2.S3 ATP synthesis as a function of cTmp. 52

Figure 2.S4 RNA extension by an RNA polymerase ribozyme using synthesized nucleotides. 53

Figure 3.1 Kinetic analysis of primer extension by tC19Z and inhibition by pyrophosphate. 63

Figure 3.2 Pyrophosphate-utilizing, ribozyme-mediate repair of terminal 2′-3′ cyclic phosphate damage. 67

Figure 3.3 Effects of pyrophosphate on terminally mismatched hairpins. 70

Figure 3.S1 ^{32}P-labeling and ribozyme mediated extension of a primer/template sequence. 84

Figure 3.S2 Magnesium dependence of tC19Z. 85

Figure 3.S3 Kinetics of tC19Z mediated extension of a primer. 86

Figure 3.S4 Pyrophosphate inhibits ribozyme mediated extension and co-precipitates RNA. 87

Figure 3.S5 Purification of 2′-3′ cyclic phosphate-cleaved hairpins. 88

Figure 3.S6 Hairpin controls and gel shift calibration. 89

Figure 3.S7 Full hairpin extension controls. 90

Figure 3.S8 Repair reactions and controls for all matched hairpins. 91

Figure 3.S9 Example time course of a U/A hairpin extension and pyrophosphate repair. 92

Figure 3.S10 Omission of a single nucleotide prevents extension of an A/U hairpin. 93

Figure 3.S11 α-^{32}P-rNTPs can be incorporated into repaired, non-labeled hairpins. 94

Figure 3.S12 An A/U hairpin is repaired and extended as seen by Illumina sequencing. 95

Figure 3.S13 An A/C mismatch is only extended following treatment with both PNK and pyrophosphate. 96

Figure 3.S14 Controls and repair of all mismatch combinations. 97

Figure 3.S15 Selected fluorescently labeled primer/template sequences show varying extension efficiencies. 98

Figure A.1 Ribozyme mediated, primer-free polymerization along a template. 106

Figure B.1 T7 RNA polymerase and tC19Z can use rNDPs as substrates for polymerization. 114

Figure C.1 Selection for ATP/ADP utilizing ribozymes. 121

Figure D.1 Monitoring cell-cell interactions utilizing a caged luciferin probe. 128

Figure D.2 NCL is stable, but is uncaged *in vitro.* 131

Figure D.3 NCL is uncaged *in cellulo* in a dose dependent manner. 133

Figure D.4 NCL is uncaged and consumed in cells. 134

Figure D.5 NCL can detect low numbers of bacteria. 136

Figure D.6 Circular luciferases can be used to monitor cell contact. 137

Figure D.S1 Synthetic route for NCL. 139

Figure D.S2 NMR stability study in D_2O. 139

Figure D.S3 Gel electrophoresis image and bioluminescence from NTR expression. 140

Figure D.S4 Fluorescence data generated from partially active NTR uncaging Luntr. 141

Figure D.S5 DB7-NTR cells selectively remove a nitro cage. 142

Figure D.S6 Gel electrophoresis images from TEV protease expression. 143

Figure D.S7 Fluorescence data for Figure D.2A. 144

LIST OF TABLES

Page

Table

Table 2.S1 Sequences. 54

Table 3.S1 Chapter 3 DNA and RNA sequences. 99

Table A.1 Appendix A sequences. 108

Table B.1 Appendix B sequences. 116

Table C.1 Appendix C sequences. 124

Table D.S1 Primers used for cloning. 146

ACKNOWLEDGMENTS

I would like to begin by thanking my advisor, Professor Andrej Lupták, for all his support, mentorship, and kindness over the past three and half years. From offering me chocolate that first day, to the day we finally saw pyrophosphorolysis, to when we got the NASA grant, with every incredible story and all of our arguments over gel smudges in between, being a graduate student in his lab has been an unforgettable experience. He constantly amazes me with his depth of knowledge and scientific creativity. He has taught me to think more critically about experiments, run every control reaction imaginable, and countless other important skills, and I feel confident moving forward in my scientific career as a direct result of the foundation he has helped build. He believed in me when many did not, and I will forever be grateful for that confidence. Andrej has become one of my standards of excellence both in science and in character, and I could not have had a better mentor.

I would also like to thank my committee members, Professor Greg Weiss and Professor Rachel Martin. Greg has been the only professor at UCI who I have consistently interacted with since beginning my PhD. His advice, mentorship, and career tips throughout the years have been invaluable. Although I have only spoken with Rachel a handful of times, I have admired her for many years. She has set an amazing example to follow in so many ways, and her enthusiasm for her science and hobbies is an inspiration.

I would like to thank the staff in Chemistry and across UCI for their support over the years; they have helped me immeasurably in navigating graduate school: Morgan Sibley, for always answering my questions, fixing all my administrative issues, and making sure I had TA positions; Tenley Dunn, for having an open door to listen and give advice; and Haleema Sood, for always asking how I was doing. I would also like to thank UC Irvine and the National Science Foundation Graduate Research Fellowship Program for financial support.

I would like to thank my undergraduate advisor, Professor Jennifer Heemstra. Jen allowed me to join her lab before I even started my freshman year at the University of Utah, and personally spent time teaching me chemistry and lab techniques. My pipetting skills are entirely due to her patience. Jen helped me grow as a person and as a scientist and continues to be one of the most influential people in my life. She always made time for me when I'd stop by to talk about science or just life, and has set the standard impossibly high for what a great mentor and person looks like. She always pushed me out of my comfort zone to try new things and is the reason I went to graduate school.

I would also like to thank Professor Tom Richmond, who convinced me to get a degree in chemistry and who has supported me and followed my career for many years, as well as Professor Matt Sigman, who mentored me as an undergraduate, opened the world of organic chemistry to me, and is one of the reasons I decided to attend UCI.

I would like to acknowledge all of the members of the Lupták lab. From my very first day, they made me feel incredibly welcome. Luiz Passalacqua, for all his helpful advice both in academics and Life 101, and for teaching me to make beautiful sequencing gels. Kyle Cole, for his friendship, for the late night science conversations, and for teaching me so much, including all the basics of PCR, transcriptions, purifications, etc. My undergraduate mentees, Lucy Yang and Ivee Lee, for giving me the opportunity to become a better mentor, helping me make buffers and gels, and for spontaneous lab sing alongs. Fabio Chizzolini, for teaching me that every day may not be Friday,

but there is a Friday in every day. Michael Wu, for throwing the best parties. Brandon Latifi and Allan Argelagos, for the most fun pandemic lunch club.

I would like to thank Joanna Laird for her friendship that started my first year of graduate school and has been going strong for the last six years through graduations, moves, new jobs, and switching labs. She has been there for me through all the highs and lows, supporting me and cheering me on. From weekly coffee dates that turned into weekly zoom dates and visits a few times a year, she has made my time here wonderful. I would also like to thank Shane Flynn, Abigail Vandewalle, and my Caltech friends for all the conversations, hiking, beach days, ski trips, breweries, museums, craft nights, and adventures.

I would like to thank my parents for instilling in me a love of learning, science fiction, and creativity, my sister, Skie, for talking to me on my long drives, listening to all my lab stories, encouraging me, and inspiring me, and my brother, Griffin, for his support.

Finally, I'd like to thank my fiancée, Chris, for sharing this journey with me from the very beginning. From our chance meeting at Stanford to today, our life over the past six years has been a beautiful adventure and I am looking forward to continuing to make amazing memories. I've loved learning about atmospheric chemistry, getting Meowshu, learning to ski, and just spending time with him every day. He has always had confidence in me even when I didn't have confidence in myself, and I can't imagine having done this without his love and support.

CURRICULUM VITAE

Alexandra D. Kent

Education

Ph.D. **University of California at Irvine**, Irvine, CA *2015–2021*
Chemistry, Emphasis in Chemical Biology
Advisor: Andrej Lupták

B.S. **University of Utah**, Salt Lake City, UT *2011–2015*
ACS Certified Honors Chemistry, Emphasis in Biology
Advisor: Jennifer Heemstra
Thesis: "DNA split aptamers as a biosensing platform for the detection of small drug molecules."

Professional Experience

Research

Graduate Research Assistant *2018–2021*
with Professor Andrej Lupták at the University of California at Irvine
- Demonstrated a polymerase ribozyme (tC19Z) can act in reverse to repair damaged and mismatched RNA.
- Characterized the reaction of cyclic trimetaphosphate with nucleosides to form nucleoside triphosphates.
- Investigated replacing nucleoside triphosphates with diphosphates in polymerization and capping reactions.

Graduate Research Assistant *2015–2017*
with Professor Jennifer Prescher at the University of California at Irvine
- Developed a nitro-caged luciferin and demonstrated its use in detecting cell contact events *in cellulo* and *in vivo*.
- Designed and synthesized a nitrile caged luciferin for imaging cell-cell contact.

Undergraduate Research Assistant *2011–2015*
with Professor Jennifer Heemstra at the University of Utah
- Developed an enzyme-linked small-molecule detection assay using DNA aptamers for cocaine.
- Engineered a split-aptamer based assay for the detections of steroids.
- Developed a split-aptamer, morpholino linkage assay for qPCR detection of steroids.

Teaching Assistantships

University of California at Irvine

General Chemistry I	Chem 1A	*Fall 2020*
Organic Chemistry III	Chem 51C	*Spring 2019, Spring 2020*
Medicinal Chemistry Laboratory	Chem 177L	*Winter 2020*
Organic Chemistry II	Chem 51B	*Winter 2019*

University of Utah

Organic Chemistry II	CHEM 2320	*Spring 2014, Spring 2015*
General Chemistry II	CHEM 1220	*Summer 2013*
General Chemistry I	CHEM 1210	*Summer 2012*

Selected Honors and Awards

RNA Society Travel Award	RNA Society	*2020*
Department of Chemistry Travel Award	University of California at Irvine	*2020*
Provost Fellowship	University of California at Irvine	*2015*
NSF Graduate Research Fellowship	**National Science Foundation**	***2015–2018***
ACS Membership Award	University of Utah, Dept. of Chemistry	*2015*
Goldwater Scholarship	**Barry M. Goldwater Foundation**	***2014***
Crocker Science Scholarship	University of Utah, Dept. of Chemistry	*2014*
Myriad Genetics Academic Excellence Award	University of Utah	*2014*
ACS Undergraduate Poster Award	Northwest Regional Meeting of ACS	*2014*
NCRC Innovation Challenge Award	Harvard NCRC	*2014*
Honors College Sweet Candy Scholarship	University of Utah	*2013*
James M. Sugihara Scholarship	University of Utah, Dept. of Chemistry	*2013*
Myriad Genetics Women in Science Award	University of Utah	*2013*
Mack Thomas Rozelle Scholarship	University of Utah, Dept. of Chemistry	*2013*
UROP Award	University of Utah	*2012 and 2013*
High School Research Award	University of Utah, Dept. of Chemistry	*2011*
Access Program Scholarship	University of Utah	*2011*
Honors at Entrance Scholarship	University of Utah	*2011–2015*

Professional Memberships

RNA Society	Member	*2020–Present*
American Chemical Society (ACS)	Member	*2012–Present*

Publications

5. **Kent, A. D.;** Yang, L. L.; Lee, A. I.; Lupták, A. Ribozyme mediated repair of damaged and mismatched RNA sequences. *(Manuscript In Preparation)*
4. Chizzolini, F.;* **Kent, A. D.;*** Passalacqua, L. F. M.;* Lupták, A. Enzymatic RNA production from NTPs synthesized from nucleosides and trimetaphosphate. *ChemBioChem* **2021**, 22, 2098-2101.
3. **Kent, A. D.;** Burke, C. R.; Lupták, A. Catalytic RNA, *eLS.* **2021**, John Wiley & Sons, Ltd (Ed.), 1-16.
2. **Kent, A. D.;*** Spiropulos, N. G.;* Heemstra, J. M. General approach for engineering small-molecule-binding DNA split aptamers. *Anal. Chem.* **2013**, 85, 9916-9923.
1. Sharma, A. K.; **Kent, A. D.**; Heemstra, J. M. Enzyme-linked small-molecule detection using split aptamer ligation. *Anal. Chem.* **2012**, 84, 6104-6109.

* Denotes equal contribution

Presentations

2021	RNA 2021, Annual Meeting of the RNA Society (Online)	*Oral Presentation*
2020	National Meeting of the American Chemical Society (Online)	*Oral Presentation*
2020	RNA 2020, Annual Meeting of the RNA Society (Online)	*Oral Presentation*
2020	RNA Enzymes for Emerging Life (REEL) Seminar Series (Online)	*Oral Presentation*
2016	Synthetic Chemical and Biology Club (University of California at Irvine)	*Poster Presentation*
2015	National Collegiate Research Conference (Harvard University)	*Poster Presentation*
2015	Research on Capitol Hill (The Capitol Building, Salt Lake City, UT)	*Poster Presentation*
2014	National Meeting of the American Chemical Society (San Francisco, CA)	*Poster Presentation*
2014	Northwest Regional Meeting of ACS (Missoula, MT)	*Poster Presentation*
2014	Undergraduate Research Symposium (University of Utah)	*Oral Presentation*
2014	Utah Conference for Undergraduate Research (Provo, UT)	*Poster Presentation*
2014	National Collegiate Research Conference (Harvard University)	*Poster Presentation*
2013	Science Day at “the U” (University of Utah)	*Poster Presentation*
2013	Undergraduate Research Symposium (University of Utah)	*Poster Presentation*
2012	Northwest Regional Meeting of ACS (Boise, ID)	*Poster Presentation*
2012	Undergraduate Research Symposium (University of Utah)	*Poster Presentation*
2012	ACCESS Symposium (University of Utah)	*Poster Presentation*

Service and Outreach

University of California at Irvine

Lupták Laboratory Safety Officer	*2018–Present*
Mentor to undergraduate students in the Lupták Lab	*2018–Present*

University of Utah

Undergraduate Research Opportunities Program Student Ambassador	*2014–2015*
Executive Member of the University of Utah ACS Student Chapter	*2014–2015*
Co-organizer of the University of Utah Chemistry Fair for public schools	*2014–2015*
Mentor to high school students in the Heemstra Lab	*2014–2015*
Student representative at “Science at Breakfast” community outreach program	*2013–2015*
“Science Day at the U” organizer and volunteer	*2012–2014*
Founder and designer of experiments for home school student outreach program	*2012–2014*
Volunteer high school chemistry teacher home school students	*2013–2014*
Volunteer chemistry tutor to low-income students	*2012–2013*

ABSTRACT OF THE DISSERTATION

Investigation of Phosphate-Based Reactions in the Context of an Evolving RNA World

By

Alexandra D. Kent

Doctor of Philosophy in Chemistry

University of California, Irvine, 2021

Professor Andrej Lupták, Chair

In light of the unique ability of ribonucleic acid (RNA) to both retain genetic information and catalyze chemical reactions, it has been proposed that life on Earth progressed through an RNA world. It has been demonstrated that phosphate was present on an early Earth and due to the prevalence of modern-day energy metabolisms that rely on phosphorylation, it is likely that phosphate played a key role in prebiotic chemistry. This work characterizes a number of phosphate-based reactions ranging from formation of triphosphates to RNA repair. Chapter 1 introduces catalytic RNA and summarizes the main types of reactivities that have been discovered or evolved. Chapter 2 characterizes the reaction of cyclic trimetaphosphate with nucleosides to form nucleoside triphosphates in concentrations sufficient to sustain polymerization. Chapter 3 demonstrates the first known example of ribozyme-mediated repair of damaged and mismatched RNA sequences using pyrophosphate. Taken in concert, these reactions provide key insight into the potential role of phosphate in an evolving RNA world.

CHAPTER 1: CATALYTIC RNA

Publication Note:

This paper was originally published in the journal *eLS.*

Kent, A. D.; Burke, C. W.; Lupták, A. Catalytic RNA. *eLS*. **2021**, 2, 1-16

© 2021, John Wiley & Sons, Ltd.

Contribution Statement

A.D.K., C.W.B and A.L. wrote the manuscript.

1.1 Introduction

Enzymes are biological macromolecules that accelerate chemical transformations, typically by many orders of magnitude above the background rates. Enzymes function in aqueous media or within lipid membranes of cells and are also characterised by the high specificity for the substrates of the catalysed reactions. Historically, only proteins were thought to be able to exhibit the high specificity and rate acceleration typical of enzymatic catalysis, but since the early 1980s, many ribonucleic acids (RNAs) have been added to the list. The discovery of catalytic RNAs, ribozymes, dramatically changed biochemistry and molecular biology, because these polyribonucleotides were capable of both information storage and catalysis.

Similar to protein enzymes, ribozymes fold into specific shapes that allow them to recognise their substrates and to accelerate chemical reactions. Typically, ribozymes consist of helical domains, stacked coaxially or next to each other and connected by single-stranded regions. The single-stranded regions dictate the topology of the molecule and are often responsible for tertiary contacts that specify which helical regions interact. In addition, the single-stranded regions – in the form of loops, internal loops, bulges or turns – can fold to create substrate-binding pockets and display individual residues necessary for catalysis. Often, it is the extruded, extrahelical residue that makes a contact with another domain of the ribozyme or positions a key chemical group

to act as hydrogen bond donor or acceptor or to coordinate a catalytic metal ion. Divalent ions, especially Mg^{2+}, are often coordinated by backbone phosphate groups in regions where a single-stranded RNA bends, thus providing specific binding sites for structural and catalytic metals. Other groups involved in catalysis are the 2′ hydroxyls and the Watson–Crick faces of nucleobases, which can promote proton transfer and stabilise transition states through hydrogen bonding or columbic interactions.

As with other biological macromolecules, the sequences of the RNAs dictate what three-dimensional folds they can assume. Because the structures of most ribozymes are dominated by helical regions, many parts of the sequences are complementary, and the uniqueness of the fold is determined by the specificity of base-pairing interactions that make up individual helices. The sequence of a well-folded ribozyme assumes only one set of base-pairing interactions, which give rise to the helical domains and allow specific single-stranded regions to form tertiary contacts, catalytic sites and so on. Because the structures are dominated by helical elements, the sequences that form these paired regions do not need to be specified beyond the pairing requirements. In other words, a helical segment can be composed of any sequence on one strand, as long as the sequence on the opposite strand can pair with it. Thus, a structured RNA is typically conserved on the level of secondary structure, and the residues required for function, for example catalysis and tertiary interactions, are sequence conserved. Many secondary structures of ribozymes are drawn without specific sequences in the paired domains, and only the key residues in active sites are shown explicitly (see below).

Because RNAs can catalyse many chemical transformations, it is possible to imagine a simplified biological system consisting of RNAs capable of self-replication and perhaps metabolism as well. *In vitro* selection experiments performed over the past 20 years have shown that RNAs can catalyse many different types of reactions. Laboratory evolution of the selected RNAs into multiturnover ribozymes supports the notion that RNA-catalysed RNA polymerisation and

metabolic transformation could exist in this hypothetical RNA World. Moreover, the fact that the templated protein synthesis is catalysed by the ribosomal RNA and that the decoding of the messenger ribonucleic acid (mRNA) by transfer ribonucleic acid (tRNA) is also performed by the RNA within ribosomes strongly suggest that the RNA world preceded the current protein-dominated biosphere.

In addition to catalysis, RNAs are capable of regulating gene expression through direct interactions with metabolites and other cellular factors. These riboswitches sense the concentration of their ligands using aptamer domains and couple the binding event to a domain that affects the expression of the nearby coding region, for example by promoting transcription termination, translation initiation or alternative splicing. Ligands binding to aptamer domains can also affect the activity of natural ribozymes, as substrates (guanosine in Group I self-splicing introns), cofactors (glucosamine-6-phosphate in *GlmS* ribozyme) or allosteric regulators (cyclic di-GMP regulation of a bacterial Group I ribozyme). Regulation of self-cleaving ribozymes by allosteric regulators has been extensively used in sensing of various small molecules in aptazymes (functional fusions of aptamers and ribozymes).

1.2 Catalytic Mechanisms

Biochemical and structural analyses of known ribozymes have shown that they promote catalysis through a variety of mechanisms. The ribosome catalyses peptide bond formation in *trans* by bringing the aminoacylated tRNA into the proximity of the peptidyl tRNA, activating the nucleophile (amino group on the aa-tRNA), stabilising the tetrahedral transition state on the peptidyl tRNA and promoting the breakdown of the intermediate to form a new peptide bond. Other biological ribozymes are phosphotransferases that act either in *trans*, as in the case of Ribonuclease P (RNase P), which processes tRNAs and other cellular RNAs from longer transcripts, or in *cis*, that is on the same RNA chain that codes the ribozyme, as is the case for self-splicing

introns and self-cleaving ribozymes. To promote phosphoryl transfer, these ribozymes act either directly, using their nucleobase and sugar moieties in general acid–base mechanisms to promote proton transfer, or coordinating divalent metal ions that stabilise the developing negative charges of transitions states. Metal ions also activate water molecules for direct nucleophilic attack or to act in proton transfer. Most ribozymes specifically use Mg^{2+}, while others act in the presence of a variety of divalent metal ions. The extreme case appears to be the hepatitis delta virus (HDV) family of self-cleaving ribozymes, which uses a metal ion to activate the nucleophile and potentially stabilise the transition state, but have little preference for the identity of the metal, accommodating Mg^{2+} , Mn^{2+} , Ca^{2+} , Co^{2+} and other divalent metals for catalysis. Two of the ribozymes described herein are thought to use a hybrid strategy, employing both metal ions and nucleobases to promote catalysis.

1.3 Classes of Catalytic RNAs

Ribosome

The ribosome is responsible for protein synthesis and is conserved throughout all cellular life forms.[1,2] The ribosome binds mRNA and translates the 4 standard nucleotides into more than 20 amino acids. The amino acid sequence is determined by groups of three nucleotides on the mRNA called codons. Each codon defines a particular amino acid that is carried to the ribosome via tRNA. The ribosome is responsible for binding the mRNA, decoding the nucleotide sequence, binding the correct tRNA and performing the peptidyl transferase reaction that lengthens the nascent peptide by one amino acid (**Figure 1.1**). This series of events happens in multiple turnovers and with a low error rate, making the ribosome the most efficient ribozyme known to date. The ribosome is composed of two subunits, large and small, each consisting of both RNA and protein moieties (**Figure 1.1**). There are three tRNA binding sites present: A-site binds the aminoacyl-tRNA, P-site binds the tRNA linked to the nascent peptide and E-site binds the free tRNA before

release. The ribosome catalyses two distinct biochemical reactions: RNA-catalysed peptide bond formation during protein synthesis and peptidyl hydrolysis during the release of the completed polypeptide mediated by protein release factors. For peptide bond formation, the carbonyl carbon of the ester bond that links the tRNA to the synthesised peptide at the P-site is nucleophilically attacked by the α amino group from the A-site-bound aminoacyl-tRNA, forming a peptide bond (**Figure 1.1c**). The reaction for peptidyl hydrolysis before the release of the completed protein is similar where the nucleophile is a water molecule, rather than the aminoacyl-tRNA. Compared to the uncatalysed peptidyl transferase reaction, the ribosome increases the rate by a factor of a million.[3] This reaction occurs in a two-step mechanism rather than a concerted proton shuttle as previously thought. The formation of the tetrahedral intermediate occurs simultaneously with the deprotonation of the amine, followed by fast break down of the intermediate yielding the products. It was once thought that the role of ribosome in pep- tide synthesis was simply to bring the reactants together; however, the ribosome does not only alter the energy landscape to allow for concerted nucleophilic attack and deprotonation, it also affects the rate of formation and decomposition of the intermediate.[4,5] It was unclear up until the later part of the twentieth century whether RNA or protein domains were responsible for peptidyl transferase activity. Early experiments in which ribosomal extracts were treated with high concentrations of protein denaturants showed that the ribosome activity remained with approximately 80% of the protein subunits removed. Further treatment with EDTA and phenol abolished the activity, suggesting that the RNA was the catalytically active component.[6] It was not until high-resolution X-ray crystallographic structures of the large subunit were available that the ribosome was proven to be a ribozyme. An inhibitor that mimics the transition state analogue of the peptidyl transferase reaction was co-crystallised with the large subunit, and its location determined the catalytic active site. There were no protein side chains within 18 Å of the active site, providing the last bit of evidence that the RNA was indeed solely responsible for catalysis.[7,8]

Ribonuclease P (RNase P)

Long before the true catalytic activity of the ribosome was understood, RNase P was discovered to be the first known naturally occurring multiple turnover ribozyme.[9] RNase P is composed of both protein and RNA subunits (S- and C-RNA domains) and is responsible for the processing of the 5′-ends of tRNA (**Figure 1.1d–f**) and several other noncoding RNAs, including 5S rRNA, U6 small nuclear RNA (snRNA) and SRP RNA.[10] Although both the protein and RNA subunits are necessary for activity *in vivo,* the RNA subunit alone can catalyse the phosphodiester cleavage reaction in low divalent magnesium conditions *in vitro*.[11] The RNA core is highly conserved through bacterial, archaeal and eukaryotic species, whereas the size and number of protein domains in the complex vary.[12] Unlike other large ribozymes, RNase P can accept a variety of pre-transfer RNA (pre-tRNA) substrates independent of primary sequence and therefore relies on interactions other than base-pairing near the active site, such as π–π stacking, hydrogen bonding and metal–ion coordination to correctly position the substrate for cleavage. The only restrictions for substrate recognition are the T-stem and loop structure that stacks with the S domain of the ribozyme and a highly conserved 3′ CCA primary sequence that forms canonical base-pairing with the L15 loop of the C domain.[13,14] Hydrolysis of the phosphodiester bond necessary for the cleavage reaction depends on the presence of divalent metal cations such as magnesium. Crystal structures of the ribozyme show two metals in the active site (M1 and M2) and suggest a transition state-based model in which the first metal (M1) interacts with the nonbridging scissile phosphate oxygens, allowing a hydroxide (or metal-coordinated hydroxide) to perform an S_N2 nucleophilic substitution. The second metal (M2) appears to be responsible for stabilisation of the transition state and proton donation to the 3′ scissile oxygen upon release of the product (**Figure 1.1g**).[14]

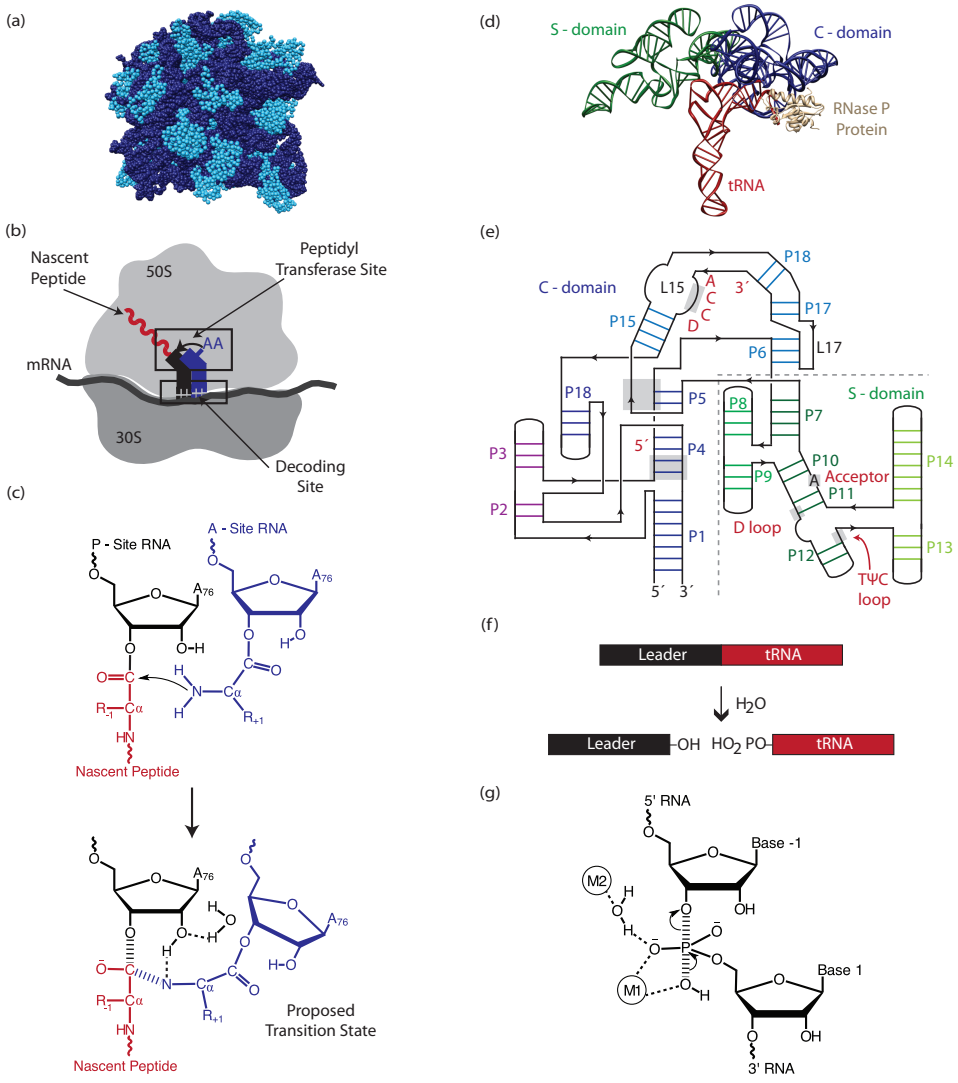


Figure 1.1. Structures and proposed transition states of the ribosome and RNase P. (a) The ribosome is responsible for mRNA-templated protein synthesis in all cellular life forms and is composed of both RNA (dark blue) and protein domains (light blue). (b) For accurate and efficient protein synthesis, the ribosome binds mRNA (dark grey) and translates the codons into amino acids (decoding site). The ribosome then binds the correct tRNA (blue) corresponding to the co-dons and preforms a peptidyl transferase reaction to increase the length of the nascent peptide

(red) by one amino acid. (c) The carbonyl carbon that links the nascent peptide to the tRNA at the P-site undergoes a nucleophilic attack by the A-site-bound aminoacyl-tRNA. The proposed transition state was determined by kinetic isotope effect analysis and shows a two-step mechanism in which the proton donation from the nucleophilic nitrogen and the formation of the tetrahedral intermediate are both in the rate-limiting step, with the fast breakdown of the tetrahedral intermediate following. (d) Crystal structure of the *Thermotoga maritima* RNase P holoenzyme (S (green) and C (blue) domains and a small protein cofactor (tan)) in complex with processed tRNA (red). The ribozyme–product complex illustrates the intermolecular contacts necessary for substrate binding and processing. (e) The secondary structure of RNase P denotes the S and C domains (shades of green and blue, respectively) along with the tRNA identification sites (red) and contacts (grey boxes). The 5′-end of the tRNA marks the location of the active site near stems P4 and P5, while the 3′ terminus is base-paired with L15 and is necessary for identification of the particular type of tRNA. (f) RNase P is mainly responsible for the processing of the 5′-ends of tRNA, although it does process other RNAs. A hydroxide, or metal-coordinated hydroxide, acts as the nucleophile attacking the phosphate at the cleavage site, resulting in a 3′-hydroxyl on the 5′-leader and a 5′-phosphate on the processed tRNA. (g) Two metal ions have been proposed to stabilise the transition state. The first metal (M1) is proposed to interact with the nonbridging scissile phosphate oxygens, allowing the nucleophilic hydroxide to attack. The second metal (M2) is necessary for stabilisation of the transition state and proton donation to the 3′ hydroxyl on the 5′ product.

Introns

Eukaryotic pre-messenger RNAs (pre-mRNAs) contain long non-coding sequences (introns) dispersed throughout protein-coding regions (exons). To form functional mRNA with correctly positioned codons, these introns must be excised from the parent RNA, and the coding exons ligated together. This cleavage and ligation is termed splicing. In eukaryotes, the

spliceosome is responsible for this process; however, self-splicing of certain introns is known to occur in some eukaryotes, bacteria and bacteriophages. Self-splicing introns are ribozymes that can be divided into two groups (Group I and Group II), based on the mechanism of the splicing reaction and secondary and tertiary structural differences. It has been shown that the eukaryotic spliceosome is a protein-directed metalloribozyme.[15,16]

Group I self-splicing introns

Group I self-splicing introns are widely distributed in a number of protists, bacteria and bacteriophages and interrupt transcripts of pre-mRNA, pre-rRNA and pre-tRNA.[9,17,18] The *Tetrahymena thermophila* ribozyme is the most extensively studied self-splicing Group I intron. The *T. thermophila* 26S pre-rRNA-coding region is interrupted by an intron that is 413-nucleotides long. This pre-rRNA is able to undergo splicing *in vitro* in the absence of any cellular extracts.[19] Pre-rRNA undergoes self-splicing via two consecutive transesterification reactions. An exogenous guanosine, or one of its 5′ phosphorylated forms, attacks the phosphate at the 5′ splice site, generating a new 3′–5′ phosphodiester bond. The resulting 5′-exon contains a free 3′ hydroxyl group which is available for attack of the phosphate at the 3′ splice site, subsequently ligating the exons producing mature rRNA (**Figure 1.2**).[20] The excised intron is often cyclised in a subsequent self-ligation reaction.[21] One of the smallest Group I self-splicing introns known to date is located in the pre-tRNA anticodon loop of the purple bacteria *Azoarcus*. Extensive crystallographic data have been obtained, including a crystal structure of the full-length intron in complex with both the 5′ and 3′ exons preceding the second step of splicing (pre-2s state) (**Figure 1.2a**). The 5′-exon has been cleaved by the first transesterification reaction and remains base-paired to the internal guide sequence (IGS) located in the P1 helix. The 3′-exon is covalently linked to the intron and is base paired to the IGS at the P10 helix. The guanosine nucleophile (ΩG) is located at the 3′-end of the intron during the pre-2s state and is attacked by the conserved uridine (U-1) of a conserved

G-U wobble pair at the IGS of the 5′-exon (**Figure 1.2b**). There are two metal ions located near the scissile phosphate. The first metal, M1, activates the nucleophile, while the second metal, M2, stabilises the leaving group. Both metals are thought to be magnesium ions and interact with the oxygens associated with the scissile phosphate (**Figure 1.2d**).[22,23]

Group II self-splicing introns

Group II self-splicing introns have been found in bacterial genomes and in the organellar genomes of fungi, plants, protists and an annelid worm.[24] These ribozymes have been shown to undergo splicing *in vitro* under high salt conditions with kinetics 10 times slower than those *in vivo*. This observation suggests that although the correctly folded RNA can be catalytically active there are likely protein cofactors necessary to facilitate the splicing reactions *in vivo*.[25-27] Group II self-splicing introns are characterised by their conserved secondary and tertiary structures consisting of six domains (DI–DVI), with domain DV containing the active site (**Figure 1.2e,f**). Although the secondary structure is similar among these self-splicing introns, the composition of the active site can vary, giving rise to three subclasses termed IIA, IIB and IIC. The mechanism for splicing is similar in all classes and to that of the Group I introns except that a catalytic adenosine, rather than a guanosine, provides the nucleophile. Two *trans*-esterification reactions are necessary for the splicing mechanism: The first is initiated by a bulged adenosine in DVI that acts as a nucleophile attacking the 5′ splice site by its 2′ hydroxyl liberating a 3′ hydroxyl that acts as a nucleophile during the second step, attacking the 3′ splice site. This second *trans*-esterification reaction results in the ligation of the coding exons and the excision of the intron (**Figure 1.2g,h**). An alternative mechanism has been known to occur in which a water molecule replaces the adenosine as the original nucleophile. There are two divalent metals located in the active site. Both metals are thought to be catalytic and agree with the two metal ion reaction mechanisms that the other large ribozymes display.[28,39] The active site also contains the catalytic triad ACG and an AY bulge, both

of which coordinate the catalytically relevant magnesium ions.[24]

Spliceosome

Splicing of eukaryotic pre-mRNA is achieved by the ribonucleoprotein machinery known as the spliceosome.[30,31] Cryo-EM studies have captured key states including assembly, activation, catalysis and disassembly, revealing that the spliceosome is, in fact, a ribozyme assembled and structurally assisted by protein cofactors (**Figure 1.2i,j**).[16] The splicing that occurs within the core of the structure is catalysed by two metal ions bound by RNA. Each splicing event comprises a two-step phosphoryl transfer reaction that is reminiscent of the group II intron self-splicing mechanism, where the lariat is formed during the first step, and ligation of the coding exons in the second step (**Figure 1.2g,h**). The spliceosome includes five snRNAs and up to approximately 100 proteins and is not a preassembled structure, in that it assembles anew for each splicing event. Both the 3′ and 5′ splice sites have conserved sequences that directly precede or follow them within the intron that are recognised by the splicing machinery. Additionally, the branch point that acts to form the lariat is conserved, with a key adenosine acting as the residue involved in the transesterification reaction, also similar to the branch point of group II introns. Based on structural analysis of group II introns and the spliceosome, it has been suggested that over time proteins evolved to replace sections of the extensive RNA machinery in group II introns, while the catalytic core remained largely unaltered.[32]

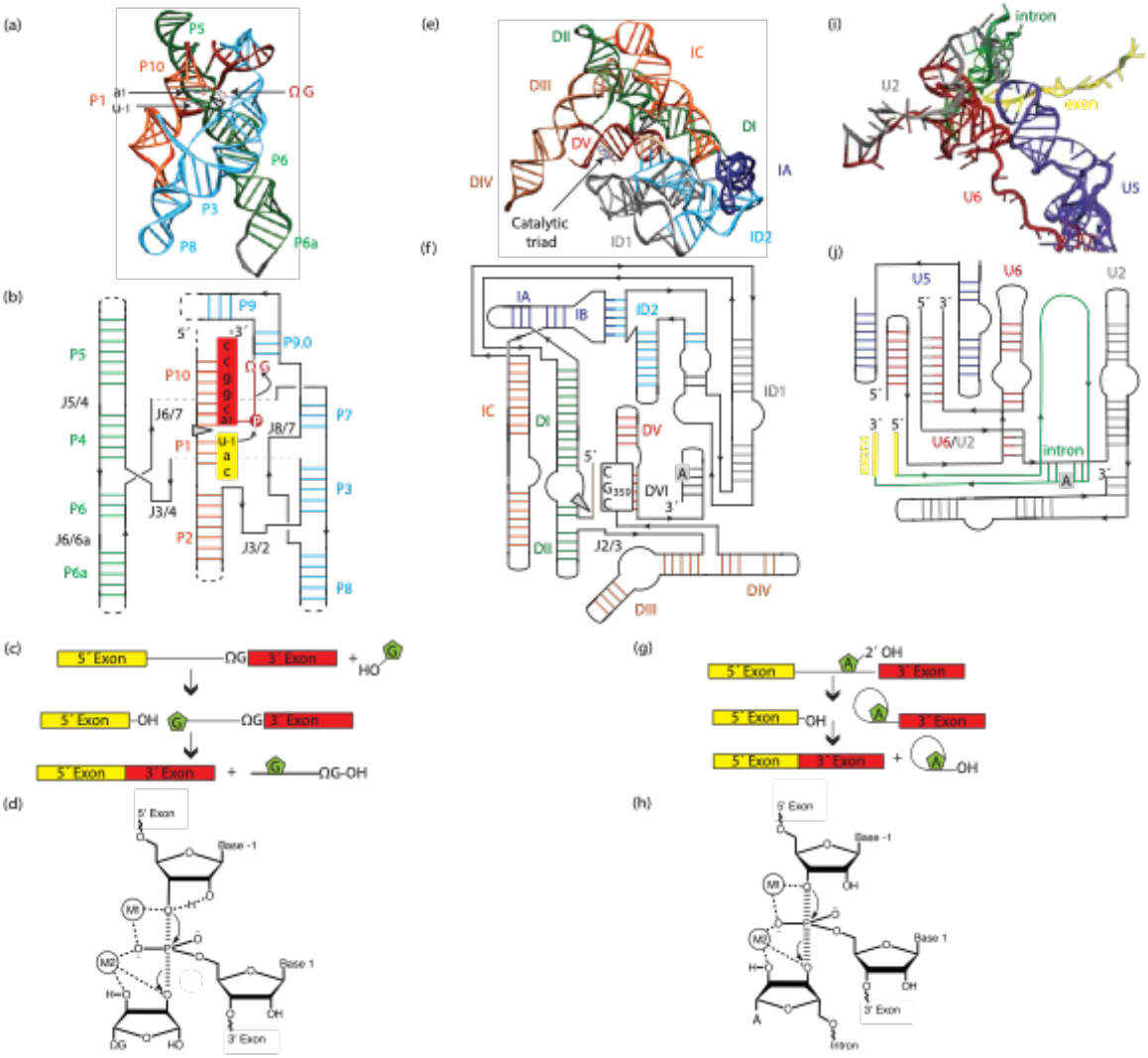


Figure 1.2. Structures and proposed transition states of the group I and group II self-splicing introns and the spliceosome. (a) Crystal structure of the purple bacterium *Azoarcus* group I self-splicing intron in complex with both the 5′ - and 3′ -exons from the tRNA anticodon loop. This structure corresponds to the intermediate after the first of two consecutive trans-esterification reactions. The crystal structure shows the 3′ hydroxyl of the 5′-exon in line for nucleophilic attack on the scissile phosphate, which results in the excision of the intron and ligation of the exons. The catalytic residues are explicitly shown in the active site. (b) There are three coaxially stacked helical elements: P5, P4, P6 and P6a (green); P10, P1 and P2 (orange) and P9.0, P7, P3 and P8

(blue). The 3′ and 5′ exons are depicted in red boxes with lower case letters and are base-paired with P10 and P1, respectively. The splice site is shown as a grey arrowhead in both the secondary and crystal structures. (c) The splicing reaction proceeds through two trans-esterification reactions in which an exogenous guanosine attacks the 5′ splice site and generates a new 3′ –5′ phosphodiester bond. The resulting 5′ -exon contains a free 3′ hydroxyl that is available for attack on the 3′ splice site, completing the splicing reaction. (d) The proposed transition state of the second *trans*-esterification reaction, corresponding to the crystal and secondary structures, is shown with two metal ions coordinated in the active site yielding the ligated exons. (e) Crystal structure of the group II intron from *Oceanobacillus iheyensis* in the precatalytic state. The G359A mutant abolishes catalytic activity by changing the conserved G·U wobble to an A–U base pair, while maintaining similar overall structure to the postcatalytic state, and represents the first crystal structure of an intron prior to splicing. The catalytic triad and A376 are explicitly shown, as well as the 5′-exon (tan), and the cleavage site with a grey arrowhead. (f) The ribozyme contains six domains (DI–DVI), with DI (green) being the largest. The most highly conserved structure and catalytic core is contained within DV (red). This includes a hairpin secondary structure and the 'catalytic triad' of conserved nucleotides and is located in close proximity to the catalytic bulged adenosine in DVI that acts as the nucleophile. (DVI is not shown in the crystal structure). (g) Two trans-esterification reactions are necessary for both group II intron and spliceosome splicing. The first is initiated by an adenosine that acts as a nucleophile in the 5′ splice reaction using its 2′ hydroxyl to liberate a 3′ hydroxyl which then attacks the 3′ splice site, resulting in ligation of the exons and excision of the intron. (h) The proposed transition state of the second *trans*-esterification reaction in which at least two metal ions are coordinated in the active site. (i) Cryo-EM structure of RNA catalytic core of the spliceosome from *Saccharomyces cerevisiae* directly following lariat formation and the first transesterification reaction. Three snRNAs involved in catalysis are shown as well as the intron (green) and exons (yellow). The branch point remains within the

U2/U6 complex close to the catalytic Mg^{2+} and the 5′ splice site. (j) Three domains of the spliceosome of the catalytic core are shown (U2, U5 and U6). Some helices are functionally similar to those of the group II introns: core helices of U6 correspond to DV, and the green base pairs between the intron and U2 can be related to DVI. The branch point adenosine is highlighted, and the exons and intron are shown in yellow and green, respectively.

Nucleolytic Ribozymes

A set of small self-cleaving ribozymes, called nucleolytic ribozymes, exist that catalyse the same internal transesterification reaction but fold into vastly different secondary structures and active sites.[33] These ribozymes have provided the foundation for studying the biological, structural and catalytic consequences of functional RNAs. The self-scission reaction proceeds via an S_N2-like mechanism with the hydroxyl at the 2′ position upstream of the cleavage site attacking the scissile phosphate. A pentagonal bipyrimidal transition state occurs and resolves into a 2′–3′ cyclic phosphate on the 5′-product and a 5′-hydroxyl on the 3′-cleavage product. This mechanism is the same in all five classes of small nucleolytic ribozymes, with active site participants and rates of the forward self-cleaving and reverse ligation reaction varying.[34]

HDV-like self-cleaving ribozymes

The HDV is a small single-stranded satellite virus of the human hepatitis B virus.[35] Rolling circle replication of the virus utilises two self-cleaving ribozymes (one on the genomic and the other on the antisense strand) to process the RNA into unit-sized transcripts. The first HDV-like ribozyme was discovered through *in vitro* selection from a human genomic library, which mapped the ribozyme to an intron of the mammalian cytoplasmic polyadenylation element-binding protein 3 (*CPEB3*) gene.[36] Sequence comparison of the *CPEB3* and HDV ribozymes and their homologs revealed only six invariant nucleotides but a conserved nested double-pseudoknot secondary

structure (**Figure 1.3a,b**). This complex fold is characterised by five base-paired regions that form two coaxial helices (P1 stacked with P1.1 and P4, and P2 stacked with P3) that are connected by single-stranded regions (J 1/2 and J 4/2). The six conserved nucleotides include an active site cytosine (C75), a guanosine–pyrimidine base pair in the P1 helix, an adenosine in J 4/2 that forms A-minor interactions with P3 and a G·U reverse wobble in L3 at the base of P3.[37] Structure-based bioinformatics searches revealed that these ribozymes are widespread across many genomes.[38] Self-scission proceeds though a general acid–base mechanism (**Figure 1.3c**). A general base activates the 2′-hydroxyl upstream of the cleavage site. This allows for nucleophilic attack on the scissile phosphate, generating a 2′–3′ cyclic phosphate on the 5′-cleavage product and a 5′-hydroxyl on the 3′-cleavage product. The cleavage reaction for HDV-like ribozymes requires low concentrations of divalent metal cations such as magnesium or calcium. A magnesium ion is coordinated to the conserved G·U reverse wobble pair and interacts directly with the scissile phosphate, likely acting as a general base to deprotonate the nucleophile and to stabilise the phosphorane transition state. The C75 nucleobase likely acts as a general acid to protonate the leaving group.[39,40]

Hammerhead ribozymes

Hammerhead ribozymes were first discovered as small satellite RNAs in plant viroids but are now known to reside in many bacterial and eukaryotic genomes.[41] The catalytically active minimal hammerhead structure is represented by three helical elements connected by a three-way junction with a highly conserved catalytic core at the centre (**Figure 1.3d,e**). There are three topologies of hammerhead ribozymes, denoted Type I, II and III, based on which stem is open to the remainder of the transcript. The fastest of these ribozymes include tertiary interactions between regions I and II, first predicted and tested *in vitro* and later confirmed in the crystal structure (**Figure 1.3d**).[42-44] Hammerhead ribozymes catalyse self-scission and ligation, with scission 100

times more likely than the reverse reaction. Catalytic activity is detectable in low concentrations of divalent metal ions such as magnesium and also in high concentrations of monovalent ions such as sodium, lithium and ammonium. Because of this, the role of the divalent metal is thought to be structural and not directly involved in catalysis.[45] Structural studies of an inhibited form of the ribozyme revealed that the 2′-hydroxyl of the conserved G8 base is hydrogen bonded to the scissile phosphate, suggesting charge stabilisation by supplying a proton to the leaving group and acting as the general acid. The conserved G12 has direct contact with the C17 cleavage site and is suspected to act as the general base deprotonating the 2′-hydroxyl of the cleavage site (**Figure 1.3f**).[44]

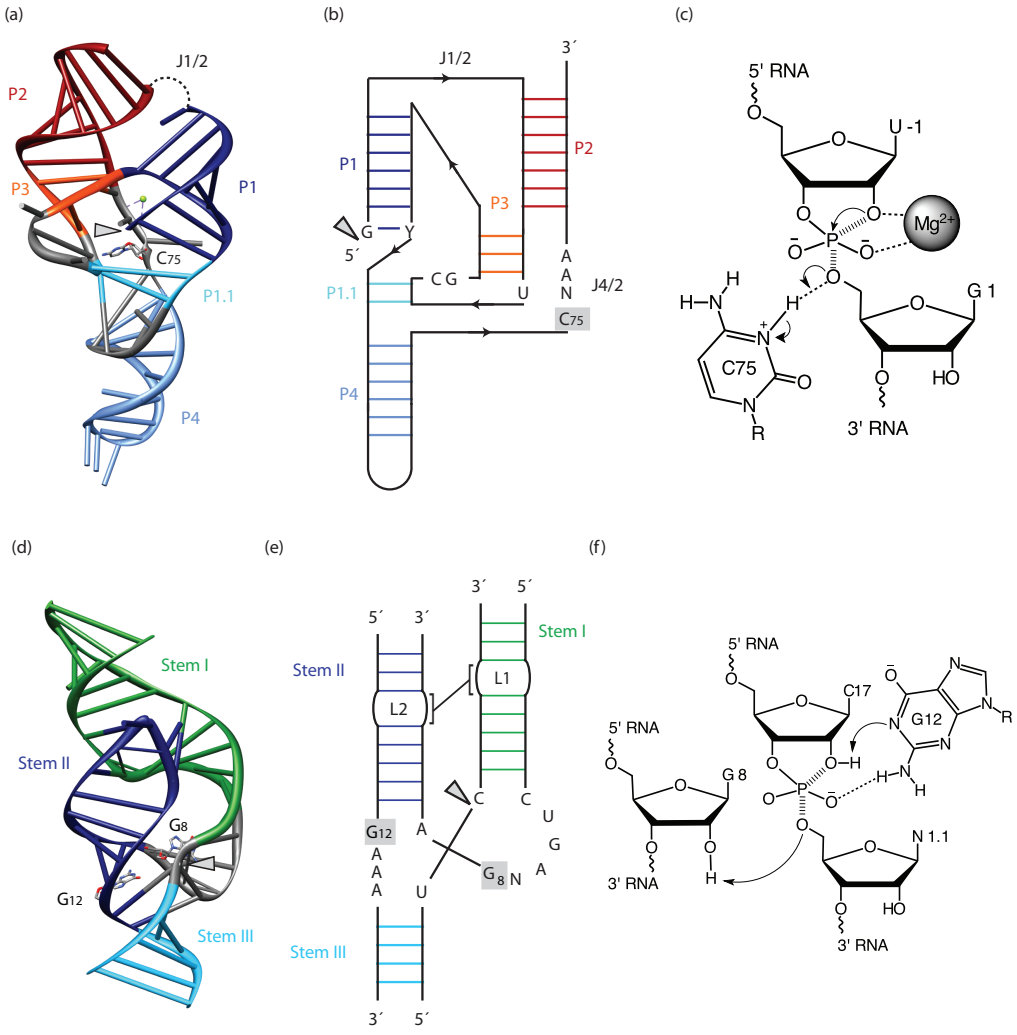


Figure 1.3. Structures and proposed transitions states of the hepatitis delta virus and hammerhead ribozymes. (a) Crystal structure of the HDV ribozyme in the precleaved state. To inhibit the ribozyme self-cleavage and to capture the precleavage structure, the ribozyme is bound to an inhibitor RNA that contains a deoxynucleotide at the cleavage site. (b) This ribozyme folds into the complex nested-double pseudoknot structure with two coaxial helices; P1 stacked with P1.1 and P4 (shades of blue), and P2 stacked with P3 (shades of red) depicted in the secondary structure. An active site cytosine is shown explicitly as is a divalent metal ion with the site of self-

scission indicated by a grey arrowhead. (c) The self-cleavage reaction most likely proceeds through a general acid–base mechanism. A hydrated magnesium ion is suggested to act as a general base, deprotonating the 2′ hydroxyl upstream of the cleavage site and stabilising the phosphorene transition state. This coordination allows for nucleophilic attack on the phosphate backbone, generating a 2′–3′ cyclic phosphate on the 5′ product and a 5′ hydroxyl on the 3′ product. The C75 nucleobase likely acts as the general acid, protonating the leaving group. (d) Crystal structure of the full-length *Schistosoma mansoni* hammerhead ribozyme. The full-length ribozyme reveals tertiary contacts between stems I and II that account for the 1000-fold catalytic enhancement relative to the minimal sequence. (e) A general secondary structure denoting the fold of the hammerhead ribozyme. There are three classes – Type I, II and III – depending on which stem is open to the remainder of the transcript. Catalytic residues G12 and G8 are shaded in grey in the secondary structure and explicitly shown in the crystal structure along with a grey arrowhead indicating the site of cleavage in both. (f) The self-scission proceeds through a general acid–base pathway in which the conserved C12 residue has direct contact with the cleavage site and acts as the general base deprotonating the 2′ hydroxyl. The conserved G8 residue is hydrogen bonded to the scissile phosphate, providing stabilisation of the transition state. G8 may also act as a general acid donating a proton to the leaving group.

Twister and twister sister ribozymes

Twister ribozymes were discovered during a bioinformatics search and were found to commonly occur in *Clostridia* bacteria as well as eukaryotic organisms.[46,47] Twister is similar in structure and function to hammerhead ribozymes and can also be found with active circular permutations unique to this class of ribozymes. Twister is composed of three essential stem loops, with the potential for three additional stems with a structure stabilised by dual pseudoknots. Crystal structures indicate that the active site is composed of a pseudoknot and a junction with a cleavage

site following a highly conserved adenine.[48] An additional highly conserved guanine has been shown to stabilise the transition state and acts as a general base in the transesterification reaction. Finally, the cleavage reaction is likely influenced by the inner sphere coordination of an Mg^{2+} ion.

The twister sister ribozyme was discovered when searching for additional conserved RNA structures in the vicinity of other known self-cleaving ribozymes. Twister sister is similar in structure and sequence to the twister ribozymes, hence the name, but cleaves at a different site. It also lacks the characteristic pseudoknot structure that is the hallmark of the twister ribozymes. Twister sister cleaves between a cytidine and adenine, with the adenine being highly conserved. It also shows a similar dependency on pH and Mg^{2+} concentration, with a water molecule coordinated to an Mg^{2+} ion found within hydrogen-bonding distance of the scissile phosphate oxygen. The crystal structure of one variant revealed a four-way junction with highly conserved residues within the junction.[49]

Pistol and hatchet ribozymes

The pistol and hatchet ribozyme classes were discovered in the same bioinformatics search that revealed the twister sister ribozyme.[47] They also primarily occur in close proximity to other self-cleaving ribozymes. The pistol ribozyme self-cleaves via an internal *trans*-esterification with a mechanism reminiscent of hammerhead ribozymes. The cleavage site is between a guanosine and uridine with the scissile phosphate on the uridine. The structure has a pseudoknot critical for cleavage and three helical stems as well as three highly conserved adenines that fix the geometry for active site formation. A purine at position 32 likely acts as the general acid during catalysis, and G40 acts as the general base to deprotonate the 2′ -OH nucleophile.[50]

Unlike pistol, hatchet cleaves at the 5′-end of its sequence and is the only class of self-cleaving ribozymes that does so other than HDV ribozymes. The cleavage site is located at the

centre of the folded ribozyme near the termini of the four stems of the structure, and self-cleavage occurs between a uridine and a cytidine. Near the cleavage site are two highly conserved guanosine bases, G31 and G30, that likely act as the general acid and base for the reaction. Although the ribozyme has an Mg^{2+} dependence, it is not yet clear what role it plays in the catalytic mechanism.[51]

Hairpin ribozyme

The hairpin ribozyme is naturally found on the negative strand of the tobacco ringspot virus satellite RNA and participates in RNA processing during replication.[52,53] This class of nucleolytic ribozymes catalyses the reversible transesterification reaction of site-specific cleavage and ligation. Unlike the hammerhead ribozyme, the hairpin ribozyme is a far better ligase than nuclease. The hairpin ribozyme structure consists of a four-helical junction with adjacent arms (stem A and stem B), each containing internal loops (**Figure 1.4a,b**). The interactions between the loops are critical for catalysis, and the minimal structural motif can be reduced to a two-stem hinge with stems A and B; however, the reaction of this minimal construct requires significantly higher concentrations of divalent cations to correctly fold the ribozyme. These interactions were verified though X-ray crystallography.[54] A conserved nucleobase G8 is positioned to hydrogen bond with the nucleophilic 2′-hydroxyl of the cleavage site, suggesting that it plays the role of general base abstracting the proton-initiating catalysis. The A38 nucleobase acts as the general acid donating a proton to the active site, resulting in self-cleavage. A transition state vanadate analogue was also crystallised and showed that additional hydrogen bonds are made with the transition state than the product or precursor states (**Figure 1.4c**).[55] This suggests that stabilisation of the transition state is important for catalysis.

Neurospora VS ribozyme

The *Neurospora* Varkud satellite (VS) ribozyme is one of the largest nucleolytic ribozymes

identified to date.[56] It is naturally found as an abundant satellite transcript from the mitochondrial DNA of a single isolate of *Neurospora crassa*. The structure of the VS ribozyme is composed of five helical elements organised by two three-way junctions, with the upper junction composed of helices III, IV and V, and the lower junction composed of helices II, III and VI. The 5′ stem loop contains the active site with an internal bulge housing the scissile phosphate. This stem loop can be easily isolated from the catalytic core, and consequently, the VS ribozyme has been shown to function in the *trans* configuration *in vitro*.[57] Unlike many of the other small nucleolytic ribozymes, the contacts made between the substrate and core ribozyme are mostly tertiary with the loop of the substrate stem contacting helix V. The catalytic activity of the cleavage and reverse ligation reactions are highly dependent on the lengths of helices III and V and on the presence of an adenosine residue in the internal bulge of helix VI. Recent X-ray scattering experiments show an additional three-way junction in the substrate domain of the *cis* form of the ribozyme, potentially aiding in the folding of the active structure. The transition state is proposed to be similar to that of the hairpin ribozyme, with conserved G638 acting as the general base abstracting a proton from the 2′-hydroxyl at the cleavage site. A conserved A756 likely acts as the general acid donating a proton, resulting in a 5′ -hydroxyl on the 3′ -product of self-cleavage.[58]

glmS ribozyme

The *glmS* ribozyme was first discovered in *Bacillus subtilis* and 17 other Gram (+) bacteria.[59] The putative 5′-untranslated region (UTR) upstream of the *glmS* gene was shown to have high sequence conservation and nucleotide covariance consistent with several conserved secondary structural elements. The *glmS* gene encodes the enzyme glutamine-fructose-6-phosphate amidotransferase, which is responsible for the production of gluclosamine-6-phsophate (GlcN6P) from fructose-6-phosphate. *In vitro* studies showed that when incubated with GlcN6P, the most highly conserved portion of the upstream *glmS* gene self-cleaves, leading to the hypothesis that

this activity of the ribozyme is dependent on the concentration of GlcN6P. *In vivo* studies confirmed that the *glmS* ribozyme is a riboswitch regulating the expression of the *glmS* gene.[60] The tertiary structure is more complex than the other nucleolytic ribozymes with several coaxial helices and a double pseudoknot, as shown by the X-ray crystallographic data (**Figure 1.4d**).[61,62] The catalytic core is contained within P2.1 and P2.2 with six nucleotides responsible for direct binding of GluN6P (**Figure 1.4e**). Core catalytic activity is independent of metal ions, leading to the hypothesis that they are structural and not catalytic. The deprotonated N1 of nucleobase G40 is modelled to be responsible for catalytic initiation by acting as the general base and deprotonating the 2′-hydroxyl at the cleavage site. The general acid is the protonated form of GlcN6P and results in completion of the self-cleavage reaction (**Figure 1.4e**).[59]

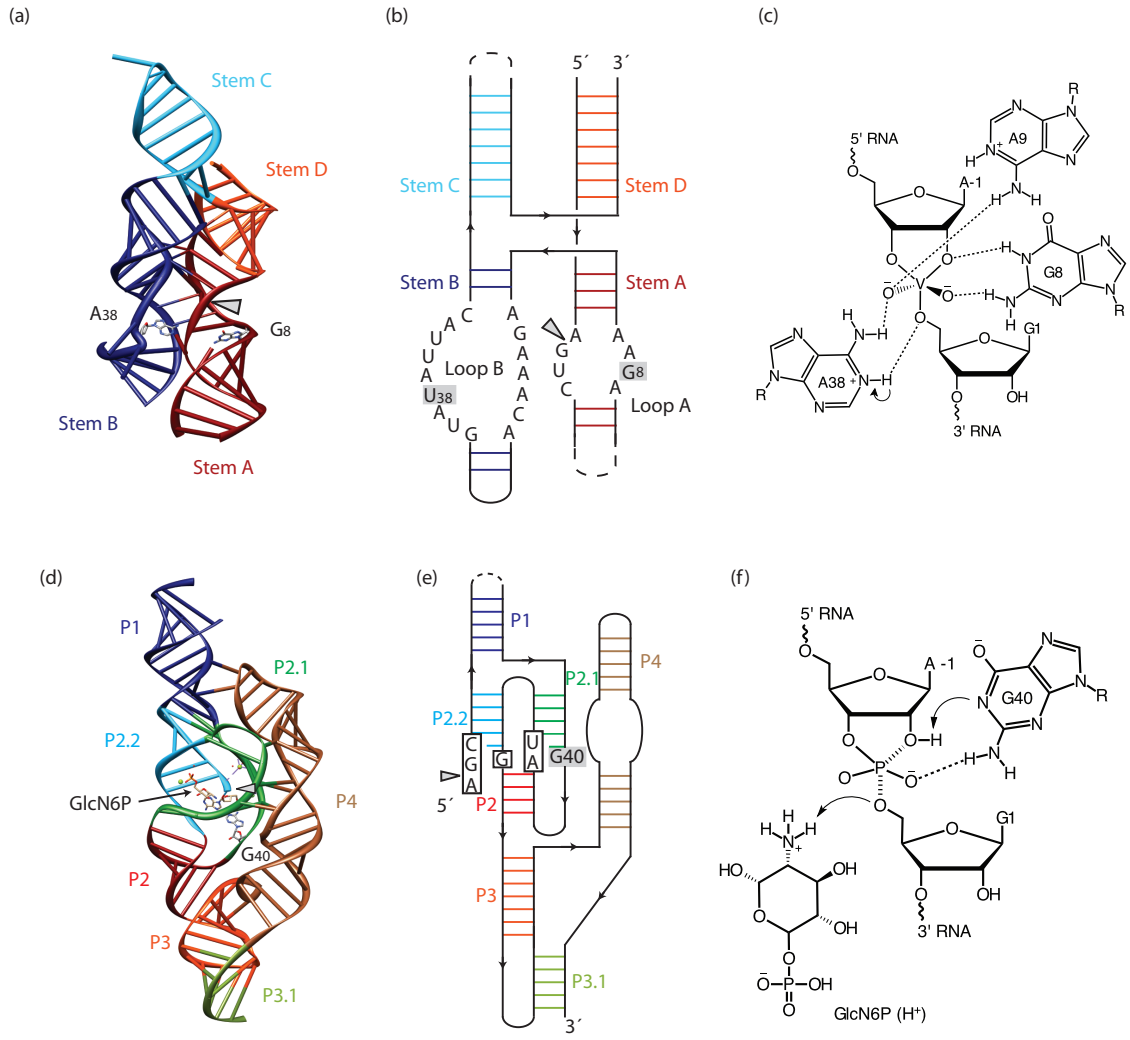


Figure 1.4. Structures and proposed transition states of the hairpin and *glmS* ribozymes. (a) Crystal structure of the hairpin ribozyme in the precleavage form. The ribozyme is trapped in the precursor state by substitution of the nucleophilic 2′ hydroxyl with a methoxy group, retaining structure but abolishing self-cleavage activity. (b) Secondary structure of the hairpin ribozyme. The ribozyme consists of four helical stems centred on a four-way junction. There are two coaxial stacks, with stem A stacked on stem D (shades of red), and stem C stacked on stem B (shades of blue). Stems A and B are docked. (c) The active form of the ribozyme catalyses reversible transesterification reactions of site-specific self-cleavage and ligation. The proposed transition

state of the self-cleavage reaction is shown with a vanadate transition state analogue. Nucleobase G8 in loop A hydrogen bonds with the nucleophile suggesting that it acts as the general base deprotonating the 2′-hydroxyl initiating the reaction. The transition state is proposed to be stabilised by nucleobase A38 through hydrogen bonding, and it is also suggested to act as the general acid donating a proton to the leaving group. (d) Crystal structure of the *Bacillus anthracis glmS* ribozyme in complex with substrate gluclosamine-6-phosphate (GlcN6P). The structure contains an active site G40A mutation abolishing self-cleavage activity but retaining the overall structure. GlcN6P is explicitly shown in the crystal structure along with the catalytic residue G40A. (e) The secondary structure of the *glmS* ribozyme consists of three coaxially stacked helices, with P1 stacked on P2.2 (shades of blue), P2 stacked on P3 (shades of red) and P2.1 stacked on P3.1 (shades of green) arranged in a double pseudoknot. The catalytic core is located in P2.1 and P2.2, with six nucleotides responsible for binding the cofactor GlcN6P, outlined in black, and the catalytic residue G40, shaded in grey. The cleavage site is indicated with a grey arrowhead. (f) The deprotonated N1 on nucleobase G40 is predicted to act as the general base, deprotonating the 2′-hydroxyl at the cleavage site. The general acid is suggested to be the protonated form of GlcN6P, which donates a proton to the leaving group, completing the self-cleavage reaction.

Artificial Ribozymes

The discovery of novel functional nucleic acids in recent years has been aided by the introduction of *in vitro* selections. This technique makes use of large nucleic acid libraries that contain upwards of 10^{16} different sequences and isolate the few functional molecules by reiterative selection and amplification. The first of these experiments isolated RNAs that were capable of binding specific target molecules with high specificity and affinity and rendered a class of functional RNAs were termed aptamers. Similarly, in vitro selections have been developed for isolation of novel ribozymes. The simplest of these selections involve some covalent modification of the

selected sequences that render them easy to separate from the remainder of the nonactive pool.[63] Examples of *in vitro* selected ribozymes described here perform a number of different prebiotically relevant reactions and advance the hypothesis that life progressed through an RNA-based world.

Diels Alderase

Carbon–carbon bond-forming reactions are the backbone of an anabolic metabolism and would have been an essential part of an evolving RNA world.[64,65] Although there are examples of ribozymes acting on other nucleotide or peptide substrates, there is a relative dearth of truly organic substrates. To expand the scope of reactions that ribozymes are capable of catalysing, *in vitro* selection was performed to select for a ribozyme that catalyses a Diels–Alder cycloaddition. Two selections resulted in Diels Alderase ribozymes, one requiring a modified nucleotide for activity, and the other consisting entirely of unmodified RNA. The modified Diels Alderase requires a pyridyl-modified uridine, and the reaction is catalysed in the presence of a Cu^{2+} ion. In the case of the native RNA ribozyme, the selection was done with the anthracene diene tethered to the ribozyme and the biotinylated maleimide in solution (**Figure 1.5c**), although a ribozyme that resulted from the selection was engineered to act in a truly bimolecular manner with all substrates free in solution. This variant adopts a structure with three helical elements and an asymmetric catalytic bubble shown in red in **Figure 1.5a,b**.[66] Unlike the majority of aptamers, the Diels Alderase has a preformed core that does not significantly change upon binding of the substrate or product. Additionally, binding is primarily dependent on stacking rather than hydrogen bond interactions. The metal ions found associated with the structure act to stabilise the folded ribozyme but do not directly participate in catalysis, because none are found in proximity to the binding pocket. The ribozyme reduces the translational and rotational degrees of freedom of the reagents and catalyses the reaction between its substrates in a stereospecific manner, which can be switched by using L- or D-RNA. This stereospecificity likely arises from binding of the maleimide in a specific orientation, allowing for only one enantiomer to be produced. Although the ribozyme

can perform multiturnover catalysis, it suffers from product inhibition due to the aforementioned preformed binding pocket.[67]

Nucleotide synthetase

The RNA World hypothesis posits that RNAs functioned not only as genetic information carriers but also as macromolecules capable of performing the catalytic events necessary to sustain simple self-replicating systems.[68,69] One key step would have been the coupling of sugars with bases to generate nucleosides, since the uncatalysed reaction has not been observed. The first *in vitro* selection to generate a nucleotide synthetase coupled a 4-thiouracil to a tethered phosphoribosyl 1-pyrophosphate (PRPP). Analysis of the selected sequences indicated a complex core structure reminiscent of the class I ligase, allowing for additional minimisation of the ribozyme. Certain nucleotides within the core are highly conserved and likely play a key role in catalysis. Interestingly, the thio-modification on the uracil was necessary to observe nucleotide synthesis. Kinetic isotope effect analysis revealed that pyrimidine synthesis proceeds through an oxocarbenium intermediate stabilised by the phosphate backbone, via chemistry that differs significantly from other ribozyme-catalysed mechanisms.[70] A subsequent *in vitro* selection for purine nucleotide synthetases revealed ribozymes that are up to 100 times more efficient than comparable pyrimidine nucleotide synthetases. This difference aligns with the disfavoured thermodynamics of pyrimidine glycosidic bond formation compared to purine synthesis.

5′-Triphosphorylating and Capping Ribozymes

Following the synthesis of nucleosides, an RNA world would have necessitated a method of activation to energetically drive polymerisation, specifically, nucleoside triphosphates.[71,72] Cyclic trimetaphosphate (cTMP) has been shown to be a promising activating agent and was likely prebiotically available. An *in vitro* selection aimed to generate a ribozyme capable of utilising cTMP yielded a sequence that could triphosphorylate the 5′-end of RNA both in *cis* and *trans*.

Secondary structure analysis revealed a central, four-helix junction, and kinetic studies indicated a dependence on Mg^{2+} .

A 5′-triphosphorylated RNA can become the substrate for RNA capping. Two capping ribozymes, Iso6 and 6.17, have been evolved, both having promiscuous substrate recognition.[73] These ribozymes have little sequence or structure similarities, yet their molecular mechanisms are quite comparable and include a two-step mechanism with a covalent intermediate. Substrate catalysis is dependent on the phosphate chain length, with triphosphate substrates reacting faster than monophosphates. However, the identity of the base does not significantly affect the capping reaction. The ribozymes are also capable of decapping and cap exchange. This substrate promiscuity is in contrast to the majority of ribozymes that have very specific active site requirements.

Flexizyme

Two of the strongest pieces of scientific evidence that support the notion of an early RNA world are the existence of the ribosome and RNase P as discussed above.[74] For RNA-based primitive translation machinery, it would be necessary to also have a ribozyme capable of charging amino acids on to tRNA, the catalytic equivalent of today's aminoacyl-tRNA synthetases. An *in vitro* selection aimed to select for RNAs capable of charging amino acids on to the 3′-end of tRNA yielded ribozymes with modest 4% activity.[75] Subsequent selections and alterations to the original ribozyme led to the evolution of a class of versatile aminoacylating ribozymes, termed flexizymes, with practically no structural limits on the amino acid or tRNA substrates. X-ray crystallographic studies of this family of ribozymes showed two internal helices, P1 and P2, with an additional three base-pair recognition helix formed between the 3′-end of the flexizyme and the tRNA. The amino acid recognition site is located in J1a/2 and J2/1a (**Figure 1.5d,e**). There are three magnesium ions present, with one in the catalytic core and the other two structural. This tRNA-charging reaction proceeds in two steps: First, the 5′ hydroxyl of the G1 position of the

ribozyme self-aminoacylates using an activated L-phenylalanine derivative. Next, the charged ribozyme recognises the correct tRNA substrate via base-pairing and exclusively aminoacylates its terminal 3′ hydroxyl (**Figure 1.5f**).[76]

Polymerase

If RNAs were to be the sole genetic and catalytic biopolymers on early Earth, as suggested by the RNA World hypothesis, there needed to be a system for self-replication, during which an RNA template could be replicated, possibly by a ribozyme catalysing the templated addition of a nucleotide to the 3′-end of an RNA strand. A single step in such a polymerisation corresponds to ligation of an activated nucleotide to an RNA terminus. The first *in vitro* selection of a ribozyme aimed to select for a ligase, which was later evolved *in vitro* into a bona fide RNA polymerase.[77-79] The ligase core has been extensively studied revealing a mechanism in which the 3′-hydroxyl on the substrate attacks the phosphate on the 5′-ribozyme, ligating the substrate and releasing pyrophosphate (**Figure 1.5g–i**).[80] The selected polymerase ribozyme (R18) is capable of extending a primer sequence by 14 nucleotides and did so with a rather high fidelity rate at about 96.7%. To further push the limits of this polymerase ribozyme, another selection was preformed that added a randomised sequence to the 5′-end of R18 to improve the primer–template interactions. The new selected ribozyme was then engineered to improve activity, and the resulting polymerase was able to extend a template to 95 nucleotides.[81] Further iterations of polymerase ribozymes have been selected to improve the length of polymerase products and fidelity of amplification. One such polymerase ribozyme, 24-3, not only performs RNA polymerisation but also acts as reverse transcriptase.[82,83] Further evolution of the 24-3 polymerase led to the 38-6 variant, which is capable of synthesising its ancestor, the class I ligase.[84] Additionally, the 38-6 ribozyme can use other nucleoside analogues based on threose or arabinose sugars.[85]

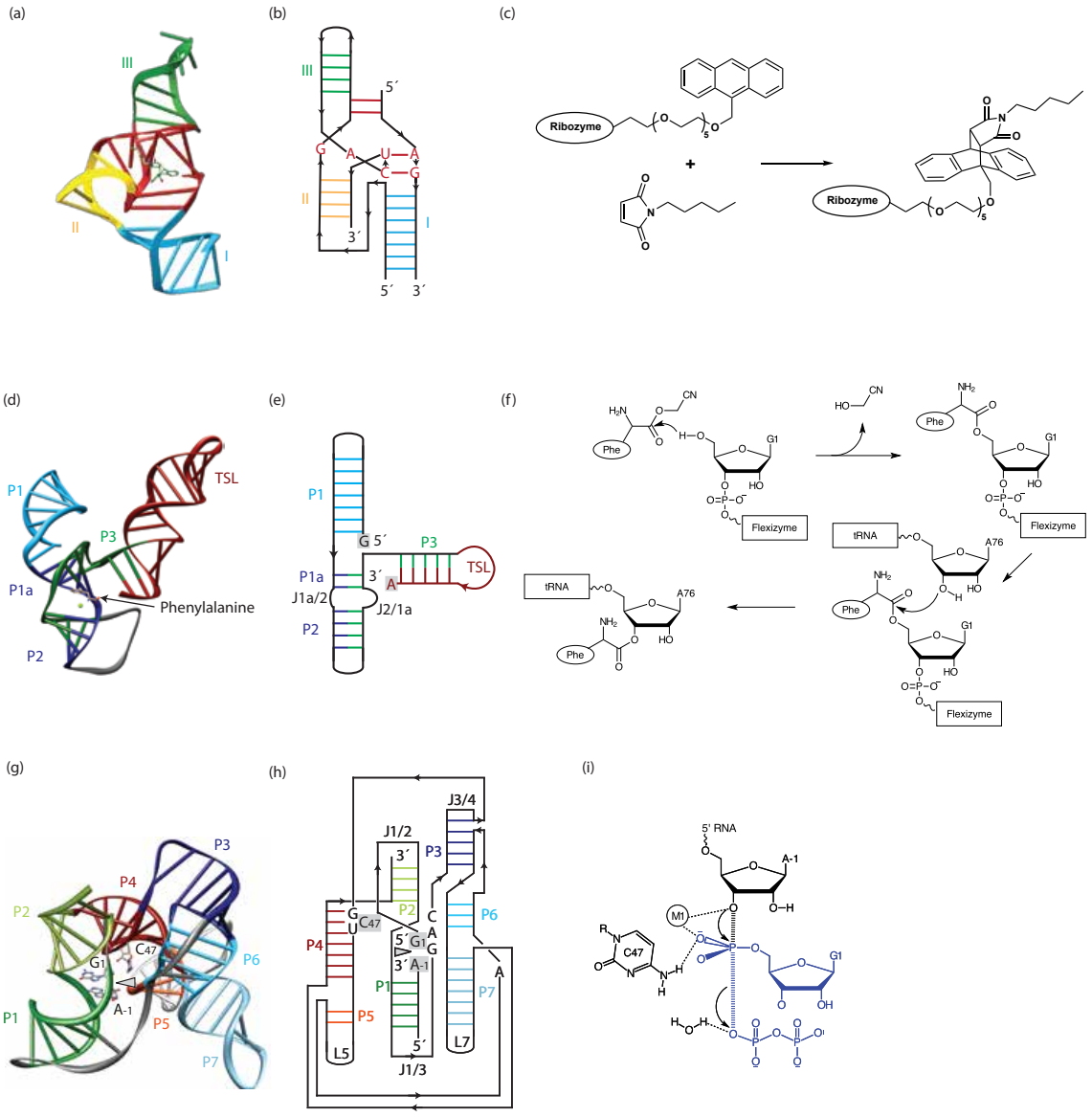


Figure 1.5. Structures and proposed reaction mechanisms of the *in vitro* selected Diels Alderase, flexizyme, and RNA polymerase ribozymes. (a) Crystal structure of an unmodified Diels Alderase, which catalyses the addition of a maleimide to an anthracene diene, the product of which is shown in the active site. (b) The ribozyme consists of three helical domains (I–III) and a catalytic core depicted in red. The selection was done with the anthracene covalently attached to the ribozyme at guanidine between helices II and III. (c) The Diels–Alder reaction catalysed by the ribozyme yields a product with only one enantiomer. (d) Crystal structure of the flexizyme,

which preforms ribozyme-catalysed tRNA aminoacylation, fused to an RNA substrate. Phenylalanine and a magnesium ion are shown in the binding pocket. (e) The core consists of two internal helices, P1 and P2, with an additional three base-pair recognition helix formed between the 3′-end of the flexizyme and the tRNA in P3. (f) The aminoacylation occurs in two steps: First, the 5′ hydroxyl of the G1 position of the ribozyme self-aminoacylates, utilising an activated phenylalanine derivative. Second, the charged ribozyme recognises the correct tRNA substrate via base-pairing and exclusively aminoacylates its terminal 3′ hydroxyl, completing the reaction. (g) Crystal structure of the postligation state of the ligase ribozyme that functions as the core of the polymerase ribozyme. (h) The core of the *in vitro* selected RNA-dependent RNA ligase contains three coaxially stacked domains, with P2 stacked on P1 (shades of green), P3 stacked on P6 and P7 (shades of blue) and P4 stacked on P5 (shades of red). The ligation site is contained within P1, whereas P4 and the catalytic C47, shown shaded in grey, for the active site. The site of ligation is indicated by a grey arrowhead. (i) The ligation reaction is initiated by an attack of the α-phosphate at the 5′-end of the ribozyme by the 3′ hydroxyl on the substrate, ligating the substrate and releasing pyrophosphate. The suggested transition state is shown with the ligation substrate in blue. One magnesium ion is shown in the active site, possibly stabilising the transition state.

1.4 Final Remarks

Ribozymes catalyse several chemical transformations central for cellular function and as such are essential for life. These reactions include ribosomal peptide synthesis, tRNA maturation and splicing. Recent discoveries of the widespread distribution of self-cleaving ribozymes, a co-factor-dependent ribozyme and an allosteric group I intron suggest that many new biological functions of ribozymes remain to be discovered. The multitude of reactions catalysed by *in vitro* selected ribozymes, including a polymerase ribozyme capable of synthesising another catalytic RNA, support the RNA World hypothesis and give hope to ever larger application in

biotechnology, synthetic biology and perhaps therapy as well.

1.5 References

1. Ramakrishnan, V. Ribosome structure and the mechanism of translation. *Cell* **2002**, 108, 557-572.

2. Leung, E. K.; Suslov, N.; Tuttle, N.; Sengupta, R.; Piccirilli, J. A. The mechanism of peptidyl transfer catalysis by the ribosome. *Annu. Rev. Biochem.* **2011**, 80, 527-555.

3. Sievers, A.; Beringer, M.; Rodnina, M. V.; Wolfenden, R. The Ribosome as an entropy trap. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, 101, 7987-7901.

4. Hiller, D. A.; Singh, V.; Zhong, M.; Strobel, S. A. A two-step chemical mechanism for ribosome-catalysed peptide bond formation. *Nature* **2011**, 476, 236-239.

5. Kuhlenkoetter, S.; Wintermeyer, W.; Rodnina, M. V. Different substrate-dependent transition states in the active site of the ribosome. *Nature* **2011** 476, 351-354.

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

7. Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 angstrom resolution. *Science* **2000**, 289, 905-920.

8. Nissen, P.; Hansen, J.; Ban, M.; Moore, P. B.; Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **2000**, 289, 920-930.

9. Lonnberg, T. Understanding catalysis of phosphate-transfer reactions by the large ribozymes. *Chem. Eur. J.* **2011**, 17, 7140-7153.

10. Reiner, R.; Ben-Asouli, Y.; Krilovetzky, I.; Jarrous, N.; et al. A role for the catalytic ribonucleoprotein RNase P in RNA polymerase III transcription. *Genes Dev.* **2006**, 20, 1621-1635.

11. Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. The RNA moiety of RNase P is the catalytic subunit of the enzyme. *Cell* **1983**, 35, 849-857.

12. Schon, A. Ribonuclease P: the diversity of a ubiquitous RNA processing enzyme. *FEMS Microbio. Rev.* **1999**, 23, 391-406.

13. Mcclain, W. H.; Guerrier-Takada, C.; Altman, S. Model substrates for an RNA enzyme. *Science* **1987**, 238, 527-530.

14. Reiter, N. J.; Osterman, A.; Torres-Larios, A.; et al. Structure of a bacterial ribonuclease P holoenzyme with complex with tRNA. *Nature* **2010**, 468, 784-789.

15. Valadkhan, S.; Mohammadi, A.; Jaladat, Y.; Geisler, S. Protein-free small nuclear RNAs catalyze a two-step splicing reaction. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, 106, 11901-11906.

16. Shi, Y. The spliceosome: a protein-directed metalloribozyme. *J. Mol. Biol.* **2017**, 429, 2640-2653.

17. Cech, T. R.; Bass, B. L. Biological catalysis by RNA. *Annu. Rev. Biochem.* **1986**, 55, 599-629.

18. Cech, T. R. Self-splicing of group I introns. *Annu. Rev. Biochem.* **1990**, 59, 543-586.

19. Kruger, K.; Grabowski, P. J.; Zaug, A. J.; et al. Self-splicing RNA auto-excision and auto-cyclization of the ribosomal-RNA intervening sequence of tetrahymena. *Cell* **1982**, 31, 147-157.

20. Cech, T. R. The chemistry of self-splicing RNA and RNA enzymes. *Science* **1987**, 236, 1532-1539.

21. Zaug, A. J.; Grabowski, P. J.; Cech, T. R. Autocatalytic cyclization of an excised intervening sequence RNA is a cleavage-ligation reaction. *Nature* **1983**, 301, 578-583.

22. Adams, P. L.; Stahley, M. R.; Kosek, A. B.; Wang, J.; Strobel, S. A. Crystal structure of a self-splicing group I intron with both exons. *Nature* **2004**, 430, 45-50.

23. Stahley, M. R.; Strobel, S. A. Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science* **2005**, 309, 1587-1590.

24. Lambowitz, A. M.; Zimmerly, S. Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol.* **2011**, 3, a003616.

25. Peebles, C. L.; Perlman, C. S.; Mecklenburg, K. L.; et al. A self-splicing RNA excises an intron lariat. *Cell* **1986**, 44, 213-223.

26. Schmelzer, C.; Schweyen, R. J. Self-splicing of group II introns in vitro: mapping of the branch point and mutational inhibition of lariat formation. *Cell* **1986**, 46, 557-565.

27. Van der Veen, R.; Arnberg, A. C.; Van der Horst, G.; et al. Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro. *Cell* **1986**, 44, 225-234.

28. Toor, N.; Keating, K. S.; Taylor, S. D.; Pyle, A. M. Crystal structure of a self-spliced group II intron. *Science* **2008**, 320, 77-82.

29. Chan, R. T.; Robart, A. R.; Rajashankar, K. R.; Pyle, A. M.; Toor, N. Crystal structure of the group II intron in the pre-catalytic state. *Nat. Struct. Mol. Biol.* **2012**, 19, 555-557.

30. Galej, W. P.; Wilkinson, M. E.; Fica, S. M.; et al. Cryo-EM structure of the spliceosome immediately after branching. *Nature* **2016**, 537, 197-201.

31. Wilkinson, M. E.; Charenton, C.; Nagai, K. RNA splicing by the spliceosome. *Annu. Rev. Biochem.* **2020**, 89, 359-388.

32. Galej, W. P.; Toor, N.; Newman, A. J.; Nagai, K. Molecular mechanism and evolution of nuclear pre-mRNA and group II intron splicing: insights from cryo-electron microscopy structures. *Chem. Rev.* **2018**, 118, 4156-4176.

33. Jimenez, R. M.; Polanco, J. A.; Luptak, A. Chemistry and biology of self-cleaving ribozymes. *Trends Biochem. Sci.* **2015**, 40, 648-661.

34. Ferre D′ Amare, A. R.; Scott, W. G. Small self-cleaving ribozymes. *Cold Spring Harb. Perspect. Biol.* **2010**, 2, a003574.

35. Webb, C. H.; Luptak, A. HDV-like self-cleaving ribozymes. *RNA Biol.* **2011**, 8, 719-727.

36. Salehi-Ashtiani, K.; Luptak, A.; Litovchick, A.; Szostak, J. W. A genome wide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science* **2006**, 313, 1788-1792.

37. Ferre D′ Amare, A. R.; Zhou, K.; Doudna, J. A. Crystal structure of hepatitis delta virus ribozyme. *Nature* **1998**, 395, 567-574.

38. Webb, C. H.; Riccitelli, N. J.; Ruminski, D. J.; Luptak, A. Widespread occurance of self-cleaving ribozymes. *Science* **2009**, 326, 953.

39. Golden, B. L. Two distinct catalytic strategies in the hepatitis delta virus ribozyme cleavage reaction. *Biochemistry* **2011**, 50, 9424-9433.

40. Veeraraghavan, N.; Ganguly, A.; Chen, J. H. Metal binding motif in the active site of the HDV ribozyme binds divalent and monovalent ions. *Biochemistry* **2011**, 50, 2672-2682.

41. Hamman, C.; Luptak, A.; Perreault, J.; de la Pena, M.; et al. The ubiquitous hammerhead ribozyme. *RNA* **2012**, 18, 871-885.

42. De la Pena, M.; Gago, S.; Flores, R. Peripheral regions of natural hammerhead ribozymes greatly increase their self-cleavage activity. *EMBO J.* **2003**, 22, 5561-5570.

43. Khvorova, A.; Lescoute, A.; Westhof, E.; Jayasena, S. D. Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nat. Struct. Mol. Biol.* **2003**, 10, 708-712.

44. Martik, M.; Scott, W. G. Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell* **2006**, 126, 309-320.

45. Leclerc, F. Hammerhead ribozymes: true metal or nucleobase catalysis? Where is the catalytic power from? *Molecules* **2010**, 15, 5389-5407.

46. Roth, A.; Weinberg, Z.; Chen, A. G. A widespread self-cleaving ribozyme class is revealed by bioinformatics. *Nat. Chem. Biol.* **2014**, 10, 56-60.

47. Weinberg, Z.; Kim, P. B.; Chen, T. H.; et al. New classes of self-cleaving ribozymes revealed by comparative genomics analysis. *Nat. Chem. Biol.* **2015**, 11, 606-610.

48. Gebetsberger, J.; Micura, R. Unwinding the twister ribozyme: from structure to ribozyme. *Wiley Interdiscip. Rev. RNA* **2017**, 8, e1402.

49. Zheng, L.; Falschlunger, C.; Huang, K.; et al. Structure-based insights into self-cleavage by a four-way junctional twister-sister ribozyme. *Nat. Commun.* **2017**, 8, 1180.

50. Nguyen, L. A.; Wang, J.; Steitz, T. A. Crystal structure of Pistol, a class of self-cleaving ribozyme. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, 114, 1021-1026.

51. Zheng, L.; Mairhofer, E.; Teplova, M.; et al. Hatchet ribozyme structure and implications for cleavage mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, 116, 10783-10791.

52. Fedor, M. J. Structure and function of the hairpin ribozyme. *J. Mol. Biol.* **2000**, 297, 269-291.

53. Wilson, T. J.; Nahas, M.; Ha, T.; Lilley, D. M. Folding and catalysis of the hairpin ribozyme. *Biochem. Soc. Trans.* **2005**, 33, 461-465.

54. Rupert, P. B.; Ferre-D′Amare, A. R. Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature* **2001**, 410, 780-786.

55. Rupert, P. B.; Massey, A. P.; Sigurdsson, S. T.; Ferre-D′Amare, A. R. Transition state stabilization by a catalytic RNA. *Science* **2002**, 298, 1421-1424.

56. Lilley, D. M. The Varkud satellite ribozyme. *RNA* **2004**, 10, 151-158.

57. Guo, H. C.; Collins, R. A. Efficient trans-cleavage of a stem-loop RNA substrate by a ribozyme derived from neurospora VS RNA. *Embo Rep.* **1995**, 14, 368-376.

58. Lipfert, J.; Ouellet, J.; Norman, D. G.; Doniach, S.; Lilley, D. M. The complete VS ribozyme in solution studied by small-angle X-ray scattering. *Structure* **2008**, 16, 1357-1367.

59. McCown, P. J.; Winkler, W. C.; Breaker, R. R. Mechanism and distribution of glmS ribozymes. *Methods Mol. Biol.* **2012**, 848, 113-129.

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

61. Klein, D. J.; Ferre-D′Amare, A. R. Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science* **2006**, 313, 1752-1756.

62. Cochrane, J. C.; Lipchock, S. V.; Strobel, S. A. Structural investigation of the glmS ribozyme bound to its catalytic cofactor. *Chem. Biol.* **2007**, 14, 97-105.

63. Wilson, D. S.; Szostak, J. W. In vitro selection of functional nucleic acids. *Annu. Rev. Biochem.* **1999**, 68, 611-647.

64. Tarasow, T. M.; Tarasow, S. L.; Eaton, B. E. RNA-catalysed carbon-carbon bond formation. *Nature* **1997**, 389, 54-57.

65. Seelig, B.; Jaschke, A. A small catalytic RNA motif with Diels-Alderase activity. *Chem. Biol.* **1999**, 6, 167-176.

66. Serganov, A.; Keiper, S.; Malinina, L.; et al. Structural basis for Diels-Alder ribozyme-catalyzed carbon-carbon bond formation. *Nat. Struct. Mol. Biol.* **2005**, 12, 218-224.

67. Stuhlmann, F.; Jaschke, A. Characterization of an RNA active site: interactions between a Diels-Alderase ribozyme and its substrates and products. *J. Am. Chem. Soc.* **2002**, 124, 3238-3244.

68. Unrau, P. J.; Bartel, P. J. RNA-catalysed nucleotide synthesis. *Nature* **1998**, 395, 260-263.

69. Martin, L. L.; Unrau, P. J.; Muller, U. F. RNA synthesis by in vitro selected ribozymes for recreating an RNA world. *Life* **2015**, 5, 247-268.

70. Chapple, K. E.; Bartel, D. P.; Unrau, P. J. Combinatorial minimization and secondary structure determination of a nucleotide synthase ribozyme. *RNA* **2003**, 9, 1208-1220.

71. Huang, F.; Yarus, M. 5′-RNA self-capping from guanosine diphosphate. *Biochemistry* **1997**, 36, 6557-6563.

72. Moretti, J. E.; Muller, U. F. A ribozyme that triphosphorylates RNA 5′-hydroxyl groups. *Nucleic Acids Res.* **2014**, 42, 4767-4778.

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

74. Morimoto, J.; Hayashi, Y.; Iwasaki, K.; Suga, H. Flexizymes: their evolutionary history and the origin of catalytic function. *Acc. Chem. Res.* **2011**, 44, 1359-1368.

75. Lee, N.; Bessho, Y.; Wei, K.; Szostak, J. W.; Suga, H. Ribozyme-catalyzed tRNA aminoacylation. *Nat. Struct. Mol. Biol.* **2000**, 7, 28-33.

76. Xiao, H.; Murakami, H.; Suga, H.; Ferre-D′Amare, A. R. Structural basis of specific tRNA aminoacylation by a small in vitro selected ribozyme. *Nature* **2008**, 454, 358-361.

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

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

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

80. Shechner, D. M.; Grant, R. A.; Bagby, S. C.; et al. Crystal structure of the catalytic core of an RNA-polymerase ribozyme. *Science* **2009**, 326, 1271-1275.

81. Wochner, A.; Attwater, J.; Coulson, A.; Holliger, P. Ribozyme-catalyzed transcription of an active ribozyme. *Science* **2011**, 332, 209-212.

82. Horning, D. P.; Joyce, G. F. Amplification of RNA by an RNA polymerase ribozyme. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, 113, 9786-9791.

83. Samanta, B.; Joyce, G. F. A reverse transcriptase ribozyme. *Elife* **2017**, 6, e31153.

84. Tjhung, K. F.; Shokhirev, M. N.; Horning, D. P.; Joyce, G. F. An RNA polymerase ribozyme that synthesizes its own ancestor. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, 117, 2906-2913.

85. Horning, D. P.; Bala, S.; Chaput, J. C.; Joyce, G. F. RNA-catalyzed polymerization of deoxyribose, threose, and arabinose nucleic acids. *ACS Synth. Biol.* **2019**, 8, 955-961.

CHAPTER 2: ENZYMATIC RNA PRODUCTION FROM NTPS SYNTHESIZED FROM NUCLEOSIDES AND TRIMETAPHOSPHATE

Publication Note:

This paper was originally published in the journal *ChemBioChem.*

Chizzolini, F.;* **Kent, A. D.;*** Passalacqua, L. F. M.;* Lupták, A. *ChemBioChem*. **2021**, 22, 2098-2101.

* Authors contributed equally to this work.

© 2021, Wiley-VCH GmbH

Contribution Statement

F.C. performed the kinetics, thermodynamics, and purification experiments. **A.D.K.** performed the ribozyme experiments. L.F.M.P. performed the enzyme experiments. **A.D.K.** and F.C. analyzed the data**. A.D.K.** generated the figures. **A.D.K.**, L.F.M.P. and A.L. wrote the manuscript.

2.1 Abstract

A mechanism of nucleoside triphosphorylation would have been critical in an evolving "RNA world" to provide high-energy substrates for reactions such as RNA polymerization. However, synthetic approaches to produce ribonucleoside triphosphates (rNTPs) have suffered from conditions such as high temperatures or high pH that lead to increased RNA degradation, as well as substrate production that cannot sustain replication. Previous reports have demonstrated that cyclic trimetaphosphate (cTmp) can react with nucleosides to form rNTPs under prebiotically-relevant conditions, but their reaction rates were unknown and the influence of reaction conditions not well-characterized. Here we established a sensitive assay that allowed for the determination of second-order rate constants for all four rNTPs, ranging from 1.7×10^{-6} to 6.5×10^{-6} M^{-1} s^{-1}. The ATP reaction shows a linear dependence on pH and Mg^{2+}, and an enthalpy of activation of 88 ± 4kJ/mol. At millimolar nucleoside and cTmp concentrations, the rNTP production rate is sufficient to facilitate RNA synthesis by both T7 RNA polymerase and a polymerase ribozyme. We

suggest that the optimized reaction of cTmp with nucleosides may provide a viable connection between prebiotic nucleotide synthesis and RNA replication.

2.2 Introduction

Efforts to elucidate the chemical processes by which life arose on Earth suggest that ribonucleic acid (RNA) was the first biopolymer capable of sustaining Darwinian evolution.[1-7] In order to establish a plausible path to an RNA-only system, it is critical to connect prebiotic formation of nucleosides and ribozyme evolution through development of reactions that yield activated substrates, particularly nucleoside triphosphates (rNTPs).[8-15] Cyclic trimetaphosphate (cTmp) has been shown to be particularly promising as both a prebiotically relevant and a highly reactive substrate among polyphosphorylating reagents (**Figure 2.1**).[16-19]

Several reports have demonstrated that cTmp is capable of phosphorylating a number of different substrates, including nucleosides.[20-27] However, these reactions typically require conditions that are not generally favorable for RNA stability, such as high pH, additional cofactors, high temperatures, and urea.[28-30] We address these limitations by optimization of a cTmp reaction to produce the canonical 5′-triphosphorylated ribonucleosides in sufficient concentrations to synthesize RNA by both protein and ribozyme RNA polymerases.

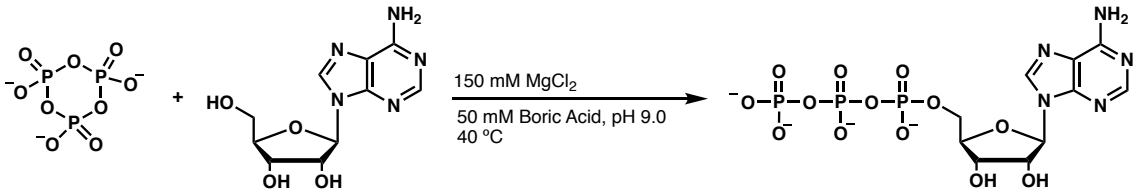


Figure 2.1. Triphosphorylation of adenosine with cTmp.

2.3 Results

In order to measure the low concentrations of nucleoside triphosphates, we developed an ultrasensitive ATP detection assay using highly purified firefly luciferase (Fluc)[31] and a cooled EMCD detector equipped with a high-numerical aperture (NA ~ 0.5) lens system (**Figure 2.2A**). This allows the detection of ATP down to pM concentrations and monitoring of multiple reactions in parallel (**Figure 2.2B**) to provide a systematic characterization of the reaction (**Figure 2.2C**).

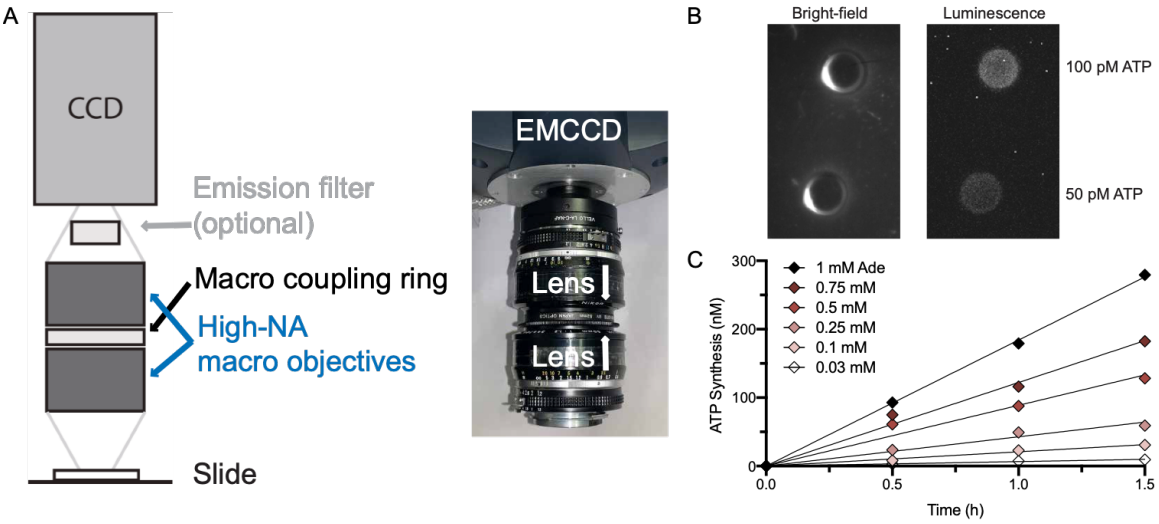


Figure 2.2. Luminescence based detection of ATP production.

(a) A macroscope setup consisting of two high-numerical-aperture lenses mounted in opposition on a cooled electron-multiplying CCD camera. The macroscope field of view is the same as the size of its CCD chip (13.3 × 13.3 mm) and allows observation of several 1 µL reactions in parallel. (b) An example of two reactions containing 100 pM (top) and 50 pM (bottom) ATP detected through a chemiluminescent reaction catalyzed by firefly luciferase. A wide range of ATP concentrations was used to calculate a calibration curve for ATP concentration measurement in cTmp reactions. (c) Time course of reactions containing 10 mM cTmp, 50 mM boric acid, 150 mM $MgCl_2$,

and varying concentrations of adenosine at pH 9.0 at 40°C. Synthesized ATP was measured using the luciferase assay at each timepoint.

We first established a calibration curve for ATP concentration measurement by relating the photon output of the chemiluminescent reaction catalyzed by the firefly luciferase to ATP over a wide range of concentrations. We then determined the ATP concentration at each time point of the triphosphorylation reactions. We confirmed previous experiments showing that the reaction between adenosine and cTmp is likely base-catalyzed (**Figure 2.3A**).[32] The reaction was independent of the buffer identity (**Figure 2.3A–C** and **Figure 2.S1–2.S2**), and the rate constants exhibited different slopes below and above neutral pH, suggesting that the rate-limiting step is different in these two regimes. We also observed that the reaction dependence on Mg^{2+} ions increases linearly up to ~ 200 mM with a slope of ~ 1.1, consistent with a Mg^{2+} -catalyzed rate-limiting step (**Figure 2.3B**). Given the relatively high concentration at which the Mg^{2+} curve reaches saturation, this rate-limiting step likely does not involve a Mg^{2+} ion coordinated by multiple phosphates of the cTMP; rather, a weakly coordinating position, such as the 5′ oxyanion, that likely acts as the nucleophile in the triphosphorylation reaction.[33-34] We also measured the enthalpy and entropy of activation of ATP production from nucleosides and cTmp. An Eyring plot (**Figure 2.3C**) revealed the enthalpy of activation of 88 $\pm$ 4 kJ/mol and entropy of –66 $\pm$ 10 J/mol · K. This relatively high energetic barrier highlights the importance of a catalyzed ATP synthesis for early self-replicating systems to emerge.

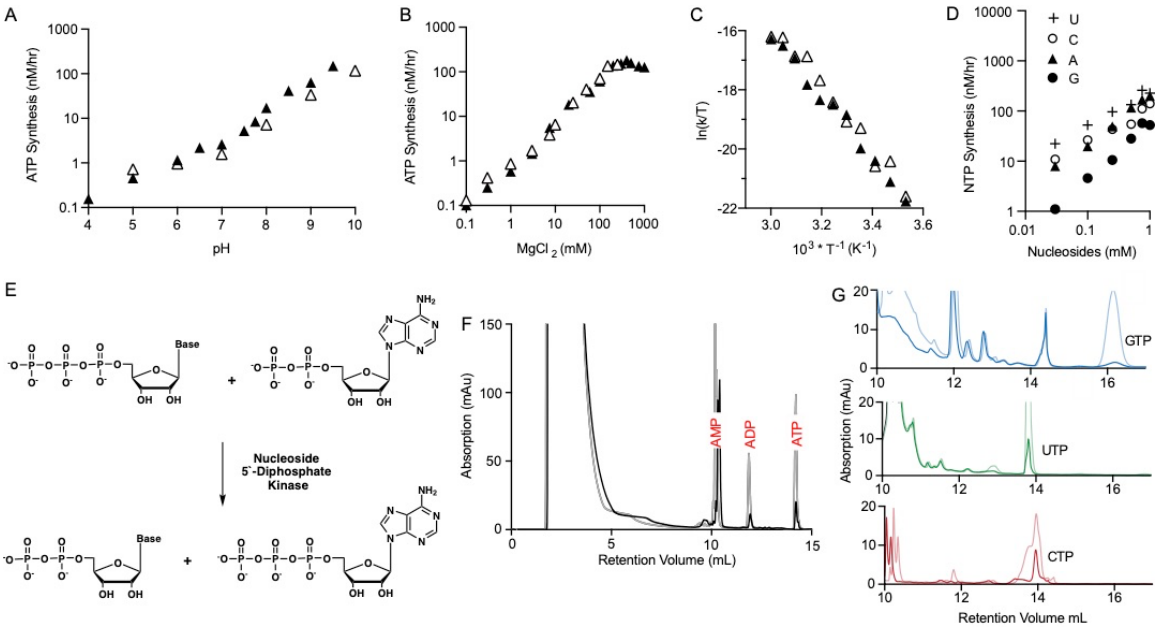


Figure 2.3. Reaction analysis of nucleoside triphosphorylation by trimetaphosphate

ATP synthesis rate measured as a function of pH (A) and Mg^{2+} ion concentration (B) in a three-buffer system (white) or a boric acid system (black), measured using a firefly luciferase assay. (C) An Eyring plot for the reaction, revealing the activation enthalpy as 88±4 kJ/mol (20.9±1.1 kcal/mol) for boric acid (black), and 87±3 kJ/mol (20.8±0.7 kcal/mol) for the three-buffer system (white). The entropy was calculated as – 66±10 J/mol · K and – 70±9 J/mol · K (white), respectively. (D) Synthesis of all four canonical ribonucleotide triphosphates measured using a firefly luciferase assay coupled to 5′-diphosphate kinase, which uses rNTPs to produce ATP. (E) Nucleoside 5′ diphosphate kinase transfers a phosphate from an rNTP to ADP to generate ATP. (F) Anion exchange chromatography showing production of ATP. The trace of the reaction products (black line) was overlaid with a chromatographic trace of the same reaction supplemented with 2 μM of AMP, ADP, and ATP (gray line) to confirm the phosphorylation state of the products. (G) Anion exchange chromatography showing production of other rNTPs (dark line) overlaid with 2 μM of each rNTP standard (light line). No additional peaks were seen at later retention volumes in any of the chromatograms.

To detect the formation of other nucleoside triphosphates (NTPs) with firefly luciferase, we developed a coupled enzymatic assay using nucleoside 5′-diphosphate kinase, which transfers the gamma phosphate of nucleoside triphosphates to ADP, yielding a molecule of ATP for each NTP (**Figure 2.3E**).[35] Triphosphorylation was confirmed for all four RNA nucleotides, with rate constants estimated to be $4.8 \times 10^{-6}\ M^{-1}\ s^{-1}$ for ATP, $1.7\times10^{-6}\ M^{-1}\ s^{-1}$ for GTP, $3.5\times10^{-6}\ M^{-1}\ s^{-1}$ for CTP, and $6.5\times10^{-6}\ M^{-1}\ s^{-}1$ for UTP (**Figures 2.2C, 2.3D, and 2.S3**). We further confirmed the synthesis of the nucleotides using anion- exchange liquid chromatography and assigned the phosphorylation states of the peaks in the chromatogram by co-injecting known amounts of AMP, ADP, and ATP (**Figure 2.3F**) or one of the other three rNTPs, as appropriate (**Figure 2.3G**). The guanosine reaction suffered both from lower levels of GTP production and from the generation of additional side products. This result corroborates the luciferase assay, which indicated that GTP synthesis is more challenging than synthesis of the other canonical rNTPs (**Figure 2.3D**).

To demonstrate the utility of this reaction, we used the synthesized nucleotides to sustain RNA transcription using T7 RNA polymerase. We chose to transcribe an HDV-like ribozyme, drz-Fpra-2, because of its small size and because its self-scission requires high-fidelity transcription.[36] Furthermore, we have recently used this system to show that omitting a single nucleotide (CTP) can stably pause the polymerase in the elongation mode without detectable production of a full-length transcript, making it a sensitive readout for incorporation of freshly synthesized rNTPs.[37] A band corresponding to the full-size transcript of drz-Fpra-2 was readily observed when using only the cTmp-derived nucleotides (**Figure 2.4A**). However, no cleavage product was detected, likely due to misincorporation of other nucleotides. To test this model, we supplemented the transcription with GTP and observed a greatly increased yield of the transcripts and the fraction of active ribozymes (**Figure 2.4B**). Similarly, we used the synthesized nucleotides with tC19Z, a polymerase ribozyme capable of extending an RNA primer along an RNA template.[11] PAGE analysis

showed the addition of three of the four nucleotides (**Figure 2.4C and Figure 2.S4**), with the polymerase activity apparently again limited by the low concentration of GTP.

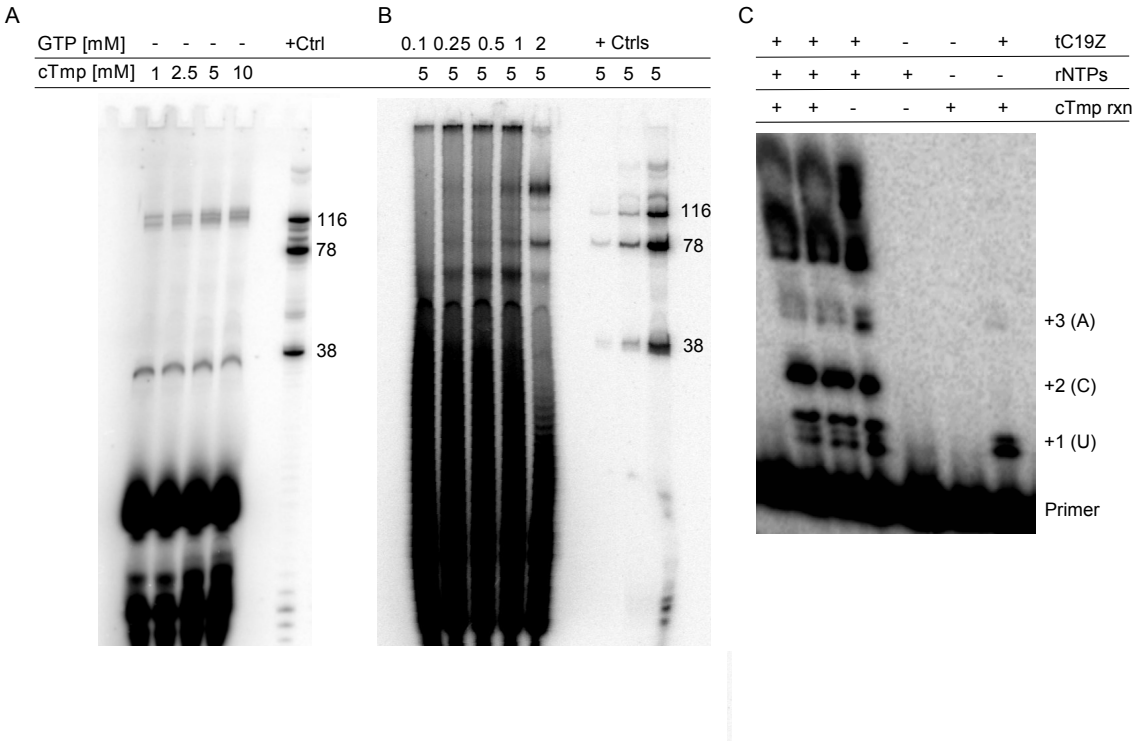


Figure 2.4. In vitro RNA production coupled to cTmp-derived rNTPs.

(A) In vitro transcriptions of a self-cleaving ribozyme. cTmp reactions were diluted into T7 RNA polymerase reactions to produce the drz-Fprau-2 self-cleaving ribozyme. A control reaction (+Ctrl) with additional rNTPs shows production of full-length ribozyme, as well as the products of self-scission. (B) Transcription reactions with cTmp-derived NTPs supplemented with GTP at concentrations indicated above the gel, showing robust transcription initiation and production of active ribozymes. Control reactions (+Ctrls) include a five-fold serial dilution series of transcriptions containing all NTPs. (C) A laboratory-evolved polymerase ribozyme, tC19Z, utilizing cTmp-derived rNTPs generated at 70 ºC to extend a primer along a template. Lanes 1–3 (left to right) show

positive controls and lanes 4–5 negative control reactions. Lane 6 resolves the reaction containing cTmp-generated nucleotides and shows extension of three nucleotides along the template.

2.4 Conclusion

In summary, we synthesized all four canonical nucleoside triphosphates from cTmp and nucleosides, and utilized them in RNA synthesis. We found that in each NTP synthesis reaction, the formation of GTP was significantly lower than formation of the other three nucleosides. We hypothesize that reduced GTP synthesis may result from relatively low solubility and the propensity for guanosine to assume the *syn* conformation in solution, perhaps hindering the 5′ oxygen from reacting with cTmp. This was further corroborated by the presence of additional peaks in the chromatograms of the guanosine-cTmp reaction, which likely represent other phosphoguanosines (**Figure 2.3F**). The low concentrations of GTP present in the reactions presented a unique challenge when transcribing and copying RNA. Although RNA was successfully produced and detected in each case, yields were low or GTP was not incorporated in appreciable amounts, as was the case in ribozyme-mediated primer-extension. Thus, robust RNA production may require an alternate route to GTP synthesis or reliance on ribozymes that are composed of only two or three types of nucleotides.[38,39] Nevertheless, rNTP synthesis in prebiotically-relevant conditions provides a crucial link between nucleoside formation and RNA production. Further development of this reaction with alternative conditions to increase GTP synthesis or by the introduction of ribozymes that enhance the rate of production would have been crucial for an evolving system to emerge.[40,41] Another challenge is represented by the high Mg^{2+} ion concentration required for both NTP and ribozyme-catalyzed RNA synthesis, because these conditions also promote RNA degradation (**Figure 2.S4**) and would likely require a shift of the reactions towards less RNA-damaging conditions, such as the presence of weak chelators of Mg^{2+} ions, for stable functional RNAs to emerge and for their encapsulation in protocells.[42,43]

2.5 Acknowledgments

This work was supported by grants from the John Templeton Foundation through the Foundation for Applied Molecular Evolution and NSF-CBET 1804220 and NASA 20-EXO20-0110 (A.L.), NSF Graduate Research Fellowship (A.D.K.), and Science Without Borders Program, CAPES Foundation, Ministry of Education of Brazil, Process 99999.013571/2013-03, and Miguel Velez Scholar- ship-UCI (L.P.).

2.6 Materials and Methods

All chemicals were purchased from Sigma-Aldrich, unless otherwise stated.

ATP detection by chemo-luminescent assay.

ATP was detected using the highly thermostable StayBrite luciferase enzyme (BioVision). In order to remove trace amounts of ATP present in the enzyme stock, StayBrite luciferase was pre-incubated for 6 hours at room temperature in a solution containing 3 μM luciferase enzyme, 50 μM D-luciferin, 200 μM ammonium sulfate, 1 mM dithiothreitol (DTT), 20 mM $MgCl_2$, and 25 mM Tris-HCl at pH 7.4. After this pre-incubation, ATP was detectable down to 10 pM concentration. ATP detection was performed by mixing 2 parts of pre-incubated enzyme solution with 1 part of the ATP-containing solution on a glass coverslip (Fisher Scientific) attached to a Frame-Seal chamber (Bio-Rad). The chamber was then closed by a second glass coverslip. The closed Frame-Seal chamber was then attached to a mirror and mounted in front of a macro lens system constructed by mounting two Nikkor f/1.2 lenses in opposition. The macro lens system images an area of the same size as the CCD chip on the camera. In our case, this system images an area approximately the size of the Frame-Seal chamber, ensuring that most of the luminescence output is directed to the chip of an EMCCD camera (iXon Ultra 888, Andor). The temperature of chip of the EMCCD camera was maintained at –100 °C during detection using a thermoelectric recirculating chiller (Thermocube 300, Solid State Cooling Systems). The sample slides were held on

the mirror using a paper clip and mounted on a micrometer connected to a sliding platform for the precise focusing. To prevent contamination from environmental light sources, the sample was isolated with two nested dark boxes (Thorlabs) and kept in a dark room. Standard curves were generated using increasing concentrations of ATP in order to convert the luminescence output of samples containing unknown amounts of ATP to molar concentration.

CTP, GTP, and UTP detection by enzymatic phosphorylation of ADP to ATP.

In order to detect CTP, GTP, and UTP using the ATP-dependent chemo-luminescent assay described above, the enzyme nucleoside 5'-diphosphate kinase (NDK) was used. The kinase used either CTP, GTP, or UTP as the substrate to phosphorylate ADP to ATP. Samples containing the triphosphorylated nucleotides to be used as substrates for the ATP production were diluted 10-fold in 10 mM 2-amino-2-(hydroxymethyl)propane-1,3-diol (Tris), 0.1 mM EDTA at pH 7.4 (buffer TE), along with 0.05 units of the kinase enzyme and 50 nM ADP. The ADP used was purified using ion-exchange chromatography to remove trace amounts of ATP. The phosphorylation reaction was incubated for 5 minutes at room temperature, after which the samples were analyzed using the luciferase-based assay described above. Standard curves were generated by using the same conditions for ATP phosphorylation as for the unknown samples and by including increasing concentrations of commercial CTP, GTP, and UTP, as appropriate. The standard curves were then used to convert the luminescence output of the luciferase assay to the molar concentration of each of the individual triphosphorylated nucleotides.

Adenosine phosphorylation reaction.

Adenosine and cyclic trimetaphosphate (cTmp) were allowed to react under different conditions, and the production of ATP was assessed using the luciferase assay described above. To derive the ATP synthesis rate, ATP production was measured multiple times during the reaction. Standard conditions were 1 mM adenosine, 10 mM cTmp, 150 mM $MgCl_2$, and 50 mM boric acid at pH 9.0. The pH was adjusted using sodium hydroxide after all the components of the reactions were mixed together. The standard temperature for the reaction was 40 °C, unless otherwise

stated. The pH dependence of ATP production was assessed using a three-component buffer system, in which the pH could be directly titrated from 4.0 to 10.0. This was done in order to maintain constant ionic strength of the solution. The system was composed of 100 mM acetic acid, 100 mM 2-(*N*-morpholino)ethanesulfonic acid (MES), and 100 mM boric acid. To assess the possible role of boric acid on the catalysis of ATP, boric acid was replaced with *N*-cyclohexyl-2-aminoethanesulfonic acid (CHES). Prior to the luciferase assay measurement of ATP concentration, the synthesis reactions were diluted 10-fold in TE buffer.

Guanosine, cytidine, and uridine phosphorylation reactions.

Guanosine, cytidine, or uridine at a concentration of 1 mM were allowed to react with 10 mM cyclic trimetaphosphate (cTmp) in a buffer containing 150 mM $MgCl_2$ and 50 mM boric acid with pH adjusted to 9.0 using sodium hydroxide. Reactions were carried out at 40 °C, and aliquots were taken out at different time points in order to estimate the synthesis rates of the different triphosphorylated nucleotides. At each time point, the concentration of GTP, CTP, or UTP, respectively, was measured through the coupled enzymatic phosphorylation of ADP and chemiluminescence measurement with the luciferase assay, as described above.

FPLC analysis of phosphorylation reaction products.

An FPLC Äkta Purifier UPC-10 system (GE Healthcare) was used to analyze the products of the reactions between nucleosides and cTmp. All reactions were performed with 50 mM cTmp and 50 mM $MgCl_2$ in 50 mM boric acid (pH 9.0). Starting concentrations of the nucleosides for the different reactions were 25 mM adenosine, 25 mM guanosine, 50 mM cytidine, and 50 mM uridine. Reactions were incubated at 40 °C for 4 hours, and afterwards diluted 20-fold in TE buffer. Approximately 1 mL of the diluted reactions was then loaded onto the FPLC system, and the reaction products were separated by charge using a DNAPac PA-100 column (Dionex) with a linear gradient of 0.9 M ammonium carbonate. Control reactions were performed using the same compositions supplemented with 2 μM of either ATP, GTP, CTP, or UTP, as appropriate. In the case of

the ATP synthesis reaction, 2 μM of ADP and 2 μM of AMP were co-injected with the ATP standard.

RNA synthesis from triphosphorylated nucleosides with T7 RNA polymerase.

A triphosphorylation reaction to synthetize the four rNTPS starting from the four nucleosides and cTmp contained 25 mM adenosine, 25 mM guanosine, 50 mM uridine, 50 mM cytidine, 10 mM cTmp, 50 mM $MgCl_2$, and 25 mM Tris-HCl at pH 9.0. Tris-HCl was used as the buffer instead of boric acid, as tests showed that the presence of boric acid inhibited the activity of T7 RNA polymerase. The reaction was then incubated at 40 °C for 4 hours. An HDV-like ribozyme, drz-Fpra-2 (**Table 2.S1**), was used as the template for the in vitro transcription. RNA was transcribed at 37 °C for one hour in a 20 μL reaction containing 40 mM Tris-HCl (pH 7.9), 10 mM $MgCl_2$, 2 mM spermidine, 1 mM dithiothreitol, 4.5 μCi [α-^{32}P]ATP (Perkin Elmer), 10 units of T7 RNA polymerase (NEB), 2 pmoles of DNA template, and 4 μL of the rNTP synthesis reaction. For the positive-control reaction, 100 μM ATP and 1 mM of each GTP, UTP, and CTP were used in place of the rNTPs synthesis reaction. The transcripts were resolved on a 10% polyacrylamide gel electrophoresis (PAGE) under denaturing conditions (7 M urea). The PAGE gel was dried, exposed to a phosphor image screen, and analyzed using a Typhoon phosphorimager (GE Healthcare).

RNA Primer phosphorylation.

A primer labeling reaction was prepared in a total volume of 20 μL containing 20 μM of RNA primer, 1x T4 Polynucleotide Kinase (PNK) ligase buffer (NEB), 1 unit of T4 PNK enzyme (NEB), and 0.5 μCi [γ-^{32}P]ATP (Perkin Elmer). The reaction was incubated at 37 °C for one hour. The radiolabeled RNA primer was purified by 7.5% polyacrylamide gel electrophoresis (PAGE) under denaturing conditions (7 M urea), and the resolved gel was visualized using a Typhoon phosphorimager (GE Healthcare). The band of interest was excised, eluted from the gel into 300 μL of 300 mM KCl, and precipitated by adding 1 μL glycoblue and 700 μL of 98% ethanol at –80 °C.

RNA synthesis with cTmp derived nucleotides using tC19Z RNA polymerase ribozyme.

A reaction containing 25 mM adenosine, 25 mM guanosine, 50 mM uridine, 50 mM cytidine, 10 mM cTmp, 25 mM Tris HCl (pH 9.0), and 50 mM $MgCl_2$ was incubated at 70 °C for 1 hour. The reaction was centrifuged to remove the unreacted nucleosides from the nucleoside triphosphates remaining in solution. Polymerase reactions were prepared in 10 µL volumes where 5 µL were added directly from the cTmp reaction. Additionally, the reactions contained 0.5 µM tc19z ribozyme (**Table 2.S1**), 0.5 µM template, and 0.5 µM [α-^{32}P] 5′-end radiolabeled RNA primer (**Table 2.S1**) in 200 mM $MgCl_2$ and 50 mM Tris HCl (pH 8.0). The RNA was added first and heated at 85 °C for two minutes, then cooled to 17 °C for 10 minutes. The remaining reagents were added and incubated at 17 °C for 5 days. The first positive control did not contain any cTmp reaction, but had 4 mM of each nucleotide. The second positive control resembled the first, but with addition of the cTmp reaction. Negative controls contained cTmp reaction, but were missing either the polymerase ribozyme or the template. After 5 days, reactions were quenched with 5 µL of 500 mM EDTA and 5 µL of 7 M urea loading dye, and then were resolved on a 20% PAGE under denaturing conditions (7 M urea). The PAGE gel was dried and exposed to a phosphor image screen and analyzed using a Typhoon phosphorimager (GE Healthcare).

2.7 Supplementary Data

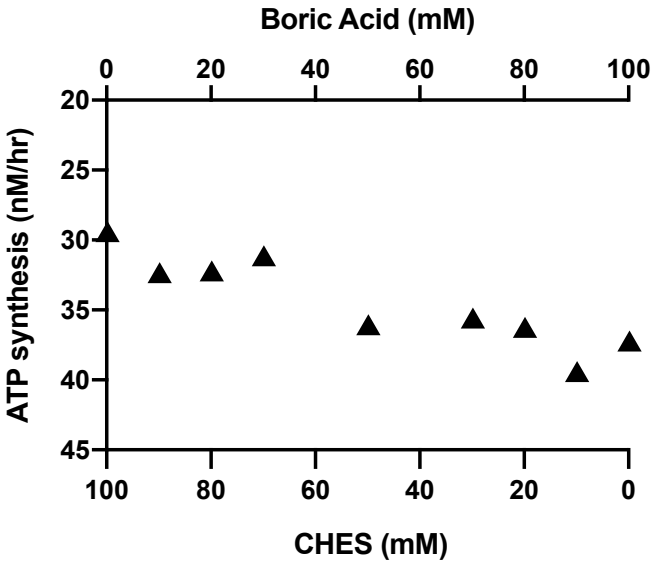


Figure 2.S1. ATP synthesis does not depend on boric acid.

ATP synthesis rates measured in the presence of varying borate/CHES ratios. Reactions containing 1 mM adenosine, 10 mM cTmp, and 50 mM $MgCl_2$ at pH 9.2 were incubated at 40 °C in a buffer composed of 100 mM final concentration of CHES and borate, mixed at different ratios.

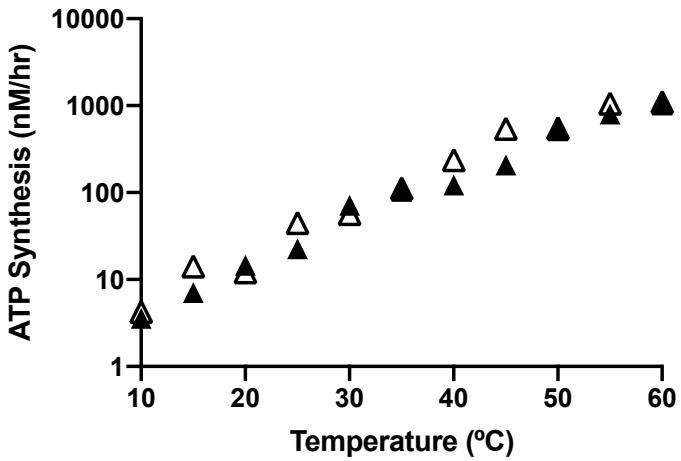


Figure 2.S2. Temperature dependence of ATP synthesis is independent of buffer identity.

Either a three-buffer system (white), or a boric acid buffer (black), was used with a reaction consisting of 1 mM adenosine, 10 mM cTmp, 150 mM $MgCl_2$ at pH 9.0. ATP production at various temperatures was measured using the optimized luciferase assay to measure synthesis rate.

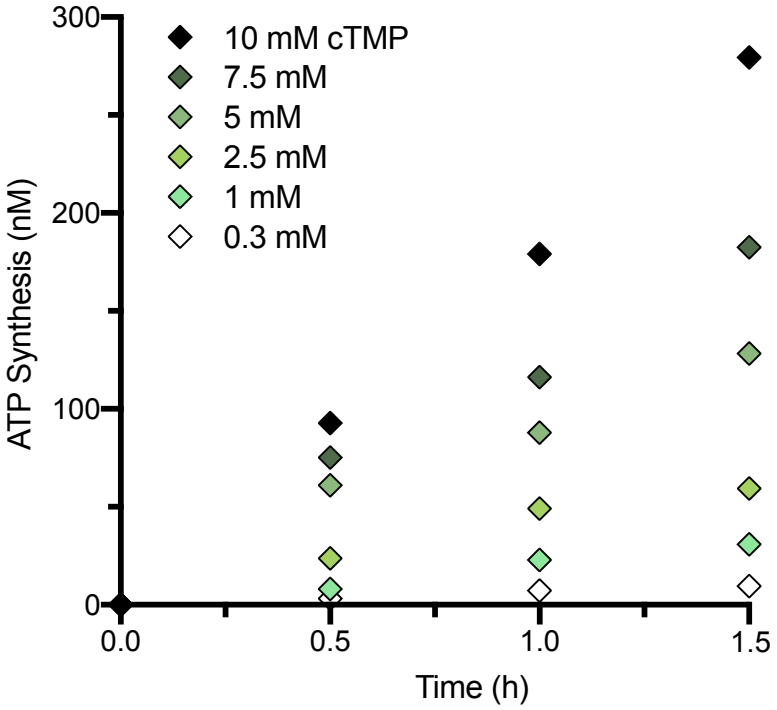


Figure 2.S3. ATP synthesis as a function of cTmp.

Reactions containing 10 mM cTmp, 50 mM boric acid, 150 mM $MgCl_2$, and varying concentrations of cTmp at pH 9.0 were allowed to react at 40 ºC for up to 1.5 hours. ATP concentrations were measured using the luciferase assay.

ATP [mM]	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1
UTP [mM]	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1
CTP [mM]	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1
GTP [mM]	-	1	1	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1
TMP [mM]	-	-	5	10	10	5	5	2.5	2.5	1	1	5	5	10	10	5	-		
Dilution	-	-	-	0.4	0.2	0.4	0.2	0.4	0.2	0.4	0.2	0.4	0.2	0.4	0.2	-	-		

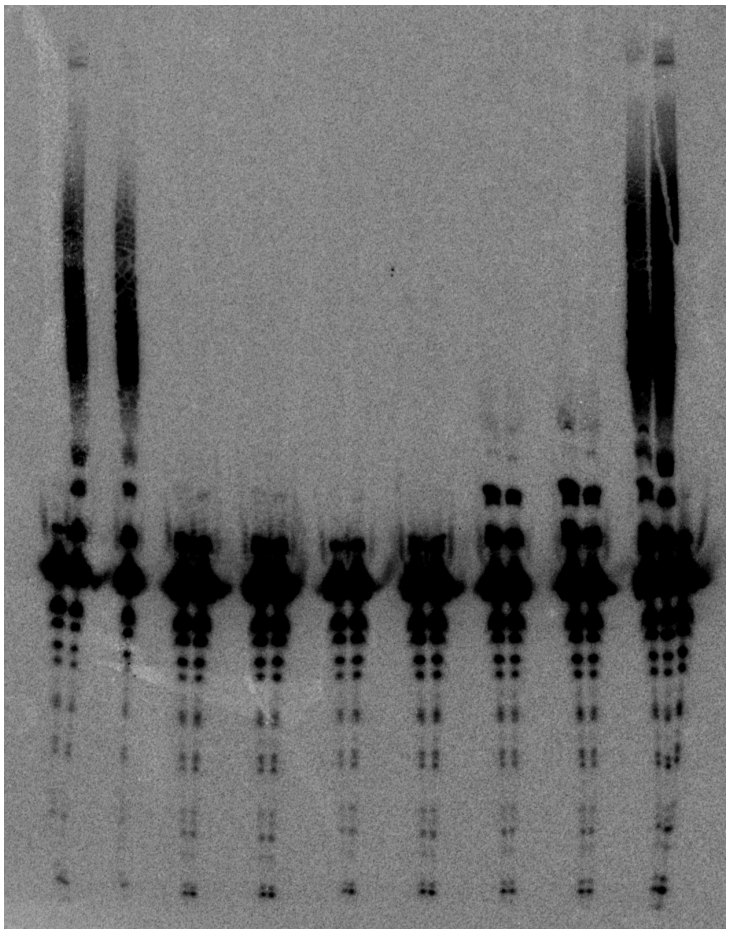


Figure S4. RNA extension by an RNA polymerase ribozyme using synthesized nucleotides.

Reactions containing 25 mM adenosine, 25 mM guanosine, 50 mM uridine, 50 mM cytidine, and indicated concentrations of cTmp in 50 mM $MgCl_2$ and 25 mM Tris-HCl at pH 9.0 were incubated at 40 °C for 4 hours. The reactions were diluted either 1:2.5 or 1:5 as in RNA polymerization reactions, composed of a radiolabeled RNA primer and the RNA polymerase ribozyme

(tC19Z). Some reactions were supplemented with either GTP or all four rNTPs, while a negative-control reaction contained no added nucleotides. The polymerization reactions were incubated for 36 hours at 17 °C and then resolved on a 20% polyacrylamide gel under denaturing conditions. While the no-reaction control showed a weak band corresponding to the one-nucleotide extension, the reactions using 5 mM and 10 mM cTmp showed extension by two nucleotides. Single nucleotide primer degradation is seen in the lower half of the gel. Consistent with what was previously observed for GTP synthesis, the reactions supplemented with GTP showed an extension of 4 nucleotides. Some reactions are shown with two different dilution factors.

Table 2.S1. Sequences

drz-Fprau-2 Primer (Fwd.)	5′-TTC CCG CGA AAT TAA TAC GAC TCA CTA
drz-Fprau-2 Primer (Rev.)	5′-GAC AAT ATT GCA TCT CAG CCA CAT
drz-Fprau-2 DNA Full Length	5′-TTC CCG CGA AAT TAA TAC GAC TCA CTA TAG GGG AGA TAT AAG GCT GTG GGA AGG TTC TCA CAG CTT TGC CGC TCG AAA CTT TGC ACA CCT CTA CGC GGT GGG TGG CAG GCA ACA CAT GAG AAT GTG GCT GAG ATG CAA TAT TGT C
tC19Z Template Elongation (Fwd.)	5′-GGG AGA AGC ACA CTA AAT TGA CCT GAC GAG GCC TAA GGG GCC GAA ACC GGG GCA ACC CCG
tC19Z Template Elongation (Rev.)	5′-CTG CCA ACC GTC AGA AGC GTG TCA TTG ACG ACG GGG TTG CCC CGG TTT CGG
tC19Z Template Primer (Fwd.)	5′-GCT AAT ACG ACT CAC TAT AGG GGA GAA GCA CAC TAA ATT G
tC19Z Template Primer (Rev.)	5′-CTG CCA ACC GTC AGA AG
tC19Z RNA Primer	5′-CUG CCA ACC G

2.8. References

1. Neveu, M.; Kim, H. J.; Benner, S. A. The “strong” RNA World hypothesis: Fifty years old. *Astrobiology* **2013**, 13, 391-403

2. Robertson, M. P.; Joyce, G. F. The origins of the RNA World. *Cold Spring Harb. Perspect. Biol.* **2012**, 13, 391-403.

3. Schrum, J. P.; Zhu, T. F.; Szostak, J. W. The origins of cellular life. *Cold Spring Harb. Perspect. Biol.* **2010**, 2, a002212.

4. Bartel, D. P.; Unrau, P. J.; Constructing an RNA World. *Trends Cell Biol.* **1999**, 12, 9-13.

5. Lehman, N. Lennarz, W. J.; Lane, M. D. *Encyclopedia of Biological* Chemistry 2nd ed., Academic Press, Elsevier, **2013**, pp 142-146.

6. Chen, X.; Li, N.; Ellington, A. D. Ribozyme catalysis of metabolism in the RNA World. *Chem. Biodivers.* **2007**, 4, 633-655.

7. Martin, L.; Unrau, P. J.; Muller, U. RNA synthesis by in vitro selected ribozymes for recreating an RNA World. *Life* **2015**, 5, 247-268.

8. Sutherland, J. D. The origin of life-out of the blue. *Angew. Chemie Int. Ed.* **2016**, 55, 104-121.

9. Sutherland J. D. Ribonucleotides. *Cold Spring Harb. Perspect. Biol.* **2010**, 2, a005439.

10. Sutherland, J. D.; Whitfield, J. N. Synthesis of potentially prebiotic RNA precursors: cytosine and guanine derivatives. *Tetrahedron Lett.* **1997**, 38, 1451-1454.

11. Wochner, A.; Attwater, J.; Coulson, A.; Holliger, P. Ribozyme-catalyzed transcription of an active ribozyme. *Science* **2011**, 332, 209-212.

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

13. Horning, D. P.; Joyce, G. F. Amplification of RNA by an RNA polymerase ribozyme. *Proc. Natl. Acad. Sci.* **2016**, 113, 9786-9791.

14. Lohrmann, R.; Orgel, L. E. Prebiotic activation processes. *Nature* **1973**, 244, 418-420.

15. Pasek, M. A.; Gull, M.; Herschy, B. Phosphorylation on the early earth. *Chem. Geol.* **2017**, 475, 149-170.

16. 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.* **2013**, 10, 10089-10094.

17. Pasek, M. A. Rethinking early earth phosphorus geochemistry. *Proc. Natl. Acad. Sci.* **2008**, 105, 853-858.

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

19. Tsuhako, M.; Fujimoto, M.; Ohashi, S.; Nariai, H.; Motooka, I. Phosphorylation of nucleosides with sodium cyclo-triphosphate. *Bull. Chem. Soc. Jpn.* **1984**, 57, 3274-3280.

20. Mohamady, S.; Taylor, S. D. Synthesis of nucleoside tetraphosphates and dinucleoside pentaphosphates via activation of cyclic trimetaphosphate. *Org. Lett.* **2013**, 15, 2612-2615.

21. Mohamady, S.; Taylor, S. D. Synthesis of nucleoside triphosphates from 2′-3′-protected nucleosides using trimetaphosphate. *Org. Lett.* **2016**, 18, 580-593.

22. Shepard, S. M.; Cummins, C. C.; Functionalization of intact trimetaphosphate: a triphosphorylating reagent for C, N, and O nucleophiles. *J. Am. Chem. Soc.* **2019**, 141, 1852-1856.

23. Sibilska, I.; Feng, Y.; Li, L.; Yin, J. Trimetaphosphate activates prebiotic peptide synthesis across a wide range of temperature and pH. *Orig. Life. Evol. Biosph.* **2018**, 48, 277-287.

24. Schwartz, A. W. Specific phosphorylation of the 2′- and 3′-positions in ribonucleosides. *J. Chem. Soc. D Chem. Comm.* **1969**, 23, 1393.

25. Saffhill, R. Selective phosphorylation of the cis-2′,3′-diol of unprotected ribonucleosides with trimetaphosphate in aqueous solution. *J. Org. Chem.* **1970**, 35, 2881-2883.

26. 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, 219-224.

27. Kim, H. J.; Benner, S. A. Abiotic synthesis of nucleoside 5′-triphosphates with nickel borate and cyclic trimetaphosphate (CTMP). *Astrobiology* **2021**, 21, 298-306.

28. Soukup, G. A.; Breaker, R. R. Relationship between internucleotide linkage geometry and the stability of RNA. *RNA* **1999**, 5, 1308-1325.

29. Li, Y.; Breaker, R. R. Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2′-hydroxyl group. *J. Am. Chem. Soc.* **1999**, 121, 5364-5372.

30. Levy, M.; Miller, S. L. The stability of the RNA bases: implications for the origin of life. *Proc. Natl. Acad. Sci.* **1998**, 95, 7933-7938.

31. Tisi, L. C.; White, P. J.; Squirrell, D. J.; Murphy, M. J.; Lowe, C. R.; Murray, J. A. H.; Development of a thermostable firefly luciferase. *Anal. Chim. Acta* **2002**, 457, 115-123.

32. Etiax, E.; Orgel, L. E. Phosphorylation of nucleosides in aqueous solution using trimetaphosphate: formation of nucleoside triphosphates. *J. Carbohydrates Nucleosides Nucleotides* **1978**, 5, 91-110.

33. Jones, H. W.; Monk, C. B.; Davies, C. W. The condensed phosphoric acids and their salts. Part IV. Dissociation constants of some trimetaphosphates. *J. Chem. Soc.* **1949**, 2693-2695.

34. Yamagata, Y.; Inoue, H.; Inomata, K. Specific effect of magnesium ion on 2′,3′-cyclic AMP synthesis from adenosine and trimetaphosphate in aqueous solution. *Orig. Life Evol. Biosph.* **1995**, 25, 47-52.

35. Lascu, I.; Gonin, P. The catalytic mechanism of nucleoside diphosphate kinases. *J. Bioenerg. Biomembr.* **2000**, 32, 237-246.

36. Passalacqua, L. F. M.; Jimenez, R. M.; Fong, J. Y.; Lupták, A. Allosteric modulation of the Faecalibacterium Prausnitzii hepatitis delta virus-like ribozyme by glucosamine 6-phosphate: the substrate of the adjacent gene product. *Biochemistry* **2017**, 56, 6006-6014.

37. Passalacqua, L. F. M.; Dingilian, A. I.; Lupták, A. Single-pass transcription by T7 RNA polymerase. *RNA* **2020**, 26, 2062-2071.

38. Rogers, J.; Joyce, G. F. A ribozyme that lacks cytidine. *Nature* **1999**, 402, 323-325.

39. Reader, J. S.; Joyce, G. F. A ribozyme composed of only two different nucleotides. *Nature* **2002**, 420, 841-844.

40. Moretti, J. E.; Muller, U. F. A ribozyme that triphosphorylates RNA 5′-hydroxyl groups. *Nucleic Acids Res.* **2014**, 42, 4767-4778.

41. Dolan, G. F.; Akoopie, A.; Muller, U. F. A faster triphosphorylation ribozyme. *PLoS One* **2015**, 10, e0142559.

42. Adamala, K.; Szostak, J. W. Nonenzymatic template-directed RNA synthesis inside model protocells. *Science* **2013**, 342, 1098-1100.

43. Yamagami, R.; Bingaman, J. L.; Frankel, E. A.; Bevilacqua, P. C. Cellular conditions of weakly chelated magnesium ions promote RNA stability and catalysis. *Nat. Commun.* **2018**, 9, 2149.

CHAPTER 3: RIBOZYME MEDIATED REPAIR OF DAMAGED AND MISMATCHED RNA SEQUENCES

Kent, A. D.; Yang, L. L.; Lee I. A.; Lupták, A. Ribozyme mediated repair of damaged and mismatched RNA sequences.

Manuscript in preparation.

Contribution Statement

A.D.K. performed the experiments, analyzed the data, generated the figures, and wrote the manuscript. L.Y. collected kinetics data and assisted with experiments. I. L. assisted with experiments. A.L. conceived the study.

3.1 Abstract

Prior to the advent of modern life, the early Earth is likely to have progressed through an RNA based world, because RNA can act both as an information carrying polymer as well as a catalyst for chemical reactions. Critical to such an evolving world would have been a system of replication that allowed for the transfer of heritable information such as an RNA polymerase. Though a substantial body of work has been published on evolving such a polymerase, many iterations still suffer from low yield of full-length sequences and low number of active sequences as a result of premature termination and non-optimal fidelity, respectively. One of the first phenotypes that could lend an evolutionary advantage to an early protocell would be the ability to replicate with higher stability and fidelity. Cells today achieve this through a variety of means including complex damage repair mechanisms and proofreading pathways, but these would likely not have had dedicated machinery in an evolving world. Herein we demonstrate the first example of ribozyme-mediated repair of damaged and mismatched RNA sequences. Under conditions of high concentrations of pyrophosphate, a polymerase ribozyme (tC19Z) is capable of repairing a 2´-3´

cyclic-phosphate terminated RNA sequence, a mismatch terminated sequence, or both to subsequently allow for further extension along a template. Increased fidelity and processivity of polymerization advances the longstanding goal of developing a truly self-replicating polymerase ribozyme.

3.2 Introduction

The cellular functions of ribonucleic acid (RNA), including critical roles in translation, catalysis, and metabolite-dependent gene expression, suggest that it may have been the first biopolymer to assume both informational and catalytic roles on an evolving Earth.[1-5] Replication is a process fundamental to life and is the mechanism by which heritable information is stored and propagated. Though today this is done through protein-based polymerase enzymes, in an RNA World, prior to the advent of genetically encoded proteins, replication would likely have been carried out by RNA polymerase ribozymes.[6,7] These ribozymes would have to have sufficient activity and fidelity to build up an information storage system, and could serve as the central molecules linking phenotype to genotype.[8,9]

Transphosphorylation of the α-phosphate of a triphosphorylated nucleoside to a 3′ hydroxyl of another nucleoside, with concomitant release of a pyrophosphate, is the central reaction catalyzed by enzymes replicating phosphodiester-based informational polymers. A large body of work using molecular evolution and engineering has been generated to isolate ribozymes that catalyze this reaction, beginning with the first *in vitro* selection of a ligase ribozyme.[10-14] This ribozyme has continued to be evolved into progressively more sophisticated polymerases capable of production of other functional nucleic acids, reverse transcription, and transcription of XNAs, among others.[15-20] However, even the most advanced polymerase ribozymes suffer from low yields of full-length products (through premature termination and degradation) and mutation burdens that reduce the activity of produced transcripts.[21-23] Thus, a mechanism of repair of both of

degraded and mismatched sequences would lend an evolutionary advantage to an evolving system. However, very few mechanisms of RNA repair have been reported, and none that are ribozyme mediated.[24-26] Additionally, there is a lack of data on the role of pyrophosphate on the activity of these ribozymes and the related stability of the transcripts.

We present below our study of one of the evolved RNA polymerases, the tC19Z ribozyme, and the effect of pyrophosphate on its activity. Herein, we began by characterizing the kinetics of the polymerization reaction, and the effect the addition of pyrophosphate had on extension. We then demonstrated that addition of pyrophosphate and tC19Z to 2′-3′ cyclic phosphate-terminated (damaged) RNA sequences facilitated damage repair and subsequent RNA extension. Finally, we show that addition of pyrophosphate to mismatched RNA sequences allowed for repair and alteration of extension profiles.

3.3 Results

Kinetic analysis of extension by tC19Z.

The tC19Z ribozyme was the product of evolution from R18, the first polymerase ribozyme generated from evolution of the class I ligase.[16,18] It was evolved to have improved sequence generality and polymerase activity, though the template sequence still plays a major role in replication efficiency. In order to extend a sequence, the 5′ end of the tC19Z ribozyme base-pairs with the 5′ end of the template sequence, which is in turn bound to a primer at its 3′ end (**Figure 3.1A**). The ribozyme consists of two domains, a catalytic domain made up of the class I ligase core and an accessory domain, and can extend the primer along the template sequence. We began by using a primer/template sequence that had been characterized by the Holliger lab in their publication of the tC19Z ribozyme (**Figure 3.1A and Figure 3.S1**), and have used variations of that sequence throughout this paper.

We began characterizing factors that influenced the extension rate including nucleoside triphosphate concentration, $MgCl_2$ concentration, and temperature (**Figure 3.S2**). To do this, we used a 5′ ^{32}P-labeled primer, and tracked the extension reaction by single nucleotide resolution polyacrylamide gel electrophoresis (PAGE). We found that the reactivity drops off sharply with decreasing $MgCl_2$ concentrations, and increases in temperature led to a substantial increase in degradation products. We further fully characterized the kinetics of the forward reaction with regard to changing concentrations of NTPs (**Figure 3.1B** and **3.1D** and **Figure 3.S3**). Initial rates were plotted as a function of substrate concentration and a K_M for rNTPs was estimated to be ~4 mM (**Figure 3.1E**).

Pyrophosphate inhibition of ribozyme mediated polymerization.

The centrality of phosphate in the origin of life stems from its presence in the backbone of RNA and DNA and is inextricably coupled to replication.[27] Pyrophosphate is the byproduct of RNA polymerization but has also been demonstrated to be a source of available phosphate on an early Earth.[28-31] Because of its prebiotic availability and likely accumulation from replication, we decided to investigate the effect of pyrophosphate on ribozyme mediated replication. We noticed a sharp decrease in extension products upon addition of high concentrations of exogenous pyrophosphate to the polymerization reaction (**Figure 3.S4A**). Using the standard reaction conditions for polymerization (200 mM $MgCl_2$ and 4 mM NTPs), we demonstrated that at 100 μM, exogenous pyrophosphate began to inhibit the forward polymerization reaction (**Figure 3.1C**). We then plotted the initial rates of tC19Z mediated polymerization as a function of pyrophosphate concentration (**Figure 3.1F**). Because we saw significant inhibition by pyrophosphate, we began investigating a possible mechanism of inhibition. To start, we noticed that upon addition of 10 mM pyrophosphate, a visible white precipitate formed which we attributed to complexation with Mg^{2+} because magnesium pyrophosphate is practically insoluble under most aqueous conditions.[32] We performed the extension reaction under standard ribozyme conditions with addition of 10 mM

pyrophosphate and following reaction completion, separated the precipitate from the solution. The precipitate was resuspended and redissolved and the two fractions were analyzed by high-resolution PAGE (**Figure 3.S4B**). We found that a substantial portion of the ^{32}P-labeled primer RNA co-precipitated with the magnesium-pyrophosphate; however, the extension reaction only occurred in solution. This led us to hypothesize that one mechanism of inhibition by pyrophosphate may be through the removal of RNA from the reaction.

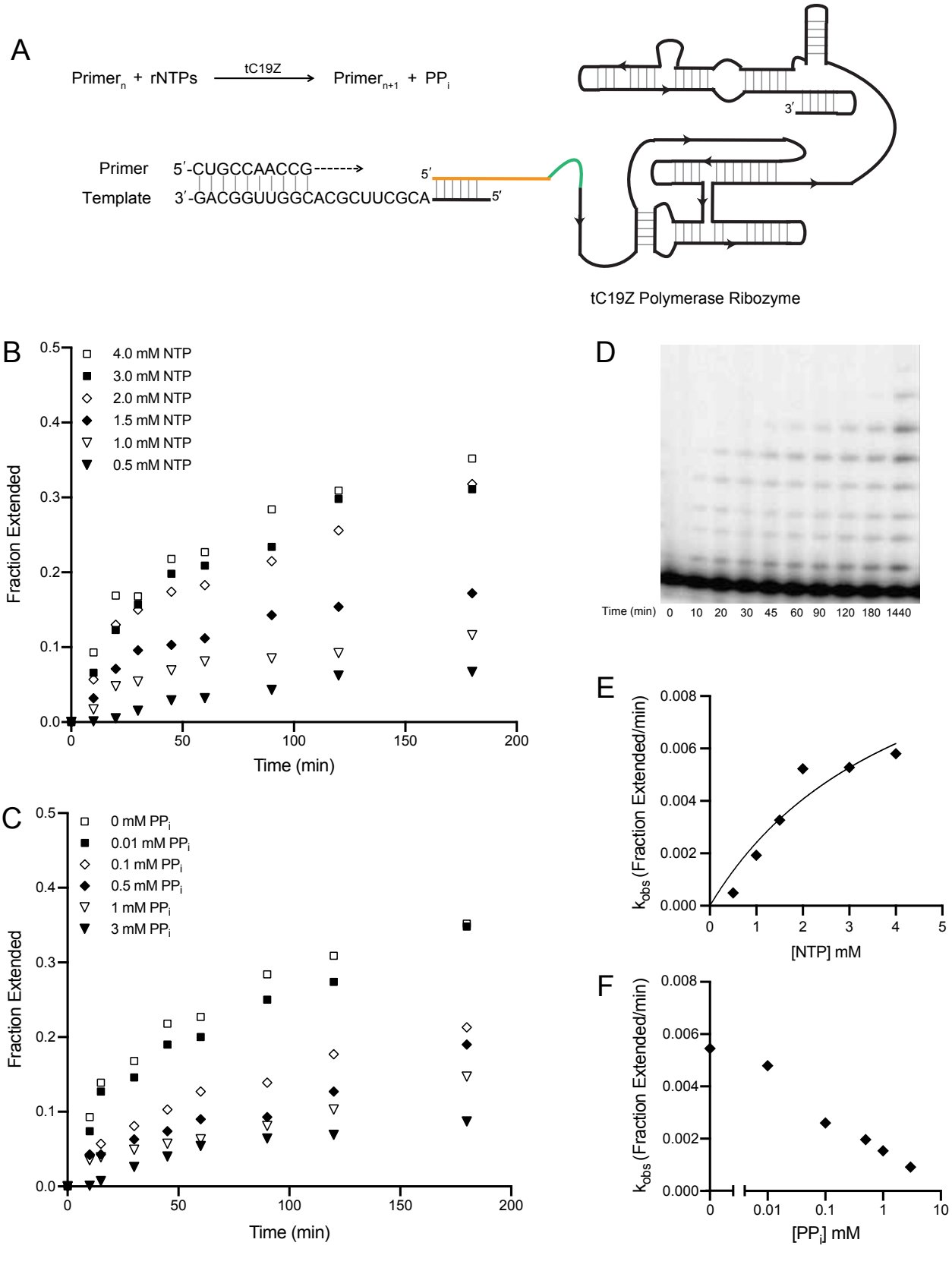

A
Primer$_n$ + rNTPs
tC19Z
Primer$_{n+1}$ + PP$_i$
Primer 5′-CUGCCAACCG------>
Template 3′-GACGGUUGGCACGCUUCGCA
5′
5′
3′
tC19Z Polymerase Ribozyme
B
4.0 mM NTP
3.0 mM NTP
2.0 mM NTP
1.5 mM NTP
1.0 mM NTP
0.5 mM NTP
Fraction Extended
Time (min)
C
0 mM PP$_i$
0.01 mM PP$_i$
0.1 mM PP$_i$
0.5 mM PP$_i$
1 mM PP$_i$
3 mM PP$_i$
Fraction Extended
Time (min)
D
Time (min) 0 10 20 30 45 60 90 120 180 1440
E
k_{obs} (Fraction Extended/min)
[NTP] mM
F
k_{obs} (Fraction Extended/min)
[PP$_i$] mM

Figure 3.1. Kinetic analysis of primer extension by tC19Z and inhibition by pyrophosphate. (A) An *in vitro* evolved polymerase ribozyme, tC19Z, base-pairs to a template strand and adds nucleotides to the 3′ end of a primer with pyrophosphate as the byproduct of the reaction. (B) The fraction extended of the ^{32}P-labeled primer (shown in A) is quantified as a function of time at varying NTP concentrations by high resolution PAGE. (C) The fraction of the ^{32}P-labeled primer extended quantified as a function of time at varying pyrophosphate conditions. (D) High-resolution PAGE of the primer extension at 4 mM rNTPs which is plotted on the graph in B. (E) The fraction of primer extended per minute is plotted as a function of NTP concentration and shows an estimated K_M of ~4 mM. (F) The fraction extended per minute is plotted as a function of the log of pyrophosphate concentration.

Generation of 2′-3′ cyclic-phosphate-terminated hairpins.

RNAs terminated with 2′-3′ cyclic phosphates are the products of degradation via a 2′ oxyanion nucleophilic attack on the adjacent phosphate—a reaction common to self-cleaving ribozymes, many RNases, and base-catalyzed RNA degradation.[33-35] This reactivity is responsible for the relatively low stability of RNA, when compared with DNA or other nucleic acids that lack a phosphate-adjacent hydroxyl group. Because this type of damage is the most common form of RNA degradation, we devised an RNA hairpin with terminal 2′-3′ cyclic phosphate damage that serves as a primer-template substrate of the polymerase ribozyme to test repair ability of the ribozyme. In order to generate the hairpin, we first engineered a four-nucleotide loop into a primer/template complex to both increase local concentration of RNA components, as well as for ease of preparation. We generated the hairpin via an *in vitro* transcription reaction of a self-cleaving HDV-like ribozyme (drz-Mbac-1) fused to the hairpin (**Figure 3.S5**).[36] The ribozyme self-cleaves from the 5′ hairpin sequence, leaving a 2′-3′ cyclic phosphate on the 3′ end of the “primer” strand. This 2′-3′ cyclic phosphate acts as a chain terminator, blocking the 3′ hydroxyl from

extension during polymerization (**Figure 3.2A**). The cyclic phosphate can be removed enzymatically by T4 polynucleotide kinase (PNK) to serve as a positive control for testing the extension efficiencies of the newly generated hairpins (**Figure 3.S6** and **3.S7**).[37,38]

Pyrophosphate-mediated repair of 2′-3′ cyclic-phosphate-terminated RNA sequences.

We generated RNA hairpins terminating on the 3′ "primer" end in each of the four canonical nucleotides with the corresponding match on the "template" strand. No extension of the sequence by the ribozyme was seen when polymerization of the terminated sequence was attempted. However, upon digestion by PNK, or addition of pyrophosphate, each of the sequences could be extended by the polymerase ribozyme (**Figure 3.2B**, **3.2C, 3.S8**). The extension efficiency, as well as repair efficiency, is largely dependent on the identity of the sequence with U/A (U on the "primer" strand and A on the "template" strand) repaired the most and G/C repaired the least efficiently (A/U = 22%, U/A = 49%, C/G = 17%, G/C = 8%). The kinetics of extension following repair with pyrophosphate can be visualized with respect to enzymatic PNK repair for each of the four hairpins (**Figure 3.2D** and **3.S9**). An additional kinetic comparison is given for the A/U hairpin where pyrophosphate is added to a PNK digested sequence to visualize the role of pyrophosphate inhibition on hairpin extension.

We hypothesized that one mechanism of repair could be through pyrophosphate acting to remove the 3′ terminal, damaged nucleotide, thus we set out to further investigate this possibility. We endeavored to visualize the removal of a nucleotide by pyrophosphate by high-resolution page; however, in every iteration and reaction condition, we were unable to see a shift that could definitively be attributed to removal of a single nucleotide from the 2′-3′ cyclic phosphate-terminated hairpin. We hypothesize that this is due to the anomalous gel shifts that arise from treatment with PNK and acid respectively (**Figure 3.S6**). It has been previously reported that treatment of a 2′-3′ cyclic phosphate-terminated sequence with PNK to remove the cyclic phosphate leads to an

upward gel shift, and treatment of a 2′-3′ cyclic phosphate-terminated sequence with acid to generate a linear phosphate leads to a downward gel shift.[39-41] We see consistently see the opposite effect and hypothesize that this is a property of the hairpin sequence, which differs structurally from the previously reported experiments with linear RNA sequences. Thus, the gel shift from removal of a nucleotide is likely being occluded by other bands.

We began by performing repair experiments that could indirectly indicate the removal of a nucleotide, first we omitted the rNTP from the reaction buffer that would correspond to the damaged nucleotide to be replaced. In the case of an A/U hairpin, we omitted ATP from the rNTPs mixture since following excision of the terminal damaged adenosine, the ribozyme would need to replace it to continue extension. We indeed found that if ATP was not present in the reaction, the sequence was not extendable (**Figure 3.S10**) To demonstrate additional evidence of repair, we transcribed hairpins without α-^{32}P-ATP and subsequently performed the polymerization reaction with α-^{32}P-NTPs. We observed that the α-^{32}P-NTP is only incorporated in sequences that have been repaired (**Figure 3.S11**). However, it is notable that the ribozyme can incorporate a mismatch when provided with certain nucleotides. This misincorporation occured with lower efficiency than the matched nucleotide, but it is nonetheless detectable, especially with α-^{32}P-UTP. Finally, we also disproved the hypothesis that the 2′-3′ cyclic phosphate was being opened by pyrophosphate to form a triphosphate by adding uridine and pyrophosphate to extension reactions. No addition to the hairpin was seen under these conditions indicating that this is not the pathway of repair (**Figure 3.S6**).

Finally, in order to demonstrate that addition of pyrophosphate to ribozyme mediated extension reaction did repair the sequence and allowed for extension of the hairpin, we purified and prepared extension reactions for Illumina sequencing. We isolated extended A/U hairpins from pyrophosphate treated polymerization reactions. Initial sequencing data showed extension

products of varying lengths indicating that the sequences are being repaired and correctly extended (**Figure 3.S12**).

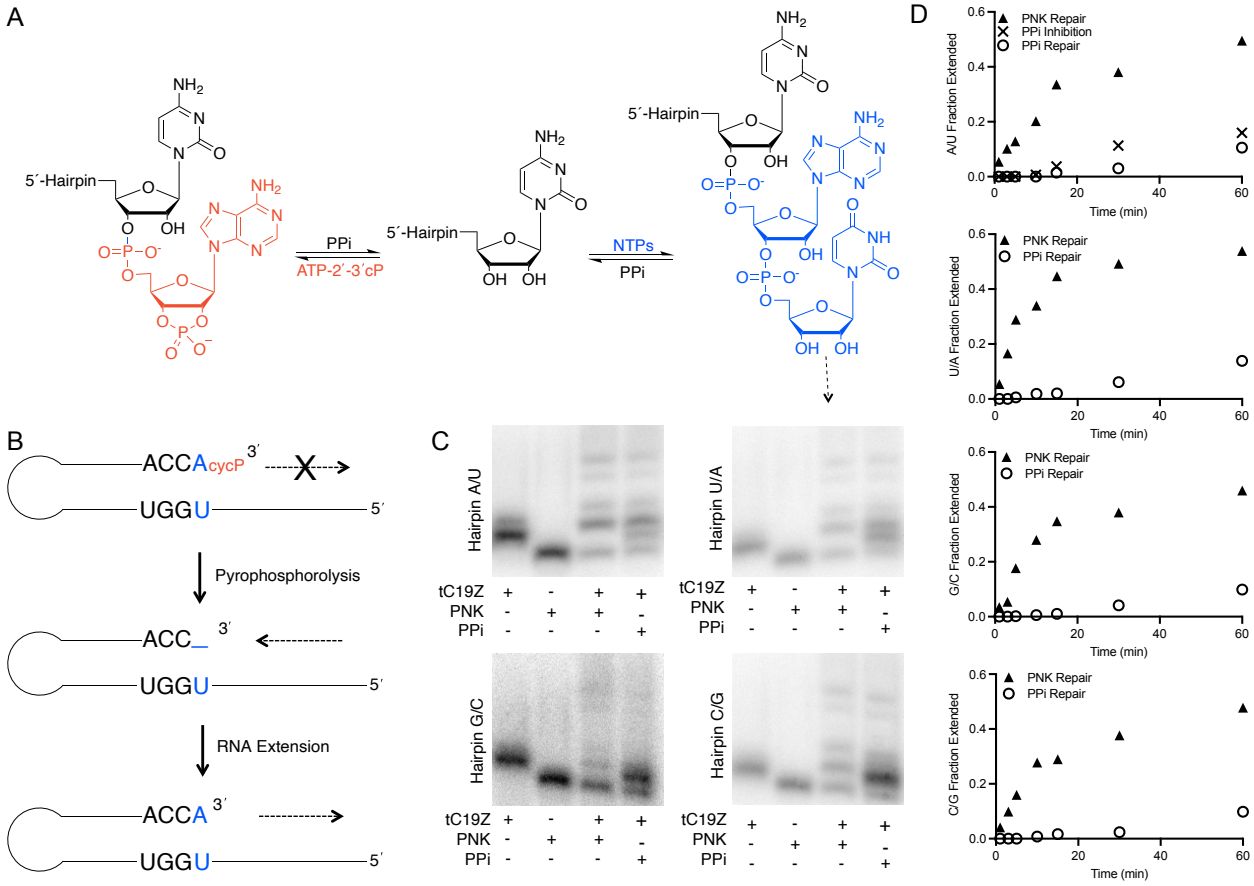


Figure 3.2. Pyrophosphate-utilizing, ribozyme-mediated repair of terminal 2′-3′ cyclic phosphate damage. (A) The 3′ end of a hairpin is damaged with a 2′-3′ cyclic phosphate (red) and is not extendable by tC19Z. Upon addition of pyrophosphate the damaged nucleotide triphosphate can be removed, and polymerization of undamaged nucleotides can occur (blue). (B) A graphical representation of the 2′-3′ cyclic phosphate damaged A/U hairpin and subsequent repair. (C) High-resolution PAGE images of α-^{32}P-ATP labeled hairpins terminated in all four base-pairs (shown left of each image). Lane 1 is a control reaction showing that the 2′-3′ cyclic-phosphate-terminated hairpin is not extended by the ribozyme. Lane 2 shows the gel mobility shift following removal of the 2′-3′ cyclic phosphate by PNK. Lane 3 is a positive control showing extension of

the hairpin by tC19Z following 2′-3′ cyclic phosphate removal by PNK. Lane 4 is the repair of the hairpin by tC19Z and pyrophosphate, and the resulting extension products. (D) Extension kinetics for each of the hairpins derived from the high-resolution PAGE gels. Closed triangles are extensions resulting from PNK repaired hairpins. Open circles are extension resulting from pyrophosphate repaired hairpins. Finally, because pyrophosphate inhibits the forward extension reaction an additional control is included for the A/U hairpin, where the crosses denote inhibition of extension of the PNK digested hairpin.

Effect of pyrophosphate on terminally mismatched hairpin sequences

In order to prevent propagation of mutations, modern replication machinery has dedicated active sites or entire protein complexes to correct mismatched base-pairs, but at the onset of living, evolvable systems, it was likely that mismatch repair was not available as a distinct biochemical activity.[42,43] In addition to damage repair, pyrophosphate-dependent mismatch repair would also have been a desirable feature of an early replicating polymerase, lending itself to increased fidelity of replication. To test the hypothesis that pyrophosphate could also correct a mismatch, we once again generated hairpin sequences using the HDV-like ribozyme; however, this time with a terminal 3′ mismatch. We began by generating an A/C mismatch and adding pyrophosphate to the cyclic-phosphate-terminated hairpin. We did not observe any repair and subsequent extension of the sequence. Upon digestion with PNK and as well as addition of NTPs, we still did not observe any extension, the mismatch effectively acting as a terminator. However, after both digestion with PNK as well as addition of pyrophosphate, we observed ribozyme-mediated extension of the sequence (**Figure 3.3A** and **Figure 3.S13**).

We generated all other possible mismatch sequences, and observed variable efficiencies of extension reactions: for certain hairpins, both damage and mismatch repair was observed with addition of pyrophosphate. Specifically, the purine/purine mismatches (A/A, G/A, and A/G) and

the G·U wobble pairs (G/U and U/G) could be repaired by the polymerase when pyrophosphate was added to the extension reaction of these cyclic phosphate terminated mismatches. (**Figure 3.3B**). Similar to the variable efficiency of extension of different terminal nucleotides in the matched hairpins, the mismatches were also extendable with varying efficiencies either following 2'-3' cyclic phosphate removal with PNK or both PNK and pyrophosphate (**Figure 3.3C** and **3.S14**). The pyrimidine/pyrimidine (C/C, U/U, C/U, and U/C) mismatches were best extended by the ribozyme following 2′-3′ cyclic phosphate removal by PNK, and addition of pyrophosphate made no substantial difference on extension, indicating that tC19Z is more tolerant of some mismatches over others. Finally, some sequences did show similar behavior to the A/C mismatch and only showed extension following treatment with PNK and pyrophosphate (C/A, U/G, G/U, and the purine/purine mismatches) although with varying efficiencies. The preference for the ribozyme to extend some sequences over others was confirmed with select primer/template constructs (**Figure 3.S15**). Removal of the mismatched base and replacement with the correct base on the 3′ end of the sequence was confirmed for the G/U mismatch by sequencing (**Figure 3.3D**).

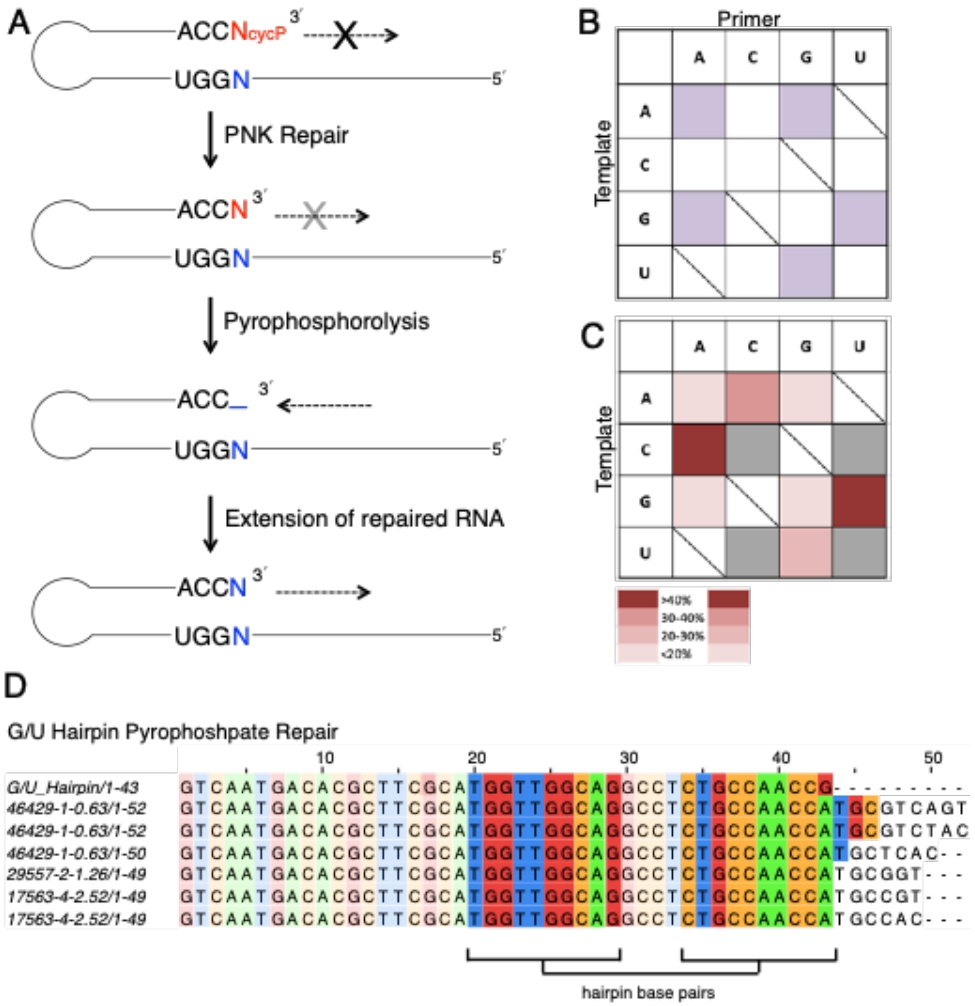


Figure 3.3. Effects of pyrophosphate on terminally mismatched hairpins. (A) A terminally mismatched and cyclic phosphate damaged hairpin can either be repaired by pyrophosphate and extended, repaired by PNK and extended, or repaired by PNK and pyrophosphate and extended dependent on the identity of the mismatch. (B) The primer (top) and template (side) nucleotide identities are shown for all combinations. Boxes in purple indicate which mismatch sequences are extended following addition of pyrophosphate when cyclic phosphate damage is also present. (C) Mismatches that have first been treated with PNK. Shades of red indicate percent of extension that occurs following addition of pyrophosphate. Grey boxes indicate the mismatches that are extended by tC19Z prior to addition of pyrophosphate, and addition of pyrophosphate does not enhance the extension. (D) Sequencing data of the G/U mismatch indicates replacement of the

mismatched base with the correct base can occur during a tC19Z polymerization reaction treated with pyrophosphate. The G/U mismatch hairpin is the first sequence shown in the 5′ to 3′ direction. Subsequent sequences are in varying states of extension and show the repaired terminal 3′ base. Bases with no color are part of the DNA ligation primer used to prepare the hairpins for sequencing.

3.4 Discussion

It would have been critical for an evolving replication system to achieve mechanisms of repair to increase genome stability and fidelity. The tC19Z ribozyme was initially selected to extend a primer along a template in the effort to generate a system of self-sustained replication. Although many ribozymes have been shown to have varying levels of replication efficiency, no ribozyme to date has been reported that has RNA repair capabilities. We have demonstrated that the tC19Z ribozyme can use pyrophosphate to repair 2′-3′ cyclic phosphate-damaged RNA sequences, effectively rescuing them for further extension. Additionally, tC19Z can repair some mismatched sequences, which may present a solution to the fidelity problems faced by early replication ribozymes.

We were particularly interested in the mechanism of repair employed by tC19Z when utilizing pyrophosphate and have made some strides to understand the pathway. Most of our experiments have focused on understanding the repair of 2′-3′ cyclic phosphate-terminated hairpins, and we hypothesize that the ribozyme is catalyzing the removal of the terminal, damaged nucleotide using pyrophosphate. We have been unable to directly visualize the removal of a single nucleotide, likely as a result of the anomalous migration on a high-resolution polyacrylamide gel exhibited when the terminal cyclic phosphate is removed from a hairpin. This has occluded our efforts to definitively conclude that removal of the terminal nucleotide is the mechanism of action and has induced us to pursue indirect methods.

When a repair reaction was not provided with exogenous terminal nucleotide (ATP), we were unable to observe primer extension. This result would indicate that the 3′ nucleotide is being removed by tC19Z utilizing pyrophosphate, and thus the extension reaction is stalled without the undamaged, terminal nucleoside triphosphate starting the polymerization. In another experiment when the hairpin was not transcribed with a ^{32}P-radiolabel, but rather provided with α-^{32}P-NTPs, the hairpins were only visualized when they had been repaired in some manner. The PNK processed hairpin was labeled because it has a terminal 3′-OH to which the ribozyme can add a ^{32}P-nucleotie . However, the reaction that contained the pyrophosphate-repaired hairpin was also labeled, indicating that it also must have a free 3′-OH. Once again, the anomalous shift of the PNK-processed hairpin occludes the result, because we were unable to determine a difference in length between the two repaired hairpins. From this experiment we also somewhat surprisingly found that tC19Z can incorporate mismatches into the template sequence in appreciable amounts. We saw that when provided with only one nucleotide, the ribozyme would incorporate it into the sequence even if it was mismatched with the template base. Overall, we observed that tC19Z is more efficient at incorporating some mismatches over others and has a preference for UTP.

To explore alternative mechanisms, we hypothesized that pyrophosphate could potentially act through opening the 2′-3′ cyclic phosphate, thus generating either a 2′ or 3′ triphosphate, then utilizing the pyrophosphate to activate an incoming nucleoside *in situ* for transesterification. We tested this mechanism by adding uridine to the pyrophosphate-repair reactions, rationalizing that the 5′-OH of the uridine (the first templated nucleotide) could displace the pyrophosphate to begin the polymerization reaction. However, we found that uridine was not incorporated into the reaction at any stage, indicating that this is likely not the mechanism of repair. Taken in concert, these findings suggest that the ribozyme is using pyrophosphate to remove the 2′-3′ cyclic phosphate terminated nucleotide from the hairpin in order to repair the sequence for further extension.

Further experimentation is necessary though, to fully characterize the mechanism of repair.

With the addition of mismatches to the hairpins, the effect of tC19Z and pyrophosphate becomes increasingly complex. For some mismatches, the addition of pyrophosphate to the extension reaction removes the 2′-3′ cyclic phosphate damage and allows for further extension of the sequence, but this primarily occurs for the purine/purine mismatches and wobble mismatches. Previously published data, as well as our observations on single nucleotide incorporation and patterns of extended sequences, showed that tC19Z prefers some nucleotides and sequences over others. For example, we have observed that the ribozyme prefers to both add and repair uridine nucleotides and terminal sequences, respectively, over the other nucleotides. After removal of the 2′-3′ cyclic phosphate, we observed that tC19Z will extend pyrimidine/pyrimidine mismatched sequences quite efficiently, not discriminating against the mismatch. Thus, addition of pyrophosphate to these sequences seems to have no effect on the extension outside of simply inhibiting the extension reaction. On the other hand, for some sequences, such as the A/C mismatch or the G/U wobble, the ribozyme does not extend the sequence unless pyrophosphate is added. Some mismatches are repaired or extended less efficiently, specifically, the purine/purine mismatches. We hypothesize that there is a structural basis for the trends in ribozyme mediated extension of mismatched sequences and the effect of pyrophosphate on extension. The trends we observed could suggest that narrowing or widening of the terminal base-pair affects the recognition and polymerization capabilities of the ribozyme.[45-49] However, for this ribozyme-hairpin RNA complex, it would be difficult to definitively conclude what the ramifications of the mismatches are on tC19Z polymerization reactivity without additional structural data.

We have shown that tC19Z has the ability to extend various mismatches with or without pyrophosphate, but critically, we wanted to establish the mechanism of repair, and thus demonstrate the potential to improve ribozyme fidelity. To this end we sequenced the PNK-processed, pyrophosphate treated G/U wobble mismatched hairpin. The sequencing data we obtained

comprised primarily of truncated sequences, and over 90% of the full-length sequences were not extended or corrected. However, we observed some sequences that were corrected, and a few sequences that were both corrected and extended within the sequencing data. The corrected base, from a guanosine to an adenosine, was statistically significant over background sequencing error. This confirms that we can correct a mismatch during ribozyme mediated polymerization with addition of pyrophosphate by removing the terminal nucleotide.

Modern biological systems have numerous, distinct complexes and pathways dedicated to repairing each different type of damage that could arise. Early replicating systems in an RNA world would not have had those functionalities or diversity available to them. However, this would have been an entirely necessary set of reactivities for an such a system in order to evolve heritability of long genomes. It has been widely hypothesized that in order for known polymerase ribozymes to truly achieve efficient self-replication capabilities, there would need to be an increase in the current standard of fidelity of replication. Unfortunately, this has been an elusive goal for even the most evolved polymerase ribozymes. Rather than further evolving polymerase ribozymes, we address this shortcoming by demonstrating that a highly evolved polymerase ribozyme (tC19Z) has repair functionalities. The ability for tC19Z to rescue 2′-3′ cyclic phosphate-terminated sequences that can be generated from base-catalyzed cleavage, lends itself to a more stable early genome. Additionally, the fact that tC19Z in concert with pyrophosphate can correct mismatched base-pairs suggests a possible mechanism for improving replication fidelity of polymerase ribozymes. These results indicate that we may be one step closer to generating truly self-replicating systems that could have evolved in an RNA world.

3.5 Materials and Methods

Cloning tC19Z into a pUC19 vector.

The insert, tC19Z, was amplified with primers containing the restriction enzyme sites BamHI and EcoRI (**Table 3.S1**). The tC19Z insert and pUC19 vector were double digested with BamHI and EcoRI for four hours. Additionally, the pUC19 vector was further treated with calf-intestinal, alkaline phosphatase (CIP) for one hour following the double digestion. The insert was ligated into the vector using T4 DNA ligase and incubating at 16 ºC for 16 hours. The resulting ligation was transformed into DH5α competent cells and subsequently plated on ampicillin containing agar plates which were grown at 37 ºC for 14 hours. Single colonies were picked and grown in 3 mL of ampicillin containing LB broth for 14 hours. Cultures were subjected to alkaline lyses and the plasmids were purified. The presence of tC19Z in the vector was confirmed by sequencing. Future PCR reactions of tC19Z were done fresh from the plasmid for each batch of ribozyme.

PCR and Transcription of tC19Z and generation of hairpin constructs.

Polymerase chain reaction (PCR) was performed to isolate tC19Z from the plasmid prepared above. 5 ng of the plasmid, 0.5 µM of each primer (**Table 3.S1**), and *Taq* DNA polymerase were used to amplify the sequence at an annealing temperature of 49 ºC. PCR products of tC19Z from the plasmid prepared above were used for in vitro transcription. Transcription was performed under the following conditions: 50 mM Tris HCl, 2 mM spermidine, 12 mM $MgCl_2$, 8 mM DTT, 2 mM each rNTP, and 0.01% Triton X-100 at a total volume of 100 µL. T7 RNA polymerase was added to the reaction and allowed to react at 37 ºC for 3 hours. The transcriptions were desalted with G15 Sephadex and then purified with polyacrylamide gel electrophoresis and subsequent elution and ethanol precipitation.

To prepare the RNA hairpins (**Table 3.S1**), the hairpin sequence (0.5 µM) was amplified with the drz-Mbac-1 HDV-like ribozyme (0.5 µM) (**Table 3.S1**). Following amplification, the sequence was transcribed and with α-^{32}P-ATP by adding 50 mM Tris HCl, 2 mM spermidine, 12 mM $MgCl_2$ 8 mM DTT, 2 mM GTP, 2 mM CTP, 2 mM UTP, 200 µM ATP and 0.01% Triton X-100 to a

50 µL reaction containing T7 RNA polymerase and subsequently reacting at 37 ºC for 2 hours. The HDV-like ribozyme self-cleaves co-transcriptionally to yield the cyclic phosphate terminated hairpins. The transcriptions were desalted with G10 Sephadex and then purified with polyacrylamide gel electrophoresis and subsequent elution and ethanol precipitation.

Adding γ-^{32}P-ATP to an RNA primer.

An RNA primer 10 nucleotides in length was obtained from Dharmacon (**Table 3.S1**). The RNA primer was added at a concentration of 5 µM to a reaction containing γ-32P ATP, 1x polynucleotide kinase buffer, and 5 units polynucleotide kinase to add a ^{32}P label to the 5`-end of the sequence. The reaction was incubated at 37 ºC for 2 hours and subsequently purified on a 12% denaturing polyacrylamide gel, followed by elution and ethanol precipitation.

Kinetic analysis of RNA primer extension by tC19Z.

The K_M of the reaction was determined by monitoring the extension of a ^{32}P-labeled primer by tC19Z over time and varying concentrations of nucleoside triphosphates. Equimolar concentrations (0.5 uM) of tC19Z, primer, and template (**Table 3.S1**) were mixed and heated to 82 ºC for 2 minutes, then cooled slowly to 17 ºC for 10 minutes. Buffer containing 200 mM $MgCl_2$, 50 mM Tris buffer, and varying concentrations of rNTPs was added to the RNA and the reaction was incubated at 17 ºC for 24 hours. Time points were taken by removing 5 µL aliquots and quenching with 250 mM EDTA and 7 M urea gel-loading loading buffer. An 18% polyacrylamide gel was prepared by generating a solution containing 18% w/v Acrylamide/Bis 19:1 solution, 7 M urea, and 0.5x TBE Buffer (50 mM Tris, 45 mM boric acid, and 0.5 mM EDTA, pH 8.4). The crosslinking reaction was initiated by addition of 0.1% w/v of ammonium persulfate, and 6.7 mM N,N,N`,N`-tetramethylethylenediamine and allowed to polymerize between glass plates to form a gel 0.4 mm in depth. The reactions were then analyzed on the denaturing polyacrylamide gel to obtain single-nucleotide resolution images. All gels were either dried or frozen and exposed to an imaging

screen for 24-96 hrs. The gels were imaged on a Typhoon 9410 Variable Mode Imager. Histograms were generated in ImageJ from the pixel density of the images and the area under the curves for each band was integrated. The value of all bands was added to generate a value for the total RNA, then the values obtained for each band corresponding to an extension was added. Finally, the values for extension were divided by the value for total RNA to generate a fraction extended value shown by **Equation 3.1** below. Initial rates were generated by calculating the slope of the line of the first four time points in Prism9. The approximate K_M with respect to nucleotide concentration was generated by fitting the initial rates to a Michaelis-Menten curve.

Equation 3.1

$$Fraction\ extended = \frac{(Primer_{+1}+\ldots+Primer_{+n})}{(Primer_{+1}+\ldots+Primer_{+n})+Primer}$$

Kinetic analysis of pyrophosphate inhibition of RNA primer extension.

Kinetic data for pyrophosphate inhibition was obtained by monitoring the extension of a ^{32}P labeled primer by tC19Z over time and varying concentrations of pyrophosphate. Equimolar concentrations (0.5 μM) of tC19Z, primer, and template were mixed and heated to 82 ºC for 2 minutes, then cooled slowly to 17 ºC for 10 minutes. Varying concentrations of pyrophosphate were then added to the reactions followed by buffer containing 200 mM $MgCl_2$, 50 mM Tris buffer, and 4 mM rNTPs. The reaction was then allowed to incubate at 17 ºC and 5 μL aliquots were taken at varying time points and were quenched with 250 mM EDTA and 7 M urea gel-loading buffer. The reactions were analyzed on an 18% denaturing polyacrylamide gel to obtain single nucleotide resolution images. The gels were imaged on a Typhoon 9410 Variable Mode Imager and the pixel density of the bands was analyzed using ImageJ to generate the kinetic data. Initial rates were calculated using the first four time points and used to demonstrate pyrophosphate inhibition.

Detection of ^{32}P-labeled RNA co-precipitation with pyrophosphate under reaction conditions.

^{32}P-labeled RNA hairpin, RNA template, and tC19Z were added together at concentrations of 0.5 μM each. Reactions were heated to 82 ºC for 2 minutes, cooled slowly to 17 ºC for 10 minutes, followed by addition of varying concentrations of pyrophosphate up to 10 mM. A buffer containing 200 mM $MgCl_2$, 50 mM Tris buffer, and 4 mM rNTPs was added and the reactions were incubated at 17 ºC for 24 hours. The reactions were then centrifuged at 2,000xg for 10 minutes to pellet the pyrophosphate precipitate, and the supernatant was removed. The supernatant was mixed with 250 mM EDTA and 7 M urea gel-loading buffer and the pellet was re-dissolved in 10 μL of water and mixed with the same gel-loading buffer. The reactions were then resolved on an 18% denaturing polyacrylamide gel and imaged on a Typhoon 9410 Variable Mode Imager.

Repair of cyclic phosphate damaged RNA hairpins.

Hairpin repair was visualized and quantified by imaging the extension of a α-^{32}P-ATP labeled hairpin by tC19Z. For negative controls, equimolar concentrations (0.5 μM) of tC19Z and 2′-3′ cyclic phosphate-terminated hairpin were mixed and heated to 82 ºC for 2 minutes then cooled slowly to 17 ºC for 10 minutes. Either a buffer containing 200 mM $MgCl_2$ and 50 mM Tris buffer, or 200 mM $MgCl_2$, 50 mM Tris buffer, and 4 mM rNTPs were added and the reactions were allowed to incubate at 17 ºC for 24 hours. The reactions were quenched with 250 mM EDTA and 7 M urea gel-loading buffer and were analyzed on a 15% denaturing polyacrylamide gel. For the repair reaction, equimolar concentrations (0.5 μM) of tC19Z and cyclic phosphate-terminated hairpin were mixed and heated to 82 ºC for 2 minutes then cooled slowly to 17 ºC for 10 minutes, followed by addition of 10 mM pyrophosphate. A buffer containing 200 mM $MgCl_2$, 50 mM Tris buffer, and 4 mM rNTPs was added and the reaction was allowed to incubate at 17 ºC for 24 hours. The reactions were quenched with 250 mM EDTA and 7 M urea gel-loading buffer and were analyzed on a 15% denaturing polyacrylamide gel. In order to generate a positive control,

the cyclic phosphate was removed from the hairpin by reacting with polynucleotide kinase in 1x PNK buffer for 1 hour at 37 ºC then inactivated at 65 ºC for 10 minutes. The PNK digested hairpin was then added to tC19Z, both at concentrations of 0.5 μM and heated to 82 ºC for 2 minutes then cooled slowly to 17 ºC for 10 minutes. Either a buffer containing 200 mM $MgCl_2$ and 50 mM Tris buffer, or 200 mM $MgCl_2$, 50 mM Tris buffer, and 4 mM rNTPs was added and the reactions were allowed to incubate at 17 ºC for 24 hours. The reactions were quenched with 250 mM EDTA and 7 M urea gel-loading buffer and were analyzed on a 15% denaturing polyacrylamide gel. The gels were imaged on a Typhoon 9410 Variable Mode Imager.

Kinetic analysis of cyclic phosphate damage repair.

Kinetics of extension by tC19Z of pyrophosphate-repaired hairpin, PNK-repaired hairpin, and pyrophosphate inhibited PNK-repaired hairpin were obtained under similar conditions as those described above. Either hairpin, or PNK repaired hairpin was incubated with tC19Z at 82 ºC for 2 minutes and then cooled slowly to 17 ºC for 10 minutes. Pyrophosphate was added at a concentration of 10 mM to the pyrophosphate repaired hairpin and pyrophosphate inhibited hairpin reactions. A buffer containing 200 mM $MgCl_2$, 50 mM Tris buffer, and 4 mM rNTPs was added and the reaction was allowed to incubate at 17 ºC. Time points were removed from the reactions in 5 μL aliquots and quenched with 250 mM EDTA and 7 M urea gel-loading buffer. The reactions were analyzed on a 15% denaturing polyacrylamide gel that was imaged on a Typhoon 9410 Variable Mode Imager. ImageJ was used to quantify band density and generate the kinetic data for tC19Z extension of hairpins under varying pyrophosphate and repair conditions.

$MgCl_2$ dependence of hairpin extension by tC19Z.

The Mg^{2+} dependence of the extension reaction was determined by quantifying the extension of a α-^{32}P-ATP labeled G/C hairpin by tC19Z at varying concentrations of $MgCl_2$. Equimolar concentrations (0.5 μM) of tC19Z and PNK treated hairpin were mixed and heated to 82 ºC for 2

minutes, then cooled slowly to 17 ºC for 10 minutes. Buffer containing varying concentrations of $MgCl_2$, 50 mM Tris buffer, and 4 mM rNTPs was added to the RNA and the reaction was allowed to incubate at 17 ºC for 24 hours. The reactions were quenched with 250 mM EDTA and 7 M urea gel-loading buffer and then analyzed on an 18% denaturing polyacrylamide gel. The gels were imaged on a Typhoon 9410 Variable Mode Imager.

Hairpin migration and extension under varying conditions.

To better understand hairpin migration and reactivity, an A/U hairpin was either left untreated, or was reacted with a phosphatase (5 μM hairpin, 1x cutsmart buffer, and shrimp alkaline phosphatase) for 30 mins at 37 ºC. The two conditions were then either left as is, treated with PNK as described above, or with acid (25 mM HCl) followed by quenching. These conditions were then analyzed on a 15% denaturing polyacrylamide gel. Furthermore, each treated hairpin (0.5 μM) was heated with tC19Z (0.5 μM) for 2 minutes at 85 ºC and then cooled to 17 ºC over 10 minutes. A buffer containing 200 mM $MgCl_2$, 50 mM Tris HCl and in some reactions 4 mM rNTPs was added. Additionally, 4 mM uridine and/or 10 mM pyrophosphate were added to some reactions. The mixtures were incubated at 17 ºC for 24 hours followed by quenching with 250 mM EDTA and 7 M urea gel-loading buffer. The reactions were analyzed on a 15% denaturing polyacrylamide gel that was imaged on a Typhoon 9410 Variable Mode Imager.

Incorporation of α-^{32}P-rNTPs into non-labeled, repaired hairpins.

To further demonstrate repair, an A/U hairpin was either transcribed with an α ^{32}P[ATP] or without (cold). Some of the hairpins were then treated with PNK as described above. The α-^{32}P-ATP labeled hairpin (0.5 μM) was incubated with tC19Z (0.5 μM) at 82 ºC for 2 mins and cooled to 17 ºC over 10 minutes. A buffer containing 200 mM $MgCl_2$, 50 mM Tris HCl and in some reactions 4 mM rNTPs was added, and to some reactions 10 mM pyrophosphate and/or α ^{32}P[ATP]

were added. The cold hairpin was treated in the same way. The mixtures were incubated at 17 ºC for 24 hours followed by quenching with 250 mM EDTA and 7 M urea gel-loading buffer. The reactions were analyzed on a 15% denaturing polyacrylamide gel that was imaged on a Typhoon 9410 Variable Mode Imager.

A U/A hairpin was transcribed without α-^{32}P-ATP and either left untreated or reacted with PNK as described above. The hairpin (0.5 μM) was heated with tC19Z (0.5 μM) for 2 minutes at 85 ºC and then cooled to 17 ºC over 10 minutes. A buffer containing 200 mM $MgCl_2$, 50 mM Tris HCl and, in some reactions, 10 mM pyrophosphate was added. Additionally, either α ^{32}P[ATP], α ^{32}P[UTP], or α ^{32}P[CTP] were added. The mixtures were incubated at 17 ºC for 24 hours followed by quenching with 250 mM EDTA and 7 M urea gel-loading buffer. The reactions were analyzed on a 15% denaturing polyacrylamide gel that was imaged on a Typhoon 9410 Variable Mode Imager.

Extension of mismatched RNA hairpin sequences.

Mismatched RNA hairpins were extended by tC19Z under conditions of PNK repair or pyrophosphate. Pyrophosphate enables extension of both a cyclic phosphate damaged and mismatched hairpin when 0.5 μM hairpin was added to 0.5 μM tC19Z. The reaction was then incubated at 82 ºC for 2 minutes and then cooled slowly to 17 ºC for 10 minutes. Pyrophosphate was added at a concentration of 10 mM followed by a buffer containing 200 mM $MgCl_2$, 50 mM Tris buffer, and 4 mM rNTPs and the reaction was allowed to react at 17 ºC for 24 hours. A reaction containing a PNK repaired mismatched hairpin and tC19Z at concentrations of 0.5 μM each were incubated at 82 ºC for 2 minutes and then cooled slowly to 17 ºC for 10 minutes. A buffer was then added containing 200 mM $MgCl_2$, 50 mM Tris buffer, and 4 mM rNTPs and the reaction was allowed to react at 17 ºC for 24 hours. A reaction containing a PNK repaired mismatched hairpin and tC19Z at concentrations of 0.5 μM each were incubated at 82 ºC for 2 minutes and then

cooled slowly to 17 ºC for 10 minutes. Pyrophosphate was added at a concentration of 10 mM followed by a buffer containing 200 mM $MgCl_2$, 50 mM Tris buffer, and 4 mM rNTPs and the reaction was allowed to react at 17 ºC for 24 hours. All reactions were quenched with 250 mM EDTA and 7 M urea gel-loading buffer and analyzed on a 15% denaturing polyacrylamide gel that was imaged on a Typhoon 9410 Variable Mode Imager. ImageJ was used to quantify band density to generate fraction extended graphs.

Ribozyme extension of mismatched primer/template sequences.

Primers having a terminal 3′ mismatch with the template sequence and a 5′-Cy3 fluorescent tag were ordered from IDT DNA (**Table 3.S1**). The primers (0.5 μM) were mixed with tC19Z (0.5 μM) and incubated at 82 ºC for 2 minutes and then cooled slowly to 17 ºC for 10 minutes. A buffer was then added containing 200 mM $MgCl_2$, 50 mM Tris buffer, and in some reactions 4 mM rNTPs and/or varying concentrations of pyrophosphate, and were allowed to react at 17 ºC for 24 hours. All reactions were quenched with 250 mM EDTA and 7 M urea gel-loading buffer and analyzed on a 18% denaturing polyacrylamide gel that was imaged on a Typhoon 9410 Variable Mode Imager.

Sequencing RNA hairpins.

RNA hairpins were sequenced by Illumina Sequencing to determine sequence identity. Cyclic phosphate repair and mismatch extension reactions as described above were performed and separated by polyacrylamide gel electrophoresis. The images obtained were analyzed for extension products and the corresponding bands were excised from the gel, followed by elution and ethanol precipitation. The RNA sequences were resuspended and then ligated to a DNA sequence containing an Illumina forward adaptor (**Table 3.S1**). The DNA sequence was designed with a 5′-phosphate followed by four randomized nucleotides to reduce ligation bias on the

unknown 3′-RNA end. To anneal the sequences, the DNA strand was added in approximately three times excess of the RNA hairpin (1.2 μM), as well as 10 mM Tris HCl and 10 mM NaCl. The reaction was heated to 85 ºC for 2 mins, then slowly cooled to 25 ºC over 1.5 hrs. The ligation was performed in 1X SplintR Ligase Buffer and with 15 units SplintR Ligase at 25 ºC for 1.5 hrs.

After ligation, the sequence was reverse transcribed using the RT primer in **Table 3.S1**. To the ligation reaction, 0.5 μM RT primer and 0.5 mM dNTPs were added, heated to 65 ºC for five mins., and then quickly chilled on ice. 1X First Strand Buffer and 1 mM DTT were added to the reaction and heated to 42 ºC for 2 mins. House RT, 100 units Superscript II, 100 units M-MuLV, and 5 units Bst DNA polymerase were added and the reaction was cycled through the following conditions: 25 ºC for 10 mins, 35 ºC for 10 mins, 42 ºC for 10 mins, 50 ºC for 10 mins, 60 ºC for 10 mins, and 70 ºC for 10 mins.

The resulting cDNA was PCR amplified with 0.5 μM of a forward primer with an Illumina adapter, 0.5 μM of a reverse primer and 1X Taq Master Mix (**Table 3.S1**). A second PCR reaction was performed to install the Illumina Barcodes and the reaction was gel purified. The libraries were submitted for Illumina Sequencing and the resulting data was counted and primer regions were trimmed using FastAptamer and CutAdapt.[50] Full length and extended sequences were isolated from data using FastAptamer and Python 3.0.

3.6 Supplementary Data

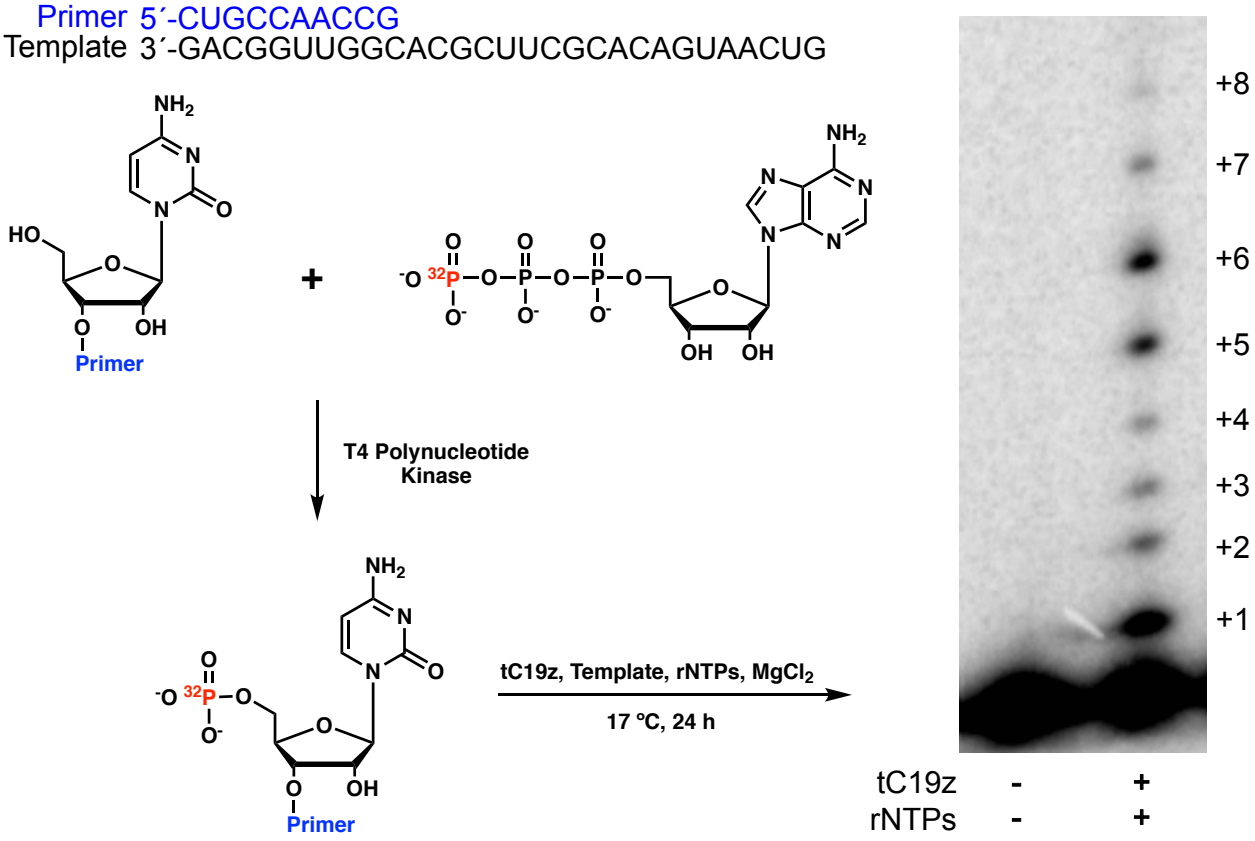


Figure 3.S1. ^{32}P labeling and ribozyme mediated extension of a primer/template sequence. A primer sequence (blue) is ^{32}P labeled (red) on the 5′-end using T4 Polynucleotide kinase. Under published conditions, the primer is then annealed to the template and tC19Z and extension occurs over 24 hours. The resulting reactions were analyzed on a high-resolution PAGE gel. A negative control shows just the ^{32}P labeled primer and the positive control shows that eight nucleotides are added to the primer sequence before the ribozyme stalls after nearing the ribozyme annealed portion of the template.

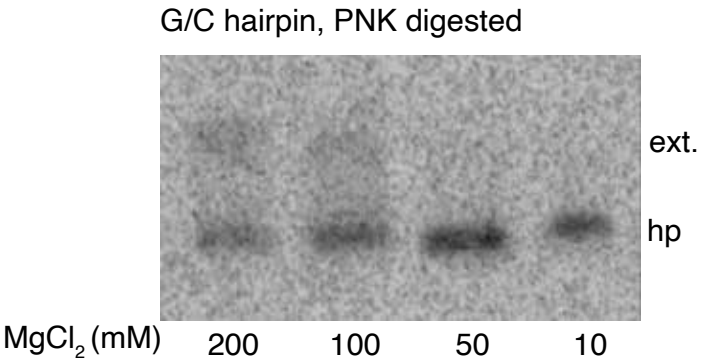


Figure 3.S2. Magnesium dependence of tC19Z. A α-^{32}P-ATP labeled G/C hairpin was treated with PNK to remove the cyclic phosphate and was allowed to react under standard conditions with varying concentrations of $MgCl_2$ for 24 hours. The resulting reactions were analyzed on a high-resolution PAGE gel. The hairpin and extension products are labeled. As magnesium concentration decreases, the extension reaction also decreases.

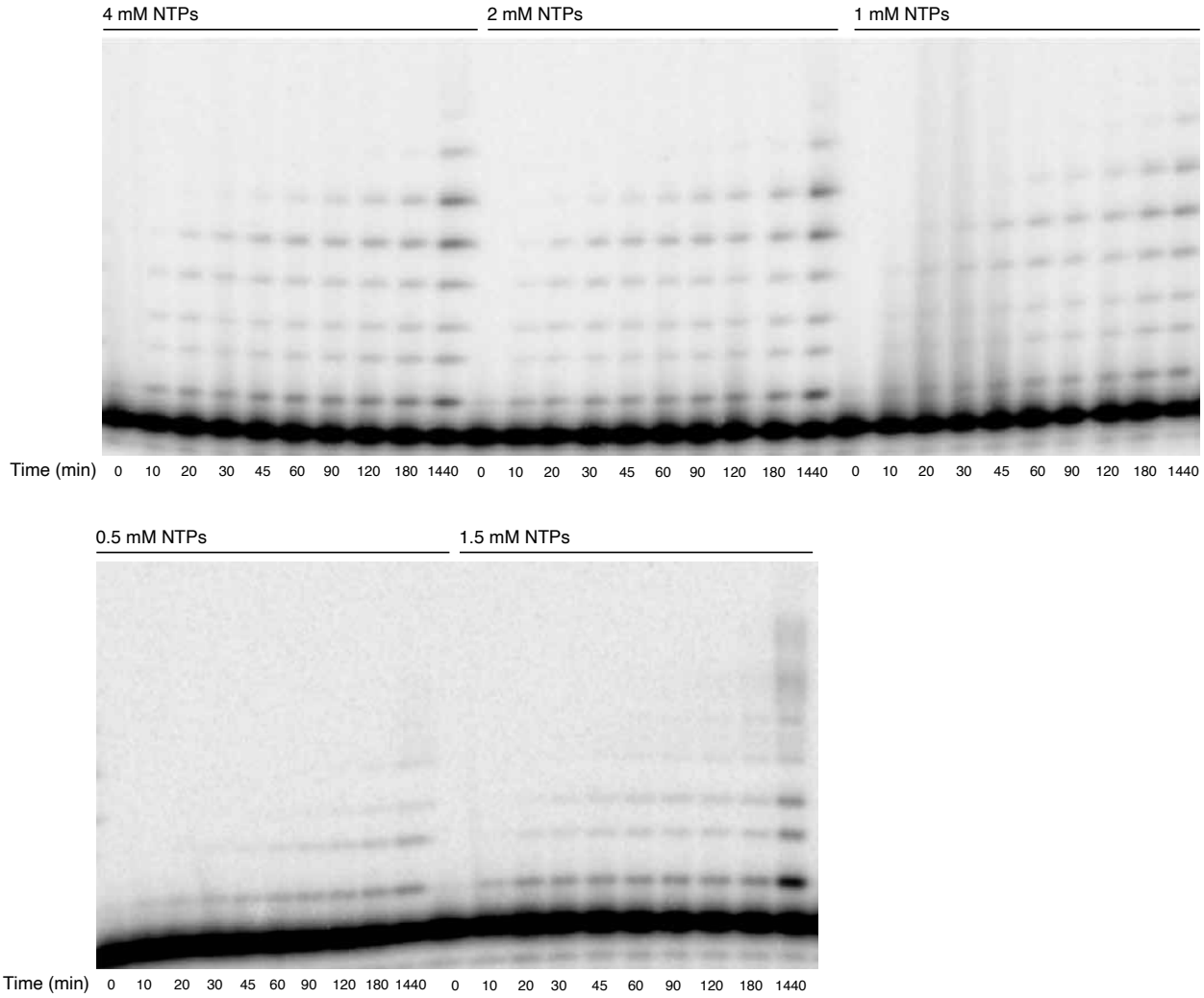


Figure 3.S3. Kinetics of tC19Z mediated extension of a primer. The primer/template sequences were reacted with tC19Z under standard conditions and variable concentrations of NTPs for up to 24 hours. Reaction aliquots were taken at ten different time points and quenched before being analyzed by high-resolution PAGE. Some representative gel images are shown here.

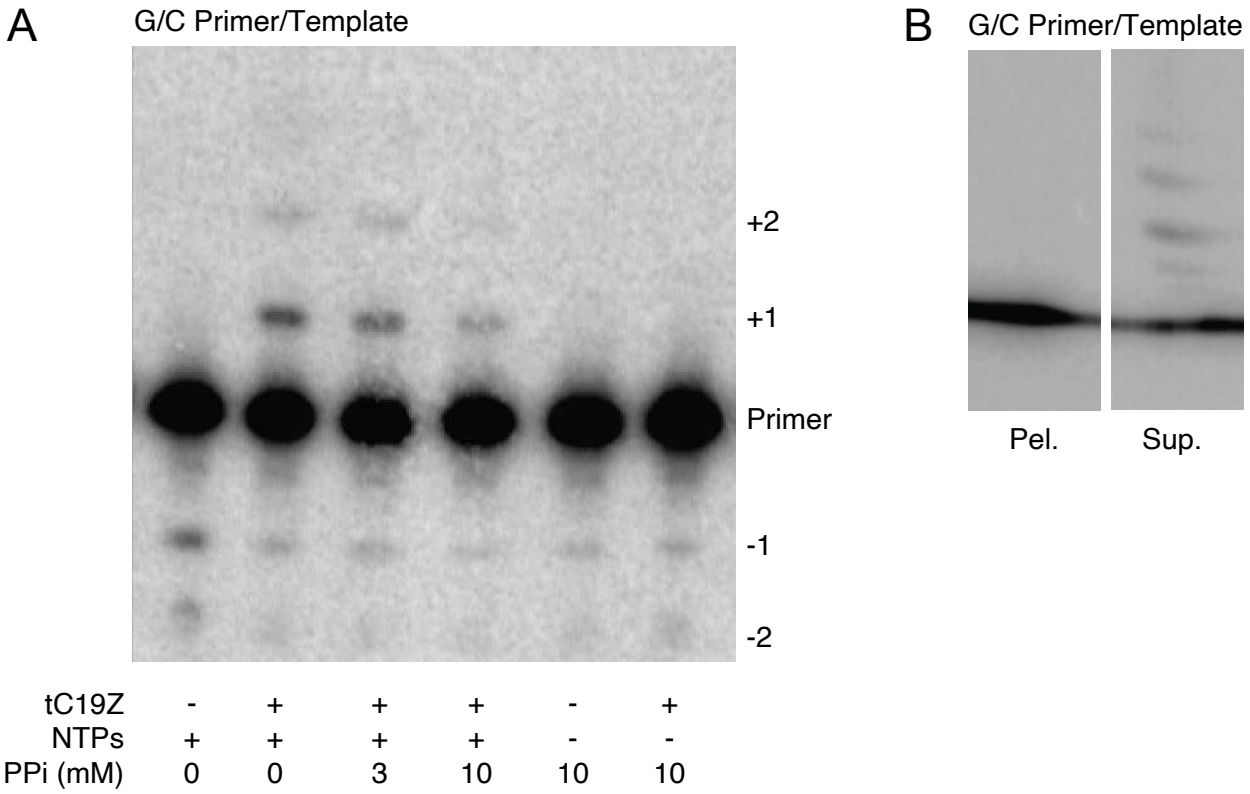


Figure 3.S4. Pyrophosphate inhibits ribozyme mediated extension and co-precipitates RNA. (A) A primer/template sequence is extended by the ribozyme under standard conditions, but the extension is inhibited when pyrophosphate is also added to the reaction as seen in lanes 2-4. However, addition of pyrophosphate either with or without tC19Z does not show any additional degradation or chewback bands. (B) Following completion of an extension reaction under standard conditions and containing 10 mM pyrophosphate, the resulting magnesium pyrophosphate precipitate was pelleted and separated from the supernatant. After resuspension and quenching, the pellet and supernatant were analyzed by high-resolution PAGE and show that although extension of the primer occurs in the supernatant, the majority of the ^{32}P labeled RNA is co-precipitating with the magnesium pyrophosphate.

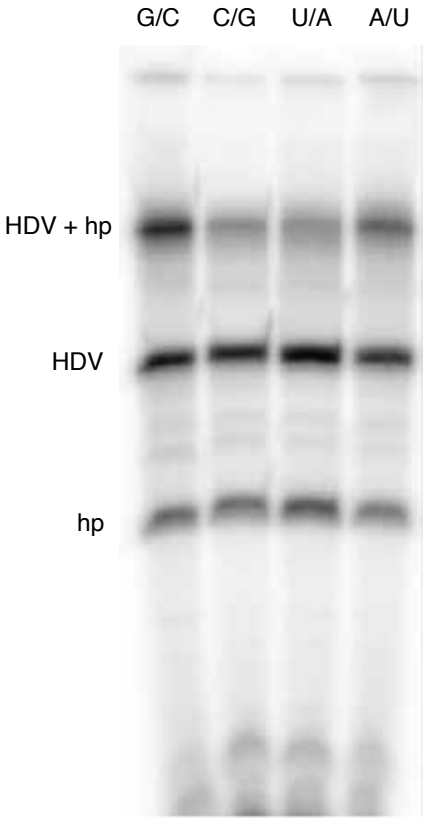


Figure 3.S5. **Purification of 2′-3′ cyclic phosphate-cleaved hairpins.** Each of the four matched, drz-Mbac-1-hairpin (HDV + hp) sequences were transcribed with an α-^{32}P-ATP and self-cleave during transcription. The transcription reactions were purified via PAGE and the hairpin (hp) band containing the 2′-3′ cyclic phosphate was excised from the gel and used in tC19Z extension reactions.

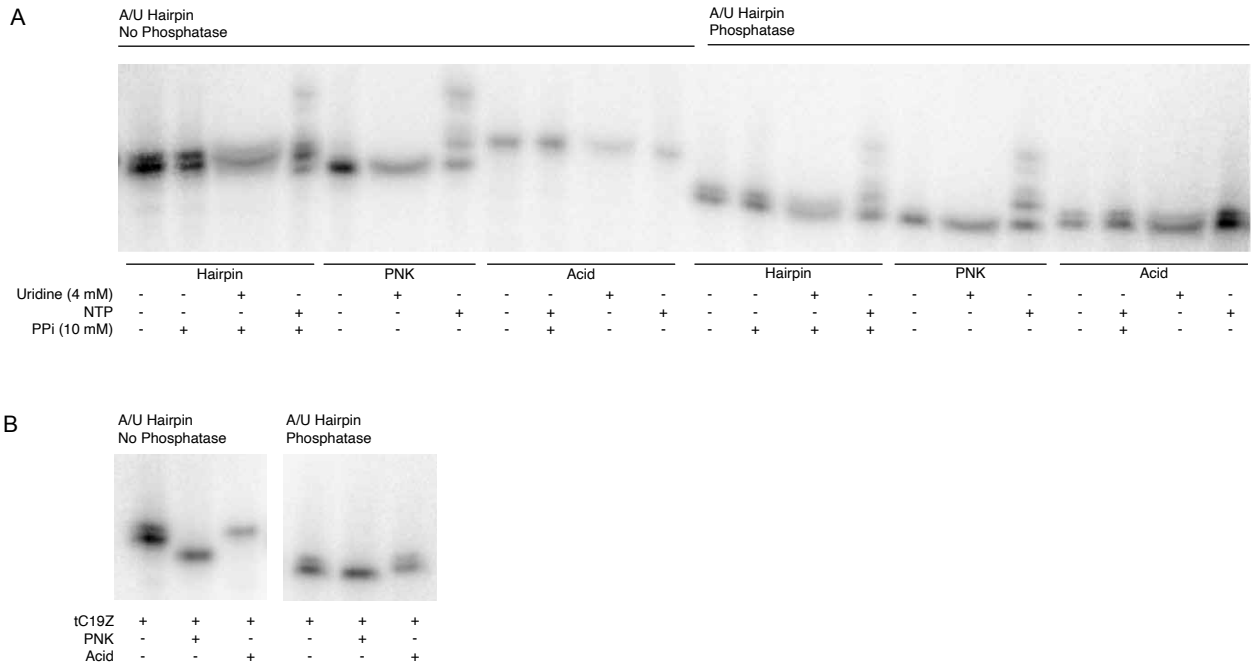


Figure 3.S6. Hairpin controls and gel shift calibration. (A) A hairpin was either subjected to conditions of phosphatase or no phosphatase, as well as digestion by PNK or acid. The resulting hairpins were reacted with tC19Z under standard reaction conditions with variations shown by the ± indications. Phosphatase removes the 5′-phosphate from the hairpin causing a downward shift in the position of the hairpin. (B) The hairpin was treated the same as above, but variations in hairpin gel shift are compared directly. Digestion with PNK (and thus removal of the cyclic phosphate) causes the hairpin to shift down on the gel. Treatment with acid (and thus opening the cyclic phosphate) causes the hairpin to shift up on the gel.

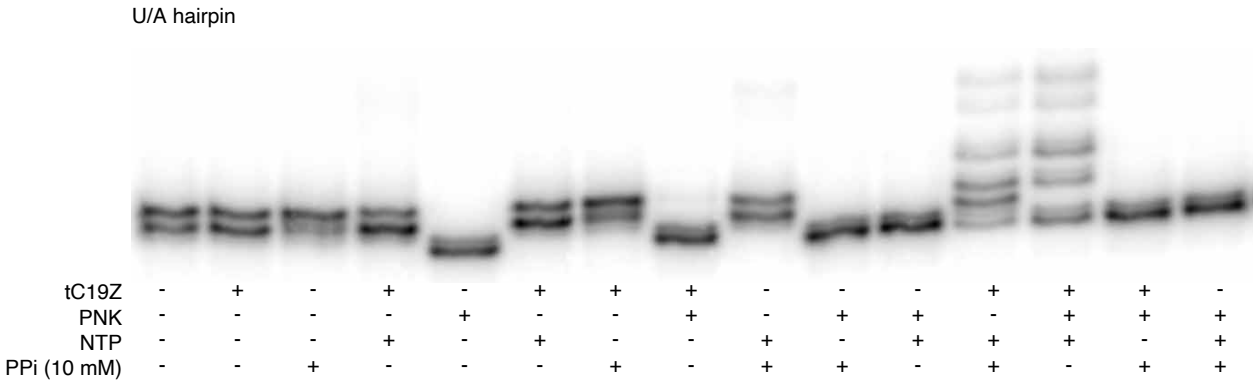


Figure 3.S7. Full hairpin extension controls. A U/A hairpin was reacted under standard extension conditions with variations shown by ± symbols above. The reactions were then analyzed by high-resolution PAGE.

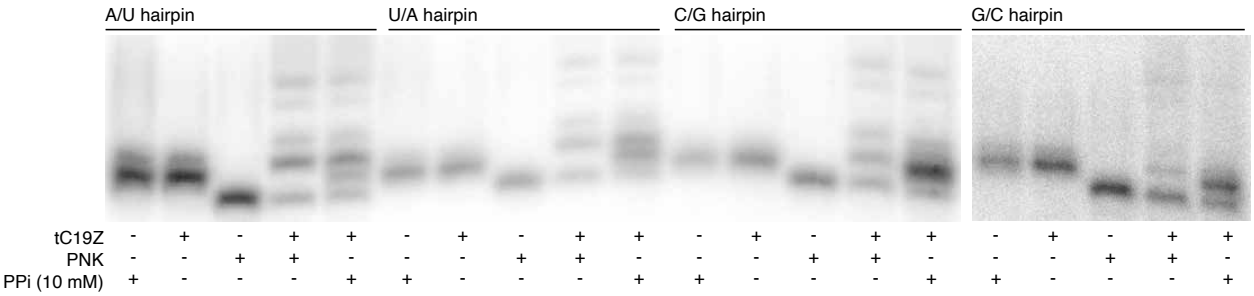


Figure 3.S8. Repair reactions and controls for all matched hairpins. All hairpins were reacted under standard conditions with variations shown by the ± symbols. Hairpins were analyzed by high-resolution PAGE. The A/U, U/A, and C/G hairpins were all run on the same gel, however, the G/C hairpin was run on a separate gel due to lower α-^{32}P-ATP incorporation efficiency.

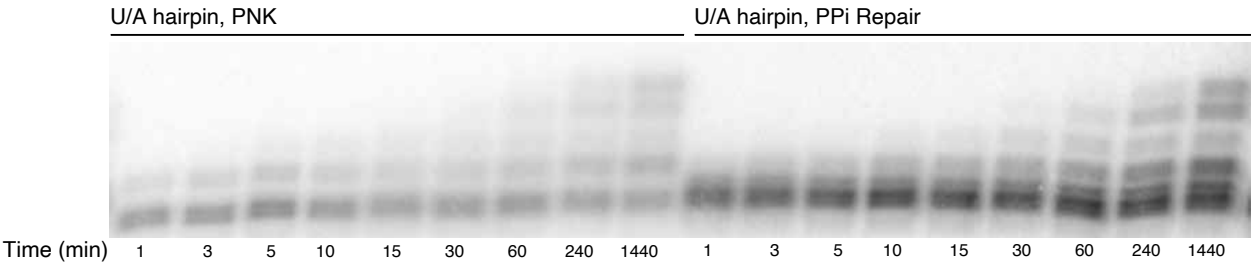


Figure 3.S9. Example time course of a U/A hairpin extension and pyrophosphate repair. An α-^{32}P-ATP U/A hairpin was either treated with PNK or pyrophosphate and then subjected to standard extension conditions. Nine time points were collected with emphasis on early time points and were analyzed by high-resolution PAGE.

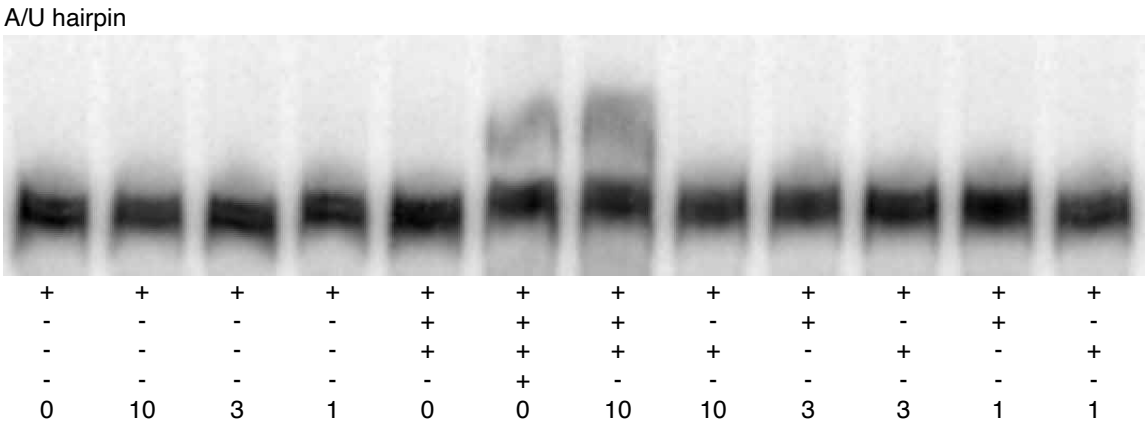


Figure 3.S10. Omission of a single nucleotide prevents extension of an A/U hairpin. An α-^{32}P-ATP labeled A/U hairpin was extended under standard reaction conditions with variations indicated by $\pm$ symbols. Lanes 1-5 are negative controls, lane 6 is a positive control, and lane 7 is the pyrophosphate repair reaction at 10 mM pyrophosphate. Lane 8 indicates that upon addition of all nucleotides except ATP, indicated by the nucleotide nomenclature (B)TP, there is no extension. Lanes 9-12 demonstrate that the repair does not occur at lower concentrations of pyrophosphate.

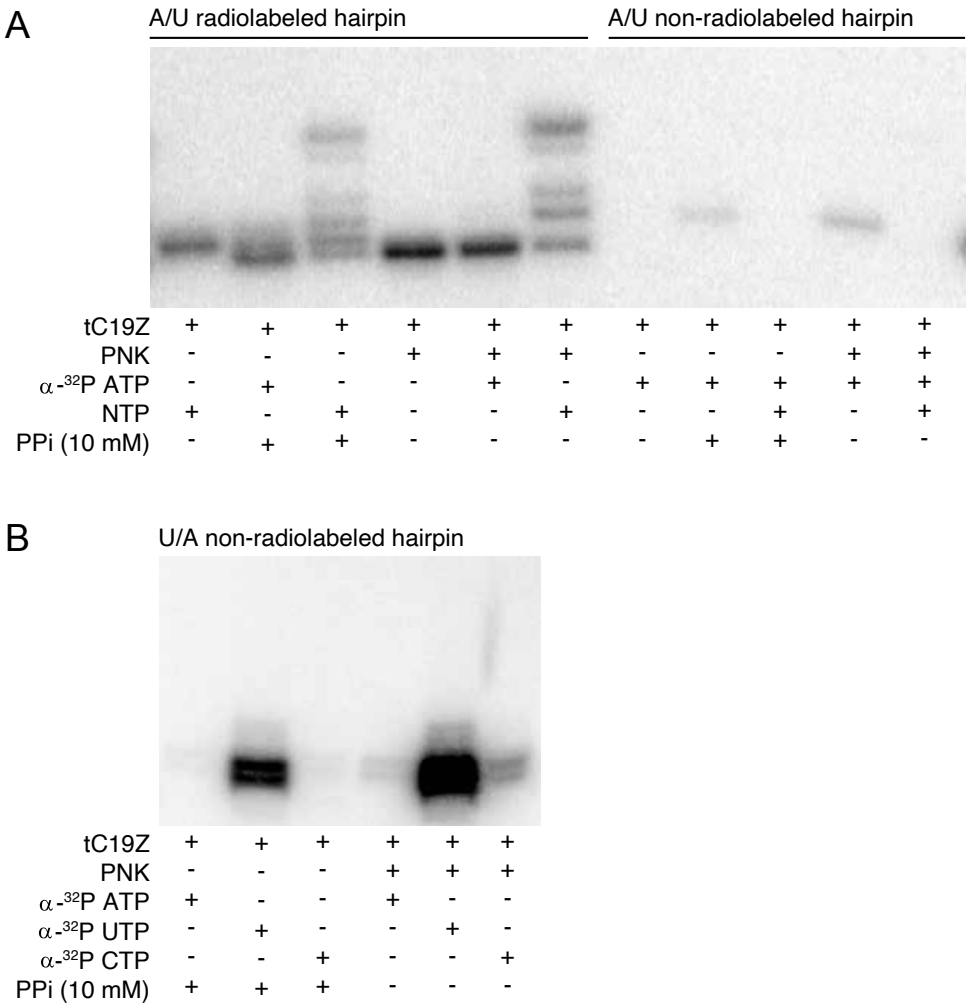


Figure 3.S11. α-^{32}P-rNTPs can be incorporated into repaired, non-labeled hairpins. (A) An A/U hairpin was either transcribed and purified with α-^{32}P-ATP, or without. The hairpin was then subjected to repair and extension conditions, largely consistent with standard reaction conditions. The exception being the addition of α-^{32}P-ATP to some reactions. In the case of the non-labeled hairpin, the α-^{32}P-ATP is only incorporated if the sequence has been repaired. (B) A U/A hairpin was transcribed without α-^{32}P-ATP and was subjected to standard extension conditions, however, an α-^{32}P-rNTP was added in place of the standard NTPs. Different α-^{32}P-rNTPs are incorporated into the repaired hairpin with different efficiencies as shown by the varying intensities of the bands. UTP is incorporated most efficiently because the first nucleotide on the template sequence is an

adenosine. However, misincorporation is seen of the other nucleotides and a second nucleotide appears to be added in the case of the addition of UTP.

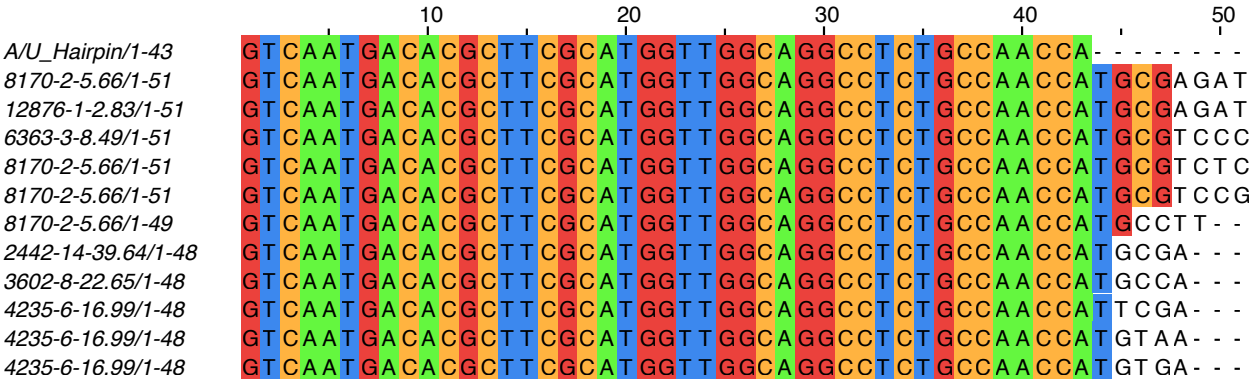


Figure 3.S12. An A/U hairpin is repaired and extended as seen by Illumina sequencing. The sequence of the A/U hairpin is shown at the top of the block in the 5′ to 3′ direction. Sequences underneath the A/U hairpin have been isolated from processed Illumina sequencing data. Colored bases extending past the A/U hairpin sequence are repair and extensions (of varying lengths) by the tC19Z ribozyme in the presence of pyrophosphate. Bases in white are four random nucleotides originating from the DNA sequence ligated to the 3′ end of the hairpin when preparing for sequencing.

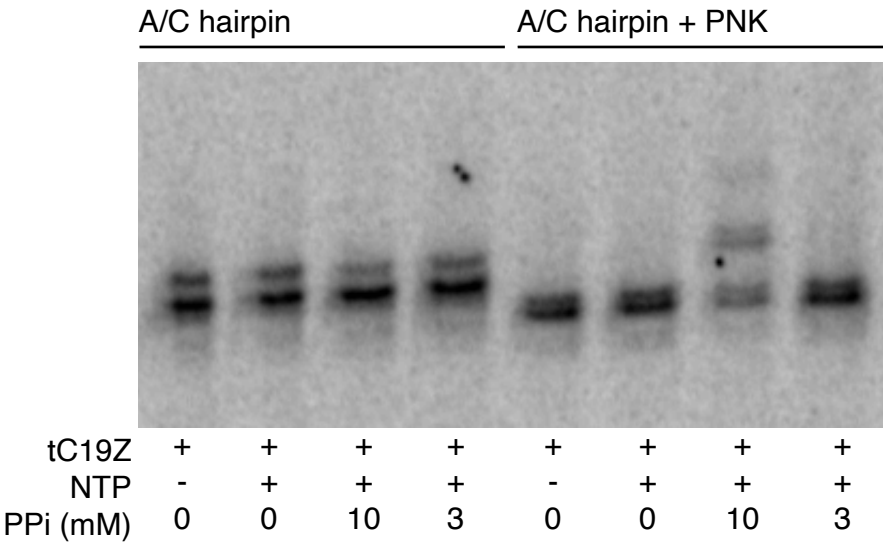
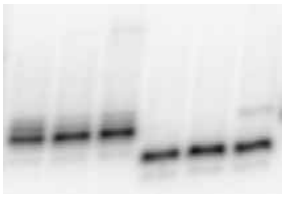


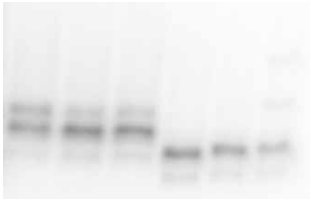
Figure 3.S13. An A/C mismatch is only extended following treatment with both PNK and pyrophosphate. An α-^{32}P-ATP A/C mismatch was treated with PNK and then subjected to standard reaction conditions and two pyrophosphate concentrations. No extension of the hairpin is seen in lanes 1-4 prior to removal of the cyclic phosphate damage, even when pyrophosphate is added. Extension is also not seen after digestion with PNK (lane 6), and it is only after both digestion with PNK and addition of pyrophosphate (lane 7) that the ribozyme is able to extend the hairpin.

A/A hairpin



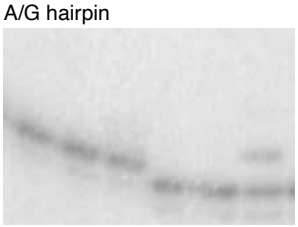
tC19Z	+	+	+	-	+	+
PNK	-	-	-	+	+	+
NTP	-	+	+	+	+	+
PPi (10 mM)	+	-	+	+	-	+

G/G hairpin



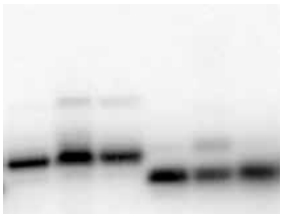
tC19Z	+	+	+	-	+	+
PNK	-	-	-	+	+	+
NTP	-	+	+	+	+	+
PPi (10 mM)	+	-	+	+	-	+

A/G hairpin



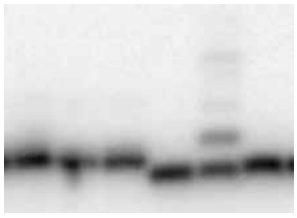
tC19Z	+	+	+	-	+	+
PNK	-	-	-	+	+	+
NTP	-	+	+	+	+	+
PPi (10 mM)	+	-	+	+	-	+

G/A hairpin



tC19Z	+	+	+	-	+	+
PNK	-	-	-	+	+	+
NTP	-	+	+	+	+	+
PPi (10 mM)	+	+	+	+	+	-

C/A hairpin



tC19Z	+	+	+	-	+	+
PNK	-	-	-	+	+	+
NTP	-	+	+	+	+	+
PPi (10 mM)	+	-	+	+	+	-

C/C hairpin



tC19Z	+	+	+	-	+	+
PNK	-	-	-	+	+	+
NTP	-	+	+	+	+	+
PPi (10 mM)	+	-	+	+	-	+

U/C hairpin



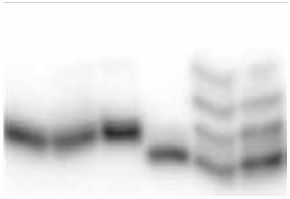
tC19Z	+	+	+	-	+	+
PNK	-	-	-	+	+	+
NTP	-	+	+	+	+	+
PPi (10 mM)	+	-	+	+	-	+

C/U hairpin



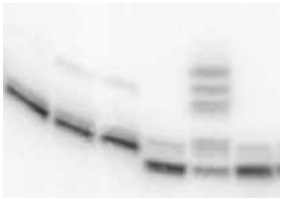
tC19Z	+	+	+	-	+	+
PNK	-	-	-	+	+	+
NTP	-	+	+	+	+	+
PPi (10 mM)	+	-	+	+	-	+

U/U hairpin



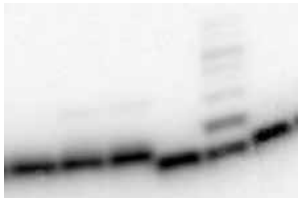
tC19Z	+	+	+	-	+	+
PNK	-	-	-	+	+	+
NTP	-	+	+	+	+	+
PPi (10 mM)	+	-	+	+	-	+

U/G hairpin



tC19Z	+	+	+	-	+	+
PNK	-	-	-	+	+	+
NTP	-	+	+	+	+	+
PPi (10 mM)	+	+	+	+	+	-

G/U hairpin



tC19Z	+	+	+	-	+	+
PNK	-	-	-	+	+	+
NTP	-	+	+	+	+	+
PPi (10 mM)	+	+	+	+	+	-

Figure 3.S14. Controls and repair of all mismatch combinations. All mismatches were α-^{32}P-ATP labeled and reacted under standard conditions with variations indicated with ± symbols. The reactions were then analyzed by high-resolution PAGE.

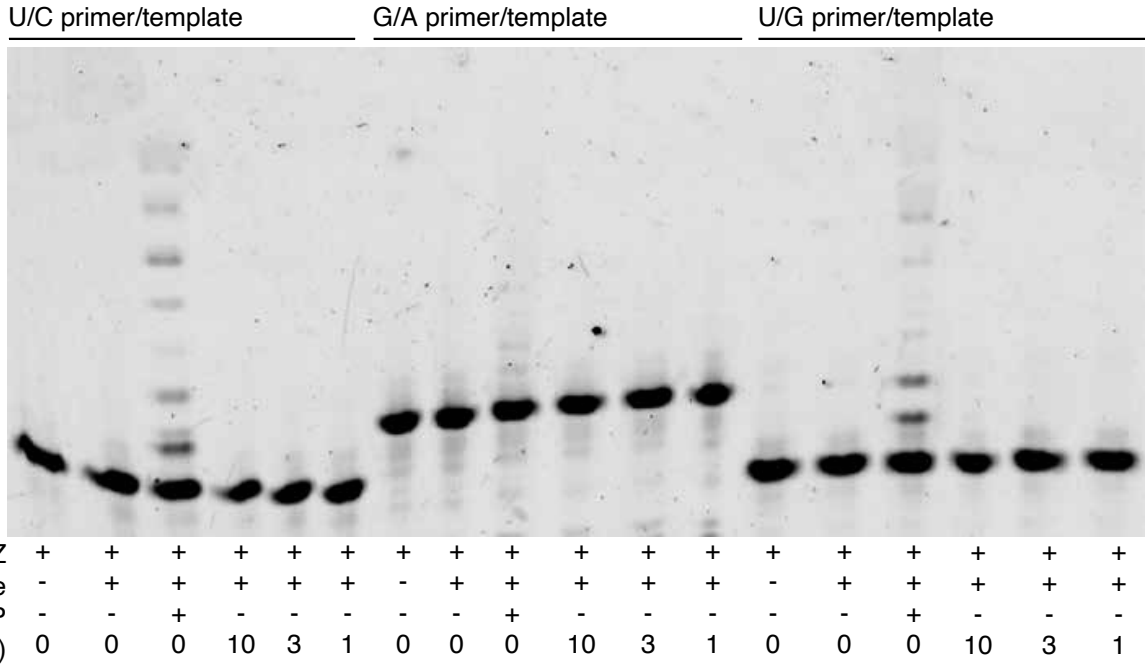


Figure 3.S15. Selected fluorescently labeled primer/template sequences show varying extension efficiencies. Fluorescently labeled primers (5′-Cy3) with terminal mismatches were incubated with the template and tC19Z under standard reaction conditions and varying concentrations of pyrophosphate. The primers were analyzed by high-resolution PAGE. The third lane of each primer sequence shows the extension efficiency of that mismatch by tC19Z. The U/C and G/A mismatches follow similar extension efficiency trends as the corresponding PNK repaired hairpins.

Table 3.S1. Chapter 3 DNA and RNA sequences.

1	tC19Z Ribozyme	GCTAATACGACTCACTATAGTCATTGAAAAAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCT TCGGATGCAGAGGAGGCAGCCTTCGGTGGCGCGATAGCGCCAACGTTCTCAACAGACACCCAAT ACTCCCGCTTCGGCGGGTGGGGATAACACCTGACGAAAAGGCGATGTTAGACACGCCCAGGTCA TAATCCCCGGAGCTTCGGCTCC
2	RNA Primer	CUGCCAACCG
3	RNA Template	GUCAAUGACACGCUUCGCACGGUUGGCAG
4	tC19Z Cloning Fwd.	ATAGGATCCGCTAATACGACTCACTATAGTCATTG
5	tC19Z Cloning Rev.	TATGAATTCGGTCTCGGGAGCCGAAGCTCCGGGGATTATGAC
6	drz-Mbac-1 Rev.	ATACGTATAACTATTAGCATGACCCATCGTCAACCACGAGGAGGTTGTTATACTATGACGAC
7	G/C Hairpin Fwd.	GCTAATACGACTCACTATAGTCAATGACACGCTTCGCACGGTTGGCAGGCCTCTGCCAACCGGTC GTCATAGTATAACAACC
8	C/G Hairpin Fwd.	GCTAATACGACTCACTATAGTCAATGACACGCTTCGCAGGGTTGGCAGGCCTCTGCCAACCCGTC GTCATAGTATAACAACC
9	A/U Hairpin Fwd.	GCTAATACGACTCACTATAGTCAATGACACGCTTCGCATGGTTGGCAGGCCTCTGCCAACCAGTC GTCATAGTATAACAACC
10	U/A Hairpin Fwd.	GCTAATACGACTCACTATAGTCAATGACACGCTTCGCAAGGTTGGCAGGCCTCTGCCAACCTGTC GTCATAGTATAACAACC
11	G/G Hairpin Fwd.	GCTAATACGACTCACTATAGTCAATGACACGCTTCGCAGGGTTGGCAGGCCTCTGCCAACCGGT CGTCATAGTATAACAACC
12	C/C Hairpin Fwd.	GCTAATACGACTCACTATAGTCAATGACACGCTTCGCACGGTTGGCAGGCCTCTGCCAACCCGTC GTCATAGTATAACAACC
13	A/A Hairpin Fwd.	GCTAATACGACTCACTATAGTCAATGACACGCTTCGCAAGGTTGGCAGGCCTCTGCCAACCAGTC GTCATAGTATAACAACC
14	U/U Hairpin Fwd.	GCTAATACGACTCACTATAGTCAATGACACGCTTCGCATGGTTGGCAGGCCTCTGCCAACCTGTC GTCATAGTATAACAACC
15	G/A Hairpin Fwd.	GCTAATACGACTCACTATAGTCAATGACACGCTTCGCAAGGTTGGCAGGCCTCTGCCAACCGGTC GTCATAGTATAACAACC
16	G/U Hairpin Fwd.	GCTAATACGACTCACTATAGTCAATGACACGCTTCGCATGGTTGGCAGGCCTCTGCCAACCGGTC GTCATAGTATAACAACC
17	C/A Hairpin Fwd.	GCTAATACGACTCACTATAGTCAATGACACGCTTCGCAAGGTTGGCAGGCCTCTGCCAACCCGTC GTCATAGTATAACAACC
18	C/U Hairpin Fwd.	GCTAATACGACTCACTATAGTCAATGACACGCTTCGCATGGTTGGCAGGCCTCTGCCAACCCGTC GTCATAGTATAACAACC
19	A/G Hairpin Fwd.	GCTAATACGACTCACTATAGTCAATGACACGCTTCGCAGGGTTGGCAGGCCTCTGCCAACCAGTC GTCATAGTATAACAACC
20	A/C Hairpin Fwd.	GCTAATACGACTCACTATAGTCAATGACACGCTTCGCACGGTTGGCAGGCCTCTGCCAACCAGTC GTCATAGTATAACAACC
21	U/C Hairpin Fwd.	GCTAATACGACTCACTATAGTCAATGACACGCTTCGCACGGTTGGCAGGCCTCTGCCAACCTGTC GTCATAGTATAACAACC
22	U/G Hairpin Fwd.	GCTAATACGACTCACTATAGTCAATGACACGCTTCGCAGGGTTGGCAGGCCTCTGCCAACCTGTC GTCATAGTATAACAACC
23	Ligation Primer for Seq.	5Phos/NNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATC/3AmMO
24	RT Primer for Seq.	GATCTACACTCTTTCCCTACACG
25	Illumina adaptor Fwd.	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCGTCAATGACACGCTTCGCATGGTT
26	TrueSeq Universal Fwd.	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
27	Illumina Barcode Rev.	GATCGGAAGAGCACACGTCTGAACTCCAGTCACNNNNNNATCTCGTATGCCGTCTTCTGCTTG
28	Fluorescent U/C Primer	Cy3-CUGCCAACCU
29	Fluorescent G/A Primer	Cy3-CUGCCAACCGG
30	Fluorescent U/G Primer	Cy3-CUGCCAACU

3.7 References

1. Leslie, E. O. Prebiotic chemistry and the origin of the RNA World. *Crit. Rev. Biochem. Mol. Biol.* **2004**, 39, 99-123.

2. Robertson, M. P.; Joyce, G. F. The origins of the RNA World. *Cold Spring Harb. Perspect. Biol.* **2012**, 13, 391-403.

3. Schrum, J. P.; Zhu, T. F.; Szostak, J. W. The origins of cellular life. *Cold Spring Harb. Perspect. Biol.* **2010**, 2, a002212.

4. Bartel, D. P.; Unrau, P. J.; Constructing an RNA World. *Trends Cell Biol.* **1999**, 12, 9-13.

5. Joyce, G. F. The antiquity of RNA-based evolution. *Nature* **2002**, 418, 214-221.

6. Ma, W.; Yu, C. Intramolecular RNA replicase: Possibly the first self-replicating molecule in the RNA world. *Orig. Life Evol. Biosph.* **2006**, 36, 413-420.

7. Chen, X.; Lin, N.; Ellington, A. Ribozyme catalysis of metabolism in the RNA World. *Chem. Biodivers.* **2007**, 633-655.

8. Benner, S. A.; Ellington, A. D. Tauer, A. Modern metabolism as a palimpsest of the RNA World. *Proc. Natl. Acad. Sci. USA* **1989**, 86, 7054-7058.

9. Neveu, M.; Kim, H. J.; Benner, S. A. The “strong” RNA World hypothesis: Fifty years old. *Astrobiology* **2013**, 13, 391-403.

10. Ekland, E. H.; Szostak, J. W. Bartel, D. P. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science* **1995**, 269, 364.

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

12. Ekland, E. H.; Bartel, D. P. The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res.* **1995**, 23, 3231-3238.

13. Jaeger, L.; Wright, M. C.; Joyce, G. F. A complex ligase ribozyme evolved in vitro from a group I ribozyme domain. *Proc. Natl. Acad. Sci. USA* **1999**, 96,14712.

14. Shechner, D. M; Grant, R. A.; Bagby, S. C.; Koldobskaya, Y.; Piccirilli, J. A. Bartel, D. P. Crystal structure of the catalytic core of an RNA-polymerase ribozyme. *Science* **2009**, 326, 1271-1275.

15. Ekland, E. H.; Bartel, D. P. RNA-catalyzed RNA polymerization using nucleoside triphosphates. *Nature* **1996**, 382, 373.

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

17. Samanta, B.; Joyce, G. F. A reverse transcriptase ribozyme. *eLife* **2017**, 6, e31153.

18. Wochner, A.; Attwater, J.; Coulson, A.; Holliger, P. Ribozyme-catalyzed transcription of an active ribozyme. *Science* **2011** 332, 209.

19. Horning, D. P.; Joyce, G. F. Amplification of RNA by an RNA polymerase ribozyme. *Proc. Natl. Acad. Sci. USA* **2016**, 113, 9786-9791.

20. Horning, D. P.; Bala, S.; Chaput, J. C.; Joyce, G. F. RNA-catalyzed polymerization of deoxyribose, threose, and arabinose nucleic acids. *ACS Synth. Biol.* **2019**, 8, 955-961.

21. Tjhung, K. F.; Shokhirev, M. N.; Horning, D. P.; Joyce, G. F. An RNA polymerase ribozyme that synthesizes its own ancestor. *Proc. Natl. Acad. Sci. USA* **2020**, 117, 2906-2913.

22. Zaher, H.; Unrau, P. Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA* **2007**, 13, 1017-1026.

23. Attwater, J.; Tagami, S.; Kimoto, M.; Butler, K.; Kool, E. T.; Wengel, J.; Herdewijn, P.; Hirao, I.; Holliger, P. Chemical fidelity of an RNA polymerase ribozyme. *Chem. Sci.* **2013**, 4, 2804-2817.

24. Schwer, B.; Sawaya, R.; Ho, C. K.; Shuman, S. Portability and fidelity of RNA-repair systems. *Proc. Natl. Acad. Sci. USA* **2004**, 101, 2788-2793.

25. Burroughs, A. M.; Aravind, L. RNA damage in biological conflicts and the diversity of responding RNA repair systems. *Nucleic Acids Res.* **2016**, 44, 8525-8555.

26. Nandakumar, J.; Schwer, B.; Schaffrath, R.; Shuman, S. RNA repair: An antidote to cytotoxic eukaryal RNA damage. *Mol. Cell* **2008**, 31, 278-286.

27. Kamerlin, S. C.; Sharma, P. K.; Prasad, R. B.; Warshel, A. Why nature really chose phosphate. *Q. Rev. Biophys.* **2013**, 46, 1-132.

28. Pasek, M. A.; Rethinking early Earth phosphorus geochemistry. *Proc. Natl. Acad. Sci. USA* **2008**, 105, 853.

29. Miller, S. L.; Parris, M. Synthesis of pyrophosphate under primitive earth conditions. *Nature* **1964**, 204, 1248.

30. Horder, M. Complex formation of inorganic pyrophosphate with magnesium: the influence of ionic strength, supporting medium and temperature. *Biochim. Biophys. Acta. Enzymol.* **1974**, 358, 319-328.

31. Holm, H. G.; Baltscheffsky, H. Links between hydrothermal environments, pyrophosphate, Na+, and early evolution. *Orig. Life Evol. Biosph.* **2011**, 41, 483-493.

32. CRC Handbook of Chemistry and Physics, 102nd ed.; Rumble, J., Ed.; CRC Press: Boca Raton, FL, **2021**; pp 8.

33. Jimenez, R. M.; Polanco, J. A.; Lupták, A. Chemistry and biology of self-cleaving ribozymes. *Trends Biochem. Sci.* **2015**, 40, 648-661.

34. Soukup, G. A.; Breaker, R. R. Relationship between internucleotide linkage geometry and the stability of RNA. *RNA* **1999**, 5, 1308-1325.

35. Li, Y.; Breaker, R. R. Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2′-hydroxyl group. *J. Am. Chem. Soc.* **1999**, 121, 5364-5372.

36. Riccitelli, N. J.; Delwart, E.; Lupták, A. Identification of minimal HDV-like ribozymes with unique divalent metal ion dependence in the human microbiome. *Biochemistry* **2014**, 53, 1616-1626.

37. Wang, L. K.; Lima, C. D.; Shuman, S. Structure and mechanism of T4 polynucleotide kinase: an RNA repair enzyme. *EMBO J.* **2002**, 21, 3873-3880.

38. Das, U.; Shuman, S. Mechanism of RNA 2′,3′-cyclic phosphate end healing by T4 polynucleotide kinase-phosphatase. *Nucleic Acids Res.* **2013**, 41, 355-365.

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

40. Roth, A.; Weinberg, Z.; Chen, A. G. Y.; Kim, P. B.; Ames, T. D. Breaker, R. R. A widespread self-cleaving ribozyme class is revealed by bioinformatics. *Nat. Chem. Biol.* **2014**, 10, 56-60.

41. Pinto, P. H.; Kroupova, A.; Schleiffer, A.; Mechtler, K.; Jinek, M.; Weitzer, S.; Martinez, J. ANGEL2 is a member of the CCR4 family of deadenylases with 2′,3′-cyclic phosphatase activity. *Science* **2020**, 369, 524-530.

42. Li, G. M. Mechanisms and functions of DNA mismatch repair. *Cell Res.* **2007**, 18, 85-98.

43. Iyer, R. R.; Pluciennik, A.; Burdett, V.; Modrich, P. L. DNA mismatch repair: functions and mechanisms. *Chem. Rev.* **2006**, 106, 302-323.

44. Cruse, W. B. T.; Saludjian, P.; Biala, E.; Strazewski, P.; Prange, T.; Kennard, O. Structure of a mispaired RNA double helix at 1.6 Å resolution and implications for the prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. USA* **1994**, 91, 4160-4164.

45. Pan, B.; Mitra, S. N.; Sundaralingam, M. Crystal structure of an RNA 16-mer duplex R(GCAGAGUUAAAUCUGC)$_2$ with nonadjacent G(Syn)A$^+$(Anti) mispairs. *Biochemistry* **1999**, 38, 2826-2831.

46. Sugimoto, N.; Kierzek, R.; Turner, D. H. Sequence dependence for the energetics of terminal mismatches in ribooligonucleotides. *Biochemistry* **1987**, 26, 4559-4562.

47. Giese, M. R.; Betschart, K.; Dale, T.; Riley, C. K.; Rowan, C.; Sprouse, K. J.; Serra, M. J. *Biochemistry* **1997**, 37, 1094-1100.

48. Sydow, J. F.; Brueckner, F.; Cheung, A. C. M.; Damsma, G. E.; Dengl, S.; Lehmann, E.; Vassylyev, D.; Cramer, P. Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol. Cell* **2009**, 34, 710-721.

49. Kahn, J. D.; Hearst, J. E. Reversibility of nucleotide incorporation by *Escherichia coli* RNA polymerase, and its effect on fidelity. *J. Mol. Biol.* **1988**, 205, 291-314.

50. Alam, K. K.; Chang, J. L.; Burke, D. H. FASTAptamer: A bioinformatic toolkit for high-throughput sequence analysis of combinatorial selections. *Mol. Ther. Nucleic Acids* **2015**, 4, e230.

APPENDIX A: TEMPLATE DIRECTED, PRIMER-FREE POLYMERIZATION OF NUCLEOSIDE TRIPHOSPHATES BY A POLYMERASE RIBOZYME

A.1 Introduction

Fundamentally, the transmission of genetic information depends heavily on Watson-Crick base-pairing between nucleosides on a template and incoming nucleoside triphosphates.[1,2] Following recognition, polymerization can occur between adjacent bases. In contemporary biological systems, this reaction is carried out by complex, protein-based polymerases.[3-5] However, in an evolving RNA world, these intricate functionalities would likely not have been available.[6-8] A few different approaches have been taken toward generating RNA in a presumed prebiotic world, beginning in its most basic form with *de novo* RNA synthesis, which largely relies on a range of activated nucleosides to achieve polymerization.[9-12] This could then have evolved into templated, non-enzymatic copying of RNA, which has been demonstrated and requires primers and/or activated nucleosides.[13-19]

Another approach to RNA replication relies on ribozymes acting as polymerases that can extend primers along templates with varying efficiencies.[20-25] These approaches require primers to initiate polymerization, and although modern biology uses enzymes such as primases to generate primers for replication, it was likely that in an early RNA world, there was no dedicated machinery for their production.[26] Thus, there are few bridges linking nonenzymatic polymerization with activated nucleosides, and primer-dependent RNA catalyzed polymerization. To begin to address this gap, we report the first instance of primer-free polymerization of nucleoside triphosphates along a template, catalyzed by a polymerase ribozyme.

A.2 Results and Discussion

Polymerization of rNTPs along a template using γ-^{32}P[ATP] as the means for visualization.

The polymerase ribozyme tC19Z was evolved to bind to a template and then extend a primer along that template (Chapter 3).[23] We hypothesized that because tC19Z base-pairs with

the template, effectively holding it in place, it may not require the primer to initiate polymerization. To test this, we began by performing a polymerization reaction under standard conditions (200 mM $MgCl_2$, 50 mM Tris HCl, 4 mM NTPs, 0.5 μM tC19Z, and 0.5 μM template strand) with the addition of γ-^{32}P[ATP] to track any potential polymerization. The ribozyme and template were mixed and heated to 82 ºC for 2 minutes before cooling to 17 ºC over 10 minutes. The remaining buffers and reagents were added and the reaction was kept at 17 ºC for 24 hours. We chose to test with γ-^{32}P[ATP] because of the position of ^{32}P on the γ-phosphate. With the ^{32}P-label at this position, the product of polymerization can only be visualized if adenosine is the first nucleotide incorporated in the sequence and becomes the 5`-end of the replication (**Figure A.1A**).

We did observe polymerization beginning with ATP and then extending at least two and possibly up to four nucleotides (**Figure A.1B**). This process is both template dependent and ribozyme dependent, since the absence of either prevents any polymerization from occurring. Additionally, supplementing the reaction with 10 mM pyrophosphate also inhibits polymerization. We found these results to be somewhat surprising because the experimental design resulted in the concentrations of other nucleoside triphosphates in the reaction substantially higher than that of the γ-^{32}P[ATP]. Thus, the visualization of the polymerization stems from a single ^{32}P-radiolabel at the 5`-end of the growing sequence and is accumulating in high enough concentrations for detection after only 24 hours, indicating this may be a robust functionality of the tC19Z ribozyme.

Polymerization of rNTPs along a template using α-^{32}P[UTP] as the means for visualization.

After visualizing primer-free polymerization with γ-^{32}P[ATP], we decided to test the polymerization with α-^{32}P[UTP] because we found previously that tC19Z has a preference for incorporating UTP over the other nucleoside triphosphates (Chapter 3). In this experiment, we used 200 mM $MgCl_2$, 50 mM Tris HCl, 0.5 μM tC19Z, 0.5 μM template strand, and 0.5 mM NTPs, as well as α-^{32}P[UTP]. We used a lower concentration of NTPs in order to allow for more incorporation of the ^{32}P-labeled UTP, and thus better visualization of the reaction. We also allowed the

reaction to incubate at 17 ºC for 72 hours rather than 24 to increase the accumulation of polymerization products. Once again, we were able to visualize primer-free polymerization of nucleotides, this time with more distinct products indicating polymerization of 4-5 nucleotides (**Figure A.1C**).

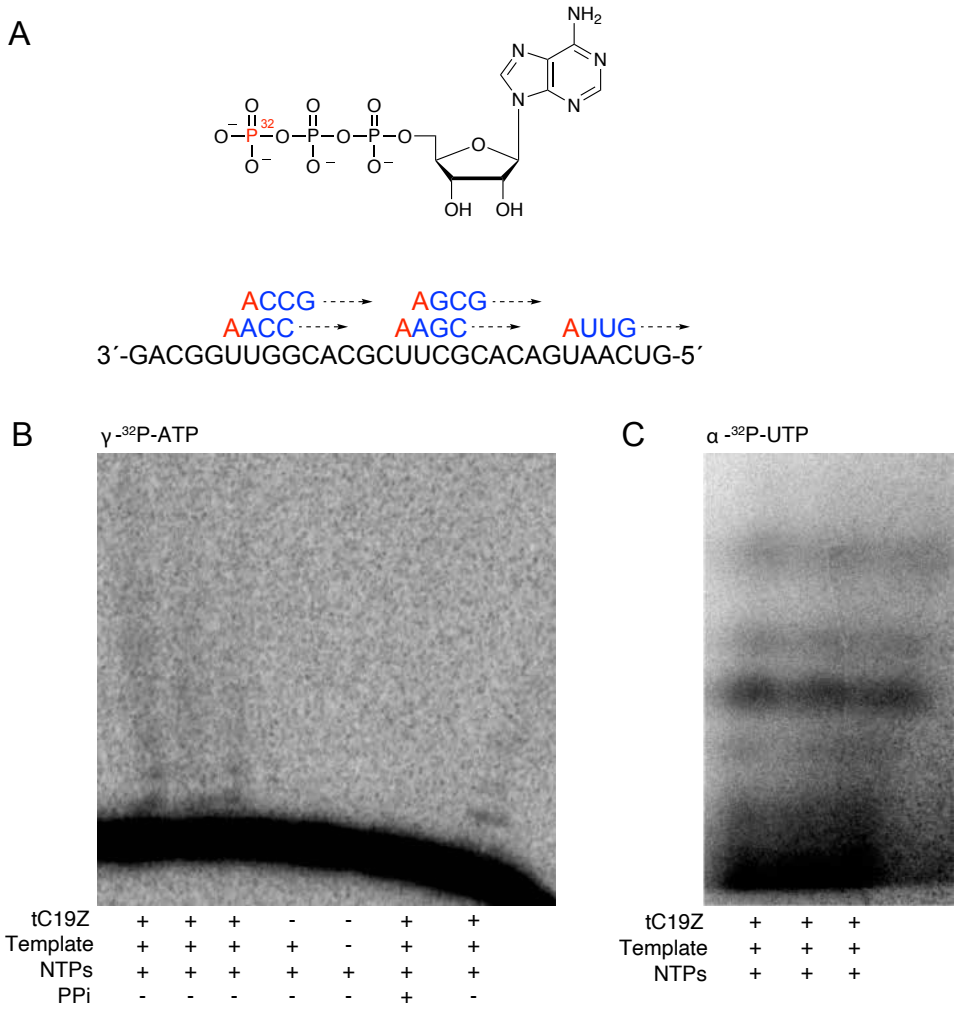


Figure A.1. Ribozyme mediated, primer-free polymerization along a template. (A) Adenosine triphosphate is shown with the ^{32}P-label on the γ-phosphate. Because the polymerization must begin with ATP to be seen on the gel, all possible places on the template where the polymerization could occur are shown, with the starting A shown in red and subsequent nucleotides shown in blue. (B) Polymerization of nucleoside triphosphates along a template by tC19Z is shown. The lowermost band is γ-^{32}P[ATP] and any subsequent upward shifts are polymerization products.

Polymerization is not seen in controls lacking the template or ribozyme, as well as upon addition of pyrophosphate. (C) Polymerization of nucleoside triphosphates along a template by tC19Z is shown with α-^{32}P[UTP] acting as the label and shown as the bottom most band.

These results taken in concert indicate that primer-free polymerization of nucleotides along a template is a functionality of the tC19Z ribozyme. This is a reactivity that has not yet been seen with polymerase ribozymes and could provide an important link in the evolution of information transfer in an RNA world. Additional experiments are necessary to further understand the limits of reactivity and characterize the kinetics. Additionally, these small RNA products should be sequenced to concretely demonstrate that they are templated polymerizations, as well as to gain information on variation of sequences produced.

A.3 Materials and Methods

Polymerization of rNTPs along a template using γ-^{32}P[ATP].

Reactions were performed in 10 μL volumes with γ-^{32}P[ATP] and with the concentration of the reagents when present as follows: 200 mM $MgCl_2$, 50 mM Tris HCl, 4 mM NTPs, 0.5 μM tC19Z, and 0.5 μM template strand. Either the template strand, tC19Z, or both (**Table A.1**) were added to a tube and heated to 82 ºC for 2 minutes before cooling to 17 ºC over 10 minutes. To one reaction was added 10 mM pyrophosphate. The remaining buffers and reagents were added and the reaction was kept at 17 ºC for 24 hours. The reactions were quenched by addition of 250 mM ethylenediaminetetraacetic acid (EDTA) and 7 M urea gel-loading buffer. Samples were analyzed on a high-resolution polyacrylamide gel (18% acrylamide/bis 19:1, 7 M urea, 0.5x Tris/Borate/EDTA buffer). The gel was then dried and imaged on a Typhoon 9410 Variable Mode Imager.

Polymerization of rNTPs along a template using α-^{32}P[UTP].

Reactions were performed in 10 µL volumes with α-^{32}P[UTP] and with the concentration of reagents as follows: 200 mM $MgCl_2$, 50 mM Tris HCl, 0.5 mM NTPs, 0.5 µM tC19Z, and 0.5 µM template strand. The template strand and tC19Z were added to a tube and heated to 82 ºC for 2 minutes before cooling to 17 ºC over 10 minutes. The remaining buffers and reagents were added and the reaction was kept at 17 ºC for 72 hours. The reactions were quenched by addition of 250 mM EDTA and 7 M urea gel loading buffer. Samples were analyzed on a high-resolution polyacrylamide gel. The gel was then frozen and imaged on a Typhoon 9410 Variable Mode Imager.

Table A.1. Appendix A sequences.

1	tC19Z Ribozyme	GCTAATACGACTCACTATAGTCATTGAAAAAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCT TCGGATGCAGAGGAGGCAGCCTTCGGTGGCGCGATAGCGCCAACGTTCTCAACAGACACCCAAT ACTCCCGCTTCGGCGGGTGGGGATAACACCTGACGAAAAGGCGATGTTAGACACGCCCAGGTCA TAATCCCCGGAGCTTCGGCTCC
2	RNA Template	GUCAAUGACACGCUUCGCACGGUUGGCAG
3	tC19Z Fwd.	GCTAATACGACTCACTATAGTCATTG
4	tC19Z Rev.	GGAGCCGAAGCTCCGGGGATTATGAC

A.4 References

1. Kervio, E.; Claasen, B.; Steiner, U. E.; Richert, C. The strength of the template effect attracting nucleosides to naked DNA. *Nucleic Acids Res.* **2014**, 42, 7409-7420.

2. Izgu, E. C.; Fahrenbach, A. C.; Zhang, N.; Li, L.; Zhang, W.; Larsen, A. T.; Blain, J. C.; Szostak, J. W. Uncovering the thermodynamics of monomer binding for RNA replication. *J. Am. Chem. Soc.* **2015**, 137, 6373-6382.

3. Lieberman, K. R.; Dahl, J. M.; Mai, A. H.; Cox, A.; Akeson, M.; Wang, H. Kinetic mechanism of translocation and dNTP binding in individual DNA polymerase complexes. *J. Am. Chem. Soc.* **2013**, 135, 9149-9155.

4. Genna, V.; Vidossich, ; Ippoliti, E.; Carloni, P.; De Vivo, M. A self-activated mechanism for nucleic acid polymerization catalyzed by DNA/RNA polymerases. *J. Am. Chem. Soc.* **2016**, 138, 14592-14598.

5. Genna,V.; Donati, E.; De Vivo, M. The catalytic mechanism of DNA and RNA polymerases. *ACS Catal.* **2018**, 8, 11103-11118.

6. Orgel, L. E.; Lohrmann, R. Prebiotic chemistry and nucleic acid replication. *Acc. Chem. Res.* **1974**, 7, 370-377.

7. Leslie, E. O. Prebiotic chemistry and the origin of the RNA World. *Crit. Rev. Biochem. Mol. Biol.* **2004**, 39, 99-123.

8. Bartel, D. P.; Unrau, P. J.; Constructing an RNA World. *Trends Cell Biol.* **1999**, 12, 9-13.

9. Jheeta, S.; Joshi, P. C. Prebiotic RNA synthesis by montmorillonite catalysis. *Life* **2014** 4, 318-330.

10. Ferris, J. P. Montmorillonite catalysis of 30-50 mer oligonucleotides: laboratory demonstration of potential steps in the origin of the RNA world. *Orig. Life Evol. Biosph.* **2002**, 32, 311-332.

11. Miyakawa, S.; Joshi, P. C.; Gaffey, M. J.; Gonzalez-Toril, E.; Hyland, C.; Ross, T.; Rybij, K.; Ferris, J. P. Studies in the mineral and salt-catalyzed formation of RNA oligomers. *Orig. Life Evol. Biosph.* **2006**, 36, 343.

12. Costanzo, G.; Saladino, R.; Botta, G.; Giorgi, A.; Scipioni, A.; Pino, S.; Di Mauro, E. Generation of RNA molecules by a base-catalysed click-like reaction. *ChemBioChem* **2012**, 13, 999-1008.

13. Kim, S. C.; O'Flaherty, D. K.; Giurgiu, C.; Zhou, L.; Szostak, J. W. The emergence of RNA from the heterogeneous products of prebiotic nucleotide synthesis. *J. Am. Chem. Soc.* **2021**, 143, 3267-3279.

14. Li, L.; Prywes, N.; Tam, C. P.; O'Flaherty, D. K.; Lelyveld, V. S.; Izgu, E. C.; Pal, A.; Szostak, J. W. Enhanced nonenzymatic RNA copying with 2-aminoimidazole activated nucleotides. *J. Am. Chem. Soc.* **2017**, 139, 1810-1813.

15. Wu, T.; Orgel, L. E. Nonenzymatic template-directed synthesis on oligodeoxycytidylate sequences in hairpin oligonucleotides. *J. Am. Chem. Soc.* **1992**, 114, 317-322.

16. Prywes, N.; Blain, J. C.; Del Frate, F.; Szostak, J. W. Nonenzymatic copying of RNA templates containing all four letters is catalyzed by activated oligonucleotides. *eLife* **2016**, 5, e17756.

17. Lohrmann, R.; Bridson, P. K.; Orgel, L. E. Efficient metal-ion catalyzed template-directed oligonucleotide synthesis. *Science* **1990**, 208, 1464-1465.

18. Weimann, B. J.; Lorhmann, R.; Orgel, L. E.; Schneider-Bernloehr, H.; Sulston, J. E. Template-directed synthesis with adenosine-5`-phosphorimidazolide. *Science* **1968**, 161, 387.

19. Walton, T.; Szostak, J. W. A highly reactive imidazolium-bridged dinucleotide intermediate in nonenzymatic RNA primer extension. *J. Am. Chem. Soc.* **2016**, 138, 11996-12002.

20. Ekland, E. H.; Bartel, D. P. RNA-catalyzed RNA polymerization using nucleoside triphosphates. *Nature* **1996**, 382, 373.

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

22. Samanta, B.; Joyce, G. F. A reverse transcriptase ribozyme. *eLife* **2017**, 6, e31153.
23. Wochner, A.; Attwater, J.; Coulson, A.; Holliger, P. Ribozyme-catalyzed transcription of an active ribozyme. *Science* **2011** 332, 209.

24. Horning, D. P.; Joyce, G. F. Amplification of RNA by an RNA polymerase ribozyme. *Proc. Natl. Acad. Sci. USA* **2016**, 113, 9786-9791.

25. Tjhung, K. F.; Shokhirev, M. N.; Horning, D. P.; Joyce, G. F. An RNA polymerase ribozyme that synthesizes its own ancestor. *Proc. Natl. Acad. Sci. USA* **2020**, 117, 2906-2913.

26. Frick, D. N.; Richardson, C. C. DNA Primases. *Annu. Rev. Biochem.* **2001**, 70, 39-80.

APPENDIX B: USING NUCLEOSIDE DIPHOSPHATES AS SUBSTRATES FOR POLYMERIZATION

B.1 Introduction

The prevalence of phosphate in the origin of life on Earth derives from its use in the backbone of RNA, and potential earlier precursors.[1-4] Phosphate utilization is inextricably coupled to replication, and it is therefore plausible that a phosphate metabolism would have been one of the first chemical transformations to be encoded in early genomes, potentially providing one of the first phenotypes of early cells.[5,6] However, relatively little has been done to delineate the energy metabolism requirements of early phosphate-based reactions. Specifically, there has been a lack of investigation into the role of phosphate energetics in replication, repair, and related prebiotic processes.[7-9] Although nucleoside triphosphates predominate the energy metabolism landscape in cells today, a lower energy counterpart, such as nucleoside diphosphates, could have been a more stable precursor in an RNA world.[10,11]

Our previous work has demonstrated that DNA polymerases can use nucleoside diphosphates as substrates for replication.[12,13] Analysis of polymerization products showed that the rate of DNA synthesis from dNDPs is a little over an order of magnitude lower than from dNTPs, and data support a transition state model that is energetically very similar for di- or tri-phosphorylated nucleosides. This indicates that polymerases do not fundamentally require triphosphorylated substrates for DNA synthesis. Because of the success with DNA polymerases, we hypothesized that both protein-based and nucleic acid-based RNA polymerases may also be able to utilize diphosphorylated substrates for transcription and replication respectively.

B.2 Results and Discussion

T7 RNA polymerase can use rNDPs to transcribe a ribozyme.

Recently it has been reported that *E. coli* RNA polymerase can use rNDPs in transcription.[14] We decided to test if T7 RNA polymerase was also capable of using nucleoside diphosphates in

transcription. To this end, we decided to use an HDV-like ribozyme, Fprau-2, because a certain level of fidelity of replication must be maintained in order to visualize the self-cleaving activity.[15] We transcribed Fprau-2 with rNDPs, then used γ-^{32}P[ATP] to label the 5`-end of the transcripts for visualization on a denaturing polyacrylamide gel. Varying time points were taken from the transcription reaction (**Figure B.1A**). A band corresponding to the full-length sequence is easily visualized on the gel and the amount of transcribed RNA increases over time. However, although the full-length sequence is present, the bands are largely missing (except at the longest time point) that would indicate the self-cleaving activity is present. This implies that perhaps there is misincorporation during transcription of one or more of the nucleobases critical to maintaining self-cleaving activity.[16,17]

Since guanine is important to the initiation of transcription, we hypothesized that we could increase transcription products by the addition of a low concentration of GTP (10 μM).[18] Indeed, upon addition of GTP, we did visualize an increase in transcription products (**Figure B.1B**). Addition of exogenous GTP also rescued the ribozyme self-cleaving activity and bands are present on the gel that correspond to the cleavage products. Once again, time points were taken from the reaction, although some variability in the ^{32}P-labeling is seen on the denaturing polyacrylamide gel between time points.

A substantial amount work needs to be done to fully characterize the reactivity, kinetics, and fidelity when using rNDPs as substrates for transcription. Sequencing experiments would have to be performed to confirm the assumptions made about the possible lower fidelity that resulted in the absence of ribozyme activity. Since the lack of self-cleaving activity was seen when only rNDPs are present in the transcription, it may suggest that one reason for the current dependence on higher energy substrates (i.e., nucleoside triphosphates) was to improve fidelity of replication.

tC19Z (a polymerase ribozyme) can use rNDPs in replication.

While polymerase ribozymes have been shown to incorporate a range of modified nucleotides, they have never been shown to use nucleoside diphosphates.[19,20] We demonstrated transcription was possible using lower energy substrates, so we wanted to test a polymerase ribozyme as well. To that end, we used the polymerase ribozyme tC19Z and standard reaction conditions (200 mM $MgCl_2$, 50 mM Tris HCl, 0.5 μM ribozyme, 0.5 μM template, and 0.5 μM primer) as well as 4 mM rNDPs.[21] The polymerization was allowed to react at 17 ºC for 48 hours before analysis using high-resolution polyacrylamide PAGE. (**Figure B.1C)** A positive control was performed with 4 mM rNTPs and addition of rNDPs did not visibly inhibit the reaction. When only provided with rNDPs, the ribozyme can add at least two nucleotides to the primer sequence. The ribozyme has difficulty incorporating guanosine (Chapter 2 and Chapter 3), which may account for the stalling after adding two nucleotides. If provided with a template that did not require incorporation of guanosine, longer extension products may be possible. Though we have demonstrated that tC19Z can utilize nucleoside diphosphates to extend a primer along a template, more experiments must be performed to gain information on kinetics, energetics, and fidelity of replication.

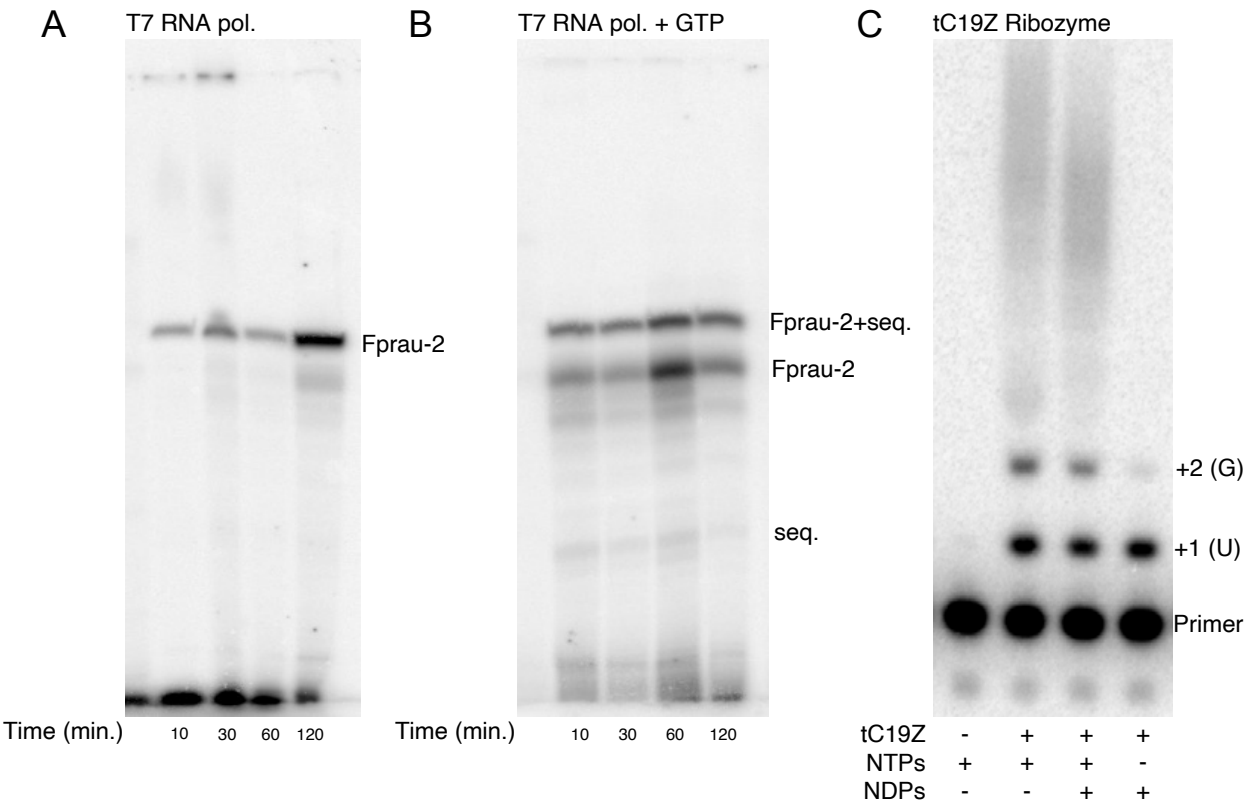


Figure B.1. T7 RNA polymerase and tC19Z can use rNDPs as substrates for polymerization. (A) T7 RNA polymerase was used to transcribe a self-cleaving ribozyme (Fprau-2) with rNDPs and visualized by 5`-end labeling with γ-^{32}P[ATP]. Aliquots were removed at 10, 30, 60, and 120 minutes and the amount of full-length ribozyme band is seen increasing over time. (B) The same transcription reaction was performed as in (A), but the reaction was supplemented with 10 μM GTP. Aliquots were taken at 10, 30, 60, and 120 minutes and the full-length sequences and products from ribozyme self-cleavage are seen. (C) A polymerase ribozyme (tC19Z) was used to extend a primer along a template. Lane 1 is a negative control showing no extension without the presence of the ribozyme. Lane 2 is a positive control showing that the primer is extended along the template in the presence of rNTPs. Lane 3 is a control showing that addition of rNDPs to the reaction does not significantly affect the primer extension. Lane 4 shows that in the presence of only rNDPs, two nucleotides are incorporated.

B.3 Materials and Methods

T7 polymerase uses rNDPs or rNDPs supplemented with GTP in transcription.

Transcription was performed under the following conditions: 50 mM Tris HCl, 2 mM spermidine, 12 mM $MgCl_2$, 8 mM DTT, 2 mM each rNDP, and 0.01% Triton X-100 at a total volume of 50 µL (for reactions in **Figure B.1B**, 10 µM GTP was added to the transcription). T7 RNA polymerase was added to the reaction and allowed to react at 37 ºC. 10 µL aliquots were removed from the reaction at 10, 30, 60, and 120 minutes and were desalted with G10 Sephadex. To each time point was added 1x Cutsmart buffer and 1 unit of shrimp alkaline phosphatase which were reacted at 37 ºC for 30 minutes followed by reaction at 65 ºC for 5 minutes. 5`-end, ^{32}P-labeling was done by adding 0.5 µL of γ-^{32}P[ATP], 0.5 mM DTT, and 20 units of T4 polynucleotide kinase (PNK) and reacting at 37 ºC for 2 hrs. The resulting mixture was again desalted with G10 Sephadex and mixed with 7 M urea gel-loading dye. The reactions were analyzed on a 10% denaturing polyacrylamide gel (10% acrylamide/bis 19:1, 7 M urea, 0.5x Tris/Borate/EDTA buffer) and imaged on a Typhoon 9410 Variable Mode Imager.

tC19Z uses rNDPs in polymerization.

An RNA primer 10 nucleotides in length was added at a concentration of 5 µM to a reaction containing γ-^{32}P[ATP], 1x polynucleotide kinase buffer, and 10 units polynucleotide kinase to add a ^{32}P label to the 5`-end of the sequence. The reaction was incubated at 37 ºC for 2 hours and subsequently purified on a 12% denaturing polyacrylamide gel, followed by elution and ethanol precipitation. Equimolar concentrations (0.5 uM) of tC19Z, ^{32}P-labeled primer, and template were mixed and heated to 82 ºC for 2 minutes, then cooled slowly to 17 ºC for 10 minutes. Buffer containing 200 mM $MgCl_2$, 50 mM Tris buffer, and either 4 mM rNTPs, 4 mM rNDPs, or both was added to the RNA and the reaction was incubated at 17 ºC for 24 hours. Reactions were quenched with 250 mM EDTA and 7 M urea gel-loading loading buffer and analyzed on an 18% denaturing polyacrylamide gel (18% acrylamide/bis 19:1, 7 M urea, 0.5x

Tris/Borate/EDTA buffer). The gel was frozen and then imaged on a Typhoon 9410 Variable Mode Imager.

Table B.1. Appendix B sequences.

1	tC19Z Ribozyme	GCTAATACGACTCACTATAGTCATTGAAAAAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATCT TCGGATGCAGAGGAGGCAGCCTTCGGTGGCGCGATAGCGCCAACGTTCTCAACAGACACCCAAT ACTCCCGCTTCGGCGGGTGGGGATAACACCTGACGAAAAGGCGATGTTAGACACGCCCAGGTCA TAATCCCCGGAGCTTCGGCTCC
2	RNA Primer	CUGCCAACCG
3	RNA Template	GUCAAUGACACGCUUCGCACGGUUGGCAG
4	drz-Fprau-2	TTCCCGCGAAATTAATACGACTCACTATAGGGGAGATATAAGGCTGTGGGAAGGTTCTCACAGCT TTGCCGCTCGAAACTTTGCACACCTCTACGCGGTGGGTGGCAGGCAACACATGAGAATGTGGCT GAGATGCAATATTGTC
5	drz-Fprau-2 Fwd.	TTCCCGCGAAATTAATACGACTCACTA
6	drz-Fprau-2 Rev.	ATACGTATAACTATTAGCATGACCCATCGTCAACCACGAGGAGGTTGTTATACTATGACGAC

B.4 References

1. Leslie, E. O. Prebiotic chemistry and the origin of the RNA World. *Crit. Rev. Biochem. Mol. Biol.* **2004**, 39, 99-123.

2. Robertson, M. P.; Joyce, G. F. The origins of the RNA World. *Cold Spring Harb. Perspect. Biol.* **2012**, 13, 391-403.

3. Schrum, J. P.; Zhu, T. F.; Szostak, J. W. The origins of cellular life. *Cold Spring Harb. Perspect. Biol.* **2010**, 2, a002212.

4. Bartel, D. P.; Unrau, P. J.; Constructing an RNA World. *Trends Cell Biol.* **1999**, 12, 9-13.

5. Chen, X.; Lin, N.; Ellington, A. Ribozyme catalysis of metabolism in the RNA World. *Chem. Biodivers.* **2007**, 633-655.

6. Benner, S. A.; Ellington, A. D. Tauer, A. Modern metabolism as a palimpsest of the RNA World. *Proc. Natl. Acad. Sci. USA* **1989**, 86, 7054-7058.

7. Kamerlin, S. C.; Sharma, P.K.; Prasad, R. B.; Warshel, A. Why nature really chose phosphate. *Q. Rev. Biophys.* **2013**, 46, 1-132.

8. Arrhenius, G.; Sales, B.; Mojzsis, S.; Lee, T. Entropy and charge in molecular evolution-the case of phosphate. *J. Theor. Biol.* **1997**, 187, 495-501.

9. Baltscheffsky, H. Major “anastrophes” in the origin and early evolution of biological energy conversion. *J. Theor. Biol.* **1997**, 187, 495-501.

10. Robertson, M. P.; Joyce, G. F. The origins of the RNA world. *Cold Spring Harbor Laboratory Press* **2012**, 4, a003608.

11. Deamer, D.; Weber, A. L. Bioenergetics and life’s origins. *Cold Spring Harb. Pespect. Biol.* **2010**, 2, a004929.

12. Burke, C. R.; Lupták, A. DNA synthesis from diphosphate substrates by DNA polymerases. *Proc. Natl. Acad. Sci. USA* **2018**, 115, 980.

13. Garforth, S. J.; Parniak, M. A.; Prasad, V. R. Utilization of a deoxynucleoside diphosphate substrate by HIV reverse transcriptase. *Plos One* **2008**, 3, e2074.

14. Gottesman, M. E.; Mustaev, A. Ribonucleoside-5`-diphosphates (NDPs) support RNA polymerase transcription, suggesting NDPs may have been substrates for primordial nucleic acid biosynthesis. *J. Biol. Chem.* **2019**, 294, 11785-11792.

15. Passalacqua, L. F. M.; Jimenez, R. M.; Fong, J. Y.; Lupták, A. Allosteric modulation of the Faecalibacterium Prausnitzii hepatitis delta virus-like ribozyme by glucosamine 6-phosphate: the substrate of the adjacent gene product. *Biochemistry* **2017**, 56, 6006-6014.

16. Olson, A. C.; Patro, J. N.; Urban, M.; Kuchta, R. D. The energetic difference between synthesis of correct and incorrect base pairs accounts for highly accurate DNA replication. *J. Am. Chem. Soc.* **2013**, 135, 1205-1208.

17. Oertell, K.; Harcourt, E. M.; Mohsen, M. G.; Petruska, J.; Kool, E. T.; Goodman, M. F. Kinetic selection vs. free energy of DNA base pairing in control of polymerase fidelity. *Proc. Natl. Acad. Sci. USA* **2016**, 113, 2277-2285.

18. Kuzmine, I.; Gottlieb, P. A.; Martin, C. T. Binding of the priming nucleotide in the initiation of transcription by T7 RNA polymerase. *J. Biol. Chem.* **2003**, 278, 2819-2823.

19. Samanta, B.; Joyce, G. F. A reverse transcriptase ribozyme. *eLife* **2017**, 6, e31153.

20. Horning, D. P.; Bala, S.; Chaput, J. C.; Joyce, G. F. RNA-catalyzed polymerization of deoxyribose, threose, and arabinose nucleic acids. *ACS Synth. Biol.* **2019**, 8, 955-961.

21. Wochner, A.; Attwater, J.; Coulson, A.; Holliger, P. Ribozyme-catalyzed transcription of an active ribozyme. *Science* **2011** 332, 209.

APPENDIX C: EFFECT OF SUBSTRATE ENERGETICS ON RIBOZYMES SELECTED FROM A RANDOM POOL

Contribution Statement

Luiz F. M. Passalacqua (L.F.M.P.) performed rounds 1-6 of the selection and first sequencing analysis. **A.D.K.** performed rounds 7-12 of the selection and second sequencing analysis. A.L. conceived the study.

C.1 Introduction

Phosphate reactions have been integrally incorporated into chemical processes fundamental to life—from maintenance of genetic information to energy transduction. Phosphoryl transfer reactions are enzymatically catalyzed and provide the chemical energy necessary for cellular homeostasis and proliferation. Nature has clearly chosen phosphate as the activating molecule for these processes, likely during the advent of an RNA-based world on an early Earth.[1-4] Because ribonucleic acids can act as genetic polymers and chemical catalysts, RNA catalyzed reactions likely arose as one of the earliest genetically encoded chemical transformations.[5-7] While much effort has been extended towards discovering and generating catalytic RNAs (ribozymes), little has been done to investigate the chemical-evolutionary rationale for their overwhelming reliance on triphosphorylated substrates.[8] In other words, nucleoside triphosphates prevail as one of the preferred high-energy molecules, but the evolutionary mechanism for choosing nucleoside triphosphates vs. diphosphates, has not yet been elucidated.[9-13]

One possibility for the preference for triphosphates is that enzymes that use them as substrates can be simple or act faster. To address these hypotheses and gain insight into the evolution of prebiotic energy metabolism, we performed a *de novo* selection of adenylylating ribozymes utilizing either ATP or ADP. We have evolved a series of ribozymes that catalyze reactions with di- or triphosphorylated substrates. We performed these selections in direct

competition to gauge ribozyme abundance in random pools, giving us a method to measure the relative requirement of high-energy phosphate substrates.[14]

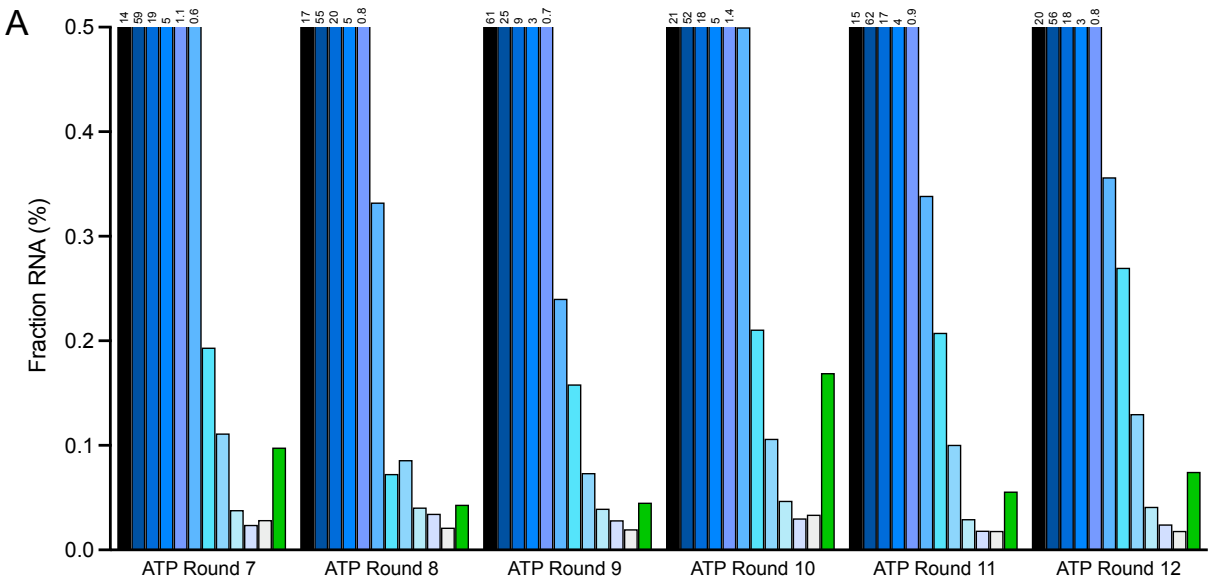
B.2 Results and Discussion

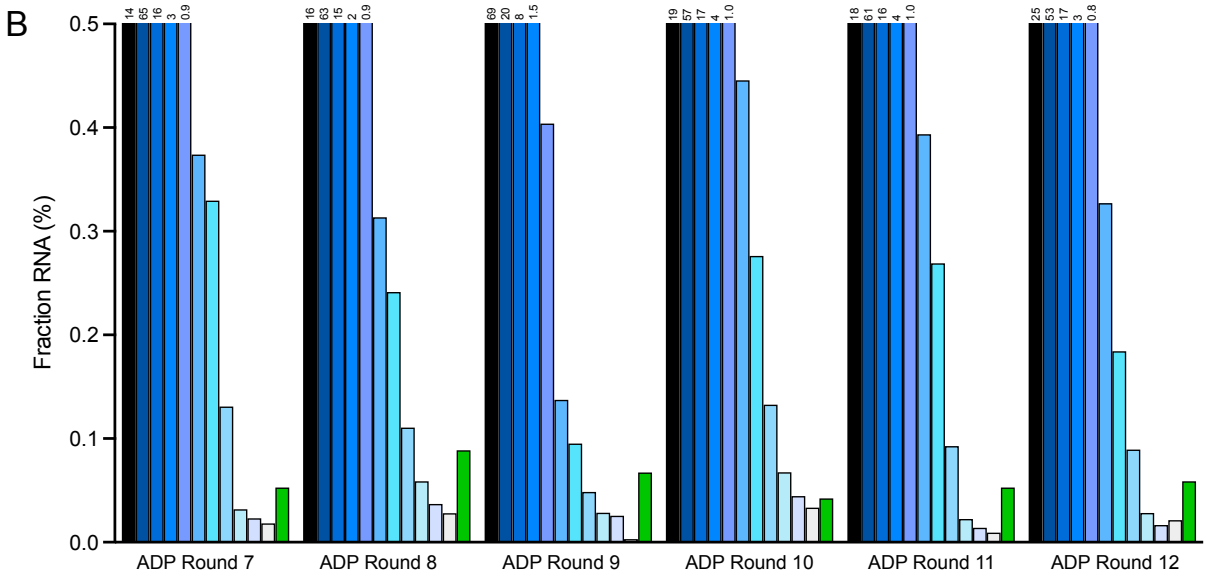
We performed an *in vitro* selection (from a random pool of ~5×10^{15} sequences) for ribozymes with reversible self-adenylylating activity with ADP and ATP substrates (addition of adenylate, i.e. AMP). At the beginning of each round, the RNA pools were transcribed with limiting ATP and limiting Mg^{2+} to prevent Mg^{2+}-dependent ribozyme adenylylation during transcription. Transcribed RNA was purified, incubated with the bead-immobilized substrate—ATP and ADP in parallel selections—and the adenylylated RNA was separated from the rest of the pool using extensive denaturing washes. Reverse transcription was performed on the beads after each round and the resulting cDNA was amplified and subjected to additional rounds of selection. The selection was sequenced at rounds 6 and 12. Previous studies with ATP and GDP substrates revealed 5′-capping ribozymes, and we expected that the majority of the selected sequences would have capping activity as well.[15-17]

In order to select for ribozyme activity with either ATP or ADP, additional stringency was added at each round of the selection. Either the concentration of ATP used in transcription was decreased, the length of incubation time was decreased, or the bead volume was decreased. For the last 6 rounds of selection, a counter selection step was added where the RNA pool was incubated with agarose beads prior to the selection round. Before the start of round 7, the pool was mutagenized by performing PCR with two modified nucleosides, 8-oxo-dGTP and dPTP used 1:1 with regular dNTPs.[18] Covalently bound RNA was detected on the ATP and ADP beads following all washes in each of the last 6 rounds of selection (**Figure C.1A-B**). The fraction of covalently bound RNA did not increase significantly between rounds; however, each round introduced new stringency which could account for the consistently low fraction of the pool RNA reacting with ATP or ADP. Following round 12 of the selection, the pool was sequenced and the relative abundancies

of the top 10 sequences were plotted (**Figure C.1C**). Sequencing the pool at different rounds of the selection allowed for comparison of relative sequence abundance between rounds, and allowed us to monitor the shifting sequence rankings as the selection progressed. We found that many sequences that appeared at the top of the selection at round six appeared much lower ranked in the round 12 selection. For example, the top sequence in the ADP selection after round 6 (ADP-6-1) appeared as the 42nd most abundant sequence in round 12 (ADP-12-42), and the second and third top sequences from round six do not appear in round 12 at all. Some sequences appear in both the ATP and ADP selections. The top sequence in the ADP selection after round 12 appears in the ATP selection as the 227th sequence, and ADP-12-2 appears as ATP-12-785. In the ATP selection, ATP-12-1 appears as ADP-12-23. The top sequences were subjected to Vienna RNAfold and many are predicted to have a highly structured hairpin on the 3`-end of the sequence. Additionally, a known adenosine binding motif appears in some of the top sequences of both selections.[19]

Our preliminary data indicate that the *in vitro* selections for both ATP and ADP utilizing ribozymes are equally successful, with no drastic difference in amplification efficiency between the two selections. Among the most abundant sequences in each selection, there were few sequences in common and they appeared as low abundance sequences in the counterpart selection indicating that the ADP/ATP reacting ribozymes are largely unique to the individual substrates. Beyond these initial observations, further analysis of the sequencing data as well as biochemical activity of individual clones is necessary to draw additional conclusions. Adenylylation activity must be measured for each of the two substrates to determine the activity of ATP vs ADP utilizing ribozymes, as well as the relative complexity of those ribozymes. This step will provide information that will further our understanding of ribozyme properties with respect to substrate energetics.[20]


A
Fraction RNA (%)
0.0
0.1
0.2
0.3
0.4
0.5
14 59 19 5 1.1 0.6
17 55 20 5 0.8
61 25 9 3 0.7
21 52 18 5 1.4
15 62 17 4 0.9
20 56 18 3 0.8
ATP Round 7
ATP Round 8
ATP Round 9
ATP Round 10
ATP Round 11
ATP Round 12


B
Fraction RNA (%)
0.0
0.1
0.2
0.3
0.4
0.5
14 65 16 3 0.9
16 63 15 2 0.9
69 20 8 1.5
19 57 17 4 1.0
18 61 16 4 1.0
25 53 17 3 0.8
ADP Round 7
ADP Round 8
ADP Round 9
ADP Round 10
ADP Round 11
ADP Round 12

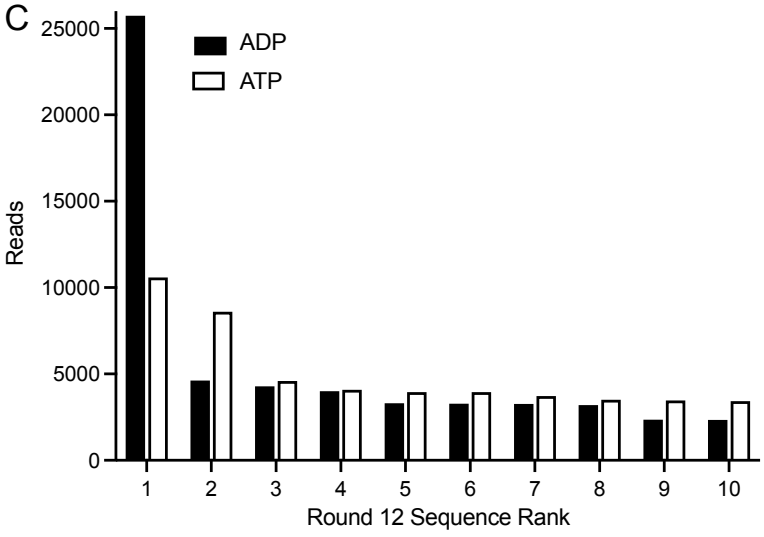

C
ADP
ATP
Reads
0
5000
10000
15000
20000
25000
1
2
3
4
5
6
7
8
9
10
Round 12 Sequence Rank

Figure C.1. Selection for ATP/ADP utilizing ribozymes. Rounds 7-12 of the selection for ATP (A) or ADP (B) utilizing ribozymes. The black column represents the counter selection step with agarose beads. All blue columns represent the different washing steps following the reaction of ribozymes with ATP-agarose beads. The green column represents the fraction of RNA covalently bound to the ATP-agarose beads, presumably after having reacted with ATP (A) or ADP (B). (C) The abundancies of the first 10 sequences of both ATP and ADP selections.

C.3 Materials and methods

PCR, transcription, binding assay, and reverse transcription of the RNA pool.

Either ATP selected pools or ADP selected pools were transcribed and ^{32}P-labeled for 1 hour at 37 ºC under the following conditions: 0.25 mM GTP, 0.25 mM CTP, 0.25 mM UTP, varying ATP concentrations, 4 mM $MgCl_2$, 8 mM DTT, 15% v/v DNA template, α-^{32}P-[ATP], 50 mM Tris HCl, 2 mM spermidine, and 0.01% Triton X-100 at 50 µL total volume. Following the transcription, 2 units of DNase I and 1x DNase I reaction buffer were added to the transcription and reacted at 37 ºC for 15 minutes. The reactions were then mixed with 7 M urea gel-loading buffer and purified on a 10% denaturing polyacrylamide gel (10% acrylamide/bis 19:1, 7 M urea, 0.5x Tris/Borate/EDTA buffer). The RNA was eluted from the gel in 300 mM KCl and then ethanol precipitated.

Agarose beads were suspended in a SpinX column and washed with 40 µL of 1x binding buffer (50 mM Tris HCl, 25 mM $MgCl_2$, 150 mM KCl, pH 7.6).[17] ATP and ADP RNA was resuspended in 40 µL of 1x binding buffer, heated to 42 ºC and added to the agarose beads. The RNA and agarose beads were rotated for 30 minutes, centrifuged (3000 rpm for 1 min), then the flow through was collected. The flow through was heated to 42 ºC and was either added to ATP or ADP-agarose beads with respect to the selection. The RNA was allowed to react with the ATP and ADP-agarose beads, then was centrifuged and the flow through collected. The columns were washed 8x with either a urea denaturing buffer (7 M urea, 50 mM EDTA) or a formamide buffer (95% formamide, 22.5 mM EDTA) heated to 42 ºC. The first four washes were incubated with the

beads for 5 minutes at room temperature with shaking. The fifth and sixth washes were incubated with the beads at room temperature for 20 minutes with shaking. The seventh and eighth washes were added to the beads with no incubation time. A final wash was done with 1x Tris buffer (50 mM Tris). The agarose counter selection beads, all the flow throughs and washes, and the ATP and ADP-agarose beads were collected and counted using a Beckman LS 6500 Multi-purpose Scintillation Counter. The fraction of RNA was calculated by dividing the counts of each condition by the total counts and multiplying by 100.

Reverse transcription was performed on the ATP and ADP-agarose beads under the following conditions: 0.5 μM RT Primer, 1x First Strand Buffer, 10 mM DTT, 0.5 mM each dNTP, and 100 units of M-MuLV reverse transcriptase. The reactions were allowed to incubate for 30 minutes at 42 ºC then centrifuged. The ATP and ADP-beads were washed with 1x Tris buffer and added to the reverse transcription reactions. The cDNA was ethanol precipitated and subsequently resuspended. PCR was performed to amplify the selected sequences with 0.5 μM each of a forward primer and reverse primer, and 1x DreamTaq PCR Master Mix. The resulting DNA was then carried on to subsequent rounds of selection. DNA from round 12 of the selection was amplified with TrueSeq adaptors, then another PCR was performed to add the Illumina barcodes to the ATP and ADP-selections and the reactions were gel purified. The libraries were submitted for Illumina Sequencing and the resulting data was counted and primer regions were trimmed using FastAptamer and CutAdapt.[21]

Conditions for each selection round.

Round 7: All conditions as above, but the transcription was performed with 0.075 mM ATP. The transcribed and purified pools were incubated with 40 μL (~5 mM) of either ATP or ADP-agarose beads for 2 hours.

Round 8: All conditions as above, but the transcription was performed with 0.050 mM ATP. The transcribed and purified pools were incubated with 40 μL (~5 mM) of either ATP or ADP-agarose beads for 2 hours.

Round 9: All conditions as above, but the transcription was performed with 0.050 mM ATP. The transcribed and purified pools were incubated with 40 µL (~5 mM) of either ATP or ADP-agarose beads for 1 hour.

Round 10: All conditions as above, but the transcription was performed with 0.050 mM ATP. The transcribed and purified pools were incubated with 20 µL (~2.5 mM) of either ATP or ADP-agarose beads for 1 hour.

Round 11: All conditions as above, but the transcription was performed with 0.050 mM ATP. The transcribed and purified pools were incubated with 20 µL (~2.5 mM) of either ATP or ADP-agarose beads for 30 minutes.

Round 12: All conditions as above, but the transcription was performed with 0.050 mM ATP. The transcribed and purified pools were incubated with 10 µL (~1.25 mM) of either ATP or ADP-agarose beads for 30 minutes.

Table C.1. Appendix C sequences.

1	Pool Primer Fwd.	GCTAATACGACTCACTATAGGGAGATCACTTACGGCACC
2	Pool Primer Rev. + RT Primer	CGACGCGCTGTCCCGAGC
3	TrueSeq Adapt. Fwd.	CACGACGCTCTTCCGATCTGGGAGATCACTTACGGCAC
4	TrueSeq Adapt. Rev.	AGACGTGTGCTCTTCCGATCTCGCGCTGTCCCGAGC
5	TrueSeq Universal Fwd.	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
6	Illumina Barcode Rev.	GATCGGAAGAGCACACGTCTGAACTCCAGTCACNNNNNNATCTCGTATGCCGTCTTCTGCTTG

C.4 References

1. Liu, Z.; Rossi, J.; Pascal, R. How prebiotic chemistry and early life chose phosphate. *Life* **2019**, 9, 26.

2. Leslie, E. O. Prebiotic chemistry and the origin of the RNA World. *Crit. Rev. Biochem. Mol. Biol.* **2004**, 39, 99-123.

3. Robertson, M. P.; Joyce, G. F. The origins of the RNA World. *Cold Spring Harb. Perspect. Biol.* **2012**, 13, 391-403.
4. Kamerlin, S. C.; Sharma, P.K.; Prasad, R. B.; Warshel, A. Why nature really chose phosphate. *Q. Rev. Biophys.* **2013**, 46, 1-132.

5. Bartel, D. P.; Unrau, P. J.; Constructing an RNA World. *Trends Cell Biol.* **1999**, 12, 9-13.

6. Chen, X.; Lin, N.; Ellington, A. Ribozyme catalysis of metabolism in the RNA World. *Chem. Biodivers.* **2007**, 633-655.

7. Joyce, G. F. RNA evolution and the origins of life. *Nature* **1989**, 338, 217-224.

8. Micura, R.; Hobartner, C. Fundamental studies of functional nucleic acids: aptamer, riboswitches, ribozymes and DNAzymes. *Chem. Soc. Rev.* **2020**, 20, 7331-7353.

9. Benner, S. A.; Ellington, A. D. Tauer, A. Modern metabolism as a palimpsest of the RNA World. *Proc. Natl. Acad. Sci. USA* **1989**, 86, 7054-7058.

10. Kamerlin, S. C.; Sharma, P.K.; Prasad, R. B.; Warshel, A. Why nature really chose phosphate. *Q. Rev. Biophys.* **2013**, 46, 1-132.

11. Arrhenius, G.; Sales, B.; Mojzsis, S.; Lee, T. Entropy and charge in molecular evolution-the case of phosphate. *J. Theor. Biol.* **1997**, 187, 495-501.

12. Baltscheffsky, H. Major “anastrophes” in the origin and early evolution of biological energy conversion. *J. Theor. Biol.* **1997**, 187, 495-501.

13. Deamer, D.; Weber, A. L. Bioenergetics and life’s origins. *Cold Spring Harb. Pespect. Biol.* **2010**, 2, a004929.

14. Chizzolini, F.; Passalacqua, L. F. M.; Oumais, M. O.; Dingilian, A. I.; Szostak, J. W.; Lupták, A. Large phenotypic enhancement of structured random RNA pools. *J. Am. Chem. Soc.* **2020**, 142, 1941-1951.

15. Huang, F.; Yarus, M. 5′-RNA self-capping from guanosine diphosphate. *Biochemistry* **1997**, 36, 6557-6563.

16. Moretti, J. E.; Muller, U. F. A ribozyme that triphosphorylates RNA 5′-hydroxyl groups. *Nucleic Acids Res.* **2014**, 42, 4767-4778.

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

18. Zaccolo, M.; Williams, D. M.; Brown, D. M.; Gherardi, E. An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *J. Mol. Biol.* **1996**, 255, 589-603.

19. Sassanfar, M.; Szostak, J. W. An RNA motif that binds ATP. *Nature* **1993**, 364, 550-553.

20. Carothers, J. M.; Oestreich, S. C.; Davis, J. H.; Szostak, J. W. Informational complexity and functional activity of RNA structures. *J. Am. Chem. Soc.* **2004**, 126, 5130-5137.

21. Alam, K. K.; Chang, J. L.; Burke, D. H. FASTAptamer: A bioinformatic toolkit for high-throughput sequence analysis of combinatorial selections. *Mol. Ther. Nucleic Acids* **2015**, 4, e230.

APPENDIX D: BIOLUINESCENT PROBES FOR IMAGING CELL-CONTACT EVENTS *IN VIVO*

Note: Previously submitted as the Second Year Research Report in partial satisfaction for the degree of Doctor of Philosophy in Chemistry.

D.1 Abstract

Cell-contact events are key to numerous biological processes and govern many functions including disease onset and cell motility. However, current methods of studying these interactions are incapable of delivering information on the macroscale. I am developing caged luciferins that can report on cell-cell interactions in tissues and whole animals. I have synthesized one probe that can be selectively released by nitroreductase (NTR)-expressing cells. The liberated luciferin is preferentially consumed by neighboring luciferase-expressing cells, providing a readout on cell proximity. I have also designed caged luciferase enzymes that provide an amplification of signal for more sensitive imaging. Herein is presented progress toward this goal employing caged luciferin molecules or circularly permutated, caged luciferases capable of imaging cell proximity *in vivo.*

D.2 Introduction

Cell-cell interactions govern numerous *in vivo* functions from immune responses to host-pathogen communication.[1-3] These cellular communications occur on the microscopic scale and cannot be visualized non-invasively on a whole animal scale.[4,5] Studying these exchanges is central to expanding our understanding of fundamental physiological processes. Thus, methods to "see" cell-cell interactions are needed. The goal of my research is to develop a toolset that gives us a window into the world of microscopic interactions on a macroscopic scale.

Much of our knowledge regarding cell-cell interactions derives from fluorescence-based imaging strategies.[4] However, these techniques are largely confined to imaging on the microscale. This is due, in part, to their need for excitation light. Researchers must know where and when to

excite the sample, which may be unclear when surveying a large area. Tissues also exhibit high autofluorescence, decreasing the sensitivity of many conventional fluorescent proteins on the macroscale.[6]

As a complement to fluorescence-based microscale approaches, bioluminescence imaging (BLI) enables examination of cellular interactions on a macroscopic scale.[7-9] BLI relies on enzymes (luciferases) that produce light with luciferin substrates. The most widely used luciferase/luciferin pair originates from the firefly.[7] Firefly luciferase (Fluc) catalyzes the oxidation of D-luciferin (D-luc), releasing light that can penetrate tissues and skin and be detected by sensitive cameras.[10] Since BLI relies on a chemical reaction and requires no excitation light, it is often more suitable for addressing biological questions on a whole animal level. Indeed, Fluc and D-luc have been used to image diverse biological phenomena such as tumor metastases and bacterial clusters in mouse models.[3, 11-13]

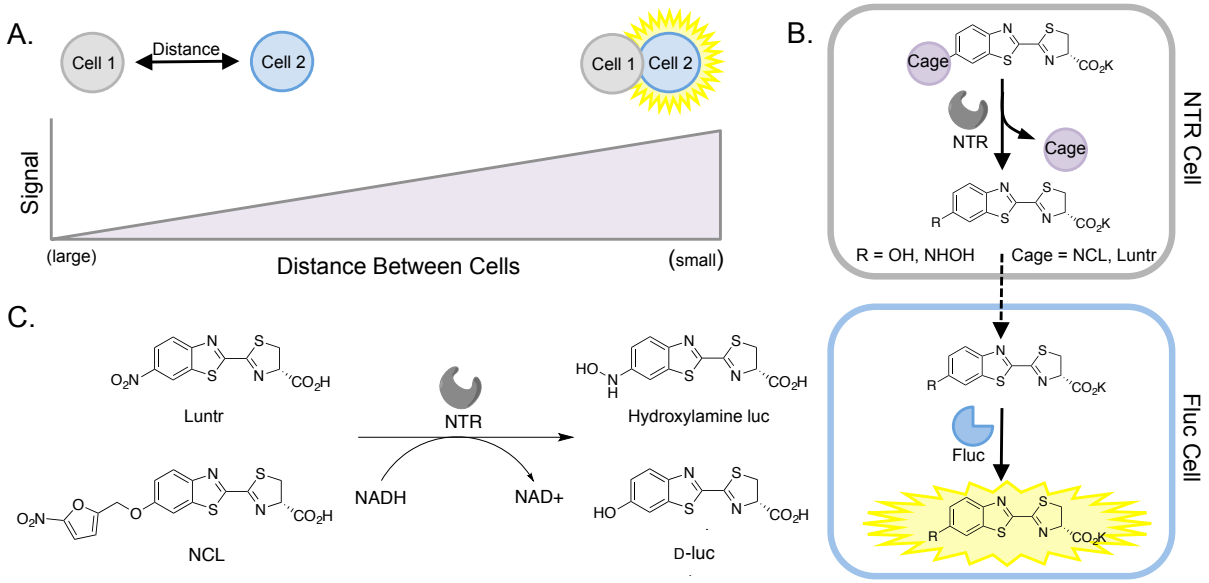


Figure D.1. Monitoring cell-cell interactions utilizing a caged luciferin probe. Figure adapted from Reference (14) (A) Strategy for visualizing cell contact. (B) After removal of the cage in an activator cell, D-luciferin (D-luc) or an analog diffuses into a firefly luciferase (Fluc) expressing

cell, and is oxidized to produce light. (C) Luntr or nitro caged luciferin (NCL) can be uncaged by nitroreductase (NTR) to produce hydroxylamine luc or D-luc respectively.

While broad in scope, BLI currently lacks the spatial resolution to report on microscopic cell contact events with adequate detail.[8] BLI tracks global cell movements and numbers, but like many macroscale techniques, can only approximate cell locations and interactions. The objective of my research is to build on the sensitivity of BLI and craft bioluminescent tools to observe cell-cell interactions by correlating cell proximity with bioluminescent photon output (**Figure D.1**).[14] To this end, I have been working to develop "caged" luciferins.[15] These probes incorporate masking groups that prevent light emission until removed by a corresponding enzyme. Such luciferins can be co-opted to visualize cell-cell interactions.[16,17] In this strategy, one cell (expressing the uncaging enzyme) can liberate the luciferin, enabling nearby Fluc-expressing cells to use the probe.[18] Light emission is greatest in regions where the two cell types are in direct contact.

Caged luciferins for imaging cell-cell contacts must meet several criteria. They should emit little to no light before enzymatic release, and thus be resistant to non-specific uncaging under physiological conditions. The luciferins should also freely diffuse between cells and produce a high level of light emission upon Fluc catalysis. Finally, the probes should be rapidly uncaged, enabling sufficient luciferin to be produced on timescales relevant to cell-cell interactions.[18] Nitroreductase (NTR) is an enzyme not expressed in many mammalian cells, making it an attractive uncaging enzyme for imaging cell-cell interactions relevant to human biology. NTR has also been widely used to remove masking groups from fluorophores and other clinically relevant imaging molecules, demonstrating its robustness for reducing nitro groups *in vivo*.[19-22]

Previously, the Prescher lab capitalized upon the selective reducing nature of nitroreductase and designed a NTR-sensitive caged luciferin (Luntr) comprising a 6'-nitro group. Although NTR could effectively uncage Luntr and liberate a reduced hydroxylamine luciferin (**Figure D.1B**), the product displayed low light emission. We hypothesized that this was due to a combination of low cell membrane permeability, slow uncaging, and Fluc turnover of Luntr itself. Low light

emission coupled with very slow *in cellulo* uncaging rates, extending 6-12 h in some experiments, severely limited *in vivo* utility, necessitating the development of new caged probes.

To identify improved probes, I am exploring alternative NTR-sensitive cages (**Figure D.1C**). I hypothesized that an alternative nitro caged luciferin (NCL) could address some of the limitations presented by Luntr. NCL is known to exhibit faster uncaging kinetics and permeate bacterial cell walls. Thus, I pursued this avenue of investigation and report my progress toward developing a more robust, caged, bioluminescent tool. I also hypothesized that with previously mentioned advantages, NCL might extend its utility to bacterial sensing and host pathogen interactions, biological interactions Luntr was unable to report on. Finally, I present my efforts toward developing a new method to visualize cell contacts with extracellularly activated luciferases. Such tools have the potential to boost bioluminescent photon output and enable more sensitive imaging of cell-cell interactions.

D.3 Results and Discussion

Design and in vitro characterization of an NTR-responsive caged luciferin.

In exploring new probe designs, I was attracted to a recently reported caged luciferin.[23] NCL was designed to sense engineered strains of bacteria expressing Fluc. Most bacterial species express high levels of NTR, such that NCL can be rapidly cleaved, liberating D-luc. Unlike Luntr, NCL uncaging provides a more cell permeant luciferin.[24] D-Luc is the native substrate of Fluc and can easily pass between cell membranes. D-Luc also possesses a higher quantum yield and photon output compared to other luciferin architectures.[25,26] These enhancements should improve imaging sensitivities, and enable NCL to be used in a wider array of cell contact applications.

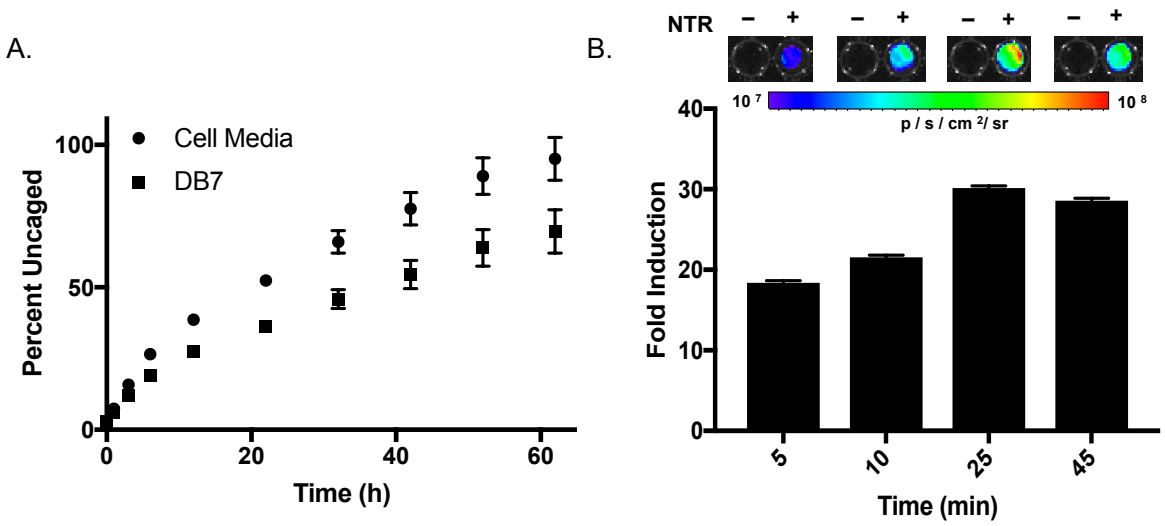


Figure D.2. NCL is stable, but is uncaged *in vitro*. (A) NCL (250 μM) was incubated with media (DMEM + 10% FBS) or 50,000 DB7 cells at 37 °C for 62 h. Non-specific uncaging was monitored by fluorescence (Ex: 350, Em: 534). Fluorescence signal was normalized for background and plotted as percent of uncaged probe present. (B) NCL (100 μM) was incubated with NADH (200 μM) and NTR ([+] = 10 μg/mL, [–] = no enzyme) at 37 °C for 0-45 min. Reactions were then imaged for 5 min following the addition of Fluc (10 μg/mL) and ATP (500 mM). The fold induction in bioluminescent signal is plotted and sample bioluminescence images are shown. Error bars for (A) and (B) represent the standard deviation of the mean for n = 3 replicates.

To test this hypothesis, I first prepared NCL according to an established protocol (**Figure D.S1**).[23] The stability of NCL was then assessed via NMR (**Figure D.S2**). No degradation was observed over several days, suggesting that it would be suitable for imaging in tissues and whole organisms.[27] The probe was also tested in more physiologically relevant conditions. NCL was incubated with cell media (DMEM + 10% FBS) or DB7 cells, and probe integrity was monitored via fluorescence over time (**Figure D.2A**). NCL and the uncaged probe (D-luc) have unique fluorescence signatures and thus the non-specific uncaging of the probe can be monitored at λ_{ex} 350 nm and λ_{em} 534 nm.[28] Minimal degradation was observed after 1 h, the time scale relevant to the majority of *in vivo* imaging experiments.

Confident in NCL's ability to capture cell-cell interactions, I evaluated its uncaging *in vitro*. NCL was incubated with purified NTR for 0-45 min and aliquots were then added to purified Fluc and imaged (**Figure D.2B**). Robust light emission was observed even after 5 min of incubation. NCL exhibits much faster uncaging kinetics compared to Luntr, where little signal was observed until 30 min of NTR exposure.[14] It should be noted that these data were acquired with NTR exhibiting 45% activity relative to previously published activity (**Figures D.S3-S4**).[14] While the relative comparisons between Luntr and NCL are most important, future experiments will utilize fully active NTR. Thus far, the initial experiments indicate that NTR provides a two-fold improvement in activation compared to Luntr, but on a much faster time scale. A detailed kinetic analysis of NCL responsiveness to NTR *in vitro* has been previously published and comparable activity will be confirmed upon expression of NTR.[23]

NCL can be uncaged by NTR-expressing mammalian cells.

The uncaging reaction of NCL by NTR was also investigated *in cellulo* using a mouse breast cancer cell line (DB7) engineered to express bacterial NTR.[14, 29] NCL was added to DB7-NTR cells. After 30 min, the media from these cultures was transferred to wells containing Fluc and ATP. As shown in **Figure D.3A**, dose-dependent light emission was observed up to 10 μM. Non-NTR expressing cell lines showed no appreciable uncaging (**figure D.S5**).[30] Less light emission was observed at higher concentrations of NCL though, likely due to product inhibition or NCL competitively binding the Fluc active site. To test this latter hypothesis, increasing amounts of NCL were added to a solution of Fluc and D-luc. As shown in **Figure D.3B**, larger concentrations of NCL inhibited light output. Oxyluciferin (the product of Fluc catalysis) is also an inhibitor of Fluc.[31,32] Thus, these two phenomena are likely impeding light emission. It is possible that NTR is exhibiting low uncaging efficiency in cells. However, the exact amount of D-luc being released is difficult to measure. More experiments will be done to test this hypothesis, but uncaging efficiency is a possible parameter to improve in my strategy. If improved uncaging is necessary in

mammalian cells, I will engineer the cells to express a codon optimized NTR.[33,34] Future experiments could also focus on evolving NTR to for improved uncaging efficiency.

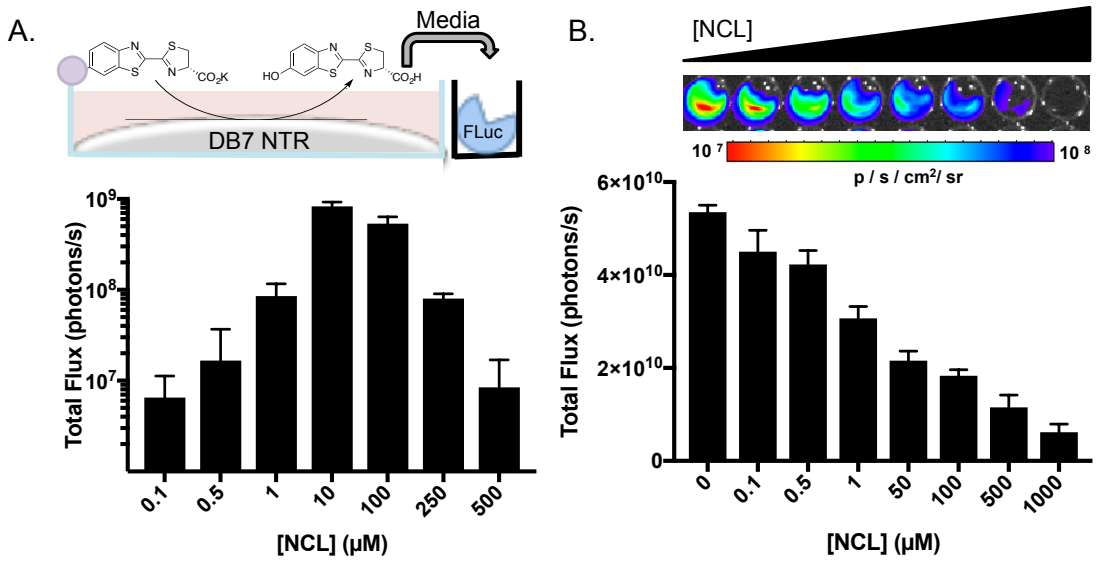


Figure D.3. NTR is uncaged *in cellulo* in a dose dependent manner. (A) NCL was added to 50,000 DB7 NTR expressing cells. Media was then transferred to a 96-well plate containing Fluc (10 μg/mL) and ATP (500 mM) after a 30 min incubation at 37 °C. The total flux is plotted on a log scale. Error bars represent the standard deviation of the mean for *n* = 3 replicates. (B) Caged probe inhibits light emission. NCL was added to a 96-well plate containing Fluc (10 μg/mL), ATP (500 mM), and D-luc (100 μM). The total flux is plotted and sample bioluminescence images are shown. Error bars represent the standard deviation of the mean for *n* = 3 replicates.

NCL selectively reports on cell-cell interactions.

Based on promising initial data, I progressed to a completely cellular regime in which NCL uncaged by one cell type could be processed by Fluc-expressing cells. DB7-NTR cells were again incubated with NCL for 0-120 min. Media samples from the cultures were then transferred to DB7 cells expressing Fluc (DB7-Fluc). Light emission was observed as early as 5 min into the incubation period, a drastic improvement over the 16 h incubation period required for light emission with Luntr (**Figure D.4A**).[14] This result confirms that the uncaged form of NCL, D-luc, can easily diffuse

between cells and solves one of the key limitations of earlier cell-contact investigations in the Prescher lab. DB7 NTR cells were then co-plated with DB7 Fluc cells, and BLI was performed periodically after addition of NCL (**Figure D.4B**). Successful uncaging of NCL was observed in the co-plating experiment and a 40-fold increase in bioluminescent signal over non-specific background was observed. Compared to the six-fold increase seen from Luntr, NCL offers a significant improvement.[14] Both media transfer and co-plating assays provide strong evidence that NCL is reporting on cell contact, which will be verified through obtaining distance dependent data by employing stencils previously pioneered in the Prescher lab.[14,35] Photon output from these co-plating data further suggests that as few as 300-700 cell-cell interaction events can be visualized with NCL in a mouse model, as calculated from bioluminescence imaging standards.[36,37]

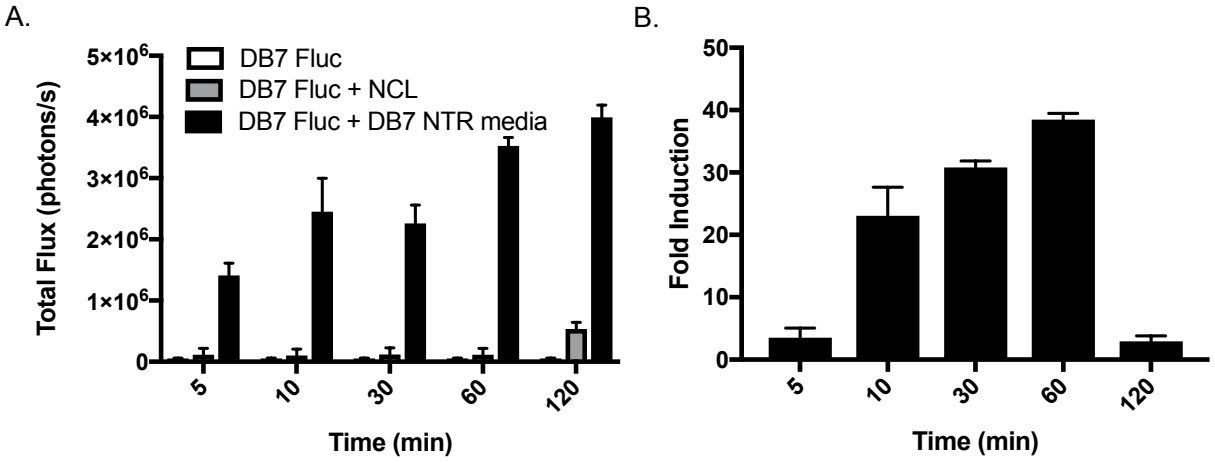


Figure D.4. NCL is uncaged and consumed in cells. (A) NCL and uncaged D-luc are cell permeant demonstrated with a media transfer assay. NCL (250 μM) was added to 50,000 DB7 NTR cells and allowed to incubate at 37 °C for 0-120 min. Media from those cells was then transferred to DB7 Fluc cells and imaged. Total photon flux is plotted. Error bars represent the standard deviation of the mean for *n* = 3 replicates and the data are representative of two independent runs. (B) NCL is uncaged and used when DB7 NTR and DB7 Fluc cells are plated together. NCL (250 μM) was added to 50,000 DB7 NTR cells and 50,000 DB7 Fluc cells. Bioluminescence

images were collected over 120 min. The fold induction in bioluminescent signal over control DB7 cells is plotted. Error bars represent the standard deviation of the mean for n = 3 replicates and the data are representative of three independent runs.

Application of NCL to host-pathogen interactions.

Earlier probes prevented sensitive imaging of host pathogen interactions. In order to determine if NCL would be a viable probe to monitor these phenomena, a model bacteria system with *Escherichia coli* was investigated. *E. coli* transformed with an ampicillin resistant vector were grown overnight from a single colony. The number of bacteria in the stock was estimated from OD_{600} measurements. Serially diluted samples were then added to DB7-Fluc cells and incubated for 0-90 min (**Figure D.5A**). Following addition of NCL, signal was observed over background in a dose-dependent manner, but with inverse dependence with respect to *E. coli* concentration. The lowest concentrations of bacteria showed the highest light emission from 0-30 min; for 100 *E. coli* cells, a 12-fold increase was measured over background indicating that very low numbers of bacteria can be detected (**Figure D.5B**). Some reports on bacterial growth indicate that NTR and other metabolic enzymes are upregulated at low *E. coli* concentrations, which could provide one hypothesis for the inverse dependency.[38,39] The high concentrations of bacteria could also be causing mammalian cell death.[40] More experiments will be conducted to evaluate these hypotheses. A sensitive assay for detection of bacteria could have applications in investigating pathogen interactions with specific cell types *in vivo* such as tracking immune responses to infection.[41]

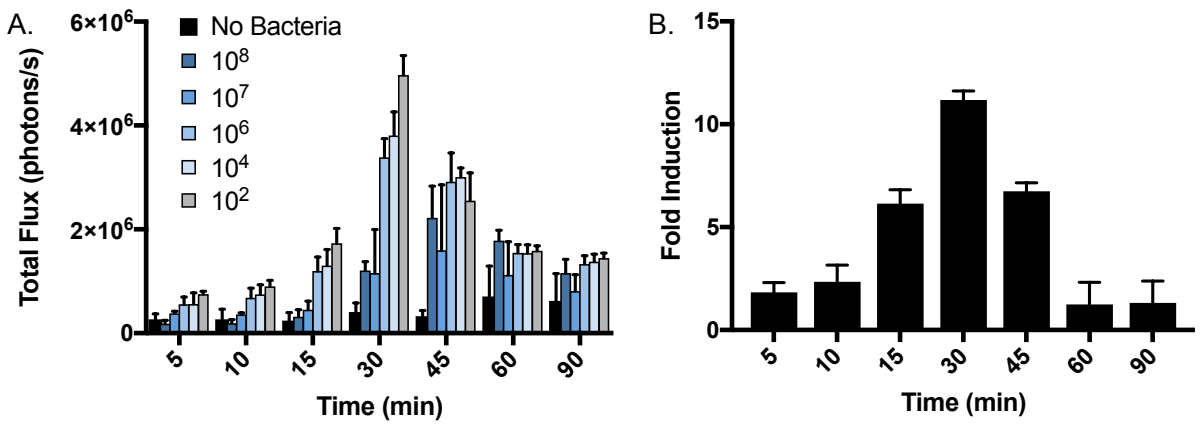


Figure D.5. NCL can detect low numbers of bacteria. (A) *E. coli* (10^2-10^8 cells) were added to DB7 Fluc cells. NCL (250 μM) was added and bioluminescence images were acquired over time. Total flux is plotted. (B) The fold induction in bioluminescent signal over background is potted for 100 *E. coli.* For (A) and (B) error bars represent the standard deviation of the mean for n = 3 experiments and data are representative of two biological repeats.

Progress towards a protease-responsive luciferase.

In conjunction with my work on NTR-responsive caged luciferins, I have also been designing a system for detection of cell contact events that does not rely on caged luciferins. This approach involves "caged" luciferase enzymes that can be expressed on cell surfaces and uncaged in response to secreted enzymes from neighboring cells (**Figure D.6A**).[42-44] The luciferases selected for this purpose were Gaussia luciferase and Nanoluc.[45,46] These enzymes can emit light in the extracellular space, as they do not require ATP to catalyze their substrates (coelantrazine and furimazine, respectively). This approach could potentially provide a more direct readout of cell contact, but more importantly, it could enhance bioluminescent photon output. Each productive uncaging results in the release of an enzyme that can catalyze multiple turnovers. In the case of NCL and other caged substrates, each uncaging event provides a maximum of one photon. Caged enzymes could thus enable imaging on a broader timescale than transient interactions

that can be studied with other caged tools.[37] It can be applied to problems such as organ development and longer-lived immune interactions.[1,47]

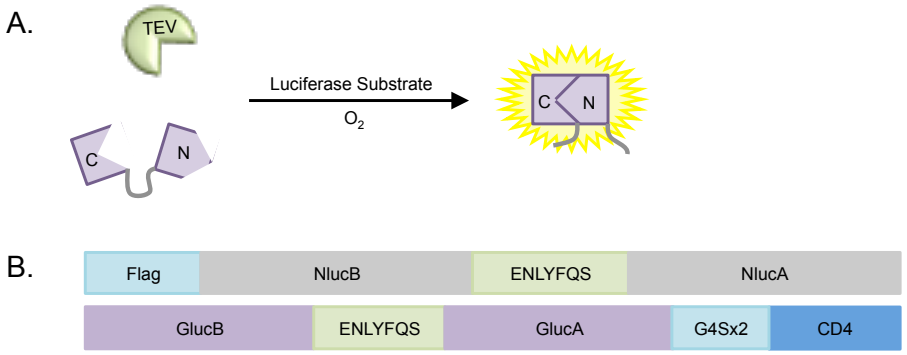


Figure D.6. Circular luciferases can be used to monitor cell contact. (A) A circular luciferase is split by a protease (TEV) recognition site. TEV cleaves the recognition site keeping the two halves of the protein apart allowing the protein to reform. (B) Sequences of the cLuc inserts in pBMN vectors. The secreted version featured on the top has a Flag recognition sequence on the N terminus followed by the C terminal portion of Nluc, a TEV recognition sequence, and the N terminal portion of Nluc at amino acid split 65/66. In the transmembrane version, the C terminus of Gluc is followed by a TEV recognition sequence, the N terminus of Gluc, a Glycine-Serine linker, and the CD4 transmembrane domain at amino acid split 93/94.

I have been working toward developing circular variants of Gaussia (Gluc) and Nano (Nluc) luciferases in which they are split at one point and encoded such that the C-terminus (B) of the protein precedes the N-terminus (A).[48-50] The two parts of the protein are then joined by a protease recognition element. When in the presence of the matching protease, that sequence is cleaved and should allow for the reformation of functional protein (**Figure D.6A**). After careful review of literature precedence for circularized proteins and use of extracellular luciferases to detect small molecules, I designed split positions for both extracellular luciferases and added a protease recognition sequence. I chose TEV protease as an initial model to work with due to its

detailed characterization and use in other circularized protein variants.[51] I expressed TEV protease according to a protocol adapted from the Martin lab (UCI) and confirmed its activity compared to a standard (**Figure D.S6**). I designed both secreted and transmembrane variants of both Gaussia and Nano luciferases. We cloned two of these constructs into PBMN vectors (**Figure D.6B**). I have been progressing toward making other circularly permutated luciferases and transfecting current plasmids into HEK293 cells.

D.4 Summary and Future Directions

A pressing need in the bio-imaging field is the visualization of microscopic events on a macroscopic scale. I have been working to address these needs on two fronts. In one area, I have made strides toward the development of a viable approach to imaging cell contact events utilizing a caged luciferin, NCL, which represents an improvement over other NTR-responsive probes in both *in vitro* and *in cellulo* regimes. NCL has also been shown to exhibit bacteria-sensing capabilities useful for host pathogen visualization due to low detection limits. Cell data suggests that only 300-700 contact events are necessary to visualize cell-cell interactions in a mouse model, demonstrating progress toward *in vivo* imaging. I can further improve on that limit through codon optimization and directed evolution of NTR. Optimized NCL imaging will be used to detect immune interactions with tumor metastases *in vivo*. Finally, I have begun progress toward developing of a fully genetic cell contact visualization system, that allows for signal amplification compared to BLI methods employing caged probes and could thus afford even greater sensitivity in the quest to visualize cell contact. In further developing these probes, I am adding more tools for bioluminescent imaging to those already available in the Prescher Lab in order to shed light on macroscale biological processes.

D.5 Supporting Information

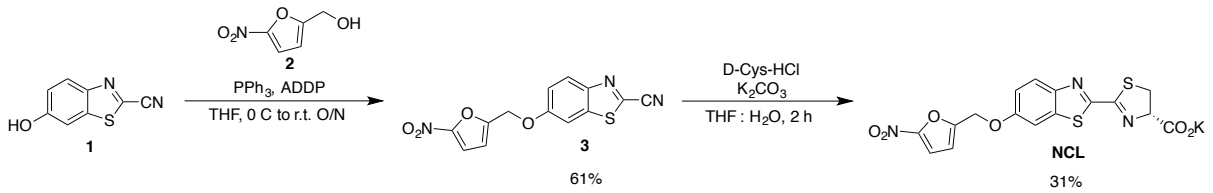


Figure D.S1. Synthetic route for NCL following the protocol reported by Vorobyeva, A. G. *et al*.[23]

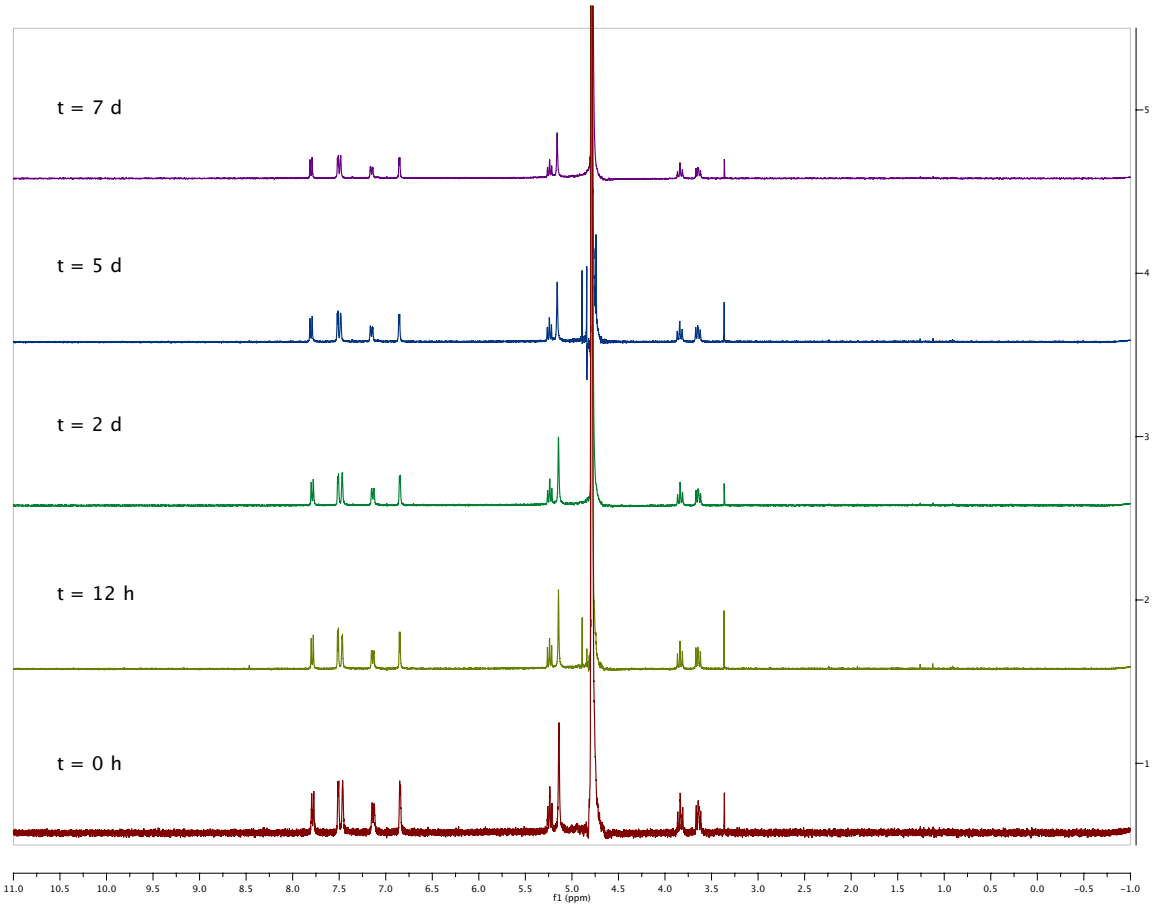


Figure D.S2. NMR stability study in D_2O. Monitoring of probe in D_2O over seven days shows no degradation products.

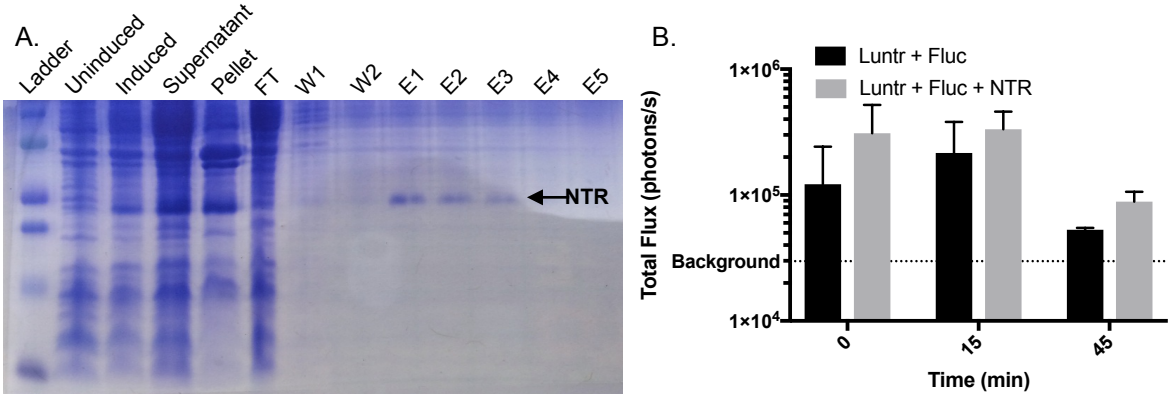


Figure D.S3. Gel electrophoresis image and bioluminescence from NTR expression. (A) Gel electrophoresis image displaying the expression of NTR as described in the Materials and Methods section. The uninduced fraction comprised BL21 cells before addition of IPTG, and the induced fraction comprised cells following addition of IPTG and growth. The supernatant and pellet fractions were collected taken after lysis of the bacteria and centrifugation. The flow through (FT) was the fraction collected after the supernatant was allowed to bind to a Ni-NTA column. Washes (W1, W2) were performed and the protein was eluted (E1-E5). NTR (26 kDa) appears in E1-E3. (B) Expressed NTR was not active and did not reduce the Luntr. Luntr (100 μM) was incubated with NADH (200 μM) and NTR (10 μg/mL) at 37 °C for 0-45 minutes. Reactions were treated with Fluc (10 μg/mL) and ATP (500 mM), then imaged. Background light emission indicates samples containing Fluc and ATP only. Note that Luntr does show some activity with Fluc over background consistent with previously published data.[14] Error bars represent the standard deviation of the mean for n = 3 replicates.

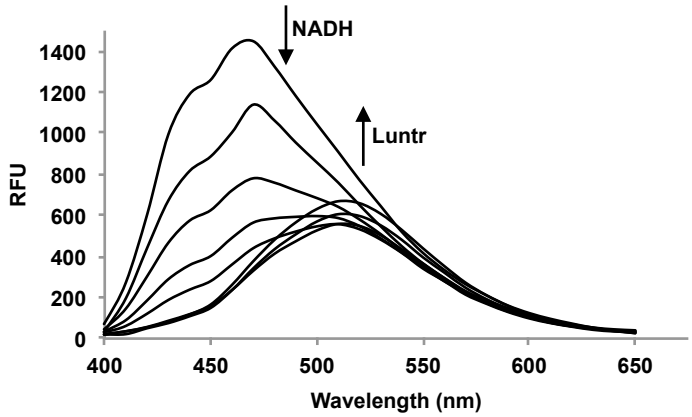


Figure D.S4. Fluorescence data generated from partially active NTR uncaging Luntr. NTR aliquot demonstrates reduced activity. Luntr (100 μM) was added to NADH (200 μM) and NTR (2 μg/mL). The reaction was monitored over 4.5 h at an excitation wavelength of 350 nm. Measurements were taken in 5 min intervals and representative spectra are shown.

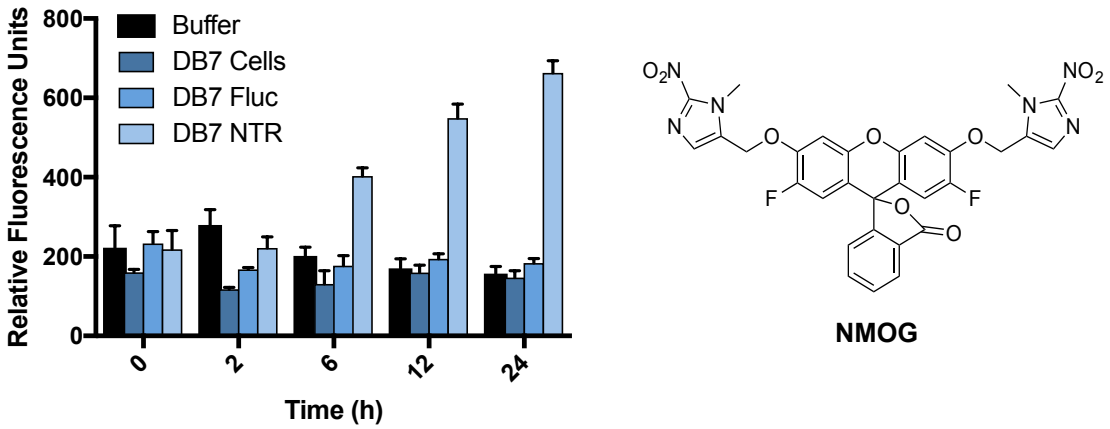


Figure D.S5. DB7-NTR cells selectively remove a nitro cage. Nitroimidizole caged Oregon Green (NMOG 100μM) (courtesy of the Lavis lab, HHMI Janelia) was added to 50,000 cells of DB7-NTR, DB7-Fluc, or DB7 cells.[30] Fluorescence was measured at λ_{ex} 500 nm and λ_{em} 528 nm and was monitored over 24 h. Error bars represent the standard deviation of the mean for n = 3 replicates and data are representative of three independent runs.

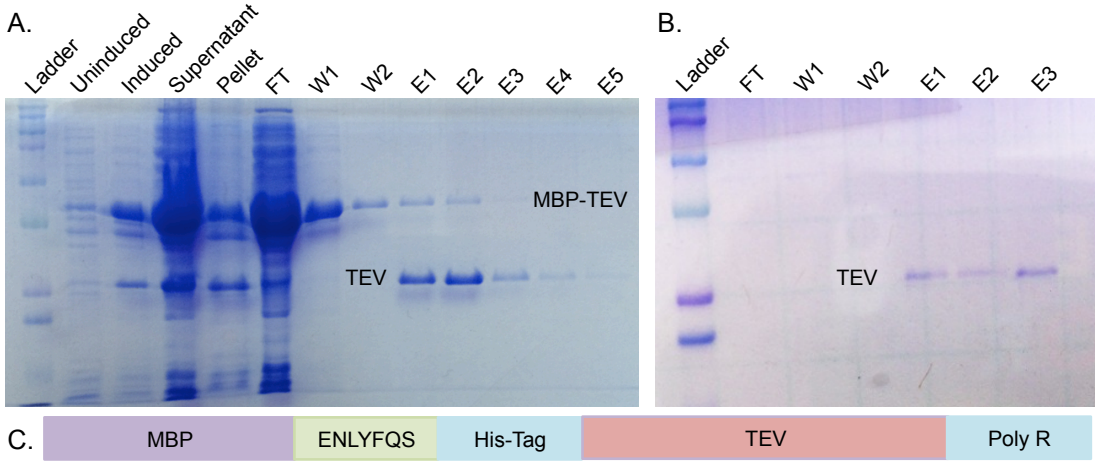


Figure D.S6. Gel electrophoresis images from TEV protease expression. (A) Gel electrophoresis image for TEV protease isolation as described in the Materials and Methods section. The uninduced fraction comprised BL21 cells before addition of IPTG, and the induced fraction comprised cells following addition of IPTG and growth. The supernatant and pellet fractions were collected taken after lysis of the bacteria and centrifugation. The flow through (FT) was the fraction collected after the supernatant was allowed to bind to a Ni-NTA column. Washes (W1, W2) were done and the protein was eluted (E1-E5). TEV protease (27 kDa) as well as the MBP-TEV fusion can be visualized. (B) Gel electrophoresis image of the second purification of TEV protease after self-cleavage from the MBP tag. (C) Sequence of expressed protein. Maltose binding protein (MBP) is followed by the protease recognition sequence for TEV. A poly histidine tag allowing for nickel column purification precedes TEV protease, and the construct ends with a poly arginine tag (poly R).

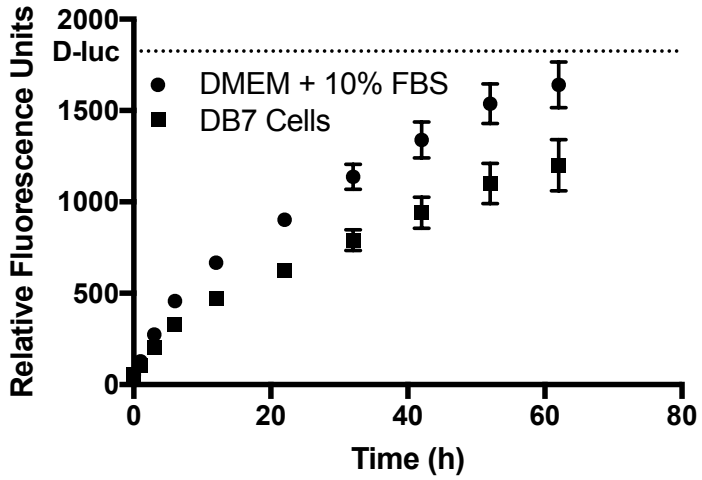


Figure D.S7. Fluorescence data for Figure D.2A. NCL (250 μM) was incubated with media (DMEM + 10% FBS) or 50,000 DB7 cells at 37 °C for 62 h. Non-specific uncaging was monitored by fluorescence (Ex: 350, Em: 534). Fluorescence signal was normalized for background. The dashed line represents 250 μM D-luc. Error bars represent the standard deviation of the mean for *n* = 3 replicates.

D.6 Materials and Methods

General biological methods.

All restriction enzymes and T4 DNA ligase were purchased from New England Biolabs. DNA polymerase (DreamTaq) was purchased from Thermo Scientific. Buffer salts and dNTPs were purchased from Fisher Scientific. Luria-Bertani medium (LB) was purchased from Genesee Scientific. Profinity IMAC Uncharged Resin was purchased from BioRad and then charged with 0.1 M $NiCl_2$ before use in protein purification. NADH was purchased from Research Products International Corp. and D-luciferin was purchased from Biosynth International Inc. All PCR reactions and restriction enzyme reactions were analyzed using agarose gel electrophoresis and ethidium bromide staining. For all experiments, NCL was prepared as a 25 mM stock by first dissolving in DMSO followed by dilution with PBS (final concentration of 10% DMSO). NCL was stored as a solid at –80 °C under argon. Solutions of NCL were stored at –80 °C for up to two weeks.

General fluorescence imaging.

All fluorescence uncaging and stability assays were performed using black Grenier-Bio 96-well microplates on a Molecular Devices SpectrMax Gemini XPS microplate reader. Sample volumes of 100 µL per well were used and all plates were incubated at 37 °C for the duration of the experiment. All measurements were performed in triplicate. Fluorescence data were exported and analyzed using GraphPad Prism 7 and Microsoft Excel.

General bioluminescence imaging.

All *in vitro* biochemical and cell culture assays were performed using Grenier-Bio 96-well microplates. Plates were imaged in a light proof chamber with an IVIS Lumina (Xenogen) CCD camera chilled to –90 °C. The imaging stage was kept at 37 °C throughout the experiment. The camera was controlled utilizing Living Image software. Exposure times ranged from 1-10 s for *in vitro* biochemical assays with data binning levels set to medium. Exposure times ranged from 10-30s for cell culture assays with data binning levels set to medium. Total flux values were quantified using Living Image software and bioluminescence images were taken from those displayed using Living Image software. Data were plotted and analyzed using GraphPad Prism 7.

General cell culture methods.

HEK 293 cells (ATCC), DB7 cells, DB7-Luc2 cells (courtesy of the Contag Lab, Stanford), and DB7-NTR cells (generated in the Prescher Lab) were cultured in complete media: DMEM containing 10% (vol/vol) FBS (Corning), 4.5 g/L glucose, 2mM glutamine, 100 U/mL penicillin, and 100 µg/mL streptomycin all from Gibco. Cells were maintained in a 5% CO_2, water-saturated incubator at 37 °C. Circularized luciferase plasmids in Figure 7B were obtained from Krysten Jones (Prescher Lab). The pET-NTR plasmid was obtained from William Porterfield (Prescher Lab).

Cloning of circularized luciferase constructs.

Table D.S1. Primers used for cloning.

Number	Primer	Sequence 5' to 3'
1	For	ATAATGCATGAGGCGATCGTCGAGATTC
2	Rev	ATACTCGAGGCCCTGGAAGTACAGGTTCTCACGCGTGTCACCAC-CGGCCCC
3	For	ATAGAATTCGGTCTGAGCGGCGACC
4	Rev	GCCCTGGAAGTACAGGTTCTCACGCGTCGCCAGAATGCGTTCGC
5	For	ACGCGTGAGAACCTGTACTTCCAGGGCATGCATGTCTTCACAC-TCGAAGATTTCGTTG
6	Rev	ATAATTAATTAATTCATACGGGATGATGACATGG
7	For	ATAATGCATGGGCTGAAGATCGACATCCA
8	Rev	GCCCTGGAAGTACAGGTTCTCACGCGTCGCCAGAATGCGTTCGC
9	For	ACGCGTGAGAACCTGTACTTCCAGGGCCTCGAGGTCTTCACAC-TCGAAGATTTCGTTG
10	Rev	ATAGCGGCCGCTTAATTTTCACCGCTCAGGACAATC
11	For	ATAGAATTCGGGCTGAAGATCGACATCCA
12	For	ACGCGTGAGAACCTGTACTTCCAGGGCATGCATGTCTTCACAC-TCGAAGATTTCGTTG
13	Rev	ATAATTAATTAAATTTTCACCGCTCAGGACAATC
14	For	GGCCGGCCGAGGCGATCGTCGACATTC
15	Rev	CTCGAGGCCCTGGAAGTACAGGTTCTCATTAATTAAGTCACCAC-CGGCCCC
16	For	GGCCGGCCGGGCTGAAGATCGACATCCA
17	Rev	CTCGAGGCCCTGGAAGTACAGGTTCTCATTAATTAAC-GCCAGAATGCGTTCG
18	For	TAATTAATGAGAACCTGTACTTCCAGGGCCTCGAGGTCTTCACAC-TCGAAGATTTCGTT
19	Rev	TCTAGATTAATTTTCACCGCTCAGGACA
20	For	GGCCGGCCGGTGAAAGCTTGTTTAAGGGACC
21	Rev	TCTAGAGGTTCATGAGTTGAGTTGCTTCTTT

Circularized secreted GLuc plasmid (pBMN)

Progress was made toward a circularized secreted Gluc construct preceded by a Kozak sequence with the structure: FLAG-GlucB-TEVRS-GlucA. The GlucB-TEVRS portion of this sequence was cloned using Primers **1** and **2** from the Prescher lab vector: pBMN Fos-GlucB. A plasmid, pBMN Jun-GlucA, previously made in the Prescher lab was digested with restriction enzymes *Nsi*I and *Xho*I to prepare for cloning of the insert.

Circularized transmembrane Nluc65/66 plasmid (pBMN)

Vector backbone and insert were prepared en route to a transmembrane secreted Nluc construct preceded by a Kozak sequence with the structure: FLAG-Nluc66-TEVRS-Nluc65-CD4. The Nluc66-TEVRS-Nluc65 portion of this sequence was cloned using Primers **3-6** from the Prescher lab vector: pBMN Nluc-FL. pBMN Jun-GlucA-CD4 was digested with restriction enzymes *Eco*RI and *Pac*I to prepare for cloning of the insert.

Circularized secreted NLuc52/53 plasmid (pBMN)

Vector backbone and insert were prepared en route to a circularized secreted Nluc construct preceded by a Kozak sequence with the structure: FLAG-Nluc53-TEVRS-Nluc52. The Nluc53-TEVRS-Nluc52 portion of this sequence was cloned using Primers **7-10** from the vector pBMN Nluc-FL. pBMN Jun-GlucA was digested with restriction enzymes *Nsi*I and *Not*I to prepare for cloning of the insert.

Circularized transmembrane Nluc52/53 plasmid (pBMN)

Vector backbone and insert were prepared en route to a transmembrane secreted Nluc construct preceded by a Kozak sequence with the structure: FLAG-Nluc53-TEVRS-Nluc52-CD4. The Nluc53-TEVRS-Nluc52 portion of this sequence was cloned using Primers **8** and **11-13** from the vector pBMN Nluc-FL. pBMN Jun-GlucA-CD4 was digested with restriction enzymes *Eco*RI and *Pac*I to prepare for cloning of the insert.

Circularized secreted GLuc plasmid (pcDNA)

Vector backbone and insert were prepared en route to a circularized secreted Gluc construct preceded by a Kozak sequence with the structure: FLAG-GlucB-TEVRS-GlucA. The GlucB-TEVRS portion of this sequence was cloned using Primers **14-15** from the Prescher lab vector: pcDNA3.1 Fos GlucB. The Prescher lab vector, pcDNA3.1 Jun-GlucA was digested with restriction enzymes *Fse*I and *Xho*I to prepare for cloning of the insert.

Circularized secreted NLuc52/53 plasmid (pcDNA)

Vector backbone and insert were prepared en route to a circularized secreted Nluc construct preceded by a Kozak sequence with the structure: FLAG-Nluc53-TEVRS-Nluc52. The Nluc53-

TEVRS-Nluc52 portion of this sequence was cloned using Primers **16-19** from the vector pBMN Nluc-FL. pcDNA3.1 Jun-GlucA was digested with restriction enzymes *Fse*I and *Xba*I to prepare for cloning of the insert.

Secreted TEV plasmid (pcDNA)

Vector backbone and insert were prepared en route to inserting TEV protease into a mammalian vector. A protein expression plasmid containing TEV (courtesy of the Martin Lab, UCI) was used to clone an insert containing TEV using Primers **20-21**. pcDNA3.1 Jun-GlucA was digested with restriction enzymes *Fse*I and *Xba*I to prepare for cloning of the insert.

NTR expression and purification.

A pET-NTR plasmid was transformed into chemically competent *E. coli* BL21 cells. Over-expression and purification were performed following the procedure of Olekhnovich, *et al.*[52] The purified protein was dialyzed into 50 mM Tris, 50 mM NaCl (pH 7.5) buffer. The sample was then concentrated to 750 µg/mL via spin column (GE Healthcare, VivaSpin, 5 kDa MWCO). Gel electrophoresis and staining with coomassie blue was done to confirm expression and the resulting image is shown in **Figure D.S3**.

TEV protease expression and purification.

TEV protease overexpression and purification were performed following a protocol provided by the Martin lab (UCI). In brief, an MBP-TEV protease plasmid (courtesy of the Martin lab) was transformed into chemically competent *E. coli* BL21 cells. A single colony was grown in 25 mL of LB with ampicillin overnight. The culture was added to 1 L of LB with ampicillin and allowed to grow at 37 °C to OD_{600} = 0.7. IPTG was added to a concentration of 1 mM, and the culture was allowed to grow at 37 °C for 6.5 h. The culture was centrifuged at 6000 x g and the resulting pellet was resuspended in 35 mL of lysis buffer (300 mM NaCl, 50 mM Na_3PO_4, 10 mM imidazole, 0.05% NaN_3, 10% glycerol, pH 7.4). Cells were lysed with sonication and centrifuged at 10,000 rpm. The supernatant was applied to a Ni-NTA column overnight. The flow through was collected and then two washes were done with wash buffer (300 mM NaCl, 50 mM Na_3PO_4, 10 mM

imidazole, 0.05% NaN_3, pH 7.8). Five elutions were done with elution buffer (300 mM NaCl, 50 mM Na_3PO_4, 250 mM imidazole, 0.05% NaN_3, pH 7.8). In order to facilitate self cleavage, 1 mM DTT and 0.5 mM EDTA were added to the elution fractions and allowed to incubate for 24 h at 4 °C. The elution fractions were then applied to a Ni-NTA column for 5 min. The flow through was collected and two washes were done with wash buffer. Elution buffer was used to collect 3 elution fractions. The elution fractions were combined and dialized for 24 h in dialysis buffer (300 mM NaCl, 50 mM Na_3PO_4, 0.05% NaN_3, pH 7.8). TEV protease was concentrated to 980 µg/mL via spin column (GE Healthcare, VivaSpin, 5 kDa MWCO). Activity was confirmed using a SensoLyte 520 TEV Protease Assay Kit.

Fluorescence assays.

Luntr uncaging by expressed NTR

Luntr (100 µM) was incubated with NADH (200 µM) and NTR (2 µg/mL). Assays were performed in PBS in black 96-well microplates. Incubation and readings were conducted at 37 °C. Fluorescence spectra were acquired every 10 min over 4.5 h. An excitation wavelength of 350 nm was used and spectra were collected over 400-650 nm. Assays were performed in triplicate and the data shown in **Figure D.S4** are representative of two independent runs.

NCL stability assay

DB7 cells (50,000 cells/well in complete media) were plated in a black 96-well microplate at 100 µL volumes. Complete media (100 µL/well) was added to a number of wells on the microplate. D-Luc (250 µM) or NCL (250 µM) was added to either DB7 cells or complete media. Fluorescence was monitored at λ_{ex} 350 nm and λ_{em} 534 nm for 60 h. The microplate was incubated at 37 °C for the duration of the experiment. Assays were performed in triplicate, and the data are shown in **Figure D.2A**.

Uncaging efficiency of DB7 cell lines

DB7, DB7-Fluc, and DB7 NTR cells were plated in separate wells on a black 96-well plate (50,000 cells/well) and allowed to adhere for 24 h. Nitroimidizole caged Oregon Green (NMOG 100 µM)

was added wells containing cells or media only. Fluorescence was measured at λ_{ex} 500 nm and λ_{em} 528 nm. The microplate was incubated at 37 °C for the duration of the experiment. Assay was performed in triplicate and the data are shown in **Figure D.S5**.

Enzymatic bioluminescent assays.

Bioluminescent assay with Luntr and expressed NTR

Luntr (200 μM) was incubated with NADH (400 μM) in the presence or absence of NTR (20 μg/mL) for 0-45 min at 37 °C. Aliquots (50 μL) were added to wells on a 96-well black microplate containing purified Fluc (expressed by Krysten Jones, Prescher lab) (20 μg/mL) and ATP (1 mM) in bioluminescence imaging (BLI) buffer (20 mM Tris, 0.5 mg/mL BSA, 0.1 mM EDTA, 1 mM TCEP, 2 mM $MgSO_4$, pH 7.6) for a total volume of 100 μL. Bioluminescent images were acquired, and the data are shown in **Figure D.S3**. Assays were performed in triplicate.

Bioluminescent assay with NCL and expressed NTR

NCL (200 μM) was incubated with NADH (400 μM) in the presence or absence of NTR (provided by William Porterfield, Prescher lab) (20 μg/mL) for 0-45 min at 37 °C. Aliquots (50 μL) were added to wells on a 96-well black microplate containing purified Fluc (expressed by Krysten Jones, Prescher lab) (20 μg/mL) and ATP (1 mM) in bioluminescence imaging (BLI) buffer for a total volume of 100 μL. Bioluminescent images were acquired and the data are shown in **Figure D.2B.** Assays were performed in triplicate.

Bioluminescent assay with NCL, D-luc, and Fluc

NCL (0-1 mM) was added to wells in a black 96-well microplate containing purified Fluc (10 μg/mL), ATP (500 μM), and D-luc (100 μM) in 100 μL BLI buffer. Bioluminescent images were acquired 5 min after addition of substrate and data are shown in **Figure D.3B**. Assays were performed in triplicate.

Bioluminescent assay with D-luc

D-Luc (0-100 nM) was added to wells in a black 96-well microplate containing purified Fluc (10 μg/mL) and ATP (500 μM) in 100 μL BLI buffer. Bioluminescent images were acquired 5 min after

addition of substrate and the data is shown in **Figure D.S6.** Assays were performed in triplicate and data are representative of two independent runs.

Cell culture bioluminescent assays.

Bioluminescent assay with NCL with DB7-NTR cells and Fluc

DB7 or DB7-NTR cells (50,000 cells/well) were plated in a 96-well, clear, flat-bottomed microplate in 100 µL of media and allowed to adhere for 24 h. NCL (0-500 µM) was added to wells containing cells and allowed to incubate for 30 min at 37 °C. Media (50 µL) from the wells were removed and added to wells in a 96-well black microplate containing Fluc (20 µg/mL) and ATP (1 mM) in 50 µL BLI buffer. Bioluminescent images were acquired and the data are shown in **Figure D.3A**. The total flux from DB7 cells was subtracted from the total flux from DB7-NTR cells at each concentration of NCL. Assays were performed in triplicate.

DB7-NTR NCL uncaging assay (media transfer)

DB7-NTR cells or DB7 cells (50,000 cells/well) were plated in a clear 96-well microplate and allowed to adhere for 24 h. DB7-Fluc cells (50,000 cells/well) were plated in a black 96-well microplate and allowed to adhere for 24 h. NCL (500 µM) was added to wells containing DB7-NTR or DB7 cells and allowed to incubate at 37 °C for 0-120 min. Media (50 µL) was transferred from wells containing NCL to wells containing DB7-Fluc cells to a final volume of 100 µL. Bioluminescent images were acquired 5 min after addition of media and data are shown in **Figure D.4A**. Assays were performed in triplicate and data are representative of two independent runs.

DB7-NTR NCL uncaging assay (co-plating assay)

DB7-NTR cells (50,000 cells/well) or DB7 cells (50,000 cells/well) and DB7-Fluc cells (50,000 cells/well) were mixed and then plated in a black 96-well microplate. Cells were allowed to adhere for 24 h. NCL (250 µM) was added to wells and allowed to incubate for 0-120 min at 37 °C. Bioluminescent images were acquired and the data are shown in **Figure D.4B**. Assays were performed in triplicate and the data are representative of three independent runs.

NMR stability studies.

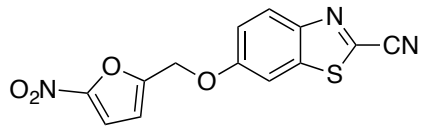
NCL was dissolved in deuterated H_2O and incubated at room temperature for 7 days. NMR spectra were collected over time. The spectra were referenced to residual solvent and data are shown in **Figure D.S2**.

General synthetic methods.

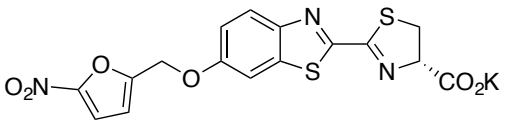
All reagents were purchased from commercial suppliers, were of analytical grade, and were used without further purification. Intermediates (**1**) and (**2**) were provided by W. B. Porterfield. Reaction progress was monitored by thin layer chromatography on EMD 60 F254 plates and visualized with UV light. Flash column chromatography with Sorbent Technologies 60 Å, 230-400 mesh silica gel was used to purify (**3**). HPLC purification of NCL was performed on a Varian ProStar equipped with 325 Dual Wavelength UV-Vis detector. Semi-preparative runs were performed using an Agilent Prep-C18 Scalar column (9.4 x 150 mm, 5 µm). Preparative runs were performed using an Agilent ZORBAX Eclipse XDB-C18 column (21.2 x 250 mm, 7 µM). Reaction vessels were flame dried prior to use. NMR spectra were acquired using a Bruker Advanced spectrometer. All spectra were acquired at 298 K and ^{1}H NMR spectra were collected at 400 Hz. Chemical shifts are reported in ppm in relation to non-deuterated NMR solvent.

Compound syntheses.

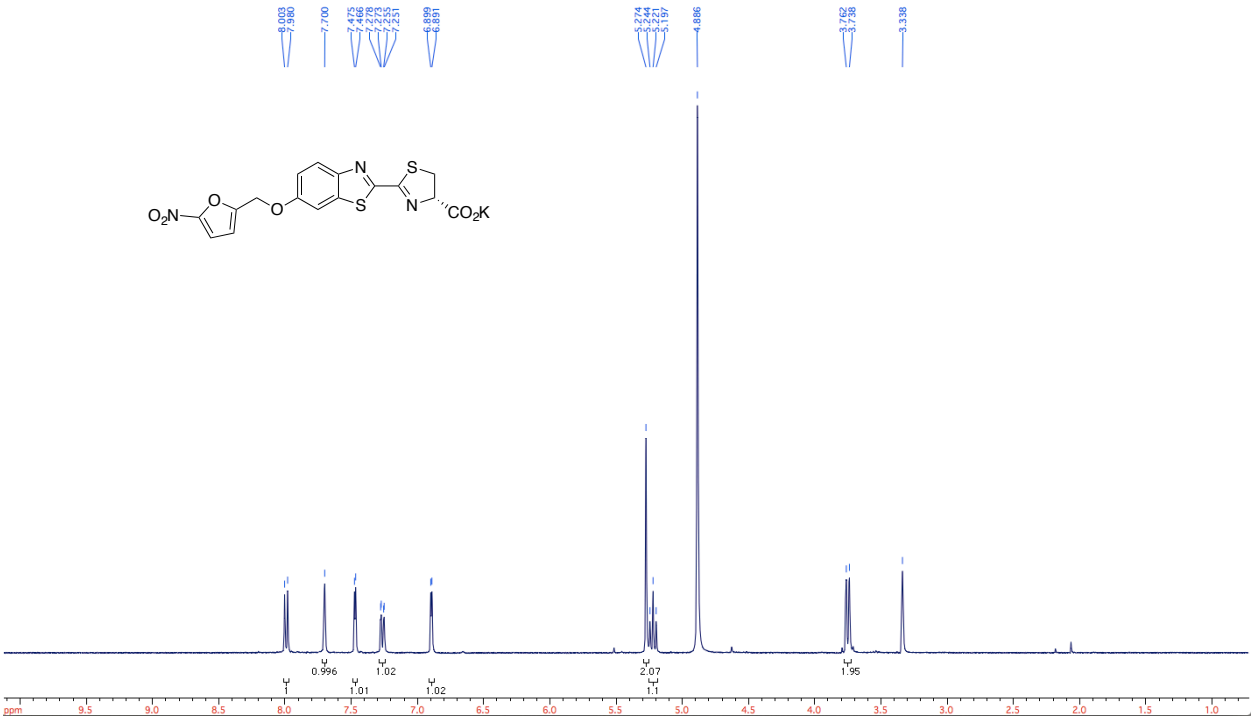
6-((5-Nitrofuran-2-yl)methoxy)benzo[*d*]thiazole-2-carbonitrile (3). Following the procedure of Vorobyeva *et al,*[23] to a flask was added triphenylphosphine (0.461 g, 1.76 mmol), 6-hydroxylbenzo[*d*]thazole-2-carbonitrile (**1**) (0.200 g, 1.13 mmol), and (5-nitrofuran-2-yl)methanol (**2**) (1.36 mL, 1.36 mmol) in 4.5 mL of THF. The reaction mixture was cooled to 0 °C. To this, 1,1'-(Azodicarbonyl)dipiperadine (0.444 g, 1.76 mmol) in 2.5 mL of THF was then added dropwise. The reaction mixture was allowed to stir at 0 °C for 1 h and then warmed to room temperature and stirred overnight under N_2. The crude material was purified by flash column chromatography eluting with 9:1 hexanes:EtOAc to provide **3** as a yellow oil (0.251 g) This material



was carried on to the next step without further purification or characterization.



Potassium(*S*)-2-(6-((5-nitrofuran-2-yl)methoxy)benzo[*d*]thiazol-2-yl)-4,5-dihydrothiazole-4-carboxylate (NCL). Following the procedure of Vorobyeva *et al,*[23] to a flask was added potassium carbonate (0.127 g, 0.916 mmol) and 6-((5-nitrofuran-2-yl)methoxy)benzo[*d*]thiazole-2-carbonitrile (**3**, crude from above) as a suspension in 2.5 mL of THF. D-Cysteine•HCl (0.160 g, 0.916 mmol) was added to the flask in 2.5 mL of H_2O. The reaction mixture was allowed to stir under N_2 at room temperature for 2 h. The reaction mixture was then concentrated *in vacuo*. The crude product was then filtered and the eluent was purified by HPLC. NCL was isolated as a light yellow powder (0.115 g, 23% over two steps). ^{1}H NMR (400 MHz, CD_3OD) δ 7.99 (d, J = 9.0 Hz, 1H), 7.70 (s, 1H), 7.47 (d, J = 3.5 Hz, 1H), 7.26 (dd, J = 8.9, 1.7 Hz, 1H), 6.90 (d, J = 3.4 Hz, 1H), 5.27 (s, 2H), 5.22 (t, J = 9.3 Hz, 1H), 3.75 (d, J = 9.3 Hz, 2H). The spectra were in accordance with reported characterization.[23]



D.7 References

1. Germain, R. N.; Robey, E. A.; Cahalan, M. D. A decade of imaging cellular motility and interaction dynamics in the immune system. *Science* **2012**, 336, 1676-1681.

2. Kjos, M.; Aprianto, R.; Fernandes, V. E.; Andrew, P. W.; van Strijp, J. A. G.; Nijland, R.; Veening, J. Bright fluorescent *Streptococcus pneumoniae* for live-cell imaging of host-pathogen interactions. *J. Bacteriol.* **2015**, 197, 807-818.

3. Cronin, M.; Akin, A. R.; Meganck, J.; Kim, J. B.; Baban, C. K.; Joyce, S. A.; van Dam, G. M.; Zhang, N.; van Sinderen, D.; O'Sullivan, G. C.; Kasahara, N.; Gahan, C. G.; Francis, K. P.; Tangney, M. High resolution *in vivo* bioluminescent imaging for the study of bacterial tumor targeting. *PloS ONE* **2012**, 7, e30940.

4. Ntziachristos, V. Fluorescence molecular imaging. *Annu. Rev. Biomed. Eng.* **2006**, 8, 1-33.

5. Alexander, S.; Weigelin, B.; Winkler, F.; Friedl, P. Preclinical intravital microscopy of the tumor-stroma interface: invasion, metastasis, and therapy response. *Curr. Opin. Cell Biol.* **2013**, 25, 659-671.

6. Close, D. M.; Hahn, R. E.; Patterson, S. S.; Baek, S. J.; Ripp, S. A.; Sayler, G. S. Comparison of human optimized bacterial luciferase, firefly luciferase, and green fluorescent protein for continuous imaging of cell culture and animal models. *J. Biomed. Opt.* **2011**, 16, e047003.

7. Prescher, J. A.; Contag, C. H. Guided by the light: visualizing biomolecular processes in living animals with bioluminescence. *Curr. Opin. Chem. Biol.* **2010**, 14, 80-89.

8. Paley, M. A.; Prescher, J. A. Bioluminescence: a versatile technique for imaging cellular and molecular features. *Med. Chem. Commun.* **2014**, 5, 255-260.

9. Contag, C. H.; Bachmann, M. H. Advances in in vivo bioluminescence imaging of gene expression. *Annu. Rev. Biomed. Eng.* **2002**, 4, 235-260.

10. Gould, S. J.; Subramani, S. Firefly luciferase as a tool in molecular and cell biology. *Anal. Biochem.* **1988**, 175, 5-13.

11. Schipper, M. L.; Patel, M. R.; Gambhir, S. S. Evaluation of firefly luciferase bioluminescence mediated photodynamic toxicity in cancer cells. *Mol. Imaging Biol.* **2006**, 8, 218-225.

12. Li, H.; Li, J. Z.; Helm, G. A.; Pan, D. Non-invasive imaging of firefly luciferase reporter gene expression using bioluminescence imaging in human prostate cancer models. *Biotechnol. Appl. Biochem.* **2007**, 46, 179-184.

13. Badr, C. E.; Tannous, B. A. Bioluminescence imaging: progress and applications. *Trends Biotechnol.* **2011**, 29, 624-633.

14. Porterfield, W. B.; Jones, K. A.; McCutcheon, D. C.; Prescher, J. A "caged" luciferin for

imaging cell-cell contacts. *J. Am. Chem. Soc.* **2015**, 137, 8656-8659.

15. Li, J.; Chen, L.; Du, L.; Li, M. Cage the firefly luciferin! A strategy for developing bioluminescent probes. *Chem. Soc. Rev.* **2013**, 42, 662-676.

16. Yao, H.; So, M.; Rao, J. A bioluminogenic substrate for in vivo imaging of B-lactamase activity. *Angew. Chem. Int. Ed.* **2007**, 46, 7031-7034.

17. Rush, J. S.; Beatty, K. E.; Bertozzi, C. R. Bioluminescent probes of sulfatase activity. *ChemBioChem* **2010**, 11, 2096-2099.

18. Sellmyer, M. A.; Bronsart, L.; Imoto, H.; Contag, C. H.; Wandless, T. J.; Prescher, J. A. Visualizing cellular interactions with a generalized proximity reporter. *Proc. Natl. Acad. Sci. USA* **2013**, 110, 8567-8572.

19. Knox, R. J.; Friedlos, F.; Boland, M. P. The bioactivation of CB 1954 and its use as a prodrug in antibody-directed enzyme producing prodrug therapy (ADEPT). *Cancer Metastasis Rev.* **1993**, 12, 195-212.

20. Pisharath, H.; Rhee, J. M.; Swanson, M. A.; Leach, S. D.; Parsons, M. J. Targeted ablation of beta cells in the embryonic zebrafish pancreas using *E. coli* nitroreductase. *Mech. Dev.* **2007**, 124, 218-229.

21. Williams, E. M.; Little, R. F.; Mowday, A. M.; Rich, M. H.; Chan-Hyams, J. V. E.; Copp, J. N.; Smaill, J. B.; Patterson, A. V.; Ackerley, D. F. Nitroreductase gene-directed enzyme prodrug therapy: insights and advances toward clinical utility. *Biochem. J.* **2015**, 421, 131-353.

22. Dachs, G. U.; Hunt, M. A.; Syddall, S.; Singleton, D. C.; Patterson A. V. Bystander or no bystander for gene directed enzyme prodrug therapy. *Molecules* **2009**, 14, 4517-4545.

23. Vorobyeva, A. G.; Stanton, M.; Godinat, A.; Lund, K. B.; Karateev, G. G.; Francis, K. P.; Allen, E.; Gelovani, J. G.; McCormack, E.; Tangney, M.; Dubikovskaya, E. A. Development of a bioluminescent nitroreductase probe for preclinical imaging. *PloS ONE* **2015**, 10, e0131037.

24. Shinde, R.; Perkins, J.; Contag, C. H. Luciferin derivatives for enhanced in vitro and in vivo bioluminescence assays. *Biochemistry* **2006**, 45, 11103-11112.

25. Ando, Y.; Niwa, K.; Yamada, N.; Enomoto, T.; Irie, T.; Kubota, H.; Ohmiya, Y.; Akiyama, H. Firefly bioluminescence quantum yield and color change by pH-sensitive green emission. *Nat. Photon.* **2008**, 2, 44-47.

26. Adams, S. T.; Miller, S. C. Beyond D-luciferin: expanding the scope of bioluminescence imaging in vivo. *Curr. Opin. Chem. Biol.* **2014**, 12, 112-120.

27. Zinn, K. R.; Chaudhuri, T. R.; Szafran, A. A.; O'Quinn, D.; Weaver, C.; Dugger, K.; Lamar, D.; Kesterson, R. A.; Wang, X.; Frank, S. J. Noninvasive bioluminescence imaging in small animals. *ILAR J.* **2008**, 49, 103-115.

28. Erez, Y.; Presiado, I.; Gepshtein, R.; da Silva, L. P.; Esteves da Silva, J. C. G.; Huppart,

D. Excited-state proton transfer of firefly dehydroxyluciferin. *J. Phys. Chem. A* **2012**, 116, 7452-7461.

29. Cardiff, R. D. Contributions of mouse biology to breast cancer research. *Comp. Med.* **2002**, 52, 12-31.

30. Grimm, J. B.; Gruber, T. D.; Ortiz, G.; Brown, T. A.; Lavis, L. D. Virginia orange: a versatile, red-shifted fluorescein scaffold for single and dual-input fluorogenic probes. *Bioconjugate Chem.* **2016**, 27, 474-480.

31. Ribeiro, C.; Esteves da Silva, J. C. G. Kinetics of inhibition of firefly luciferase by oxyluciferin and dehydroluciferyladenylate. *Photochem. Photobiol. Sci.* **2008**, 7, 1085-1090.

32. Leitao, J. M. M.; Esteves da Silva, J. C. G. Firefly luciferase inhibition. *J. Photochem. Photobiol. B* **2010**, 101, 1-8.

33. Grohmann, M.; Paulmann, N.; Fleischhaur, S.; Vowinckel, J.; Priller, J.; Walther, D. J. A mammalianized synthetic nitroreductase gene for high-level expression. *BMC Cancer* **2009**, 9, 2407-2414.

34. Slomovic, S.; Collins, J. J. DNA sense-and-respond protein modules for mammalian cells. *Nat. Methods* **2015**, 10, 933-938.

35. Jones, K. A.; Li, D. J.; Hui, E.; Sellmyer, M. A.; Prescher, J. A. Visualizing cell proximity with genetically encoded bioluminescent reporters. *ACS Chem. Biol.* **2015**, 10, 933-938.

36. Troy, T.; Jekic-McMullen, D.; Sambucetti, L.; Rice, B. W. Quantitative comparison of the sensitivity of detection of fluorescent and bioluminescent reporters in animal models. *Mol. Imaging* **2004**, 3, 9-23.

37. Zhao, H.; Doyle, T. C.; Coquoz, O.; Kalish, F.; Rice, B. W.; Contag, C. H. Emission spectra of bioluminescent reporters and interaction with mammalian tissue determines the sensitivity of detection in vivo. *J. Biomed. Opt.* **2005**, 10, e041210.

38. Rolfe, M. D.; Rice, C. J.; Lucchini, S.; Pin, C.; Thompson, A.; Cameron, A. D. S.; Alston, M.; Stringer, M. F.; Betts, R. P.; Baranyi, J.; Peck, M. W.; Hinton, J. C. D. Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation. *J. Bacteriol.* **2012** 194, 686-701.

39. Laasko, K.; Koskenniemi, K.; Koponen, J.; Kankainen, M.; Surakka, A.; Salusjarvi, T.; Auvinen, P.; Savijoki, K.; Nyman, T. A.; Kalkkinin, N.; Tynkkynen, S.; Varmanan, P. Growth phas-associated changes in the proteome and transcriptome of *Lactobacillus rhamnosus* GG in industrial-type whey medium. *Microb. Biotechnol.* **2011**, 4, 746-766.

40. Lancellotti, M.; Brocchi, M.; da Silveira, W. D. Bacteria-induced apoptosis: an approach to bacterial pathogenesis. *Braz. J. Morphol. Sci.* **2006**, 23, 75-86.

41. Furman, D.; Davis, M. M. New approaches to understanding the immune response to vaccination and infection. *Vaccine* **2015**, 33, 5271-5281.

42. Cheraghi, R.; Hosseinkhani, S.; Davoodi, J.; Nazari, M.; Amini-Bayat, Z.; Karimi, H.;

Shamseddin, M.; Gheidari, F. Structural and functional effects of circular permutation on firefly luciferase: in vitro assay of caspase 3/7. *Int. J. Biol. Macromol.* **2013**, 58, 336-342.

43. Fan, F.; Binkowski, B. F.; Butler, B. F.; Stecha, P. F.; Lewis, M. K.; Wood, K. V. Novel genetically encoded biosensors using firefly luciferase. *ACS Chem. Biol.* **2008**, 3, 346-351.

44. Kim, S. B.; Sato, M.; Tao, H. Split Gaussia luciferase-based bioluminescence template for tracing protein dynamics in living cells. *Anal. Chem.* **2009**, 81, 67-74.

45. Tannous, B. A.; Kim, D.; Fernandez, J. L.; Weissleder, R.; Breakefield, X. O. Codon-optimized Gaussia luciferase cDNA for mammalian gene expression in cluture and in vivo. *Mol. Ther.* **2005**, 11, 435-443.

46. Hall, M. P.; Unch, J.; Binkowski, B. F.; Valley, M. P.; Butler, B. L.; Wood, M. G.; Otto, P.; Zimmerman, K.; Vidugiris, G.; Machleidt, T.; Robers, M. B.; Benink, H. A.; Eggers, C. T.; Slater, M. R.; Meisenheimer, P. L.; Klaubert, D. H.; Fan, F.; Encell, L. P.; Wood, K. V. Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS Chem. Biol.* **2012**, 7, 1848-1857.

47. Roh-Johnson, M.; Bravo-Cordero, J. J.; Patsialou, A.; Sharma, V. P.; Guo, P.; Liu, H.; Hodgson, L.; Condeelis, J. Macrophage contact induces RhoA GTPase signaling to trigger tumor cell intravasation. *Oncogene* **2014**, 33, 1676-1681.

48. Remy, I.; Michnick, S. W. A highly sensitive protein-protein interaction assay based on Gaussia luciferase. *Nat. Methods* **2006**, 3, 977-979.

49. Zhao, J.; Nelson, T. J.; Vu, Q.; Truong, T.; Stains, C. I. Self-assembling NanoLuc luciferase fragments as probes for protein aggregation in living cells. *ACS Chem. Biol.* **2016**, 11, 132-138.

50. Verhoef, L. G. G. C.; Mattioli, M.; Ricci, F.; Li, Y.; Wade, M. Multiplex detection of protein-protein interactions using a next generation luciferase reporter. *BBA Mol. Cell Res.* **2016**, 1863, 284-292.

51. Kapust, R. B.; Tozser, J.; Copeland, T. D.; Waugh, D. S. The P1 specificity of tobacco etch virus protease. *Biochem. Biophys. Res. Commun.* **2002**, 294, 949-955.

52. Olekhnovich, I. N.; Goodwin, A.; Hoffman, P. S. Characterization of the NAD(P)H oxidase and metronidazole reductase activities of the RdxA nitroreductase of Helicobacter pylori. *FEBS J.* **2009**, 12, 3354-3356.