UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Chemical Activation of Nucleic Acids Catalyzed by RNA: Implications for Origin of Life Scenarios

Permalink

https://escholarship.org/uc/item/2st659w2

Author

Arriola, Joshua

Publication Date

2022

Peer reviewed|Thesis/dissertation

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

UNIVERSITY OF CALIFORNIA SAN DIEGO

Chemical Activation of Nucleic Acids Catalyzed by RNA: Implications for Origin of Life Scenarios

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

in

Biochemistry & Molecular Biophysics

by

Joshua T. Arriola

Committee in charge:

Professor Ulrich F. Müller, Chair
Professor Simpson Joseph
Professor Jens Lykke-Andersen
Professor Dionicio Siegel
Professor Navtej Toor

2022

Copyright

Joshua T. Arriola, 2022

All rights reserve

The Dissertation of Joshua T. Arriola is approved and is acceptable in quality and form for publication on microform and electronically.

University of California San Diego

2022

Epigraph

So long and thanks for all the fish.

-Douglas Adams, The Hitchhiker's Guide to the Galaxy, Ch. 23

Table of Contents

List of Figures

Acknowledgements

This dissertation was made possible by many different people. First, I want to thank the graduate students who were directly involved in making the Müller lab a wonderful place to work and learn: Arvin Akoopie, Kevin Sweeney, Xu Han, Debolina Sarkar, and Tommy Le. Next, I want to thank Uli Müller for his mentorship. Uli approaches his role as a Principal Investigator and as a teacher with professionalism, scientific rigor, and most of all kindness.

I would not have made it this far without the love and support of my parents, Edwin and Maureen Arriola, who fostered in me a curiosity for the natural world from a young age and gave me the best education they could no matter the cost. Thank you, Jake Arriola, for being my best friend since the day you were born, for understanding me better than anyone else, and for introducing me to the Cosmere: Life before Death, Strength before Weakness, Journey before Destination!

I'm not sure I would have made it through graduate school if I had not met my ex-girlfriend/current fiancée, Anna Guzikowski, in the neighboring Zid Lab. I have not met anyone with her gentle, and loving spirit. She shares my love of nature and has the remarkable ability to pack-bond with anything. The pets we have welcomed into our pack include a black cat named Herbert, a tabby cat named Filbert, a dog named Wally, a water dragon named Bruce, and countless houseplants and succulents. Every day with her is an adventure and I can't wait for the next chapter.

Thank you to all the friends I've made in graduate school. Even if you are not mentioned here by name, know that am grateful for your friendship.

Chapter 2, in full, is a reprint of the material as it appears in Arriola, J. T.; Müller, U. F. “A combinatorial method to isolate short ribozymes from complex ribozyme libraries”, *Nucleic Acids Res.*, 2020. The dissertation author was the first author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in Lin, H.; Jimenez, E. I.; Arriola, J. T.; Müller, U. F.; Krishnamurthy, R.; “Concurrent Prebiotic Formation of Nucleoside-Amidophosphates and Nucleoside-Triphosphates Potentiates Transition from Abiotic to Biotic Polymerization”, *Angew. Chem. Int. Ed.*, 2021. The dissertation author was the third author of this paper. He contributed by investigating if ribozymes could use crude mixtures of nucleoside triphosphates generated by diamidophosphate.

Chapter 4, in full, is a reprint of the material as it appears in Akoopie, A.; Arriola, J. T.; Magde, D.; Müller, U. F. “A GTP synthesizing ribozyme selected by metabolic coupling to an RNA polymerase ribozyme”, *Sci. Adv.,* 2021. The dissertation author was the second author of this paper. He contributed by characterizing a novel guanosine triphosphate synthase ribozyme.

Vita

Education

University of California San Diego — December 2022
Ph.D. in Biochemistry & Molecular Biophysics

University of California San Diego — March 2018
M.S. in Chemistry

University of California Santa Barbara — June 2014
B.S. in Chemistry

Publications

Arriola, J. T.; Valdivia, E. M.; Schellinger, J. L.; Müller, U. F. “Prebiotically plausible Peptides can support Ribozyme Function”, manuscript in preparation

Akoopie, A.; Arriola, J. T.; Magde, D.; Müller, U. F. “A GTP synthesizing ribozyme selected by metabolic coupling to an RNA polymerase ribozyme”, *Sci. Adv.,* 2021

Lin, H.; Jimenez, E. I.; Arriola, J. T.; Müller, U. F.; Krishnamurthy, R.; “Concurrent Prebiotic Formation of Nucleoside-Amidophosphates and Nucleoside-Triphosphates Potentiates Transition from Abiotic to Biotic Polymerization”, *Angew. Chem. Int. Ed.*, 2021

Arriola, J. T.; Müller, U. F. “A combinatorial method to isolate short ribozymes from complex ribozyme libraries”, *Nucleic Acids Res.*, 2020

Lubin, P.; Hughes, G. B.; Bible, J.; Bublitz, J.; Arriola, J.; Motta, C.; Suen, J.; Johansson, I.; Riley, J.; … O’Neille, H. “Towards Directed Energy Planetary Defense”, *Opt. Eng.*, 2014

Abstract of the Dissertation

Chemical Activation of Nucleic Acids by Catalytic RNA: Implications for Origin of Life Scenarios

by

Joshua T. Arriola

Doctor of Philosophy in Biochemistry & Molecular Biophysics

University of California San Diego, 2022

Professor Ulrich F. Müller, Chair

An early stage of life likely used catalytic RNAs (ribozymes) to support self-replication and a metabolism. Because the abiotic polymerization of RNA is challenging, the formation of long RNA polymers is unlikely on the prebiotic Earth. In addition, ribozymes likely co-evolved with peptides since the formation of amino acids under prebiotic conditions occurs readily and peptide bond formation occurs under prebiotic conditions. Like in today's organisms, nucleoside 5'-triphosphates (NTPs) may have been used as central building blocks for self-replication and in metabolism. In early

stages of life, NTPs could have been generated by reacting nucleoside 5'-hydroxyl groups with the prebiotically plausible molecule cyclic trimetaphosphate (cTmp). Our lab previously showed that this chemistry can be catalyzed by ribozymes, but these ribozymes generated only RNA 5'-triphosphates and not free NTPs. In this work, a combinatorial method that can be used to identify short catalytic motifs from pools of active ribozymes is demonstrated. Diamidophosphate is investigated as an alternative reactive phosphorus species to generate NTPs and it is shown that ribozymes can use NTPs from this reaction. A ribozyme capable of synthesizing GTP from cTmp and guanosine is characterized. And finally, a novel interaction between a prebiotically plausible peptide and a self-triphosphorylating ribozyme is identified and characterized.

Chapter 1

Introduction

1.1 The Early Earth

Any discussion of the origin of life must first contend with the environmental conditions of the early Earth. According to geological evidence, the Earth-Moon system formed around 4.5 billion years ago and it is hypothesized that a large impact formed the Moon.[1–4] Any life that existed on the surface of the planet would have been extinguished because Earth's surface was converted to molten rock by the impact. Analysis of ancient minerals called zircons suggests the presence of liquid water on the surface of the planet 4.4 billion years ago.[5,6] By about 4 billion years ago, the surface of the planet was likely submerged by a global ocean. The elevation of continental mountain ranges was limited on the early Earth because continental crust was hotter and weaker than it is now. Since the core was hotter, there would have been a greater flux of mantle plumes that could cause volcanic hotspots to form ocean islands on sufficiently old seafloor.[7,8] In addition, there is evidence supporting the presence of microbial life between 3.5-3.4 billion years ago.[9,10] Determining the biogenicity of microfossils is difficult and claims of biogenicity must be supported by multiple lines of evidence that show that the morphology and chemical composition of the fossil are biogenic in origin. The biogenicity of 3.4 billion year old fossils in Wacey, et al. (2011) is supported by cell-like morphology of the fossils and geochemical signatures in the fossils including the co-occurrence of carbon and nitrogen, the disordered structure of

carbon (inconsistent with abiotic graphite), and a carbon isotope ratio consistent with biological carbon fixation. The biogenicity of microbially induced sediment structures in Nofke, et al. (2011) is supported by the presence of disordered carbon and relies more on morphological evidence: the structures closely resemble modern and ancient structures, they are distinct from surrounding sediment structures, and microscopic biotextures are consistent with other fossils from the Archean and Proterozoic eons. Given these time constraints, a transition from non-life to life must have occurred within 1-1.5 billion years after the formation of the Earth-Moon system.

In the 1920s, Oparin and Haldane independently proposed origin of life scenarios to explain what this transition could have looked like.[11–15] While there are some differences between the two proposals, the important similarity is that a wide range of complex organic molecules could be generated under a reducing atmosphere in the presence of a sufficient supply of energy (UV or lightning for example). These complex organic molecules would provide the “prebiotic soup” in which life as we know it could arise. There is some question of whether or not the early Earth’s atmosphere was composed of reduced species, and recent work suggests that the atmosphere was comprised of oxidized CO_2 and either H_2O or N_2.[16–19] This paradox could be resolved by the idea that a large impactor could have generated a sufficiently reducing atmosphere to generate organic molecules.[20] The seminal work by Miller is laboratory evidence in support of the Oparin-Haldane hypothesis. Miller showed that under a reducing atmosphere containing ammonia, methane, and water amino acids could be generated in the presence of a spark discharge.[21] In the presence of sulfur and using modern

chemical analysis techniques, the Bada group showed that Miller-type spark discharge experiments could generate a variety of canonical and non-canonical amino acids.[22]

1.2 RNA World

The central dogma of biology is that DNA encodes genetic information which is transcribed into RNA and translated into proteins. It describes an information transfer from DNA, which is resistant to hydrolysis but catalytically inert (for the most part), via mRNA to proteins, which are capable of catalysis. Rich, Woese, Crick, and Orgel proposed the idea that the first self-replicating system consisted of RNA.[23–26] Such a system would rely on catalysts made of RNA and would have preceded genome encoded protein catalysts. The RNA World hypothesis – as it is used by most labs in this field of research today – states that RNA played an important role in the emergence of life on Earth: RNA acted both as genetic storage and as the only genome-encoded catalyst. Because RNA possesses both a genotype and a phenotype, it can be subject to direct evolution. In the early 1980s, the Cech and Altman groups showed that nucleic acids were capable of catalysis.[27,28] These experiments showed for the first time that biomolecular catalysis was not restricted to proteins. Second, peptide bond formation in the ribosome is catalyzed by RNA.[29] This suggests that RNA was involved in the emergence of encoded peptide synthesis in early stages of life. Additional support comes from the fact that coenzymes – organic molecules essential for the function of many enzymes – are usually nucleotides or can be derived from nucleotides.[30] This implies there was an early stage in the evolution of life during which RNA was more

dominant than today Since RNA can have both a genotype and phenotype, it could fill both roles that DNA and proteins fulfill today.

Nucleobases were likely present on the prebiotic Earth. Guanine, adenine, cytosine, uracil, and thymine have been detected in carbonaceous meteorites.[31] Nucleobases could also be generated from prebiotically plausible reactions, for example, Oró showed in 1960 that adenine could be synthesized from cyanide.[32] While the nucleobases appear to readily form from abiotic processes, nucleoside synthesis is more complicated because the addition of a nucleobase to ribose is inefficient, and a new stereocenter is generated. The Sutherland group showed that pyrimidine ribonucleotides could be synthesized from cyanamide, cyanoacetylene, glycolaldehyde, glyceraldehyde and inorganic phosphate.[33] In addition, the Carell group showed that the purine nucleoside could be synthesized by the condensation of formamidopyrimidines and sugars.[34] Recent work by the Carrell group has demonstrated a unified synthesis of pyrimidines and purines from small prebiotically plausible molecules and driven by wet-dry cycles.[35]

Despite the many observations that support it, there are several problems that underlie the RNA world hypothesis. One issue is the unlikelihood of the prebiotic polymerization of long catalytic RNA (ribozymes). Since the polymerization of RNA monomers is thermodynamically unfavorable in aqueous environments, RNA monomers must be activated by a chemical activating group. Activating groups that have been explored include adenylate, cyanide, imidazole, 2-methyl imidazole, and triphosphates.[36] Of these, triphosphates are the most prebiotically likely[37] and show the clearest evolutionary path to modern organisms who rely on NTPs in metabolism.

Huang and Ferris showed that RNA polymers up to 40 nucleotides in length could be synthesized on montmorillonite mineral surfaces when uridine-5'-phosphoro-1-methyladenine was used as a monomer. The activating group, 1-methyladenine, could be synthesized prebiotically from methylamine and adenine.[38] A follow up experiment showed that heteropolymers up to 50 nucleotides in length containing adenine and uridine could be synthesized.[39] The lengths of oligomers containing A, U, G, and C could not be determined by gel electrophoresis but the HPLC data indicated that these heteropolymers were shorter than the polymers containing only A and U. A 50 nt RNA polymer then represents the longest so far achieved under lab-based model systems for a prebiotic environment. In contrast, the only known RNA polymerase is 200 nt long, but can only synthesize RNA polymers 96 nt long in a sequence general manner.[40–42]

Additionally, RNA is subject to base catalyzed hydrolysis and degradation due to in-line attack by its 2' hydroxyl on the vicinal phosphodiester bond, leading to a 2',3'-cyclic phosphate and a free 5'-hydroxyl group.[43] Longer polymers would be more susceptible to inactivation due to hydrolysis and RNA polymers would require half-lives significantly longer than the doubling times of RNA world organisms. Therefore, it is important to investigate the shortest possibly ribozymes. Chapter 2 details a method that can be used to identify short catalytic motifs from pools of ribozymes.

1.3 Ribozymes

Natural ribozymes include the ribosome, RNase P, HDV ribozyme, hammerhead ribozyme, self-splicing introns, hairpin ribozyme, twister ribozyme, pistol ribozyme,

hatchet ribozyme, the spliceosome, and SAM-dependent methyltransferase ribozyme.[44,45] Apart from the ribosome and the SAM-dependent methyltransferase ribozyme, other natural ribozymes catalyze phosphate transfer reactions. *In vitro* selections have allowed researchers to identify ribozymes with much more varied chemical catalysis capabilities. These ribozymes include RNA polymerase ribozymes, ligase ribozyme, Diels-Alderase ribozyme, GTP synthase, self-triphosphorylating ribozyme, self-alkylating ribozymes, and methyltransferase ribozymes.[36,46–48] Catalytic DNA have also been identified with *in vitro* selections.[49] Riboswitches may illustrate how RNA could catalyze a variety of reactions in early life. One hypothesis states that riboswitches may be remnants of ancient ribozymes whose functions have now been taken over by proteins, and what remains are the binding and gene expression controls.[50] The identification of a pre-Q1 riboswitch with self-methylation activity supports the idea that some riboswitches can possess catalytic ability.[51] Intriguingly most natural riboswitches bind RNA derived compounds further supporting the idea that riboswitches played some role in the RNA World.[52]

1.4 Energy Sources and Phosphate

In extant biology, phosphorous plays a key role in many biological processes. For example, organisms use adenosine triphosphate (ATP) as a form of energy currency where the dephosphorylation of ATP is coupled to energetically unfavorable reactions to drive those reactions forward. Even if the earliest forms of life relied on a different energy source to catalyze replication of their genomes, there would have been a

transition to phosphorylated nucleosides in the course of evolution. Most of the phosphorus on Earth's surface is in the +5 oxidation state, which forms compounds of low solubility in the presence of calcium, and therefore exists mostly in the form of a highly insoluble mineral called apatite.[53] It is important to consider the availability of water soluble phosphorus on the surface of the Earth, for example, phosphorus in the +3 oxidation state is about 1000 times more soluble than phosphorus in the +5 oxidation state.[54] Reduced phosphorus species have been identified on meteorites, so it is likely that most of the water soluble phosphorus was delivered to Earth exogenously.[53,55] Geological evidence suggests that phosphite (HPO_3^{2-}) with phosphorus in the +3 state was abundant in the ocean prior to 3.5 billion years ago.[54] There are few natural processes that can reduce phosphate, with the exception of lightning strikes and hydrothermal processes.[55]

Given the importance of NTPs in extant biology, it is important to show how NTPs could be formed on the early Earth. Pasek, et al. (2008) have investigated how phosphate could have been produced from meteoric phosphorus (figure 1.1). Iron nickel phosphide is a component of iron meteorites (~0.5% by weight).[56] When in contact with water, this iron nickel phosphide (oxidation state formally <0) erodes from the meteorite and is oxidized to phosphite. Linear and circular orthophosphates can be generated via a series of radical reactions.[57] Reactive oxygen species, such as hydroxyl radicals, can be generated from the photolysis of atmospheric water[58] or from the hydrological abrasion of quartz[59]. Cyclic trimetaphosphate generated via this reaction pathway can be used to generate nucleoside triphosphates under prebiotic conditions.[60] Recent work has measured the nonenzymatic synthesis of NTPs at near-neutral pH and found it is

inefficient (on the order of 10^{-6} M^{-1} s^{-1})[61], so RNA World organisms would benefit from a ribozyme that could synthesize NTPs. Chapter 4 details a novel GTP synthase ribozyme.

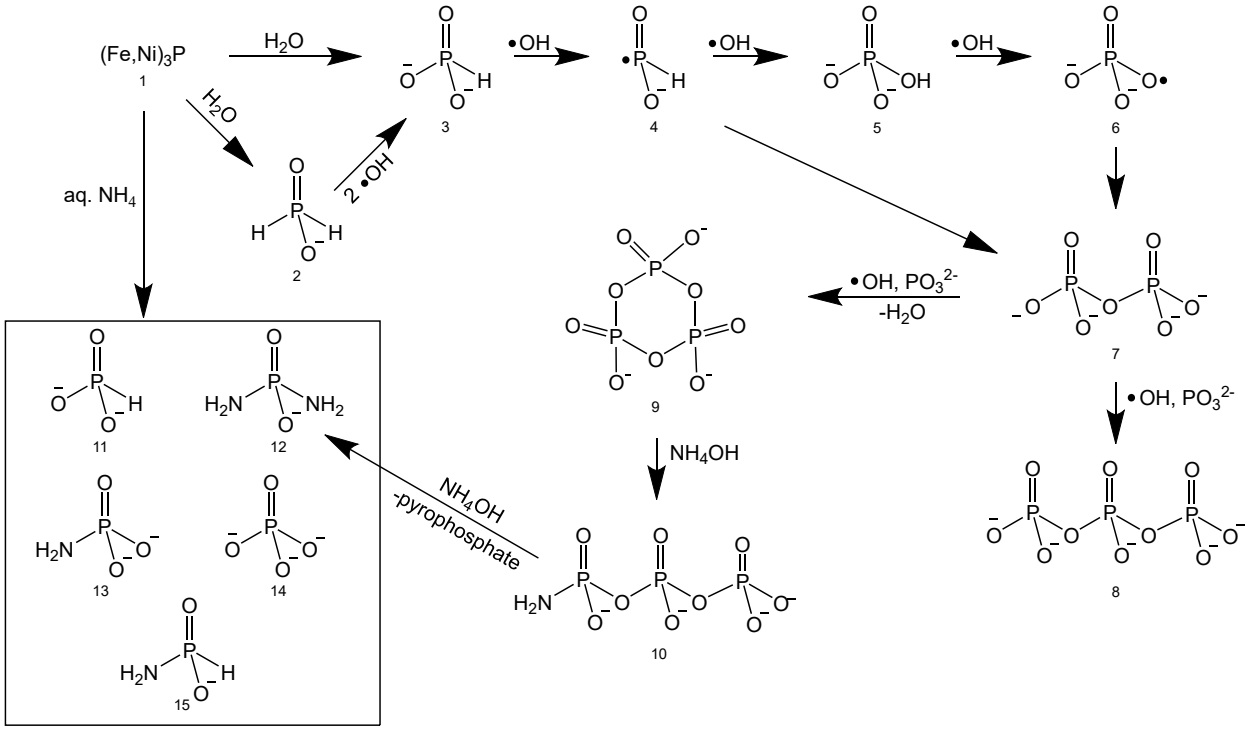


Figure 1.1 The synthesis of condensed phosphates and amidophosphates from schreibersite. (Fe, Ni)$_3$P, commonly known as schreibersite (1), has been identified on iron meteorites and could have been an exogenous source of reactive phosphorus species. Weathering of iron meteorites containing schreibersite could produce reduced phosphorus species (2-3), which could then be oxidized to phosphate (5) and condensed phosphates (7-9). Cyclic trimetaphosphate (9) is a reactive orthophosphate that is used in this work as a substrate to triphosphorylate nucleosides. In the presence of aqueous ammonium, schreibersite can produce amidophosphates (11-15), including diamidophosphate (12). Diamidophosphate can also be formed from cyclic trimetaphosphate in the presence of ammonium hydroxide and can mediate the triphosphorylation of nucleosides. The pathways illustrated above are adapted from Pasek, et al. (2008) and Gibard, et al. (2019).

Orthophosphate (figure 1.1, 5) is insoluble when in the presence of high concentrations of divalent cations as would have been present in the Earth's ocean. Cyclic trimetaphosphate (figure 1.1, 9) is soluble in water. Reduced phosphate is much more soluble in the presence of divalent cations and one form of reduced phosphate is

diamidophosphate. Diamidophosphate can be generated from nickel iron phosphide in the presence of liquid ammonia and from cyclic trimetaphosphate in the presence of ammonium hydroxide (figure 1.1).[62] In Chapter 3, DAP is shown to generate nucleoside triphosphates from nucleoside monophosphates. While the synthesis of NTPs from nucleosides using DAP appears to be challenging, DAP does not suffer from the insolubility issues that cyclic trimetaphosphate suffers from. However, recent work has shown that the presence of reduced iron ($Fe^{2+)}$ may substantially increase the solubility of most phosphate species.[63] Therefore, the presence of aqueous and reactive phosphate species was likely available on the early Earth and early organisms could have used these phosphate species to generate triphosphorylated nucleosides.

1.5 Peptides and RNA/Peptide Interactions

As discussed previously, amino acids can be synthesized under prebiotic conditions using Miller-type spark discharge experiments. Since Miller's original spark discharge experiments, several groups have investigating amino acid synthesis under a variety of assumed prebiotic atmospheres with different energy sources and have demonstrated the production of amino acids.[64] A reanalysis of some of Miller's original samples by the Bada group showed the production of over 20 different amino acids including noncanonical and natural amino acids.[22,65] In addition, amino acids could have been delivered exogenously since multiple amino acids have been identified on carbonaceous chondrite meteorites.[66] Racemic mixtures of amino acids were likely present in the early stages of life: spark discharge experiments produce racemic

mixtures of amino acids and both D and L amino acids have been identified on meteorites. This may pose a problem for the interaction between prebiotically plausible peptides and RNA because binding of peptides to double stranded DNA is stereospecific.[67] However, it is not clear if peptides that contain non-canonical amino acids such as α,β,γ-aminobutyric acid and α-aminoisobutyric acid have different stereochemical requirements. Second, if only a few residues need to be oriented in the correct way to aid, in RNA folding then only a few amino acids may be required to be in a specific stereoisomer. Third, prebiotically plausible peptides contain a large fraction of glycine, which does not have a stereocenter. Tagami, et al. (2017) have suggested that racemic mixtures of amino acids do not aid function of a specific ribozyme; however, this conclusion was drawn from experiments with a ribozyme selected and evolved in the absence of peptides.[68] It is notable that the biological cationic amino acids (Arg, Lys, His) and most aromatic amino acids (Tyr, Trp) were not detected in these Miller type experiments suggesting that these amino acids did not appear prior to the evolution. The non-biological, cationic amino acid ornithine was detected in Miller-type experiments although in low yields. Phenylalanine was detected in small quantities in Johnson, et al. (2008) so may have been very rare on the early Earth.[65]

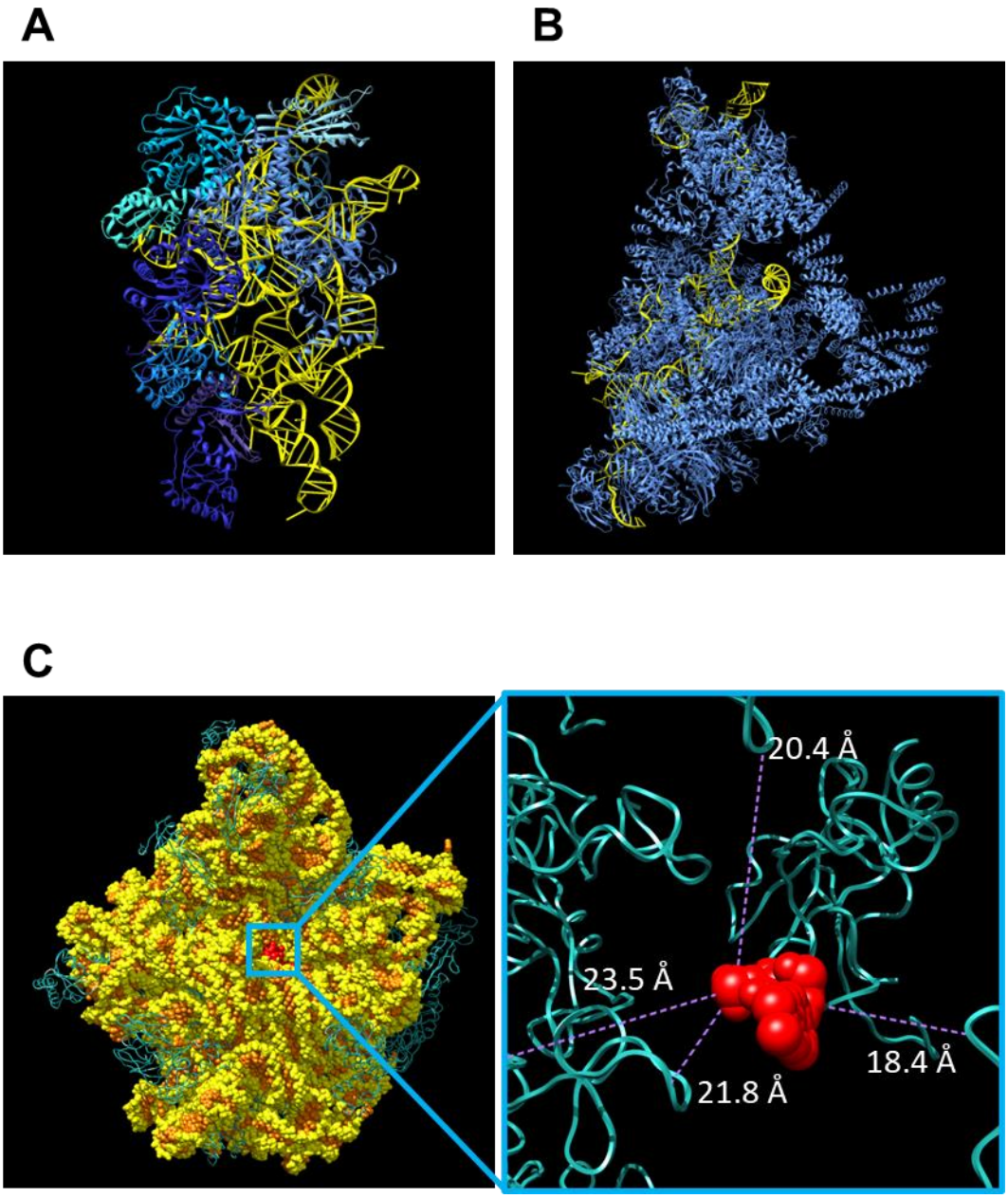


Figure 1.2 Examples of ribonucleoproteins in modern biology. (A) In yeast RNase P, the protein (blue ribbons) component wraps around RNA (yellow ribbons) thereby stabilizing it. (B) In the spliceosome, a protein called Prp8 (blue ribbons) acts a scaffold for RNA (yellow ribbons). (C) The ribosome consists of ribosomal proteins (blue ribbons) and ribosomal RNA (yellow and orange spheres) and is responsible for the synthesis of proteins. The inset shows the catalytic site (red) of peptide bond formation. There are no ribosomal proteins within 18 Å of the catalytic site indicating that only ribosomal RNA mediate peptide bond formation. The structure of the ribosome indicates that it is a ribozyme, which suggests that RNA catalysts preceded genome-encoded proteins.

Several macromolecules in extant biology are ribonucleoproteins (RNPs) in which RNA is bound to RNA-binding proteins (figure 1.2). In some RNPs, the RNA acts as a catalyst and protein acts as a structural element. For example, RNase P is responsible for processing the 5' end of pre-tRNAs. In eukaryotes and archea, the protein component of RNase P likely plays a structural role and helps to stabilize RNA.[69] The ribosome carries out protein synthesis in all living organisms and the structure of the ribosome shows that no proteins are involved in the catalytic step of peptide bond formation but dozens of ribosomal proteins aid in ribosomal function.[70] Additionally, the spliceosome is a highly dynamic RNP responsible for processing pre-mRNA. The spliceosomal RNA (snRNA) coordinates two metal ions in order to catalyze two transesterification reactions that splice together exons, thereby, producing mature mRNA. Spliceosomal proteins are involved in rearrangment of the snRNA during each step of splicing and play a structural and regulatory role in splicing.[71] As discussed previously, amino acids and peptides are synthesized under prebiotic conditions. Several examples exist in biology where proteins support in the catalytic activity of RNA. Since modern ribozymes closely associate with proteins, did ancient ribozymes associate with prebiotically plausible peptides?

RNA binding proteins, such as the ones discussed above, must necessarily contain structures capable of recognizing and binding to RNA. Perhaps the most common and best characterized domain is the RNA recognition motif (RRM), which recognizes RNA through interaction with the surface of a β-sheet (figure 1.3). The pre-mRNA splicing factor U2AF65 (figure 2) contains an RRM that recognizes the U-tract of the 3’ splice site.[72] This RRM displays three modes of RNA interaction: base stacking

(His 230), hydrogen bonding with a nucleobase (Arg 150), and electrostatic interaction with the phosphate backbone (Lys 195). A fourth mode not shown here is interaction of aliphatic side chains with hydrophobic nucleobases

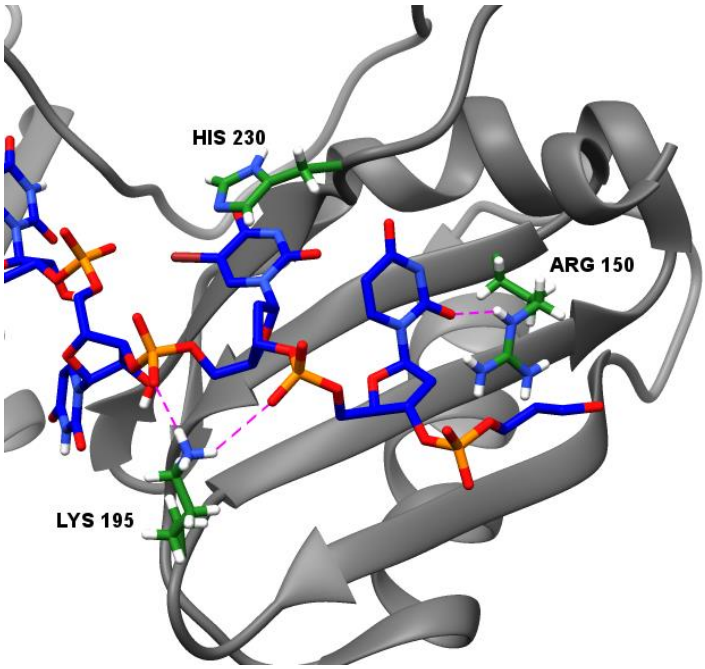


Figure 1.3 The RNA recognition motif of a pre-mRNA splicing factor. Shown here are three different modes by which this RRM interacts with RNA: base stacking (His 230), hydrogen bonding (Arg 150), and electrostatics (Lys 195). Modern proteins typically exploit the cationic amino acids, Arg and Lys, to interact with RNA; however, prebiotic peptides would need to use different amino acids to interact with RNA because the cationic amino acids, Lys and Arg, would not have been present on the early earth or would have been extremely rare.

The cationic amino acids Arg and Lys are known to interact with RNA the most in modern biology[73–75]; however, they have not been detected in the products of Miller-type discharge experiments or in meteorites. This represents a major hurdle in understanding how prebiotic peptides could interact with RNA. One possibility is that aliphatic amino acids such as alanine, α-aminobutyric acid, β-aminobutyric acid, α-aminoisobutyric acid or valine may interact with hydrophobic surfaces of the nucleobases.[76] For example, γ-aminobutyric acid, α-aminoisobutyric acid and β-alanine

are known to prevent the formation of β-sheets by interacting with the hydrophobic surface of β-sheets[77,78] and it may be possible that prebiotic peptides that contain these amino acids could interact with hydrophobic surfaces of the nucleobases. Moreover, α-aminoisobutyric acid is restricted in conformational space and may confer structural rigidity to RNA stabilizing favorable structures.[78]

In addition, the high proportion of glycine may aid RNA interactions because glycine is small enough to allow its peptide bond to more efficiently interact with RNA.[76] For example, glycine has been observed to aid in binding of peptides to the major groove of A-form RNA.[79] It is possible that amino acids with extended backbones, such as β-alanine or γ-aminobutyric acid, could interact electrostatically with the phosphodiester backbone of RNA if these amino acids were located at the N terminus of the peptide.

Additionally, serine can participate in hydrogen bonding with nucleobases, and both aspartate and the carboxy terminus of the peptide can chelate and coordinate Mg^{2+} for binding to the negatively charged phosphodiester backbone of RNA.[76,80–82] Prebiotically plausible amino acids may contribute to ribozyme function by destabilizing unfavorable structures and stabilizing favorable ones, or by coordinating metal cofactors. In addition, prebiotically plausible amino acids such as α, β , γ -aminobutyric acid, β-alanine, and α-aminoisobutyric acid may show novel modes of interaction with RNA that make the corresponding peptides beneficial to ribozyme activity.

The polymerization of amino acids into peptides appears to be prebiotically plausible. Recent work has investigated how peptides could have formed prebiotically: from a mixture of alpha-hydroxy acids and amino acids when subjected to wet dry cycles; from wet dry cycling in the presence of deliquescent mineral mixtures; and by

mechanical energy.[83–85] The largest body of work investigating prebiotic peptide bond formation involves the use of condensing agents. Most of these experiments showed synthesis of dipeptides, tripeptides and in few cases tetrapeptides.[86] Peptide synthesis using diamidophosphate as an activating agent showed the synthesis of homoglycine octamers.[87] Oligomers up to 11 residues in length were observed by Campbell, et al. (2019) when deliquescent minerals and wet-dry cycles were used to drive polymerization of glycine.[84] In addition, wet-dry cycles with alpha hydroxy acids (which are prebiotically plausible) generated depsipeptides up to 10 residues in length.[83] In the study by Forsythe, et al. (2015), lactic acid was incubated with amino acids. In the dry phase of the cycle, evaporation favored ester bond formation and ester-amide bond exchange. During the wet phase of the cycle, ester bonds are preferentially hydrolyzed; therefore, the pool of oligomers should enrich with amide bond containing polymers over time.

Several groups have attempted to investigate how peptides could interact with ribozymes in the context of the RNA world. Previous work has shown that polymers containing lysine can enhance ribozyme activity by concentrating RNA and Mg^{2+}.[68,88–90] In addition, groups have shown that the hammerhead ribozyme and ligase ribozyme activities can be promoted by a cationic amino acid rich peptide fragments from HIV1 Tat, HIV1 Rev, and HDV antigen proteins.[91–93] Much of the previous work has relied on cationic amino acids to effect ribozyme activity, even though these amino acids were not available or very rare on the early Earth. A recent study used *in vitro* evolution to identify a variant of the RNA binding domain of the ribosomal uL11 protein that was

composed of prebiotically plausible amino acids and could still bind RNA.[82] The protein variant could bind to RNA using salt bridges mediated by glutamate residues.

1.6 *In Vitro* Selection

In vitro selection is a method in which a large pool of randomized DNA or RNA is used to select for rare functional molecules. *In vitro* selections were used to great effect to identify RNA aptamers capable of binding a wide variety of small molecules.[94–96] The first artificial ribozyme was selected by evolving a group I intron to cleave single-stranded DNA.[97] But the first ribozyme selected from a random pool of sequences was the class I ligase ribozyme, which was further evolved to generate the first RNA polymerase ribozymes.[40,98–100] The method has been used to identify RNA capable of a wide variety of catalysis as discussed in section 1.3.

Previous work in our group showed that *in vitro* selection could be used to identify a ribozyme capable of self-triphosphorylation.[101] In this study, a large pool of randomized RNA sequences was challenged to use cyclic trimetaphosphate to triphosphorylate its own 5'-hydroxyl group. A hammerhead ribozyme upstream of the randomized sequences was used to ensure that the RNA transcripts had a 5'-OH group. The selection step was carried out by ligating sequences with a 5'-triphosphate modification to a biotinylated oligomer using a ligase ribozyme. Streptavidin bead pulldown was then used to isolate active sequences. From a library of 1.7×10^{14} randomized sequences, this selection identified more than 300 ribozymes via high-throughput sequencing.[102]

A variation of the selection scheme used in Moretti & Müller (2014) is used in this dissertation as a model system to test whether or not prebiotically plausible peptides could aid the activity of self-triphosphorylation ribozymes as detailed in Chapter 5. By using an established selection system that was already shown to be successful and by changing one variable – the presence of peptides, we can evaluate the effect of prebiotic peptides on the emergence of self triphosphorylating ribozymes. In contrast to previous studies, the selection scheme used here will use a set of peptides composed of prebiotically plausible amino acids in an *in vitro* selection starting from random sequence. Therefore, ribozymes will have the opportunity to evolve in the presence of prebiotically plausible peptides. Furthermore, the activity being selected for is chemical catalysis as opposed to simple binding of a peptide.

1.7 References

1. Bowring, S. Closing the gap. *Nat. Geosci.* **7**, 169–170 (2014).
2. Harrison, T. M., Schmitt, A. K., McCulloch, M. T. & Lovera, O. M. Early (≥4.5 Ga) formation of terrestrial crust: Lu–Hf, δ18O, and Ti thermometry results for Hadean zircons. *Earth Planet. Sci. Lett.* **268**, 476–486 (2008).
3. Valley, J. W., Cavosie, A. J., Ushikubo, T., Reinhard, D. A., Lawrence, D. F., Larson, D. J., Clifton, P. H., Kelly, T. F., Wilde, S. A., Moser, D. E. & Spicuzza, M. J. Hadean age for a post-magma-ocean zircon confirmed by atom-probe tomography. *Nat. Geosci.* **7**, 219–223 (2014).

4. Nakajima, M. & Stevenson, D. J. Melting and mixing states of the Earth's mantle after the Moon-forming impact. *Earth Planet. Sci. Lett.* **427**, 286–295 (2015).
5. Watson, E. B. & Harrison, T. M. Zircon Thermometer Reveals Minimum Melting Conditions on Earliest Earth. *Science* **308**, 841–844 (2005).
6. Wilde, S. A., Valley, J. W., Peck, W. H. & Graham, C. M. Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. *Nature* **409**, 175–178 (2001).
7. Bada, J. & Korenaga, J. Exposed Areas Above Sea Level on Earth >3.5 Gyr Ago: Implications for Prebiotic and Primitive Biotic Chemistry. *Life* **8**, 55 (2018).
8. Korenaga, J. Was There Land on the Early Earth? *Life* **11**, 1142 (2021).
9. Wacey, D., Kilburn, M. R., Saunders, M., Cliff, J. & Brasier, M. D. Microfossils of sulphur-metabolizing cells in 3.4-billion-year-old rocks of Western Australia. *Nat. Geosci.* **4**, 698–702 (2011).
10. Noffke, N., Christian, D., Wacey, D. & Hazen, R. M. Microbially Induced Sedimentary Structures Recording an Ancient Ecosystem in the *ca.* 3.48 Billion-Year-Old Dresser Formation, Pilbara, Western Australia. *Astrobiology* **13**, 1103–1124 (2013).
11. Oparin, A. *Proiskhozhedenie Zhizni*. (Reprinted and translated in Bernal JD (1967) The origin of life. Weidenfeld and Nicolson, London, 1924).
12. Haldane, J. The origin of life. *Ration. Annu.* **148**, 3–10 (1929).
13. Lazcano, A. Historical Development of Origins Research. *Cold Spring Harb. Perspect. Biol.* **2**, a002089–a002089 (2010).

14. Miller, S. L., Schopf, J. W. & Lazcano, A. Oparin's ``Origin of Life'': Sixty Years Later. *J. Mol. Evol.* **44**, 351–353 (1997).
15. Tirard, S. J. B. S. Haldane and the origin of life. *J. Genet.* **96**, 735–739 (2017).
16. Lupu, R. E., Zahnle, K., Marley, M. S., Schaefer, L., Fegley, B., Morley, C., Cahoy, K., Freedman, R. & Fortney, J. J. THE ATMOSPHERES OF EARTHLIKE PLANETS AFTER GIANT IMPACT EVENTS. *Astrophys. J.* **784**, 27 (2014).
17. Armstrong, K., Frost, D. J., McCammon, C. A., Rubie, D. C. & Boffa Ballaran, T. Deep magma ocean formation set the oxidation state of Earth's mantle. *Science* **365**, 903–906 (2019).
18. Sossi, P. A., Burnham, A. D., Badro, J., Lanzirotti, A., Newville, M. & O'Neill, H. St. C. Redox state of Earth's magma ocean and its Venus-like early atmosphere. *Sci. Adv.* **6**, eabd1387 (2020).
19. Solomatova, N. V. & Caracas, R. Genesis of a CO2-rich and H2O-depleted atmosphere from Earth's early global magma ocean. *Sci. Adv.* **7**, eabj0406 (2021).
20. Zahnle, K. J., Lupu, R., Catling, D. C. & Wogan, N. Creation and Evolution of Impact-generated Reduced Atmospheres of Early Earth. *Planet. Sci. J.* **1**, 11 (2020).
21. Miller, S. L. A Production of Amino Acids Under Possible Primitive Earth Conditions. *Science* **117**, 528–529 (1953).
22. Parker, E. T., Cleaves, H. J., Dworkin, J. P., Glavin, D. P., Callahan, M., Aubrey, A., Lazcano, A. & Bada, J. L. Primordial synthesis of amines and amino acids in a 1958 Miller H2S-rich spark discharge experiment. *Proc Natl Acad Sci U A* **108**, 5526–31 (2011).

23. Rich, A. On the Problems of Evolution and Biochemical Information Transfer. in *Horizons in Biochemistry* 103–126 (Academic Press, 1962).
24. Woese, C. R. *The genetic code: the molecular basis for genetic expression , Carl R. Woese.* (Harper & Row, 1967).
25. Crick, F. H. C. The origin of the genetic code. *J Mol Biol* **38**, 367–379 (1968).
26. Orgel, L. E. Evolution of the genetic apparatus. *J Mol Biol* **38**, 381–393 (1968).
27. Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E. & Cech, T. R. Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* **31**, 147–157 (1982).
28. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857 (1983).
29. Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–30 (2000).
30. White, H. B., 3rd. Coenzymes as fossils of an earlier metabolic state. *J Mol Evol* **7**, 101–4 (1976).
31. Oba, Y., Takano, Y., Furukawa, Y., Koga, T., Glavin, D. P., Dworkin, J. P. & Naraoka, H. Identifying the wide diversity of extraterrestrial purine and pyrimidine nucleobases in carbonaceous meteorites. *Nat. Commun.* **13**, 2008 (2022).
32. Oró, J. Synthesis of adenine from ammonium cyanide. *Biochem. Biophys. Res. Commun.* **2**, 407–412 (1960).
33. Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239–42 (2009).

34. Becker, S., Thoma, I., Deutsch, A., Gehrke, T., Mayer, P., Zipse, H. & Carell, T. A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. *Science* **352**, 833–836 (2016).
35. Becker, S., Feldmann, J., Wiedemann, S., Okamura, H., Schneider, C., Iwan, K., Crisp, A., Rossa, M., Amatov, T. & Carell, T. Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides. *Science* **366**, 76–82 (2019).
36. Martin, L. L., Unrau, P. J. & Müller, U. F. RNA Synthesis by in Vitro Selected Ribozymes for Recreating an RNA World. *Life* **5**, 247–268 (2015).
37. Robertson, M. P. & Joyce, G. F. The origins of the RNA world. *Cold Spring Harb. Perspect. Biol.* **4**, a003608 (2012).
38. Levy, M. & Miller, S. L. The Prebiotic Synthesis of Modified Purines and Their Potential Role in the RNA World. *J. Mol. Evol.* **48**, 631–637 (1999).
39. Huang, W. & Ferris, J. P. One-Step, Regioselective Synthesis of up to 50-mers of RNA Oligomers by Montmorillonite Catalysis. *J. Am. Chem. Soc.* **128**, 8914–8919 (2006).
40. 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* **292**, 1319–1325 (2001).
41. Wochner, A., Attwater, J., Coulson, A. & Holliger, P. Ribozyme-Catalyzed Transcription of an Active Ribozyme. *Science* **332**, 209–212 (2011).
42. 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.* **117**, 2906–2913 (2020).

43. Li, Y. & Breaker, R. R. Kinetics of RNA Degradation by Specific Base Catalysis of Transesterification Involving the 2‘-Hydroxyl Group. *J. Am. Chem. Soc.* **121**, 5364–5372 (1999).
44. Jimenez, R. M., Polanco, J. A. & Lupták, A. Chemistry and Biology of Self-Cleaving Ribozymes. *Trends Biochem. Sci.* **40**, 648–661 (2015).
45. Müller, S., Appel, B., Balke, D., Hieronymus, R. & Nübel, C. Thirty-five years of research into ribozymes and nucleic acid catalysis: where do we stand today? *F1000Research* **5**, F1000 Faculty Rev-1511 (2016).
46. Chen, X., Li, N. & Ellington, A. D. Ribozyme Catalysis of Metabolism in the RNA World. *Chem. Biodivers.* **4**, 633–655 (2007).
47. Breaker, R. R. & Joyce, G. F. The Expanding View of RNA and DNA Function. *Chem. Biol.* **21**, 1059–1065 (2014).
48. Silverman, S. K. Nucleic Acid Enzymes (Ribozymes and Deoxyribozymes): In Vitro Selection and Application. in *Wiley Encyclopedia of Chemical Biology* wecb406 (John Wiley & Sons, Inc., 2008). doi:10.1002/9780470048672.wecb406.
49. Brandsen, B. M., Hesser, A. R., Castner, M. A., Chandra, M. & Silverman, S. K. DNA-Catalyzed Hydrolysis of Esters and Aromatic Amides. *J. Am. Chem. Soc.* **135**, 16014–16017 (2013).
50. Breaker, R. R. Imaginary Ribozymes. *ACS Chem. Biol.* **15**, 2020–2030 (2020).
51. Flemmich, L., Heel, S., Moreno, S., Breuker, K. & Micura, R. A natural riboswitch scaffold with self-methylation activity. *Nat. Commun.* **12**, 3877 (2021).
52. McCown, P. J., Corbino, K. A., Stav, S., Sherlock, M. E. & Breaker, R. R. Riboswitch diversity and distribution. *RNA* **23**, 995–1011 (2017).

53. Walton, C. R., Shorttle, O., Jenner, F. E., Williams, H. M., Golden, J., Morrison, S. M., Downs, R. T., Zerkle, A., Hazen, R. M. & Pasek, M. Phosphorus mineral evolution and prebiotic chemistry: From minerals to microbes. *Earth-Sci. Rev.* **221**, 103806 (2021).
54. Pasek, M. A., Harnmeijer, J. P., Buick, R., Gull, M. & Atlas, Z. Evidence for reactive reduced phosphorus species in the early Archean ocean. *Proc Natl Acad Sci U A* **110**, 10089–94 (2013).
55. Todd, Z. R. Sources of Nitrogen-, Sulfur-, and Phosphorus-Containing Feedstocks for Prebiotic Chemistry in the Planetary Environment. *Life* **12**, 1268 (2022).
56. Pasek, M. & Lauretta, D. Extraterrestrial Flux of Potentially Prebiotic C, N, and P to the Early Earth. *Orig. Life Evol. Biospheres* **38**, 5–21 (2008).
57. Pasek, M. A., Kee, T. P., Bryant, D. E., Pavlov, A. A. & Lunine, J. I. Production of potentially prebiotic condensed phosphates by phosphorus redox chemistry. *Angew Chem Int Ed Engl* **47**, 7918–20 (2008).
58. Haqq-Misra, J., Kasting, J. F. & Lee, S. Availability of O2 and H2O2 on Pre-Photosynthetic Earth. *Astrobiology* **11**, 293–302 (2011).
59. He, H., Wu, X., Xian, H., Zhu, J., Yang, Y., Lv, Y., Li, Y. & Konhauser, K. O. An abiotic source of Archean hydrogen peroxide and oxygen that pre-dates oxygenic photosynthesis. *Nat. Commun.* **12**, 6611 (2021).
60. Etaix, E. & Orgel, L. E. Phosphorylation of Nucleosides in Aqueous Soultion Using Trimetaphosphate: Formation of Nucleoside Triphosphates. *J Carbohydr. Nucleosides Nucleotides* **5**, 91–110 (1978).

61. Chizzolini, F., Kent, A. D., Passalacqua, L. F. M. & Lupták, A. Enzymatic RNA Production from NTPs Synthesized from Nucleosides and Trimetaphosphate**. *ChemBioChem* **22**, 2098–2101 (2021).
62. Gibard, C., Gorrell, I. B., Jiménez, E. I., Kee, T. P., Pasek, M. A. & Krishnamurthy, R. Geochemical Sources and Availability of Amidophosphates on the Early Earth. *Angew. Chem. Int. Ed.* **58**, 8151–8155 (2019).
63. Brady, M. P., Tostevin, R. & Tosca, N. J. Marine phosphate availability and the chemical origins of life on Earth. *Nat. Commun.* **13**, 5162 (2022).
64. Rode, B. M. Peptides and the origin of life. *Peptides* **20**, 773–786 (1999).
65. Johnson, A. P., Cleaves, H. J., Dworkin, J. P., Glavin, D. P., Lazcano, A. & Bada, J. L. The Miller Volcanic Spark Discharge Experiment. *Science* **322**, 404–404 (2008).
66. Sephton, M. A. Organic compounds in carbonaceous meteorites. *Nat. Prod. Rep.* **19**, 292–311 (2002).
67. Gabbay, E. J., Adawadkar, P. D., Kapicak, L., Pearce, S. & Wilson, W. D. The Interaction Specificity of Peptides with DNA. Evidence for Peptide /3-Sheet-DNA Binding"^. 6.
68. Tagami, S., Attwater, J. & Holliger, P. Simple peptides derived from the ribosomal core potentiate RNA polymerase ribozyme function. *Nat Chem* **9**, 325–332 (2017).
69. Mondragón, A. Structural Studies of RNase P. *Annu. Rev. Biophys.* **42**, 537–557 (2013).
70. Voorhees, R. M. & Ramakrishnan, V. Structural Basis of the Translational Elongation Cycle. *Annu. Rev. Biochem.* **82**, 203–236 (2013).

71. Wahl, M. C., Will, C. L. & Lührmann, R. The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell* **136**, 701–718 (2009).
72. Cléry, A., Blatter, M. & Allain, F. H.-T. RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.* **18**, 290–298 (2008).
73. Hoffman, M. M. AANT: the Amino Acid-Nucleotide Interaction Database. *Nucleic Acids Res.* **32**, 174D – 181 (2004).
74. Krüger, D. M., Neubacher, S. & Grossmann, T. N. Protein–RNA interactions: structural characteristics and hotspot amino acids. *RNA* **24**, 1457–1465 (2018).
75. Liu, L., Xiong, Y., Gao, H., Wei, D.-Q., Mitchell, J. C. & Zhu, X. dbAMEPNI: a database of alanine mutagenic effects for protein–nucleic acid interactions. *Database* **2018**, (2018).
76. Helene, C., Maurizot, J.-C. & Wagner, K. G. Interactions of Oligopeptides With Nucleic Acid. *Crit. Rev. Biochem.* **10**, 213–258 (1981).
77. Madine, J., Wang, X., Brown, D. R. & Middleton, D. A. Evaluation of β-Alanine- and GABA-Substituted Peptides as Inhibitors of Disease-Linked Protein Aggregation. *ChemBioChem* **10**, 1982–1987 (2009).
78. Gilead, S. & Gazit, E. Inhibition of Amyloid Fibril Formation by Peptide Analogues Modified with α-Aminoisobutyric Acid. *Angew. Chem. Int. Ed.* **43**, 4041–4044 (2004).
79. Chen, L. & Frankel, A. D. A peptide interaction in the major groove of RNA resembles protein interactions in the minor groove of DNA. *Proc. Natl. Acad. Sci.* **92**, 5077–5081 (1995).

80. Weiner, L. M., Backer, J. M. & Rezvukhin, A. I. Participation of manganese ions complexed with tRNA in the interaction with amino acids and dipeptides. *Biochim. Biophys. Acta* **383**, 316–324 (1975).
81. Szostak, J. W. The eightfold path to non-enzymatic RNA replication. *J. Syst. Chem.* **3**, 2 (2012).
82. Giacobelli, V. G., Fujishima, K., Lepšík, M., Tretyachenko, V., Kadavá, T., Makarov, M., Bednárová, L., Novák, P. & Hlouchová, K. In Vitro Evolution Reveals Noncationic Protein–RNA Interaction Mediated by Metal Ions. *Mol. Biol. Evol.* **39**, msac032 (2022).
83. Forsythe, J. G., Yu, S.-S., Mamajanov, I., Grover, M. A., Krishnamurthy, M., Fernandez, F. M. & Hud, N. V. Ester-mediated amide bond formation driven by wet-dry cycles: A possible path to polypeptides on the prebiotic Earth. *Angew Chem Int Ed Engl* **54**, 9871–9875 (2015).
84. Campbell, T. D., Febrian, R., McCarthy, J. T., Kleinschmidt, H. E., Forsythe, J. G. & Bracher, P. J. Prebiotic condensation through wet–dry cycling regulated by deliquescence. *Nat. Commun.* **10**, 4508 (2019).
85. Stolar, T., Grubešić, S., Cindro, N., Meštrović, E., Užarević, K. & Hernández, J. G. Mechanochemical Prebiotic Peptide Bond Formation. *Angew. Chem. Int. Ed.* **60**, 12727–12731 (2021).
86. Frenkel-Pinter, M., Samanta, M., Ashkenasy, G. & Leman, L. J. Prebiotic Peptides: Molecular Hubs in the Origin of Life. *Chem. Rev.* **120**, 4707–4765 (2020).

87. Gibard, C., Bhowmik, S., Karki, M., Kim, E.-K. & Krishnamurthy, R. Phosphorylation, oligomerization and self-assembly in water under potential prebiotic conditions. *Nat. Chem.* **10**, 212–217 (2018).
88. Le Vay, K., Song, E. Y., Ghosh, B., Tang, T. -Y. D. & Mutschler, H. Enhanced Ribozyme-Catalyzed Recombination and Oligonucleotide Assembly in Peptide-RNA Condensates. *Angew. Chem. Int. Ed.* **60**, 26096–26104 (2021).
89. Li, P., Holliger, P. & Tagami, S. Hydrophobic-cationic peptides modulate RNA polymerase ribozyme activity by accretion. *Nat. Commun.* **13**, 3050 (2022).
90. Iglesias-Artola, J. M., Drobot, B., Kar, M., Fritsch, A. W., Mutschler, H., Dora Tang, T.-Y. & Kreysing, M. Charge-density reduction promotes ribozyme activity in RNA–peptide coacervates via RNA fluidization and magnesium partitioning. *Nat. Chem.* **14**, 407–416 (2022).
91. Herschlag, D., Khosla, M., Tsuchihashi, Z. & Karpel, R. L. An RNA chaperone activity of non-specific RNA binding proteins in hammerhead ribozyme catalysis. *EMBO J.* **13**, 2913–2924 (1994).
92. Huang, Z.-S. & Wu, H.-N. Identification and Characterization of the RNA Chaperone Activity of Hepatitis Delta Antigen Peptides. *J. Biol. Chem.* **273**, 26455–26461 (1998).
93. Robertson, M. P., Knudsen, S. M. & Ellington, A. D. In vitro selection of ribozymes dependent on peptides for activity. *RNA* **10**, 114–127 (2004).
94. Tuerk, C. & Gold, L. Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase. *Science* **249**, 505–510 (1990).

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

96. Wilson, D. S. & Szostak, J. W. In Vitro Selection of Functional Nucleic Acids. *Annu. Rev. Biochem.* **68**, 611–647 (1999).

97. Robertson, D. L. & Joyce, G. F. Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature* **344**, 467–468 (1990).

98. Bartel, D. & Szostak, J. Isolation of new ribozymes from a large pool of random sequences. *Science* **261**, 1411–1418 (1993).

99. Ekland, E. H., Szostak, J. W. & Bartel, D. P. Structurally Complex and Highly Active RNA Ligases Derived from Random RNA Sequences. *Science* **269**, 364–370 (1995).

100. Ekland, E. H. & Bartel, D. P. RNA catalysed RNA polymerization using nucleoside triphosphates. *Nature* **382**, 373–376 (1996).

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

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

Chapter 2

A Combinatorial Method to Isolate Short Ribozymes from Complex Ribozyme Libraries

A combinatorial method to isolate short ribozymes from complex ribozyme libraries

Joshua T. Arriola and Ulrich F. Müller*

Department of Chemistry & Biochemistry, University of California San Diego, La Jolla, California 92093, USA

Received June 06, 2020; Revised August 28, 2020; Editorial Decision September 16, 2020; Accepted October 01, 2020

ABSTRACT

In vitro selections are the only known methods to generate catalytic RNAs (ribozymes) that do not exist in nature. Such new ribozymes are used as biochemical tools, or to address questions on early stages of life. In both cases, it is helpful to identify the shortest possible ribozymes since they are easier to deploy as a tool, and because they are more likely to have emerged in a prebiotic environment. One of our previous selection experiments led to a library containing hundreds of different ribozyme clusters that catalyze the triphosphorylation of their 5′-terminus. This selection showed that RNA systems can use the prebiotically plausible molecule cyclic trimetaphosphate as an energy source. From this selected ribozyme library, the shortest ribozyme that was previously identified had a length of 67 nucleotides. Here we describe a combinatorial method to identify short ribozymes from libraries containing many ribozymes. Using this protocol on the library of triphosphorylation ribozymes, we identified a 17-nucleotide sequence motif embedded in a 44-nucleotide pseudoknot structure. The described combinatorial approach can be used to analyze libraries obtained by different in vitro selection experiments.

INTRODUCTION

In vitro selections were originally established to identify RNAs that bind to specific target molecules (1,2), and developed further to identify catalytic RNAs (ribozymes) from large, combinatorial libraries (3). *In vitro* selections for catalytic RNAs have usually employed libraries with 10^{12} to 10^{15} different sequences, in which a portion of the library between 40 and 228 nucleotides in length is randomized (3–30). To select for active sequences, the libraries of randomized RNA sequences are usually covalently coupled to one of the reaction partners of the desired reaction, while the second reaction partner is coupled to a functional group that serves as a handle. When incubating the modified library with the functionalized substrate, catalytically active sequences can covalently link themselves to the handle. The handle is then used to isolate the active sequences. Because the enrichment for active sequences in a single experiment is limited, multiple cycles of incubation with substrate, isolation of active molecules, reverse transcription, PCR amplification, and transcription are necessary to obtain libraries that are dominated by active sequences. The resulting, active library is then analyzed for individual, active sequences.

After active sequences have been enriched by selection steps, it is often desirable to identify ribozymes with specific properties. To enrich for the most active sequences, usually selection cycles with higher selection stringency are added (e.g. (3,12,13)). To identify ribozyme variants that use different metal ion cofactors, further rounds of in vitro selection or in vitro evolution can be employed in the presence of different metal cations, or metal cation concentrations (31). To generate, and identify smaller variants of *individual* ribozymes a combinatorial method is available that can remove unnecessary fragments from the ribozyme (32). However, we are not aware of a method to identify the smallest ribozymes from complex libraries with *many* ribozymes.

The desire to identify short ribozymes is partially motivated by attempts to explore how an RNA world could have emerged as an early stage in the origin of life. Short ribozymes are especially helpful because they are more likely to emerge from a prebiotic setting for two reasons: First, short RNA polymers are more abundant than long RNA polymers in model prebiotic reactions (33). Second, short ribozymes require less information content, and are therefore found more frequently in a given set of randomized sequences. Previously selected ribozymes for prebiotically relevant reactions typically have sizes around 70 nucleotides or larger (e.g. (11,13,20,30,34,35), but see (36)). In contrast, model reactions for prebiotic RNA polymerization in the absence of enzymes or sophisticated ribozymes generates little material above 40 nucleotides, even under optimal conditions (33). Because RNAs with lengths of 30–40 nucleotides are more prebiotically plausible, the involvement of catalytic RNAs in early stages of life would be more convincing if short ribozymes could be found to emerge from random sequences.

*To whom correspondence should be addressed. Tel: +1 858 534 6823; Email: ufmuller@ucsd.edu
Present address: Ulrich Müller, Department of Chemistry & Biochemistry, University of California San Diego, La Jolla, CA 92093, USA.

© The Author(s) 2020. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

We previously performed an in vitro selection for ribozymes that triphosphorylate their own 5′-hydroxyl group with cyclic trimetaphosphate (Tmp) (30), a prebiotically plausible energy source (37,38). This selection identified >300 different ribozymes (39), each with a total length of 182 nucleotides. These results showed that catalytic RNA systems are able to use Tmp as energy source. Here we developed a combinatorial method to identify the shortest ribozyme sequences in complex libraries. We show how combinatorial truncation at the pool's 3′-terminus, size fractionation, and re-selection of each size fraction can identify the shortest ribozymes in this complex library. The shortest ribozyme had a truncated length of 44 nucleotides, and its 17-nucleotide conserved core was also identified in two additional sequences of the same selected ribozyme library by the re-analysis of high throughput sequencing data. The method may be useful for the isolation of short ribozyme motifs from other libraries containing hundreds of different ribozymes.

MATERIALS AND METHODS

Generation of truncated pools with a randomized primer

The truncated sub-pools were generated from template DNA using a randomized primer consisting of an N11 randomized region and a new 3′ constant region (5′-TAAG TCGTAGTTACATCANNNNNNNNNNN-3′). The template DNA was a pool consisting of hundreds of active triphosphorylating ribozymes flanked at the 5′ end by the T7 promoter sequence, a hammerhead ribozyme sequence, and a 5′ constant region and at the 3′ end by a 3′ constant region. The annealing and extension conditions were chosen to obtain a broad size distribution of the desired extension products. A mixture containing 30 pmol of template DNA, 5 pmol randomized primer in a volume of 0.2 ml, and a trace amount of radiolabeled randomized primer were denatured at 94°C for 2 min then immediately placed on ice. The Klenow fragment (New England Bioscience) was used to extend this primer according to manufacturer's instructions. The mixture was allowed to incubate overnight at room temperature. Formamide loading buffer containing 30 mM sodium/EDTA was added to the reaction mixture to quench the reaction. The mixture was heat denatured at 80°C for 2 min then separated on a denaturing 15% polyacrylamide gel. A 246 nt and 96 nt marker were used to define the region of interest. This region of interest was divided into 10 fragments and each fragment was excised. Each gel slice was eluted and ethanol precipitated. The resulting DNA sub-pools were PCR amplified using Taq DNA polymerase and transcribed using T7 RNA polymerase. The RNA sub-pools were gel purified, ethanol precipitated, and resuspended in water.

In vitro selection

One round of selection was performed on each purified RNA sub-pool according to previously published methods (30). 100 nM of purified RNA sub-pool was incubated with 50 mM Tris/HCl, 3.3 mM NaOH, 100 mM $MgCl_2$ and 50 mM of freshly dissolved trimetaphosphate. This mixture was allowed to incubate for 3 h at room temperature. The reaction was quenched by ethanol precipitation. The RNA pellet was washed with chilled water and desalted using P30 Tris/HCl spin columns (Bio-Rad). Active ribozymes were ligated to a biotinylated oligomer (biotin-*d*(GAACTGAAGTGTATG)rU) using the R3C ligase ribozyme whose arms were designed to anneal to the 5′ constant region of the pool RNA and the biotinylated oligomer. A mixture containing 1.3 uM of desalted pool RNA, 1 uM ligase ribozyme, 1.2 uM biotinylated oligomer, 100 mM KCl, 100 mM Tris/HCl, and 60 pM triphosphorylated RNA was heated to 65°C for 2 min then cooled to 30°C at a rate of 0.1°C per second. The triphosphorylated RNA was used to generate a small amount of ligation product and thereby reduce the number of PCR cycles, and prevent artefacts during PCR amplification. After heat renaturation, an equal volume of a mix consisting of 40% (w/v) PEG8000, 4 mM Spermidine, and 50 mM $MgCl_2$ was added to the ligase reaction mixture. This mixture was incubated at 30°C for 3 h. After the incubation step, the reaction was quenched by adding final concentrations of 13.9 mM sodium/EDTA, 50 mM Tris/HCl, 50 mM KCl, 0.012% (w/v) Triton X-100, and 1.19 uM of a 61 nt oligomer that was complementary to the ligase ribozyme (5′GAACTGAAGTGTATGCTTCAACCCATTCAA ACTGTTCTTACGAACAATCGAGCAAGATGTT-3′). This mixture was heated at 50°C for 10 min. Streptavidin magnetic beads (Promega) were washed thrice with 20 mM HEPES/KOH pH7.2, 0.01% (w/v) Triton X-100 and 50 mM KCl. Biotinylated RNA was then captured by mixing the RNA with the magnetic beads and rotating end-over-end at room temperature for at least 30 min. A magnetic rack was used to focus the beads and the beads were washed twice with a solution containing 0.01% (w/v) Triton X-100 and 20 mM NaOH. Captured RNA was eluted from the beads by incubating the beads in a solution of 25 mM Tris/HCl pH 8.5, 1.56 mM EDTA and 96% formamide at 65°C for 3 min. The beads were removed by immediate centrifugation, and the supernatant was concentrated by ethanol precipitation. 1 ug of tRNA was used as precipitation carrier. Captured RNA was resuspended in 10 mM Tris/HCl, pH 8.3. The RNA was reverse transcribed using Superscript III (Invitrogen) and a reverse transcription primer (5′-TAAGTCGTAGTTACATCA-3′) complementary to the new 3′ constant region, which was implemented after the Klenow extension using the randomized primer. The products were then PCR amplified with 5′ and 3′ primers that could bind to the sequence of the biotinylated oligomer and the 3′ constant region of the pool respectively. A second PCR amplification was performed to add the T7 promoter sequence and the hammerhead ribozyme sequence to the DNA.

Triphosphorylation assays

Purified ribozyme (5 uM) was incubated with 50 mM Tris/HCl, 100 mM $MgCl_2$, and 50 mM of freshly dissolved trimetaphosphate. This mixture was incubated for 3 h at room temperature. The products were ligated to a radiolabeled oligomer (γ^{32}P-d(GAACTGAAGTGTATG)rU) using an R3C ligase ribozyme whose arms were designed to anneal to the 5′ constant region of the pool RNA and the

oligomer. To do this, the above reaction mixture was diluted 1:10 in a solution that contained a final concentration of 0.5 uM ligase ribozyme, 0.5 uM biotinylated oligomer, 100 mM KCl, 100 mM Tris/HCl, and 15 mM sodium EDTA. This mixture was heat renatured at 65°C for 2 min then cooled to 30°C at a rate of 0.1°C per second. After heat renaturation, an equal volume of a mix consisting of 40% PEG8000 (w/v), 4 mM Spermidine, and 50 mM $MgCl_2$ was added to the ligase reaction mixture. This mixture was incubated at 30°C for 3 h. After the incubation step, the reaction was quenched by ethanol precipitation. The reaction products were resolved on a denaturing 10% polyacrylamide denaturing gel.

Generation of 3′ truncations

3′ truncations were generated by PCR using Taq DNA polymerase and a series of 3′ primers. The 3′ primers annealed to different portions of the ribozyme sequence of interest to generate a series of truncations with desired deletions from the 3′ end. DNA templates of each truncation were transcribed using T7 RNA polymerase, and purified by denaturing PAGE.

Secondary structure analysis using SHAPE

Selective 2′Hydroxyl Acylation Analyzed by Primer Extension (SHAPE) was performed on the 44-nucleotide long ribozyme isolated in this study, with 1-Methyl-7-nitro-2H-3,1-benzoxazine-2,4(1H)-dione (1M7) as the chemical probe (40). An adaptor region was added to the 3′ end of the ribozyme by PCR amplification. The sequence of this primer is (5′-GTGTGCTAGGATCACAATGATGTCTC TTTAATAAGA-3′) where the underlined portion is the adaptor region. 20 pmol of ribozyme were heat renatured in 10 ul at 80°C for 2 min, cooled to 50°C for 5 min then left at room temperature for 5 min. To this solution, HEPES/NaOH pH 8.0 and $MgCl_2$ were added to a final concentration of 50 mM each. This mixture was incubated at room temperature for 2 min. A 20 mM stock solution of 1M7 was prepared in DMSO. 1M7 was added to this solution such that its final concentration was 2 mM and the concentration of DMSO was 10% of the solution. A negative control was prepared with ribozyme, HEPES, $MgCl_2$ and 10% final DMSO. Both samples were incubated at room temperature for 3 min. The reaction was quenched by ethanol precipitation and resuspended in 10 ul of 5 mM Tris/HCl pH 8.0. The products were reverse transcribed using Superscript III reverse transcriptase (Invitrogen) according to manufacturer's instructions and trace amounts of a radiolabeled primer that annealed to the aforementioned adaptor region. The sequence of this primer is (5′- GTGTGCTAGGATCACAAT-3′). The RNA template in each sample was degraded by alkaline hydrolysis, by incubating for 5 min at 80°C in a solution containing 750 mM NaOH. The reaction was quenched by adding a two-fold stoichiometric excess of acetic acid to generate a NaOAc/HOAc buffer with a final concentration of 300 mM. The products were ethanol precipitated, resuspended in formamide loading buffer and resolved on a denaturing 20% polyacrylamide gel. Signals were quantified using the 'rectangles' function in the software Quantity One (Bio-Rad) and background rectangles were subtracted.

To predict the secondary structure based on these SHAPE data the software vsFold5 was chosen because it is able to predict pseudoknots, and because it outperformed four other algorithms in predicting a secondary structure consistent with the SHAPE probing data. Specifically, iterative HFold (41) was unable to predict the pseudoknot, IPKnot (42) predicted a pseudoknotted structure lacking a part of the P2 helix and had an additional 2-base pair helix that were inconsistent with the SHAPE data, HotKnots (43) and CCJ (44) introduced two different 2-base pair helices that were not supported by the SHAPE data. None of the five algorithms predicted a structure that showed an extension of the P2 duplex between bases C16/C17 and G37/A38, which would have been consistent with the SHAPE data (Figure 5).

RESULTS

To identify the shortest ribozymes from our previously selected library of self-triphosphorylating ribozymes (30) with > 300 different ribozyme clusters (39), we followed a five-step procedure (figures 1 - 3). In the first step, 3′-truncated variants of the DNA pool were generated with the help of partially randomized primers and Klenow enzyme (Figure 1A). To do this, the DNA pool was annealed to a DNA primer that contained 11 nt of a randomized sequence at its 3′-terminus (red in Figure 1A). This randomized sequence allowed the primer's 3′-terminus to anneal at any position of the templating DNA library molecules. The primer was then extended by a Klenow Fragment enzyme that lacked exonuclease activity. A portion of the DNA primer was radiolabeled at its 5′-terminus to allow the detection of extension products of this primer.

In the second step, the primer extension products were separated into size fractions, using denaturing polyacrylamide gel electrophoresis (PAGE) (Figure 1B). Two radiolabeled markers indicated the sizes of long Klenow extension products that included the full 150-nucleotide long random region (total length 246 nt), and short Klenow extension products that omitted any portion of the randomized region (total length 96 nt). The extension products with sizes between these markers were separated into ten evenly sized segments, and the DNA was eluted from each segment (red brackets in Figure 1B). As a negative control, a reaction that omitted any DNA pool was performed (Figure 1B, lane labeled 'Rxn-T'). Extension products <96 nt in length were detected but little to no products with >96 nt in length. The short products can be explained by a scenario where the randomized portions of two primers were complementary to each other or the randomized portion of one primer was complementary to the constant region of another. This did not generate problems because the length of the desired products was between 96 and 246 nt in length, and the segments excised from the gel contained little or no contamination from short, non-templated extension products.

In the third step, the DNA molecules eluted from each of the ten size fractions were amplified by PCR using primers complementary to the T7 RNA polymerase promoter at the

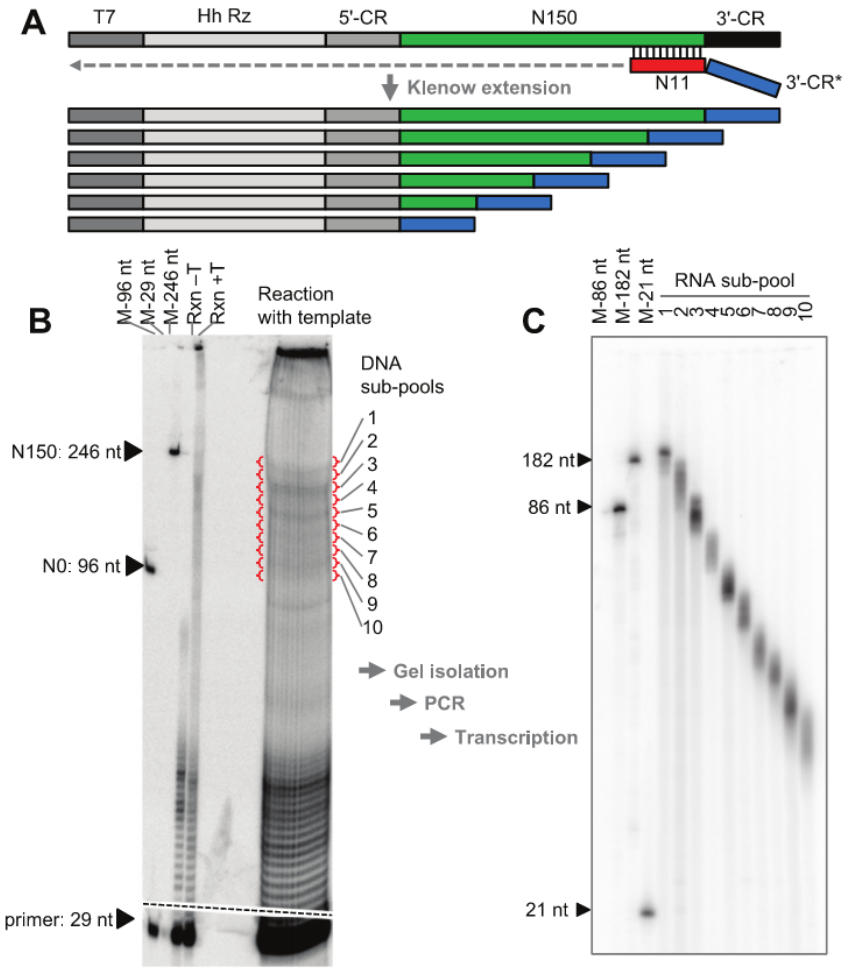


Figure 1. Workflow of generating ribozyme sub-pools with truncations at their 3′-end. (**A**) Schematic describing the extension of a primer with a randomized region (red) that can anneal anywhere on the template pool, and that introduces a new 3′-constant region (3′-CR*, blue). The annealed primers were then extended by the Klenow fragment. Only extension products with a random pool region (green) of 0–150 nucleotides are shown. (**B**) Autoradiogram of a denaturing 15% PAGE of Klenow extension products using 5′-radiolabeled primers. The first three lanes show markers that indicate the position of the unextended primer (29 nt), an extension product corresponding to the full-length pool (246 nt), and an extension product corresponding to the pool without randomized region (96 nt). The next two lanes show the reaction products of a small-scale reaction without template (Rxn-T) and with template (Rxn+T). The broad lane on the right shows the separation of a large-scale extension reaction with template (200 ul). Note that many products are shorter that the length of the full-length random region because the DNA library also contains a promoter for T7 RNA polymerase, a sequence encoding a hammerhead ribozyme, and a 5′-constant region. The dashed line indicates the position where two parts of the same exposure were assembled. We are unclear about the origin of the bands in the gel pockets; they may be an aggregation of template, hot primer, and Klenow fragment. The positions where ten sub-pools were excised from the gel are indicated with red brackets. After elution, each sub-pool was processed separately using PCR with a 5′-primer complementary to the T7 promoter, and a 3′-primer complementary to the new 3′-CR (blue in (A)). (**C**) Autoradiogram of 5′ end radiolabeled RNA sub-pools after transcription of the ten sup-pools, and separation by 8% denaturing PAGE. Three size markers indicate the position of full-length pool RNA (182 nt) and shorter fragments (86 nt and 21 nt).

5′-terminus (dark grey in Figure 1A), and the newly introduced 3′-constant region at the 3′-terminus (blue in Figure 1A). This PCR generated ten sub-pools with distinct size distributions. To test whether this procedure would give rise to RNA sub-pools with distinct sizes, each of the ten DNA sub-pools was transcribed, and a sample of the resulting RNA sub-pool was 5′ end radiolabeled with $\gamma[^{32}P]$ATP. The products were separated by PAGE (Figure 1C) and showed the desired size distributions. All ten sub-pools showed an expected overlap with the neighboring size fraction but no significant overlap with other size fractions. Note that the RNA sub-pools are significantly shorter than the DNA sub-pools because the hammerhead ribozyme at the 5′-terminus of all transcripts cleaves co-transcriptionally, generating a 5′-hydroxyl group (30).

In the fourth step, the shortest sub-pool with active ribozymes was identified. To do this, all ten RNA sub-pools were subjected to one round of in vitro selection, using the same in vitro selection procedure that was used to select the ribozyme library from random sequence (30)

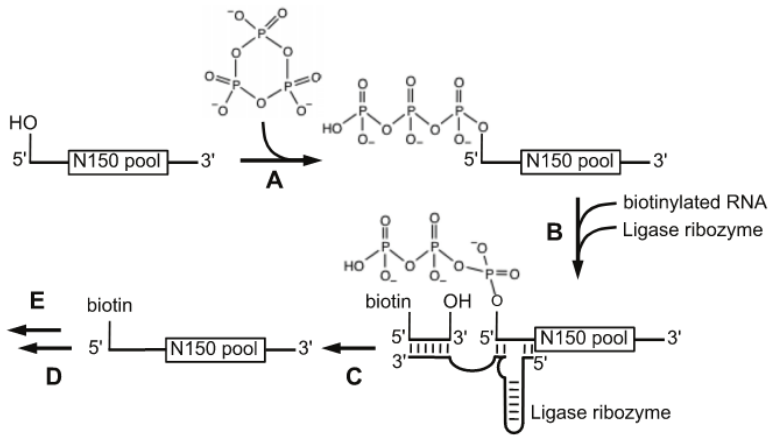


Figure 2. Schematic for selecting active ribozymes from each of the ten RNA sub-pools. (**A**) The RNA sub-pool is incubated with Tmp to allow active ribozymes to triphosphorylate their 5′-hydroxyl group, generating a 5′-triphosphate. (**B**) Pool molecules are annealed with a biotinylated RNA and a ligase ribozyme. The ligase ribozyme covalently links 5′-triphosphorylated pool molecules to biotin. (**C**) After capture of biotinylated RNAs on streptavidin beads, RNAs not linked covalently are washed away. Isolated Sequences are then (**D**) reverse transcribed and (**E**) PCR amplified.

(Figure 2). Specifically, the RNA sub-pools were incubated in the presence of Tmp such that catalytically active pool molecules could convert their 5′-hydroxyl to a 5′-triphosphate group. Pool molecules with a 5′-triphosphate were covalently linked to a biotinylated primer, using the R3C ligase ribozyme (19). Covalently linked pool molecules were isolated via their new biotin 5′-terminus using streptavidin beads, reverse transcribed, and PCR amplified. Importantly, only sub-pools 1–7 gave rise to a clean PCR product of the expected size (Supplementary Figure S1). Sub-pools 8 and 9 gave rise to a diffuse product distribution without a band at the expected size. This suggested that after one round of selection, sub-pool 7 contained the shortest pool sequences with self-triphosphorylation activity.

In the fifth step, catalytically active RNAs were identified from sub-pool 7. To do this, sub-pool 7 was cloned, and 21 clones were arbitrarily chosen for biochemical analysis and sequencing (Supplementary Figures S2 and S3). To measure the biochemical activity of each clone, an assay was used in which a radiolabeled primer was ligated to an equimolar concentration of individual RNAs after their incubation with Tmp, thereby quantifying the fraction of ribozymes that triphosphorylated their 5′-terminus (30) (Figure 3 and Supplementary Figure S3). After denaturing PAGE separation of the products, the quantified bands showed that ten of the 21 clones had biochemical activity (Figure 3A, B). When the sequences of these ribozymes were aligned, all ten active ribozymes fell within one cluster, and all 11 inactive clones belonged to other clusters (Figure 3B, C and Supplementary Figure S2). These results suggest that the cluster with these ten ribozymes represents the shortest ribozymes in the selected library. The individual sequence with highest activity, clone 78, was chosen for further biochemical analysis.

To identify the catalytic core of clone 78, additional truncations were made at its 3′-terminus, and the biochemical activity was tested for each truncation variant (Figure 4). Truncations at the 5′ end were not made since the 5′ terminus is the catalytic site for the selected activity. The average activity was highest when the ribozyme was truncated to 54 nucleotides, and the average activity of clone 78 (with a length of 72 nucleotides) was maintained when the ribozyme was truncated down to 44 nucleotides. Analysis of secondary structures predicted by vsFold5 (45) found that the ribozymes between 57 and 44 nucleotides in length could fold into a conserved pseudoknot structure, whereas longer and shorter variants of the ribozyme did not (Supplementary Figure S4).

To probe the secondary structure of the 44-nucleotide ribozyme we used selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) (40) (Figure 5). The SHAPE probing data were mostly consistent with the formation of the same pseudoknot as predicted by vsFold5. While the existence of the P2 stem is confirmed, the P1 stem may be shorter than the expected six base pairs, or its ends may be dynamic under the conditions of the SHAPE assay. The P2 stem may be slightly longer than the expected 6 bp through interactions of the bases C16/C17 with A38/G39. However, the SHAPE data of these bases may also reflect a more complex interaction, for example involving the protected bases C23/G24. A tertiary structure determination may help reveal, for example, interesting noncanonical base pair interactions or the three-dimensional shape of the Tmp binding pocket. Previous efforts in RNA 3D structure prediction have not yet been successful in predicting the structure of large ribozymes and most successful predictions have been made on short aptamers, especially on those where the structure of a known homolog has been determined (46–48). The ribozyme identified in this study may be a good candidate for 3D structure prediction due to its small size since the quality of a model correlates with size. Nevertheless, the pseudoknot shown in Figure 5B represents the best working model for the ribozyme's secondary structure.

When the sequence of the 44-nucleotide ribozyme was tested for similarities in high throughput sequencing data of the original selection (39), two additional clusters were identified that contained the same 17-nucleotide long motif

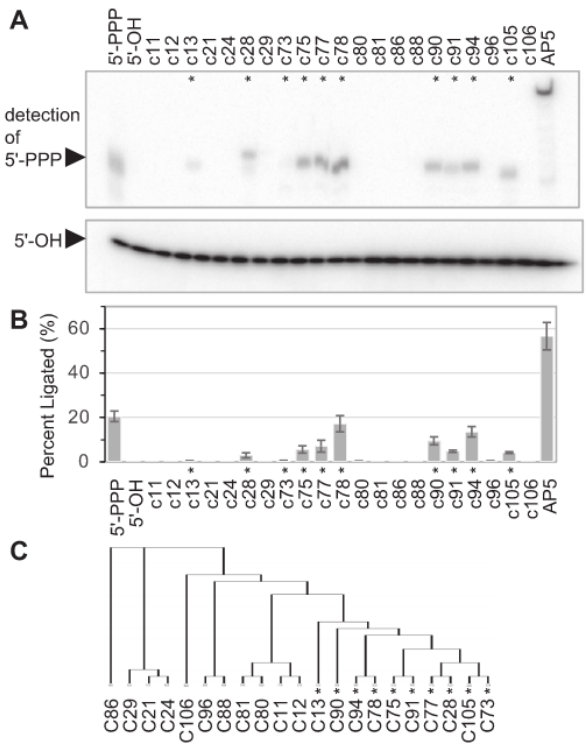


Figure 3. Biochemical and phylogenetic analysis of 21 cloned sequences from sub-pool 7. (**A**) Autoradiograph of a gel-shift assay to detect self-triphosphorylation activity. The lower panel shows unreacted RNAs, whereas bands in the upper panel show triphosphorylation activity. Triphosphorylated RNA (5′-PPP) and RNA with a 5′-hydroxyl group (5′-OH) served as positive and negative control, respectively. The highly active ribozyme AP5 (39) was used as additional positive control. AP5 is 182 nucleotides in length. Clones with an average activity above the detection limit of 0.1% are marked with an asterisk. (**B**) Quantitation of triphosphorylation assays as shown in (A). Error bars are standard deviations from triplicate experiments. (**C**) Phylogenetic comparison of the 21 sequences from sub-pool 7. The phylogenetic tree was generated with the software Geneious, using the neighbor-joining treebuild method. All clones with detectable biochemical activity (higher than 0.1%) are marked with an asterisk.

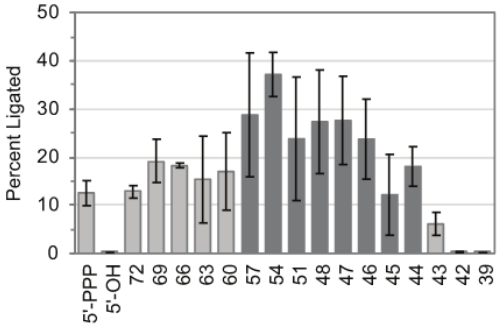


Figure 4. Biochemical activity of variants of ribozyme C78 that are successively truncated at their 3′-terminus. In addition to positive control (5′-PPP) and negative control (5′-OH), the length of the 3′-truncation variants is used as label. Clone C78 has a length of 72 nucleotides. Error bars are standard deviations from triplicate experiments. Columns highlighted in dark grey are predicted to fold into a pseudoknot structure (see Supplementary Figure S4).

(Figure 6). All three sequences are consistent with a model in which the RNA can fold back on itself to generate helix P1, immediately followed by an eight-nucleotide sequence that was identical between the three clusters (red in Figure 6). The six nucleotides preceding the P1 helix contained an additional three conserved nucleotides (green in Figure 6), forming an element of 21 nucleotides with 17 conserved positions. The three clusters containing this motif differed in the sequence and length of insertions before and after this 21-nucleotide long element. Additionally, the 3′-terminus of the pseudoknot forming the P2 helix was heterogeneous in sequence and appeared to be sufficient with 5 base pairs. At the 5′-terminus, the first 14 nucleotides of the ribozyme are defined by the constant region, which includes a two-nucleotide linker between helices P1 and P2 (purple). Because the P2 helix could also form with other sequences downstream, the length of this linker—two nucleotides—is likely required for activity.

To identify the optimal reaction conditions of the 44-nucleotide ribozyme we measured the biochemical activity at different concentrations of Tmp and Mg^{2+}, at different pH values and at different temperatures (Figure 7). At a Mg^{2+} concentration of 100 mM, the optimum of the Tmp concentration saturates at or above 200 mM (Figure 7A). In contrast, at a Tmp concentration of 200 mM, the optimal Mg^{2+} concentration does not exceed 100 mM (Figure 7B). This is different from the other well-characterized triphosphorylation ribozyme TPR1, which shows optimal activity only when the Mg^{2+} concentration exceeds the Tmp concentration (30), and suggests that these two ribozymes use different catalytic mechanisms. The pH dependence of the 44-nucleotide ribozyme shows increased activity at higher pH values (Figure 7C) as expected when the deprotonation of the 5′-hydroxyl group is the likely rate-limiting step of the reaction. The temperature optimum of the 44-nucleotide ribozyme is around 10°C (Figure 7D), different from the optimum of a TPR1 variant at 40°C (49), underlining the different characteristics of these two ribozymes. Together, the optimum reaction conditions for the 44-nucleotide long ribozyme were 200 mM Tmp, 100 mM Mg^{2+}, a pH of 8.5, and a temperature of 10°C.

Together, the results establish a new technique to isolate short ribozymes from combinatorial libraries containing many ribozymes, and provide an initial characterization of the smallest identified self-triphosphorylation ribozyme. The technique to identify small ribozymes from large ribozyme libraries may be employed for different ribozyme libraries and thereby help to identify short ribozymes as tools, as well as explore the feasibility of RNA systems that are small enough to have emerged from a prebiotic environment. For instance, the 44-nucleotide ribozyme identified in this study is close in length to products generated by model prebiotic reactions of RNA polymerization (33) and the 17-nucleotide conserved motif is well within this range.

DISCUSSION

The method described in this study identified a self-triphosphorylation ribozyme with a conserved motif of 17 nucleotides, and likely additional constraints at positions 21, 23, 24 and up to 5 positions that form the P2 duplex

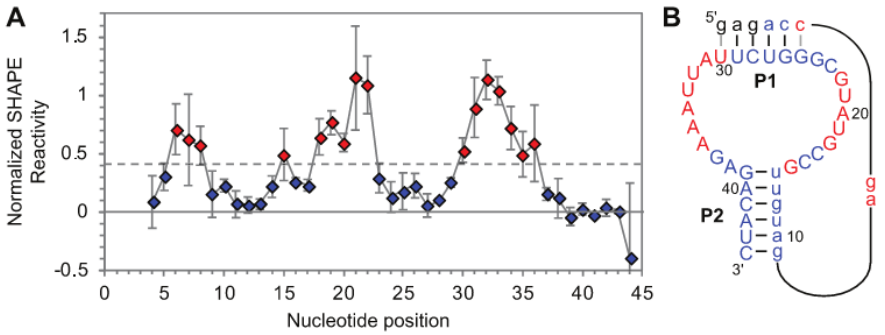


Figure 5. SHAPE probing of the secondary structure of the 44-nucleotide long ribozyme. (**A**) Normalized SHAPE reactivity as a function of the nucleotide position in the ribozyme. The reactivities of the average of the five most reactive positions were set to 1.0. The reactivity of 0.4 was set as the cutoff between 'low reactivity'(blue) and 'high reactivity' (red). The first three nucleotides (black) did not yield SHAPE data because they were too close to the radiolabeled 5′-terminus to be separated on the used denaturing polyacrylamide gels. Error bars are standard deviations from triplicate experiments. (**B**) Pseudoknot structure that is mostly consistent with the SHAPE probing data. Nucleotides with low SHAPE reactivity are shown in blue, positions with high reactivity in red.

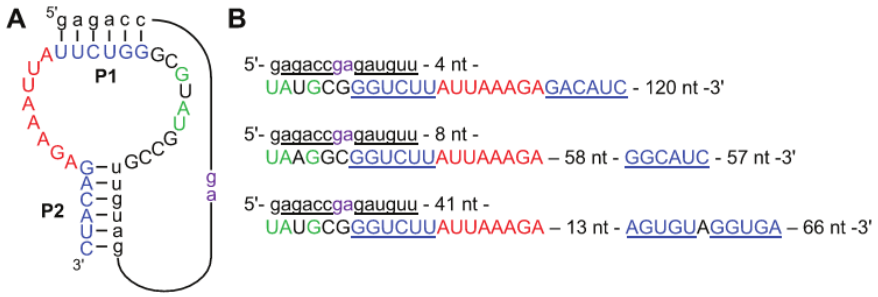


Figure 6. Sequence conservation among three different clusters that appear to form the same catalytic core. (**A**) Secondary structure with annotation of nucleotides that are within the primer binding site of the library (lower case), a two-nucleotide linker (purple), three conserved nucleotides in the L2/1 loop (green), sequences that form P1 and P2 (blue), and an eight-nucleotide sequence that is completely conserved among the three clusters (red). (**B**) Three peak sequences from clusters of the original selection. They could form the same secondary structure but with different sizes of loops between the P1 and P2 duplexes.

(see figures 5 and 6). For a given RNA sequence of length, n, the total sequence space is 4^n. Thus, the conserved motif identified in this study is likely to be found about once in 4^{20} ($\sim 1.1 \times 10^{12}$) sequences. The original library contained an effective complexity of 1.7×10^{14} sequences (30), therefore we would expect to find about 150 such ribozymes. However, we identified only three such ribozymes in the original library. There are several possibilities to explain this discrepancy. First, there may be additional sequence determinants in the ribozyme. For example, the nucleotides at position 15–18 may be constrained, and could contribute up to a 256-fold reduction in complexity. Second, the nucleotides forming the P2 stem (positions 39–44) may not be as flexible as the sequences of the second and third clones in the original library would suggest (Figure 6B). Third, a larger size of the loops between conserved elements may reduce the fraction of ribozyme folding into the active conformation. Therefore, most sequences with the 17-nucleotide motif near the 3′-terminus of the 150 randomized nucleotides may be inactive. This idea is supported by the finding that the conserved regions of all three putative ribozymes are within the first 100 nucleotides of the pool. Awaiting a detailed characterization of the sequence requirement of this small ribozyme, an incidence number of three in the original library is therefore within the expected range.

It is possible that there exist other, shorter triphosphorylation ribozymes that were not discovered. However, the 17-nucleotide motif of the 44-nucleotide ribozyme was identified three times in the library, suggesting that any smaller motif would have been contained even more frequently, and identified in the used procedure. Two types of self-triphosphorylation ribozymes would not have been discovered by the original selection: First, the random sequence library contained 14 nucleotides of fixed sequence at its 5′-terminus. Any ribozyme with different sequence requirements within the first 14 nucleotides would not have been contained in the original library. Second, the selection procedure requires that after self-triphosphorylation, the ribozyme 5′-terminus is unfolded and based paired to the 5′-terminus of a ligase ribozyme. Any ribozyme with a very stable structure at their 5′-terminus that does not allow annealing with the ligase ribozyme would not be selected. The procedure of this study relies on the same selection principle as the original selection, therefore the procedure described in this study would likely have found any ribozymes in the original library with a smaller sequence motif than the iden-

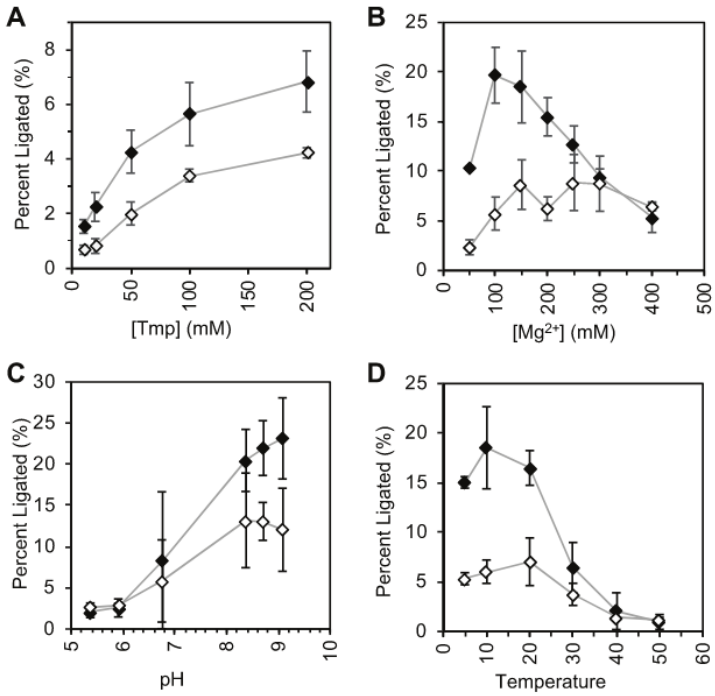


Figure 7. Dependence of ribozyme activity on reaction conditions. (**A**) Dependence of ribozyme activity on the concentration of trimetaphosphate (Tmp) at 100 mM $MgCl_2$ and 50 mM Tris/HCl pH 8.5. Empty diamonds correspond to reaction times of 2 min, filled diamonds to 20 min. (**B**) Dependence of ribozyme activity on the concentration of $MgCl_2$ at 200 mM Tmp and 50 mM Tris/HCl pH 8.5. (**C**) Dependence of ribozyme activity on the pH at 200 mM Tmp and 100 mM $MgCl_2$. (**D**) Dependence of ribozyme activity on reaction temperature at 200 mM Tmp, 100 mM $MgCl_2$, and 50 mM Tris/HCl pH 8.5. In all cases, error bars are standard deviations of triplicate experiments.

tified ribozyme. Similarly, other libraries of in vitro selected ribozymes can now be tested with the same approach, each using their own selection scheme.

It is interesting to deliberate whether computational tools would be able to identify catalytic motifs within High Throughput Sequencing data of in vitro selected RNA populations. One possibility for such a procedure is to use sequence conservation data in each of the clusters, and thereby isolate those sequence fragments that are highly conserved. HTS data on our own in vitro selection were not deep enough to allow such analysis (39). However, the high mutagenic rates, and defined nucleotide bias used in 'DNA shuffling' are able to tease out the secondary / tertiary structure of a specific ribozyme with a limited depth of sequencing (50). With the advance of more powerful High Throughput Sequencing platforms it may become possible to efficiently identify co-variations even from in vitro selected libraries with low mutagenesis. While this may generate a computational alternative to the experimental approach described here, it would be limited by identifying only conserved structural motifs as opposed to catalytic function.

The experimental method described in this study could generate a bias to identify motifs that end in, or are followed by GC-rich sequences. The reason is that such sequences would generate more stable duplexes with the randomized N11-region on the primer used in the Klenow extension reaction (see Figure 1A). Indeed, most of the identified clones carry a C-rich sequence immediately after the motif. While this could reduce the chance of identifying motifs with AU-rich sequences it may not be a serious concern, especially if randomized regions with more than 11 nucleotides, and/or High Throughput Sequencing is used to analyze the selected sequences.

The described procedure uses truncations at the library 3′-terminus, and therefore depends on the 5′-terminus of the library to be the position where the reaction occurs. The majority of in vitro selections for ribozymes have used the 5′-terminus as the reaction site (3,6–10,13,15–20,25,28,29), therefore the described procedure should be widely applicable. However, several in vitro selections focused on internal 2′-hydroxyl groups as reactive groups; for these selections the proposed method would be unsuitable (12,26,27). Other in vitro selections have used the 3′-terminus of the library as the position where the reaction occurs (11,24,36). In these instances, a variation of the described procedure could be successful: Instead of the 3′-primer with a randomized 3′-terminus and a new 3′-constant region for Klenow extension (see Figure 1A), a 5′-primer could be used that truncates the ribozyme 5′-terminus. After PAGE purification of the extension products, necessary elements at the 5′-terminus such as the promoter for T7 RNA polymerase could be added by PCR. In this way, the described procedure would be useful for the isolation of short ribozymes for many types of selections, with reaction centers positioned at the library 5′-terminus and 3′-terminus.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Arvin Akoopie for initial discussions and pilot experiments with Klenow enzyme.

FUNDING

National Aeronautics and Space Administration [80NSSC 19K0467] (in part) issued through the Science Mission Directorate, in Astrobiology/Exobiology (to U.F.M.); partial support for J.T.A. was provided by an NIH Molecular Biophysics Training Grant [T32 GM008326 to E.K.]. Funding for open access charge: NASA.

Conflict of interest statement. None declared.

REFERENCES

1. Ellington,A.D. and Szostak,J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818–822.
2. Tuerk,C. and Gold,L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.
3. Bartel,D.P. and Szostak,J.W. (1993) Isolation of new ribozymes from a large pool of random sequences. *Science*, **261**, 1411–1418.
4. Illangasekare,M., Sanchez,G., Nickles,T. and Yarus,M. (1995) Aminoacyl-RNA synthesis catalyzed by an RNA. *Science*, **267**, 643–647.
5. Wilson,C. and Szostak,J.W. (1995) In vitro evolution of a self-alkylating ribozyme. *Nature*, **374**, 777–782.
6. Lohse,P.A. and Szostak,J.W. (1996) Ribozyme-catalysed amino-acid transfer reactions. *Nature*, **381**, 442–444.
7. Wecker,M., Smith,D. and Gold,L. (1996) In vitro selection of a novel catalytic RNA: characterization of a sulfur alkylation reaction and interaction with a small peptide. *RNA*, **2**, 982–994.
8. Tarasow,T.M., Tarasow,S.L. and Eaton,B.E. (1997) RNA-catalysed carbon-carbon bond formation. *Nature*, **389**, 54–57.
9. Wiegand,T.W., Janssen,R.C. and Eaton,B.E. (1997) Selection of RNA amide synthases. *Chem. Biol.*, **4**, 675–683.
10. Zhang,B. and Cech,T.R. (1997) Peptide bond formation by in vitro selected ribozymes. *Nature*, **390**, 96–100.
11. Unrau,P.J. and Bartel,D.P. (1998) RNA-catalysed nucleotide synthesis. *Nature*, **395**, 260–263.
12. Jenne,A. and Famulok,M. (1998) A novel ribozyme with ester transferase activity. *Chem. Biol.*, **5**, 23–34.
13. Seelig,B. and Jaschke,A. (1999) A small catalytic RNA motif with Diels-Alderase activity. *Chem. Biol.*, **6**, 167–176.
14. Jaeger,L., Wright,M.C. and Joyce,G.F. (1999) A complex ligase ribozyme evolved in vitro from a group I ribozyme domain. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 14712–14717.
15. Robertson,M.P. and Ellington,A.D. (1999) In vitro selection of an allosteric ribozyme that transduces analytes to amplicons. *Nat. Biotechnol.*, **17**, 62–66.
16. Lee,N., Bessho,Y., Wei,K., Szostak,J.W. and Suga,H. (2000) Ribozyme-catalyzed tRNA aminoacylation. *Nat. Struct. Biol.*, **7**, 28–33.
17. Kumar,R.K. and Yarus,M. (2001) RNA-catalyzed amino acid activation. *Biochemistry*, **40**, 6998–7004.
18. Sengle,G., Eisenfuhr,A., Arora,P.S., Nowick,J.S. and Famulok,M. (2001) Novel RNA catalysts for the Michael reaction. *Chem. Biol.*, **8**, 459–473.
19. Rogers,J. and Joyce,G.F. (2001) The effect of cytidine on the structure and function of an RNA ligase ribozyme. *RNA*, **7**, 395–404.
20. Tsukiji,S., Pattnaik,S.B. and Suga,H. (2003) An alcohol dehydrogenase ribozyme. *Nat. Struct. Biol.*, **10**, 713–717.
21. Hati,S., Boles,A.R., Zaborske,J.M., Bergman,B., Posto,A.L. and Burke,D.H. (2003) Nickel2+-mediated assembly of an RNA-amino acid complex. *Chem. Biol.*, **10**, 1129–1137.
22. Ikawa,Y., Tsuda,K., Matsumura,S. and Inoue,T. (2004) De novo synthesis and development of an RNA enzyme. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 13750–13755.
23. Yoshioka,W., Ikawa,Y., Jaeger,L., Shiraishi,H. and Inoue,T. (2004) Generation of a catalytic module on a self-folding RNA. *RNA*, **10**, 1900–1906.
24. Lau,M.W., Cadieux,K.E. and Unrau,P.J. (2004) Isolation of fast purine nucleotide synthase ribozymes. *J. Am. Chem. Soc.*, **126**, 15686–15693.
25. Fusz,S., Eisenfuhr,A., Srivatsan,S.G., Heckel,A. and Famulok,M. (2005) A ribozyme for the aldol reaction. *Chem. Biol.*, **12**, 941–950.
26. Saran,D., Nickens,D.G. and Burke,D.H. (2005) A trans acting ribozyme that phosphorylates exogenous RNA. *Biochemistry*, **44**, 15007–15016.
27. Li,N. and Huang,F. (2005) Ribozyme-catalyzed aminoacylation from CoA thioesters. *Biochemistry*, **44**, 4582–4590.
28. Ryu,Y., Kim,K.J., Roessner,C.A. and Scott,A.I. (2006) Decarboxylative Claisen condensation catalyzed by in vitro selected ribozymes. *Chem. Commun. (Camb.)*, **13**, 1439–1441.
29. Wang,T.P., Su,Y.C., Chen,Y., Liou,Y.M., Lin,K.L., Wang,E.C., Hwang,L.C., Wang,Y.M. and Chen,Y.H. (2012) In vitro selection and characterization of a novel Zn(II)-dependent phosphorothiolate thiolesterase ribozyme. *Biochemistry*, **51**, 496–510.
30. Moretti,J.E. and Muller,U.F. (2014) A ribozyme that triphosphorylates RNA 5′-hydroxyl groups. *Nucleic. Acids. Res.*, **42**, 4767–4778.
31. Lehman,N. and Joyce,G.F. (1993) Evolution in vitro of an RNA enzyme with altered metal dependence. *Nature*, **361**, 182–185.
32. Wang,Q.S. and Unrau,P.J. (2005) Ribozyme motif structure mapped using random recombination and selection. *RNA*, **11**, 404–411.
33. Huang,W. and Ferris,J.P. (2003) Synthesis of 35–40 mers of RNA oligomers from unblocked monomers. A simple approach to the RNA world. *Chem. Commun. (Camb.)*, **12**, 1458–1459.
34. Coleman,T.M. and Huang,F. (2005) Optimal random libraries for the isolation of catalytic RNA. *RNA Biol*, **2**, 129–136.
35. Sabeti,P.C., Unrau,P.J. and Bartel,D.P. (1997) Accessing rare activities from random RNA sequences: the importance of the length of molecules in the starting pool. *Chem. Biol.*, **4**, 767–774.
36. Chumachenko,N.V., Novikov,Y. and Yarus,M. (2009) Rapid and simple ribozymic aminoacylation using three conserved nucleotides. *J. Am. Chem. Soc.*, **131**, 5257–5263.
37. Pasek,M.A., Kee,T.P., Bryant,D.E., Pavlov,A.A. and Lunine,J.I. (2008) Production of potentially prebiotic condensed phosphates by phosphorus redox chemistry. *Angew. Chem. Int. Ed. Engl.*, **47**, 7918–7920.
38. Pasek,M.A., Harnmeijer,J.P., Buick,R., Gull,M. and Atlas,Z. (2013) Evidence for reactive reduced phosphorus species in the early Archean ocean. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 10089–10094.
39. Pressman,A., Moretti,J.E., Campbell,G.W., Muller,U.F. and Chen,I.A. (2017) Analysis of in vitro evolution reveals the underlying distribution of catalytic activity among random sequences. *Nucleic Acids Res.*, **45**, 8167–8179.
40. Mortimer,S.A. and Weeks,K.M. (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.*, **129**, 4144–4145.
41. Jabbari,H. and Condon,A. (2014) A fast and robust iterative algorithm for prediction of RNA pseudoknotted secondary structures. *BMC Bioinformatics*, **15**, 147.
42. Sato,K., Kato,Y., Hamada,M., Akutsu,T. and Asai,K. (2011) IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming. *Bioinformatics*, **27**, i85–93.
43. Andronescu,M.S., Pop,C. and Condon,A.E. (2010) Improved free energy parameters for RNA pseudoknotted secondary structure prediction. *RNA*, **16**, 26–42.
44. Chen,H.L., Condon,A. and Jabbari,H. (2009) An O(n(5)) algorithm for MFE prediction of kissing hairpins and 4-chains in nucleic acids. *J. Comput. Biol.*, **16**, 803–815.
45. Dawson,W.K., Fujiwara,K. and Kawai,G. (2007) Prediction of RNA pseudoknots using heuristic modeling with mapping and sequential folding. *PLoS One*, **2**, e905.
46. Miao,Z., Adamiak,R.W., Blanchet,M.F., Boniecki,M., Bujnicki,J.M., Chen,S.J., Cheng,C., Chojnowski,G., Chou,F.C., Cordero,P. *et al.* (2015) RNA-Puzzles Round II: assessment of RNA structure

prediction programs applied to three large RNA structures. *RNA*, **21**, 1066–1084.
47. Miao,Z., Adamiak,R.W., Antczak,M., Batey,R.T., Becka,A.J., Biesiada,M., Boniecki,M.J., Bujnicki,J.M., Chen,S.J., Cheng,C.Y. *et al.* (2017) RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme. *RNA*, **23**, 655–672.
48. Miao,Z., Adamiak,R.W., Antczak,M., Boniecki,M.J., Bujnicki,J., Chen,S.J., Cheng,C.Y., Cheng,Y., Chou,F.C., Das,R. *et al.* (2020) RNA-Puzzles Round IV: 3D structure predictions of four ribozymes and two aptamers. *RNA*, **26**, 982–995.
49. Akoopie,A. and Muller,U.F. (2016) Lower temperature optimum of a smaller, fragmented triphosphorylation ribozyme. *Phys. Chem. Chem. Phys.*, **18**, 20118–20125.
50. Curtis,E.A. and Bartel,D.P. (2013) Synthetic shuffling and in vitro selection reveal the rugged adaptive fitness landscape of a kinase ribozyme. *RNA*, **19**, 1116–1128.

Figure S1

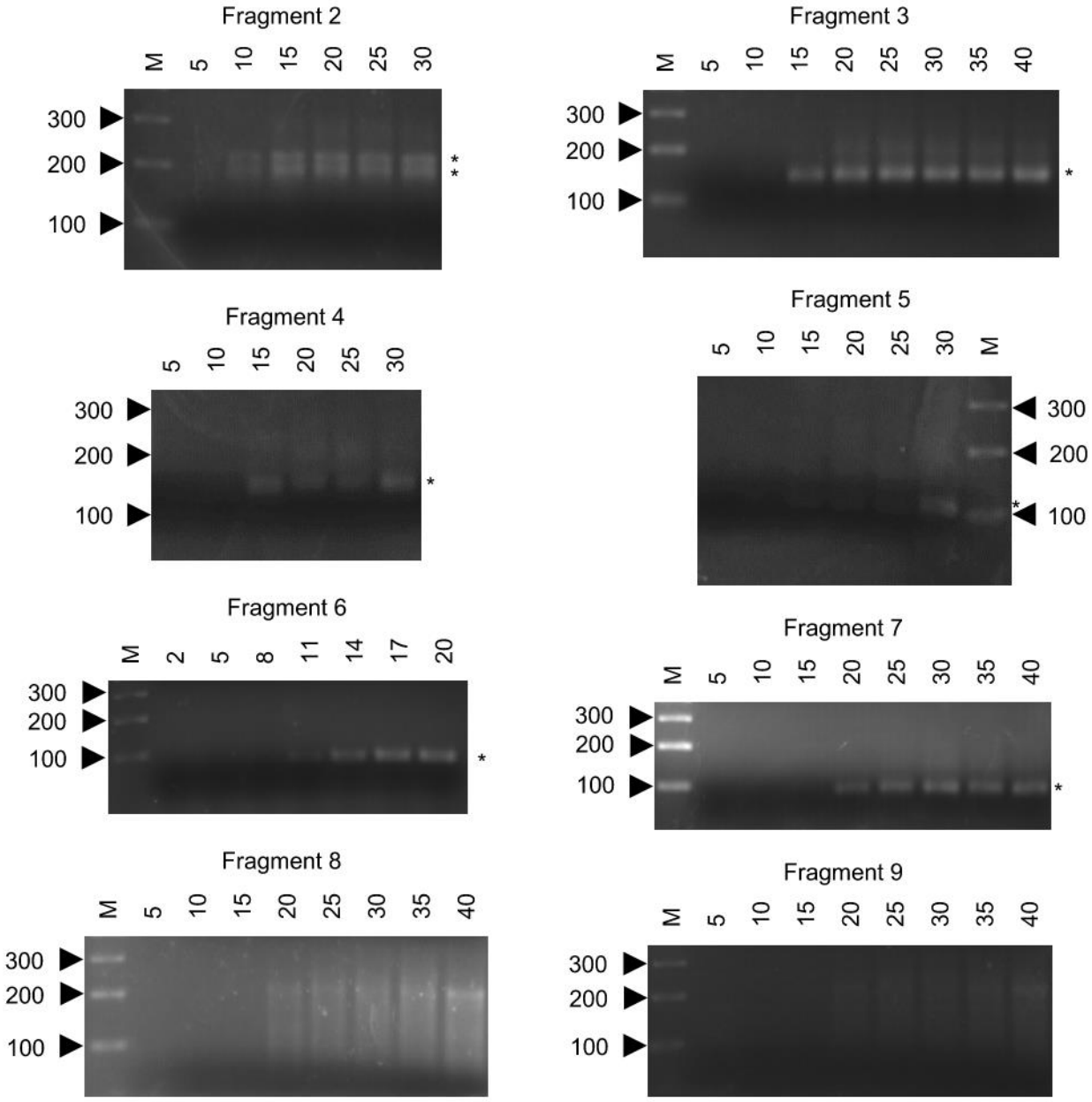


Figure S1. Agarose gel separations of PCR products from gel fragments after one round of selection. The fragment is annotated on the left for each agarose gel. For each PCR amplification, 5 uL samples were taken from a 100 uL PCR reaction after the indicated number of cycles, and separated on 2% Agarose Tris/Acetate/EDTA gel electrophoresis. Gels were stained with ethidium bromide and visualized by UV illumination. Asterisks next to the gels indicate the position of the identified PCR product. Note that the size of the PCR product decreases as expected from ~200 bp (fragment 2) to ~100 bp (fragment 7), while fragment 8 and 9 do not give rise to a clean PCR product band.

Figure S2

Sequences obtained by cloning from the selected pool of fragment 7. The clone names are shown on the left, their sequence on the right. The sequences are shown in the same order as figure 3C to allow easier comparison of the sequences. The 5'-constant region is shown in green, the previously randomized sequence is shown in black, and the 3'-constant region is shown in purple. Biochemically active sequences are labeled with an asterisk. Mutations that distinguish the biochemically active sequences are shown in red.

clone name	sequence
C86	GAGACCGAGATGTT**AACGTCTGCATTGAGTTCTGTACTCTGCCCGCGC**TGATGTAACTACGACTTA
C29	GAGACCGAGATGTT**ATAGGATGAGGATCCGATAGACGGGCTGCTCCCCCC**TGATGTAACTACGACTTA
C21	GAGACCGAGATGTT**CAGGGCGACGCCTATTTAATGTGGGCGTCAGACAAAAACGTATTTTTGCTTCCCATAC CTCTAGCCGTGCATGATGCATGATCTATGTGTCCCCCTTGCCC**TGATGTAACTACGACTTA
C24	GAGACCGAGATGTT**AACGACTTACTTCCCCGGAGACCACTGCGGTTACGTCCAAAAGCACCGAATCTCTTTT AAGGGATGCAAATATGAGATCGGGAACTTATCCCACACTAT**TGATGTAACTACGACTTA
C106	GAGACCGAGATGTT**AACGACTTACTACCCCCCCTTGA**TGTAACTACGACTTA
C96	GAGACCGAGATGTT**TACCGACTTAACAAACTCAGAGTCAAACTCCCGT**TGATGTAACTACGACTTA
C88	GAGACCGAGATGTTT**ACCGACTTAACAAACTCAGAGTCAAACTCCTGAATTATTGCCTCTCCCGCT**TGATGT AACTACGACTTA
C81	GAGACCGAGATGTT**CGTGCGCGGGCCGCGCTACCCTCCCCCCC**TGATGTAACTACGACTTA
C80	GAGACCGAGATGTT**CGTGCGCGGGCCGCGCTACCCTCCCCTCC**TGATGTAACTACGACTTA
C11	GAGACCGAGATGTT**CGTGCGCAGGCCGCGCTACCCTCACCTGA**TGATGTAACTACGACTTA
C12	GAGACCGAGATGTT**AACGCTCGCTGTGGGCAACCTGGTGACCCCTACCCTGCA**TGATGTAACTACGACTTA
C13*	GAGACCGAGATGTT**GCCGTATGCGGGTCTTATTAAAGAGACAC**C**TACTCCC**C**CC**TGATGTAACTACGACTTA
C90*	GAGACCGAGATGTT**GCCGTATGCGGGTCTTATTAAAGAGACATCTACTCC**C**CCCC**TGATGTAACTACGACTTA
C94*	GAGACCGAGATGTT**GCCGTATGCGGGTCTTATTAAAGAGACATCTAC**CA**C**C**ACC**TGATGTAACTACGACTTA
C78*	GAGACCGAGATGTT**GCCGTATGCGGGTCTTATTAAAGAGACATCTAC**C**CCACCC**TGATGTAACTACGACTTA
C75*	GAGACCGAGATGTT**GCCGTATGCGGGTCTTATTAAAGAGACATCTACTCCACC**TGATGTAACTACGACTTA
C91*	GAGACCGAGATGTT**GCCGTATGCGGGTCTTATTAAAGAGACATCTACTCCACCC**TGATGTAACTACGACTTA
C77*	GAGACCGAGATGTT**GCCGTATGCGGGTCTTATTAAAGAGACATCTACT**G**CACCT**TGATGTAACTACGACTTA
C28*	GAGACCGAGATGTT**GCCGTATGCGGGTCTTATTAAAGAGACATCTACT**A**CACCTTCCGCA**TGATGTAACTAC GACTTA
C105*	GAGACCGAGATGTT**GCCGTATGCGGGTCTTATTAAAGAGACAC**C**TAC**C**T**GATGTAACTACGACTTA
C73*	GAGACCGAGATGTT**GCCGTATGCGGGTCTTATTAAAGAGACATC**C**ACTCC**T**CC**TGATGTAACTACGACTTA

Figure S3

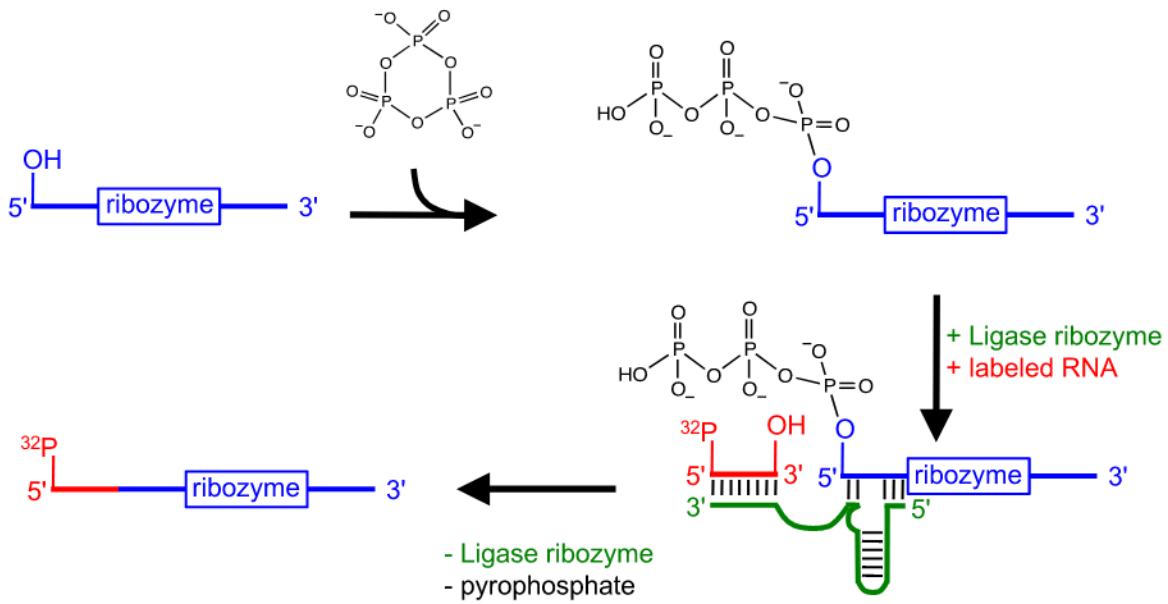


Figure S3. Schematic for the assay on triphoshorylation activity of triphosphorylation ribozymes. After the selected RNA clone (blue) was incubated with trimetaphosphate, ligase ribozyme (green) and 5'-[^{32}P]-radiolabeled substrate RNA (red) were added, and heat renatured. After an additional incubation those ribozymes that carried a 5'-triphosphate were covalently linked to the short, radiolabeled RNAs. After denaturing PAGE, active ribozymes thereby generated a gel shifted band (see figure 3A).

Figure S4

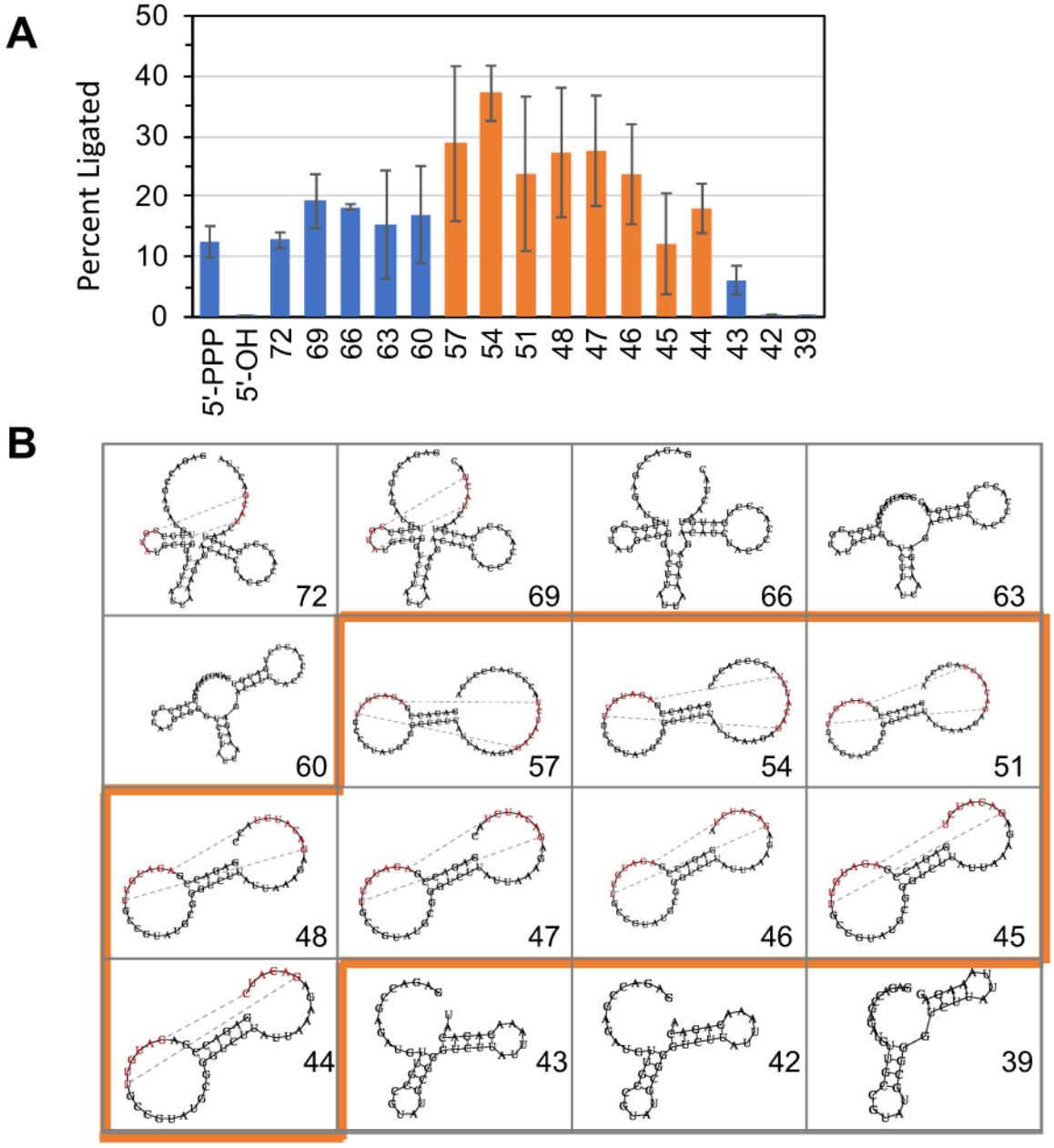


Figure S4. Comparison of the biochemical activity of ribozyme 3'-truncations with their predicted secondary structure. **(A)** Biochemical activity of truncated variants of ribozyme clone c78. In addition to positive control (5'-PPP) and negative control (5'-OH), the length of the 3'-truncation variants is used as label. Error bars are standard deviations from triplicate experiments. Columns are highlighted in orange to correlate with the predicted secondary structure. **(B)** Secondary structure of the truncation variants preditced by vsFold5 (Dawson et al 2007, PLoS ONE 9, e905). This program is able to identify pseudoknots. Each structure is labeled with the ribozyme truncation variant. Structures highlighted in orange correspond to the columns in (A) highlighted in orange.

Chapter 2, in full, is a reprint of the material as it appears in Arriola, J. T.; Müller, U. F. “A combinatorial method to isolate short ribozymes from complex ribozyme libraries”, *Nucleic Acids Res.*, 2020. The dissertation author was the first author of this paper.

Chapter 3

Concurrent Prebiotic Formation of Nucleoside-Amidophosphates and Nucleoside-Triphosphates Potentiates Transition from Abiotic to Biotic Polymerization

DNA Very Important Paper

How to cite: *Angew. Chem. Int. Ed.* **2022**, *61*, e202113625
International Edition: doi.org/10.1002/anie.202113625
German Edition: doi.org/10.1002/ange.202113625

Concurrent Prebiotic Formation of Nucleoside-Amidophosphates and Nucleoside-Triphosphates Potentiates Transition from Abiotic to Biotic Polymerization

Huacan Lin, Eddy I. Jiménez, Joshua T. Arriola, Ulrich F. Müller, and Ramanarayanan Krishnamurthy**

In memory of Leslie Orgel

Abstract: *Polymerization of nucleic acids in biology utilizes 5′-nucleoside triphosphates (NTPs) as substrates. The prebiotic availability of NTPs has been unresolved and other derivatives of nucleoside-monophosphates (NMPs) have been studied. However, this latter approach necessitates a change in chemistries when transitioning to biology. Herein we show that diamidophosphate (DAP), in a one-pot amidophosphorylation-hydrolysis setting converts NMPs into the corresponding NTPs via 5′-nucleoside amidophosphates (NaPs). The resulting crude mixture of NTPs are accepted by proteinaceous- and ribozyme-polymerases as substrates for nucleic acid polymerization. This phosphorylation also operates at the level of oligonucleotides enabling ribozyme-mediated ligation. This one-pot protocol for simultaneous generation of NaPs and NTPs suggests that the transition from prebiotic-phosphorylation and oligomerization to an enzymatic processive-polymerization can be more continuous than previously anticipated.*

Introduction

The prebiotic synthesis of RNA and DNA is considered to be an important stage in the abiotic origins of life.[1] Thus, there have been numerous investigations into the prebiotic formation of nucleosides, nucleotides and subsequent non-enzymatic polymerization and replication of nucleic acids.[2] While the bulk of the studies have focused on RNA,[1a,3] there have been recent investigations of DNA.[1c,4] Central to emergence of both systems is the prebiotic activation of the nucleotides to form longer polymers, both by oligomerization and ligation. Though extant biochemistry uses NTPs,[1a] and there have been attempts to synthesize NTP-analogs under plausible prebiotic conditions,[5] the difficulty in generating NTPs and the fact that they are fairly unreactive for non-enzymatic polymerization[6] has led to the reliance on other activation processes and reactive species such as phosphorimidazolide derivatives.[7] However, in a scenario where the transition from prebiotic chemistry to biotic chemistry via (proto)enzymes capable of handling NTPs is supposed to happen,[8] the question remains as to how NTPs or oligo-TP substrates can be accessed prebiotically in an efficient manner.[1a] Thus, the simultaneous generation of NTPs under plausibly prebiotic conditions can provide a solution to the problem of switching from completely different and incompatible prebiotic chemistries and pathways to what biology is using today and enable a straightforward prebiotic to biotic transition.

Diamidophosphate (DAP) has been proposed as plausible prebiotic phosphorylating reagent,[9] shown to phosphorylate a wide variety of biological molecules and, in the same pot, lead to higher order structures.[10] Of direct interest to nucleic acids, DAP was shown to convert ribo- and deoxyribonucleosides to the corresponding nucleotides and further to corresponding short oligonucleotides.[4a,10a] However, the progression to longer oligonucleotides was not achieved suggesting a limitation of the DAP-mediated chemistries alone in the absence of efficient catalysts. Further, DAP was shown to react with 5′-NMPs and 5′-NDPs to make primarily the corresponding 5′-amidodiphosphate and 5′-amidotriphosphate derivatives, respectively, and in the case of 2′,3′-ribomononucleotides to convert them primarily to the corresponding 2′,3′-cyclic-phosphates, since the phosphate is a better nucleophile than the hydroxyl groups.[10a] Recently, this DAP mediated 2′,3′-cyclic-phosphate formation was utilized to enable an (a) in situ activation of RNA substrates under reaction conditions compatible with, and energetically driving their ligation by, the hairpin ribozyme and (b) iterative enzymatic assembly of long RNAs.[11] While this study showed the compatibility of DAP-generated 2′,3′-cyclic-phosphates with ribozyme catalyzed ligation from the 3′-end of the nucleotides and oligonucleotides, it would be important if these transformations can also proceed from the 5′-end of nucleosides and oligonucleotides which would enable a transition to the chemistries observed in extant biology.

In all of the above DAP-mediated transformations, imidazole additives were shown to increase the efficiency of amidophosphorylation reactions of nucleosides, nucleotides and oligonucleotides by forming the amidophosphorimidazolide intermediate in situ.[10a] These combined results create

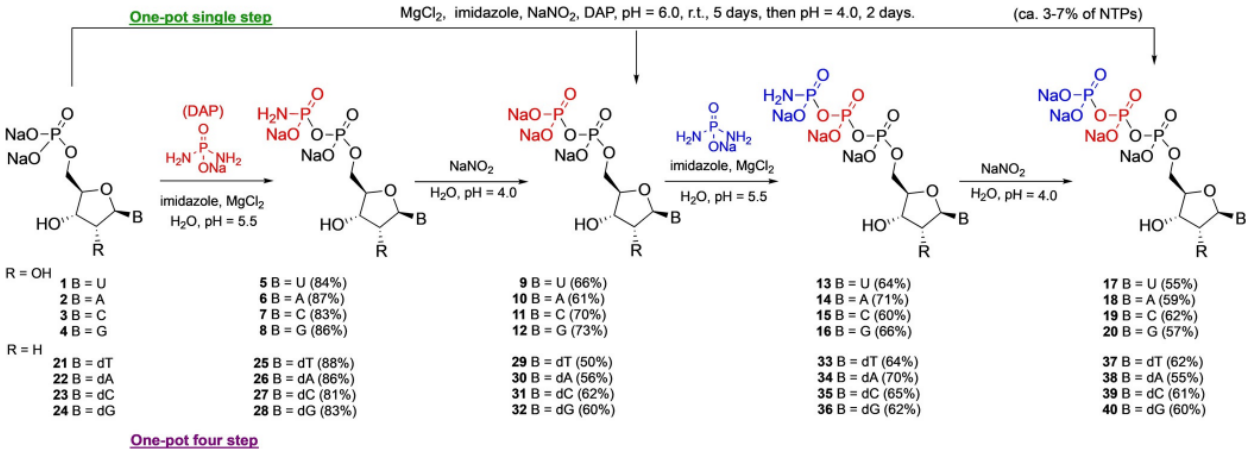
[*] Dr. H. Lin, Dr. E. I. Jiménez, Prof. Dr. R. Krishnamurthy
Department of Chemistry, The Scripps Research Institute
10550 North Torrey Pines Road, La Jolla, CA 92037 (USA)
E-mail: rkrishna@scripps.edu
J. T. Arriola, Prof. Dr. U. F. Müller
Department of Chemistry and Biochemistry, UC San Diego
9500 Gilman Drive, La Jolla, CA 92037 (USA)
E-mail: ufmuller@ucsd.edu
Dr. H. Lin, Prof. Dr. R. Krishnamurthy
NSF-NASA Center for Chemical Evolution
Atlanta, GA 30332 (USA)
Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.202113625.

a plausible setting for the one-pot generation of high energy substrates for prebiotic nucleic acid synthesis starting from nucleosides.[4a, 10a, 11] Providing support for such a scenario, we now show that the same DAP enabled transformations -starting from nucleosides and NMPs- can be used to generate the corresponding crude mixtures of 5′-ribo- and 5′-deoxy-ribo-NTPs, which can act as substrates for proteinaceous-enzymes and ribozymes. The 5′-phosphorylation of nucleosides is achieved by using 2-aminoimidazole[12] instead of the imidazole as the activator.[4a] Furthermore, this one-pot activation chemistry is also shown to be compatible with 5′-phosphate containing oligonucleotides, and the resulting crude oligonucleotide 5′-triphosphate can be used as substrate by a ligase ribozyme. These demonstrations give credence for a model scenario where, by using the same phosphorylating reagent and reaction processes, the plausible transition from prebiotically activated nucleotides to NTPs suitable for more evolved catalysts (polymerases/ligases) is relatively straightforward—avoiding the need to switch from prebiotic processes that are incompatible with biotic molecules and their chemistries.

Results and Discussion

We first investigated the potential of conducting two consecutive reactions in the same pot: (a) amidophosphorylation of the 5′-NMP substrate with DAP which would generate the nucleoside amidodiphosphate derivative followed by (b) hydrolysis of the thus formed amidodiphosphate nucleotide to produce the corresponding nucleoside diphosphate (Scheme 1). Repeating this set of reactions in the same pot was expected to form NTP with (a) the just formed nucleoside diphosphate serving as the substrate for the next round of amidophosphorylation with DAP to form the higher order nucleoside amido-triphosphates (NaTPs) and (b) the hydrolysis of NaTPs, providing the NTPs (Scheme 1). We began by exploring the reaction in the ribo-series of 5′-AMP **2** with DAP in water (in the presence of $MgCl_2$ and imidazole)[10a] to form the corresponding adenosine amidodiphosphate (AaDP, **6** Scheme 1) and followed the reaction by NMR, ion-exchange liquid chromatography (IE-LC) and electrospray ionization mass spectrometry (ESI-MS). The amidophosphorylation reaction of AMP proceeded slowly (days) but with high yield (**6**, 87 %, Scheme 1, Figure 1 a) with no discernible 2′,3′-cyclic-phosphate species as seen by ^{31}P-NMR (Figures S3–S6), which was consistent with our previous observations.[10a] We expected that simply lowering of the pH may help the subsequent hydrolysis of the AaDP to ADP **10**.[13] However, previous studies[10a,14] cautioned that the hydrolysis of amidophosphates, in the lower pH ranges (≈4–6) that are prebiotically plausible,[15] would be inefficient. Inspired by the observations of nitrous acid induced hydrolysis of phosphoramidates,[16] combined with the prebiotic plausible availability of nitrites[17] and their use in other prebiotic transformations,[1e,18] we investigated the hydrolysis of **6** by HNO_2. Treatment of the crude reaction mixture containing **6** at pH 4 with nitrite, generated **10** in 61 % yield. This set the stage for the next step of repeating the one-pot DAP addition followed by HNO_2 induced hydrolysis to see if that would lead to ATP **18** formation. Accordingly, the above crude reaction mixture containing ADP was treated with DAP and it led to the formation of the amidotriphosphate derivative AaTP **14** in 71 % yield. Subsequent hydrolysis with HNO_2 gave ATP (59 %), for an overall yield of 22 % for the one-pot, four-step sequence of reactions starting from AMP **2** (Scheme 1, Figure 1 a). Since the amidodiphosphate **6** and amidotriphosphate **14** have same overall charges as **2** and **10** respectively, they have similar retention times in the anion exchange chromatogram. Thus, in each step of the reaction the identity of the products was established by ^{31}P-NMR and mass-spectral data and for the final ATP (Figures S3–S6, S33–S48), an ESI-HRMS (high resolution MS) confirmed it was the triphosphate derivative (Figure 1 c).

We then applied this one-pot four-step reaction protocol to UMP **1**, CMP **3** and GMP **4** and observed that they also



Scheme 1. The formation of the canonical ribo- and deoxyribo-nucleoside triphosphates via one-pot sequential phosphorylation-hydrolysis scenario. General Scheme of the formation of nucleoside triphosphates via nucleoside amidophosphates and hydrolysis via the one-pot four step (bottom arrow) and one-pot single step (top arrow) sequence of reactions. For details see Figures S1 and S2.

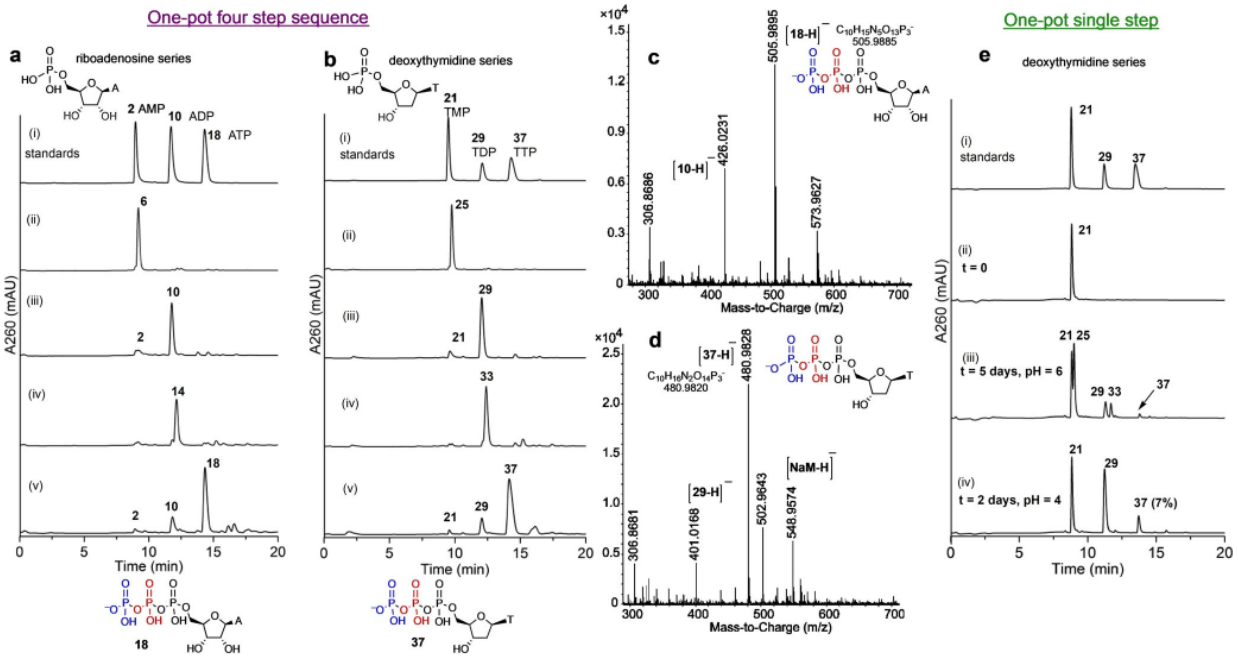


Figure 1. Analytical data confirming the formation of the canonical ribo- and deoxyribo-nucleoside triphosphates by one-pot sequential and single-step phosphorylation-hydrolysis scenario. a and b) Anion-exchange chromatography showing the standards (i) and the progress and production (ii–v) of different adenosine and deoxythymidine phosphate species towards final triphosphate. c and d) High resolution ESI-MS (*m*/*z*) spectrum (-ve mode) of the reaction crude of **18** and **37** respectively. e) Anion-exchange chromatography showing the standard (i) and the production (ii–iv) of the deoxythymidine phosphate species **29** and **37** by a one-pot single-step reaction. % Yield based on spiking respective nucleosides as internal standards in IE-LC. For details, see the Supporting Information (SI).

afforded the corresponding UTP **17**, CTP **19** and GTP **20** respectively, in good overall yields (≈23%) in one-pot, with each of the individual transformation proceeding with efficiencies (average 69%) similar for AMP to ATP conversion (Figures S11–S14). In the final step of forming the ribo-NTP, the main side products identified were the corresponding ribo-NMP (1–4%) and ribo-NDP (6–15%), and though some peaks at higher retention times were observed the identity of these could not be assigned (such as a tetraphosphate derivative which was not observed in the ESI-MS spectra). Furthermore, in the case of AMP, GMP and CMP, monitoring each step of the reaction sequence by ESI-MS showed that there was no discernible hydrolysis of the exocyclic-amino groups to give the corresponding oxo-derivative (inosine, hypoxanthine and uridine derivatives, respectively) under these reaction conditions (Figures S37–S39, S41–S43, S45–S47). This was also confirmed by obtaining the ESI-HRMS of the ATP, GTP and CTP produced from the one-pot reactions (Figures S40, S44, and S48). The one-pot four-steps amidophosphorylation-hydrolysis protocol was tested with the deoxynucleotide dTMP **21** and was found to produce the dTTP **37** in overall 22% yield (Scheme 1, Figure 1b,d). Extending this to all the other dNMPs (**22**–**24**), the efficiencies for each step (average 69%), as well as the overall conversions (average 22%) to the corresponding dNTPs (**38**–**40**), paralleled those observed in the ribonucleotide series (Scheme 1). However, in a prebiotic scenario, a one-pot single-step wherein all the reagents are present together would be considered more realistic—which raises the question of whether the DAP is decomposed by the in situ generated HNO_2[16b] such that it is not available for phosphorylation reactions? Three different conditions were investigated (Figures S19–S29) using dTMP **21** as representative to test this single-step scenario and observed that in all cases, dTTP **37** was formed to the tune of 4–7% along with 18–37% of dTDP **29** as determined by IE-LC (Figure 1e); a tetraphosphate-derivative could have been formed but was not detectable. Other deoxyribo-NMPs and all four ribo-NMPs behaved similarly (Figures S19–S32), suggesting that even in this single step approach the formation of ribo- and deoxyribo-NTP is significant.[5e,f,19]

We next turned to the question as to whether the quality of these crude deoxyribo-NTPs (produced by this one-pot four-step procedure) would be suitable as substrates for biological polymerases, even though they have excess reagents, side-products (deoxyribo-NMP and deoxyribo-NDP) and salts. This was designed as a proof-of-principle experiment to assess the quality of the crude deoxyribo-NTPs produced via this method and should not be interpreted that such primordial chemistries occurred when these highly evolved polymerases were present. We first tested the extension of primer dT_{10} on a dA_{14} template with the crude dTTP generated by the four-step sequence in the presence of Klenow Fragment KF(exo-) DNA polymerase and monitored the reaction by IE-LC (Figure 2a). The efficiency of the reaction and the yields of the extended products paralleled

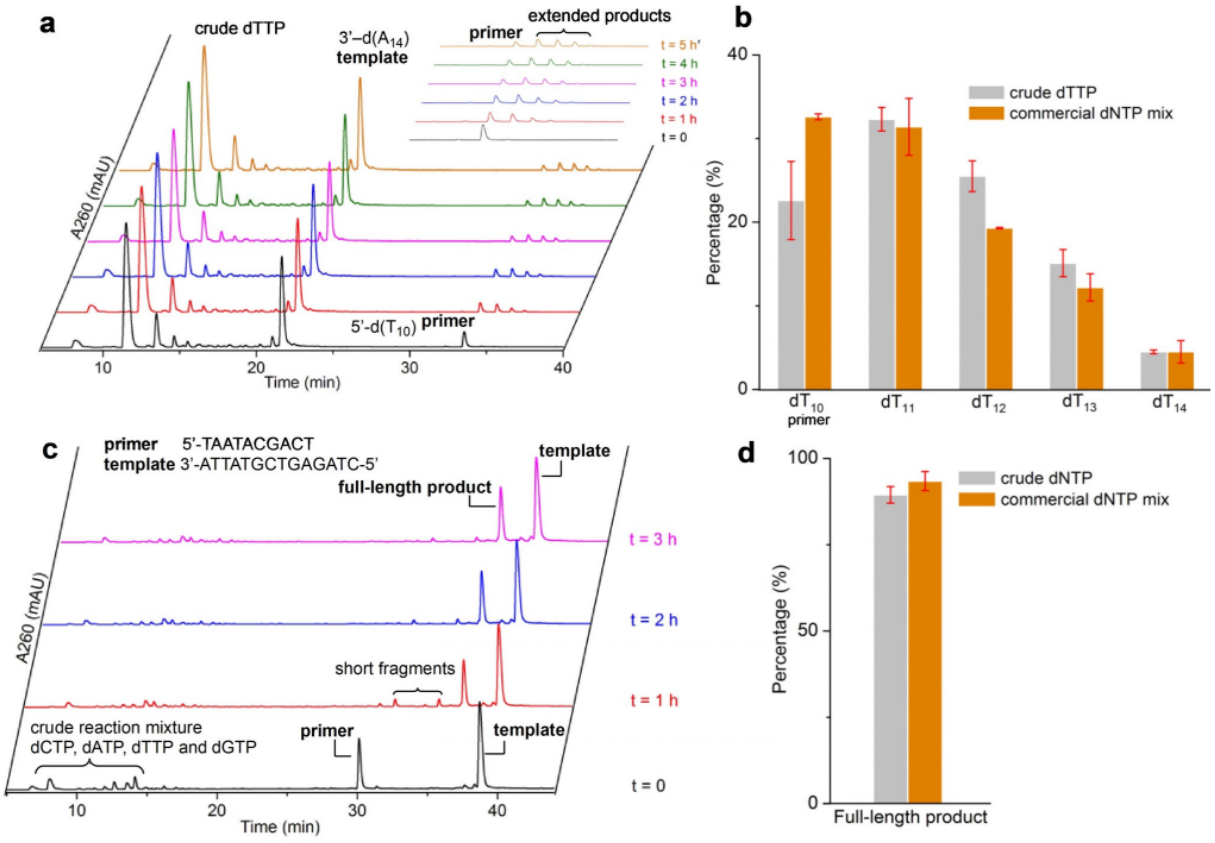


Figure 2. Compatibility of the crude deoxyribo-NTP (from Scheme 1) as substrates for template-mediated primer extension by KF (exo-) DNA polymerase. a) Anion-exchange chromatography of primer extension by KF (exo-) DNA polymerase using crude dTTP reaction mixture. b) Comparison of the yields of products after 5 hours from the primer extension experiment using the crude dTTP reaction mixture versus commercial dTTP mix. c) Anion-exchange chromatography of primer extension by KF (exo-) DNA polymerase using the mix of four deoxyribo-NTP crude mixtures. d) Comparison of the yields of products after 3 hours from the primer extension experiment using the deoxyribo-NTP crude mixture versus commercial deoxyribo-NTP mixture. % Yield of the full-length products was calculated by the anion-exchange chromatogram integration peak ratio of the full-length product and the sum of extended products normalizing the integration area with the total extinction coefficients for each oligonucleotide. Experiments were run in duplicate. The error bars represent s.d. For details see the SI.

those from when the commercial mixture of dNTPs was used under identical conditions (Figure 2b). The reverse combination of using crude dATP synthesized from the four-step single pot scenario with dT_{14} template and dA_{10} primer produced similar results (Figures S70–S72). This was also true for the dNTPs obtained by the one-pot single-step procedures (Figures S77–S80). The partial primer extension of the homo-strand template as opposed to full extension on the mixed sequence template can be explained by the primer annealing to the homo-strand template along the length of the template but only near the 3′-terminus of the mixed sequence template. The general nature of the compatibility was shown by using a mix of the four deoxyribo-NTPs (**37** + **38** + **39** + **40**) crude reactions with a heterogeneous template and the complementary primer (Figure 2c), which in the presence of the DNA polymerase formed the full-length product with efficiency similar to that obtained from the commercial deoxyribo-NTP mixture (Figure 2d). The identity of the full length product was confirmed by isolation and mass spectral data (Figure S74). The replication fidelity was not verified by sequencing since the focus of these studies is on demonstrating the suitability of these crude NTPs as substrates for the modern polymerases.

We extended this line of investigation of compatibility of the crude NTP reaction mixtures (**17**–**20**, generated from the ribose-series **1**–**4**) to include T7 RNA polymerase and observed the transcription to proceed but not as efficiently when compared to the crude deoxyribo-NTP mixture (Figures S81–S83). The T7 RNA polymerase is known to be sensitive to the salt content,[20] which is primarily responsible for the lesser efficiency observed here as documented by the inhibitory effect of the added inorganics to the commercial ribo-NTP reaction (Figures S84–S87). We endeavored to get some quantification of % yields of the truncated and full length products but the low intensity of the peaks coupled with unsatisfactory base-line separation caused issues in getting reliable data.

We then considered whether these crude ribo-NTPs could support ribozyme-mediated RNA polymerization.[21] The polymerase ribozyme was inhibited by higher concentrations

of the crude ribo-NTP mixture, with an optimal concentration around 0.5 mM for the crude ribo-NTPs (Figure S88) On a template that encoded the addition of each of the four ribo-NTPs, the primer was extended over all four positions, suggesting the incorporation of each of the crude ribo-NTPs (Figure 3). Because some of these extensions could be due to misincorporation,[21] four additional variants of this reaction were performed, each lacking a different rNTP (Figure S89). Each of these four resulted in decreased incorporation efficiency at the corresponding positions, supporting the idea that most of the ribozyme-utilized raw ribo-NTPs were incorporated at their templated positions.

Though the prebiotic availability of 5′-nucleosides has been shown by many prebiotic routes,[2,22] we wanted to address the question of whether using the same DAP-based chemistry, starting from the respective nucleosides, would also produce 5′-nucleotides. While our initial work[10a] with ribonucleosides used DAP and imidazole as an activator, subsequent work[4a] showed that DAP is able to convert deoxyribonucleosides efficiently to the corresponding 5′-amidophosphates but in the presence of 2-aminoimidazole (and not imidazole). Thus, we utilized the 2-aminoimdiazole/DAP conditions and investigated dT **41** anticipating first the formation of the 5′-amidophosphate derivative[4a] dTaMP (Figure 4a) which then could be hydrolyzed using HNO_2.

When dT was reacted with DAP and 2-aminoimidazole at 55 °C with wet-dry cycling[4a] (Figure 4a), we noticed peaks and retention times in the IE-LC trace (Figure 4b) that were strikingly similar to those observed from the single-step DAP-HNO_2 procedure (Figure 1e). This suggested that the crude reaction mixture, apart from producing dTaMP, had formed oligonucleotides and poly-phosphorylated species,[4a] which could contain dTTP **37**—implying that the phosphorylation, hydrolysis, and re-phosphorylation (depicted in Scheme 1) may have happened in a single step, without the need for a separate hydrolysis step mediated by HNO_2. To confirm that dTTP **37** may have formed, this reaction crude was subjected to the DNA polymerase catalyzed template mediated primer extension experiment as in Figure 4a and the primer extended products with comparable efficiency was observed (Figure 4c). Use of dT alone showed no extension indicating the absence of any dTTP contamination in the starting nucleoside. Use of either crude reaction containing dTaDP **25** or dTaTP **33** showed very little primer extension products indicating that these are not substrates for the polymerases and any extension products observed were due to the small amounts of dTTP produced in the phosphorylation steps (Figures S96–S100). These control experiments substantiated that dTTP **37** was indeed formed in the dT/DAP/2-aminoimdiazole reaction and that a concurrent amidophosphorylation-hydrolysis-rephosphorylation cycle had taken place. Stimulated by this result, we reacted a mixture of the four deoxynucleosides **41**–**44** under the DAP/2-aminoimidazole-phosphorylation wet-dry cycling protocol at 55 °C (Figure 4d), to see whether this resulting crude reaction mixture is also capable of acting as dNTP sources for the DNA polymerase catalyzed primer extension. And indeed, the full-length primer extension product was observed (Figure 4e) along with the other truncated products. The lower efficiency of primer extension using this crude reaction mixture is attributed to the lesser amount of dNTPs formed directly from the deoxynucleosides by DAP alone when compared to the DAP-HNO_2 protocol.

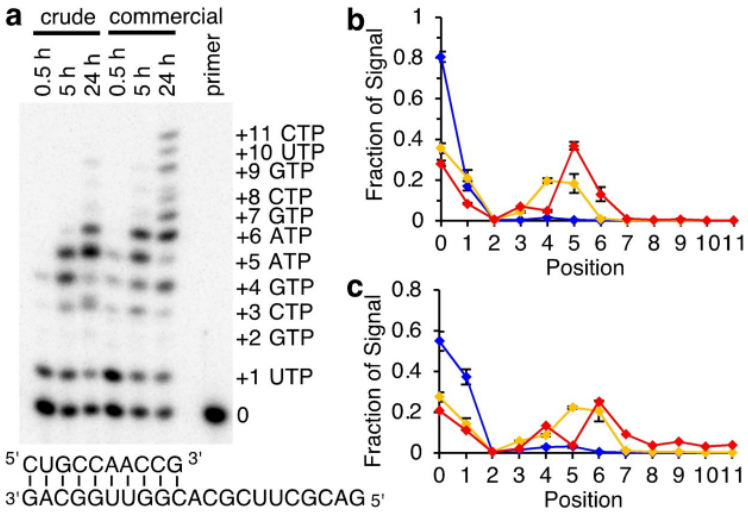


Figure 3. Compatibility of the crude ribo-NTP as substrates for a ribozyme polymerase. a) Ribozyme-mediated polymerization of ribo-NTPs prepared from ribo-NMPs. The autoradiogram shows 3′-extension products of the shown RNA primer (upper strand), templated by the template (lower strand), after separation by denaturing 20% polyacrylamide gel electrophoresis (PAGE). The number of nucleotides added to the primer, as well as the templated ribo-NTPs added at each position, are shown on the right of the autoradiogram. The incubation time with the polymerase ribozyme (0.5 h to 24 h) is shown on top of the autoradiogram. b) Quantitation of the polymerization products shown in (a) for crude ribo-NTPs. The incubation times are indicated in color with 0.5 hours (blue), 5 hours (orange), and 24 hours (red). c) Quantitation of polymerization products shown in (a) for commercial ribo-NTPs. Error bars are standard deviations from triplicate experiments.

The hydrolysis of the 5′-amidophosphate (in the absence of HNO_2) is slow[13] and, this in turn limits the formation of the 5′-dNMP which is needed to form the next intermediates leading finally to dNTP. Importantly, the crude dNTPs synthesized in a single step are functional as substrates for polymerases, despite being in a mixture of unreacted nucleosides, polynucleotide and oligonucleotides (Figure 4a). We observed (^{31}P-NMR, Figures S104–S109) that indeed the corresponding 5′-amidophosphates (along with 2′-, 3′- and 5′-phosphates and 2′,3′-cyclicphosphates) of these ribonucleosides were formed and the IE-LC traces showed peaks with retention times that suggested the formation of species with 3 negative charges, such as NTPs (Figure S110).

Inspired by this observation, we turned our attention to the ribonucleoside series and applied the DAP one-pot procedure with 2-aminoimidazole under paste conditions[4a,10a]

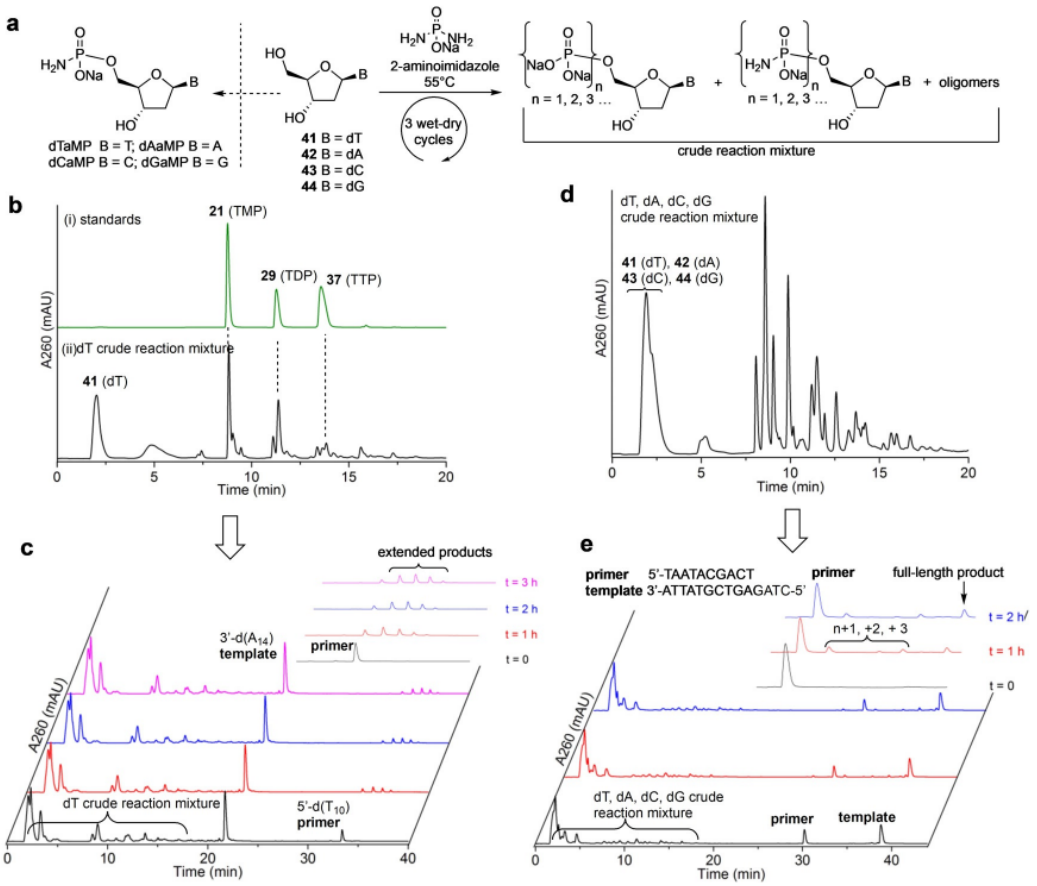


Figure 4. Crude dNTPs formed from deoxyribonucleosides are substrates for DNA polymerases. a) General Scheme of DAP mediated nucleoside phosphorylation via wet-dry cycling producing dNTPs among other products. b) Anion-exchange chromatography showing the production of **37** and other deoxythymidine phosphate species from **41**. c) Anion-exchange chromatography of primer extension by KF (exo-) DNA polymerase using crude dTTP reaction mixture. d) Anion-exchange chromatography indicating the production of **37** + **38** + **39** + **40** (peaks 12–15 min.) and other deoxynucleoside phosphate species from a mixture of deoxynucleosides. e) Anion-exchange chromatography of primer extension by KF (exo-) DNA polymerase using the dNTP crude mixture. For details see the SI.

at room temperature to the individual nucleosides (rA, rU, rG and rC). Since the use of 2-aminoimidazole had produced the 5′-phosphorylated species in the deoxyribose-series[4a] and above, we anticipated that in the ribonucleoside series we may see more of a bias towards the 5′-phosphorylation over the previously reported[10a] 2′,3′-cyclicphosphate formation with imidazole as the activator. We subjected the mixture of ribonucleosides (rA + rU + rG + rC) to the same DAP wet-dry cycle one-pot procedure that was used for deoxynucleosides (as in Figure 4). An LC/MS analysis of the mixture showed the formation of corresponding monophosphates and diphosphates with little triphosphates (Figures S112–S114). When this crude reaction mixture was tested with the either the T7 RNA polymerase or the ribozyme polymerase set-up, no polymerization was detected (Figures S115–S116). The absence of polymerization from ribonucleoside-generated reaction mixture can be attributed to (a) the presumably lower amount of ribo-NTP in ribonucleoside-generated reaction mixture, which caused a larger fraction of inhibitory components in the crude ribo-NTP mixture, and (b) the greater sensitivity of the enzyme and ribozyme to salts and these side-products in the crude reaction mixture.[20,23] These observations suggest that other conditions, such as (a) increasing the hydrolysis of the 5′-amidophosphate derivatives to the corresponding 5′-NMPs (with HNO_2) producing more ribo-NTPs directly from the ribo-nucleosides and (b) use of more promiscuous and tolerant polymerases,[23,24] need to be explored.

We then investigated as to whether the DAP protocol with the HNO_2 hydrolysis step, could be used to activate the 5′-phosphorylated oligomers. As a proof-of-principle experiment we tested each of the protocols with a 5′-phosphorylated DNA (dT_{12}) and RNA (rU_{12}) oligomer and examined the results by IE-LC and electrospray ionization time of flight high resolution mass spectrometry (ESI-TOF HRMS) (Figures S118–S131), both of which showed the efficient conversion to the oligo-5′-triphosphate via the various intermediate phosphorylated species (Figure 5). We then tested, briefly, the DAP wet-dry cycle one-pot procedure on 5′-phosphorylated DNA (dT_{12}) and observed by IE-LC and

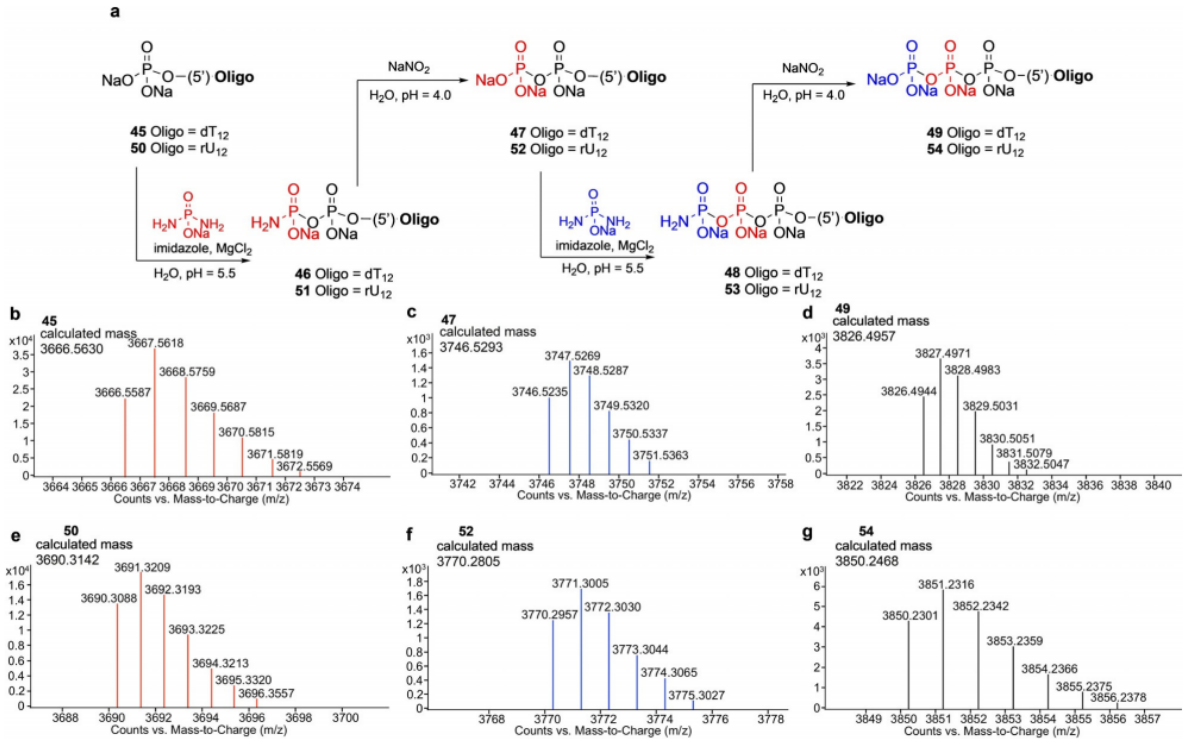


Figure 5. Formation of dT_{12} and rU_{12} 5′-triphosphate by one-pot sequential phosphorylation-hydrolysis scenario under plausible prebiotic conditions. a) General Scheme of the formation of dT_{12} and rU_{12} 5′-triphosphate via sequential amidophosphorylation and hydrolysis. b–d) ESI-TOF-HRMS (*m*/*z*) spectrum (-ve mode) of **45**, the reaction crude of **47** and **49** respectively confirming the identity of the phosphorylated oligonucleotides dT_{12} 5′-mono-, di- and triphosphate. e–g) High resolution mass-to-charge (*m*/*z*) spectrum (-ve mode) of **50**, the reaction crude of **52** and **54** respectively confirming the identity of the phosphorylated oligonucleotides rU_{12} 5′-mono-, di- and triphosphate. For details see the SI.

Matrix Assisted Laser Desorption/Ionization-Time of Flight (MALDI-TOF) mass spectrometry, the formation of the corresponding diphosphate and triphosphate derivatives (Figures S133–S136). These results suggests that the single-pot phosphorylation protocol is compatible also at the level of oligonucleotides without deleterious effect, which bodes well for the in situ activation for polymerization or ligation.[11,25] Therefore, as a proof-of-principle experiment, we subjected an RNA 14mer with a 5′-phosphate (5′-^{p}GAGACCGAGAUGUU-3′) to the one-pot DAP-HNO_2 protocol (Scheme 1) and produced the crude reaction mixture containing the corresponding 5′-triphosphorylated (5′-^{ppp}GAGACCGAGAUGUU-3′) oligonucleotide, whose identity was confirmed by ion-exchange and mass-spectroscopy (Figures S138–S143).

We then considered whether this crude mixture of 5′-phosphorylated oligonucleotides contained enough polyphosphorylated species to be used as substrate by a ribozyme. To test this idea, the crude reaction products from the DAP-mediated phosphorylation of the RNA 14mer 5′-phosphate were incubated with a variant of the R3C ligase ribozyme[26] and a 5′-radiolabeled 16mer oligonucleotide (Figure 6). After separation of the products by denaturing PAGE, the crude reaction products resulted in 48 % of the 16mer being shifted to the expected position of the ligation product, while no detectable shifts resulted from the corresponding 5′-hydroxyl

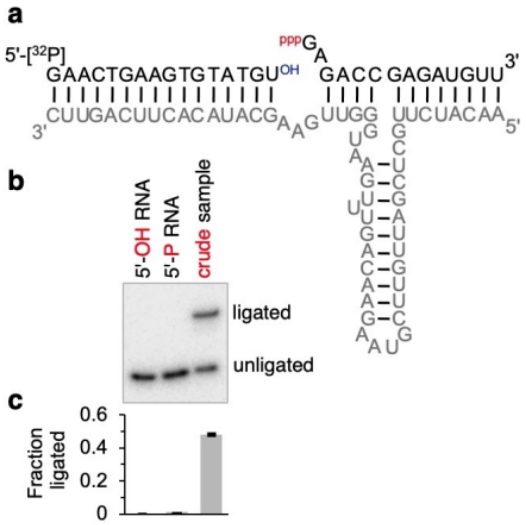


Figure 6. Compatibility of the crude ribo-oligonucleotide-5′-TPs as substrates with a ribozyme ligase. a) Secondary structure schematic for the ligase ribozyme catalyzing the nucleophilic attack of the 3′-hydroxyl group (blue) of a 5′-radiolabeled oligonucleotide (top left) to the 5′-polyphosphate (red) of an RNA 14mer. b) Autoradiogram of PAGE-separated ligase ribozyme reaction products with a control of 5′-hydroxyl RNA, 5′-phosphate RNA, and RNA from a crude reaction with DAP-mediated phosphorylation of the same RNA 5′-phosphate. c) Quantitation of the reaction products. Error bars are standard deviations from triplicate experiments, and may be smaller than the resolution of the Figure.

or 5′-phosphate RNA. These results showed that the crude mixture of 5′-phosphorylated oligonucleotides generated from the DAP-mediated activation of oligomer 5′-phosphates were compatible with ribozyme-mediated ligation.

Since we observed that cyclic-trimetaphosphate (cTMP) is also formed in the above reactions starting with DAP (Figures S3–S10),[10a] we considered whether cTMP,[27] can also be the triphosphorylating agent under these reaction conditions. Thus, dT was subjected to the same reaction conditions as in Figure 4a but using cTMP (in place of DAP). As seen by ^{31}P-NMR and IE-LC, cTMP was unable phosphorylate the nucleosides under the wet-dry cycling conditions (where DAP was able to do so) and the reaction mixtures produced from cTMP were not successful in enzymatic primer-extension studies (Figures S144–S149). This confirmed that DAP, which can be formed under prebiotically plausible conditions,[10] is the active phosphorylating agent and not the cTMP formed during the reaction, though it does not imply that cTMP cannot play a role under different reaction conditions.[5f,25,28]

Conclusion

The observations reported here show the generation of 5′-amidophosphorylated nucleosides, NMPs, NDPs and NTPs, and corresponding RNA oligomeric species in a one-pot scenario using the same prebiotically plausible reactants and reaction conditions. The nucleosides are first amidophosphorylated to give the corresponding 5′-amidophosphates, which are then hydrolyzed to the 5′-NMPs, which upon repeated amidophosphorylation-hydrolysis leads to 5′-NTPs and polyphosphate species.[4a] In the reactions with NMPs, little 2′- or 3′-phosphorylation products were observed, presumably due to higher nucleophilicity of the 5′-phosphate compared to the OH groups.[4a] The 5′-phosphorylated products formed under these conditions are found to be suitable for both ribozyme-mediated oligomerization, ligation and enzymatic polymerization. While the amidophosphorylated and cyclic-phosphate nucleosides have been shown to form short oligomers under non-enzymatic and plausible prebiotic conditions,[4a,10a] the concurrently formed NTPs have now been shown to be suitable substrates for both proteinaceous- and ribozyme-polymerase mediated polymerization.

The demonstration of forming a spectrum of phosphorylated substrates that are relevant both from prebiotic- as well as biotic-chemistries, under the same scenario, increases the chance of the same type of chemistries and reactions to enable the transition to biology (Figure 7).[29] In other words, the results demonstrated here gain significance in the context of how non-enzymatic prebiotic reactions can give rise to scenarios capable of transitioning to biotic chemistries—chemistries that can lead to abiotic processes conducive for the emergence of life's biochemistry.

In this specific case, the DAP-mediated prebiotic phosphorylation operates not only at the monomer level (such as non-enzymatic phosphorylation and oligomerization of nucleosides[4a,10a]), but also at the level of oligomers (as activation of oligonucleotides for ribozyme-mediated ligation[11]) and, at the same time, produce the NTPs that are compatible with the ribozyme- and enzyme-polymerases. The chemistry of DAP spans the whole spectrum, thus enabling a scenario that the same type of phosphorylation chemistry can operate from bottom-to-top and eliminate the need for a change in the phosphorylating versus activating agent and reaction conditions, with different stages of chemical evolution (Figure 7). The appearance of RNA and peptide based proto-enzyme catalysts capable of using the range of phosphorylated substrates can act as factors (in addition to permeability in protocellular environments[30]) determining which substrates will be selected. It is noteworthy that the DAP-chemistry is operative on both ribo- and deoxy-ribonucleosides which augurs well for a side-by-side co-evolution of

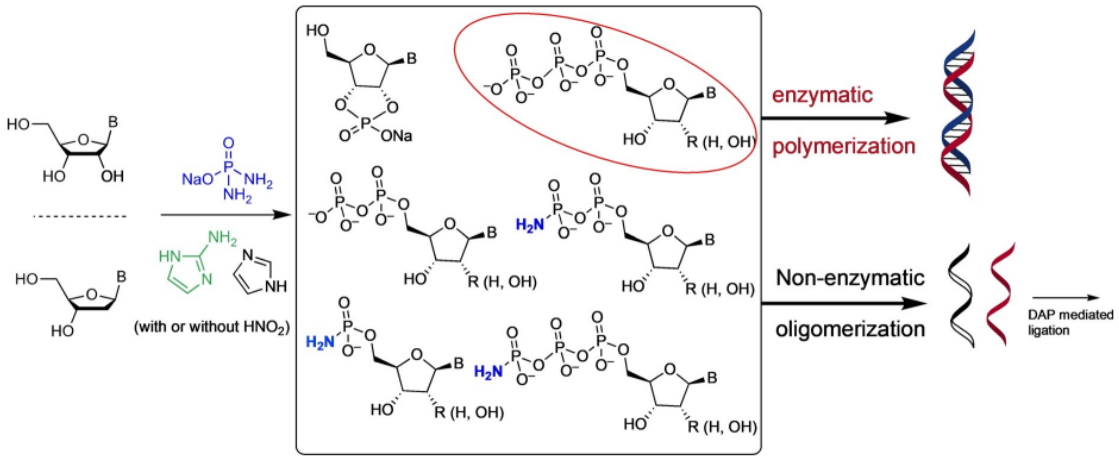


Figure 7. Prebiotic phosphorylation and activation chemistry of DAP is compatible over a range of monomers to oligomers enabling a "simplification of the transition from chemistry to biology".[33] And, thus, has the potential to enable a continuity from prebiotic to biotic processes by the same type of chemistry—starting from monomers to oligomeric systems—involving organocatalysts, proto-enzymes and polymerases.

RNA and DNA, an idea that has found recent experimental support.[1e,4,31] And, furthermore, the chemistry described here can be exploited from an application point of view to other substrates complementing existing methods[32] and may become a general paradigm for preparing higher phosphorylated analogs of a variety of biomolecules in aqueous medium for screening purposes.

Acknowledgements

This work was jointly supported by NSF and the NASA Astrobiology Program under the Center for Chemical Evolution, CHE-150421, grant from the Simons Foundation to R.K. (327124FY19), and a NASA grant 80NSSC19K0467 to U.F.M.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: diamidophosphate · DNA · phosphorylation · prebiotic chemistry · RNA

[1] a) M. P. Robertson, G. F. Joyce, *Cold Spring Harbor Perspect. Biol.* **2012**, *4*, a003608; b) T. R. Cech, *Cold Spring Harbor Perspect. Biol.* **2012**, *4*, a006742; c) A. Lazcano, *BIO Web of Conferences* **2015**, *4*, 00013; d) J. P. Dworkin, A. Lazcano, L. Miller Stanley, *J. Theor. Biol.* **2003**, *222*, 127 – 134; e) J. Xu, V. Chmela, N. J. Green, D. A. Russell, M. J. Janicki, R. W. Gora, R. Szabla, A. D. Bond, J. D. Sutherland, *Nature* **2020**, *582*, 60 – 66.

[2] M. Yadav, R. Kumar, R. Krishnamurthy, *Chem. Rev.* **2020**, *120*, 4766 – 4805.

[3] a) L. E. Orgel, *Origins Life Evol. Biospheres* **2003**, *33*, 211 – 218; b) A. Pressman, C. Blanco, I. A. Chen, *Curr. Biol.* **2015**, *25*, R953 – R963.

[4] a) E. I. Jiménez, C. Gibard, R. Krishnamurthy, *Angew. Chem. Int. Ed.* **2021**, *60*, 10775 – 10783; *Angew. Chem.* **2021**, *133*, 10870 – 10878; b) J. Xu, N. J. Green, C. Gibard, R. Krishnamurthy, J. D. Sutherland, *Nat. Chem.* **2019**, *11*, 457 – 462.

[5] a) R. Lohrmann, *J. Mol. Evol.* **1975**, *6*, 237 – 252; b) G. J. Handschuh, R. Lohrmann, L. E. Orgel, *J. Mol. Evol.* **1973**, *2*, 251 – 262; c) R. Österberg, L. E. Orgel, R. Lohrmann, *J. Mol. Evol.* **1973**, *2*, 231 – 234; d) R. Reimann, G. Zubay, *Origins Life Evol. Biospheres* **1999**, *29*, 229 – 247; e) H. J. Kim, S. A. Benner, *Astrobiology* **2021**, *21*, 298 – 306; f) F. Chizzolini, A. D. Kent, L. F. M. Passalacqua, A. Luptak, *ChemBioChem* **2021**, *22*, 2098 – 2101.

[6] R. Rohatgi, D. P. Bartel, J. W. Szostak, *J. Am. Chem. Soc.* **1996**, *118*, 3340 – 3344.

[7] a) J. Oró, E. Stephen-Sherwood in *Cosmochemical Evolution and the Origins of Life*, *Vol. 1* (Eds.: J. Oró, S. L. Miller, C. Ponnamperuma, R. S. Young), Springer, Dordrecht, **1974**, pp. 159 – 172; b) L. E. Orgel, *Nature* **1992**, *358*, 203 – 209; c) L. E. Orgel, *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99 – 123; d) T. Walton, W. Zhang, L. Li, C. P. Tam, J. W. Szostak, *Angew. Chem. Int. Ed.* **2019**, *58*, 10812 – 10819; *Angew. Chem.* **2019**, *131*, 10926 – 10933; e) E. Kervio, M. Sosson, C. Richert, *Nucleic Acids Res.* **2016**, *44*, 5504 – 5514; f) P. G. Higgs, *Life* **2016**, *6*, 24.

[8] a) D. Bartel, J. Szostak, *Science* **1993**, *261*, 1411 – 1418; b) G. F. Dolan, A. Akoopie, U. F. Muller, *PLoS One* **2015**, *10*, e0142559.

[9] a) R. Krishnamurthy, S. Guntha, A. Eschenmoser, *Angew. Chem. Int. Ed.* **2000**, *39*, 2281 – 2285; *Angew. Chem.* **2000**, *112*, 2369 – 2373; b) M. Karki, C. Gibard, S. Bhowmik, R. Krishnamurthy, *Life* **2017**, *7*, 32; c) A. Osumah, R. Krishnamurthy, *ChemBioChem* **2021**, *22*, 3001 – 3009.

[10] a) C. Gibard, S. Bhowmik, M. Karki, E. K. Kim, R. Krishnamurthy, *Nat. Chem.* **2018**, *10*, 212 – 217; b) C. Gibard, I. B. Gorrell, E. I. Jimenez, T. P. Kee, M. A. Pasek, R. Krishnamurthy, *Angew. Chem. Int. Ed.* **2019**, *58*, 8151 – 8155; *Angew. Chem.* **2019**, *131*, 8235 – 8239.

[11] E. Y. Song, E. I. Jimenez, H. Lin, K. Le Vay, R. Krishnamurthy, H. Mutschler, *Angew. Chem. Int. Ed.* **2021**, *60*, 2952 – 2957; *Angew. Chem.* **2021**, *133*, 2988 – 2993.

[12] L. Li, N. Prywes, C. P. Tam, D. K. O'Flaherty, V. S. Lelyveld, E. C. Izgu, A. Pal, J. W. Szostak, *J. Am. Chem. Soc.* **2017**, *139*, 1810 – 1813.

[13] J. Tomasz, *J. Carbohydr. Nucleosides Nucleotides* **1981**, *8*, 557 – 572.

[14] A. Simoncsits, J. Tomasz, *Nucleic Acids Res.* **1975**, *2*, 1223 – 1233.

[15] J. Kua, J. Bada, *Origins Life Evol. Biospheres* **2011**, *41*, 553—558.

[16] a) W. P. Jencks, M. Gilchrist, *J. Am. Chem. Soc.* **1964**, *86*, 1410 – 1417; b) Z. Jerman, H. Rathánová, F. Markalous, *Collect. Czech. Chem. Commun.* **1969**, *34*, 3991 – 3994.

[17] a) D. P. Summers, S. Chang, *Nature* **1993**, *365*, 630 – 633; b) D. P. Summers, *Origins Life Evol. Biospheres* **1999**, *29*, 33 – 46; c) S. Ranjan, Z. R. Todd, P. B. Rimmer, D. D. Sasselov, A. R. Babbin, *Geochem. Geophys. Geosyst.* **2019**, *20*, 2021 – 2039.

[18] S. Becker, J. Feldmann, S. Wiedemann, H. Okamura, C. Schneider, K. Iwan, A. Crisp, M. Rossa, T. Amatov, T. Carell, *Science* **2019**, *366*, 76 – 82.

[19] Y. Yamagata, T. Matsukawa, T. Mohri, K. Inomata, *Nature* **1979**, *282*, 284 – 286.

[20] M. Maslak, C. T. Martin, *Biochemistry* **1994**, *33*, 6918 – 6924.

[21] W. K. Johnston, P. J. Unrau, M. S. Lawrence, M. E. Glasner, D. P. Bartel, *Science* **2001**, *292*, 1319 – 1325.

[22] a) R. Lohrmann, L. E. Orgel, *Science* **1971**, *171*, 490 – 494; b) R. Lohrmann, L. E. Orgel, *Science* **1968**, *161*, 64 – 66.

[23] J. Attwater, S. Tagami, M. Kimoto, K. Butler, E. T. Kool, J. Wengel, P. Herdewijn, I. Hirao, P. Holliger, *Chem. Sci.* **2013**, *4*, 2804 – 2814.

[24] E. Janzen, C. Blanco, H. Peng, J. Kenchel, I. A. Chen, *Chem. Rev.* **2020**, *120*, 4879 – 4897.

[25] J. E. Moretti, U. F. Müller, *Nucleic Acids Res.* **2014**, *42*, 4767 – 4778.

[26] J. Rogers, G. F. Joyce, *RNA* **2001**, *7*, 395 – 404.

[27] a) A. W. Schwartz, *J. Chem. Soc. D* **1969**, 1393a; b) C. Cheng, C. Fan, R. Wan, C. Tong, Z. Miao, J. Chen, Y. Zhao, *Origins Life Evol. Biospheres* **2002**, *32*, 219 – 224.

[28] K. Gao, L. E. Orgel, *Origins Life Evol. Biospheres* **2000**, *30*, 45 – 51.

[29] R. Krishnamurthy, *Chem. Eur. J.* **2018**, *24*, 16708 – 16715.

[30] P. A. Monnard, D. W. Deamer, *Origins Life Evol. Biospheres* **2001**, *31*, 147 – 155.

[31] a) S. Bhowmik, R. Krishnamurthy, *Nat. Chem.* **2019**, *11*, 1009 – 1018; b) J. Xu, N. J. Green, D. A. Russell, Z. Liu, J. D. Sutherland, *J. Am. Chem. Soc.* **2021**, *143*, 14482 – 14486.

[32] a) I. Zlatev, T. Lavergne, F. Debart, J. J. Vasseur, M. Manoharan, F. Morvan, *Org. Lett.* **2010**, *12*, 2190 – 2193; b) J. Singh, A. Ripp, T. M. Haas, D. Qiu, M. Keller, P. A. Wender, J. S. Siegel, K. K. Baldridge, H. J. Jessen, *J. Am. Chem. Soc.* **2019**, *141*, 15013 – 15017.

[33] M. W. Powner, S. L. Zheng, J. W. Szostak, *J. Am. Chem. Soc.* **2012**, *134*, 13889 – 13895.

Manuscript received: October 8, 2021
Accepted manuscript online: November 4, 2021
Version of record online: November 26, 2021

Chapter 3, in full, is a reprint of the material as it appears in Lin, H.; Jimenez, E. I.; Arriola, J. T.; Müller, U. F.; Krishnamurthy, R.; “Concurrent Prebiotic Formation of Nucleoside-Amidophosphates and Nucleoside-Triphosphates Potentiates Transition from Abiotic to Biotic Polymerization”, *Angew. Chem. Int. Ed.*, 2021. The dissertation author was the third author of this paper. He contributed by investigating if ribozymes could use crude mixtures of nucleoside triphosphates generated by diamidophosphate.

Chapter 4

A GTP-Synthesizing Ribozyme Selected by Metabolic Coupling to an RNA Polymerase Ribozyme

BIOCHEMISTRY

A GTP-synthesizing ribozyme selected by metabolic coupling to an RNA polymerase ribozyme

Arvin Akoopie, Joshua T. Arriola, Douglas Magde, Ulrich F. Müller*

Copyright © 2021 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY).

Synthesis of RNA in early life forms required chemically activated nucleotides, perhaps in the same form of nucleoside 5′-triphosphates (NTPs) as in the contemporary biosphere. We show the development of a catalytic RNA (ribozyme) that generates the nucleoside triphosphate guanosine 5′-triphosphate (GTP) from the nucleoside guanosine and the prebiotically plausible cyclic trimetaphosphate. Ribozymes were selected from 1.6×10^{14} different randomized sequences by metabolically coupling 6-thio GTP synthesis to primer extension by an RNA polymerase ribozyme within 10^{16} emulsion droplets. Several functional RNAs were identified, one of which was characterized in more detail. Under optimized reaction conditions, this ribozyme produced GTP at a rate 18,000-fold higher than the uncatalyzed rate, with a turnover of 1.7-fold, and supported the incorporation of GTP into RNA oligomers in tandem with an RNA polymerase ribozyme. These results are discussed in the context of early life forms.

INTRODUCTION

Early life forms likely used catalytic RNAs in central roles before the evolution of encoded protein synthesis (*1*), as judged by the existence of the ribosome (*2*) and ribonuclease (RNase) P (*3*) in every known biological organism. To investigate what functions ribozymes could have fulfilled in early stages of life, in vitro selection experiments (*4*, *5*) were developed to generate ribozymes with functions that seem important for such molecular systems (*6*). One such ribozyme catalyzes template-dependent RNA polymerization using nucleoside 5′-triphosphates (NTPs), with recent ribozyme variants that are capable of synthesizing functional ribozymes (*7*–*9*).

NTPs were likely important in early life because they are used in every known biological organism as the central energy currency and as activated monomers for RNA polymerization, among many other roles (*10*). As activating agents, polyphosphates can be formed by several different prebiotically available routes (*11*–*15*). The prebiotic existence of phosphite—an intermediate in several of these routes—has been confirmed by its direct identification in 3.5-billion-year-old marine sediments (*16*). The most reactive polyphosphate is cyclic trimetaphosphate (cTmp) (*17*), which can react with nucleoside 5′-hydroxyl groups to form NTPs (*18*–*20*). However, the rate of the uncatalyzed reaction is low, with an observed rate of $4 \times 10^{-5}\ M^{-1}\ hour^{-1}$ for adenosine 5′-triphosphate (ATP) at 1 mM adenosine, 10 mM cTmp (pH 7), 150 mM $MgCl_2$, and 20°C; the rate for GTP is about threefold lower (*21*). While this rate increases with higher pH and temperature, those conditions also decrease the lifetime of RNA polymers. Therefore, there would have been an evolutionary benefit for ribozymes to catalyze the 5′-triphosphorylation of nucleosides at neutral pH and moderate temperature. In vitro selection experiments have shown that some ribozymes are able to catalyze the chemistry of this reaction, but those ribozymes catalyze the 5′-triphosphorylation of RNA oligomers (*22*) and do not generate free NTPs.

Ribozymes that generate free NTPs are central for origin-of-life model systems based on catalytic RNAs. The reason is that any metabolism requires metabolites that can freely be exchanged between multiple catalysts within a compartment, such as a cell or a protocell. However, the development of such coupled catalysts by in vitro selection is difficult because freely diffusing metabolites—in this case, NTPs—can escape the ribozymes that generated them and therefore not tag the active RNA sequence for selection. To overcome this hurdle, we established a coupled in vitro selection system in emulsion, where active sequences were tagged within the emulsion droplets by a second ribozyme. In this coupled system, one ribozyme generated guanosine 5′-triphosphate (GTP) from guanosine and cTmp, and a second ribozyme used the GTP to extend an RNA polymer (Fig. 1). Thereby, this study represents an important advance in establishing a metabolism for an RNA-based model system of early life.

Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA 92093, USA.
*Corresponding author. Email: ufmuller@ucsd.edu

RESULTS

In vitro selection

An in vitro selection procedure was designed to obtain ribozymes that catalyze the triphosphorylation of 6-thio guanosine (6sGsn) with cTmp to yield 6′-thio guanosine triphosphate (6sGTP) (Fig. 2A). The RNA library was prepared with a 5′-hydroxyl group, a 5′-terminal constant region, 150 nucleotides of randomized region, a 3′-terminal constant region, and a recognition site for a DNAzyme. To track these pool molecules, a fraction of them were 5′-[^{32}P]–radiolabeled. Because T7 RNA polymerase generates heterogeneous 3′-termini (*23*–*25*) and a precise 3′-terminus with 2′,3′-diols was required for the selection procedure, transcribed RNA pool molecules were processed with a catalytic DNA that generates a 2′,3′-hydroxyl terminus with a defined length (*26*). The resulting gel-purified, weakly radiolabeled pool molecules were used for the selection step. To mediate tagging of RNA pool molecules that catalyzed the reaction of 6sGsn with cTmp to 6sGTP, the pool molecules were heat-renatured with an excess of a polymerase ribozyme variant that had been selected to efficiently use 6sGTP for tagging of an RNA 3′-terminus (*27*).

To compartmentalize individual pool molecules, the aqueous solution was emulsified in an oil phase containing mineral oil and the emulsifier ABIL EM 90 (*28*) at final concentrations of 0.5 μM RNA pool, 3 μM polymerase ribozyme, 50 mM tris-HCl (pH 8.3), 150 mM $MgCl_2$, 200 mM KCl, 50 mM cTmp, and 1 mM 6sGsn. The emulsion droplet diameter was in the range of 150 nm (Fig. 2B),

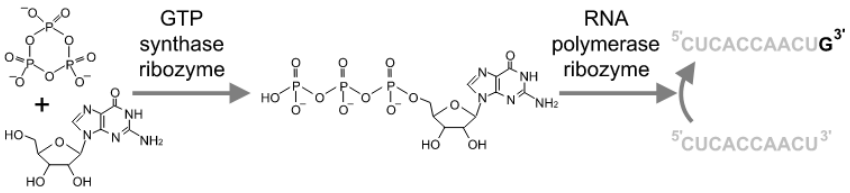


Fig. 1. Metabolic coupling of two ribozymes by GTP. The reaction between cTmp (top left) and guanosine (bottom left) is catalyzed by a ribozyme developed in this study (GTP synthase, GTR1). The resulting GTP is used by a polymerase ribozyme to extend the 3′-terminus of an RNA primer (gray).

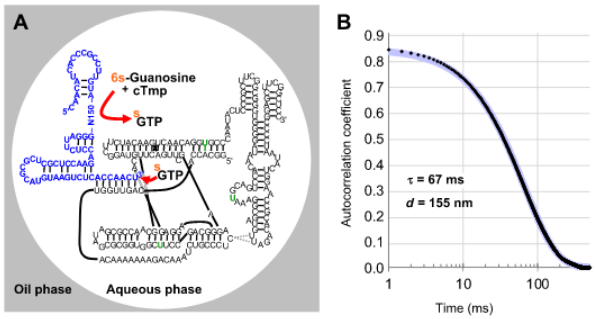


Fig. 2. In emulsio selection setup. (**A**) Secondary structure representation of the in emulsio selection setup. The emulsion consisted of an oil phase matrix (gray) and aqueous droplets (white). Two RNA strands were base-paired to each other within the droplet. The pool RNA (blue) contained 150 nucleotides of a randomized region (N150) and a 3′-terminus that base-paired to a polymerase ribozyme (black). Desired pool molecules catalyzed the reaction of 6sGsn with cTmp to form 6sGTP (upper red arrow). The polymerase ribozyme then used 6sGTP to extend the pool 3′-terminus, thereby thio-tagging the 3′-terminus of the pool RNA (lower red arrow). Three mutations in the polymerase ribozyme increased the rate of 6sGTP ligation (green letters). The secondary structure of the pool construct terminus was predicted by mfold (*32*). The secondary structure of the polymerase ribozyme is based on previous studies (*33*, *34*). (**B**) Autocorrelation function from dynamic light scattering (DLS) analysis of the emulsion used in the selection. The autocorrelation values are plotted as a function of time, with time displayed logarithmically. The data points are shown as small black diamonds. A single-exponential fit (blue) had a time constant of 67 ms, correlating to an aqueous droplet diameter of 155 nm.

which resulted in an average of 0.5 RNA pool molecules and 3 polymerase ribozyme molecules per droplet. This droplet diameter also ensured that the generation of a single molecule of 6sGTP would result in 6sGTP concentration of about 1 μM in that particular droplet. Because the polymerase ribozyme used in this study displayed a 6sGTP ligation rate of 0.012 min^{-1} at 1 μM 6sGTP concentration (*27*), the production of a single 6sGTP molecule was expected to lead to the 3′-tagging of 50% of the co-compartmentalized pool molecules in less than 1 hour. The emulsion was incubated for 6 hours, and the RNA molecules were extracted.

To isolate those pool RNA molecules that were tagged with 6sGTP at their 3′-terminus, they were separated from untagged RNAs by three-layered aminophenyl mercury (APM)–polyacrylamide gel electrophoresis (PAGE) (*29*, *30*). The thio modification of 6sGTP efficiently immobilized the active pool molecules at the mercury interface during APM-PAGE, as observed in a previous in vitro selection experiment using 6sGsn (*31*). After the RNA pool molecules were isolated from the APM interface using dithiothreitol (DTT), they were reverse-transcribed, polymerase chain reaction (PCR)–amplified, and appended with overhanging PCR primers to regenerate the pool, now enriched for active sequences.

Progress of the in emulsio selection

During the selection, several parameters were adjusted (fig. S1). In the first round of selection, the highest concentration of pool RNA molecules was used (500 nM) to be able to cover a large sequence space (1.6×10^{14} independent sequences). An average of 30 copies were used for each pool molecule sequence, corresponding to a total of 5×10^{15} molecules. Because the emulsion in the first round of selection contained 10^{16} droplets, this means that every second droplet contained one pool molecule, on average. In subsequent rounds of selection, the concentration of pool RNA was successively decreased to 50 nM to reduce the likelihood of selecting for interacting pool molecules. The concentration of the polymerase ribozyme was kept constant so that, in all rounds, statistically about 95% of the emulsion droplets contained at least one polymerase ribozyme molecule. The first selection round used 1 mM 6sGsn, about half of its solubility limit, and 50 mM cTmp, to capture even weakly active pool molecules. To enrich for active pool sequences with increased affinity to 6sGsn and cTmp, the concentrations were successively reduced to 50 μM 6sGsn and 0.5 mM cTmp.

The average 6sGsn triphosphorylation activity of the RNA pool during the selection was monitored by 5′-[^{32}P]–radiolabeling a fraction of the pool and phosphorimaging of the APM gels. By measuring the fraction of pool molecules captured at the APM interface, and normalizing for the concentrations of 6sGsn and cTmp, an average pool ligation rate was calculated and plotted as a function of the selection rounds (Fig. 3A). After 12 cycles of selection, the pool showed a marked increase in activity, to about 1500-fold above the background rate of the reaction. Because activity plateaued in selection round 14, mutagenic PCR was used in round 15 to explore sequence variants of the initially selected sequences. The corresponding reduction in pool activity in round 15 was followed by continued, high activity in selection rounds 16 to 18.

To identify active sequences, the pools of all selection rounds were subjected to high-throughput sequencing (HTS) analysis (Fig. 3B). About 400,000 sequences were analyzed for each selection round, and sequences that showed more than 75% sequence identity were clustered. The resulting data show that the pool was dominated by five clusters during the last rounds of selection. The most active sequence variants within each of the active clusters were identified by the highest enrichment of their mutations over the last five rounds of the selection (fig. S2) (*35*). On the basis of these data, four of the most promising sequences were chosen from clusters 1 to 3, and the two most promising sequences from clusters 4 and 5. One

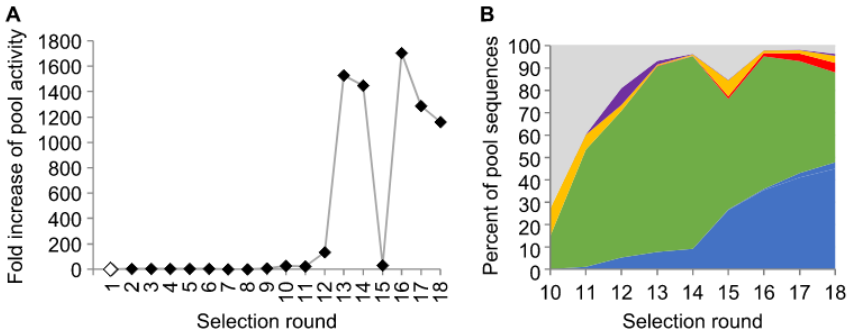


Fig. 3. Progress of the in emulsio selection. (**A**) Progress of the selection, as judged by the pool activity as a function of the selection round. Pool activity was measured as the fraction of the radiolabeled pool that was retained at the APM-PAGE interface. To account for the decreased concentrations of cTmp and 6sGsn over the selection rounds, the retained pool fraction was multiplied by both decreases in concentration. (**B**) Abundance of sequence clusters as judged by HTS analysis of the selected pools. Sequence clusters are indicated by color, with clusters 1 (blue), 2 (green), 3 (red), 4 (yellow), and 5 (purple) and unassigned sequences (gray). Selection round 15 used mutagenic PCR, correlated with a decrease in activity that recovered in round 16.

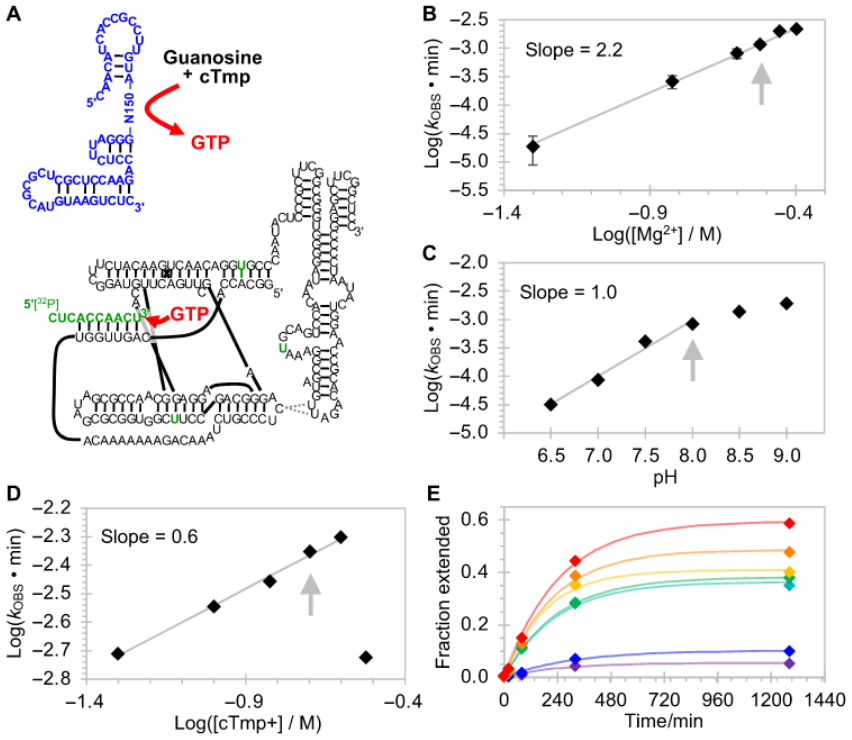


Fig. 4. Optimization of reaction conditions for GTR1. (**A**) Schematic for the trans-activity assay. The isolated pool sequence (blue) was tested for catalyzing the formation of GTP (red), physically separated from the polymerase ribozyme (black) with mutations for enhanced GTP processing (green). A 5′-radiolabeled RNA 10-mer (green) was extended by the polymerase ribozyme from 10 to 11 nucleotides, and the 10-mer and 11-mer were separated by PAGE and quantified by phosphorimaging. The fraction of elongated RNA was determined for reaction time courses, and the reaction rates were determined by single-exponential curve fitting. These rates are plotted as a function of the reaction conditions. Starting from an initial condition of 125 mM Mg^{2+}, 100 mM KCl, 50 mM tris-HCl (pH 8.3), 22°C, 25 mM cTmp, 1 mM Gsn, and 1 μM ribozyme, the conditions were successively optimized to (**B**) 300 mM Mg^{2+} (error bars are SDs from three experiments), 1 M K^{+}, (**C**) pH 8.0, 22°C, (**D**) 200 mM cTmp, (**E**) 2 mM guanosine, and 9 μM GTR1. Graphs not shown in this figure appear in fig. S6. Gray arrows indicate the chosen conditions. The gray lines show linear least-squares fits to data points, with slopes given in the inserts. In (E), the guanosine concentrations are 0.1 mM (purple), 0.2 mM (blue), 0.5 mM (teal), 0.75 mM (green), 1 mM (yellow), 1.5 mM (orange), and 2 mM (red), and the colored lines are least-squares fits of single-exponential functions to the data of the corresponding color.

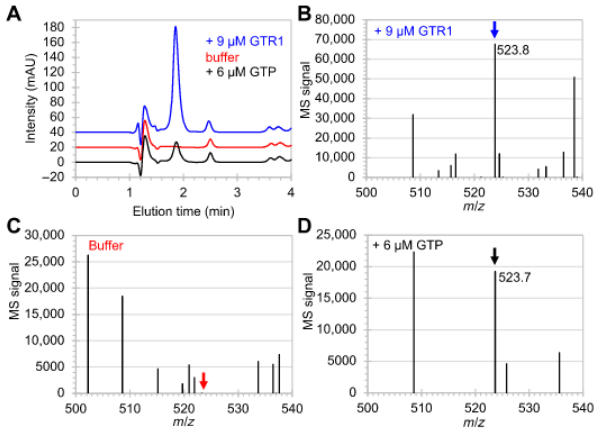


Fig. 5. LC-MS analysis of reaction products between guanosine and cTmp. (**A**) Trace at 260 nm during reversed-phase HPLC of reaction products in reaction buffer containing guanosine and cTmp. In the presence of 9 μM GTR1 (blue trace), a strong peak at the elution time of 1.9 min resulted, which was absent without added ribozyme (red trace) and appeared when 6 μM GTP was added to the reaction buffer (black trace). mAU, milli-absorption units at 260 nm. Positive ion mode ESI-MS at the elution time of 1.9 min resulted in (**B**) a signal at *m/z* of 523.8 when GTR1 was present (blue arrow), (**C**) no such signal when the ribozyme was absent (red arrow), and (**D**) a signal at 523.7 when 6 μM GTP was added to the reaction buffer (black arrow).

additional sequence was generated each for clusters 2 and 4, which combined multiple enriched mutations.

Biochemical analysis of selected sequences

To analyze the chosen sequences for their generation of freely diffusing NTPs, an assay was developed in which the produced NTP was used to extend the 3′-terminus of a 5′-radiolabeled RNA 10-mer (Fig. 4A). At the same time, the pool molecules were truncated at their 3′-termini so that they would not base pair to the polymerase ribozyme. In this format, the NTP molecules generated by triphosphorylation ribozymes were required to diffuse from one ribozyme to the other. The reaction products were separated by PAGE, and the fraction of extended radiolabeled RNAs was measured.

When the ribozymes were challenged to triphosphorylate 6sGsn, the strongest signal for primer extension was seen for ribozymes from cluster 5 (fig. S3, A and B). In contrast, the formation of GTP from guanosine and trimetaphosphate showed the most efficient primer extension with ribozymes from cluster 1 (fig. S3, C and D). Sequence 59 of cluster 1 showed the strongest average signal. When this sequence was truncated at the 5′-terminus or the 3′-terminus, the signal dropped markedly, whereas an internal region of the ribozyme (nucleotides 47 to 76) could be removed without a loss in signal (fig. S4). The resulting ribozyme was termed GTR1 (guanosine triphosphorylation ribozyme 1), had a length of 175 nucleotides, and was chosen for further analysis.

Characterization of the winning ribozyme GTR1

For a more detailed biochemical analysis of GTR1, the assay was modified to quench the GTR1 ribozyme after incubation with guanosine and cTmp before the detection of GTP (fig. S5). This allowed recording reaction kinetics for the GTP synthesis reaction and optimizing the reaction conditions (Fig. 4 and fig. S6). The dependence of the observed rate on the Mg^{2+} concentration showed a log-linear relation with a slope of 2.2, suggesting that 2 or 3 magnesium ions are limiting for the reaction kinetics (*36*). The dependence on pH showed a slope of 1.0, consistent with a single deprotonation step as the rate-limiting step in the reaction. This is likely the deprotonation of the nucleoside 5′-hydroxyl group, analogous to our earlier studies on self-triphosphorylation ribozymes (*22*). The dependence on cTmp concentration at 300 mM Mg^{2+} was consistent with Mg^{2+} forming a 1:1 complex with cTmp and leaving additional free Mg^{2+}, as observed for self-triphosphorylating ribozymes (*22*). An increase in the guanosine concentration did not lead to a substantial increase in the triphosphorylation rate but to an increase in the amplitude. The chosen conditions were 300 mM $MgCl_2$, 1 M KCl, 50 mM tris-HCl (pH 8.0), 22°C, 200 mM cTmp, 2 mM guanosine, and 9 μM GTR1.

To test whether GTP was produced by GTR1, the reaction products were analyzed by liquid chromatography (LC), followed by mass spectrometry (MS). During reversed-phased high-performance liquid chromatography (HPLC; Fig. 5A), a peak at elution time at 1.9 min was visible when reaction buffer containing guanosine and cTmp was incubated overnight with 9 μM GTR1 but not when the ribozyme was absent. The peak appeared again when 6 μM GTP was added to the reaction buffer. When this peak was analyzed by electrospray ionization (ESI)–MS in positive ion mode, the sample with 9 μM GTR1 (Fig. 5B) showed a strong peak with mass/charge ratio (*m/z*) of 523.8, as expected for GTP (calculated 524.2). This signal was absent for the sample without GTR1 (Fig. 5C) and appeared again in the sample containing 6 μM GTP (Fig. 5D). The fragmentation patterns of the peak around *m/z* of 524 gave *m/z* values of 152.2 and 363.8 for the sample containing GTP, and *m/z* values of 152.2 and 363.9 for the sample incubated with GTR1 (fig. S7). The sample containing only reaction buffer did not show these signals. These results confirmed that GTR1 catalyzed the formation of GTP.

To determine the catalytic rate enhancement by the GTR1 ribozyme, the formation of GTP from guanosine and cTmp was measured under optimal conditions as determined above (Fig. 6A). The rate for the catalyzed reaction with 9 μM GTR1 was 1.9 $hour^{-1}$, compared to a rate of 1.1×10^{-4} $hour^{-1}$ for the uncatalyzed reaction under the same conditions but without GTR1 (Fig. 6B). This corresponded to an 18,000-fold rate enhancement. The turnover number of the ribozyme was determined by reacting varying concentrations of GTR1 with a fixed concentration of substrate that was elongated with the produced GTP (Fig. 6C). The end points of the reaction kinetics showed that each GTR1 ribozyme molecule produced, on average, only 1.7 GTP molecules under these conditions. This number is a lower estimate because not every generated GTP molecule is detected by extension of the primer.

We tested whether GTP formed by the GTR1 ribozyme could be incorporated into an RNA polymer using an RNA polymerase ribozyme (Fig. 7A). The template dictated the extension by cytidine 5′-triphosphate (CTP), GTP, and ATP, but only CTP and ATP were added as NTPs. In contrast, GTP (position +2) was provided from a GTR1-catalyzed guanosine triphosphorylation reaction. Quantification of the reaction product pattern (Fig. 7B) and comparison with the same reaction lacking GTR1 confirmed that most of the nucleotides incorporated at position +2 originated from catalysis by GTR1. The background rate without GTR1 likely resulted from

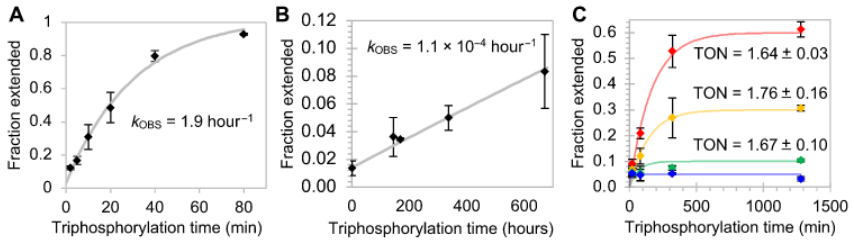


Fig. 6. Rate enhancement and turnover of ribozyme-catalyzed GTP synthesis. The rates of the reaction catalyzed by 9 μM GTR1 (**A**) and the uncatalyzed reaction (**B**) were obtained by single-exponential fits (gray) to the kinetic data (diamonds). Note the difference in the scales of the axes. Error bars are SDs from triplicate experiments. (**C**) Determination of turnover number with a fixed concentration of radiolabeled primer with increasing concentrations of GTR1. The GTR1 concentration was 6 μM (red), 3 μM (yellow), 1 μM (green), or 0 μM (blue). The ratio of extended primer over GTR1 after ~1280 hours equaled the turnover number (TON), which is given as inserts.

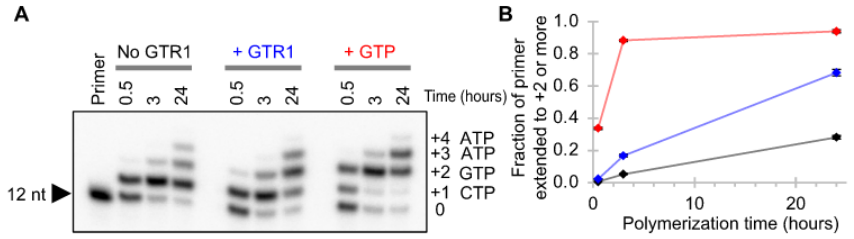


Fig. 7. Ribozyme-mediated RNA polymerization coupled to ribozyme-catalyzed GTP synthesis. (**A**) Autoradiogram of polymerization products separated by denaturing 20% PAGE. The first reaction lacked GTP and GTR1 (No GTR1, black), the second reaction lacked GTP but contained the products from a GTP synthesis reaction with GTR1 (+ GTR1, blue), and the third reaction contained GTP (+ GTP, blue). For each of the three different reactions, samples from three polymerization time points were analyzed (0.5, 3, and 24 hours). The identity of the templated NTP is shown on the right for nucleotide (nt) additions +1 to +4. (**B**) Quantification of extension products. The fraction of primers extended by at least two nucleotides is plotted as a function of reaction time for three different conditions described in (A). Error bars are SDs from three experiments and mostly smaller than the symbols. The values at 24-hour polymerization time are 28.1 ± 1.1% (no GTR), 68.3 ± 2.0% (+ GTR), and 94.0 ± 0.9% (+ GTP).

misincorporation by the polymerase ribozyme (*7*). These results confirmed that GTR1-generated GTP can be incorporated during ribozyme-mediated RNA polymerization.

The secondary structure of GTR1 was analyzed using SHAPE (selective 2′ hydroxyl acylation analyzed by primer extension) probing with 1-methyl-7-nitroisatoic anhydride (1M7) (*37*) at the optimal ribozyme reaction conditions, with tris-HCl buffer replaced by Hepes-NaOH (Fig. 8). The SHAPE reaction products were analyzed by reverse transcription with a 5′-[^{32}P]–radiolabeled primers and separation by denaturing PAGE. To allow probing the complete sequence of the ribozyme, the ribozyme was extended by 20 nucleotides that did not interfere with the ribozyme's function (gray 3′-tail in Fig. 8B). The protection pattern suggested a secondary structure with a central, tri-helical junction. Tertiary contacts appear to form between the protected nucleotides 138 to 141 and either nucleotide positions 47 to 49 or 96 to 99. Future structural analysis and sequence optimization will reveal how this ribozyme is able to catalyze the reaction between guanosine and cTmp to GTP.

DISCUSSION

This study demonstrated a de novo in emulsio selection of ribozymes from random sequence. In vitro selection in emulsion has been used previously for selecting trans-active RNAs. However, these previous selections used the emulsion system to optimize a previously selected ribozyme. One study (*39*) selected variants of the previously developed class I ligase ribozyme (*40*), one study (*41*) selected variants of a previously selected Diels-Alderase ribozyme (*42*), and two studies optimized the polymerase ribozyme (*7*) for higher polymerization efficiency (*43*, *44*). These studies relied on in vitro selection experiments in bulk to generate ribozymes and later improved their performance by subsequent evolution rounds in emulsion. In contrast, in this present study, the completely random sequence RNAs were incubated in emulsion from the first selection round on, to identify RNAs generating a freely diffusing reaction product. The study thereby directly selected for a small metabolic system, in which two ribozymes—the GTP synthase and the polymerase ribozyme—were metabolically coupled through their common metabolite GTP.

The selected ribozyme GTR1 catalyzes the triphosphorylation of free guanosine and generates free GTP. This is a challenging task for a ribozyme because the ribozyme needs to bind two small-molecule substrates, catalyze their reaction, and release the product. The catalytic rate enhancement of 18,000-fold suggests that the first three of these requirements are met well. The low turnover number of 1.7-fold may originate from a slow product release. A slow product release could be caused by the requirement during the selection to bind guanosine and cTmp such that the produced GTP may have become bound tightly at the catalytic site.

The choice of guanosine—or specifically, 6sGsn—as substrate for the in emulsio selection was based on (i) the successful previous use of 6sGsn in an in vitro selection based on the gel shift in an APM polyacrylamide gel (*31*), (ii) the commercial availability of the thio-modified nucleoside, and (iii) the fact that guanosine provides better stacking than the otherwise also useful 4-thio uridine (*45*). Nature's choice for guanosine as external substrate in group I

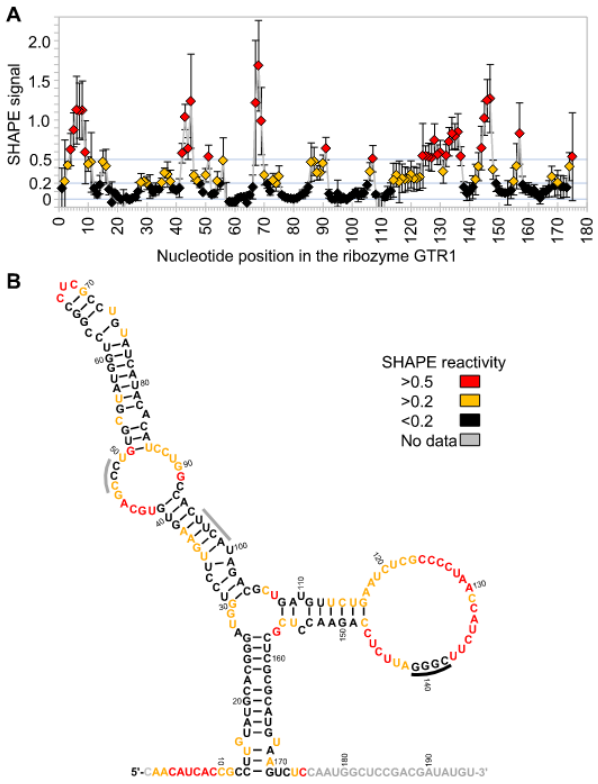


Fig. 8. Secondary structure analysis of the ribozyme GTR1. (**A**) SHAPE reactivity profile of the GTR1 ribozyme under optimal reaction conditions. The SHAPE reactivity was determined by reverse transcription of two radiolabeled primers with the SHAPE-reacted RNA. Cutoffs of 0.2 and 0.5 (light blue lines) were chosen to distinguish positions with low reactivity (black), medium reactivity (orange), and high reactivity (red). Error bars are SDs from three independent probing experiments. (**B**) Secondary structure resulting from calculation with RNAfold in the ViennaRNA package 2.0 (*38*), using the constraints of the SHAPE reactivity. Color coding is as in (A). The black bar at positions 138 to 141 denote a SHAPE-protected region that may pair with positions 47 to 49 or 86 to 89 (gray bars). Positions labeled in gray did not have SHAPE signals because a signal corresponding to position 1 was too close to full-length extension and because positions 176 to 195 were covered by the reverse transcription primer.

intron ribozymes showed that guanosine is well suited to be bound tightly and specifically and then serve as reactant (*46*). However, guanosine is harder to activate at its 5′-hydroxyl group (*21*) and its 2′- and 3′-hydroxyl groups than the other canonical nucleosides (*47*). The slower uncatalyzed rate for guanosine to react with cTmp appears to stem, at least in part, from the tendency of the guanine base to occupy the anti-conformation in the nucleoside (*21*). On the basis of these low rates for the uncatalyzed reactions, GTP may have been in shorter supply than the other NTPs for an early life form. Therefore, a ribozyme generating GTP from guanosine and cTmp would likely have had a large evolutionary benefit. While our results showed that ribozyme-generated GTP can be incorporated into ribozyme-generated RNA polymers, the low turnover of GTR1 constrains the polymers that can be generated with the current ribozyme. We anticipate that future, improved versions of GTR1 will be able to supply GTP with an increased turnover and that additional nucleoside triphosphorylation ribozymes will also generate ATP, CTP, and uridine 5′-triphosphate (UTP) to allow ribozyme-mediated synthesis of RNA polymers from nucleosides and cTmp.

MATERIALS AND METHODS

Generation of the pool

A 188-mer DNA pool consisting of a 150-nucleotide randomized region flanked by PCR primer constant regions was ordered commercially [Integrated DNA Technologies (IDT)], with the sequence 5′-CAACATCACCGCCTTGTA-N150-GGGATTCTC-CAGAACCTCGC-3′. Because only 2.2% of the pool was amplifiable by quantitative PCR, we assumed an upper limit on the pool complexity at 1.6×10^{14}. PCR amplification (45 ml) used the constant regions as primer binding sites. Products were purified on commercial PCR clean-up columns (Macherey Nagel) and used as 30 nM template in a second PCR. During this PCR, primers added the T7 transcription promoter sequence and a hammerhead ribozyme sequence to the 5′ end of the DNA pool construct, and a DNAzyme recognition site to the 3′ end of the DNA pool construct. The hammerhead ribozyme would cleave cotranscriptionally to generate a free 5′-hydroxyl; this was done to prevent the possibility that pool molecules would use their own 5′-triphosphate to activate guanosine. The DNAzyme recognition site allows recognition and cleavage by a DNAzyme, which generates a free 3′-hydroxyl group. This 3′-hydroxyl can then be recognized by the polymerase ribozyme. This resulted in a final sequence of the DNA library of 5′-AATTTAATACGACT-CACTATAGGGATGTTGCTGACGAGCTAAGCGAAACTGCG-GAAACGCAGTCCAACATCACCGCCTTGTA-N150-GGGAT-TCTCCAGAACCTCGCTCGCGCATGTAAGTCTCACCAACT-TATATGTTCTAGCGCGGA-3′, with underlined sequences acting as recognition sites for the DNAzyme (below). The PCR products were transcribed with T7 RNA polymerase and purified by denaturing 5% PAGE. A small amount of RNA pool was 5′ end radiolabeled using polynucleotide kinase (PNK) (New England Biolabs) and γ-[^{32}P]–ATP and purified by denaturing 5% PAGE. The radiolabeled and unlabeled RNA pool were combined and processed together during the subsequent DNAzyme reaction.

DNAzyme reaction

The RNA pool molecules were processed at their 3′-termini with a DNAzyme to generate a defined 3′-end with 2′,3′-hydroxyl groups. While 2′,3′-cyclic phosphates could have been generated much easier by self-cleaving ribozymes at the 3′-terminus, this was not done because of a possible selection artifact: Because 2′,3′-cyclic phosphates are chemically activated phosphates that can be used to drive the formation of a phosphodiester bond (see the reversibility of hairpin ribozyme reactions), they could also have led to the selection of a ligase ribozyme that catalyzes the nucleophilic attack of the 5′-hydroxyl group of 6sGsn to the 2′,3′-cyclic phosphate and thereby tag the RNA pool molecule with 6-thio guanosine 5′-monophosphate (6sGMP). While it might have been possible to remove 2′,3′-cyclic phosphates with alkaline phosphatase, we were concerned that some pool molecules might have been able to hide their 3′-termini from alkaline phosphatase. Therefore, the most reliable solution appeared to be processing by the DNAzyme. The DNAzyme reaction was a modified protocol for DNAzyme 9SK17, which was previously

selected to cleave RNAs by generating a 5′-phosphate and 2′,3′-hydroxyl groups (*26*). The DNAzyme sequence was 5′-GCGCTAGAA-CATGCCAGCGATCAAAGACGGCGAGTTGTACCCATAGGT-GTCTAGTTGGTGAGACTT-3′, where the underlined sequences were complementary to the 3′-constant region of the RNA library. A competitor DNA (5′-GCGAGGTTCTGGAGAATCCC-3′), complementary to the pool molecule upstream of the DNAzyme binding site, was used to mitigate inhibition of cleavage. The nucleic acids were prepared by heat renaturing 4 μM pool RNA, 8 μM DNAzyme, and 6 μM competitor DNA in 5 mM Hepes-NaOH (pH 7.5), 15 mM NaCl, and 0.1 mM EDTA and immediately placing them on ice for 10 min. An equal volume of reaction buffer was added to give final concentrations of 70 mM Hepes-NaOH (pH 7.5), 150 mM NaCl, 1.25 mM $ZnCl_2$, 15 mM $MnCl_2$, and 10 mM $MgCl_2$. The $ZnCl_2$ for the reaction buffer was prepared as an 8× stock from $ZnCl_2$ powder, by adding in order the compounds to give final concentrations of 10 mM $ZnCl_2$, 20 mM HNO_3, and 200 mM Hepes-NaOH (pH 7.5). The reaction was incubated at 37°C for 3 hours. The reaction was stopped by adding a stoichiometric excess of EDTA and a DNA oligonucleotide that is reverse complementary to the competitor DNA (5′-GGGATTCTCCAGAACCTCGC-3′) in formamide loading buffer and purified using denaturing 5% PAGE.

Selection step in emulsion

The oil phase consisted of 4% ABIL EM 90 (v/v) with 96% (v/v) heavy mineral oil. The oil phase was stirred at room temperature and then degassed in oil vacuum until no bubbles were present and stirred until the aqueous phase was prepared. To prepare the aqueous phase (2.5 ml in a 50-ml emulsion), RNAs were incubated with 100 mM tris-HCl (pH 8.3) and heat-denatured using a thermocycler at 80°C for 2 min before being immediately chilled on ice for 5 min. The RNAs were then mixed with an equal volume of buffer containing $MgCl_2$, KCl, cTmp, and 6sGsn, resulting in final concentrations of 3 μM polymerase ribozyme, 0.5 μM RNA pool, 50 mM tris-HCl (pH 8.3), 150 mM $MgCl_2$, 200 mM KCl, 50 mM cTmp, and 1 mM 6sGsn (later selection rounds contained less cTmp and 6sGsn). The reaction was started by quickly mixing the aqueous phase while on ice and then adding the entire mixture to the stirring oil phase. This generated a raw emulsion, which was mixed for 30 s before loading into the microfluidizer.

The emulsion was passed through a microfluidizer (Microfluidics, M110L) seven times at a pressure of 6000 psi (flow cell H10Z). The emulsion was then incubated at room temperature for 6 hours. The droplet size of the emulsion was measured by dynamic light scattering (DLS). In summary, a neodymium laser beam (532 nm) was passed through a 300-fold dilution of the emulsion in water-saturated oil phase. Light scattered at 90° was detected by a photomultiplier tube specially designed to tolerate large photon count rates. The measurements used a 0.5-ms dwell time in a multichannel scalar (FAST ComTec GmbH). Autocorrelation functions were calculated offline and fit to obtain droplet diameters as illustrated in Fig. 2B for one typical case.

The incubated emulsion was processed by first adding a molar excess of EDTA to chelate Mg^{2+} ions, followed by addition of 5 ml of water to increase the volume of the aqueous layer. The addition of 5 ml of diethyl ether then broke the emulsion. To isolate the RNA, the emulsion was centrifuged at 15,000*g* for 30 min at 4°C. The aqueous phase was extracted with an excess of diethyl ether and centrifuged again. The pellet was vortexed with 5 ml of phenol until it dissolved, and centrifuged at 15,000*g* for 10 min at 4°C. The aqueous phase was extracted with an excess of chloroform and centrifuged at 15,000*g* for 5 min at 4°C. RNA was recovered from the aqueous layer by ethanol precipitation.

To purify active sequences by 5% APM-PAGE, the polyacrylamide gel was made in three layers. The bottom and top layers both consisted of 5% polyacrylamide. The middle layer contained 0.1 mM APM (from a stock solution of 3 mM APM in *N*,*N*′-dimethylformamide). The sample was loaded in the middle of the gel alongside a positive control that consisted of RNA pool spiked with 6sGTP. After electrophoresis for 4 to 6 hours, RNA sequences at the APM interface were visualized using the 5′ radiolabeled pool molecules introduced during the DNAzyme processing step. The interface band was excised and eluted in 300 mM NaCl and 5 mM DTT overnight at 4°C. Following ethanol precipitation, the pool molecules were reverse-transcribed using SuperScript III and the RT primer 5′-GCGAGGTTCTGGAGAAT-3′. Reverse transcriptase (RT) products were PCR-amplified using the same primers given above to generate the selection library, completing one round of selection.

HTS analysis

PCR products were barcoded for each selection round and sequenced on an Illumina MiSeq machine using paired-end 250 sequencing (UCSD IGM Genomics Center). The output of MiSeq data was given as a demultiplexed FASTQ file format. The data were uploaded to the online Galaxy bioinformatics website, where the constant regions were clipped and the files were converted to FASTA format. Each round of selection contained its own FASTA file sorted by sequence abundance. Sequence clusters were generated in the USEARCH suite (*48*), with an identity threshold of 0.75 in the fast function. Clustering was done using a greedy algorithm, and so, the cluster centers became defined as the most abundant sequence due to the sequence order within the FASTA file. A combination of Python, terminal commands, and Excel was used to track the abundance of each cluster through the rounds of selection. An R script was used on each cluster to track the progression of individual sequences within each cluster through each round. Abundant sequences were aligned using the MUSCLE suite (*49*). This allowed the identification for highly enriching mutations within a cluster.

Guanosine triphosphorylation assay

The guanosine triphosphorylation assay follows the same principle as the reaction in emulsion during the selection procedure, but using the final concentrations 1 μM GTR, 0.5 μM radiolabeled primer (5′-CUCACCAACU-3′), 1 μM polymerase ribozyme, 0.5 mM 6sGsn or 1.0 mM Gsn, 25 mM cTmp, 50 mM tris-HCl (pH 8.3), 150 mM $MgCl_2$, and 200 mM KCl. To initiate the assay, the GTR was heat-renatured in 50 mM tris-HCl (pH 8.3) by incubating at 80°C for 2 min and then allowed to cool at room temperature for 5 min. A premix containing the polymerase ribozyme and radiolabeled primer was heat-renatured in 50 mM tris-HCl (pH 8.3) according to the same heating profile. The guanosine triphosphorylation ribozyme premix and the polymerase ribozyme premix were added together in equal volume. A reaction premix containing tris-HCl, $MgCl_2$, KCl, cTmp, and 6sGsn was then added to the ribozyme mix in equal volume. The reaction was incubated at room temperature for 6 hours. Products were mixed with a loading buffer containing 40% (v/v) formamide and 1.5 μM final concentration of a DNA oligomer complementary to the RNA primer. After heat renaturing

(2 min at 80°C), the samples were separated by 7 M urea 20% PAGE. The separations were exposed to phosphorimager screens and scanned with a phosphorimager (Typhoon NIR Plus, Amersham), and bands were quantified using Quantity One software. The data were processed in Excel worksheets.

For the optimization of reaction conditions, the guanosine triphosphorylation assay was modified to quench GTR1 catalysis at specific time points and thereby allow measuring reaction kinetics. All reactions were performed at room temperature. A premix was made as follows: 1 μM ribozyme in 50 mM tris-HCl (pH 8.3) was incubated with $MgCl_2$, KCl, cTmp, and Gsn at the indicated concentration. For the titrations of $MgCl_2$, KCl, cTmp, Gsn, and ribozyme, only one component was variable, while the rest were held constant. For the pH optimization, different buffers were used: tris-HCl (pH 9.0, 8.5, and 8.0), Hepes-NaOH (pH 7.0), and MES/NaOH (pH 6.5 and 6.0). At a given time point, the reaction was quenched by adding a mix of three DNA inhibitors (fig. S5), each at a 19.4-fold molar excess over GTR1 (GTR1 at 1 μM in the reaction). The excess was lower at higher GTR1 concentrations (twofold at 9 μM GTR1).

To determine reaction kinetics under optimized conditions, GTP generated by the GTR was used by a polymerase ribozyme to extend a substrate RNA by one nucleotide. A mixture containing RNA primer (5′-CUCACCAACU-3′, 0.5 μM final), trace-radiolabeled RNA primer, and polymerase ribozyme (1 μM final) was added to the GTR mix. The mixture also included KCl such that the overall KCl concentration was maintained at 200 mM, and it included tris-HCl (pH 8.3) such that the concentration of tris-HCl was maintained at 50 mM. In addition, the magnesium concentration was 150 mM in the polymerase ribozyme reaction regardless of magnesium concentration in the GTR mix. If the concentration of magnesium after mixing the GTR mix and the polymerase ribozyme mix was greater than 150 mM, excess magnesium was chelated using EDTA to bring the concentration of free Mg^{2+} down to 150 mM. If the concentration of magnesium after mixing the GTR mix and the polymerase ribozyme mix was less than 150 mM, additional magnesium was added to bring the concentration up to 150 mM. Reaction products were mixed with a loading buffer containing 40% (v/v) formamide and 1.5 μM final concentration of a DNA oligomer complementary to the RNA primer. After heat renaturing (2 min at 80°C), the samples were separated by 7 M urea 20% PAGE.

The separations were exposed to phosphorimager screens and scanned with a phosphorimager (Typhoon NIR Plus, Amersham), and bands were quantified using Quantity One software. The data were processed in Excel worksheets. Single exponential fits were adjusted to all kinetic data (each with five time points 5, 20, 80, 320, and 1280 min) using the Excel subroutine Solver to the equation "fraction ligated" = $A^*(1 - \mathrm{EXP}(-k_{OBS}^*t)) + C$, where fraction ligated is the experimentally determined fraction of 10-mer RNA primer extended by the polymerase ribozyme, t is triphosphorylation time, A is the amplitude of the curve, k_{OBS} is the observed rate, and C describes the constant background.

Truncation of the winning isolate

The sequence C0-59 was chosen as the best initial isolate from the in vitro selection. A series of 5′, 3′, and internal truncations were made and tested for activity. 5′ and 3′ truncations were generated by PCR using Taq DNA polymerase and a series of different 5′ and 3′ primers for each truncation. DNA templates of each truncation were transcribed using T7 RNA polymerase and gel-purified. The truncated ribozymes were then tested for activity using the previously described guanosine triphosphorylation assay. None of the 5′ or 3′ truncations were active.

A series of truncations, as well as two combinations of internal truncations, was made (fig. S4). DNA templates containing the internal truncations were obtained commercially (IDT) as gBlocks. Plasmids for each truncation were generated by cloning into pUC19 and confirmation by sequencing. DNA templates of each truncation were generated by PCR from each plasmid, transcribed using T7 RNA polymerase, and PAGE-purified. The truncated ribozymes were then tested for activity using the guanosine triphosphorylation assay described above.

Secondary structure probing using SHAPE

SHAPE was performed on GTR1 using previously published methods (*37*). A 20-nucleotide extension that did not interfere with ribozyme function was added to the 3′ end of the ribozyme by PCR amplification. The sequence of this primer is (5′-ACATATCGTCGGAGCCATTGGAGACTTACATGCGCGAG-3′), where the underlined portion is complementary to the ribozyme. 1M7 was used as the chemical probe. Ribozyme (90 pmol) was heat-renatured at 80°C for 2 min, cooled to 50°C for 5 min, and then left at room temperature for 5 min. To this solution, a final concentration of 50 mM Hepes-NaOH (pH 8.0), 2 mM Gsn, 1 M KCl, and 300 mM $MgCl_2$ was added. This mixture was incubated at room temperature for 2 min. A 100 mM stock solution of 1M7 was prepared in dimethyl sulfoxide (DMSO) and added as $^1/_{50}$ of the ribozyme solution, resulting in final concentrations of 2 mM 1M7 and 2% (v/v) DMSO. A control without 1M7 was prepared with ribozyme, Hepes, Gsn, KCl, $MgCl_2$, and 2% (v/v) DMSO. Both samples were incubated at room temperature for 3 min. The reaction was quenched by ethanol precipitation and resuspended in 10 μl of 5 mM tris-HCl (pH 8.0). The products were reverse-transcribed using SuperScript III reverse transcriptase (Invitrogen) according to the manufacturer's instructions and trace amounts of a radiolabeled primer that annealed to the aforementioned 3′-extension of the ribozyme. The sequence of this primer is (5′-ACATATCGTCGGAGCCATTG-3′). It was difficult to resolve products larger than 60 nucleotides in length using this primer, so the protocol above was repeated using one additional radiolabeled primer during the reverse transcription step. The other primer (5′-GTATGATACAGGCGAGGC-3′) annealed to positions 65 to 82 of the ribozyme. Following the reverse transcription, the RNA template in each sample was degraded by alkaline hydrolysis. To do that, the samples were incubated for 5 min at 80°C in a solution containing 750 mM NaOH. The reaction was quenched by addition of acetic acid to a final concentration of 300 mM NaOAc/HOAc (pH 5). Then, the products were ethanol-precipitated, resuspended in formamide loading buffer, and resolved by 7 M urea 20% PAGE.

MS identification of GTP

A reaction mixture consisting of 50 mM tris-HCl (pH 8.0), 200 mM KCl, 125 mM $MgCl_2$, 1 mM Gsn, 25 mM cTmp, and 9 μM GTR1 was incubated at room temperature for 24 hours. GTR1 was heat-renatured in water for 10 min at 50°C and then cooled for 5 min at 22°C before the other compounds were added. A mixture containing water in place of GTR1 was used as a negative control, and a sample containing 6 μM GTP in place of GTR1 was used as a positive control. Twenty-five microliters of each sample was analyzed by LC-MS. Twenty microliters of 6 μM GTP in water was used as

standard. After incubation, each sample was centrifuged at 10,000*g* for 5 min to remove possible particles, and the clear supernatant was submitted for LC-MS [University of California San Diego (UCSD) facility].

Samples were loaded on a 150-mm-long C18 reversed-phase column at 37°C (Scherzo SM-C18, metal-free column, Imtakt USA) in 50 mM ammonium formate (pH 8.6) and eluted in a gradient over 10 min (3 ml) with 100 mM ammonium formate (pH 8.6). Elution profiles were recorded at 260 nm with 4-nm bandwidth. The HPLC system (Agilent 1260 Infinity) was coupled to a Thermo LCQdeca MS, which used positive ion mode electrospray ionization as the ion source. The source voltage was 5 kV, the sheath gas rate was 80 units, the auxiliary gas flow was 20 units, and the capillary gas temperature was 250°C.

Ribozyme-mediated RNA polymerization

Ribozyme-mediated RNA polymerization was performed essentially as described (*7*), but the GTP was supplied from a preceding reaction between guanosine and cTmp. Specifically, 9 μM of ribozyme GTR1 was incubated with 50 mM tris-HCl (pH 8.0), 200 mM KCl, 125 mM $MgCl_2$, 25 mM cTmp, and 1 mM Gsn at 22°C for 24 hours. These conditions are not the optimal conditions for GTR1 but were adjusted to work together with the polymerase ribozyme. The polymerase ribozyme 8.12.23 was heat-renatured 2 min at 80°C in water together with RNA template 5′-GACGCUUCGCACGGUUGG-CAG-3′, 5′-[^{32}P]–radiolabeled RNA primer 5′-CUGCCAACCGUG-3′, and P2 oligonucleotide 5′-GGCACC-3′. The reaction was started by combining the GTR1 reaction mixture, RNAs, and a premixed buffer to result in the final concentrations of 2.5 μM P2 oligonucleotide, 2 μM polymerase ribozyme, 1 μM template, 0.5 μM primer, 10 μM CTP, 30 μM UTP, 30 μM ATP, 50 mM tris-HCl (pH 8.3), 200 mM $MgCl_2$, and 50 mM KCl, with one-quarter of the reaction volume being the GTR1 reaction mixture. The reaction was incubated at 20°C for the given times and quenched by adding a final concentration of 40% (v/v) formamide, a molar excess of Na_2EDTA over the Mg^{2+} in the reaction, a 20-fold excess of an RNA complementary to the template, and heat denaturation 2 min at 80°C. The separations were exposed to phosphorimager screens and scanned with a phosphorimager (Typhoon NIR Plus, Amersham), and bands were quantified using Quantity One software. The data were processed in Excel worksheets.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abj7487

REFERENCES AND NOTES

1. A. Rich, On the problems of evolution and biochemical information transfer, in *Horizons in Biochemistry* (Academic Press, 1962), pp. 103–126.
2. P. Nissen, J. Hansen, N. Ban, P. B. Moore, T. A. Steitz, The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).
3. C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857 (1983).
4. A. D. Ellington, J. W. Szostak, In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).
5. C. Tuerk, L. Gold, Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).
6. X. Chen, N. Li, A. D. Ellington, Ribozyme catalysis of metabolism in the RNA world. *Chem. Biodivers.* **4**, 633–655 (2007).
7. W. K. Johnston, P. J. Unrau, M. S. Lawrence, M. E. Glasner, D. P. Bartel, RNA-catalyzed RNA polymerization: Accurate and general RNA-templated primer extension. *Science* **292**, 1319–1325 (2001).
8. K. F. Tjhung, M. N. Shokhirev, D. P. Horning, G. F. Joyce, An RNA polymerase ribozyme that synthesizes its own ancestor. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 2906–2913 (2020).
9. R. Cojocaru, P. J. Unrau, Processive RNA polymerization and promoter recognition in an RNA World. *Science* **371**, 1225–1232 (2021).
10. S. A. Benner, A. D. Ellington, A. Tauer, Modern metabolism as a palimpsest of the RNA world. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7054–7058 (1989).
11. Y. Yamagata, H. Watanabe, M. Saitoh, T. Namba, Volcanic production of polyphosphates and its relevance to prebiotic evolution. *Nature* **352**, 516–519 (1991).
12. R. Osterberg, L. E. Orgel, Polyphosphate and trimetaphosphate formation under potentially prebiotic conditions. *J. Mol. Evol.* **1**, 241–248 (1972).
13. M. A. Pasek, T. P. Kee, D. E. Bryant, A. A. Pavlov, J. I. Lunine, Production of potentially prebiotic condensed phosphates by phosphorus redox chemistry. *Angew. Chem. Int. Ed. Engl.* **47**, 7918–7920 (2008).
14. B. Herschy, S. J. Chang, R. Blake, A. Lepland, H. Abbott-Lyon, J. Sampson, Z. Atlas, T. P. Kee, M. A. Pasek, Archean phosphorus liberation induced by iron redox geochemistry. *Nat. Commun.* **9**, 1346 (2018).
15. C. Gibard, S. Bhowmik, M. Karki, E. K. Kim, R. Krishnamurthy, Phosphorylation, oligomerization and self-assembly in water under potential prebiotic conditions. *Nat. Chem.* **10**, 212–217 (2018).
16. M. A. Pasek, J. P. Harnmeijer, R. Buick, M. Gull, Z. Atlas, Evidence for reactive reduced phosphorus species in the early Archean ocean. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 10089–10094 (2013).
17. W. Feldmann, Zur Chemie der kondensierten Phosphate und Arsenate, LIII. Das Trimetaphosphat als Triphosphorylierungsmittel für Alkohole und Kohlenhydrate in wäßriger Lösung. Seine Sonderstellung unter den kondensierten Phosphaten. *Chem. Ber.* **100**, 3850–3860 (1967).
18. E. Etaix, L. E. Orgel, Phosphorylation of nucleosides in aqueous solution using trimetaphosphate: Formation of nucleoside triphosphates. *J Carbohydr. Nucleosides Nucleotides* **5**, 91–110 (1978).
19. Y. Yamagata, H. Inoue, K. Inomata, Specific effect of magnesium ion on 2′,3′-cyclic AMP synthesis from adenosine and trimeta phosphate in aqueous solution. *Orig. Life Evol. Biosph.* **25**, 47–52 (1995).
20. C. Cheng, C. Fan, R. Wan, C. Tong, Z. Miao, J. Chen, Y. Zhao, Phosphorylation of adenosine with trimetaphosphate under simulated prebiotic conditions. *Orig. Life Evol. Biosph.* **32**, 219–224 (2002).
21. F. Chizzolini, A. D. Kent, L. F. M. Passalacqua, A. Lupták, Enzymatic RNA production from NTPs synthesized from nucleosides and trimetaphosphate. *Chembiochem* **22**, 2098–2101 (2021).
22. J. E. Moretti, U. F. Muller, A ribozyme that triphosphorylates RNA 5′-hydroxyl groups. *Nucleic Acids Res.* **42**, 4767–4778 (2014).
23. J. F. Milligan, O. C. Uhlenbeck, [5] Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol.* **180**, 51–62 (1989).
24. F. J. Triana-Alonso, M. Dabrowski, J. Wadzack, K. H. Nierhaus, Self-coded 3′-extension of run-off transcripts produces aberrant products during in vitro transcription with T7 RNA polymerase. *J. Biol. Chem.* **270**, 6298–6307 (1995).
25. H. S. Zaher, P. J. Unrau, T7 RNA polymerase mediates fast promoter-independent extension of unstable nucleic acid complexes. *Biochemistry* **43**, 7873–7880 (2004).
26. D. J. Parker, Y. Xiao, J. M. Aguilar, S. K. Silverman, DNA catalysis of a normally disfavored RNA hydrolysis reaction. *J. Am. Chem. Soc.* **135**, 8472–8475 (2013).
27. A. Akoopie, U. F. Muller, The NTP binding site of the polymerase ribozyme. *Nucleic Acids Res.* **46**, 10589–10597 (2018).
28. F. J. Ghadessy, P. Holliger, A novel emulsion mixture for in vitro compartmentalization of transcription and translation in the rabbit reticulocyte system. *Protein Eng. Des. Sel.* **17**, 201–204 (2004).
29. G. L. Igloi, Interaction of tRNAs and of phosphorothioate-substituted nucleic acids with an organomercurial. Probing the chemical environment of thiolated residues by affinity electrophoresis. *Biochemistry* **27**, 3842–3849 (1988).
30. E. Biondi, D. H. Burke, Separating and analyzing sulfur-containing RNAs with organomercury gels. *Methods Mol. Biol.* **883**, 111–120 (2012).
31. M. W. Lau, K. E. Cadieux, P. J. Unrau, Isolation of fast purine nucleotide synthase ribozymes. *J. Am. Chem. Soc.* **126**, 15686–15693 (2004).
32. M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
33. E. H. Ekland, D. P. Bartel, The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res.* **23**, 3231–3238 (1995).
34. Q. S. Wang, P. J. Unrau, Ribozyme motif structure mapped using random recombination and selection. *RNA* **11**, 404–411 (2005).
35. A. Pressman, J. E. Moretti, G. W. Campbell, U. F. Muller, I. A. Chen, Analysis of in vitro evolution reveals the underlying distribution of catalytic activity among random sequences. *Nucleic Acids Res.* **45**, 8167–8179 (2017).
36. M. T. Record Jr., M. L. Lohman, P. De Haseth, Ion effects on ligand-nucleic acid interactions. *J. Mol. Biol.* **107**, 145–158 (1976).
37. S. A. Mortimer, K. M. Weeks, A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.* **129**, 4144–4145 (2007).

38. R. Lorenz, S. H. Bernhart, C. Höner zu Siederdissen, H. Tafer, C. Flamm, P. F. Stadler, I. L. Hofacker, ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
39. M. Levy, K. E. Griswold, A. D. Ellington, Direct selection of trans-acting ligase ribozymes by in vitro compartmentalization. *RNA* **11**, 1555–1562 (2005).
40. N. H. Bergman, W. K. Johnston, D. P. Bartel, Kinetic framework for ligation by an efficient RNA ligase ribozyme. *Biochemistry* **39**, 3115–3123 (2000).
41. J. J. Agresti, B. T. Kelly, A. Jaschke, A. D. Griffiths, Selection of ribozymes that catalyse multiple-turnover Diels–Alder cycloadditions by using in vitro compartmentalization. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16170–16175 (2005).
42. B. Seelig, A. Jaschke, A small catalytic RNA motif with Diels-Alderase activity. *Chem. Biol.* **6**, 167–176 (1999).
43. H. S. Zaher, P. J. Unrau, Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA* **13**, 1017–1026 (2007).
44. A. Wochner, J. Attwater, A. Coulson, P. Holliger, Ribozyme-catalyzed transcription of an active ribozyme. *Science* **332**, 209–212 (2011).
45. P. J. Unrau, D. P. Bartel, RNA-catalysed nucleotide synthesis. *Nature* **395**, 260–263 (1998).
46. B. L. Bass, T. R. Cech, Specific interaction between the self-splicing RNA of Tetrahymena and its guanosine substrate: Implications for biological catalysis by RNA. *Nature* **308**, 820–826 (1984).
47. E. Y. Song, E. I. Jiménez, H. Lin, K. Le Vay, R. Krishnamurthy, H. Mutschler, Prebiotically plausible RNA activation compatible with ribozyme-catalyzed ligation. *Angew. Chem. Int. Ed.* **60**, 2952–2957 (2021).
48. R. C. Edgar, Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
49. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

Acknowledgments: We thank J. Moretti for pilot experiments on emulsions, M. Magde for help with the DLS analysis of the emulsion, B. Paegel for helpful suggestions for the composition of the emulsion, S. Y. Chen for helping to write an R script for the HTS analysis pipeline, and Y. Su (UCSD Molecular Mass Spectrometry Facility) for outstanding technical work on the LC-MS detection of GTP in the reaction mixture; Dr. Su carried out the LC-MS analysis and assisted in interpretation of data in this paper. **Funding:** This work was supported by the NASA [NNX16AJ27G] issued through the Science Mission Directorate, in Astrobiology/ Exobiology to U.F.M. A.A. was supported, in part, by the U.S. Department of Education through the GAANN fellowship P200A150251. J.T.A. was supported, in part, by a Roger Tsien Fellowship from the Department of Chemistry and Biochemistry at UCSD. **Author contributions:** U.F.M. designed the in vitro selection and biochemical experiments. A.A. set up and performed the in emulsio selection and the bioinformatic analysis from HTS data. D.M. developed and performed the DLS analysis with samples prepared by A.A. and U.F.M. J.T.A. performed the biochemical analysis of selected ribozymes. Data were interpreted by U.F.M., A.A., J.T.A., and D.M. The manuscript was written by U.F.M., A.A., and J.T.A. and edited by U.F.M., A.A., J.T.A., and D.M. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 30 May 2021
Accepted 17 August 2021
Published 6 October 2021
10.1126/sciadv.abj7487

Citation: A. Akoopie, J. T. Arriola, D. Magde, U. F. Müller, A GTP-synthesizing ribozyme selected by metabolic coupling to an RNA polymerase ribozyme. *Sci. Adv.* **7**, eabj7487 (2021).

Figure S1

A

Round	[cTmp] mM	[6sGsn] mM	[RNA pool] nM
1	50	1	500
2	50	1	500
3	25	0.5	500
4	12.5	0.25	500
5	12.5	0.25	500
6	12.5	0.25	500
7	12.5	0.25	500
8	12.5	0.25	500
9	12.5	0.25	500
10	12.5	0.25	100
11	12.5	0.25	100
12	12.5	0.25	100
13	2.5	0.1	100
14	1	0.05	50
15	2.5	0.1	25
16	2.5	0.1	10
17	0.5	0.05	10
18	0.5	0.05	10

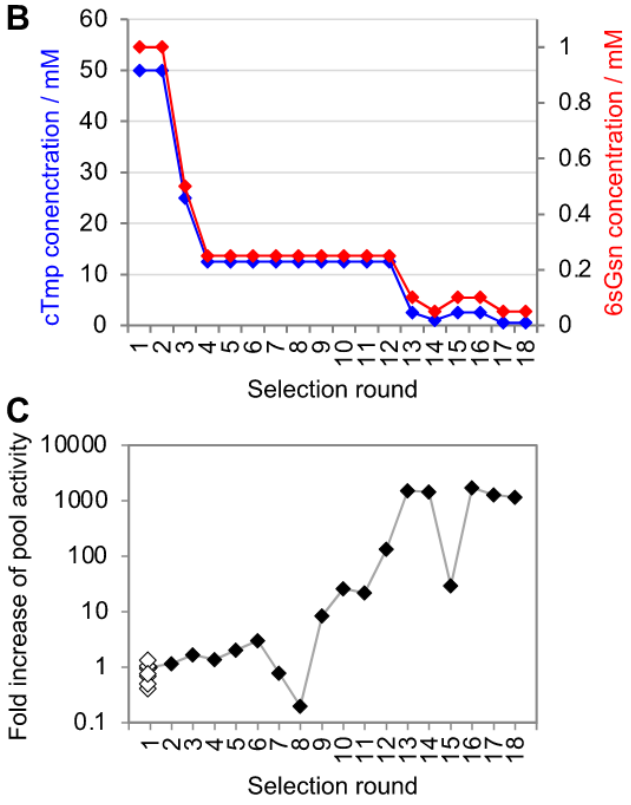


Figure S1. Parameters that were adjusted over the selection rounds. **(A)** For each selection round (left column), the concentrations of cTmp and 6sGsn in the aqueous phase of the emulsion are given in mM, and the concentration of RNA library in nM. Note that the concentration of the polymerase ribozyme stayed constant at 3 micromolar. While only a stoichiometry of 1:1 would be needed to anneal to all pool molecules in bulk, the total concentration of 3 micromolar made sure that about 95% of the emulsion droplets contained at least one polymerase ribozyme molecule, to allow tagging of the pool molecules with produced 6sGTP. **(B)** Plot of the two adjustments affecting the selection pressure, cTmp concentration and 6sGsn concentration, over the selection rounds. **(C)** Plot of the normalized pool activity over the selection rounds. In contrast to figure 2C, the pool activity was plotted logarithmically. The drop in activity in selection rounds 7 and 8 was caused by a selection artifact that appeared to form dimers of the RNA pools. These artifacts were successfully removed by increasing the time for electrophoresis during APM-PAGE purification of active pool molecules. The logarithmic plot emphasizes minor changes in pool activity over earlier selection rounds, and thereby makes clear why parameters in the selection conditions were changed at given rounds.

Figure S2

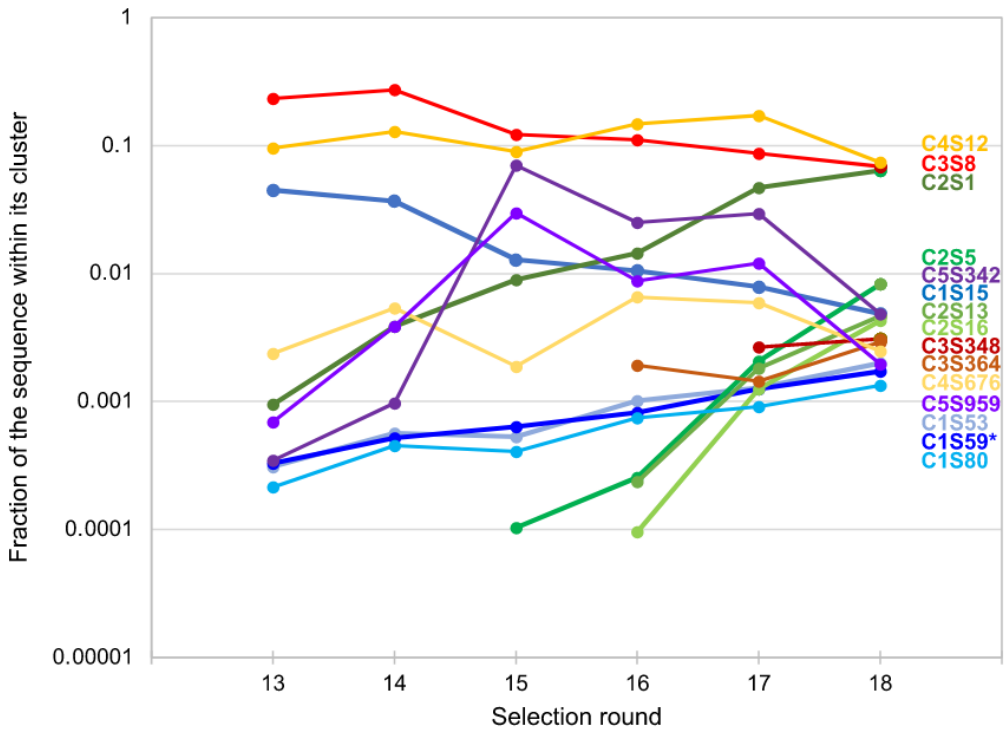


Figure S2. Abundance of individual sequences in the High Throughput Sequencing data for rounds 13 to 18 of the selection. The abundance of each sequence *within its cluster* is plotted as a function of the selection round. The name of the sequence is given on the right, in the same color as the corresponding data set. The colors correspond to the colors used for each cluster in figure 2B, with cluster 1 blue, cluster 2 green, cluster 3 red, cluster 4 yellow, and cluster 5 purple. Different shades of these colors are used to discriminate individual sequences with a given cluster. The sequence with highest biochemical activity for GTP synthesis (cluster 1 sequence 59) is marked with an asterisk, and steadily increased 5.2-fold in frequency from round 13 to 18. The two sequences that showed highest biochemical activity for 6sGTP synthesis (C5S342 and C5S959; see figure S3) also showed the highest enrichment factor in the earlier rounds R13 to R15 (200-fold and 43-fold, respectively) but decreased in abundance after mutagenic PCR was used in round 15.

Figure S3

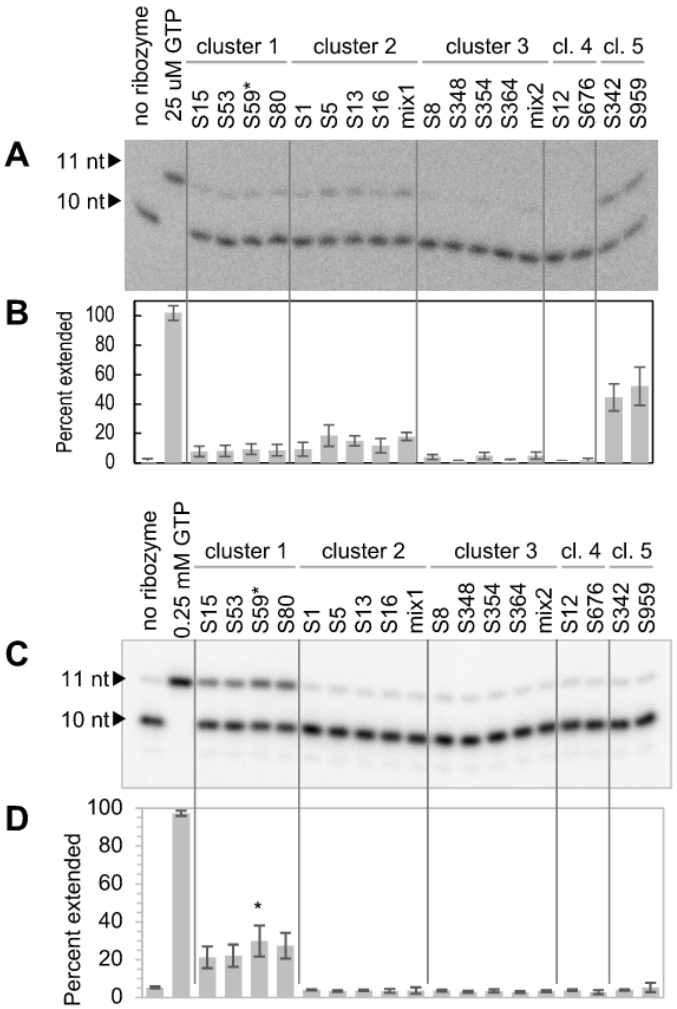


Figure S3. Biochemical assay for nucleoside triphosphorylation activity. The assay was conducted as shown in figure 3A. **(A)** Autoradiogram of reaction products after separation by denaturing 20% PAGE. Individual sequences (S1-S959) from five sequence clusters were tested alongside a positive control (0.25 mM GTP) and a negative control (no ribozyme). A 10 nucleotide long 5'-radiolabeled radiolabeled primer was extended with the produced 6sGTP, utilizing a polymerase ribozyme. The resulting gel shift can be seen in the autoradiogram. **(B)** Analysis of the signals in (A). The percent of radioactive primer that was extended from 10 nt to 11 nt is shown as the average of three replicates. Error bars are standard deviations from three replicates. Ribozyme clones from clyster 5 show the highest activity, followed by ribozymes from cluster 2 and cluster 1. **(C)** The same analysis as in (A) but with guanosine instead of 6-thio guanosine, monitoring the production of GTP instead of 6sGTP. **(D)** Analysis of the signals in (C), as in (B). Ribozyme clones from cluster 1 showed the highest activity. The ribozyme clone S59 in cluster 1 showed the highst average activity and was chosen for further analysis.
The assay was so sensitive that the background reaction between 6sGsn and cTmp at near-neutral pH led to detectable signal. This sensitivity emerged because (i) the assay employed on the order of 1,000-fold more 6sGsn (mM) than pool molecules (mM), because (ii) even micromolar concentrations of 6sGTP are sufficient to efficiently extend the radiolabeled RNA 10-mers (*27*), and because (iii) the shift of 0.1% of the bands in PAGE analysis was within the quantifiable range.

Figure S4

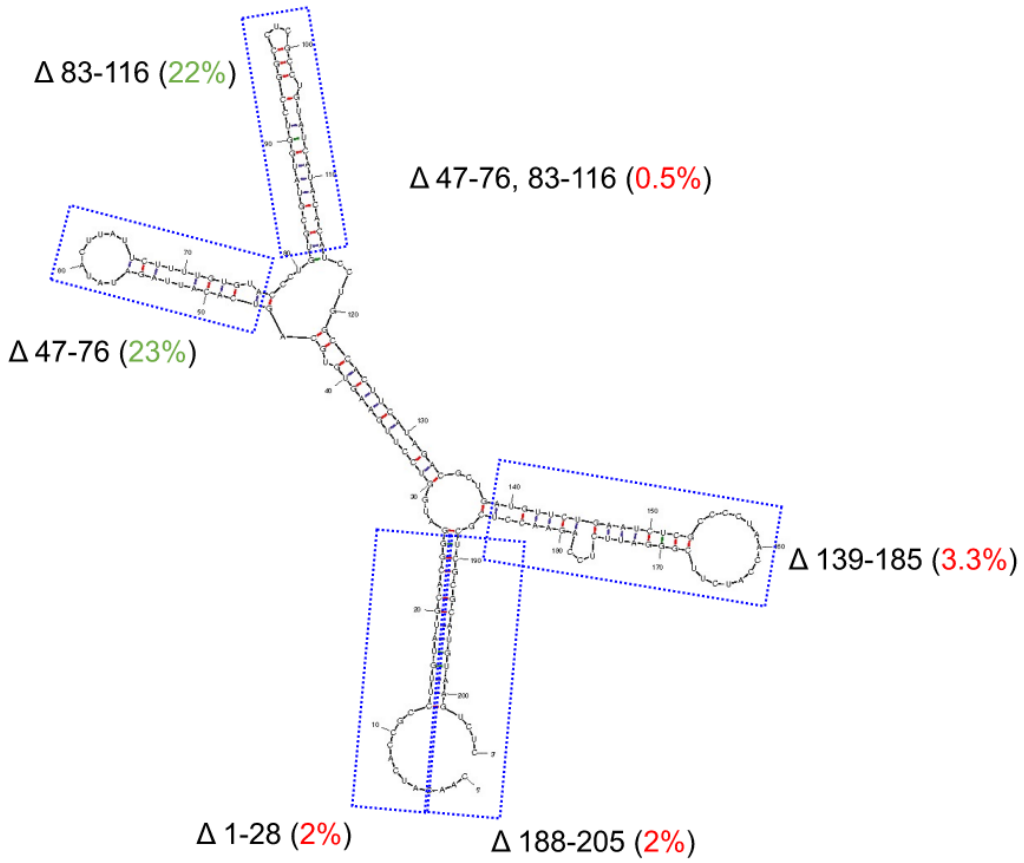


Figure S4. Truncation analysis of the most active isolate from the selection, cluster 1 clone 59. The activities of different truncation constructs are projected on a preliminary secondary structure (Zuker 2003, Nucl. Acids Res 31, 3406). The deleted regions in the ribozyme are shown in blue boxes for each deletion construct. The activities using the trans-assay (figure 3) are included with the label of each deletion variant. The full-length ribozyme generated enough GTP so that 25% of the radiolabelled primer were extended. The values of 22% and 23% were within experimental error of the 25% activity measured for the wild-type. The results suggested that the regions 47-76 and 83-116 were individually dispensable but not in combination. All other tested regions appeared essential to the ribozyme.

Figure S5

A

GTR1:
CAACATCACCGCCTTGTATGCACGGGATGGTCCTTGAAGTGTGCAGCCCTGTGCGTATGGTCC
GGCCTCGCCTGTATCATACACATCCTGGCCACTTCATAGACGCTGATGTTCTGAATCTCGCCC
CTAACCATCTTCGGGATTCTCCAGAACCTCGCTCGCGCATGTAAGTCTC-3'

D60.198 (anneals to position 1 – 60):
CCATACGCACAGGGCTGCACACTTCAAGGACCATCCCGTGCATACAAGGCGGTGATGTTG-3'

D60.199 (anneals to position 70 – 130):
TAGGGGCGAGATTCAGAACATCAGCGTCTATGAAGTGGCCAGGATGTGTATGATACAGGC-3'

D46.32 (anneals to position 131 – 175):
GAGACTTACATGCGCGAGCGAGGTTCTGGAGAATCCCGAAGATGGT-3'

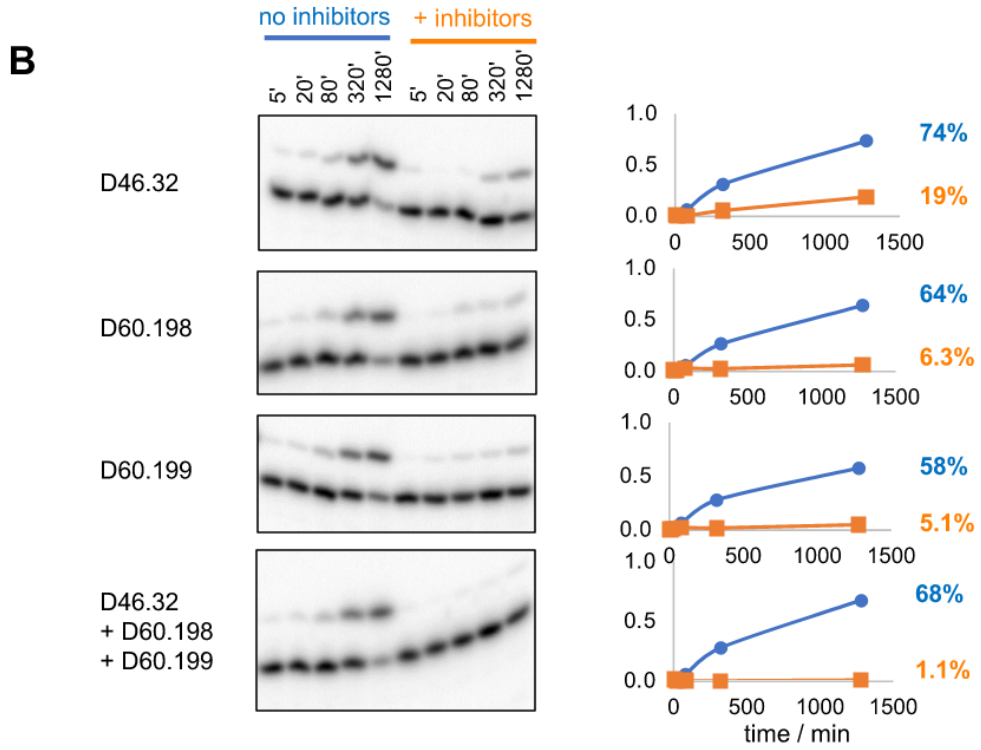


Figure S5. Quenching the activity of the guanosine triphosphorylation ribozyme GTR1 using complementary DNA oligonucleotides. **(A)** sequence of the GTR1 ribozyme and each of the three inhibitor oligonucleotides, with the complementarity indicated in color. **(B)** Activity of the ribozyme with and without each of the inhibitors. For each inhibitor, or combination of inhibitors, a triphosphorylation time course was measured with over 21 hours (1280 min). Because each individual inhibitor allows >5% of activity within 21 hours, the combination of all three DNA oligonucleotides D60.198, D60.199, and D46.32 was chosen to quench the activity of GTR1 in the assays for the optimization of reaction conditions, and the measurement of the catalyzed rate.

Figure S6

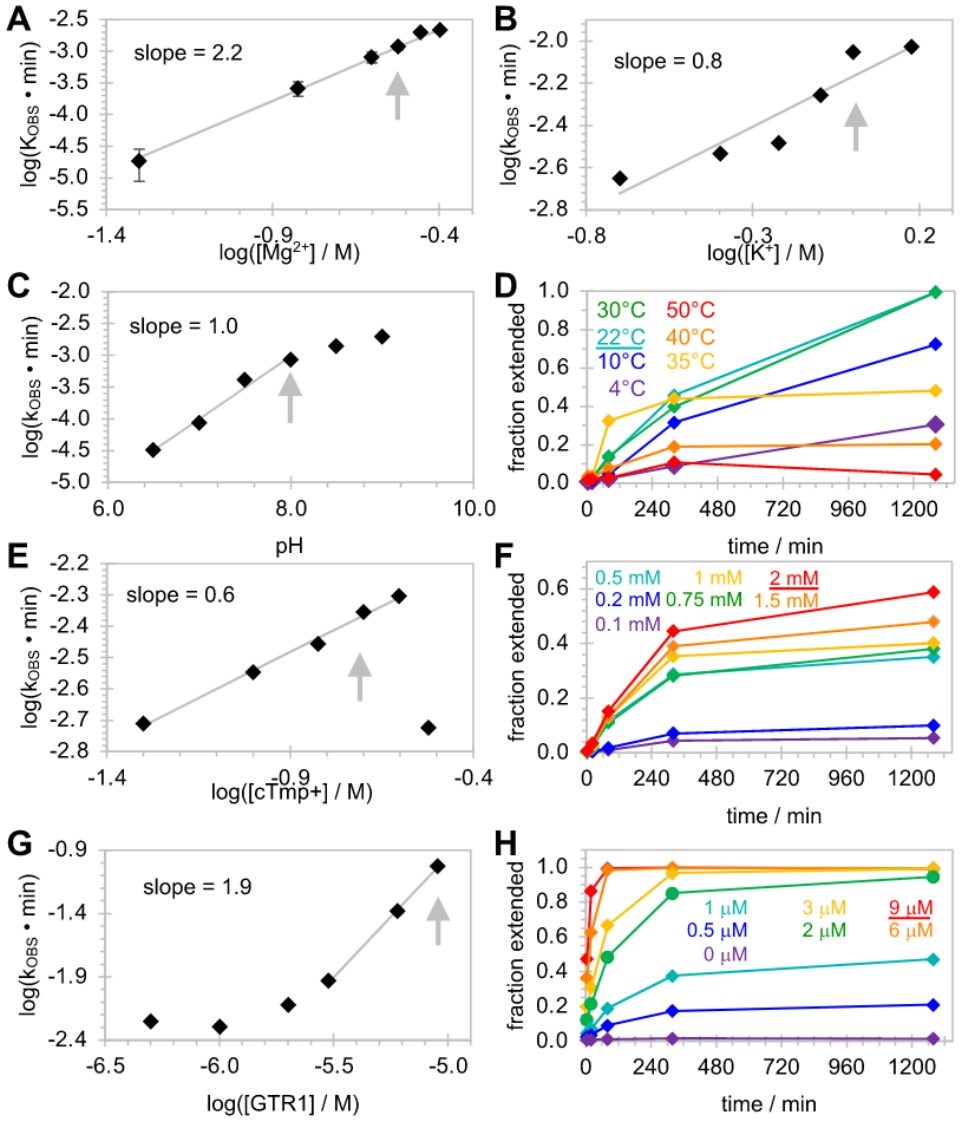


Fig. S6. Optimization of reaction conditions for GTR1. The reaction rates were determined by single-exponential curve fitting to reaction time courses and plotted as function of the reaction conditions. The condition in one assay was then used for the next assay. Starting from an initial condition of 125 mM Mg^{2+}, 100 mM KCl, 50 mM Tris/HCl pH 8.3, 22°C, 25 mM cTmp, 1 mM Gsn, and 1 µM ribozyme the conditions were successively optimized to **(A)** 300 mM Mg^{2+} (error bars are standard deviations from three experiments), **(B)** 1 M K^+, with a 95% confidence interval of +/-0.3 for the slope of 0.8, **(C)** pH 8.0, **(D)** 22°C, **(E)** 200 mM cTmp, **(F)** 2 mM guanosine, and **(G,H)** 9 µM GTR1. The sigmoid dependence on the GTR1 concentration suggested that at low GTR1 concentration a portion of the ribozyme was inactive. On the other hand, at GTR1 concentrations above 2 µM ribozyme catalysis slowed but a larger fraction of the ribozyme folded into the active conformation. Because high concentrations of KCl, $MgCl_2$, and cTmp led to precipitation the chosen 'optimal conditions' do not always represent the highest points in the graphs. Panels D, F, and H are shown as kinetic plots to illustrate the reduced amplitude at certain conditions. Gray arrows and underlined legends indicate the chosen conditions. The gray lines show linear least-squares fits to data points, with slopes given in the inserts. The data for the dependence on temperature, guanosine concentration and GTR1 concentration are shown as kinetic plots to clarify that the reaction amplitudes were sometimes significantly below 100% (panel D,F,H).

Figure S7

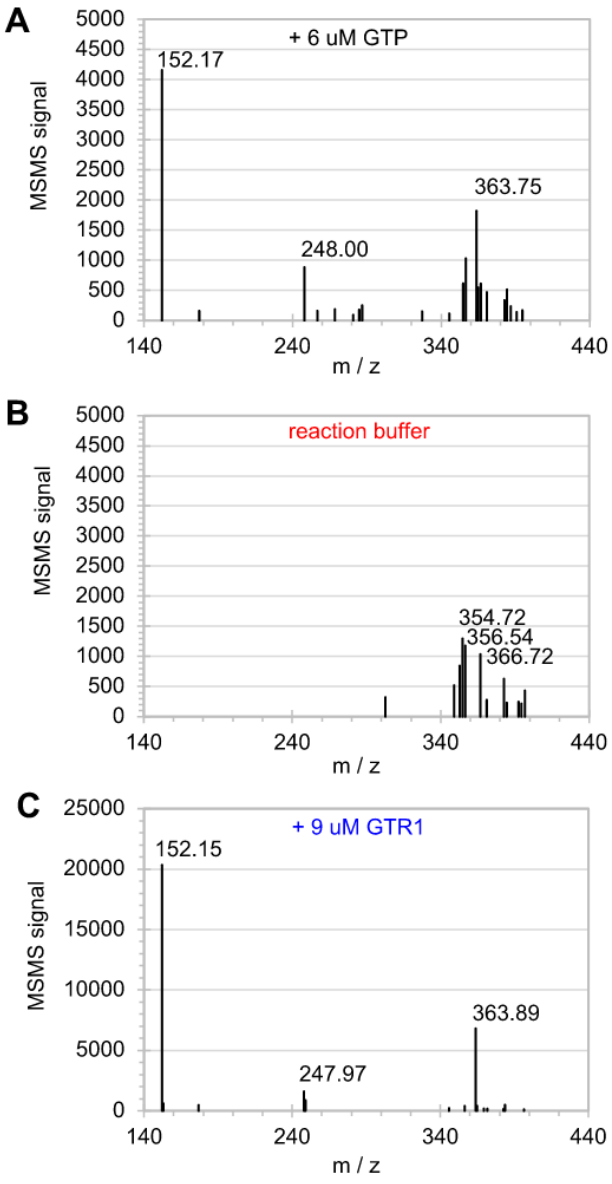


Figure S7: MSMS analysis of the fraction eluting at 1.9 minutes, the HPLC fraction where GTP showed a peak. **(A)** MSMS of reaction buffer containing 6 uM GTP. The characteristic signals are labeled. **(B)** MSMS of reaction buffer. Peaks in the vicinity of the characteristic GTP signals are labeled. No GTP characteristic signals were detected. **(C)** MS of the reaction with 9 uM GTR1. The three GTP characteristic signals are labeled.

Chapter 4, in full, is a reprint of the material as it appears in Akoopie, A.; Arriola, J. T.; Magde, D.; Müller, U. F. “A GTP synthesizing ribozyme selected by metabolic coupling to an RNA polymerase ribozyme”, *Sci. Adv.,* 2021. The dissertation author was the second author of this paper. He contributed by characterizing a novel guanosine triphosphate synthase ribozyme.

Chapter 5

Prebiotically Plausible Peptides Can Support Ribozyme Function

5.1 Abstract

Catalytic RNAs (ribozymes) were likely central in supporting early stages of life on Earth. The first ribozymes probably emerged in the presence of prebiotically generated peptides because amino acids are generated under conditions that lead to the formation of nucleotides, and amino acids can oligomerize into peptides under prebiotically plausible conditions. Here we tested whether the presence of such unencoded, prebiotic peptides could have aided the emergence of ribozymes, by conducting an *in vitro* selection of self-triphosphorylation ribozymes from random sequence. The selection was done in the presence of ten different octapeptides composed of ten different, prebiotically plausible amino acids, each as a mixture of D- and L-stereoisomers. After five rounds of selection and high throughput sequencing analysis, biochemical assays measured the catalytic activity of 30 selected RNA sequences. One ribozyme that benefitted from at least five of the ten octapeptides was analyzed in more detail. The most beneficial peptide increased ribozyme activity up to 20-fold at pH 5.9 and reduced the need for high Mg^{2+} concentrations. The peptide's effect on the ribozyme's secondary structure mirrored in part that of a pre-incubation, suggesting that the peptide may help the ribozyme overcome kinetic folding barriers. The results are discussed in the context of origins of life.

5.2 Introduction

Early stages of life on Earth likely relied on catalytic RNAs to promote self-replication and evolution [1–4]. This idea is supported by the finding that the machinery for encoded protein synthesis in all contemporary life forms is catalyzed by RNA [5,6], that many of today's enzyme cofactors are nucleotides or variants thereof [7,8], and by the idea that RNA's bi-functionality as informational molecule and as catalyst could have allowed an interdependent DNA-RNA-protein system to arise from an earlier stage dominated by catalytic RNAs [1–4]. However, it is unclear how such an RNA-based self-replicating, and evolving system could have emerged in a prebiotic environment [9]. One approach to address this challenge is to investigate the possible help of molecules that likely existed on prebiotic Earth but that don't have that function in today's biochemistry. An example lies in α-hydroxy acids, which could have catalyzed peptide bond formation in wet-dry cycling scenarios but don't have that role in protein biosynthesis today [10,11]. Analogously, we explore here whether prebiotically plausible peptides containing non-biological amino acids, in mixtures of D- and L-isomers could have aided the function of catalytic RNAs.

Amino acids and peptides can be generated under prebiotically plausible conditions, much easier than the building blocks of RNA (compare [10,12–14] to [15–18]). Since amino acids appear as side products in reactions that may have given rise to nucleotides [19], it is reasonable to assume that the first catalytic RNAs emerged in the presence of amino acids and peptides. Such peptides might have benefitted early ribozymes: Prebiotically plausible, cationic peptides increase the thermodynamic

stability of duplex RNA [20], RNA/peptide coacervates modulate RNA structure [21], and peptides can aid the function of ribozymes by co-localizing the ribozymes with their substrate RNA in coacervates [22], and by helping a ribozyme interact sequence-nonspecifically with its RNA substrate [23]. However, these previous studies tested the effect of peptides on individual, existing ribozymes that evolved in the absence of peptides. In contrast, a recent study from our lab identified new ribozymes that benefit from octapeptides composed of the 20 biological amino acids, all as the biological L-stereoisomers. This study showed that peptides consisting of canonical L-amino acids generated can benefit newly emerged ribozymes up to ~10-fold (depending on conditions), and that conjugates of such peptides with polyaromatic hydrocarbons can lead to ~100-fold enhancements of ribozyme activity [24]. To directly address the question whether *prebiotically plausible* peptides could have aided the emergence of the first ribozymes it is necessary to simulate the emergence of the first ribozymes - for example by *in vitro* selection from random sequence - in the presence of prebiotically plausible peptides that consist of prebiotically plausible amino acids and include both D- and L-stereoisomers.

The model system used in this study for the emergence of ribozymes is an *in vitro* selection for self-triphosphorylation ribozymes.[25] This *in vitro* selection system yielded more than 300 different ribozyme clusters from $1.7 \bullet 10^{14}$ different sequences[26], and has performed robustly in multiple different selection contexts [24,25,27,28]. Self-triphosphorylation ribozymes are relevant for prebiotic scenarios because they catalyze the reaction between their 5'-hydroxyl group and the prebiotically plausible energy source cyclic trimetaphosphate (cTmp) [29,30] to generate 5'-triphosphates. These

characteristics make this *in vitro* selection an ideal model system to study what molecules could have helped the emergence of prebiotically relevant ribozymes from random sequence.

Here we describe the *in vitro* selection of self-triphosphorylation ribozymes in the presence of prebiotically plausible peptides. The analysis of one ribozyme / peptide pair showed an activity increase of up to 20-fold in the presence of the peptide and reduced the magnesium concentration required for activity. The effects appeared to be mediated by the peptide's help to overcome kinetic folding barriers. The results are discussed in the context of origins of life.

5.3 Results

To test whether prebiotically plausible peptides could help the emergence of ribozymes, we performed an *in vitro* selection of self-triphosphorylating ribozymes in the presence of peptide octamers containing ten different prebiotically plausible amino acids, covering five biological and five non-biological amino acids (figure 5.1). The ten octapeptides were composed of ten prebiotically amino acids, in frequencies that mirror their abundance in prebiotic model reactions [12,13] (22 x glycine, 14 x alanine, 9 x aspartate, 8 x b-alanine, 6 x a-amino butyric acid, 6 x b-amino butyric acid, 5 x valine, 5 x a-amino isobutyric acid, 3 x g-amino butyric acid, 2 x serine). All amino acids with a stereocenter were represented as mixtures of D- and L-isomers. Since the peptides included a-, b-, and g-amino acids and D- and L-isomers at stereocenters they had heterogeneous backbones, with a total of 232 configurations. The length of eight amino

acids per peptide was chosen because this length is accessible via wet-dry cycling reactions in the presence of a-hydroxy acids [10,11].

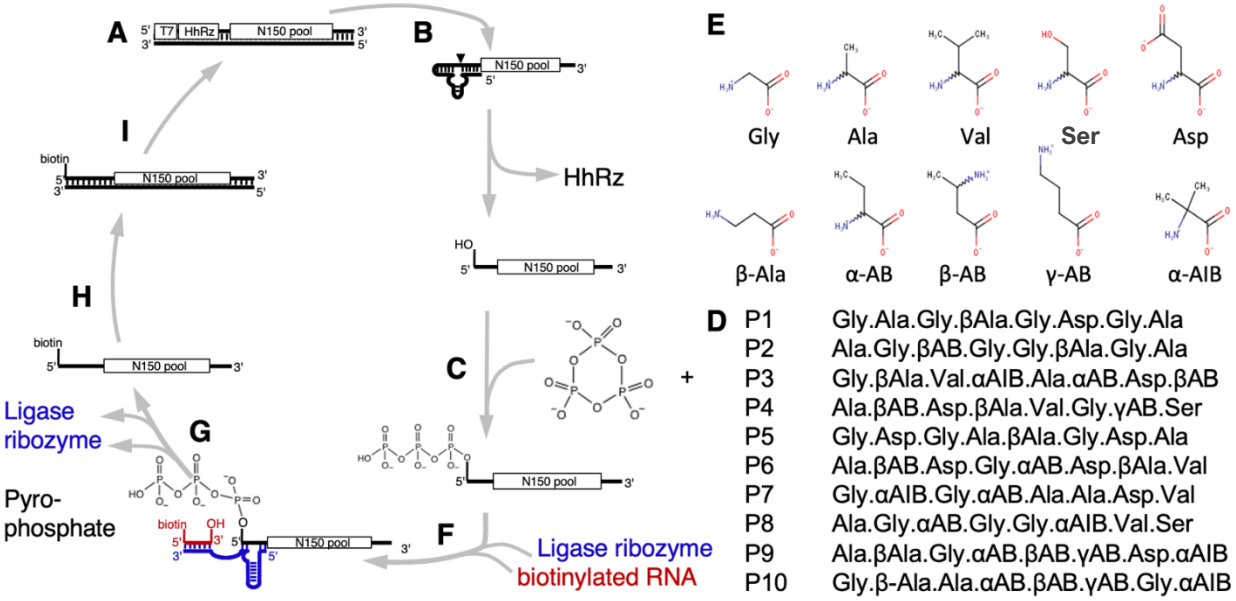


Figure 5.1 Schematic of the *in vitro* selection procedure for ribozymes that triphosphorylate their 5'-hydroxyl group in the presence of prebiotically plausible octapeptides. (A) The DNA library contained a promoter for transcription by T7 RNA polymerase (T7), the sequence encoding a hammerhead ribozyme (HhRz), and 150 nucleotides of randomized sequence (N150) flanked by constant regions. (B) Transcription into RNA generated a 5'-terminal hammerhead ribozyme, which self-cleaved co-transcriptionally and thereby generated a 5'-hydroxyl group at the RNA library. (C) The RNA library was incubated in the presence of 50 mM Tris/HCl pH 8.3, 100 mM $MgCl_2$, 50 mM cTmp, and each 1 mM of (D) ten different octapeptides that were composed of (E) ten different, prebiotically plausible amino acids, each as mixtures of their D- and L- isomers, at 1 mM each. (F) Library RNAs that triphosphorylated their 5'-hydroxyl group were ligated to a 5'-biotinylated oligonucleotide (red) using the R3C ligase ribozyme (blue). This allowed their capture and isolation via streptavidin coated magnetic beads. (F) Reverse transcription to DNA and (G) PCR amplification re-generated the DNA library, now enriched for RNAs that are able to self-triphosphorylate the the presence of the ten peptides.

The *in vitro* selection was conducted as described earlier [24,25], with an effective complexity of $4.2 \bullet 10^{14}$ sequences in the first selection round (figure 5.1). The RNAs were challenged to catalyze the nucleophilic attack of their 5'-hydroxyl group to the phosphorus of cyclic trimetaphosphate (cTmp), thereby generating a 5'-triphosphate. The incubation with cTmp was done in the presence of the described ten peptides, with each at a concentration of 1 mM. Importantly, there was no selection pressure to utilize

the peptides: the procedure selected for self-triphosphorylation activity, and any use of peptides would result only if there was a natural tendency of the ribozymes to benefit from the peptides. RNAs that successfully catalyzed the reaction, and therefore carried a 5'-triphosphate, were then isolated via ligation to a biotinylated oligonucleotide and streptavidin-coated magnetic beads, reverse transcribed, and PCR amplified to complete one cycle of selection. After four rounds of selection, the RNA library was dominated by active sequences as judged by the decrease of PCR cycles required after reverse transcription (figure S5.1).

High throughput sequencing (HTS) analysis showed that several sequence clusters were enriched in the selection protocol (figure 5.2A). To identify whether individual RNAs were benefitting from the presence of peptides, a fifth selection round was appended, in 22 different variants that differed in the presence of peptides. The first variant was done as previously, in the presence of all ten peptides. The second variant was done in the absence of peptides. The remaining variants were done in the presence of each one of the ten peptides, at either 1 mM or 10 mM concentration. All RNA libraries from these 22 variants of selection round 5, and the three preceding selection rounds were then subjected to HTS, and the enrichment of sequence clusters determined (figure S5.2). Based on the enrichment factor with individual peptides, and the number of reads (for statistical significance), a total of 30 RNA sequences representing 30 different sequence clusters was chosen for further analysis to identify RNAs that benefitted from peptides at 1 mM or 10 mM peptide concentration (figure 5.2B).

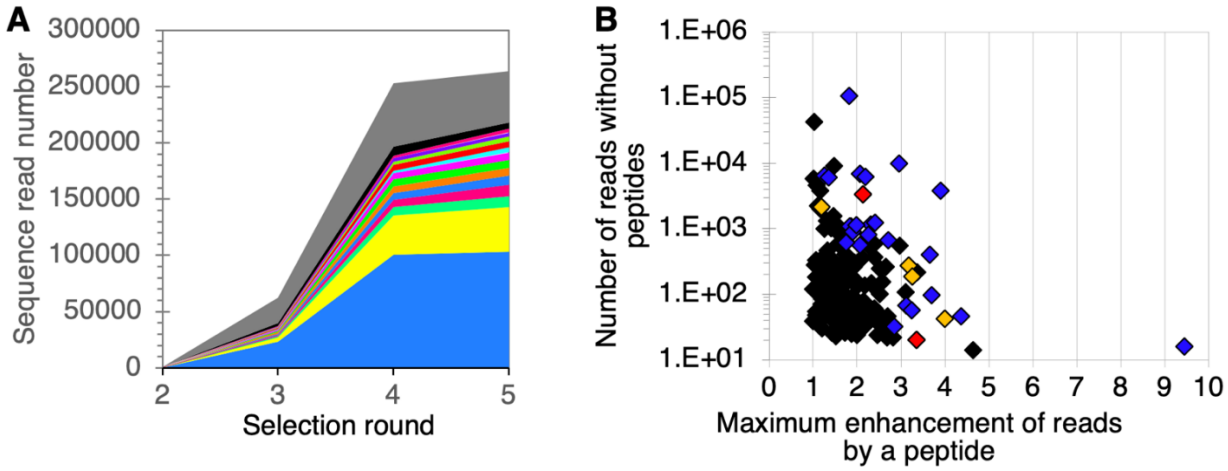


Figure 5.2. High Throughput Sequencing analysis. (A) Enrichment of sequence clusters during the selection. The number of sequence reads is shown in a stacked format, with the 13 most abundant sequence clusters indicated in color (each at least 45,000 reads in all 25 sub-libraries). The ~100 sequence clusters with lower frequency (but above 500 reads in all 25 sub-libraries) are shown in the grey section. (B) Plot of the number of reads in the selection sub-round 5 without peptides, as a function of the fold enhancement by the most beneficial peptide. Each diamond represents one sequence cluster. Sequence clusters were chosen for biochemical analysis (colored diamonds) if they displayed a combination of high enhancement from peptides, and a high number of reads (for statistical confidence). Yellow and red diamonds represent the six sequences that were found promising in biochemical assays, and red diamonds mark the two clusters S2 and S36 showed the strongest, reproducible benefit from peptides.

During biochemical testing, the chosen 30 RNAs showed minimal peptide benefit under selection conditions at pH 8.3 (figure S5.3). In contrast, higher peptide benefit was detected at pH 7.0 (figure S5.4). Six RNAs (c36, c337, c2, c10, c235, c238) showed a benefit from peptides in this initial screen and were tested in triplicate. A reproducible peptide benefit existed for ribozyme S2 at 10 mM peptide P7, and a likely benefit existed for ribozyme S36 with 10 mM peptide P9. When ribozymes S2 and S36 were tested individually with all ten peptides, the most beneficial effects came not from peptides P7 and P9 (as predicted by HTS analysis) but from peptides P1, P2, and P4 (figure 5.3). Importantly, the benefits of individual peptides were well-correlated between ribozymes S2 and S36 (R^2 = 0.921), suggesting that these ribozymes were benefitting from the peptides with the same mechanism. Since ribozyme S2 showed a stronger

peptide benefit than ribozyme S36, and peptide 4 benefitted ribozyme S2 slightly stronger than peptide 2, on average, ribozyme S2 in combination with peptide 4 was chosen for further analysis.

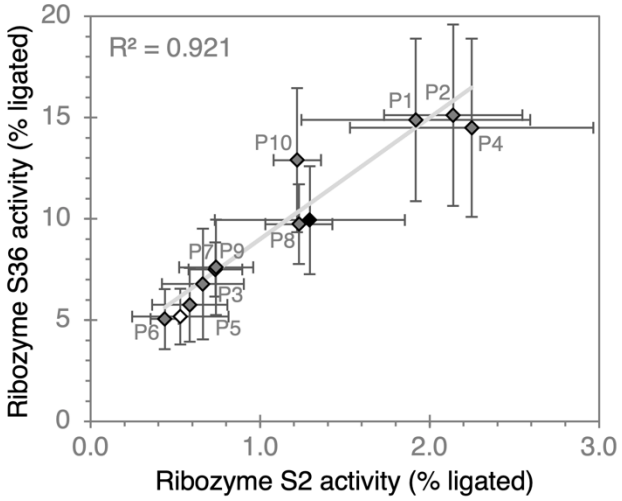


Figure 5.3. Correlation between the activities of ribozyme S2 and ribozyme S36. Both ribozymes were allowed to react in the presence of ten individual peptides (P1-P10, each 10 mM), no peptides (empty diamond), or all ten peptides, each 1 mM (black diamond). The reactions were performed at 100 mM $MgCl_2$, 50 mM cTmp, and 50 mM HEPES / NaOH pH 7.0. A linear fit (grey line) hadf an R^2 value of 0.921. Error bars are standard deviations from three independent experiments.

The length of ribozyme S2's initial isolate was 201 nucleotides. To reduce the size of this ribozyme for easier analysis, 3'-terminal truncations were generated and tested for activity (the 5'-terminus cannot be truncated because it contains the 5'-hydroxyl group at the active site). When the 3'-terminus was truncated to 180 nucleotides, ribozyme activity increased about 3-fold, with a benefit from peptide 4 remaining at 4.4-fold (figure S5.5). Specific truncation lengths gave a non-trivial pattern of ribozyme activity and peptide benefit, and ribozymes shorter than 177 nucleotides showed decreased activity. The ribozyme S2T18 (180 nucleotides in length) was chosen for further analysis.

The reaction conditions of ribozyme S2T18 in the absence and presence of 10 mM peptide 4 were successively optimized regarding the Mg^{2+} concentration, cTmp concentration, pH, and reaction temperature. Higher Mg^{2+} concentration led to an increase in ribozyme activity, and Mg^{2+} concentrations of 200 mM and 400 mM resulted in increased activity even in the absence of peptide 4 (figure 5.4). Since the focus of this study was to determine how peptides can benefit ribozyme activity, a total Mg^{2+} concentration of 100 mM was chosen, where 10 mM peptide 4 resulted in a ~13-fold stronger signal at 3 hours reaction time (figure 5.4A). Similarly, the cTmp concentration with the largest benefit of peptide was at an intermediate 50 mM concentration (figure 5.4B), therefore a combination of 100 mM $MgCl_2$ and 50 mM cTmp was chosen for the next experiments. Variation of the reaction pH (figure 5.4C) showed that the peptide benefit completely disappeared at pH 9, was weak at pH 8.3, and was strongest at pH 5.9. A pH of 5.3 did not give rise to ribozyme activity, even in the presence of peptide 4, therefore the pH of 5.9 was chosen for further experiments. The reaction temperature resulted in strongest product formation at 35°C, while the highest ratio +/- peptide 4 resulted at 25°C (figure 5.4D), therefore a reaction temperature of 25°C was used for further experiments. Under these optimized conditions, the reaction kinetics of ribozyme S2T18 were measured in triplicated with, and without 10 mM peptide 4 (figure 5.4E). Interestingly, there was a lag time of about 30 minutes for peptide 4 to exert its benefit. In the time between 3 hours and 12 hours - before the reaction kinetics reached saturation - peptide 4 resulted in an average of 22-fold increase in product formation.

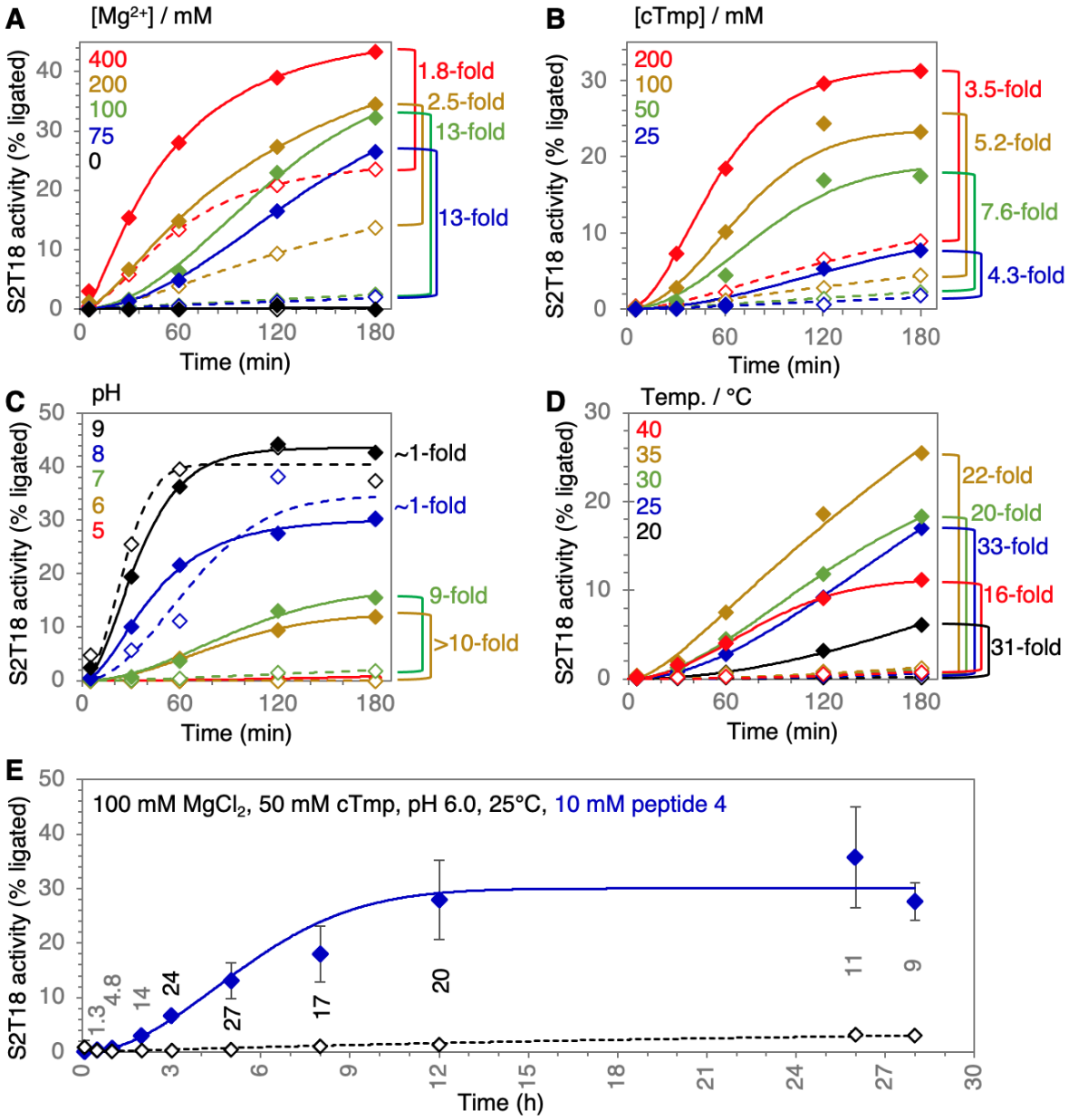


Figure 5.4: Optimization of reaction conditions for ribozyme S2T18 and its benefit from 10 mM peptide 4. Each reaction condition is represented in a different color (inserts). The reaction with 10 mM peptide 4 is represented as filled diamonds and solid lines; the reaction without peptide is shown with empty diamonds, and dashed lines. Kinetics were fitted with a double-exponential equation to reflect two processes. (A) Kinetics of ribozyme S2T18 at different Mg^{2+} concentrations, and 50 mM cTmp, 50 mM HEPES/NaOH pH 7.0, and 22°C. The enhancement of activity at 3 hours is shown on the right. (B) Kinetics of ribozyme S2T18 at different cTmp concentration, and 100 mM $MgCl_2$, 50 mM HEPES/NaOH pH 7.0, and 22°C. (C) Kinetics of ribozyme S2T18 at different pH values and 100 mM $MgCl_2$, 50 mM cTmp, at 22°C, with 50 mM buffer with the respective pH (Tris/HCl pH 9.0, Tris/HCl pH 8.3, HEPES/NaOH pH 7.0, MES/NaOH pH 5.9, MES/NaOH pH 5.3). (D) Kinetics of ribozyme S2T18 at different temperatures and 100 mM $MgCl_2$, 50 mM cTmp, 50 mM MES/NaOH pH 5.9. (E) Kinetics under optimized conditions, with (blue) and without (black) 10 mM peptide P4, at 100 mM $MgCl_2$, 50 mM cTmp, 50 mM MES/NaOH pH 5.9, and 25°C. Error bars are standard deviations of triplicate experiments. If error bars cannot be seen they are smaller than the symbols. The fold increase in the presence of peptide at the 3 hour time point is shown in grey for the time points of 0.5 to 28 hours.

To determine whether peptide 4 modulates the secondary structure of ribozyme S2T18, we performed chemical probing with the SHAPE reagent 1M7 (figure 5.5A). The 1M7 reagent acetylates the 2'-hydroxyl groups of flexible nucleotides, which allowed constructing a secondary structure, in which flexible nucleotides were positioned outside of duplexes. [31] The software RNAstructure was used to obtain a secondary structure model that best matched the SHAPE data [32]. The ribozyme contains two multi-helix junctions, at least three loops and at least three bulges (figure 5.5B). Unlike the secondary structures of two previously characterized self-triphosphorylation ribozymes [24,25,28], the reacting 5'-terminus of ribozyme S2T18 is mostly flexible. Positions 25-27 may be forming a long-range tertiary contact with positions 127-129. The possibility of positions 127-129 to alternatively pair with positions 118-120 may pose a kinetic hurdle in arriving at the catalytically active structure.

The pattern of accessibility changed in the presence of 10 mM peptide 4 (figure 5.5C, S5.6), suggesting that peptide 4 modulates the ribozyme's secondary structure. The effect of the peptide was similar to a 2-hour incubation for the region 114-124, but the effect of 10 mM peptide 4 and a 2-hour incubation different in most other parts of the ribozyme (figure S5.6). This suggests that peptide 4 helps the ribozyme overcome kinetic folding barriers in the specific region 114-124, but that peptide 4 does not have the additional effect of increasing the flexibility of the regions 5-11, 38-42, and nucleotides 95-99. Together, these data show that a prebiotically plausible peptide is able to increase the activity of a ribozyme more than 20-fold effect, with distinct effects on the ribozyme's secondary structure.

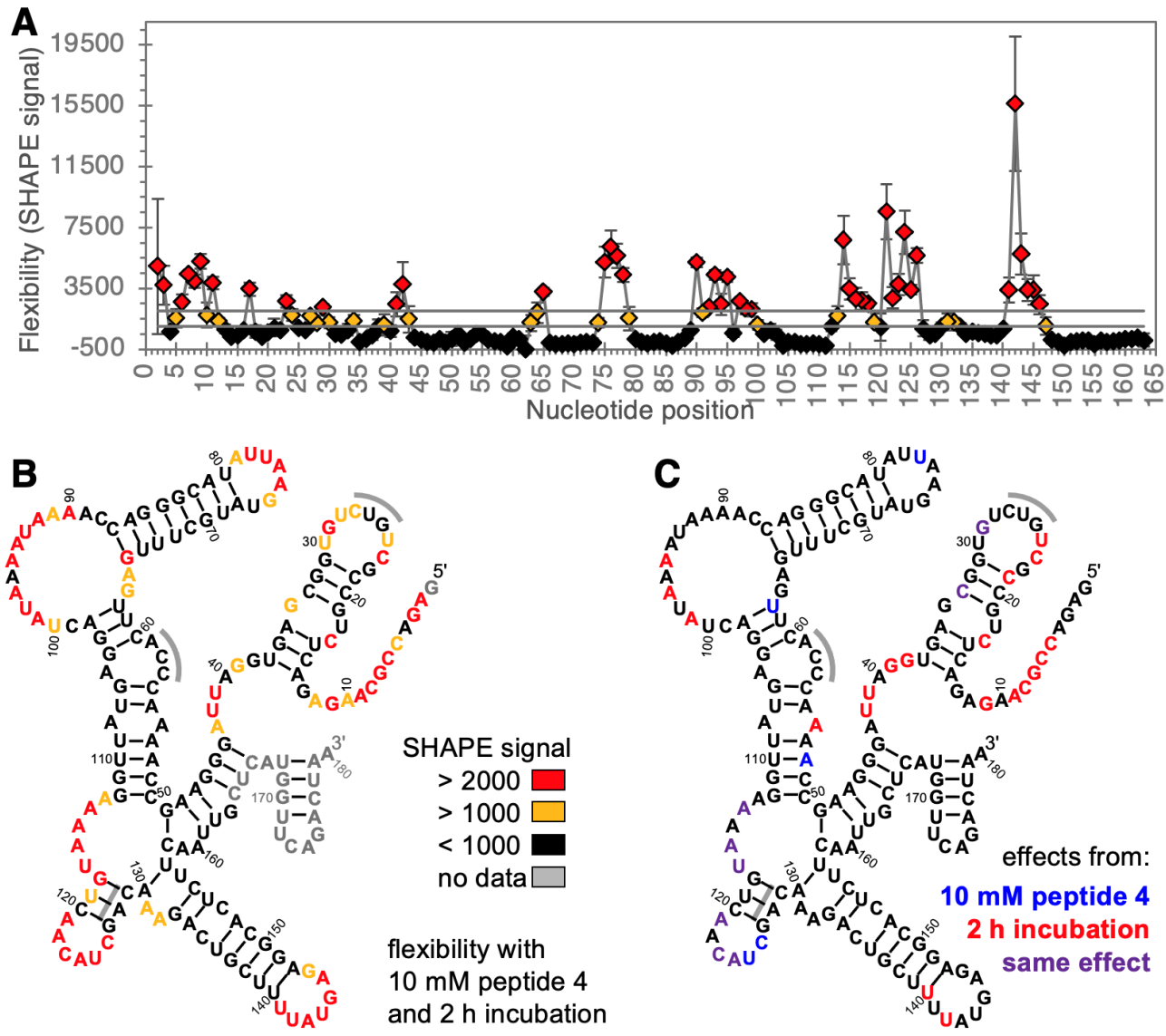


Figure 5.5 Secondary structure of the ribozyme S2T18, and its response to peptide 4 or a 2-hour pre-incubation. (A) The nucleotide flexibility, derived from SHAPE probing (MorrisonWeeks07), is shown as a function of the nucleotide position of the ribozyme. The ribozyme was probed after 2 hours of incubation in 10 mM peptide 4, 100 mM $MgCl_2$, 50 mM cTmp, 50 mM MES/NaOH pH 5.9, and 25°C. Symbols are the averages of triplicate experiments, with error bars showing their standard deviation. Values higher than 2000 are shown in red, values between 1000 and 2000 are shown in yellow, and values lower than 1000 are shown in black, as indicated by the cutoff lines (horizontal grey lines). (B) Secondary structure model constrained by the SHAPE data in (A), predicted using the software RNAstructure [32]. Color symbols use the same code as in (A). grey lines indicate positions that may be forming tertiary contacts within the ribozyme. (C) Influence of 10 mM peptide 4 or 2 hours of incubation in 100 mM $MgCl_2$, 50 mM cTmp, 50 mM MES/NaOH pH 5.9, and 25°C on the secondary structure. Positions that were affected in their SHAPE signal outside the standard deviation (Figure S7) are shown in blue if their deviation resulted from the presence of 10 mM peptide 4, and in red if the change resulted from 2 hours of incubation in reaction buffer. Positions that were affected by both of these changes are marked in purple.

5.4 Discussion

To test whether prebiotically plausible peptides could have aided the emergence, and function of ribozymes in early stages of life, we conducted an *in vitro* selection for self-triphosphorylation ribozymes in the presence of ten prebiotically plausible peptides. The ribozyme S2T18 benefitted stronger from peptide P4 (>20-fold) than a previously discovered ribozyme Rz23 / peptide pair (<10-fold), which consisted exclusively of biological L- amino acids [24]. We hypothesize that both ribozymes Rz23 and S2T18 use their peptides to overcome kinetic folding barriers because their secondary structure shows changes in the accessibility of specific nucleotides in the presence of their beneficial peptide, and because in both cases, multiple completely different peptides can mediate the benefit, showing that no high sequence specificity is required. It is unclear whether the stronger benefit from prebiotically plausible peptides is meaningful because these ribozymes have a very different secondary structure, but it is instructive that the ribozyme using prebiotic peptides shows a stronger benefit than the ribozyme based on biological L-amino acids. One possible explanation could be that a less regular backbone might help the RNA to explore different conformations better.

Several ribozymes benefit in their *in vitro* activity from the presence of spermidine [33–35]. However, the observation that increasing the temperature above 25°C did not lead to a strong activity increase in the absence of P4, and the lack of an effect by spermidine (figure S5.7) suggested that the mechanism of peptide 4's benefit for ribozyme S2T18 is more complicated that simply overcoming a folding barrier. A possible explanation is that the ribozyme can access at least two alternative folds

accessible, and peptide 4 can guide ribozyme S2T18 to the catalytically active structure. The observation that high Mg^{2+} concentrations remove the need for a peptide cofactor could be explained if the close association between two negatively charged backbones is necessary to achieve the active structure. Future studies will try to identify the mechanism how peptides can aid ribozyme S2T18 achieve its catalytically active fold.

The present study is an early step to explore how prebiotically plausible peptides could benefit the rise of an RNA-based organism. It is important to note that the benefits of such peptides would likely be very different those of encoded peptide sequences that evolved to benefit ribozyme function. Prebiotically plausible peptides contain a wide diversity of amino acids [13] with mixed stereochemistry, and likely different abundances of amino acids depending on the local microenvironment, the time of year and day, and the weather. With ribozyme / peptide model systems it will now be possible to tease out what amino acid compositions, and therefore which environmental conditions, would be promising to help the emergence of an RNA-based system as an early stage of life.

5.5 Materials and Methods

5.5.1 Synthesis and Purification of Peptides

Peptides were synthesized on rink amide PS resin (0.77 mmol/g, Novabiochem) and Rink Amide ChemMatrix (0.45 mmol/g, Biotage) following the Fmoc/tBu strategy using a microwave-assisted peptide synthesizer (Alstra, Biotage). Side chain protecting groups for amino acids were as follows: Ser(tBu), His(Trt), Lys(Boc), Asp(OtBu), Glu(OtBu),

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

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

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

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

5.5.2 *In Vitro* Selection

A 183 nucleotide (nt) long DNA oligomer that included 150 positions with randomized sequence and flanked by 5'- and 3'-constant regions was purchased from Integrated DNA Technologies (IDT). 5'-GAGACCGCAAGAGAC(N150)CATGGTTCAGACTACAAC-3' A 82 nt long oligomer was used to add the promoter sequence of T7 RNA polymerase and the sequence of the hammerhead ribozyme (underlined) to the 5’ end of the 183-mer. 5’-AATTTAATACGACTCACTATAGGGAGCGGTCTCCTGACGAGCTAAGCGAAACTGCGGAAACGCAGTCGAGACCGCAAGAGAC-3’. Amplification with the 67-mer and a short 3’ primer generated the double stranded DNA pool. qPCR was used to determine the

amount of amplifiable DNA and this number was used to estimate that the starting DNA pool consisted of 4.2 x 10^14 sequences.

The DNA pool was transcribed into RNA using T7 RNA Polymerase under standard conditions. 100 nM of purified RNA sub-pool was incubated with 50 mM Tris/HCl, 3.3 mM NaOH, 100 mM $MgCl_2$, 50 mM of freshly dissolved trimetaphosphate and 1 mM each of ten prebiotically plausible peptides (10 mM total) . This mixture was allowed to incubate for 3 hours at room temperature. The reaction was quenched by ethanol precipitation. The RNA pellet was washed with chilled water and desalted using P30 Tris/HCl spin columns (Bio-Rad). Active ribozymes were ligated to a biotinylated oligomer (biotin-d(GTAACATAATGA)rU) using the R3C ligase ribozyme whose arms were designed to anneal to the 5' constant region of the pool RNA and the biotinylated oligomer. A mixture containing 1.3 uM of desalted pool RNA, 1 uM ligase ribozyme, 1.2 uM biotinylated oligomer, 100 mM KCl, 100 mM Tris/HCl, and 60 pM triphosphorylated RNA was heated to 65°C for 2 minutes then cooled to 30°C at a rate of 0.1°C per second. The triphosphorylated RNA was used to generate a small amount of ligation product and thereby reduce the number of PCR cycles, and prevent artifacts during PCR amplification. After heat renaturation, an equal volume of a mix consisting of 40% (w/v) PEG8000, 4 mM Spermidine, and 50 mM $MgCl_2$ was added to the ligase reaction mixture. This mixture was incubated at 30°C for 3 hours. After the incubation step, the reaction was quenched by adding final concentrations of 13.9 mM sodium/EDTA, 50 mM Tris/HCl, 50 mM KCl, 0.012% (w/v) Triton X-100, and 1.19 uM of a 61 nt oligomer that was complementary to the ligase ribozyme (5' AGTAACATAATGACTTCAACCCATTCAAACTGTTCTTACGAACAATCGAGCACAAGA

GAC-3'). This mixture was heated at 50°C for 10 minutes. Streptavadin magnetic beads (Promega) were washed thrice with 20 mM HEPES/KOH pH7.2, 0.01% (w/v) Triton X-100 and 50 mM KCl. Biotinylated RNA was then captured by mixing the RNA with the magnetic beads and rotating end-over-end at room temperature for at least 30 minutes. A magnetic rack was used to focus the beads and the beads were washed twice with a solution containing 0.01% (w/v) Triton X-100 and 20 mM NaOH. Captured RNA was eluted from the beads by incubating the beads in a solution of 25 mM Tris/HCl pH 8.5, 1.56 mM EDTA and 96% formamide at 65°C for 3 minutes. The beads were removed by immediate centrifugation, and the supernatant was concentrated by ethanol precipitation. 1 ug of tRNA was used as precipitation carrier. Captured RNA was resuspended in 10 mM Tris/HCl, pH 8.3. The RNA was reverse transcribed using Superscript III (Invitrogen) and a reverse transcription primer (5'-GTTGTAGTCTGAACCATG-3') complementary to the 3' constant region. The products were then PCR amplified with 5' and 3' primers that could bind to the sequence of the biotinylated oligomer and the 3' constant region of the pool respectively. A second PCR amplification was performed to add the T7 promoter sequence and the hammerhead ribozyme sequence to the DNA.

5.5.3 HTS Analysis

Illumina sequencing primers were added to the ends of rounds 2-5 pools. Pools were sequenced on MiSeq with 10% PhiX. FastQ files were processed using custom Python scripts for rounds 2-5 (including all sub-rounds in round 5). The usearch merge function

was used to merge forward and revers reads. Sequences not containing both constant regions or short randomized regions (<135 nt in length or 10% of 150) were first filtered out. The file was reformatted to Fasta for easier downstream processing. The total counts for each sequence were calculated. The usearch clustering algorithm was used to group similar sequences into clusters. Clusters were defined as sequences that shared at least 75% sequence identity and had at least 100 reads in a given round. Clusters and sequences were tracked through each round. 189 sequences were evaluated further. For each sequence, the fold enrichment in the round 5 sub-rounds were determined by normalizing to the no peptide sub-round.

5.5.4 Generation of 3' Truncations

3' truncations were generated by PCR using Taq DNA polymerase and a series of 3' primers. The 3' primers annealed to different portions of the ribozyme sequence of interest to generate a series of truncations with desired deletions from the 3' end. DNA templates of each truncation were transcribed using T7 RNA polymerase, and purified by denaturing PAGE.

5.5.5 Self-triphosphorylation Activity Assays

Purified ribozyme (5 uM) was incubated with 50 mM Tris/HCl, 100 mM $MgCl_2$, 50 mM of freshly dissolved trimetaphosphate, and either a mixture of peptides, individual peptides or water as a control. This mixture was incubated for 3 hours at room temperature. The

products were ligated to a radiolabeled oligomer (γ^{32}P-d(GTAACATAATGA)rU) using an R3C ligase ribozyme whose arms were designed to anneal to the 5' constant region of the pool RNA and the oligomer. Following the incubation with buffer, Tmp and peptides, the mixture was diluted 1:10 in a solution that contained a final concentration of 0.5 uM ligase ribozyme, 0.5 uM biotinylated oligomer, 100 mM KCl, 100 mM Tris/HCl, and 15 mM sodium EDTA. This mixture was heat renatured at 65°C for 2 minutes then cooled to 30°C at a rate of 0.1°C per second. After heat renaturation, an equal volume of a mix consisting of 40% PEG8000 (w/v), 4 mM Spermidine, and 50 mM $MgCl_2$ was added to the ligase reaction mixture. This mixture was incubated at 30°C for 2 hours. After the incubation step, the reaction was quenched by ethanol precipitation. The reaction products were resolved on a denaturing 15% polyacrylamide denaturing gel.

5.5.6 Secondary Structure Determination Using SHAPE

Selective 2'Hydroxyl Acylation Analyzed by Primer Extension (SHAPE) was performed on the 180-nucleotide long ribozyme isolated in this study, with 1-Methyl-7-nitro-2*H*-3,1-benzoxazine-2,4(1*H*)-dione (1M7) as the chemical probe. 50 pmol of ribozyme were incubated in buffer containing 50 mM MES/NaOH pH 6.0, 100 mM $MgCl_2$, 3.3 mM NaOH, 50 mM Tmp, and 10 mM peptide 4 at room temperature for 2 minutes. This mixture was incubated at room temperature for 2 min or 2 hours. 1M7 was added to this solution such that its final concentration was 2 mM and the concentration of DMSO was 1% of the solution. A negative control was prepared with no 1M7 and 1% DMSO. Both samples were incubated at room temperature for 3 minutes. The reaction was

quenched by ethanol precipitation and resuspended in 10 uL of 5 mM Tris/HCl pH 8.0. The products were reverse transcribed using Superscript III reverse transcriptase (Invitrogen) according to manufacturer's instructions and trace amounts of a radiolabeled primers were added. Three reverse transcription primers were used to probe different portions of the ribozyme. The sequences of these primers are (5'-TTAGTCTGAACCATGA-3', annealed to pos. 165-180); (5'-CGATGTTGACATTTTC-3', annealed to pos. 112-127); and (5'-AAACTCAAGTGGGTTT-3', annealed to pos. 53-68). The RNA template in each sample was degraded by alkaline hydrolysis, by incubating for 5 minutes at 80°C in a solution containing 750 mM NaOH. The reaction was quenched by adding a two-fold stoichiometric excess of acetic acid to generate a NaOAc/HOAc buffer with a final concentration of 300 mM. The products were ethanol precipitated, resuspended in formamide loading buffer and resolved on a denaturing 10% polyacrylamide gel. The RNA secondary structure was determined using a secondary structure prediction from RNAstructure to help guide the secondary structure determination. The SHAPE data was mapped onto an initial secondary structure prediction and manually adjusted as needed to better represent the SHAPE data.

5.6 Supplementary Data

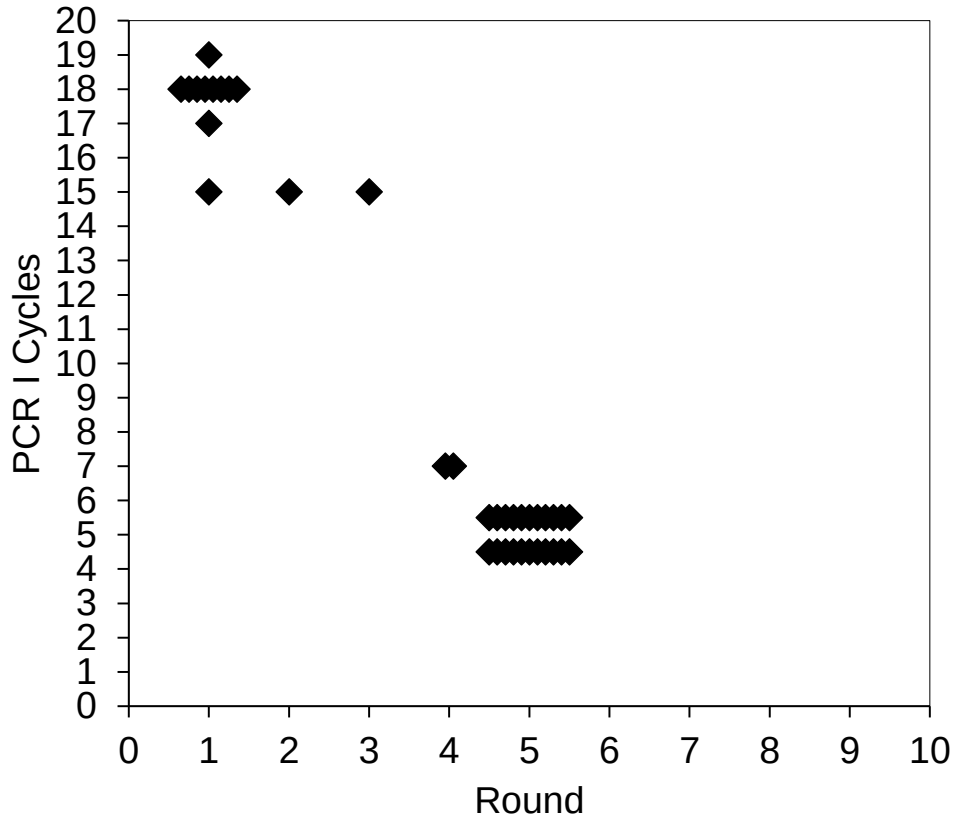
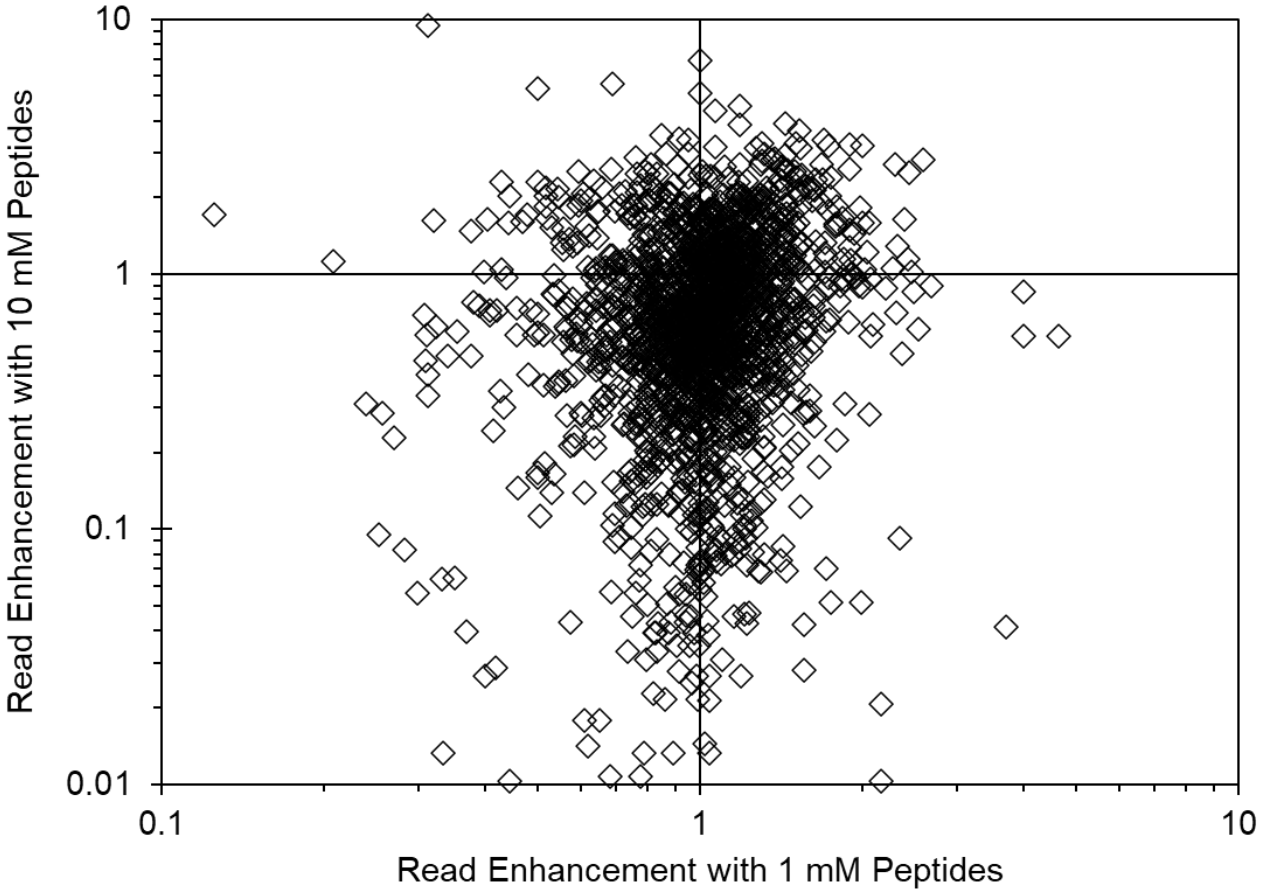


Figure S5.1 Number of PCR I cycles over the course of the selection. 11 sub-rounds were used in round 1 to cover a large portion of the sequence space. Throughout the selection a small amount of pre-selected pool was spiked into a given selection round to account for losses in material. In round 4, a parallel round was performed in which pre-selected pool was not spiked into the sub-round. Since the spiked and non-spiked rounds required the same number of PCR I cycles, we believed the pool was now sufficiently enriched with active sequences that spiking was no longer necessary. In round 5, 22 sub-rounds were performed in which all peptides, no peptides, or one of 10 peptides at two different concentrations was used during the cTmp incubation step.



1 mM Peptide	10 mM Peptide	Number	Percentage (%)
>1	>1	358	18.9
<1	>1	193	10.2
>1	<1	639	33.8
<1	<1	648	34.3
=1	>1	14	0.7
=1	<1	28	1.5
>1	=1	7	0.4
<1	=1	3	0.2

Figure S5.2 Effect of peptides on sequence enrichment over the course of Round 5 sub-rounds. In the plot above, each diamond represents a single sequence and peptide pair. The table displays the number and percentage of sequence/peptide pairs in each quadrant. Pairs with a neutral effect (=1) are counted separately. Most sequences showed lower enrichment as peptide concentration increased.

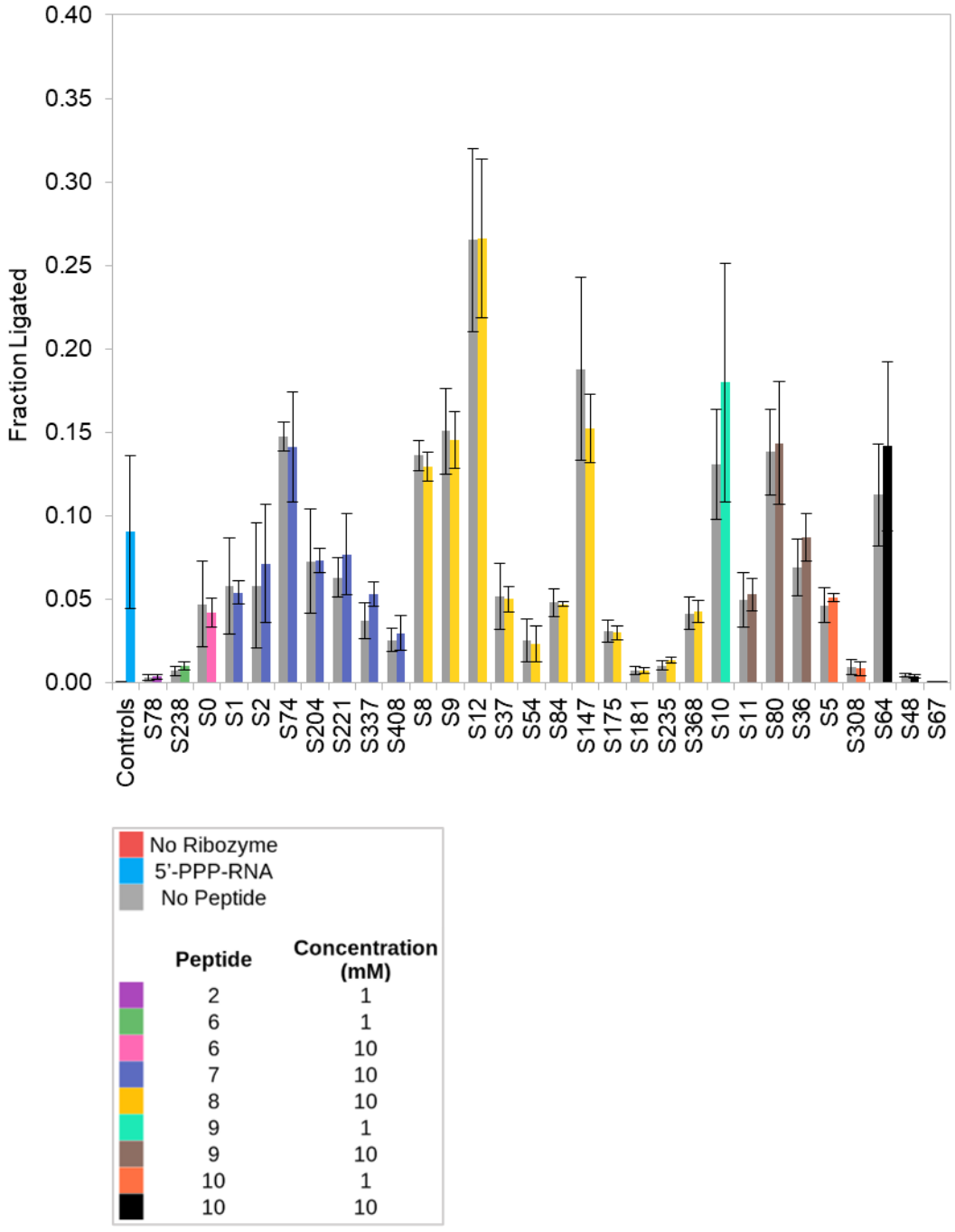


Figure S5.3 Biochemical screen of 30 candidate ribozymes. Pairs of sequences and peptides were selected based on analysis of the HTS data and tested biochemically. Error bars are the standard deviation of triplicate experiments.

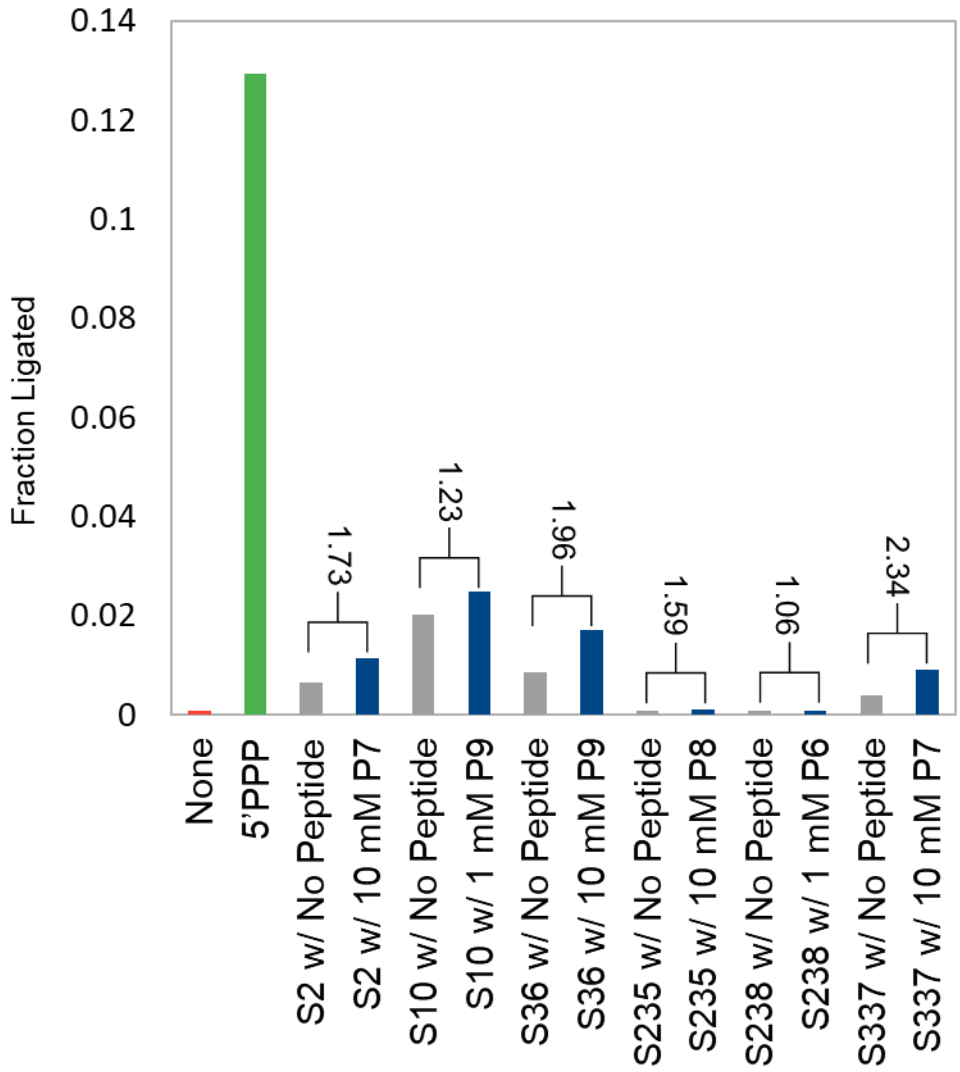


Figure S5.4 Biochemical screen under reduced pH conditions. Based on the initial biochemical screen (Figure S3), 6 sequences were tested under different conditions (not shown). The condition that showed the highest benefit had a reaction buffer with 50 mM HEPES/NaOH pH 7.0.

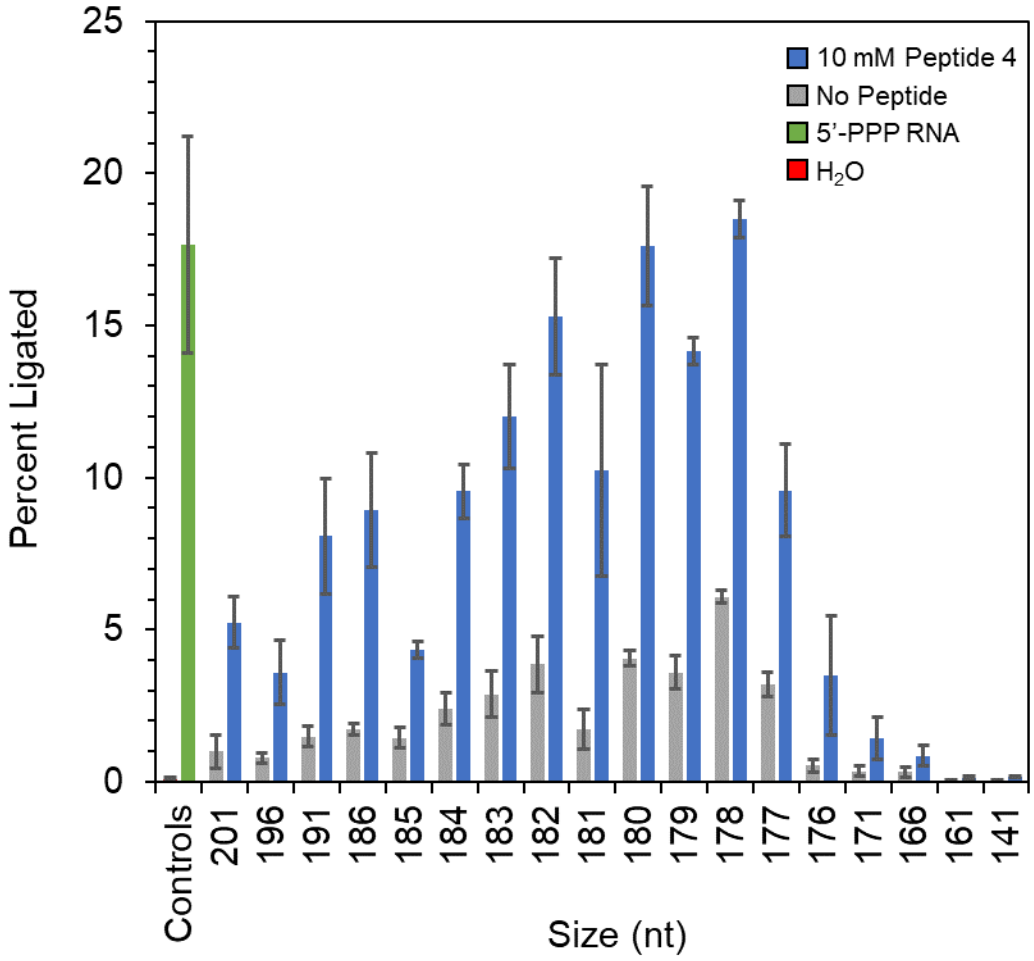


Figure S5.5 Ribozyme S2 truncation series. Pairs of sequences and peptides were selected based on analysis of the HTS data and tested biochemically. Error bars are the standard deviation of triplicate experiments.

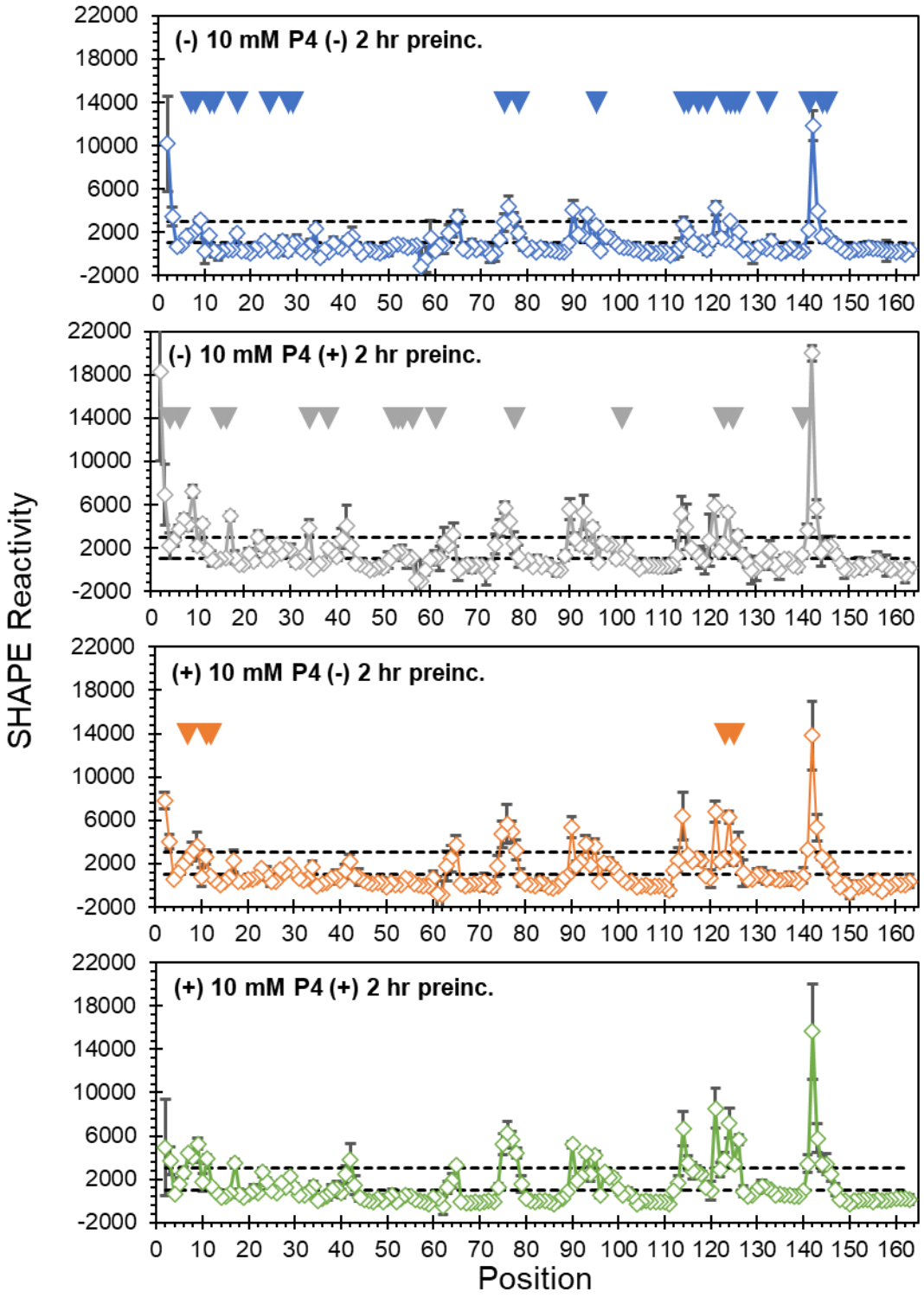


Figure S5.6 Ribozyme S2T18 SHAPE. SHAPE experiments were performed in which no peptide or 2 hour pre-incubation was used (blue), 2 hour pre-incubation was used (grey), peptide was used (orange), or both peptide and 2 hour pre-incubation were used (green). In the 2 hour preincubation, the reaction was allowed to proceed before the 1M7 chemical probe was added. The reactivity of the (+) 10 mM P4 and (+) 2 hr preincubation was compared to all other conditions and the triangles mark where there is a statistically significant change in reactivity.

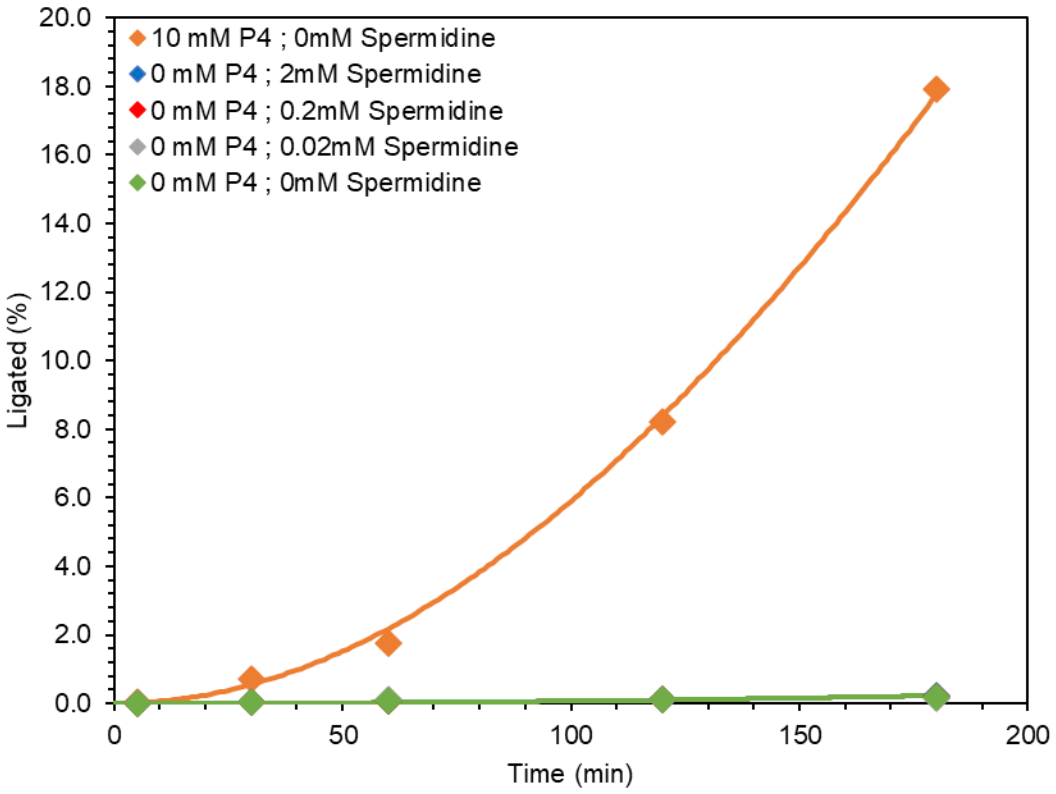


Figure S5.7 Spermidine titration. To test if a polyamine could replicate the beneficial effect of peptide 4 on ribozyme S2, spermidine was titrated with ribozyme S2. There was no perceivable benefit when spermidine was added to the reaction instead of peptide 4 indicating that peptide 4 aids the activity of ribozyme S2 through some other mechanism than hydrogen bonding between the peptide's amino groups and the ribozyme's phosphodiester backbone.

5.7 References

1. Rich, A. On the problems of evolution and biochemical information transfer. *Horiz. Biochem.* **New York, NY**, 103–126 (1962).
2. Woese, C. R. The fundamental nature of the genetic code: prebiotic interactions between polynucleotides and polyamino acids or their derivatives. *Proc Natl Acad Sci U A* **59**, 110–7 (1968).
3. Orgel, L. E. Evolution of the genetic apparatus. *J Mol Biol* **38**, 381–393 (1968).
4. Crick, F. H. C. The origin of the genetic code. *J Mol Biol* **38**, 367–379 (1968).
5. Noller, H. F. Peptidyl transferase: protein, ribonucleoprotein, or RNA? *J Bacteriol* **175**, 5297–300 (1993).
6. Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–30 (2000).
7. White, H. B., 3rd. Coenzymes as fossils of an earlier metabolic state. *J Mol Evol* **7**, 101–4 (1976).
8. Goldman, A. D. & Kacar, B. Cofactors are Remnants of Life's Origin and Early Evolution. *J. Mol. Evol.* **89**, 127–133 (2021).
9. Pavlinova, P., Lambert, C. N., Malaterre, C. & Nghe, P. Abiogenesis through gradual evolution of autocatalysis into template-based replication. *FEBS Lett.* 1873-3468.14507 (2022) doi:10.1002/1873-3468.14507.
10. Forsythe, J. G., Yu, S.-S., Mamajanov, I., Grover, M. A., Krishnamurthy, M., Fernandez, F. M. & Hud, N. V. Ester-mediated amide bond formation driven by wet-

dry cycles: A possible path to polypeptides on the prebiotic Earth. *Angew Chem Int Ed Engl* **54**, 9871–9875 (2015).

11. Krishnamurthy, R. Giving Rise to Life: Transition from Prebiotic Chemistry to Protobiology. *Acc. Chem. Res.* **50**, 455–459 (2017).
12. Miller, S. L. A production of amino acids under possible primitive earth conditions. *Science* **117**, 528–9 (1953).
13. Parker, E. T., Cleaves, H. J., Dworkin, J. P., Glavin, D. P., Callahan, M., Aubrey, A., Lazcano, A. & Bada, J. L. Primordial synthesis of amines and amino acids in a 1958 Miller H2S-rich spark discharge experiment. *Proc Natl Acad Sci U A* **108**, 5526–31 (2011).
14. Foden, C. S., Islam, S., Fernández-García, C., Maugeri, L., Sheppard, T. D. & Powner, M. W. Prebiotic synthesis of cysteine peptides that catalyze peptide ligation in neutral water. *Science* **370**, 865–869 (2020).
15. Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239–42 (2009).
16. Becker, S., Feldmann, J., Wiedemann, S., Okamura, H., Schneider, C., Iwan, K., Crisp, A., Rossa, M., Amatov, T. & Carell, T. Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides. *Science* **366**, 76–82 (2019).
17. Huang, W. & Ferris, J. P. Synthesis of 35-40 mers of RNA oligomers from unblocked monomers. A simple approach to the RNA world. *Chem Commun Camb* 1458–9 (2003).

18. 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* **139**, 1810–1813 (2017).
19. Islam, S., Bučar, D.-K. & Powner, M. W. Prebiotic selection and assembly of proteinogenic amino acids and natural nucleotides from complex mixtures. *Nat. Chem.* **9**, 584–589 (2017).
20. Frenkel-Pinter, M., Haynes, J. W., Mohyeldin, A. M., C, M., Sargon, A. B., Petrov, A. S., Krishnamurthy, R., Hud, N. V., Williams, L. D. & Leman, L. J. Mutually stabilizing interactions between proto-peptides and RNA. *Nat. Commun.* **11**, 3137 (2020).
21. Cakmak, F. P., Choi, S., Meyer, M. O., Bevilacqua, P. C. & Keating, C. D. Prebiotically-relevant low polyion multivalency can improve functionality of membraneless compartments. *biorxiv* (2020) doi:10.1101/2020.02.23.961920v1.
22. Poudyal, R. R., Guth-Metzler, R. M., Veenis, A. J., Frankel, E. A., Keating, C. D. & Bevilacqua, P. C. Template-directed RNA polymerization and enhanced ribozyme catalysis inside membraneless compartments formed by coacervates. *Nat Commun* **10**, 490 (2019).
23. Tagami, S., Attwater, J. & Holliger, P. Simple peptides derived from the ribosomal core potentiate RNA polymerase ribozyme function. *Nat Chem* **9**, 325–332 (2017).
24. Sweeney, K.J. & Muller, U.F. In vitro selection of ribozyme / peptide models for Early life's catalysts. *Manuscr. Prep.*
25. Moretti, J. E. & Muller, U. F. A ribozyme that triphosphorylates RNA 5'-hydroxyl groups. *Nucleic Acids Res* **42**, 4767–78 (2014).

26. Pressman, A., Moretti, J. E., Campbell, G. W., Müller, U. F. & Chen, I. A. Analysis of in vitro evolution reveals the underlying distribution of catalytic activity among random sequences. *Nucleic Acids Res.* **45**, 8167–8179 (2017).
27. Dolan, G. F., Akoopie, A. & Muller, U. F. A Faster Triphosphorylation Ribozyme. *PLoS One* **10**, e0142559 (2015).
28. Arriola, J. T. & Müller, U. F. A combinatorial method to isolate short ribozymes from complex ribozyme libraries. *Nucleic Acids Res.* **48**, e116–e116 (2020).
29. Pasek, M. A., Kee, T. P., Bryant, D. E., Pavlov, A. A. & Lunine, J. I. Production of potentially prebiotic condensed phosphates by phosphorus redox chemistry. *Angew Chem Int Ed Engl* **47**, 7918–20 (2008).
30. Pasek, M. A., Harnmeijer, J. P., Buick, R., Gull, M. & Atlas, Z. Evidence for reactive reduced phosphorus species in the early Archean ocean. *Proc Natl Acad Sci U A* **110**, 10089–94 (2013).
31. Mortimer, S. A. & Weeks, K. M. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J Am Chem Soc* **129**, 4144–5 (2007).
32. Bellaousov, S., Reuter, J. S., Seetin, M. G. & Mathews, D. H. RNAstructure: Web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res.* **41**, W471-474 (2013).
33. Heilman-Miller, S. L., Pan, J., Thirumalai, D. & Woodson, S. A. Role of counterion condensation in folding of the Tetrahymena ribozyme II. Counterion-dependence of folding kinetics. *J. Mol. Biol.* **309**, 57–68 (2001).

34. Suh, Y.-A., Kumar, P. K. R., Taira, K. & Nishikawa, S. Self-cleavage activity of the genomic HDV ribozyme in the presence of various divalent metal ions. *Nucleic Acids Res.* **21**, 3277–3280 (1993).
35. Hammann, C. A spermidine-induced conformational change of long-armed hammerhead ribozymes: ionic requirements for fast cleavage kinetics. *Nucleic Acids Res.* **25**, 4715–4722 (1997).

Chapter 6

Discussion and Future Directions

6.1 A Combinatorial Method to Identify Short Catalytic Motifs

The development of a method to identify short triphosphorylating ribozymes can be useful to researchers using combinatorial methods to identify functional RNA. Often the largest possible randomized pool is used in an *in vitro* selection to increase the effective complexity of the starting pool. Theoretically, if a catalytic motif for a given function is shorter than the randomized region than this motif can exist in multiple registers in the starting pool. The tradeoff is that any resulting active sequences may have extraneous sequences that can be harmful to activity. In at least one case, the advantage of using a longer randomized pool outweighed the disadvantages.[1] Other self-triphosphorylating ribozymes identified in our lab could be truncated to lengths shorter than the length of the randomized pool. The minimized ribozymes are more efficient than the full length ribozymes showing that extraneous sequences are detrimental to the activity of the ribozyme but do not abolish activity (see Chapters 4 and 5 of this dissertation and Moretti & Muller, 2014). The method used in the short TPR selection can be used to identify the shortest possible catalytic core of a ribozyme or other functional RNA. One important caveat is that since the method relies on a randomized 3' primer, catalytic motifs must be close to the 5' end of the sequence.

Multiple selection rounds using the randomized primer method could be used to slowly enrich the pool for shorter ribozymes. In Chapter 2, the shortest sub-pool

identified corresponded to sequences with a length of ~70 nt, while the minimized ribozyme identified in the 70 nt sub-pool could be truncated to 44 nt. Any active sequences in sub-pools shorter than 70 nt may not have survived the selection step or were so rare as to be undetectable by the selection scheme. Instead of performing serial truncations of sequences from the 70 nt pool, it may have been possible to perform multiple selection steps with the randomized primer method to isolate progressively shorter sub-pools.

The identification of a short self triphosphorylating catalytic motif has implications for the RNA world. Since the abiotic polymerization of RNA has not yet been shown to exceed 50 nucleotides in length, the identification of short ribozymes shows that smaller ribozymes can have catalytic function relevant to supporting the metabolism of an RNA World organism.

6.2 Ribozyme Utilization of Nucleoside Triphosphates Generated by Diamidophosphate

The fact that ribozymes can utilize NTPs generated from DAP mediated triphosphorylation indicates that the DAP reaction is efficient enough to generate NTPs in sufficient quantities that ribozymes can utilize them. In addition, the DAP reaction may form some side products that are inhibitory to ribozymes, but these side products are not formed in sufficient quantities to completely abolish activity. Since DAP is soluble in aqueous conditions in the presence of divalent cations, is prebiotically plausible, and can generate NTPs, it may be that early life used DAP to generate NTPs

and then transitioned to a system in which triphosphorylation of nucleosides was mediated by a ribozyme that instead utilized cTmp. Two reaction pathways were tested in this project: NTPs were generated from 5'monophosphate nucleosides or NTPs were generated from unphosphorylated nucleosides. An RNA polymerase ribozyme was able to utilize NTPs generated via DAP when the starting material was a nucleoside monophosphate, but when the starting material was a nucleoside, the polymerase ribozyme could not utilize NTPs in the crude mixture. Future work will need to be done to improve the efficiency of DAP mediated triphosphorylation from nucleosides.

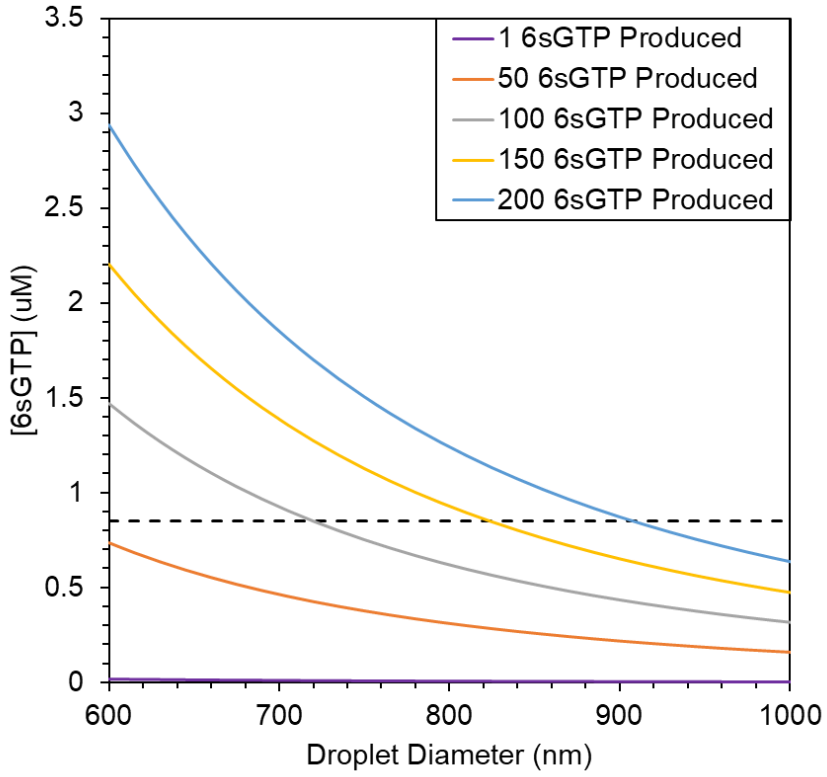


Figure 6.1 The effect of droplet diameter on the concentration of 6sGTP. To select for higher multiple turnover GTP synthase ribozymes, an *in emulsio* evolution will be performed in which the droplet diameter is increased to ~900 nm. In contrast, the previous selection used a droplet diameter of 200 nm. Under optimal conditions, the current GTP synthase ribozyme has a turnover number of ~1.7 corresponding to a concentration of 0.86 uM in a 200 nm droplet (black dotted line). By increasing the droplet diameter, we increase the selective pressure for higher turnover since ribozymes must generate 100s of molecules of 6sGTP to be successfully tagged by the polymerase ribozyme.

6.3 The Identification and Characterization of a Guanosine Triphosphorylating Ribozyme

One issue with the GTP synthase ribozyme is the low turnover number. Future studies will focus on generating a GTP synthase ribozyme with higher turnover. Other *in vitro* selected ribozymes appear to have low turnover number when compared to natural ribozymes or proteins.[2,3] One possibility is that *in vitro* selected ribozymes have not had the benefit of millions of years of evolution to evolve higher turnover numbers. *In vitro* selected ribozymes have been selected to survive a selection step and have not been challenged to generate multiple turnover reactions. Another possibility is that RNA is not capable of high turnover number and requires protein cofactors to enable high turnover. Besides the nucleolytic ribozymes, other natural ribozymes with high turnover have protein components. For example, while active without its protein components, RNase P only exhibits multiple turnover in the presence of its protein components.[4]

A future study will focus on evolving the GTP synthase to increase turnover by increasing the droplet diameter during the *in emulsio* selection. The larger droplets will select GTP synthase variants that produce multiple 6sGTP molecules (figure 6.1). The selection scheme relied on a polymerase ribozyme mediated 3' ligation of thio-modified guanine to a pool molecule. 6sGTP generated by active pool molecules was used as a substrate by the polymerase ribozyme. Because substrate concentration affects enzyme activity and because a larger droplet size decreases the concentration of any 6sGTP generated by active pool molecules, we can tune the selective pressure for multiple turnover ribozymes by increasing the droplet diameter. In addition, it may be

possible to use the *in emulsio* selection scheme to improve the turnover numbers for other classes of ribozymes by adjusting the droplet diameter.

The GTP synthase selection required the use of a DNAzyme to process the pool to ensure sequence and chemical homogeneity of the pool 3'ends. T7 RNA Polymerase was used to transcribe the DNA pool and is known to add poly-A to the 3' end of transcripts.[5–7] The sequence of a small nucleolytic ribozyme such as the HDV ribozyme could be added to the 3'-ends of pools. During transcription, the ribozyme would cleave itself and any poly-A sequences added by T7 RNA Pol; however, most small nucleolytic ribozymes will generate a 2'-3'-cyclic phosphate after cleavage.[8] For this reason, a DNazyme was used since most of the products generated after cleavage by DNazymes would have a 3'OH group.[9] Cleavage with the DNazyme was found to be inefficient and represented a major bottleneck in the *in emulsio* selection. To increase efficiency, the use of M1 Rnase P from *e. coli* to process the 3' end of pool molecules was explored (figure 6.3). Using this method, 70% product conversion in just 30 minutes was observed. Preliminary data (not shown) suggests that pool molecules processed using RNase P could still be used as a substrate by the polymerase ribozyme; however, it is still unclear what fraction of RNase P processed pool molecules have a 3'OH group.

The identification of a GTP synthase ribozyme is a significant advancement in the field of origins of life because it demonstrates that ribozymes are capable of chemically activating nucleosides using cyclic trimetaphosphate. One limitation of this study is that it only demonstrated triphosphorylation of guanine, so future studies will need to focus on identifying ribozymes that can synthesize ATP, UTP, or CTP. Ideally, a general NTP synthase ribozyme could be identified. One might be able to identify such

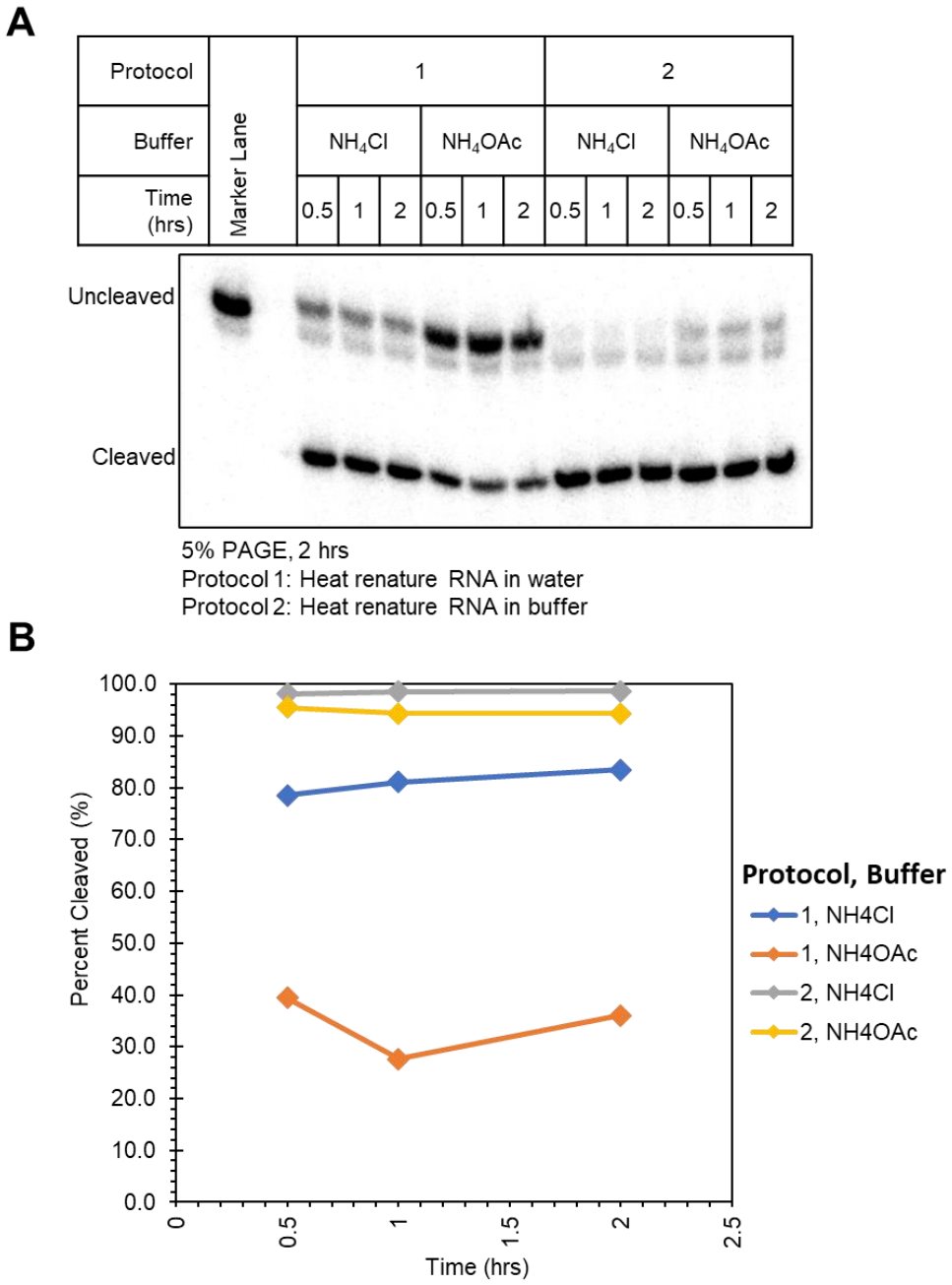


Figure 6.2 Processing of RNA pool molecules using M1 RNase P. (A) PAGE showing the ability of M1 RNase P to process RNA under a variety of conditions. In protocol 1, RNase P and the pool substrate was heat renatured in water (80°C for 2 min, 55°C for 10 min, and 22°C for 5 min). Next, the RNA was incubated for 10 min at 37°C in a pre-incubation buffer containing 50 mM Tris/HCl pH 8.0, 10 mM $MgCl_2$, and either NH_4Cl or NH_4OAc. The pH was decreased by addition of MES/NaOH pH 5.1. The final reaction conditions were 25 mM Tris/HCl pH 8.0, 25 mM MES/NaOH pH 5.1, 15 mM $MgCl_2$, 800 mM NH_4Cl or NH_4OAc, 1 uM RNase P, and 1 uM pool substrate. In protocol 2, the RNA was heat renatured in the pre-incubation buffer and the incubation at 37°C for 10 min was skipped. The rest of protocol 2 was the same as in protocol 1. (B) Quantification of the experiment shown in (A). The optimal condition included a heat renaturation step of the RNA in buffer and a reaction buffer that contained NH_4Cl.

a ribozyme by performing an evolution on the existing GTP synthase ribozyme and simply changing the thio-modified nucleoside after each round. Alternatively, instead of changing the substrate between rounds, three parallel evolutions using either thio-modified adenine, uracil, or cytosine could be performed to identify sequences capable of generating the three other NTPs. In addition, by comparing the sequences of NTP synthase ribozymes, one might be able to determine the NTP binding site of the ribozymes since mutations should occur more frequently where the NTP binds to the ribozyme.

The *in emulsio* selection scheme used in this project could be a powerful tool to identify other trans-acting ribozymes that catalyze bimolecular reactions. For example, the selection system could be used to identify a ribozyme capable of synthesizing a small molecule methyl group carrier. Nucleic acid methylation is a common post-transcriptional modification of rRNA, tRNA, and (sn)RNA.[10] Post-transcriptional modifications have a significant impact on the activities of the associated RNA, for example, tRNA modifications affect translation rate, accurate decoding of mRNA, contribute to maintaining tRNA structure, and control the specificity of aminoacyl-tRNA synthetases.[11,12] In extant biology, methylation of RNA is typically accomplished by protein methyltransferase enzymes where a methyl group is transferred from a small molecule methyl carrier, such as S-adenysl methionine, to the 2'-hydroxyl group of the nucleic acid.[13] Recent work has identified ribozymes capable of methyltransferase reactions using O^6-methyl preQ$_1$, O^6-methylguanine, and S-adenosyl methionine (SAM).[14–16]

Since the structure of the ribosome implies that rRNA directly participates in protein synthesis, this implies that in early life, protein synthesis was entirely mediated by RNA. If a proto-ribosome relied on the use of tRNAs for decoding of mRNA and delivery amino acids, it would stand to reason that such a system would also rely on ribozymes capable of mediating the post-transcriptional modification of tRNA. The majority of tRNA modifications involve a methylation step[10], so methyltransferase ribozymes would have an important role in proto-ribosomal protein synthesis. The identification of methyltransferase ribozymes show that RNA is capable of methylating nucleic acids and if similar ribozymes existed during the RNA World, there would have been a selective advantage for organisms to produce methyl carrier substrates such as O^6-methyl preQ$_1$, O^6-methylguanine, and SAM. A SAM synthase ribozyme would illustrate the most direct connection to modern biology. Today, SAM is produced biosynthetically via a protein enzyme called S-adenosyl methionine synthetase from methionine and ATP.[17] Methionine can be produced from Miller-type spark discharge experiments and ATP could be synthesized by an ATP synthase ribozyme or a general NTP synthase ribozyme. A selection *in emulsio* could be used to identify SAM synthase ribozymes (figure 6.3). A SAM synthase ribozyme along with a trans-acting methyl transferase ribozyme would show how post-transcriptional modification of RNA could be entirely mediated by ribozymes and would support the idea that early protein synthesis was mediated entirely by RNA.

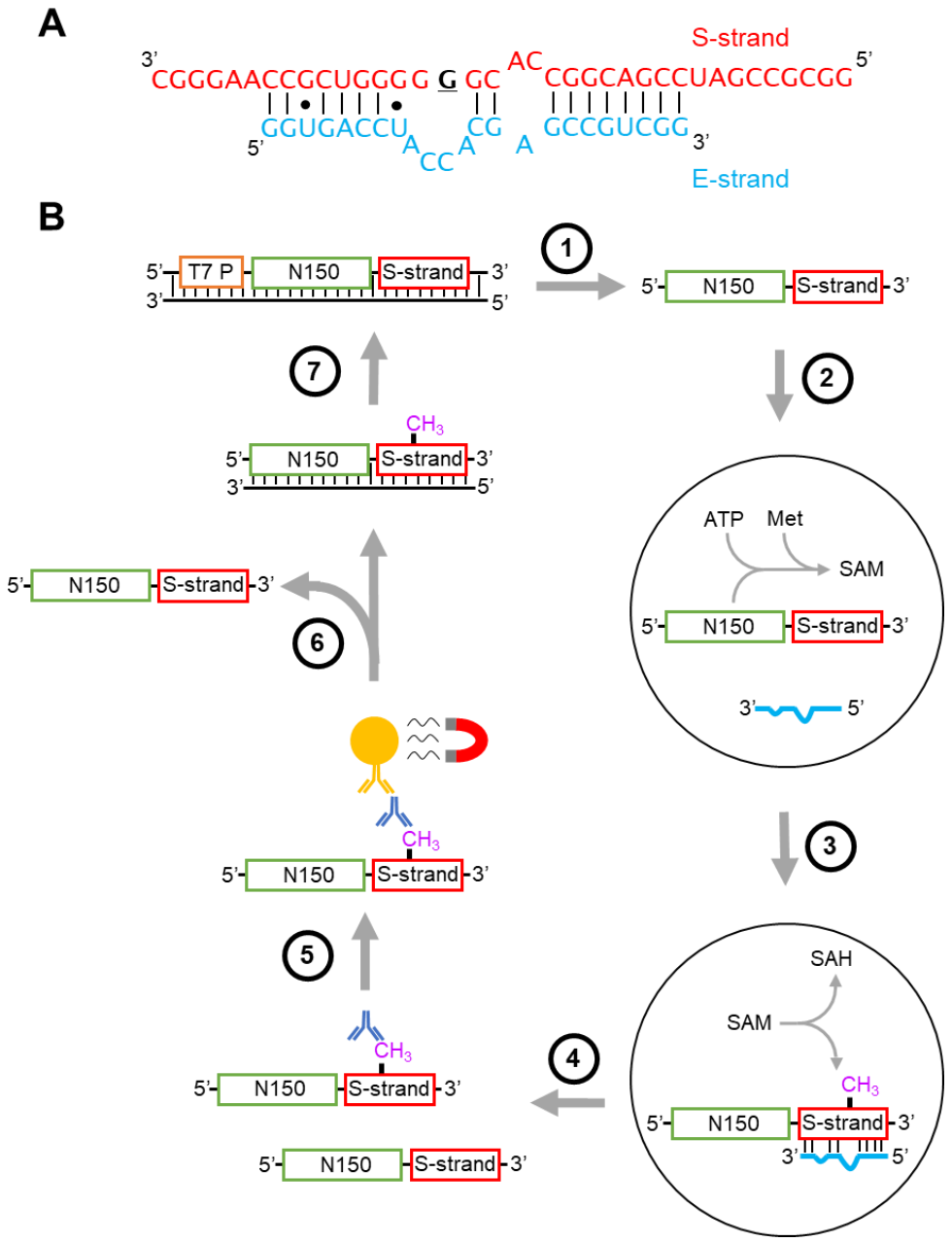


Figure 6.3 A selection scheme for a SAM synthase ribozyme. (A) The secondary structure of a trans-acting SAM dependent methyltransferase ribozyme (SMRz) identified by Jiang, et al. (2021). (B) A DNA pool composed of the sequences for the T7 promoter, a 150 nt randomized region, and the S-strand of the trans-acting SMRz is transcribed into RNA (1). RNA pool molecules are sequestered in a droplet along with ATP, methionine, and the E-strand of the trans-acting SMRz. Active sequences would generate SAM (2). The SMRz acts as a molecular SAM detection platform analogous to the polymerase ribozyme in the selection for GTP synthase ribozymes. SAM synthesized by active pool molecules are used by the trans-acting SMRz to methylate the 3' end of active pool molecules (3). Next droplets are broken open and an antibody against m6 guanine binds to sequences that could produce SAM (4). Immunoglobulin G coated magnetic beads are used to isolate active pool molecules (5). The pool is reverse transcribed (6) and PCR amplified (7) to regenerate the DNA pool for further rounds of selection. A similar strategy could be used to isolate ribozymes capable of synthesizing O^6-methyl preQ$_1$, or O^6-methylguanine.

6.4 A Prebiotically Plausible Peptide that Promotes Ribozyme Activity

The peptide selection utilized a self triphosphorylation selection scheme as a model system to investigate if a set of prebiotically plausible peptides could benefit the activity of ribozymes. We were able to identify a ribozyme peptide pair in which the ribozyme's activity increased 20-fold in the presence of the peptide. The peptides used in this study were composed of amino acids identified in Miller-type spark discharge experiments, noncanonical amino acids, and racemic mixtures of amino acids. In addition, they lacked cationic and aromatic amino acids. The fact that these prebiotically plausible peptides could benefit the function of ribozymes indicates that RNA World ribozymes could have recruited peptides to aid in chemical catalysis.

Intriguingly, the peptide benefit observed was enhanced when the pH was reduced to 5.9. Modeling of environmental conditions on the early Earth suggest that the Hadean ocean had a pH ~6 so ribozymes that could recruit peptides to enhance their activity under these conditions would have an evolutionary advantage.[18–20] The identification of a peptide that benefits ribozymes at less than neutral pH increases the potential prebiotic environments in which ribozymes could function.

Future work will attempt to determine the mechanism in which peptide 4 aids the activity of ribozyme S2. The SHAPE data and kinetics determination seem to indicate that peptide 4 guides ribozyme S2 to the active conformation. The RNA may require some time to bind the peptide and then to fold into the correct conformation; this may explain why the kinetics time course shows a lag period between 0-2 hours. The SHAPE data indicates differential base accessibility depending on the presence of

peptide. Additionally, the finding that the presence of peptide and the use of a pre-incubation step result in similar SHAPE profiles supports the idea that the peptide is helping the RNA to fold. However, key differences in the flexibility of certain regions may indicate important sites that allow accessibility to the active conformation. The positions 5-11, 38-42, and 95-99 are less flexible in the presence of peptide 4 and are more flexible when only a pre-incubation is used. Since the presence of peptide 4 dramatically increases catalytic activity, these regions are likely constrained in the active conformation. In addition, the peptide benefit is attenuated when higher concentrations of Mg^{2+} are used in the self-triphosphorylation reaction. Mg^{2+} aids in RNA folding, so high concentrations of Mg^{2+} may compensate for the lack of peptide if the peptide is indeed helping the ribozyme to fold.

The sequence and stereo specificity of peptide 4 will be teased apart in future studies. Since both L and D amino acids were used in this study and since peptide 4 has 4 amino acids with stereocenters, there are a total of 16 unique peptides that could be responsible for improved ribozyme S2 activity. It is unclear which amino acids are interacting with ribozyme S2. A series of peptide variants and truncations will be synthesized to determine which residues are needed to aid in ribozyme activity. Peptide 4 contains both aspartic acid and serine residues, and these two residues may be crucial to the observed increase in ribozyme activity because serine is known to hydrogen bond with nucleobases[21] and aspartic acid can form a salt bridge using Mg^{2+} to the phosphodiester backbone of RNA[22]. A series of peptide variants and truncations will be synthesized to identify which residues and which stereoisomers contribute to the peptide dependent ribozyme activity observed in this study.

6.5 References

1. Sabeti, P. C., Unrau, P. J. & Bartel, D. P. Accessing rare activities from random RNA sequences: the importance of the length of molecules in the starting pool. *Chem. Biol.* **4**, 767–774 (1997).
2. Seelig, B. & Jaschke, A. A small catalytic RNA motif with Diels-Alderase activity. *Chem. Biol.* **6**, 167–176 (1999).
3. Ryckelynck, M., Baudrey, S., Rick, C., Marin, A., Coldren, F., Westhof, E. & Griffiths, A. D. Using droplet-based microfluidics to improve the catalytic properties of RNA under multiple-turnover conditions. *RNA* **21**, 458–469 (2015).
4. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857 (1983).
5. Milligan, J. F. & Uhlenbeck, O. C. [5] Synthesis of small RNAs using T7 RNA polymerase. in *Methods in Enzymology* vol. 180 51–62 (Academic Press, 1989).
6. Triana-Alonso, F. J., Dabrowski, M., Wadzack, J. & Nierhaus, K. H. Self-coded 3′-Extension of Run-off Transcripts Produces Aberrant Products during in Vitro Transcription with T7 RNA Polymerase (∗). *J. Biol. Chem.* **270**, 6298–6307 (1995).
7. Zaher, H. S. & Unrau, P. J. T7 RNA Polymerase Mediates Fast Promoter-Independent Extension of Unstable Nucleic Acid Complexes. *Biochemistry* **43**, 7873–7880 (2004).

8. Weinberg, C. E., Weinberg, Z. & Hammann, C. Novel ribozymes: discovery, catalytic mechanisms, and the quest to understand biological function. *Nucleic Acids Res.* **47**, 9480–9494 (2019).
9. Parker, D. J., Xiao, Y., Aguilar, J. M. & Silverman, S. K. DNA Catalysis of a Normally Disfavored RNA Hydrolysis Reaction. *J. Am. Chem. Soc.* **135**, 8472–8475 (2013).
10. Suzuki, T. The expanding world of tRNA modifications and their disease relevance. *Nat. Rev. Mol. Cell Biol.* **22**, 375–392 (2021).
11. Tuorto, F., Liebers, R., Musch, T., Schaefer, M., Hofmann, S., Kellner, S., Frye, M., Helm, M., Stoecklin, G. & Lyko, F. RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nat. Struct. Mol. Biol.* **19**, 900–905 (2012).
12. Hori, H. Methylated nucleosides in tRNA and tRNA methyltransferases. *Front. Genet.* **5**, (2014).
13. Struck, A.-W., Thompson, M. L., Wong, L. S. & Micklefield, J. S-Adenosyl-Methionine-Dependent Methyltransferases: Highly Versatile Enzymes in Biocatalysis, Biosynthesis and Other Biotechnological Applications. *ChemBioChem* **13**, 2642–2655 (2012).
14. Scheitl, C. P. M., Ghaem Maghami, M., Lenz, A.-K. & Höbartner, C. Site-specific RNA methylation by a methyltransferase ribozyme. *Nature* **587**, 663–667 (2020).
15. Flemmich, L., Heel, S., Moreno, S., Breuker, K. & Micura, R. A natural riboswitch scaffold with self-methylation activity. *Nat. Commun.* **12**, 3877 (2021).

16. Jiang, H., Gao, Y., Zhang, L., Chen, D., Gan, J. & Murchie, A. I. H. The identification and characterization of a selected SAM-dependent methyltransferase ribozyme that is present in natural sequences. *Nat. Catal.* **4**, 872–881 (2021).
17. Takusagawa, F., Kamitori, S., Misaki, S. & Markham, G. D. Crystal Structure of S-Adenosylmethionine Synthetase. *J. Biol. Chem.* **271**, 136–147 (1996).
18. Morse, J. W. & Mackenzie, F. T. Hadean Ocean Carbonate Geochemistry. *Aquat. Geochem.* **4**, 301–319 (1998).
19. Halevy, I. & Bachan, A. The geologic history of seawater pH. *Science* **355**, 1069–1071 (2017).
20. Krissansen-Totton, J., Arney, G. N. & Catling, D. C. Constraining the climate and ocean pH of the early Earth with a geological carbon cycle model. *Proc. Natl. Acad. Sci.* **115**, 4105–4110 (2018).
21. Helene, C., Maurizot, J.-C. & Wagner, K. G. Interactions of Oligopeptides With Nucleic Acid. *Crit. Rev. Biochem.* **10**, 213–258 (1981).
22. Giacobelli, V. G., Fujishima, K., Lepšík, M., Tretyachenko, V., Kadavá, T., Makarov, M., Bednárová, L., Novák, P. & Hlouchová, K. In Vitro Evolution Reveals Noncationic Protein–RNA Interaction Mediated by Metal Ions. *Mol. Biol. Evol.* **39**, msac032 (2022).