Towards a Prebiotically Plausible Mechanism for the Emergence of RNA, Ribozymes and the RNA World

by

Lyssa Lynn Martin

B.Sc., Thompson Rivers University, 2011

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Molecular Biology and Biochemistry
Faculty of Science

SIMON FRASER UNIVERSITY

Summer 2015

Approval

Name:	**Lyssa Lynn Martin**
Degree:	**Master of Science**
Title:	***Towards a Prebiotically Plausible Mechanism for the Emergence of RNA, Ribozymes and the RNA World***
Examining Committee:	**Chair:** Dr. Fredric Pio Associate Professor

Peter Unrau
Senior Supervisor
Professor

Dipankar Sen
Supervisor
Professor

Niles Lehman
Supervisor
Professor
Portland State University
Department of Chemistry

Nancy Forde
Internal Examiner
Associate Professor
Department of Physics

Date Defended/Approved: August 25, 2015

Abstract

Available evidence, both *in vitro* and *in vivo,* attests to the descent of life from the RNA World; however, the prebiotic genesis of such RNA life remains ambiguous. How did The RNA World emerge from the abiotic chemistry on the Archean Earth? Montmorillonite clays have been shown to catalyze the polymerization of activated nucleotides (eg. adenosine 5′ phosphorimidazolide) into RNA, but polymerization has not previously been demonstrated for a prebiotically plausible nucleotide such as cyclic 2′, 3′-adenosine monophosphate (A>p).

I reacted A>p in the presence of montmorillonite clay and could detect RNA polymers up to 5 nucleotides in length using a combination of HPLC, enzymatic labeling and mass spectrometry. This chemistry was found to be sensitive to pH and temperature. Reactions at pH 6 were found to produce more polymerization products than reactions at pH 7 or 8. Similarly A>p was found to be very stable at pH 6 with a hydrolysis rate at 25ºC of 6.2×10^{-9} sec^{-1} ($t_{1/2}$ = 3.5 years) and at 4ºC of 4.4×10^{-10} sec^{-1} ($t_{1/2}$ = 50 years).

Also discussed are the requirements for the transition from abiotic chemistry to life. Important considerations for the emergence of replicating RNA networks from randomly synthesized RNA are outlined. Finally, the progress towards recreating the RNA world in the laboratory is summarized.

Keywords: RNA, RNA World Hypothesis; abiogenesis; origin of life; nucleotide chemistry; montmorillonite clay

For Sean, the light of my life and my partner for all adventure. For my parents, who gave me this life and whose love, dedication and wisdom I am grateful for beyond measure. For my family and friends, without whom I would surely have lost my sanity. And for my mentors past and present, who make me better by challenging me, even though I rarely appreciate it at the time.

Acknowledgements

Thank you Sean, Mom, Dad and family for your unwavering love and support, I would be lost without you. I would like to thank Dr. Peter Unrau and the members of the Unrau lab: Marianna, Lena, Neil, Shyam, Sunny, Razvan, Matt, Tim, Irene, Vy, Stephanie, Bashe, Ibrahim, and Amir for their support and comradery. Thank you to my committee members Dr. Dipankar Sen and Dr. Niles Lehman, and also to Dr. Clint Spigel for their help feedback and suggestions. I would also like to thank the American Colloid Company for their gift of clay samples and Dr. James Ferris for advice in preparing those samples. Thank you to all of the MBB grad caucus members: Kaylee, Even, Vil, Laura, Jesper, Julie R, Dipa, Sharron, Aarti, Mike, Bruno, Thea, Julie S, Saeideh, Zoreh, Razvan and Navi for all the good times and for providing a safe haven to hide from experiments that just won't work. A very special thank you to Evan and Katie Quon for providing power and wi-fi the weekend before this thesis was due, after half the city lost electricity for days. Thank you to the Sen, Cornell, Beh and Hawkins labs; the staff in the MBB office and anyone else I may have missed.

Table of Contents

List of Tables

List of Figures

List of Acronyms

2′AMP	2′ Adenosine monophosphate
3′AMP	3′ Adenosine monophosphate
A>p	2′, 3′ cyclic adenosine monophosphate
ATP	Adenosine triphosphate
C>p	2′, 3′ cyclic cytosine monophosphate
CMP	Cytosine monophosphate
DNA	Deoxyribonuclic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FAD	Flavin adenine dinucleotide
GNA	Glycol nucleic acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High performance liquid chromatography
ImpA	Adenosine 5′ phosphorimidazolide
ImpN	Nucleotide 5′ phosphorimidazolide
LN_2	Liquid nitrogen
MALDI-TOF	Matrix assisted laser desorption ionization- Time of flight mass spectroscopy
MeImpN	Adenosine 5′ phosphormethylimidazolide
MES	2-(N-morpholino)ethanesulfonic acid
NADH	Nicotinamide adenine dinucleotide
NMP	Nucleotide monophosphate
NTP	Nucleotide triphosphate
PAGE	Polyacrylamide gel electrophoresis
PNA	Peptide nucleic acid
PRPP	5-phosphoribosyl 1-pyrophosphate
R5P	Ribose 5-phosphate
RNA	Ribonucleic acid
SAM	S-Adenosyl methionine
T4 PNK	T4 Polynucleotide kinase
TCA	Tricarboxylic acid

TEAA	Triethyammonium acetate
TLC	Thin layer chromatography
Tmp	Trimetaphosphate
TNA	Threose nucleic acid
TRIS	Tris(hydroxymethyl)aminomethane
U>p	2′, 3′ cyclic uridine monophosphate
UMP	Uridine monophosphate
WT	Wild type

Chapter 1. The Origin of Life

1.1. The RNA World Hypothesis: a biochemical theory of our origins

Abiogenesis is the process by which non-living matter becomes a living system. This process exists at the intersection of organic, bio- and geo- chemistry. Life on Earth emerged though abiogenesis between 3.8 and 4.1 billion years ago (Oberbeck and Fogleman 1989), and remains the only known example of life in the universe.

The RNA World hypothesis is a leading theory for explaining the origin of life on Earth (Crick 1958) (Gilbert 1986). According to the RNA World hypothesis the origin of life occurred via the emergence of self-replicating RNA sequences. These RNA sequences evolved into protocells and eventually into the DNA/RNA/protein life of the modern world. This robust theory stands on several key pieces of evidence. Probably the strongest evidence of all is the dual ability of RNA to store genetic information and act as a catalyst by folding into complex secondary structures. The combination of these features in a single molecule represents the simplest possible link between code and function, between a mutation and natural selection, making evolution possible. This ability to inherit mutations that alter phenotype makes RNA an ideal candidate to assemble the first living thing.

Most evidence for the RNA World hypothesis comes from the shared biochemistry of all Earth′s organisms (Robertson and Joyce 2012). When a molecule or pathway is found in all living organisms it indicates that molecule or pathway originated before the last universal common ancestor, since the independent emergence of a pathway in all organisms is vastly unlikely. These ancient clues to the earliest lifeforms are known as “molecular fossils” because like fossils they are snapshots of the past. This is especially true for the most central and important pathways such as protein synthesis or the tricarboxylic acid (TCA) cycle, because they are under the most selection pressure to remain unchanged. Interestingly molecular fossils of our deepest and most highly conserved biochemistry provide evidence of decent from RNA world organisms.

Proteins do nearly all catalysis in an organism but conspicuously it is RNA not proteins that are the most important players in protein synthesis (Cech 2000). It is mRNA which carries the code that is translated into protein; while tRNA carries the anti-codon with the corresponding amino acid. Most importantly the rRNA of the ribosome recognizes the codon-anticodon pairing and catalyzes the peptidyl transferase reaction. More simply stated: the ribosome catalyzes protein synthesis (Cech 2000); thus, the ribosome is an RNA enzyme or “ribozyme”. If all protein enzymes in all organisms are made by one universal ribozyme, then logically that ribozyme must predate protein enzymes. Similarly bacterial RNase P is a highly conserved ribozyme that is essential for the processing of tRNA (Guerrier-Takada et al. 1983), without which RNA dependent protein synthesis could not proceed.

It is difficult to replace a molecule that is used as a common substrate by many essential enzymes because the replacement molecule must be compatible with all of the enzymes that require it. The more enzymes that require the molecule, the lower the probability that a substitute will be compatible with all of them. Disrupting even one of the pathways could be catastrophic, therefore there is strong selection pressure to retain the original molecule. Most coenzymes and cofactors such as Nicotinamide adenine dinucleotide (NAD^+), Flavin adenine dinucleotide (FAD), co-enzyme A and S-Adenosyl methionine (SAM), are nucleotide derivatives (White 1976) (Figure 1.1). They are essential for thousands of metabolic reactions, most prevalently in highly conserved pathways such as the TCA cycle. The fact that these highly utilized cofactors are derived from RNA indicates that there was a time when RNA dominated metabolism, these cofactors remained frozen while proteins took over metabolism one reaction at a time.

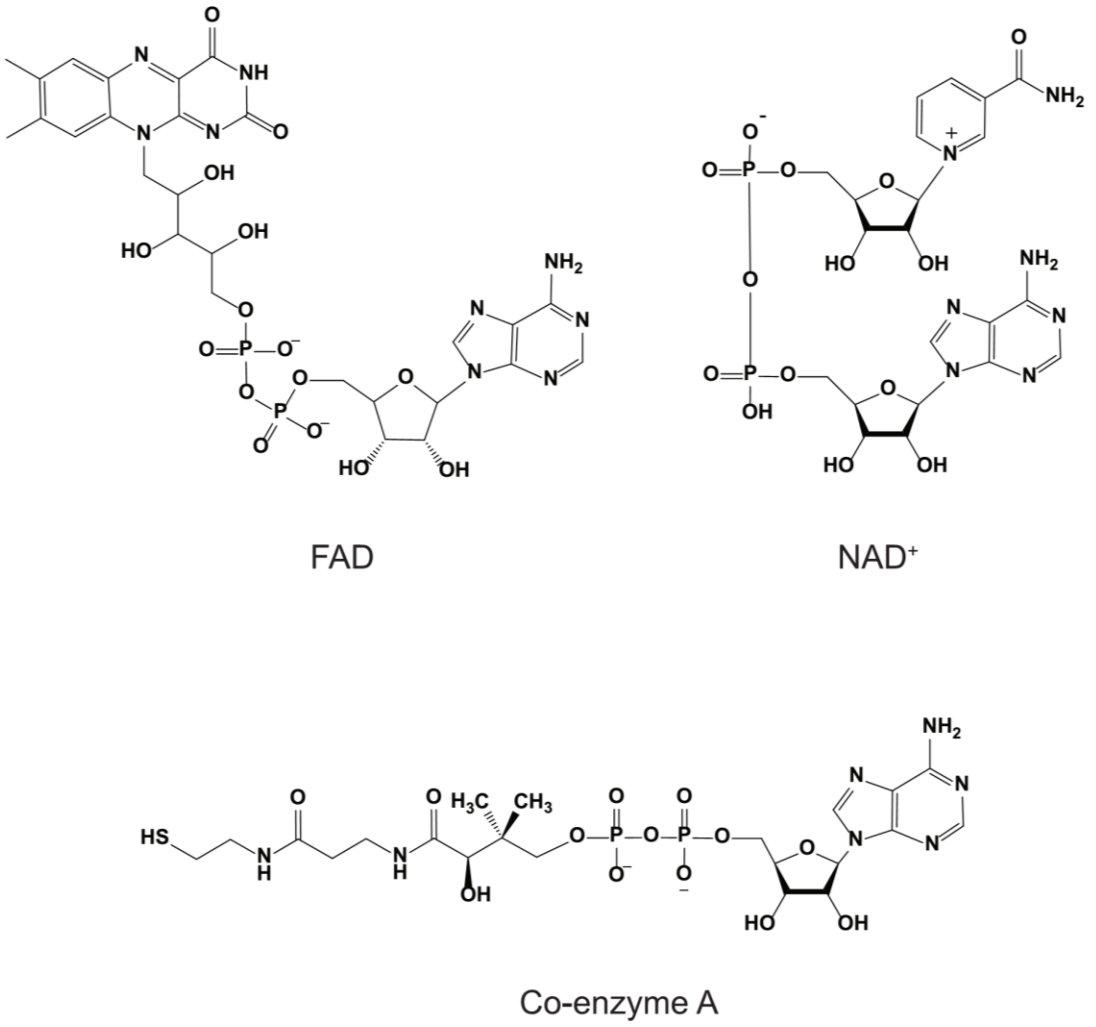

Figure 1.1. Examples of cofactors that are nucleotide derived.

Not only is there evidence that RNA predates proteins, there is also evidence that RNA is more ancient than DNA. DNA synthesis, for both deoxynucleotides and DNA polymers, in today′s organisms proceeds via RNA intermediates (Sprengel and Follmann 1981). In DNA bases the 2′-hydroxyl group is removed from nucleoside triphosphates to generate 2′-deoxynucleotide triphosphates, and the C5-methyl group is attached to 2′-deoxyuridine triphosphate to generate 2′-deoxythymidine triphosphate (REICHARD

1955). Further, nearly all extant DNA polymerases require the synthesis of RNA primers by DNA primase to commence strand synthesis during DNA replication (Eliasson and Reichard 1978), for the lagging strand multiple RNA primers are used. This biochemical evidence supports the idea that RNA is evolutionarily older than DNA.

While only a limited number of ribozymes still exist in modern organisms, synthetic ribozymes have been developed through *in vitro* selection that can perform diverse chemistries such as ligation (Bartel and Szostak 1993a; Ekland and Bartel 1995; Rogers and Joyce 2001a; Paul and Joyce 2002; Levy et al. 2005; Seelig and Szostak 2007; Voytek and Joyce 2007), recombination (Hayden and Lehman 2006), polymerization of RNA monomers (Ekland and Bartel 1996; Johnston et al. 2001a; Lawrence and Bartel 2005; Zaher and Unrau 2007; Attwater et al. 2010; Wochner et al. 2011; Wang et al. 2011), nucleotide synthasis (Unrau and Bartel 1998; Chapple et al. 2003b; Lau et al. 2004a; Lau and Unrau 2009), Diels-Alder reactions (Seelig and Jäschke 1999), decarboxylation (Cernak and Sen 2013), and amino acid transfer (Lohse and Szostak 1996). It seems that the possibilities of ribozyme function are only limited by the imaginations of the scientists who develop them; this tells us that a complicated RNA driven metabolism could certainly have existed.

1.2. Arguments against alternative hypotheses

There are other theories regarding the origin of life on earth, many of which do not provide a plausible mechanism for heritability and therefore could not have been the origin of modern life. To be clear, in order to be the origin of life on Earth a system must have been able to evolve into life as we know it, there must have been information storage, and that information must have been passed forward to progeny.

In “metabolism first” models, a chemical cycle (eg. TCA cycle) develops spontaneously in the ancient abiotic environment (Martin and Russell 2007) and the emergence of this cycle is called “the origin of life”. While these early chemical cycles certainly existed, and likely played a role in the evolution of life (Anet 2004; Chen et al. 2007), they themselves do not represent the origin of life because they do not contain a heritable code that is transmitted to the next generation. There are countless examples of

chemical reactions or cycles existing in the absence of biologic material, but they are not considered to be alive. Without reproduction chemistry is not life, no matter how complex.

The pre-RNA World hypothesis postulates that some alternate nucleic acid was the predecessor to RNA, and that RNA was a later invention of life-forms based on an alternate genetic system. An alternate genetic system having greater thermo-stability or monomers being easier to synthesize (Orgel 2004) are the most common reasons for evoking the pre-RNA world. Candidates include peptide nucleic acid (PNA), threose nucleic acid (TNA), and glycol nucleic acid (GNA), however there is no evidence that these molecules ever existed in nature let alone predated RNA in metabolism.

TNA is very similar to RNA, the difference being that the four-carbon sugar threose is used, in contrast to the five-carbon ribose sugar of RNA; for this reason it is thought that TNA may have been easier to synthesize on the prebiotic Earth (Robertson and Joyce 2012). It is known that TNA can fold into aptamers and also base pair with RNA and DNA. It was hypothesized that single stranded TNA would make a better template for the polymerization of nucleotides than RNA because the structure of the molecule is pre-organized for favourable nucleotide binding and extension (Adamala and Szostak 2013). This is because TNA residues adopt a unique 4′-exo sugar pucker with the 2′- and 3′-hydroxyl groups axial, which maximizes the interphosphate distance. However, contrary to expectation, it was found that TNA was a poor system for non-enzymatic templated polymerization compared to both RNA and DNA; the inferiority of TNA to RNA in templated polymerization may explain why RNA but not TNA lead to the emergence of life even though both RNA and TNA monomers may have been produced through prebiotic chemistry.

PNA is based on a backbone of repeating N-(2-amnioethyl)-glycine units linked by peptide bonds (Nielsen et al. 1991); as such the name peptide nucleic acid is actually a misnomer since it is not an acid. Like other nucleic acids PNA has been used to make aptamers. Unlike nucleic acids with sugar-phosphate backbones PNA does not have a negative charge. While PNA is more stable than RNA it is also much more difficult to melt a PNA duplex than an RNA duplex due to the lack of electrostatic repulsion. Without strand

displacement copying of PNA for replication would be difficult: this is major point against PNA as a pre-RNA genetic system.

GNA is another nucleic acid; it is even simpler than TNA, since it is built from three-carbon glycol units with phosphodiester bonds. For that reason it is appealing as an alternate genetic system to RNA. It has been shown to be more thermostable than RNA. However, like PNA it is more difficult to melt a GNA duplex, so strand displacement for abiotic replication may not be possible. Further, GNA is only able to form a duplex with RNA if there are no G:C bonds present, making it difficult to conceive a mechanism for transition from a GNA world to the RNA world.

The need for more thermostable nucleic acids such as PNA or GNA is probably overstated; though there is some evidence of hot oceans, early climate models vary (Kasting and Ono 2006; Wolf and Toon 2013). Like today′s Earth surface temperature would not have been uniform, thus colder regions certainly existed probably near the poles. Further, the young sun could only output 70% of the energy that it does today, providing significantly less solar energy to warm the earth. The oceans would have been acidic, stabilizing to any RNA polymers present at the time (Coooper et al. 1986). Also recent work has shown that formation of active nucleotides is not as difficult under prebiotic conditions as once thought (Powner et al. 2009a; Powner et al. 2010). Together the available evidence leaves the pre-RNA world hypotheses unnecessary.

The Panspermia hypothesis is theory that life originated elsewhere in the universe, and was spread here by micro-organisms in stellar dust, meteors or comets (Wickramasinghe 2004). Panspermia does not directly address the origin of life, but instead moves it elsewhere. As such it is not incompatible with the RNA World hypothesis, but until evidence is found that demonstrates space-born organisms did seed our world with life, most origins scientists will continue to use the conditions of the primordial Earth as a starting place.

1.3. Before RNA: the prebiotically plausible synthesis of nucleotides

In order for the RNA world to have emerged from the prebiotic Earth a chemical pathway must have existed that synthesised RNA monomers (Figure 1.2). Those monomers then were polymerized abiotically into RNA either by continuous polymerization or by the synthesis of short fragments that were then ligated. Only a fraction of these RNA polymers were functional ribozymes, and only a fraction of those could have catalyzed a reaction that promoted the formation of more ribozymes (nucleotide synthesis, polymerization, ligation or recombination). Nonetheless if a ribozyme emerged that catalyzed any of these reactions it would have pushed the system towards the formation of more ribozymes, which in turn were potentially able to catalyse more of the key reactions. Over time this would have led to the take-over of RNA synthesis metabolism from geo-chemistry by RNA catalysts and the establishment of a far from equilibrium system: life.

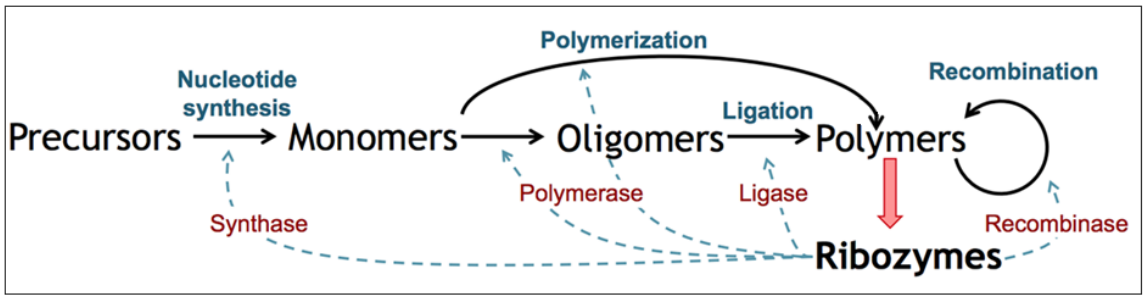

Figure 1.2. Route from abiotic to living system.
A non-living processes leads to the emergence of ribozymes. If a ribozyme emerges that can catalyze any of the reactions that lead to its synthesis it increases the probability of the emergence of more ribozymes. The ribozymes can begin to evolve and transition to a living state.

Modern organisms use nucleotide tri-phosphates (NTPs) as the monomer for RNA synthesis (Figure 1.3). They have a high activation barrier for hydrolysis and high free energy available to an organism with highly developed catalysts (Acevedo and Orgel 1987). But for the same reasons they are unlikely to have been the first RNA monomer

because they are not suited to polymerization in the absence of specific catalysts. For decades plausible prebiotically relevant nucleotides remained elusive, so work on abiotic RNA synthesis used artificially activated nucleotides such as 5′-phosphorimidazolides (ImpN) (Ferris and Ertem 1992; Ferris and Ertem 1993; Monnard and Deamer 2003; Huang and Ferris 2006; Miyakawa et al. 2006; Adamala and Szostak 2013). These ImpN nucleotides are not considered prebiotically relevant, but have a low activation barrier so they react readily and are useful for investigating abiotic polymerization. Recently 2′,3′ cyclic cytosine (C>p) and 2′,3′ cyclic uridine (U>p) were synthesized in prebiotically plausible conditions (Powner et al. 2009), making 2′,3′ cyclic nucleotide monophosphates (N>p) excellent candidates for the first RNA monomers.

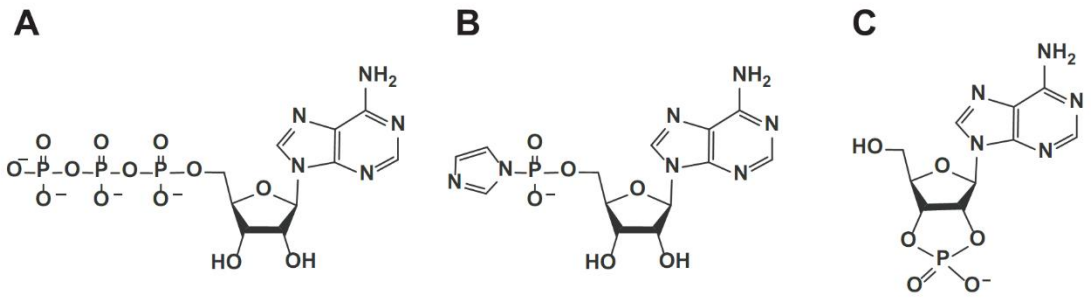

Figure 1.3. Comparison of nucleotide structure: A) ATP, B) ImpA, and C) A>p.

The synthesis of C>p and U>p by Powner et al. (2009) took a different approach to nucleotide synthesis than others had tried. Previous attempts at plausible aqueous nucleotide synthesis attempted to synthesize the base and sugar separately, then fuse them together. This approach was unsuccessful because the fusion step suffers from very low yields. This new strategy bypassed the base to sugar fusion step, instead synthesizing the base on the sugar by proceeding through arabinose amino-oxazoline (12) and anhydronucleoside (13) intermediates (Figure 1.4). The resulting C>p can be converted to U>p by irradiation with UV light at 254 nm, demonstrating plausible prebiotic synthesis of two of the four standard RNA nucleotides. Further, this group also demonstrated the

synthesis of plausible precursors to 2′,3′ cyclic purines under similar conditions to the C>p and U>p synthesis (Powner et al. 2010). This is exciting because it hints at a common mechanism for the generation of all four nucleotides from the same starting material. Similarly it was shown that from these starting materials not only are nucleotides generated but also amino acids and lipid precursors (Patel et al. 2015).

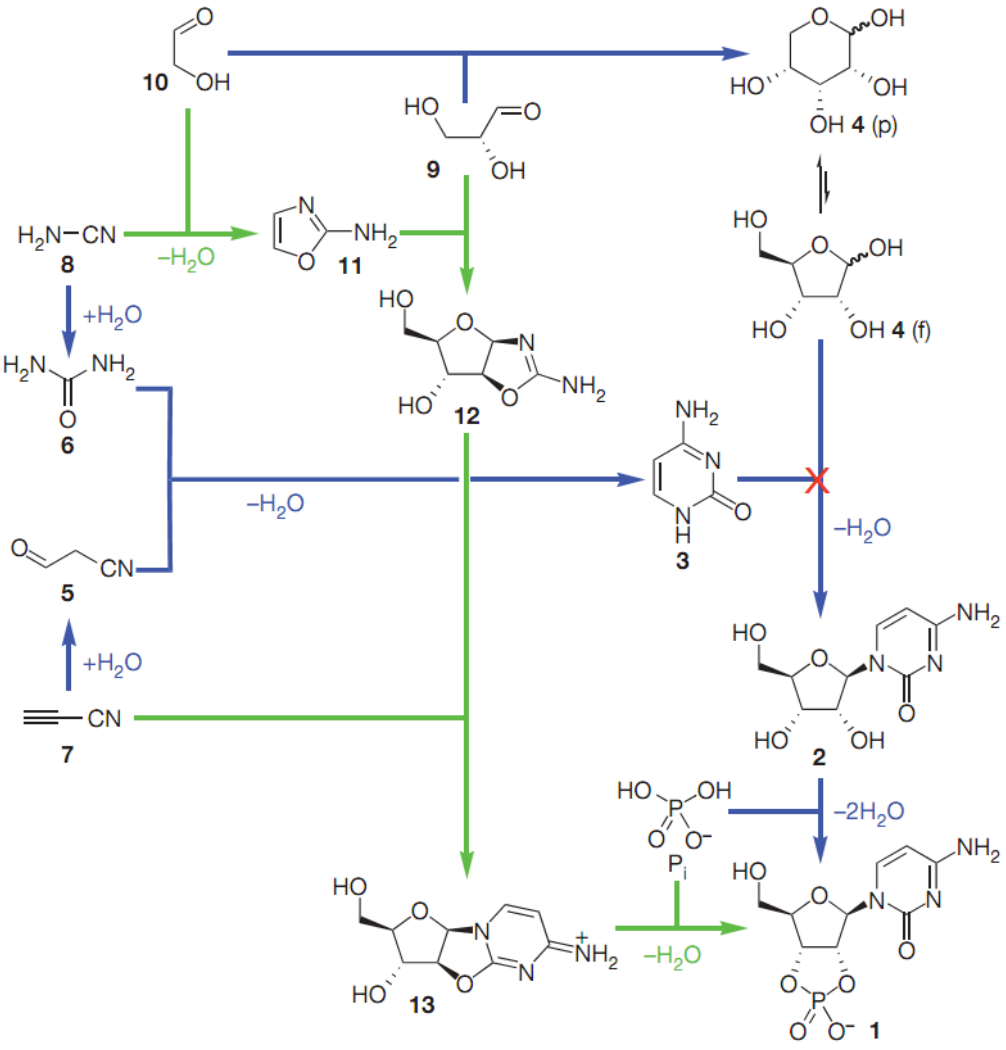

Figure 1.4. Prebiotically plausible synthesis of a pyrimidine ribonucleotide.
Previously assumed synthesis of beta-ribocytidine-2′,3′-cyclic phosphate 1 (blue; note the failure of the step in which cytosine 3 and ribose 4 are proposed to condense together) and the successful new synthesis described here (green). p, pyranose; f, furanose. Reprinted by permission from Macmillan Publishers Ltd: [Nature] Powner et al. (2009).

Chapter 2. The Origin of RNA polymers

2.1. From nucleotides to RNA: the role of mineral surfaces

It has been well established in the literature that RNA polymers as long as 50 nucleotides can be formed from active monomers under several different conditions; many of these rely on mineral catalysts (Ferris and Ertem 1992; Ferris and Ertem 1993; Ferris 2002; Monnard and Deamer 2003; Huang and Ferris 2006; Miyakawa et al. 2006; Adamala and Szostak 2013. Existing RNA templates bound on hydroxyapatite (Acevedo and Orgel 1986) can serve as a template for 5′-phosphormethylimidazole activated nucleotides (MeImpN). Similarly, ImpN will polymerize on an RNA template in eutectic ice in the presence of Mg^{2+} (Monnard and Deamer 2003). 5′- phosphorimidizolides (ImpN) can condense into RNA polymers (with mixed 2′-5′ and 3′-5′ linkages) without a template on montmorillonite clays (Ferris and Ertem 1992; Ferris 2002; Miyakawa et al. 2006), representing the only known mechanism for prebiotically plausible *de novo* synthesis of long RNA oligomers. These imidazole-activated nucleotides are not thought to be prebiotically relevant since they have not been demonstrated to be generated under plausible prebiotic conditions.

Montmorillonite clay is a 2:1 phyllosilicate mineral formed from the weathering of volcanic rock and as such would have been common on the prebiotic Earth (Meunier et al. 2010). Each layer is built from a planar sheet of aluminum oxide octahedrons, coated on both sides with a layer of silicon hydroxide tetrahedrons (Figure 2.1). Between each phyllosilicate layer is a layer of water and exchangeable cations; the spacing between the layers is large, 1-2 nm or more allowing organic compounds such as nucleotides to enter and bind. Montmorillonite is used as an industrial catalyst for the production of many organic chemicals (Bigi et al. 1999; Srinivas and Das 2003; Bahulayan et al. 2003).

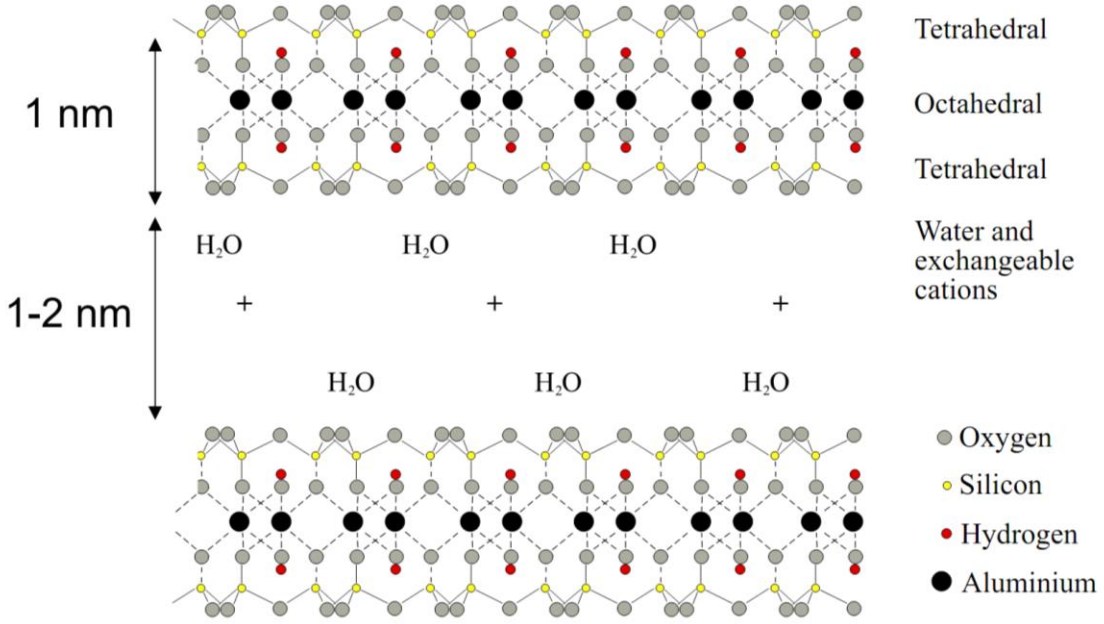

Figure 2.1. Montmorillonite clay is a phylosilicate mineral.
Each sheet has an octohedral layer of Aluminium hydroxide coated on each side with a layer of Silcon oxide tetrahedrons. Between each sheet is a layer of water and salts. The size of the water layer is such that nucleotides and other organics can enter. (Modified from Wikipedia)

The polymerization of ImpN in the presence of montmorillonite has been extensively studied, as such the optimal conditions for long RNA products have been identified (Miyakawa et al. 2006). Temperature was found to be inversely related to RNA length with 0.48% yield of 8 mers at 4° but only 0.15% yield of 8 mers at 25°C, presumably because hydrolysis of monomers is depressed at lower temperature. It was found that pH plays an important role in polymerization, with pH between 7-8 being optimal for the polymerization of ImpA. Lower pH promoted the formation of 5′-5′ dimers and no polymerization was observed at high pH. Salts were also found to play an important role, with small monovalent salts significantly promoting polymerization of long RNAs.

The polymerization of nucleotides into RNA by montmorillonite is a promising abiotic synthesis route; however, it has never been demonstrated with a prebiotically plausible nucleotides such as (A>p). In this work I explore the polymerization of A>p on montmorillonite clay and the implications of such synthesis for the RNA World Hypothesis.

2.2. Nomenclature of RNA oligomers

There are a variety of RNA oligomers in this study, some were used as controls and some that are formed in the polymerization experiments. Most control oligomers are poly(A) and as such are named as "A_n" where n is the number of nucleotides; these oligomers have 3′→5′ phosphodiester linkages and 2′, 3′ and 5′ hydroxyls. If an oligomer contains other nucleotides it will be named according to its sequence, for example "AGC".

Oligomers that are synthesised experimentally do not necessarily contain 3′→5′ linkages, there are other possibilities. The rules for labelling oligomers are: subscript gives oligomer length, while superscript gives the position of the phosphodiester (Figure 2.2). A dinucleotide with a 3′→5′ phosphodiester linkage and a 2′, 3′ and 5′ hydroxyl is named $A^{3'}A$, while the corresponding 2′→5′ species is named $A^{2'}A$; if the name A_2 is used it is denoting a mixture of 2′→5′ and 3′→5′ species or that it is not possible to distinguish between these two isomers. Similarly A_3 represents all or a subset of: $A^{2'}A^{2'}A$, $A^{2'}A^{3'}A$, $A^{3'}A^{2'}A$ and $A^{3'}A^{3'}A$. Polymerization products can also have phosphates on their 2′, 3′ and 5′ ends. A_2>p represents a dinucleotide with a terminal 2′,3′ cyclic phosphate and includes both $A^{2'}A$>p and $A^{3'}A$>p isomers. Likewise A_2p represents $A^{2'}A^{2'}$p, $A^{2'}A^{3'}$p, $A^{3'}A^{2'}$p and $A^{3'}A^{3'}$p; while $A_2{}^{3'}$p only represents dinucleotides with a 3′ phosphate: $A^{2'}A^{3'}$p and $A^{3'}A^{3'}$p. A dinucleotides with a 5′ phosphate but hydroxyls at the 2′ and 3′ position would be called pA_2, unless the phosphodiester linkage was also know then it could be written as either $pA^{2'}A$ or $pA^{3'}A$. Heterocyclic species are also possible, in a dinucleotide this would be a mutual phosphodiester link between the 5′ of each nucleotide and the 2′ or 3′ positions of the other nucleotide. To complicate matters it is also possible to generate 2′→2′, 2′→3′ and 3′→3′ diphosphate linkages. Dinucleotides of this type are named as $A^{2'\text{-}2'}A$, $A^{2\text{-}'3'}A$ and $A^{3'\text{-}3'}A$ respectively.

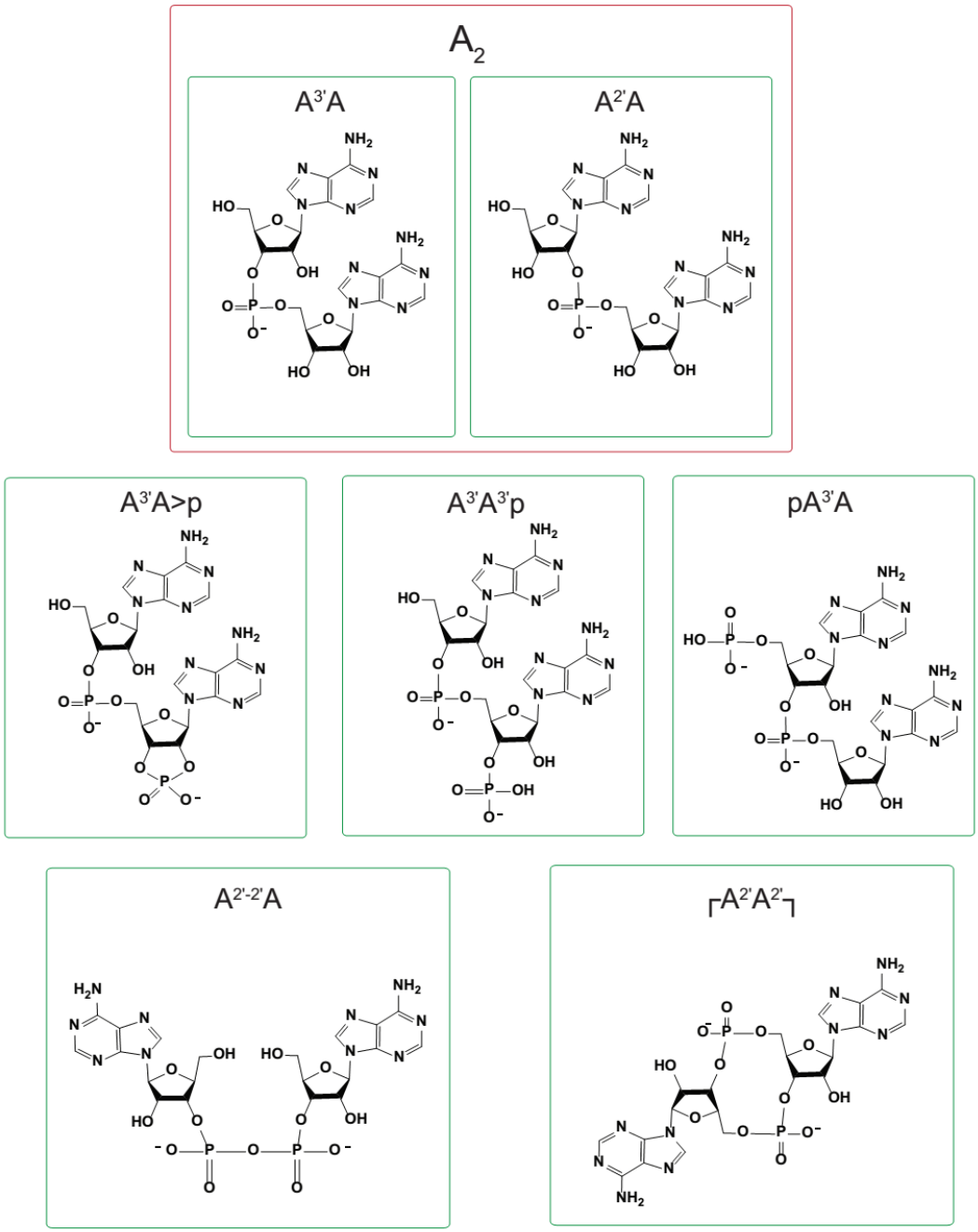

Figure 2.2. Examples of RNA nomenclature.
A_2 can indicate a mixture of $A^{3'}A$ and $A^{2'}A$ or that it is not possible to distinguish the two. $A^{3'}A>p$ denotes a dinucleotide with a 3′→5′ phosphodiester bond and a 2′, 3′ cyclic phosphate. $A^{3'}A^{3'}p$ has a 3′→5′ phosphodiester bond and a terminal 3′ phosphate. $pA^{3'}A$ has a 5′ phosphate and a 3′→5′ phosphodiester bond. $A^{2'-2'}A$ has a diphosphate linked to each nucleotide at the 2′ position. ┌$A^{2'}A^{2'}$┐ is a cyclic oligomer with two 2′→5′ bonds.

2.3. Selection of Clay

2.3.1. Introduction

Three samples of clay (WDG1, SPV, and S/C Yellow) were received from the American Colloid Company. There is evidence in the literature that there is variation in catalytic ability by montmorillonite clays (Joshi *et al.* 2009) and so the properties of these three clays were investigated to select the best clay for subsequent experiments.

2.3.2. Methods

Description of clays

Sodium montmorillonite clay samples were a gift from the American Colloid Company. Their elemental compositions are summarized in Table 2.1.

Table 2.1. Summary of composition for each clay as provided by the American Colloid Company.

	SPV	WDG1	S/C Yellow
Location	Wyoming, U.S.A		
Chemical Formula	$(Na,Ca)_{0.33}(Al_{1.67}Mg_{0.33})Si_4O_{10}(OH)_2 \cdot nH_2O$		
Main Trace Minerals	Feldspar, Calcite and Quartz		Feldspar, Calcite and Gypsum
SiO_2	63.02%	61.3%	63.02%
Al_2O3	21.08%	19.8%	21.08%
Fe_2O3	3.25%	3.9%	3.25%
FeO	0.35%	>0.1%	0.35%
MgO	2.67%	1.3%	2.67%
Na_2O	2.57%	2.2%	2.57%
CaO	0.65%	0.6%	0.65%
TiO_2	>0.1%	0.1%	>0.1%
K_2O	>0.1%	0.4%	>0.1%
Trace	0.72%	3.2%	0.72%

The S/C yellow clay was yellow, the SPV was a brownish grey and the WDG1 was a concrete grey colour. All three clays were prepared by the Banin saturation method (Banin et al. 1985) (Banin 1989). A 6 g aliquot of each clay was washed alternately with chilled 0.5 M HCl then chilled ddH_2O three times to exchange the clay's interlayer cations with H^+. This is followed by three washes with chilled ddH_2O to remove excess HCl. The pH of the clay solution was ~2 at this stage. The clay slurry was then split into three aliquots and titrated to pH 6, 7, or 8 with 1 M NaOH. The slurry was centrifuged, the supernatant was discarded and the clay pellet was dried under a vacuum. The dry clay was weighed into 10 mg aliquots.

Optical properties of clays

Clays were investigated to determine if clay would complicate the detection of RNA concentrations by absorbance. For each clay 10 mg aliquots were weighed out, then suspended in 100 µl of ddH_2O. The resulting clay slurries were pelleted by centrifugation (12,000 RCF, 2 min) and the supernatants were transferred to fresh tubes carefully, as to not disturb the clay pellet. The solutions were clear to the eye. On a nanodrop spectrophotometer each of the supernatants were scanned for absorbance between 220 and 340 nm.

Nucleic acid binding capacity

The Adenosine triphosphate (ATP) binding capacity of each clay was investigated. Sodium montmorillonite clay samples (WDG1, SPV, and S/C Yellow) were prepared as described above. An ATP stock was diluted in a solution containing 200 mM NaCl and 75 mM $MgCl_2$ in ddH_2O to give 135, 90, 67.5, 45, 22.5, and 11.25 mM ATP solutions. Further, a trace amount of [γ-^{32}P] ATP was added to each. A 200 µl aliquot of each ATP solution was added to 10 mg of pH 6 clay in a 1.7 ml eppendorf tube. The mixtures were vortexed to form a cloudy clay slurry and incubated overnight at room temperature. A control solution was also prepared in the absence of clay. The clay was pelleted by centrifugation at 12,000 RCF for 30 minutes.

Fifty microlitres of each supernatant and the no clay control was transferred to a scintillation counting vial. To each vial 2.5 ml of scintillation counting fluid (Amersham) was added; a blank vial with only scintillation fluid was prepared. The supernatants and

controls were counted on a Beckman LS6000SC scintillation counter for 1 minute. The fraction bound was calculated by subtracting the counts of the unbound control and blank (both had the same counts) from each supernatant. Data was averaged from three replicates.

Polymerization of ImpA

It has been reported that there is variability in the catalytic capacity of montmorillonite clays (Joshi et al. 2009). In order to asses which, if any, of our clays were catalytically active I performed polymerization experiments using a nucleotide that has been shown to polymerize in the presence of clay: ImpA.

ImpA was synthesized by dissolving 174 mg 5'-Adenosine monophosphate (5′AMP) free acid into 15 ml dimethylformamide to make a 0.5 mM 5′AMP solution. Next 262 mg triphenylphosphine, 220 mg of 2,2′-dithiodipyridine, and 170 mg of imidazole were dissolved into a mixture of 0.9 ml triethylamine and 15 ml dimethylformamide, with 1 mmol each of triphenylphosphine and 2,2′-dithiopyridine, and 2.5 mmol imidazole. To this the 5′AMP solution was slowly added while stirring for 1-1.5 hours at room temperature. The combined solution turned yellow-green. $NaClO_4$ (1102 mg) was dissolved into a solution of 60 ml triethyamine, 900 ml acetone and 450 ml ether and this was added dropwise while stirring vigorously. Stirring was ceased and the mixture was allowed to settle. The solvent phase was removed down to 60 ml volume. The remaining suspension was centrifuged in acetone twice, then once in ether. The ether centrifugation was performed in corex tubes at 5000 RCF for 20 min. The ImpA product was dried overnight in a vacuum. Analysis was performed on TLC cellulose plates with a saturated Ammonium sulphate solution as the mobile phase to check synthesis quality.

Polymerization reactions of ImpA were prepared. To each 10 mg aliquot of pH 8 clay 200 µl of polymerization solution (200 mM NaCl, 75 mM $MgCl_2$, 100 mM HEPES pH 8 and 10 mM ImpA) was added as described in (Joshi et al. 2009). The samples were vortexed into a slurry and incubated for 3 days at 22^{o}C. In addition 100 µl of polymerization solution was incubated in the absence of clay and 100 µl was storage at -22^{o}C as a control.

Reactions were centrifuged at 12,000 RCF for 30 min to pellet clay. Supernatants were removed and stored at -22°C. Each pellet was resuspended in 200 µl of Joshi et al.

elution solution (100 mM NaCl, 30% CH_3CN in ddH_2O). Elutions were twice incubated for 2 hr at room temperature, pelleted and replaced with fresh solution. A third elution was incubated overnight at room temperature; a fourth and final elution was incubated for 1 hr. The eluents were pooled after each collection.

The polymerization reactions were diluted to 100 µM and treated with FastAP thermosensitive alkaline phosphatase (Thermoscientific). Controls of unreacted ImpA, ImpA reacted without clay, 5′AMP, and ddH_2O were diluted to 100 µM and treated alongside polymerization reactions. A FastAP reaction contained 100 pmol ImpA, 50 mM Tris HCl (pH 7.6), 10 mM $MgCl_2$, 10 mM KCl and 0.5 units of FastAP enzyme in ddH_2O. The FastAP reactions were incubated at 37°C for 30 min then heat inactivated at 75°C for 10 minutes. Each FastAP reaction was adjusted to contain 100 pmol of ImpA, 50 mM Tris HCl (pH 7.6), 10 mM $MgCl_2$, 10 mM KCl, 10 mM Dithothreitol (DTT), 100 pmol of [γ-^{32}P] ATP and 5 units of T4 polynucleotide kinase (PNK; New England Biolabs) in 20 µl. The PNK reactions were incubated at 37°C for 30 minutes, then heat inactivated at 65°C for 25 minutes.

A urea-PAGE sequencing gel was prepared with 20% acrylamide in 1x TBE. PNK reactions were diluted 1:1 in 2x denaturing loading dye (0.025% (w/v) Xylene cyanol and 0.025% (w/v) Bromophenol blue in formamide) and heat denatured at 95°C for 5 min before loading. The gel was run in 1X TBE buffer at 72 W for approximately 2 hrs. The gels were exposed to phosphorimager plates (Fuji) and scanned using a Storm 825 phosphorimager (GE Healthcare). Gel images were analyzed using ImageQuantTL 8 (GE healthcare) to determine the intensities of each band.

2.3.3. Results

Optical properties of clay

The clear supernatants from the clays were found to absorb at 260 nm. WDG1 has the highest absorbance at 260 nm (7.5), followed by S/C yellow (3.5), and SPV (2.9). This background absorbance from the clays interfered in reading RNA recovery by absorbance on the Nanodrop spectrophotometer.

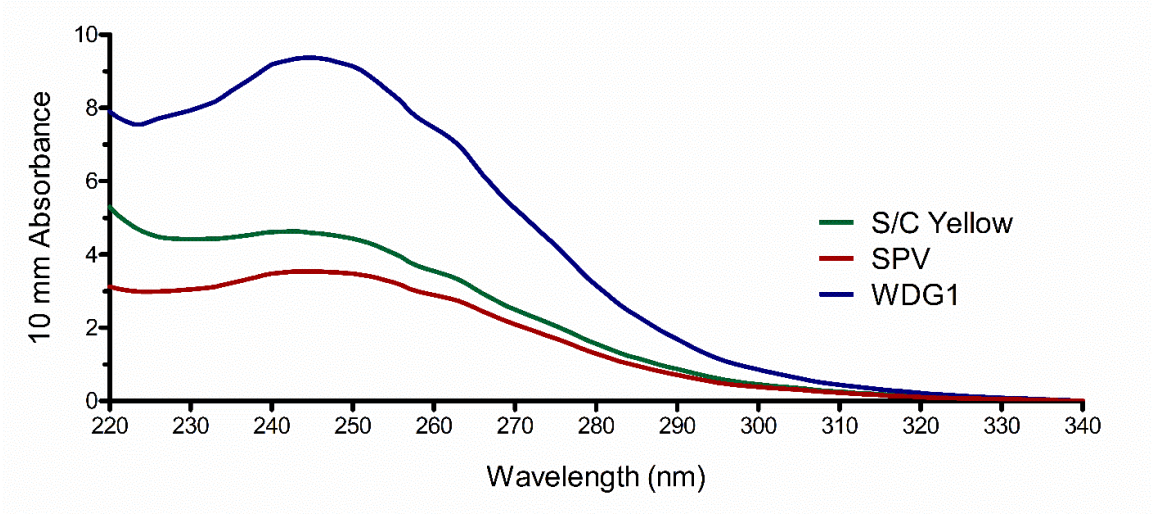

Figure 2.3. Absorbance of water supernatants from pelleted clay.
The solutions were clear to the eye but UV spectroscopy revealed that there is residual clay in the solutions. The absorbance of that clay at 260 nm could interfere with absorbance measurements of recovered polymerization reactions.

Nucleic acid binding capacity

It has been shown that strength of RNA binding and catalytic capacity of clay are correlated (Joshi et al. 2009). Our clays were investigated for strength of binding to ATP to determine which was the overall best binder to RNA. It was found that S/C yellow was the best ATP binder, saturating at 5.5 µmol of ATP bound to 10 mg of clay. SPV saturated at 5 µmol ATP bound per 10 mg of clay. WDG1 was found to be the poorest binder with saturation occurring at approximately 4 µmol of ATP bound to 10 mg of clay.

Polymerization of ImpA

ImpA was synthesised and found to be >95% pure by TLC. There was a small spot of low mobility material remaining where the ImpA was spotted.

ImpA was incubated for 3 days at room temperature in the presence and absence of clay. Recovered material was 5′ end labelled using PNK and analysed using PAGE (Figure 2.4). The A_2 dinucleotide band is present in samples but is ten times more intense in samples containing ImpA and twenty times as intense in samples containing clay relative to the 5′AMP control. It was not unexpected to find some dimer in the ImpA control samples since there was a small amount of immobile product on the TLC of the ImpA synthesis. The observation of increased A_2 intensity in the presence of clay is evidence for the clay catalysed formation of dinucleotide from ImpA. S/C yellow clay showed more A_2 formation than the other clays. Further S/C yellow showed higher intensity of upper bands in the gel including A_3.

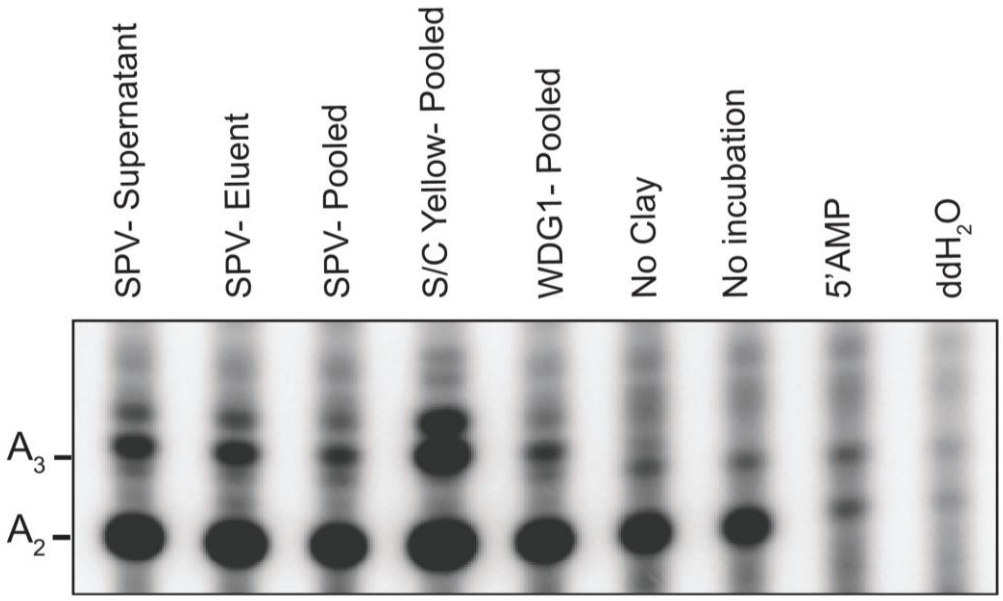

Figure 2.4. PAGE of the PNK labeled products from ImpA polymerization on clay.
The S/C yellow clay showed significantly more polymerization products than the other clays. All of the clays showed more products than the no clay control. Further the unincubated control indicates that there is a small amount of dinucleotide in the ImpA starting material.

From the above results I concluded that S/C yellow was the most catalytically active of the clays. Further, the ratio of the bands in the SPV supernatant and eluent were the same, suggesting that the supernatant and the eluent were exchanging readily. This lead me to believe that there was no difference between the distribution of products in the supernatant and eluent when I should have suspected a problem with the elution method. Later I show that the literature elution method of 100 mM NaCl in 30% CH_3CN was not removing bound RNA longer than three to four nucleotides in length.

The observed products are significantly shorter than the longest 50 nt polymers reported in the literature, however this is not unexpected. For one, to achieve the 50 nt length fresh ImpA was added multiple times (Ferris 2002), which I did not do. Also it is known that there is variability in the catalytic activity of clays (Joshi et al. 2009), the longest products are only detected on the most optimal clays. Finally even if longer RNAs were present here I would not have observed them because the elution method was ineffective (XX).

2.3.4. Selection of clay: S/C Yellow

The S/C yellow clay was selected for all subsequent experiments for a few reasons. First it was found that the WDG1 clay had about twice the absorbance of SPV and S/C yellow clay in pelleted solutions, ruling out WDG1 for potential interference with optical measurements. Second, S/C yellow was found to have the highest capacity for binding ATP, and a higher affinity for nucleic acid should increase the polymerization of nucleotides. Finally S/C yellow was found to be the best clay for the polymerization of ImpA, producing nearly twice as much A_2 and three times as much A_3 as the other clays.

2.4. Determination of A>p stability

2.4.1. Introduction

A>p can hydrolyse into either 2′AMP or 3′AMP (Figure 2.5). Both of these products can hydrolyze further to A. Hydrolysis of the A>p will affect polymerization reactions because the loss of the 2′,3′ cyclic phosphate is the loss of the activating group, though the 5′ hydroxyl of the hydrolysed products could still attack the active phosphate of an intact A>p.

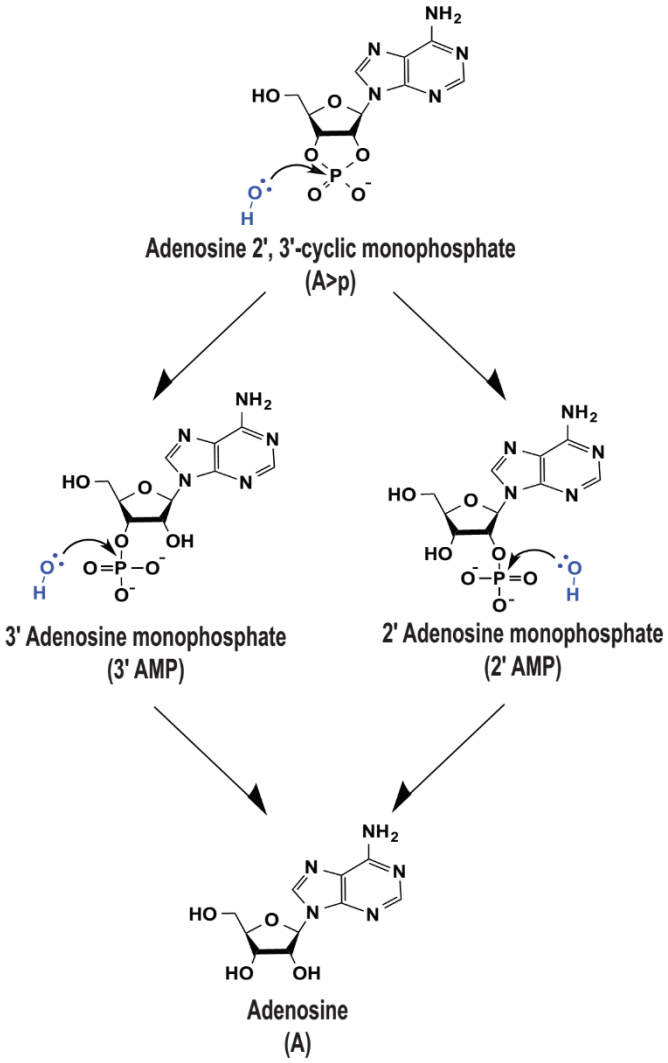

Figure 2.5. The hydrolysis of A>p. The first attack generates a 3′AMP or 2′AMP, the second attack generates A.

2.4.2. Methods

Determination of A>p pH stability

An A>p hydrolysis reaction was prepared to imitate the incubation conditions of polymerization experiments; it contained 10 mM A>p (Sigma-Aldrich), 250 mM NaCl, 75 mM $MgCl_2$, and 100 mM of a pH buffer (pH 6 MES, pH 7 HEPES, pH 8 HEPES, or pH 9 Tris). A hydrolysis reaction contained 100 µl of A>p hydrolysis mix and was incubated at 4ºC, time points were taken at 1, 3, 9 and 27 days of incubation. Time points were frozen in liquid nitrogen (LN_2) and stored at -80ºC until they were analysed by HPLC. An aliquot of unincubated input material was also frozen in LN_2 and stored at -80ºC before high performance liquid chromatography (HPLC) analysis.

A>p hydrolysis time points were diluted 1/5 in ddH_2O, and 20 µl of each dilution was injected into a C14 Bonus-RP (Agilent) column with a Triethylammonium acetate (TEAA) to CH_3CN gradient (0% TEAA at 0 min to 30% TEAA at 40 min). Absorbance at 260 nm was recorded. A standard solution of 2 mM A>p, 1 mM 2′AMP and 0.5 mM 3′AMP was also analyzed by HPLC to determine the mobility of the expected peaks.

Hydrolysis of A>p at pH 6

An A>p hydrolysis reaction was prepared containing 10 mM A>p, 100 mM of pH 6 MES and either 75 mM or 0 mM $MgCl_2$. A hydrolysis reaction contained 100 µl of A>p hydrolysis mix and was incubated at 95, 80, 72, 65 and 47º C without $MgCl_2$ and at 72, 65 and 47º C with $MgCl_2$. Time points varied by the temperature in the range of 15 min to 24 hours. Time points were frozen in LN_2 and stored at -80ºC until they were analysed by HPLC. An aliquot of unincubated input material was also frozen in LN_2 and stored at -80ºC before HPLC analysis.

A>p hydrolysis time points were diluted 1/5 in ddH_2O, and 20 µl of each dilution was injected into a new Zorbex-C18 (Agilent) column with a TEAA to CH_3CN gradient. Absorbance at 260 nm was recorded. A standard solution of 2 mM A>p, 1 mM 2′AMP, 0.5 mM 3′AMP and 0.25 mM A was also analyzed by HPLC to determine the mobility of the expected peaks.

The time points of hydrolysis at each temperature were analyzed by reverse phase HPLC using a C18 column. The percentage of A>p remaining was fit to a first order rate equation. The fits and errors for each temperature were plotted on a van ′t Hoff plot.

2.4.3. Results

Determination of A>p pH stability

After 27 days of incubation at 4° only pH 8 and 9 showed detectable hydrolysis (Figure 2.6). At pH 9 the half-life of A>p was found to be 14 days; at pH 8 the half-life was 138 days, an order of magnitude longer than at pH 9 (Table 2.2). Assuming a hydroxyl-dependent rate-limiting step this was predicted to result in hydrolysis rate of 1400 days (3.8 years) and 14000 (38 years) for pH 7 and 6 respectively. Interestingly there was a 2:1 bias for forming 3′AMP vs 2′AMP.

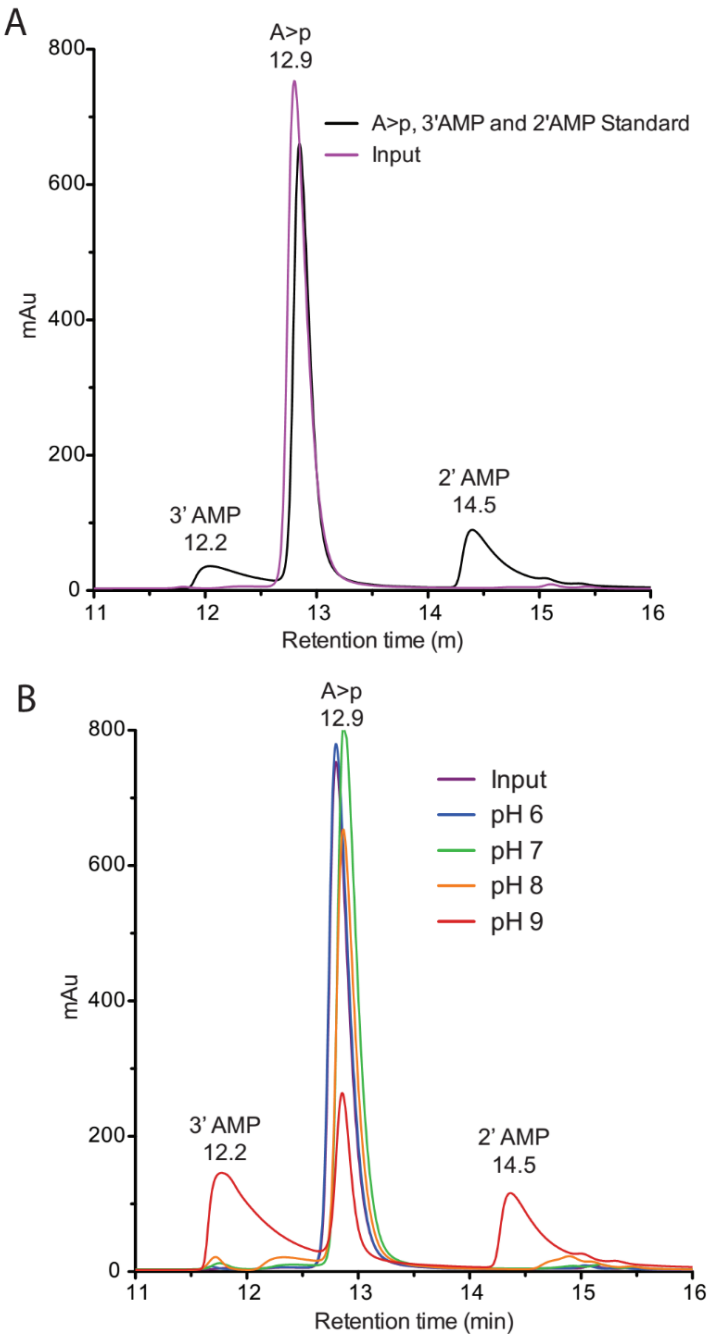

Figure 2.6. A>p was hydrolyzed at pH 6-9 to investigate the effect of pH.
A) Reverse phase HPLC (Zorbax Bonus RP column) of A>p, 3′AMP, and 2′AMP standard and un-hydrolyzed input material to the hydrolysis experiment. **B)** A>p hydrolysis at pH 6-9 after 27 days at 4ºC compared to a control stored at -80ºC. Only pH 8 and 9 showed significant hydrolysis, hydrolysis at pH 6 and 7 was not detected

Table 2.2. Summary of A>p half-life for pH 6 to 9 at 4°C.
Rows with a white background are experimental values; rows with a grey background are extrapolated

Condition	Half-life (days)
pH 9	14
pH 8	138
pH 7	1,400 (3.8 y)
pH 6	14,000 (38 y)

Hydrolysis of A>p at pH 6

The previous results determined that A>p was very stable at pH 6 at 4°C. Hydrolysis at pH 6 was investigated further (Figure 2.7). As expected hydrolysis rate was found to increase with temperature; the rate of hydrolysis was not affected by the presence or absence of $MgCl_2$. The rate and half-life at 95°C were 2.0 x 10^{-5} sec^{-1} and 0.40 days; at 80°C were 5.0 x 10^{-6} sec^{-1} and 1.60 days; at 72°C were 2.31 x10^{-6} sec^{-1} and 3.47 days; at 65°C were 1.16 x10^{-6} sec^{-1} and 6.94 days; and 47°C were 5.78 x10^{-7} sec^{-1} and 13.89 days. These rates were used to make a van ′t Hoff plot and extrapolate the rates and half-lives of A>p at lower temperatures. The half-life of A>p at 22° and 4° is extrapolated to be 3.5 and 49.5 years respectively, implying that polymerization experiments can have long time courses. The half-life of 49.5 years determined by the more detailed analysis is even longer than, but consistent with, the 38 years estimated from the previous experiment. The 3′AMP vs 2′AMP formation showed the same 2:1 bias as was observed previously.

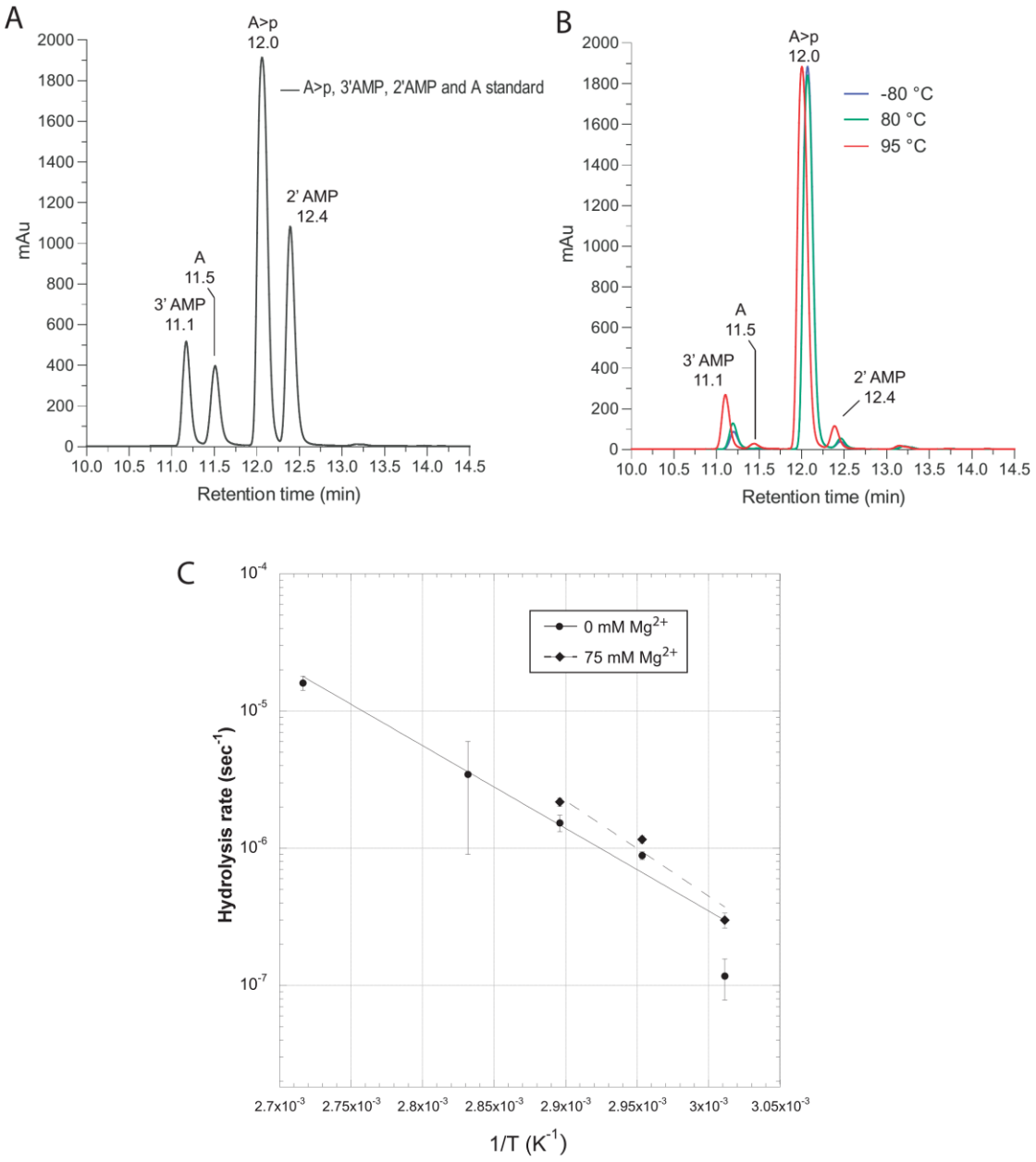

Figure 2.7. Investigating the rate of A>p hydrolysis at pH 6.
Hydrolysis was performed over a temperature gradient, then analyzed using reverse phase HPLC with a C18 column. **A)** Reverse phase HPLC (C18 column) of A>p, 3′AMP, 2′AMP and A standards. Note the baseline separation between the peaks. **B)** Sample time points (120 min) of A>p hydrolysis at 80°C and 95°C compared to a control stored at -80°C. After 120 min of hydrolysis at 80°C there was almost no difference from the control, in contrast to 95°C which yielded about twice as much 3′AMP and 2′AMP. **C)** A van 't Hoff plot of A>p hydrolysis at pH 6. The hydrolysis reaction is inversely dependent on temperature and unaffected by the concentration of Mg^{2+} relevant to our study.

Table 2.3. Summary of the half-life and hydrolysis rates of A>p hydrolysis.
Temperatures with a white background are experimental data points while those with a grey background are the extrapolated values.

Temperature (°C)	Half-life (days)	Hydrolysis Rate (sec^{-1})
95	0.40	2.0×10^{-5}
80	4.60	5.0×10^{-6}
72	3.47	2.31×10^{-6}
65	6.94	1.16×10^{-6}
47	13.9	5.78×10^{-7}
22	1,290 (3.5 y)	6.24×10^{-9}
4	18,100 (49.5 y)	4.44×10^{-10}
0	34,100 (93.4 y)	2.35×10^{-10}
-17	63,2000 (1730 y)	1.27×10^{-5}

2.5. The pH dependence of A>p polymerization on clay

2.5.1. Introduction

This experiment sought to determine the optimal pH for the polymerization of A>p on clay. ImpA polymerization is known to be optimal at pH 8 (Miyakawa et al. 2006), in part because the ImpA hydrolyses at acidic pH. In this experiment the pH was varied from 6 to 8. Unlike ImpA, A>p is stable in acidic conditions rather than basic so pH 6 was expected to yield more polymerization products. I hypothesized that A_2>p would be the dominant product since it is the result of a single attack of one A>p on another (Figure 2.8).

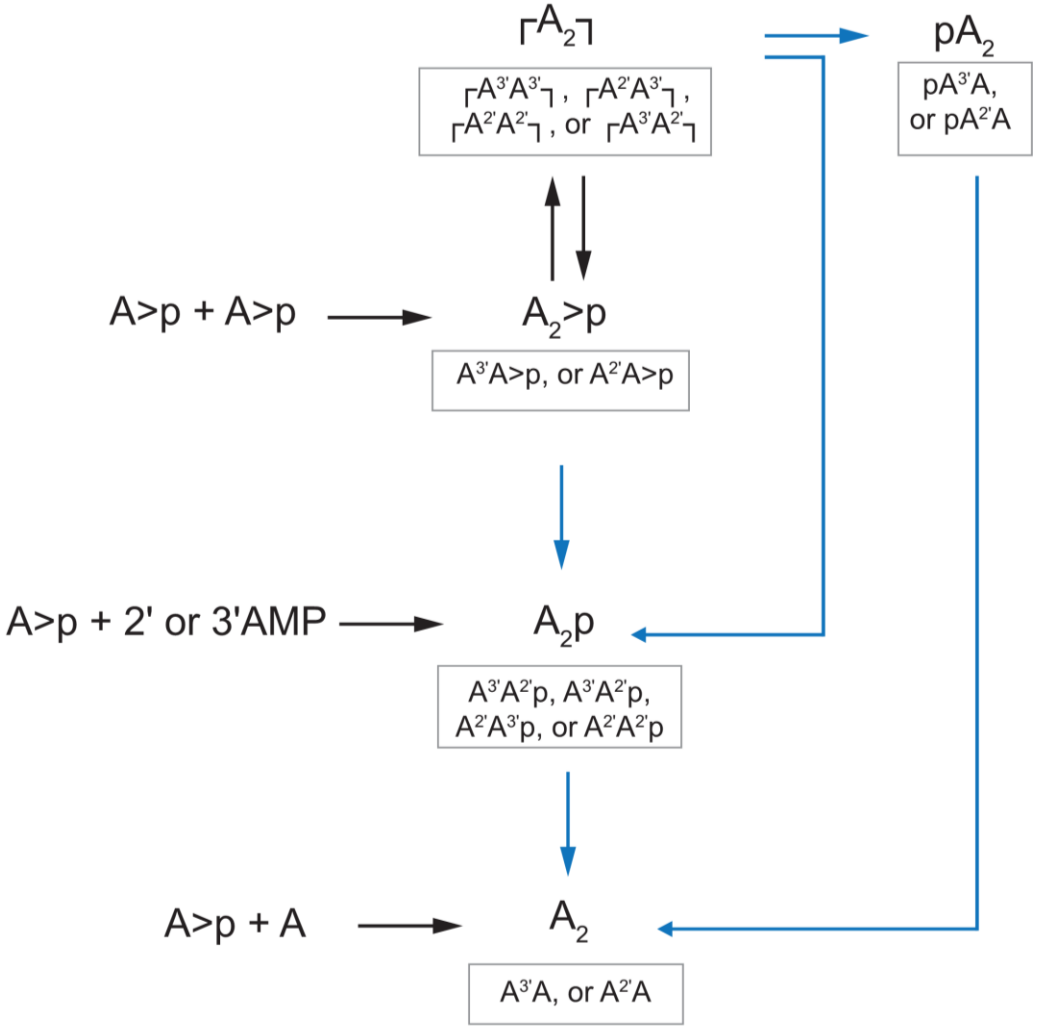

Figure 2.8. Possible dinucleotide products from A>p polymerization.
Polymerization and rearrangements are indicated by black arrows; hydrolysis reactions are shown by blue arrows. The 5′ hydroxyl of an A>p, 2′AMP, 3′AMP or A can attack the phosphate of an A>p to generate dinucleotide products. If the 5′hydroxyl of A_2>p attacks the 2′,3′ cyclic phosphate result is a cyclic dinucleotide ┌A_2┐. If the ┌A_2┐ is hydrolysed pA_2 or A_2>p will be formed; pA_2 can be hydrolysed to A_2. The attack of the 5′ hydroxyl of one A>p on the cyclic phosphate of another results in the formation of A_2>p which can hydrolyse to A_2p. The attack of 2′AMP or 3′AMP on an A>p results in A_2p, which can hydrolyse to A_2. The attack of the 5′ hydroxyl of A on an A>p will form an A_2. A hydroxyl of the terminal phosphate of 2′AMP or 3′AMP can attack the cyclic phosphate of an A>p to generate a dinucleotide joined by a diphosphate.

2.5.2. Methods

Each polymerization reaction contained 10 mM A>p, 200 mM NaCl, 75 mM $MgCl_2$, 100 mM buffer solution, 10 mg of pH 6 Banin-treated S/C yellow clay in 100 µl of ddH_2O. Control reactions were performed in the absence of clay. Reactions were incubated at 4°C for 30 days. Time points were taken at 1 and 30 days. Each 20 µl time point was centrifuged as previously described to pellet clay. Supernatants were transferred to fresh tubes and stored at -22°C before analysis.

T4 PNK was used to label the 5′ end of polymerization products; only the supernatants were investigated since the elution method available to me at this time showed no difference between supernatants and eluents. Each PNK reaction contained 70 mM Tris-HCl at pH 7.6, 15 mM DTT, 10 mM $MgCl_2$, 100 pmol of [γ-^{32}P] ATP, 1 mM polymerization sample and 5 units of PNK enzyme. PAGE was performed as described previously. Yield of polymerizations were calculated based on the quantification of gel bands.

2.5.3. Results

After 1 day of incubation only pH 6 reactions showed an accumulation of dinucleotide above the background in the incubated control. The main product detected was A_2. No trimer or longer oligomers were detected.

After 30 days of incubation with clay both dimers and trimers had formed as observed by PAGE analysis. The pH 6 condition produced the most dimers (1.5%) and trimers (0.1%) followed by pH 7 (1.3% and 0.08%), then by pH 8 (0.5% and 0.05%). After 30 days none of the no clay samples contained dinucleotides above the background, indicating that the presence of clay was essential for the synthesis and persistence of oligomers. All of the products detected lacked a terminal phosphate (A_2 and A_3) when they were expected to have terminal 2′, 3′ cyclic phosphates.

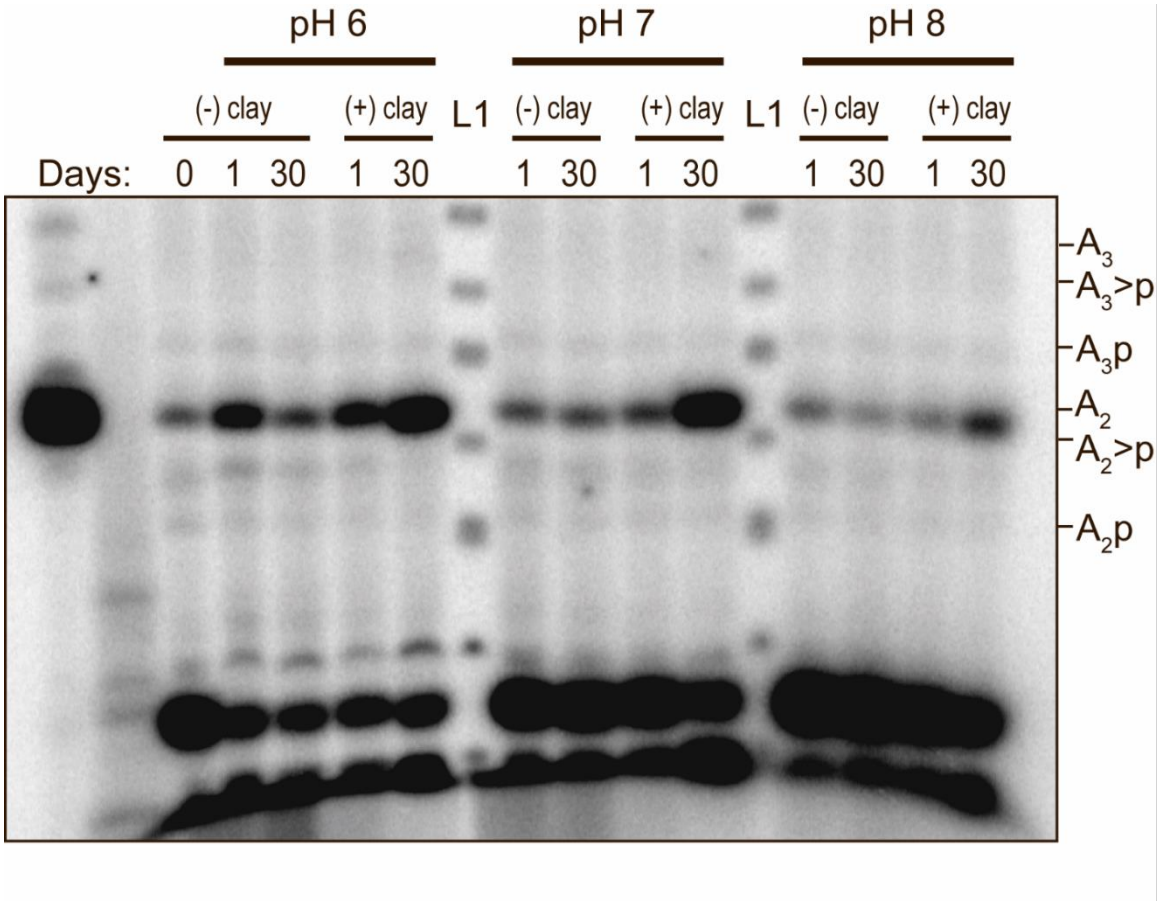

Figure 2.9. PAGE of the PNK labeled products from A>p polymerization on clay. pH 6, 7 and 8 were tested to determine the best condition for polymerization. After 1 day of incubation pH 6 with and without clay showed dinucleotide formation. After 30 days all pH conditions with clay show dinucleotide formation with pH 6 showing the most. After 30 days of incubation the pH 6 without clay sample had lost all of the dinucleotide that was present after 1 day. L1 and an A_2 commercial oligomer were used as size markers.

2.5.4. Selection of pH 6

All subsequent experiments were conducted at pH 6. This was due to the high stability of A>p at pH 6 with a hydrolysis rate at 25ºC of 6.2 x 10^{-9} sec^{-1} ($t_{1/2}$ = 3.5 years) and at 4ºC of 4.4 x 10^{-10} sec^{-1} ($t_{1/2}$ = ~50 years). Also, pH 6 was found to produce more polymerization products, with 1.5% of the A>p input converted to dinucleotide while only 1.3% and 0.8% was converted at pH 7 and 8 respectively.

2.6. Elution of RNA from Clay

2.6.1. Introduction

I developed a fast, quantitative, and non-sized biased method for eluting RNA from clay that also does not degrade RNA. I found that washing with 100 mM EDTA pH 11 could recover nearly 100% of bound oligomers without degradation. It was important to develop a reliable method for recovering RNA products from clay because I found that the method described in the literature (Joshi *et al.* 2009) was not reliable or quantitative

2.6.2. Methods

Binding of RNA oligomers to clay

Three ssRNA oligomers were selected for this study: AGC, A_6 and A_{15} (IDT). Each oligomer was 5′ end labeled using [γ-^{32}P] ATP (Perkin Elmer) and T4 PNK then gel purified using 23% urea-PAGE. A poly(A) ladder was generated by elongating the 5′-^{32}P labelled A_{15} with Poly(A) polymerase (NEB) followed by partial alkaline hydrolysis in 50 mM Na_2CO_3 at 95°C for 5 min. Montmorillonite clay was prepared as described previously; clay titrated to pH 6 was used in this experiment.

Binding experiments were designed to mimic the conditions of an abiotic polymerization reaction such as those described in Miyakawa and Ferris 2003, and Joshi *et al.* 2009. Unless otherwise stated 100 µl of a 1x binding solution (200 mM NaCl, 75 mM $MgCl_2$, 100 mM MES pH 6, 10 mM A>p and 1 µM AGC, A_6, A_{15} or poly(A) ladder) was added to 10 mg of prepared clay, with 10 µl set aside as an unbound control (I). The mixture was vortexed to form a slurry and incubated with agitation at 4 °C.

A time course of binding was performed for each AGC, A_6, A_{15}, and poly(A) ladder where 200 µl of 1x binding solution was added to 20 mg of prepared clay. At 0.33, 2, 5, 15, 30, 60 and 120 min, 15 µl of the clay slurry was transferred to a fresh tube. The slurry was immediately pelleted by centrifugation at 12,000 RCF for 2 min. Each supernatant was then transferred to new tube. Five microliters of each supernatant was added to 1 ml

of scintillation counting fluid (Amersham) and counted on a Beckman LS6000SC scintillation counter in triplicate.

Elution of RNA oligomers from clay

Oligomers AGC, A_6, A_{15} and poly(A) ladder were bound to clay as described above, with 100 µl of solution left aside before clay was added as an unbound control (I).

After binding, the clay slurries of AGC, A_6 and A_{15} were pelleted (12,000 RFC, 2 min) and the supernatants (S) were transferred to fresh tubes. Each pellet was resuspended in 100 µl of 100 mM EDTA pH 8 or pH 11 and incubated with agitation at 4°C for 30 min. The slurries were then pelleted and the eluent (E1) was transferred to a fresh tube. The pellets were resuspended in 100 µl of 100 mM EDTA at pH 8 or 11; this was repeated 3 times to give E2, E3 and E4. The pH was measured at each step.

In order to assess if there was an upper limit to the oligomer lengths that could be eluted, a poly(A) ladder was bound to the clay. The ladder was eluted with six 30 min incubations in 100 mM EDTA at pH 11 (E1-E6). Each S, I, and E1-E6 were diluted in 2x denaturing loading dye and resolved using 23% sequencing PAGE to look for size bias in the elution method. I, S, and E1-E6 were also scintillation counted to determine the percent bound and total recovery for each elution.

Comparison to Joshi et al. 2009 method of elution

To compare my method to the method by Joshi et al in 2009, A_{15} was bound to the clay as described previously, with 100 µl of binding solution left aside (I). Clay slurries were pelleted and the supernatants were removed and kept (S). Elution was then performed as described above using 100 mM EDTA pH 11, 100 mM NaCl in 30% CH_3CN or ddH_2O. Elution solutions were incubated with the clay for either 30 min for each elution step (E1-E4) or incubated for 2 hr for E1, 2 hr for E2, overnight for E3 and 1 hr for E4 as described by Joshi *et al.*

Five microliters of I, S and E1-E4 were added to 1 ml of scintillation counting fluid and counted as described above. Further, 5 µl of I, S, and E1-E4 were added to 5 µl of 2x denaturing loading dye (90% formamide, 10 mM EDTA pH 8, 0.025% Xylene Cyanol, and

0.025% bromophenol blue) and resolved using 23% sequencing PAGE to check for RNA integrity.

2.6.3. Results

Binding of RNA oligomers to clay

Oligomers longer than 6 nt were found to bind quantitatively. By scintillation counting I found that 85% of AGC, 99% of A_6, 99% of A_{15} and 96% of a poly(A) ladder bound to the clay after two hours. Binding of the poly(A) ladder showed that oligomers 6 nt and longer were completely bound after 2 min (Figure 2.10). More time did not increase the fraction bound. This result indicated that shorter oligomers do not bind completely, leaving some in the supernatant of polymerization reactions. I was able to detect some dimers and trimers in the early polymerization experiments (both ImpA and A>p polymerization) by centrifuging the clay out of solution, but without preforming elution or when preforming the flawed elution method because short oligomers do not bind completely.

Elution of RNA oligomers from clay suggests two factors in RNA binding

Elution of clay-bound oligomers with 100 mM EDTA pH 8 shows a bias towards recovering shorter oligomers: recovering 84% of ACG, 68% of A_6, and 43% of A_{15}. Elution with 100 mM EDTA at pH 11 does not show the same bias, with 98% recovery of AGC, 93% of A_6 and 94% of A_{15}. Further, the elution in pH 11 showed between 38 and 48% recovery in the first elution step, with each subsequent elution step recovering about half of the remaining oligomer. In contrast elution in pH 8 EDTA shows approximately equal recovery across all elution steps.

The pH measurements taken after each elution step show that the clay acts as a buffer, since the pH of the clay slurry does not match the pH of the EDTA solution that was applied. The pH of the solution does rise with each subsequent elution step. I hypothesize that EDTA dissociates RNA from clay by chelating divalent metal ions that are coordinating the RNA clay complex. The pH 11 solution may be more effective than pH 8 because the RNA and clay are also being deprotonated, eliminating their ability to make hydrogen bonds. The bias of pH 8 elutions for recovering short oligomers suggests

that divalent metal ions are more important for short oligo binding; in contrast as the oligomers get longer the pH 8 EDTA elution becomes decreasingly effective, suggesting that as length increases so does the importance of hydrogen bonding for clay binding.

Binding and elution of a poly(A) ladder from clay

A poly(A) ladder was bound to clay, then was eluted with six 30 min incubations in 100 mM EDTA at pH 11 (E1-E6) (Figure 2.10). The total recovery was 96% by scintillation counting (Table 2.4). Each I, S, and E1-E6 were run in 23% denaturing PAGE. Individual bands were resolved up to 30 nt. Quantification of each gel band showed that the relative intensities of each band were unchanged between each E1 though E6 indicating that there was no size bias when eluting with the 100 mM EDTA pH 11 method.

Table 2.4. Recovery of a poly(A) ladder bound to clay.
Elution was performed using 100 mM EDTA pH 11, replaced six times to generate E1-E6. After E6 96% of the bound material was recovered.

	Recovery %
S	2
E1	50
E2	25
E3	12
E4	4
E5	2
E6	1
Total	**96**

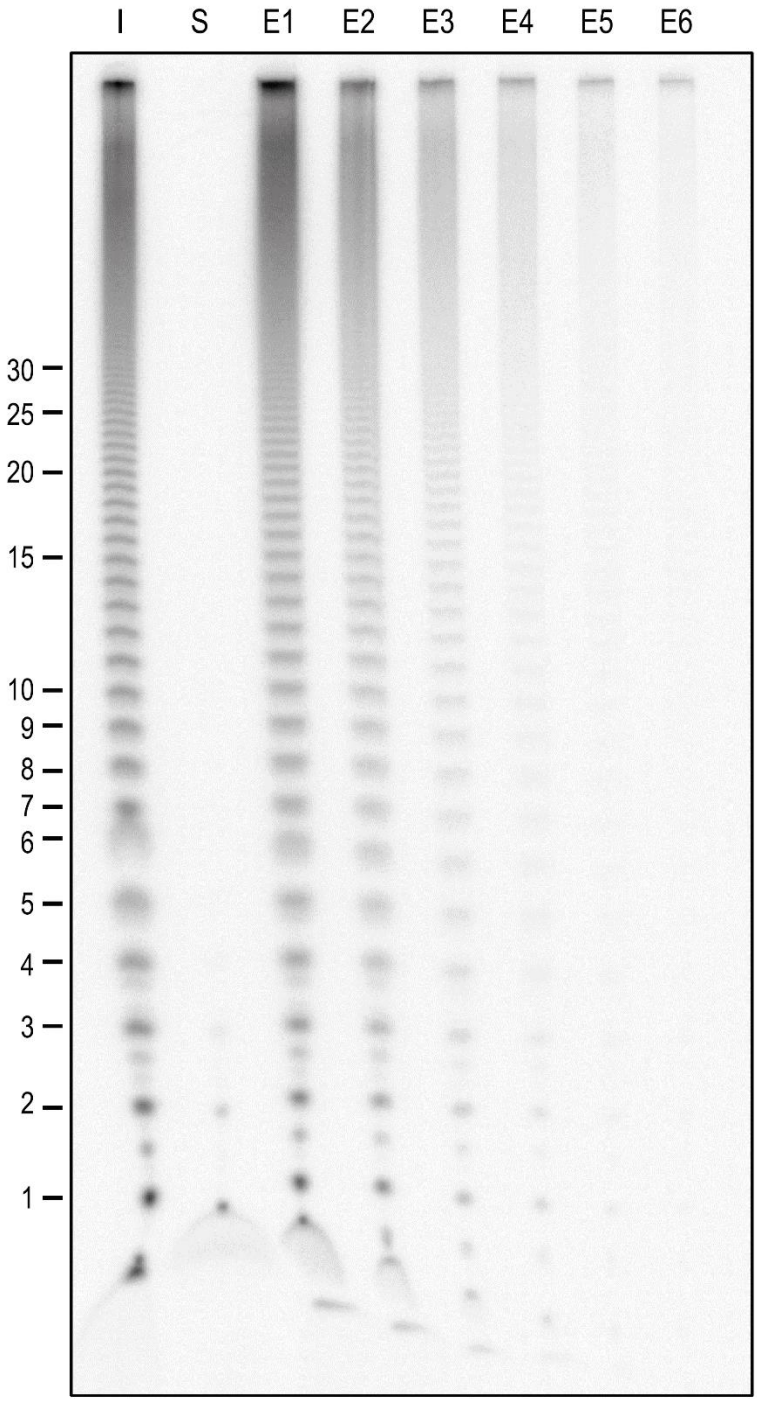

Figure 2.10. Binding and elution of a poly(A) ladder from montmorillonite clay. Elution was preformed using 100 mM EDTA pH 11, replaced six times to generate E1-E6. As seen on the gel there is no bias towards recovering shorter RNA oligomers, nor does there appear to be an upper limit of oligomer length that can be recovered.

Comparison to Joshi et al. (2009) method of elution

I demonstrated that my 100 mM EDTA pH 11 method was more effective than the elution method from Joshi *et al.* 2009. The Joshi *et al.* method uses 100 mM NaCl in 30% CH_3CN that is incubated 2 hr for E1, 2 hr for E2, overnight for E3 and 1 hr for E4. In contrast my elution steps are all 30 minutes. I compared both the elution solutions and the incubation times for the elution of A_{15} (Figure 2.11). It was found that there was very little difference between 30 min incubations and the longer Joshi *et al.* incubation times, indicating I was safe to use the much faster 30 min per elution protocol. My 100 mM EDTA pH 11 elution solution was by far the most effective, recovering 94% of input A_{15} after four 30 min incubations and after the four longer elutions. Surprisingly, the ddH_2O control was more effective than 100mM NaCl in 30% CH_3CN, though neither eluted more than 20% of bound A_{15}. None of these solutions degrade the input RNA as confirmed by denaturing PAGE.

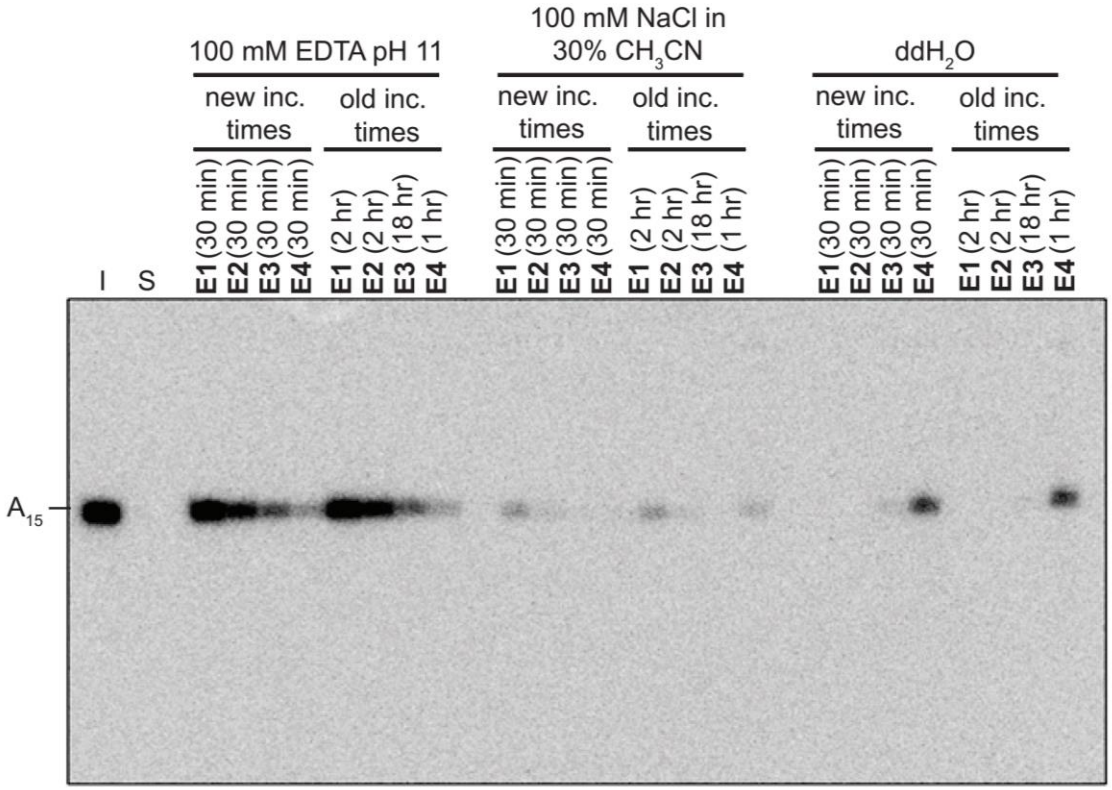

Figure 2.11. Comparison of elution solutions and steps for recovering A_{15} from clay.

Elution of 5′ ^{32}P labelled A_{15} in 100 mM EDTA at pH 11 was the most effective solution. 100 mM NaCl in 30% CH_3CN was the least effective. Short 30 min elution steps were as effective as the long overnight elution from Joshi et al. No hydrolysis was observed in any solution.

Table 2.5. Comparison of elution solutions and steps for recovering A_{15} from clay. Elution of 5′ ^{32}P labelled A_{15} in 100 mM EDTA at pH 11 was the most effective solution. 100 mM NaCl in 30% CH_3CN was the least effective. Short 30 min elution steps were as effective as the long overnight elution from Joshi et al.

	Recovery %					
	S	**E1**	**E2**	**E3**	**E4**	**Total**
		30 min	*30 min*	*30 min*	*30 min*	
100 mM EDTA pH 11	0	48	23	13	10	94
100 mM NaCl in 30% CH_3CN	0	5	1	0	1	7
ddH_2O	0	0	0	3	15	18
	S	**E1**	**E2**	**E3**	**E4**	**Total**
		2 hr	*2 hr*	*ON*	*1 hr*	
100 mM EDTA pH 11	1	54	24	9	6	94
100 mM NaCl in 30% CH_3CN	0	4	1	1	3	9
ddH_2O	2	1	0	2	16	20

2.7. Polymerization of A>p on clay

2.7.1. Introduction

Montmorillonite clay has been shown to polymerize ImpA into RNA (Ferris and Ertem 1992; Ferris 2002; Miyakawa et al. 2006, Joshi et al 2009); in (2.5) the prebiotically plausible nucleotide A>p was polymerized into RNA by clay. After developing a quantitative method for eluting RNA from clay the polymerization of A>p by montmorillonite clay was investigated in more detail. The products of these polymerizations are analyzed by PAGE, HPLC and MALDI-TOF.

2.7.2. Methods

RNA ladders

Two RNA ladders were designed to identify polymerization products by gel mobility (Figure 2.12). L1 was made by selecting a poly(A) oligomer (for example A_{15}) then 5′ end labelling with [γ-^{32}P] ATP and T4 PNK (3′ phosphatase minus, NEB). T4 (3′ phosphatase minus) PNK is a mutant version of the T4 PNK enzyme that lacks the 3′ phosphatase activity of the WT. This was followed by partial alkaline hydrolysis to generate a ladder with a 5′ ^{32}P labelled phosphate and 2′, 3′ cyclic, 3′ or 2′ phosphates. The longest and second longest RNAs in the ladder always have 2′ and 3′ hydroxyls; for the longest RNA this is because it was purchased that way, but it is unknown why the second longest RNA always has 2′ and 3′ hydroxyls.

L2 is made by selecting a poly(A) oligomer, doing partial alkaline hydrolysis followed by alkaline phosphatase (Fast AP or CIP, Thermoscientific) treatment to remove all phosphates then finally 5′ end labelling with [γ-^{32}P] ATP and T4 PNK. This generates a ladder with a 5′ ^{32}P labelled phosphate but no 2′ or 3′ phosphates. The phosphatase enzyme does not always completely dephosphorylate the di and trinucleotides so in some examples of L2 there are di and trinucleotides with 2′,3′ cyclic, 3′ or 2′ phosphates.

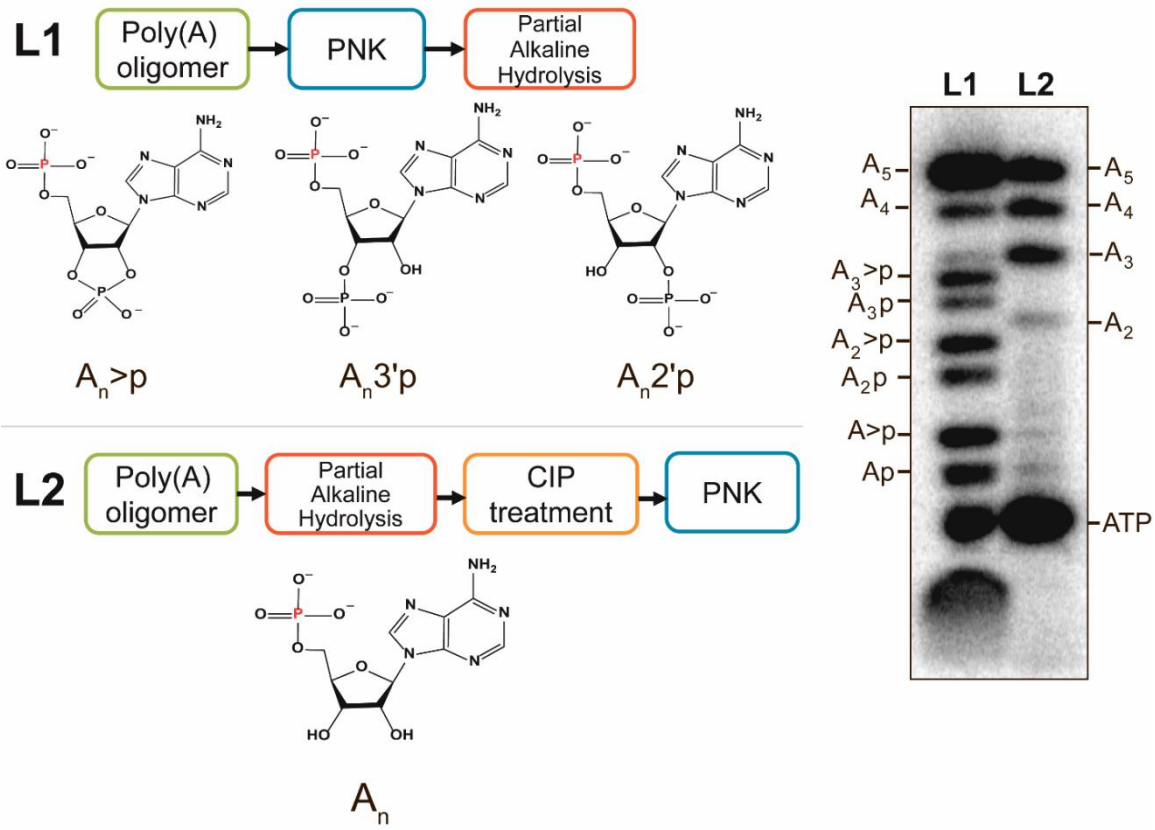

Figure 2.12. Preparation of RNA ladders.
L1 was made by selecting a poly(A) oligomer (A_5 here) then 5′ end labelling with [γ-^{32}P] ATP and T4 PNK (3′ phosphatase minus), followed by partial alkaline hydrolysis to generate a ladder with a 5′ ^{32}P labelled phosphate and 2′,3′ cyclic, 3′ or 2′ phosphates. L2 is made by selecting a poly(A) oligomer (A_5 here), doing partial alkaline hydrolysis followed by calf intestinal phosphate treatment to remove all phosphates then finally 5′ end labelling with [γ-^{32}P] ATP and T4 PNK. This generates a ladder with a 5′ ^{32}P labelled phosphate but no 2′ or 3′ phosphates. In the figures the 5′ ^{32}P labelled phosphate is not indicated because it a result of the enzymatic treatment and not an intrinsic property of the starting oligomer.

Polymerization of 10 mM A>p

A polymerization reaction contained 10 mM A>p, 200 mM NaCl, 75 mM $MgCl_2$, 100 mM MES pH 6, 10 mg of pH 6 Banin-treated S/C yellow clay in 100 µl of ddH_2O. Control reactions were performed in the absence of clay. Reactions were incubated at 4°C for 176 days, note that this reaction was the same as the reaction in section 2.4 that was left incubating while the elution method was developed. There was 20 µl remaining from the previous experiment, this time point was centrifuged as previously described to pellet clay. The supernatant was transferred to a fresh tube and stored at -22°C before analysis.

The pellet was resuspended in 100 µl of 100 mM EDTA pH 11 and incubated with agitation at 4°C for 30 min. The slurry was then pelleted and the eluent (E1) was transferred to a fresh tube. The pellet was resuspended in 100 µl of 100 mM EDTA at pH 11; this was repeated 3 times to give E2, E3 and E4. S and E1-E4 were pooled.

T4 PNK or T4 (3′ phosphatase minus) PNK was used to label the 5′ end of the pooled polymerization products. In some cases Fast AP alkaline phosphatase was used in combination with PNK. A FastAP reaction contained 100 pmol ImpA, 50 mM Tris HCl (pH 7.6), 10 mM $MgCl_2$, 10 mM KCl and 0.5 units of FastAP enzyme in ddH_2O. The FastAP reactions were incubated at 37°C for 30 min then heat inactivated at 75°C for 10 minutes. Each FastAP reaction was adjusted to contain 100 pmol of ImpA, 50 mM Tris HCl (pH 7.6), 10 mM $MgCl_2$, 10 mM KCl, 10 mM DTT, 100 pmol of [γ-^{32}P] ATP and 5 units of T4 polynucleotide kinase (New England Biolabs) in 20 µl. The PNK reactions were incubated at 37°C for 30 minutes, then heat inactivated at 65°C for 25 minutes.

Polymerization of 300 mM A>p

A polymerization reaction contained 300 mM A>p, 200 mM NaCl, 75 mM $MgCl_2$, 100 mM MES pH 6, 50 mg of pH 6 Banin-treated S/C yellow clay in 500 µl of ddH_2O. Control reactions were performed in the absence of clay. Reactions were incubated at 4°C and 22ºC for 90 days. Time points were taken at 5, 30 and 90 days. Each 20 µl time point was centrifuged as previously described to pellet clay. The supernatants were transferred to a fresh tube and stored at -22°C before analysis.

The pellets were resuspended in 100 µl of 100 mM EDTA pH 11 and incubated with agitation at 4°C for 30 min. The slurry was then pelleted and the eluents (E1) were transferred to a fresh tube. The pellets were resuspended in 100 µl of 100 mM EDTA at pH 11; this was repeated 3 times to give E2, E3 and E4. S and E1-E4 were pooled.

A>p polymerization time points were diluted to 10 and 100 µM in ddH_2O, 20 µl of each dilution was injected multiple times into a Zorbex-C18 (Agilent) column with a TEAA to CH_3CN gradient. Absorbance at 260 nm was recorded. A standard solution of 2 mM A>p, 1 mM 2′AMP, 0.5 mM 3′AMP, 0.25 mM A and 0.1 mM A_2 was also analyzed by HPLC to determine the mobility of unpolymerized monomers, which were expected to account

for the bulk of the absorbance, and to determine the mobility of a dinucleotide. Fractions were collected at 1 min intervals from 12 to 26 min. Fractions were dried in a rotary evaporator, resuspended in 20 µl ddH_2O and stored at -22°C. HPLC traces were used to calculate yields.

From each fraction of the 100 µM HPLC injections 1 µl was 5′ end-labelled using T4 PNK (3′ phosphatase minus). Each 1x PNK reaction contained 70 mM Tris-HCl at pH 7.6, 15 mM DTT, 10 mM $MgCl_2$ and 1 µl of an HPLC fraction sample. PAGE was performed as described previously.

Fractions from the 100 µM HPLC injections were resuspended in 5 µl of ddH_2O. Of this 1 µl was diluted into 4 µl of matrix solution (1x Matrix solution: 160 mM 3-Hydropicolinic acid and 16 mM L-Tartaric acid). Each sample in matrix solution was spotted onto a MALDI plate and analyzed on a Bruker Autoflex Matrix Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometer (MALDI-TOF-MS) in negative ion mode. A mix of 1 mM A_2 and A_6 were used as mass calibrants; matrix-only controls were also analyzed.

It was suspected that $┌A_2┐$ was present in the 14-15 min fraction from the incubation at 22°C in the presence clay for 90 days. To confirm this, partial alkaline hydrolysis of the fraction was performed with time points taken at 0, 30 and 60 minutes. These were then ethanol precipitated and PNK labelled as previously described. The time course was analyzed by PAGE.

2.7.3. Results

Polymerization of 10 mM A>p

Once the new elution method was developed samples that had been incubating from the previous A>p polymerization were eluted and analyzed (Figure 2.13). A_2 was the dominant band detected after 176 days of incubation with clay. Other products identified were A_3 and A_2p. There was no difference between the unincubated control and the sample incubated for 176 days without clay. The identity of A_2 and A_3 were confirmed by

comigration in PAGE with the A_2 and A_3 bands from L2 ladder. Faint bands corresponding to the mobility of A_4 and A_5 were detected after long exposures of the gels (Figure 2.15).

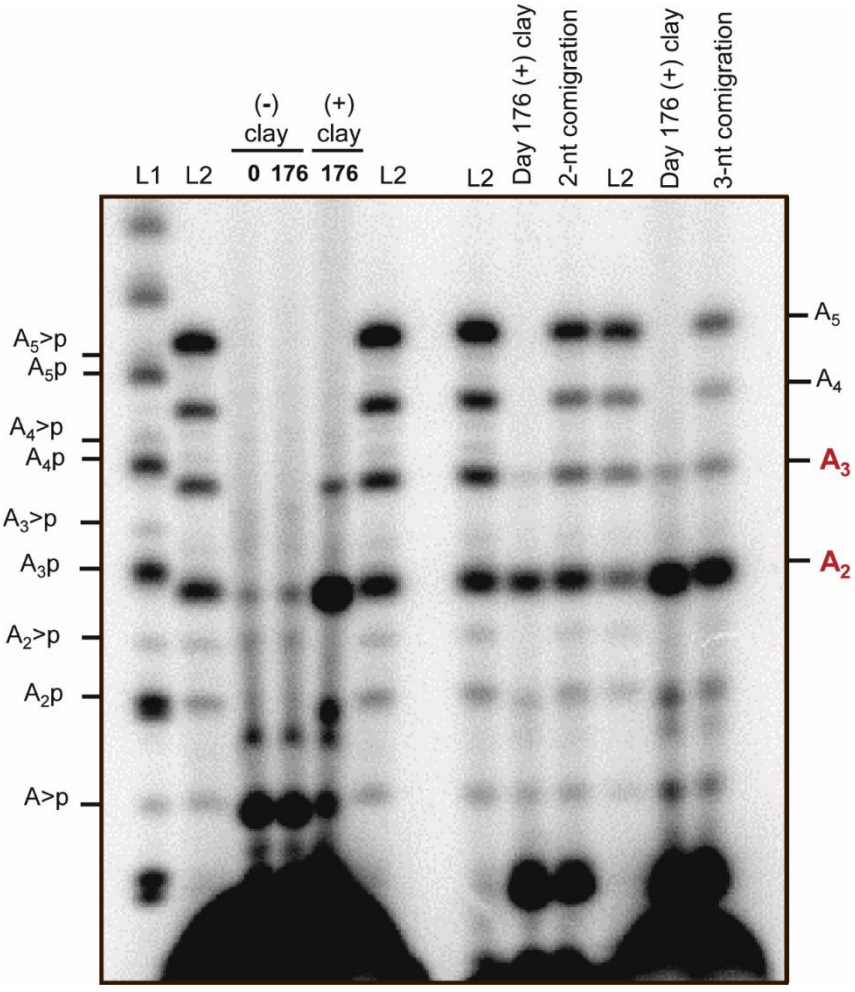

Figure 2.13 A>p (10 mM) was incubated at 4°C for 176 days with and without clay. Products were eluted from clay and PNK labeled before PAGE analysis. The presence of clay was essential for the formation of di and trinucleotides. The (+) clay samples were mixed with L2 ladder to show that A_2 and A_3 comigrate with the A_2 and A_3 from the ladder.

It was expected that the dominant polymerization product would be A_2>p because it is the product of two A>p monomers, but this species was not observed on the gel. A_2 was the dominant band seen on the PAGE, with A_2p also detected. It was hypothesized that the WT T4 PNK was removing the phosphate from A_2>p (so appearing as A_2 in the gel) since it is known to have 3′ phosphatase activity. To test this a T4 PNK mutant that lacks 3′ phosphatase activity was used in a labeling reaction in parallel with WT T4 PNK

(Figure 2.14). Before PNK labelling samples were treated with Fast AP alkaline phosphatase or ddH_2O.

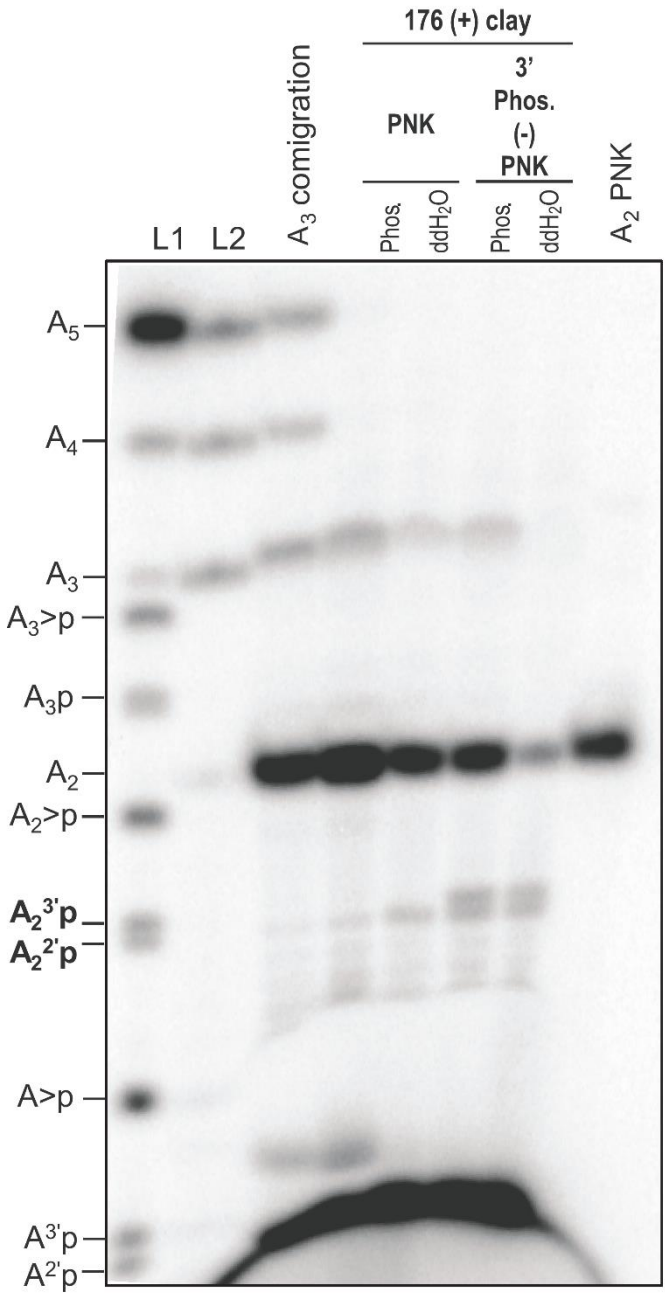

Figure 2.14. Analysis of A>p polymerization by PNK variants and Fast AP.
Pretreatment with phosphatase increased the intensity of the A_2 and A_3 bands. A_2p separated into a double band; the upper band is only visable in samples labelled with the T4 PNK (3′ phosphatse minus) enzyme indicating that the upper band is $A_2{}^{3'}$p and the lower band is $A_2{}^{2'}$p.

Just as in the previous gel (Figure 2.13) the sample treated with ddH_2O and WT T4 PNK showed A_2 as the dominant band, with less intense bands at A_2p and A_3. A_2p ran as a double band in this gel; in lanes with WT T4 PNK only the lower of the double band is detected. When Fast AP was used in combination with the WT T4 PNK A_2 was still the dominant band, but both A_2 and A_3 were significantly darker than the ddH_2O/WT T4 PNK combination. Interestingly the sample prepared with ddH_2O and the T4 PNK (3′ phosphatase minus) showed only a slight dominance of A_2 over the A_2p product; the A_3 band was not detected using this combination. When Fast AP was used in combination with T4 PNK (3′ phosphatase minus) the A_2 band was dominant and the A_3 band was detected. In lanes with T4 PNK (3′ phosphatase minus) both bands of the A_2p double band were visible. The upper band of the A_2p double band must be $A_2^{3'}p$ because it only appears when T4 PNK (3′ phosphatase minus) is used.

Since A_2>p was not observed with either T4 PNK enzyme I concluded that A_2>p was not being converted to A_2 by the PNK enzyme. At the same time it appears that A_2 is not the dominant product either, given that it labelled weakly with T4 PNK (3′ phosphatase minus) in the absence of Fast AP. I propose that the dominant product is actually pA_2. The T4 PNK is known to catalyze the exchange of a 5′ phosphate, albeit at a slower rate than it catalyzes the phosphorylation of a 5′ hydroxyl. pA_2 being the dominant product is consistent with the observation that Fast AP increased A_2 band intensity; if Fast AP dephosphorylates pA_2 to A_2 there would be more A_2 signal in the gel because PNK can more rapidly catalyse phosphorylation than phosphate transfer. There is further evidence for the pA_2 being the dominant product in later experiments (Table 2.9)

There are two possible reactions between A>p monomers (Figure 2.8): polymerization to A_2>p, which was not observed, or cyclization to $\ulcorner A_2 \urcorner$, which would not label by PNK because it has no available 5′ end. However when $\ulcorner A_2 \urcorner$ hydrolyzes to either pA_2 or A_2>p it can be visualized by PAGE. As such, I suspect that the most dominant product is actually $\ulcorner A_2 \urcorner$, which I see by PAGE only when it hydrolyzes to pA_2. Hydrolysis of $\ulcorner A_2 \urcorner$ to pA_2 was observed to be strongly preferred over hydrolysis to A_2>p (alternatively when hydrolysis generates A_2>p it may rapidly re-cyclize). Similarly the band at A_3 is enhanced in treatments with Fast AP suggesting its source was pA_3. Intriguingly, long exposure of a gel revealed bands at A_4 and A_5, but only in samples treated with Fast AP,

revealing the presence of pA_4 and pA_5. The pA_3, pA_4 and pA_5 oligomers cannot be made from the extension of pA_2 by A>p because the pA_2 has neither a 5′ hydroxyl or a 2′,3′ cyclic phosphate. This suggests a mechanism for generating longer products where the 5′ hydroxyl of an A>p attacks a phosphate of cyclic oligomer, increasing its length by one, followed by rapid re-cyclization. A subset of these cyclic oligomers get trapped in the pA_n form after hydrolysis.

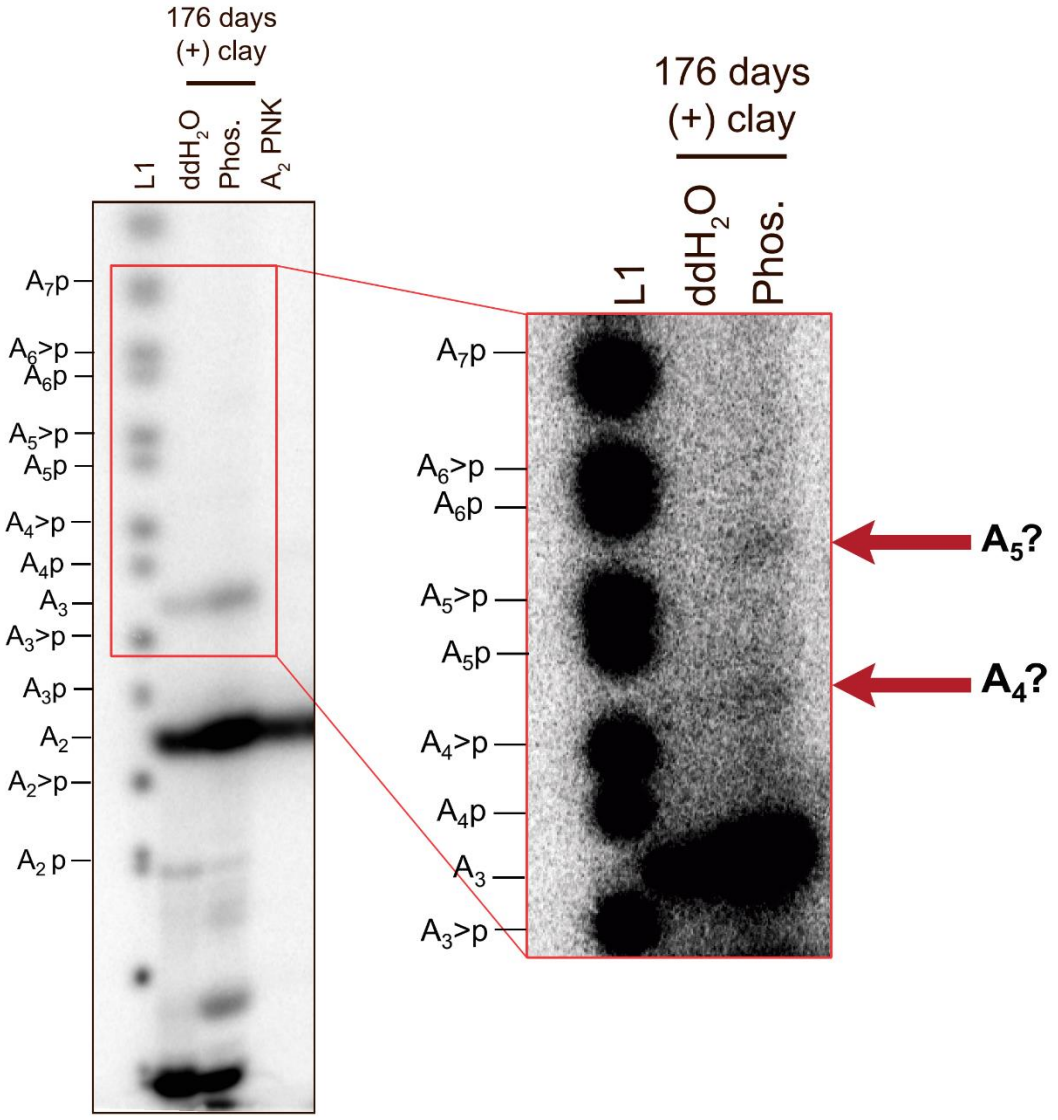

Figure 2.15. Increasing the exposure time of 176 days with clay PNK reactions revealed faint bands at A_4 and A_5 that can only be seen in samples that were treated with Fast AP.

Polymerization of 300 mM A>p

New polymerization reactions were conducted at 300 mM A>p because more material was needed for more detailed analysis. Increasing the concentration should also allow the detection of longer products. This is because the rate of formation of an oligomer of length n is directly dependent on the concentration of oligo n-1, and so the result is that longer oligomer synthesis is dependent on the concentration of the A>p monomer. The polymerization reactions of 300 mM A>p showed complex mixture of RNA oligomers up to 5 nucleotides in length; product formation was highly dependent on the presence of clay.

HPLC

Reverse phase HPLC (C18 column) allowed for the quantification of product species (Table 2.6). HPLC traces showed that A>p was the most common species after 90 days at 4°C. In contrast incubations with clay at 22°C A>p was largely converted to hydrolysed monomer species, and without clay at 22°C the hydrolyzed monomers were nearly as common as A>p monomers. Peaks with higher retention than monomers are thought to be mainly oligonucleotide products due to the A$^{3'}$A HPLC standard retention time relative to the monomers, oligonucleotide detection in gels and MALDI-TOF evidence. Incubation at 22^{o}C generated more oligomers than at 4^{o}C; while incubation in the presence of clay significantly enhanced the formation of oligomers. In all conditions longer incubations produced more oligomer products, although this gain is marginal at 4^{o}C without clay. The best total oligomer yield was 1.89% from incubating at 22°C with clay for 90 days.

Table 2.6. A summary of yields from 300 mM A>p incubations.
Incubations of 300 mM A>p were analyzed by reverse phase HPLC (C18 column) to determine yields. For simplicity all hydrolyzed monomers are grouped, as are all oligomer products. Yields are given as a percentage of the total sample absorbance at 260 nm.

		Incubation (days)	Hydrolyzed monomers (2′AMP, 3′AMP and A)	A>p	Oligomers
	no clay	**0**	2.9%	96.9%	0.27%
4°C	**clay**	**5**	4.9%	94.9%	0.17%
		30	7.1%	92.5%	0.34%
		90	14.1%	85.5%	0.42%
	no clay	**5**	3.0%	96.7%	0.29%
		30	4.6%	95.1%	0.28%
		90	6.3%	93.4%	0.31%
22°C	**clay**	**5**	7.9%	91.7%	0.37%
		30	26.3%	73.0%	0.76%
		90	74.1%	24.1%	1.89%
	no clay	**5**	4.4%	95.3%	0.28%
		30	10.5%	89.2%	0.31%
		90	46.4%	52.7%	0.94%

PAGE

The 22°C incubations were investigated further by T4 PNK (3′ phosphatase minus) labeling of the HPLC fractions that were collected from 12 to 26 min in 1 min intervals. Side by side labelling and gel electrophoresis confirmed that there were more oligomers in the incubated versus unincubated samples. Further, PAGE analysis confirmed that incubations with clay showed significantly more accumulation and longer products than both incubations without clay and no incubation (Figure 2.16). The ratio of hydrolysis to polymerization was found to be 39 at 22°C and 33 at 4°C, suggesting that polymerization at lower temperature will yield slightly more oligomers given longer incubations. Comparison of the HPLC to the PAGE shows that some peaks contain multiple oligomer species.

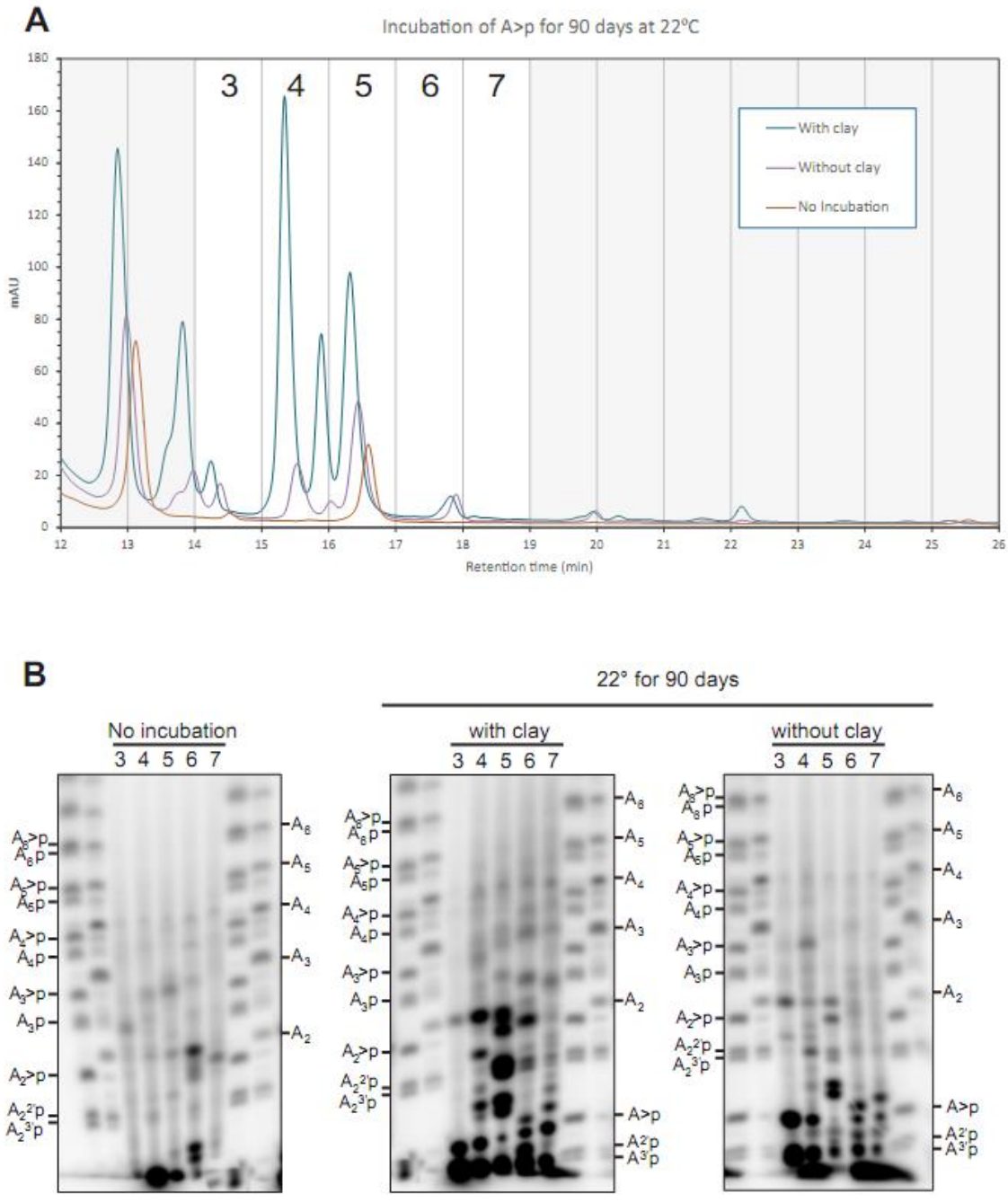

Figure 2.16. Comparison of 300 mM A>p incubations with and without clay.
A) A>p was incubated with and without clay for 90 days, after which the eluted material was fractionated using reverse phase HPLC. **B)** The PNK reactions of fractions from 14-19 min (numbered 3-7) of each clay, no clay and no incubation samples were run together on PAGE for comparison. Significantly more dinucleotide and longer product can be observed in the presence of clay than either control.

In the 22°C A>p incubation with clay the most abundant band observed by PAGE was A_2 (Figure 2.17), just as was found in the 10 mM A>p polymerization experiment at 4°C. A_2>p, $A_2{}^{2'}$p and $A_2{}^{3'}$p were also highly abundant. Trimers A_3p, $A_{3>}$p, and A_3p; tetramers A_4p, A_4>p, and A_4p; as well faint bands of pentamers A_5p, A_5>p, and A_5p were detected. Many of these products appeared in multiple fractions, this is likely due to the large number of possible isomers (Figure 2.8). The A_2, pA_2 and A_2>p dinucleotides each have two isomers owing to the two possible phosphodiester bond conformations: $A^{2'}A$ and $A^{3'}A$; $pA^{2'}A$ and $pA^{3'}A$; and $A^{2'}A$>p and $A^{3'}A$>p. Likewise A_2p has four possible isomers: $A^{2'}A^{2'}$p, $A^{2'}A^{3'}$p, $A^{3'}A^{2'}$p, and $A^{3'}A^{3'}$p. There are also four possible heterocyclic dinucleotides that in principle can be detected by HPLC but not by PNK: $\lceil A^{2'}A^{2'}\rceil$, $\lceil A^{2'}A^{3'}\rceil$, $\lceil A^{3'}A^{2'}\rceil$, and $\lceil A^{3'}A^{3'}\rceil$. Three species: $A^{2'\text{-}2'}A$, $A^{2'\text{-}3'}A$ and $A^{3'\text{-}3'}A$ are also possible and have two available 5′ hydroxyls each; it is unknown if they would label by PNK. This is seventeen possible dinucleotide isomers with phosphodiester bonds; the number of isomers increases with every nucleotide gained. Further, there are numerous related species such as depurinated oligomers, polymers with non-phosphodiester bonds and possibly others that may or may not PNK label and be detected by PAGE.

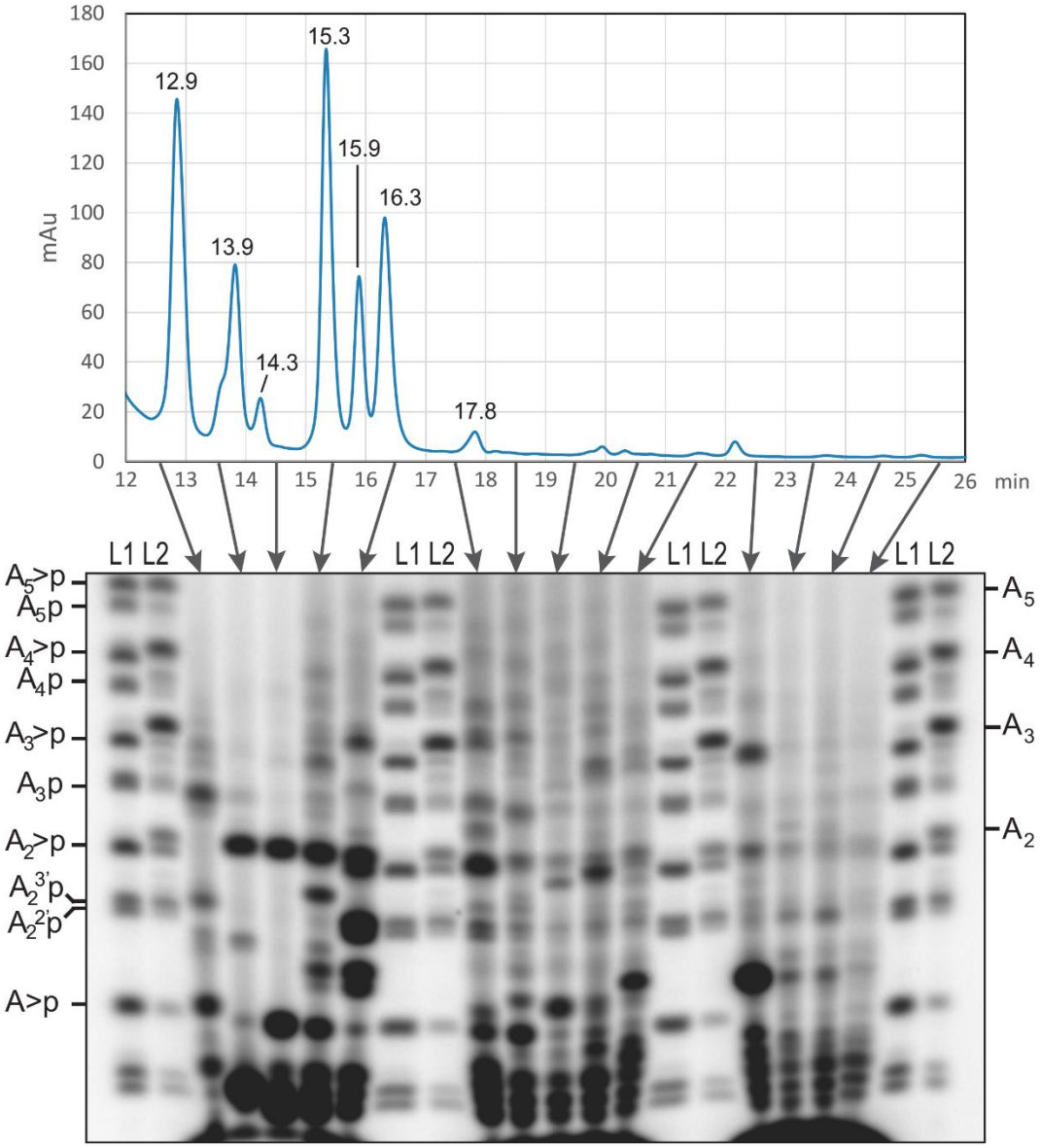

Figure 2.17. PAGE of HPLC fractions from 300 mM A>p with clay, 90 days at 22°C. A>p was reacted in the presence of clay for 90 days, after which the eluted material was fractionated using reverse phase HPLC. Each fraction was then 5′ end labeled with ^{32}P-γATP, using PNK (3′ phosphatase minus). Products of up to five nucleotides long were seen by PAGE. Here the HPLC trace is compared to the PAGE.

MALDI-TOF

HPLC fractions were sent for MALDI-TOF mass spectroscopy in parallel to the PAGE analysis. The power of the MALDI to differentiate oligonucleotide isomers is limited, for example the 17 dinucleotide isomers have only three possible masses (Table 2.7). Nonetheless it was hoped MALDI could help confirm HPLC peak identities. Unfortunately the sample concentration was too low to allow for the detection of less abundant species such as trinucleotides. Dinucleotides were detected in fractions collected from 12 to 17 min in samples incubated with clay (Table 2.8); there were no oligonucleotides detected by MALDI-TOF in unincubated controls or samples incubated without clay.

Table 2.7. Most common isotope single ion masses of all dinucleotide species.

Species (number of isomers)	Single ion mass of most common isotope
A_2 (2)	595.1415
A_2>p (2), ┌A_2┐ (4)	657.0972
pA_2 (2), A_2p (4), AppA (3)	675.1078

Table 2.8. Correlation of dinucleotide bands detected by PAGE and masses observed by MALDI-TOF in the fractions collected by HPLC.

HPLC Fraction	Bands in PAGE	Observed Masses
12-13 min	A_2p	675.388
13-14 min	A_2	675.445
14-15 min	A_2	657.351, 675.431
15-16 min	A_2	675.309
16-17 min	A_2, A_2>p, A_2p	675.320
17-18 min	A_2>p	675.358

Detection of cyclic RNA oligomers

Interestingly, MALDI-TOF detected a mass of 675.3786 ± 0.0557 in all fractions that show A_2 by PAGE (Table 2.8). Further, a standard of $A^{3'}A$ was shown to have a

retention of 17.0 min on the HPLC and there is no peak of that retention observed (Figure 2.17), nor is a mass of 595.1415 observed in that corresponding fraction. This is consistent with my hypothesis that pA_2 not A_2 is present in these samples. In the 14-15 min fraction a mass of 657.351 is observed but there is no A_2>p band detected in the PAGE, indicating the presence of $\ulcorner A_2 \urcorner$. To test this partial alkaline hydrolysis was performed on the 14-15 min fraction followed by 5′ ^{32}P labelling by PNK because hydrolysis of $\ulcorner A_2 \urcorner$ gives A_2>p and A_2p which label by PNK. Remarkably not only were A_2>p and A_2p shown to emerge with hydrolysis but so did A_3>p, A_3p, A_4>p and A_4p (Figure 2.18), indicating that $\ulcorner A_2 \urcorner$, $\ulcorner A_3 \urcorner$, and $\ulcorner A_4 \urcorner$ are all present in this fraction. Finding cyclic oligomers longer than $\ulcorner A_2 \urcorner$ implies there is a mechanism for cyclic polymer elongation, potentially by the attack of an A>p monomer on a cyclic polymer followed by rapid re-circularization.

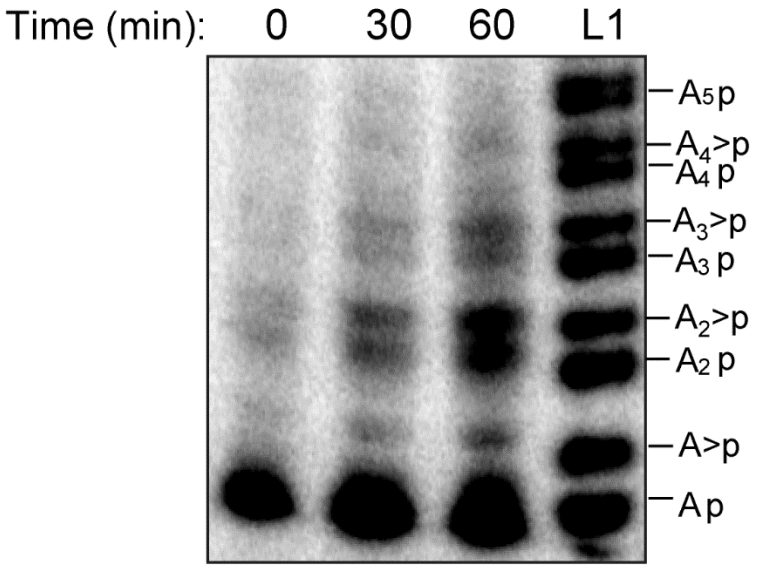

Figure 2.18. Partial alkaline hydrolysis of the 14-15 min HPLC fraction showed the emergence of A_2>p, A_2p, A_3>p, A_3p, A_4>p and A_4p over time.

2.7.4. Identification of A>p polymerization products

Most of the major HPLC peaks from the 300 mM A>p incubations with clay at 22°C for 90 days were identified based on evidence from PAGE (Figure 2.16) and MALDI-TOF (Table 2.8). The 13.9 min peak was identified as an isomer of pA_2 based on the detection of a 675.445 mass and the detection of an A_2 band by PAGE. This peak overlaps the fraction collection boundary and so can also be visualized in the neighboring lane. The peak at 14.3 min was shown by MALDI and PAGE to contain mainly $┌A_2┐$. The peak at 15.3, like the 13.9 min peak was shown to be mainly pA_2. The detection of pA_2 in two separate peaks suggests that both isomers are being produced. The peak at 16.3 showed a mixture of $A_2{}^{3'}p$, $A_2{}^{2'}p$, A_2>p bands on the gel but only the A_2p species were confirmed by MALDI. Likewise the 17.8 peak correlates to A_2>p in the PAGE but the MALDI detects a mass of 675.358. Because the A_2>p species is represented by strong bands in the PAGE but a mass of about 675.35 is detected by MALDI, it was suspected that it is being hydrolyzed to A_2p either when mixed with the MALDI matrix solution or during ionization.

Table 2.9. Identities and yields of the major HPLC peaks from incubations with 300 mM A>p incubations with clay at 22°C for 90 days.

Yields were determined by HPLC. Identities are assigned based on PAGE and MALDI evidence. A question mark indicates the identity is uncertain.

Peak (min)	Yield (%)	Identity
12.9	0.66	$A_2{}^{3'}p$
13.9	0.37	pA_2
14.3	0.08	$┌A2┐$
15.3	0.67	pA_2
15.9	0.31	?
16.3	0.46	$A_2{}^{3'}p$, $A_2{}^{2'}p$, (A_2>p?)
17.8	0.06	A_2>p?

There is evidence that HPLC is separating isomers with different phosphodiester linkages as pA_2 and A_2>p both appear in multiple fractions. Further, the commercial standard $A^{3'}A$ is known to run at about 17 min but there is no A_2 band in the PAGE in the 16-17 min or 17-18 min fractions, there is an A_2 band detected in the 22-23 min fraction

which makes it likely to be the $A^{2'}A$ isomer. This would also be consistent with the pattern of monomer retention on the C18 column, given that 2′AMP is more strongly retained than A>p and 3′AMP respectively. The proximity of the phosphate to the base decreases the retention, so it follows that $A^{3'}A$ (phosphate on the same side as the bases) would elute earlier than $A^{2'}A$ (phosphate on the opposite side as the bases). Another example of species indistinguishable by PAGE occurring in multiple HPLC fractions is A_3>p. It appears four times on the gels with varying signal intensity: 12-13 min, 15-16 min, 20-21 min and 22-23 min, intriguing since A_3>p also has four isomers. We can see oligomers of many lengths found in all fractions. The PAGE of the 14-15 min HPLC fraction hydrolysis revealed $\ulcorner A_2 \urcorner$, $\ulcorner A_3 \urcorner$ and $\ulcorner A_4 \urcorner$ were all present in the same fraction. Thus, reverse phase HPLC does not discriminate well based on oligomer length beyond separating monomers from oligomers,

Most interesting of all, the oligomers that are least strongly retained by HPLC have significantly higher signal both on the HPLC and PAGE. If the hypothesis that isomers with internal 5′ to 3′ phosphodiester bonds are less strongly retained on the HPLC than their 2′ to 5′ counterparts, then polymerization of A>p has a strong preference for producing 5′ to 3′ bonds. Given that there is a preference for the formation 3′AMP over 2′AMP during A>p hydrolysis this idea is not implausible. This would be a significant finding because it is still mysterious as to how life got its 5′ to 3′ flavour.

2.8. Conclusion

I demonstrated the RNA oligomers up to 5 nt in length could be generated by polymerization on montmorillonite clay, representing the first report of *de novo* RNA synthesis of prebiotically plausible nucleotides by a plausible mechanism. This reaction was dependent on pH, temperature and monomer concentration. The dominant oligomer product was found to be pA_2 (Table 2.9), which was surprising because the formation of this species requires the formation of a cyclic oligonucleotide then an attack by water; in contrast A_2>p was expected to be the dominant product because it is the result of one hydroxyl of A>p attacking the 2′, 3′ cyclic phosphate of another. Further, $\ulcorner A_3 \urcorner$, $\ulcorner A_4 \urcorner$, and $\ulcorner A_5 \urcorner$ were detected, suggesting a mechanism where an A>p attacks a cyclic oligomer and the resulting polymer rapidly recyclizes. Interestingly cyclic oligonucleotides were at

least as common as oligomers with 2′,3′ cyclic phosphates, indicating an equilibrium between the two forms. The polymerization of 2′,3′ cyclic nucleotides on montmorillonite clay is a promising route for the abiotic synthesis of RNA, potentially leading to the emergence of life.

In the future oligomer yields may be improved by the repeated addition of fresh N>p solution. Yields may also be increased by supressing A>p hydrolysis by lowering the temperature of the reaction, given that 4°C gave a slightly more favourable ratio of oligomers to hydrolyzed monomers compared to 22°C. It would also be interesting to determine the abundance of 3′→5′ vs 2′→5′ phosphodiester bonds; given that 3′AMP is preferred 2:1 over 2′AMP during A>p hydrolysis this system has the potential to be biased towards the formation of 3′→5′ linkages.

Chapter 3. From polymers to life

3.1. The emergence of ribozymes

RNA world theories postulate that self-replicating RNA systems were the first life on Earth (Gilbert 1986) (Joyce 2002) but the source of that RNA is debated. In Chapter 2 I demonstrated that montmorillonite is capable of catalysing the polymerization of A>p into RNA oligomers. It has been shown that RNA templates bound on hydroxyapatite can serve as a template for 5′-phosohomethylimidizole activated nucleotides (2MeImpN)(Acevedo and Orgel 1987). Similarly, ImpN will polymerize on an RNA template in eutectic ice in the presence of Mg^{2+} (Monnard and Deamer 2003) and 2MeImpG will polymerize on a polyC template in lipid vesicles (Adamala and Szostak 2013). Since RNA can be copied directly from a template RNA polymer, could the same principle apply to a non-continuous template? Could single nucleotides, spatially oriented like a polymer template act as a template for polymerization, say, if the nucleotides were closely bound on a mineral surface? For that matter can short RNA polymers, bound "sequentially" on a mineral surface, act as one long template? It has been demonstrated that two RNA strands hybridized to a template can be ligated in absence of a biological catalyst (Bowler et al. 2013).

In a mixed system, RNA polymers synthesized *de novo* could serve as the template for copying on hydroxyapatite in eutectic ice, or in lipid vesicles, provided there was a mechanism for removing the polymers from the clay. If so, longer RNA oligomers could be made by ligation, or by continuous copying across a gap between two shorter RNA oligomers; short sequences that were synthesized first would be conserved and mixed with each other in various combinations. RNA 100 nt or even longer could be generated this way. This is significant because these are long enough to be functional ribozymes; for perspective ligase ribozymes are about 60 nt long (Paul and Joyce 2002), recombinases about 130 nt long (Hayden and Lehman 2006) and polymerases about 200 nt long (Zaher and Unrau 2007). The resulting population of RNA oligomers has an important feature: it would be biased towards variations on the founding sequences (Figure 3.1).

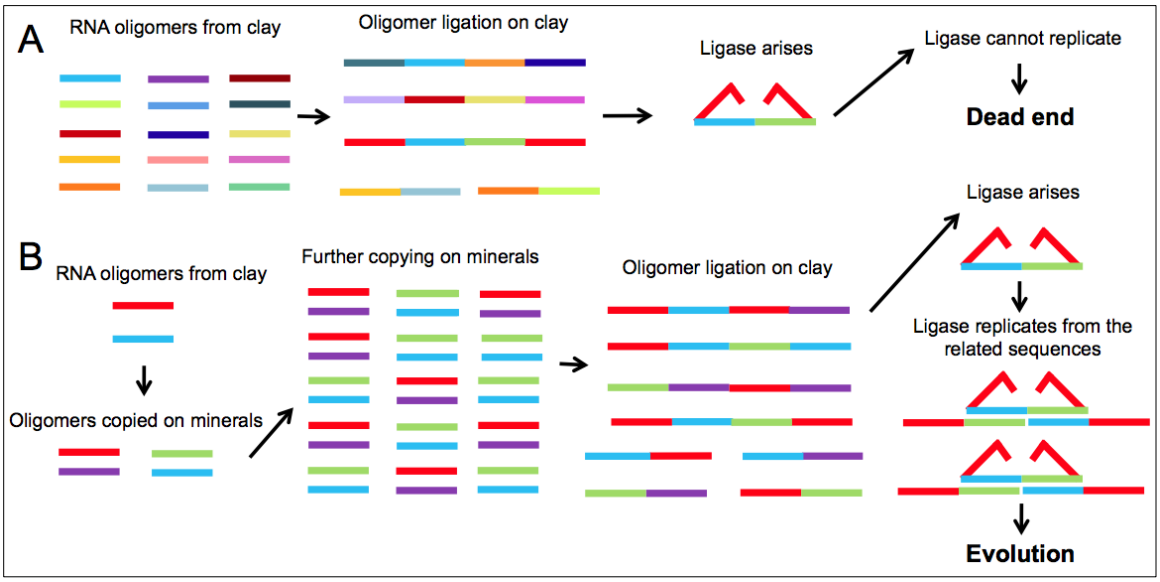

Figure 3.1. The effect of RNA population composition is important.
A) The starting oligomers are totally random so the RNA polymers formed are also random. A self-replicating ribozyme is unlikely to find complementary pieces to replicate. B) The starting oligomers were copied repeatedly: the possible polymer combinations are fewer. An emergent, self-replicating ribozyme finds more copies of its constituent pieces and replicates.

This population would be dominated by many related copies of the original starting sequences since it was made by copying and recombining the founding sequences. Thus, it is possible to make more of any full length ribozyme that arises by recombination or ligation of some of the smaller pool constituents. Consider the self-assembling ribozyme selected by Paul and Joyce (2002). The full length ribozyme T base-pairs to pieces A and B, when A and B bind T it catalyzes the formation of a 3′ to 5′ phosphodiester linkage between A and B producing a second T ribozyme. A ribozyme like this would be able to make more of itself, it may also be able to assemble pieces with some homology to A and B; potentially building other useful ribozymes such as nucleotide synthases, other ligases, recombinases or some other possibility. This is not an unlikely scenario, other ribozymes have been shown to be modular; for example, the R18 RNA polymerase ribozyme is built from the fusion of a ligase core ribozyme with an accessory domain (Johnston et al. 2001).

If a replicating ribozyme like the Paul and Joyce system had arisen from a totally random pool of RNA the likelihood that it could find even a single A and B to make another T are vanishingly small. That is to say: *de novo* synthesis of RNA isn't enough, there needs to be a mechanism for generating multiple copies of sequences to increase the likelihood of establishing a replicating population.

As convenient as the above situation is, there remains a few obvious problems. Once an RNA is copied how does it dissociate from its template strand? If the strands are immobilized on clay or other mineral surfaces how do they interact with each other? If there was a free RNA pool containing small templates and ribozymes, what would stop them from diffusing away from each other?

There are a few scenarios that address these questions, and that can incorporate the clay required to do the initial RNA polymerization. One such scenario is known as a "little warm pond", named because they are reminiscent of the little warm pond theorized by Charles Darwin as the place where life emerged. In "little warm pond" scenarios there is a small body of water, somewhat reminiscent of a tide pool, which experiences a cycle of wetting from tidal action and drying caused by solar radiation (Lathe 2005). RNA is polymerized in the pool and remains on a mineral surface. As water evaporates out of the pond the concentration of solutes increases, the high solute concentration interrupts the bond between the mineral and the RNA. The RNAs then enter into solution where they can interact and because they are trapped in the pool, the molecules have limited potential for diffusion. There is a glaring problem with this scenario: the UV radiation essential to concentrate the solutes by evaporation is harmful to RNA polymers (Cockell 1998); any ribozymes that arose could not have existed for long before being broken down by the intense solar radiation.

Drying is not the only mechanism capable of concentrating solutes; freezing can also achieve this. The freezing of salt water solutions results in the formation of eutectic ice (MacKenzie 1977). As a salt solution cools below the water can form pure ice crystals, which increases the concentration of salts in the remaining liquid depressing the freezing point of that water further (Figure 3.2). The eutectic point of a solution is where the concentration of the salt solution is no longer enough to keep liquid brine present. For a NaCl solution this point is -21°C, for $MgCl_2$ this is -34°C (Patel and Hurwitz 1972). This can happen naturally when super-cooled salt water makes contact with ocean water, such as during brinicle formation (Figure 3.2) (Paige 1970). Because the liquid water of a eutectic phase is so cold, hydrolysis of nucleotide monomers and RNA is suppressed. The half-life of an A>p monomer at pH 6 in a eutectic solution at -17°C would be 1730 years, this is 494 and 35 times longer than its half-life and 22° and 4°C, assuming the 2.2 M concentration of $MgCl_2$ in the eutectic brine does not significantly affect the nucleotide lifetime.

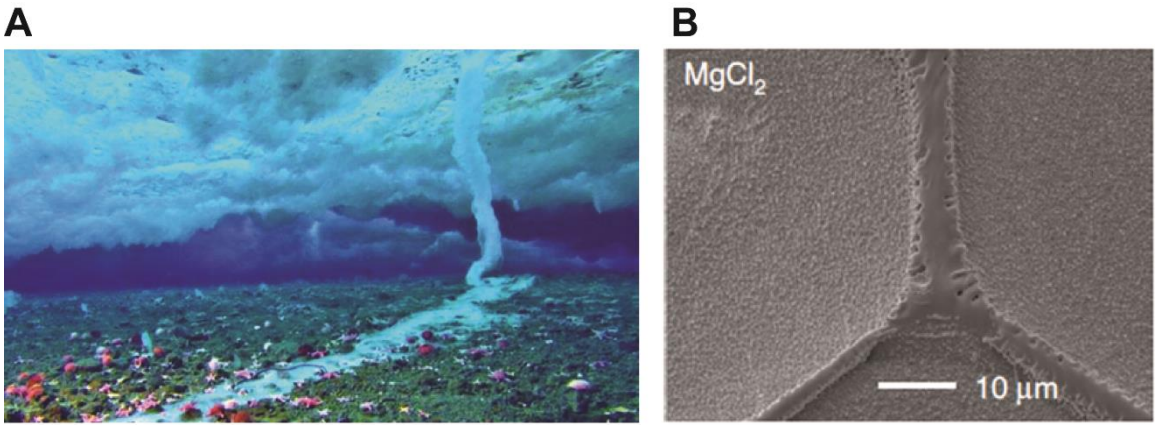

Figure 3.2. Brinicles and eutectic ice.
A) A brinicle that formed in today's arctic. The super-cooled salt water escaping from the brine channels in the sea ice above is denser than the surrounding ocean water; it sinks to the bottom, freezing everything in its path. Note the frozen trail it leaves on the ocean floor, killing extant sea stars (BBC 2011). B) Transmission electron microscopy of brine channels in a eutectic ice formed by cooling an aqueous solution of $MgCl_2$ to -17°C. Reprinted by permission from Macmillan Publishers Ltd: [Nature Communications] Atwater et al. (2010).

Excitingly, a freeze-thaw system that promotes the assembly of ribozymes has been experimentally demonstrated. A population was constructed that contained as many as 7 RNA oligomers, each containing hairpin ribozyme ligation sites and terminal 2',3' cyclic phosphates, that when assembled formed a polymerase ribozyme ~200 nt in length (Mutschler et al. 2015); This population was then cycled between -30°C and 37°C (holding at a eutectic temperature of -9°C between minimal and maximal temperatures). Full length ribozymes were detected after only a single freeze-thaw cycle when four oligomers were used and 5 freeze-thaw cycles when assembling from 7 oligomers. Polymerase ribozyme activity was detected after only 1 freeze-thaw cycle, when assembling from 4 oligomers, and was found to reach optimal performance after only 3 cycles. Assembly from four non-activated (i.e. 2', and 3' OH groups) RNA oligomers with 7-8 nt of additional sequence (cleavable 'tails') was attempted (Figure 3.3). In this case full length ribozymes were detected after 3 freeze-thaw cycles and reached near maximal yield (~10%) after 6 cycles. Polymerase ribozyme activity was detected after 3 cycles and reached peak performance after 12 cycles. This finding is exciting because the oligomers are not only able to ligate when pre-activated, but can be activated and ligated in the same system. No assembly was detected in samples that were not freeze-thaw cycled. Combining this freeze-thaw mediated ribozyme assembly with the synthesis of RNA by minerals (Figure 3.1) could have generated the first ribozymes and stared the RNA World.

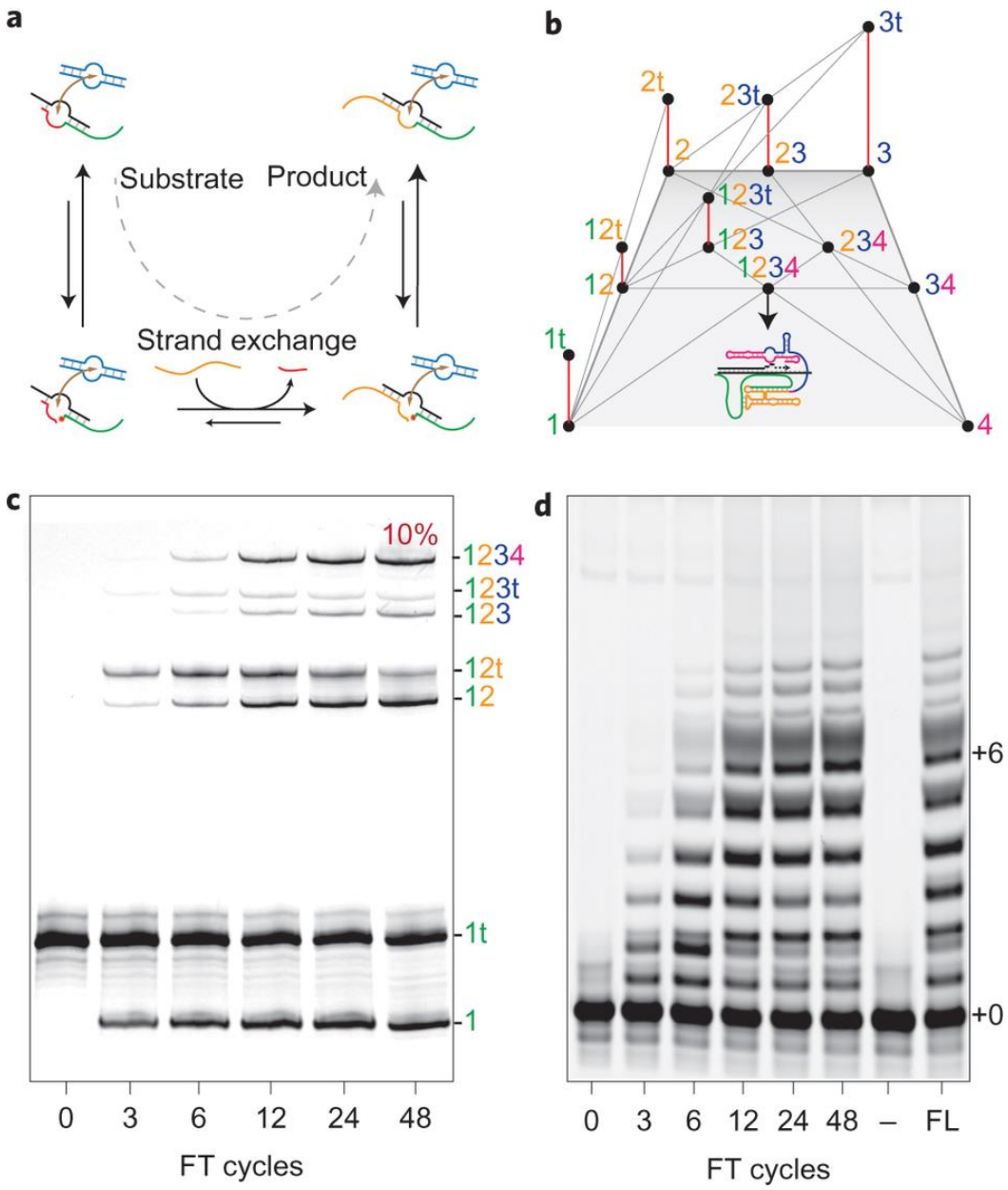

Figure 3.3. Freeze–thaw cycles as drivers of complex ribozyme assembly.
A), Schematic representation of the assembly trajectory that involves (anticlockwise from top left) cleavage of a short 3′ tail (red) to generate a >p (red dot), dissociation of the cleaved tail and strand exchange to the cognate substrate (orange) followed by ligation of substrate 5′-OH with >p. **B)** Network diagram of the RNA polymerase ribozyme (RPR4) assembly from four tailed fragments. The tailed input fragments can ligate to their cognate 5′ fragments, but have to be cleaved (red lines) before ligation to 3′ fragments. **C)** FT-cycle driven assembly of RPR4 by hairpin ribozymes from four inert fragments (1 μM) that proceeds from the fluorescently labelled 5′ fragment (1t) through several intermediates to the full-length RPR4 (1234). **D)** Primer extension activity of the reaction mixtures shown in C. The primer-only control (–) and the activity of full-length RPR4 from *in vitro* transcription (FL) are shown for comparison. Reprinted by permission from Macmillan Publishers Ltd: [Nature Chemistry] Mutschler et al. (2015).

As an alternative to a “little warm pond” I propose a cold muddy ocean (Figure 3.4). In this scenario a prebiotic body of water is lined with clay and minerals. Nucleotides in the water bind to clay and polymerize into RNA (Ferris *et al.* 1996). As the water freezes dissolved solutes are concentrated between the ice crystals, increasing salt concentration and destabilizing interactions between the clay. The RNA dissociates and diffuses in the brine channels between the ice crystals. Trapping the RNA between the ice crystals acts a precellular compartmentalization mechanism, limiting the diffusion of RNA. When the solution thaws water from the ice crystals dilutes the brine. The dilution of the salt allows the RNA to bind the clay or depending on diffusion, a hydroxyapatite or other mineral crystal. RNA bound to hydroxyapatite acts as a template for reverse complimentary copying. On the clay re-bound RNA may be lengthened further and new RNA is polymerized. Further freeze-thaw cycles would facilitate RNA replication, ligation, recombination and interaction in brine channels. Copying of the existing sequences on the hydroxyapatite generates multiple copies of some sequences. Copying could also take place in the eutectic phase. Eventually populations of related sequences emerge; this would greatly increase the probability of an emergent ribozyme that could make more of itself from pool constituents.

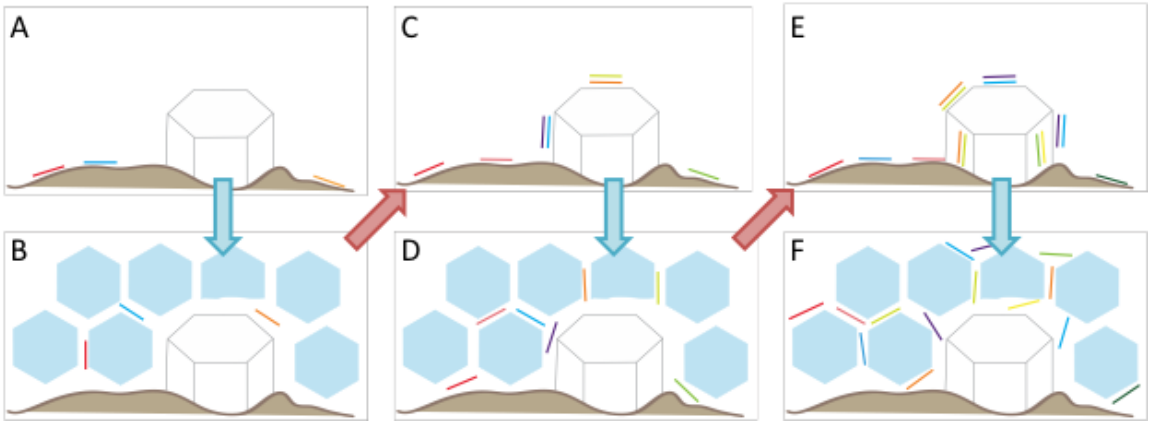

Figure 3.4. A cold, muddy scenario for the emergence of RNA.
A) A prebiotic body of water is lined with clay (brown) and minerals like hydroxyapatite (white crystal). Nucleotides in the water polymerize into RNA (coloured strands) on the clay. B) As the water freezes, the RNA dissociates and diffuses in the brine channels between the ice crystals (blue). C) When the solution thaws water from the ice crystals dilutes the brine, allowing the RNA to bind the clay or a hydroxyapatite crystal. RNA bound to hydroxyapatite acts as a template for reverse complementary copying. On the clay re-bound RNA may be lengthened further and new RNA is polymerized. D-F) Further freezing and thawing cycles facilitate RNA replication, recombination and interaction in brine channels. Copying of the existing sequences on the hydroxyapatite generates multiple copies of some sequences (Purple, blue, green and yellow strands). Eventually pools of related sequences emerge.

Chapter 4. In vitro selected ribozymes for the re-creation of the RNA world

This chapter was previously published in Life (Martin et al. 2015). I was the lead author and contributed figures 4.2, 4.3 and 4.6 as well as text in sections 4.2, 4.2.1, 4.2.3 and 4.3. Peter J. Unrau and Ulrich Muller were my coauthors. Peter contributed text throughout. Ulrich contributed figures 4.1, 4.4 and 4.5 as well as text in 4.1, 4.2, 4.2.2 and 4.3. All authors contributed to manuscript editing.

4.1. *In vitro* selection

Though the existence of an early RNA world is well-supported, several central problems remain unsolved. An important step in validating the RNA World hypothesis would the demonstration in the lab of a self-replicating and evolving system based on catalytic RNAs (Joyce 2002).

Using the method known as *in vitro* selection (Tuerk and Gold 1990) (Ellington and Szostak 1990) more than a dozen research groups have demonstrated that catalytic RNAs (ribozymes) have the potential to catalyze the diverse chemical reactions required to sustain a metabolism (for a review, see (Chen et al. 2007)). Such *in vitro* selections start from large, combinatorial libraries of RNA that are incubated with their substrate / ligand molecules (Figure 4.1). For the *in vitro* selection of catalytic RNAs (Bartel and Szostak 1993), one of the substrate molecules usually contains a handle, which is covalently coupled by the ribozyme to itself. This self-tagging allows reacted RNA molecules to be isolated from the library. Subsequently, the isolated RNA is reverse transcribed into DNA, amplified by PCR, and transcribed into an RNA pool that is now enriched for functional sequences. Multiple cycles of this selection and amplification scheme are necessary to obtain libraries from which active sequences can be isolated by analyzing individual clones. Note that some *in vitro* selections employ more complicated schemes for isolating active sequences from the library instead of a simple handle on the substrate (for examples, see (Zaher and Unrau 2007) and (Wochner et al. 2011)) or bypass the step of reverse transcription (Zaher and Unrau 2007). With this method of *in vitro* selection,

ribozymes were developed that could catalyze, among many others, reactions that include RNA ligation, peptide coupling, Diels-Alder bond formation, redox reactions and recently decarboxylation ((Bartel and Szostak 1993), (Zhang and Cech 1997), (Seelig and Jaschke 1999), (Tsukiji et al. 2003), (Cernak and Sen 2013); for a review, see (Chen et al. 2007)). While the catalytic rate enhancements of ribozymes are usually lower than those of highly evolved protein enzymes, the repertoire of reactions catalyzed by ribozymes appears sufficient to mediate a complex metabolism. This supports the plausibility of an RNA World during early stages of life (Chen et al. 2007). This review focuses on a subset of *in vitro* selected ribozymes whose activities could directly mediate the synthesis of RNA from simpler building blocks.

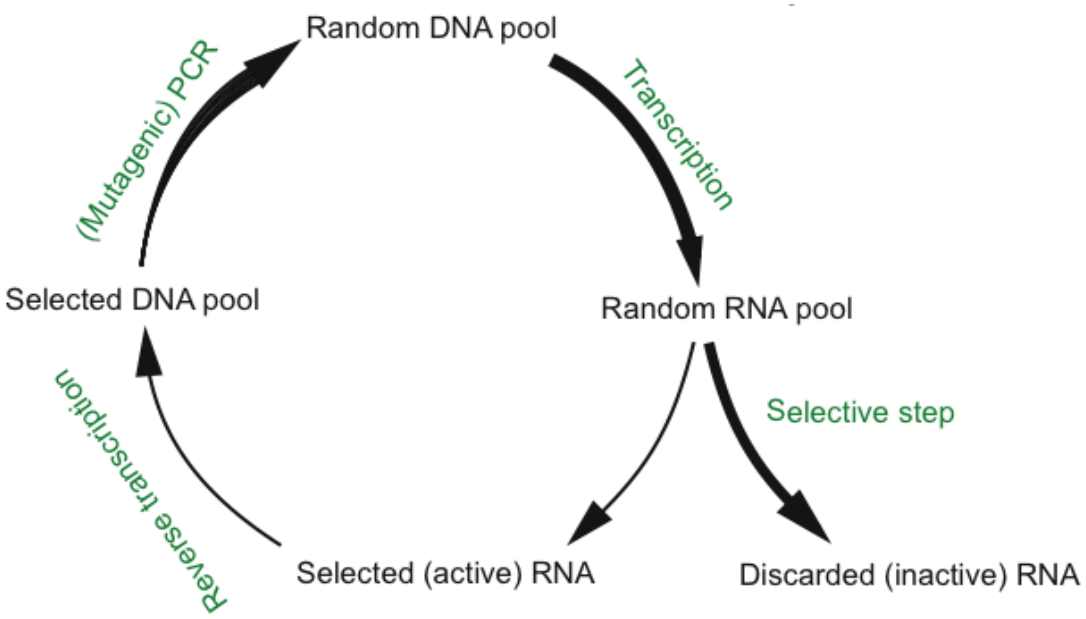

Figure 4.1.Schematic for the *in vitro* selection of catalytic RNAs.
A pool of double-stranded DNA molecules containing random sequence is transcribed into the corresponding RNA pool. Due to the differences in sequence and self-complementarity, the individual sequences of the RNA pool fold into distinct three-dimensional structures. To identify the tiny proportion of these structures that are able to catalyze the reaction with a substrate the RNA pool is incubated with the substrate and the unreacted RNAs are removed in a selection step. The isolated RNAs are then reverse transcribed into DNA molecules, and amplified by PCR to obtain a DNA pool that is now enriched for sequences that are able to catalyze the desired reaction. Because the complexity of the initial DNA pool (usually 10^{14}-10^{16}) vastly outnumbers the enrichment of active sequences per selection cycle (usually 10 - 10^4-fold) it is necessary to conduct multiple cycles before a detectable fraction of the pool shows catalytic activity.

4.2. RNA synthesis in an RNA World

The first catalytic RNA must have arisen in an environment that already executed all steps in a chemical pathway from prebiotically available molecules to RNA polymers - otherwise this catalytic RNA could not have existed. The chemistry of RNA polymer synthesis before the emergence of RNA World catalysts is poorly understood and has been proposed to occur in several alternative ways. Nucleotide synthesis has been suggested to have proceeded via the glycosidic bond formation between ribose and nucleobases (Chen et al. 2013), (Fuller et al. 1972) or the stepwise assembly of a nucleobase on ribose (Sanchez et al. 1966), (Powner et al. 2009). Many routes of prebiotic nucleotide polymerization have been explored. Some of these routes do not require activation groups on the nucleotides: such as the formation of polymers by the dehydration of nucleoside 5′-monophosphates in lipid environments (Rajamani et al. 2008), and the formation of dinucleotides by the reaction of adenosine 2′,3′-cyclic phosphate on poly(U) templates (Renz et al. 1971). The latter of these studies analyzed the balance between dinucleotide formation and hydrolysis under different conditions, illustrating the thermodynamic challenges of RNA polymerization. In aqueous solution, RNA polymerization is entropically disfavoured, making activating groups necessary for the production of long RNA polymers. A range of 5′-phosphate activating groups with variable prebiotic plausibility have been investigated: adenylate (Prabahar and Ferris 1997), cyanide (Lohrmann and Orgel 1968), imidazole (Sawai and Orgel 1975), 2-methyl imidazole (Inoue and Orgel 1982) and triphosphates (Etaix and Orgel 1978). While these forms of chemical activation have provided insight into RNA polymerization, polyphosphates are considered the most prebiotically likely activation groups (Schwartz 2006), (Robertson and Joyce 2012). The polymerization of activated nucleotides into RNA polymers has been explored on the surface of clay minerals (Huang and Ferris 2006) or using cations such as Zn^{2+} (Sawai and Orgel 1975). It is currently unclear which of these possibilities, if any, preceded the chemical pathways of the RNA World. The first RNA world organisms likely used existing abiotic chemical pathways and improved on their efficiency and/or selectivity by catalyzing rate-limiting reactions. Later evolutionary stages of RNA World organisms may have modified these pathways to use better suited metabolites essential for RNA replication (Fig. 2A). With the onset of the RNA World, vesicle-encapsulated aqueous droplets would have served as the center for an RNA

metabolism sustained by ribozymes encoded by an RNA genome (Bartel and Unrau 1999), (Chen et al. 2004), placing constraints on the chemistry and kinetic stability of nucleotide activation groups in aqueous solution.

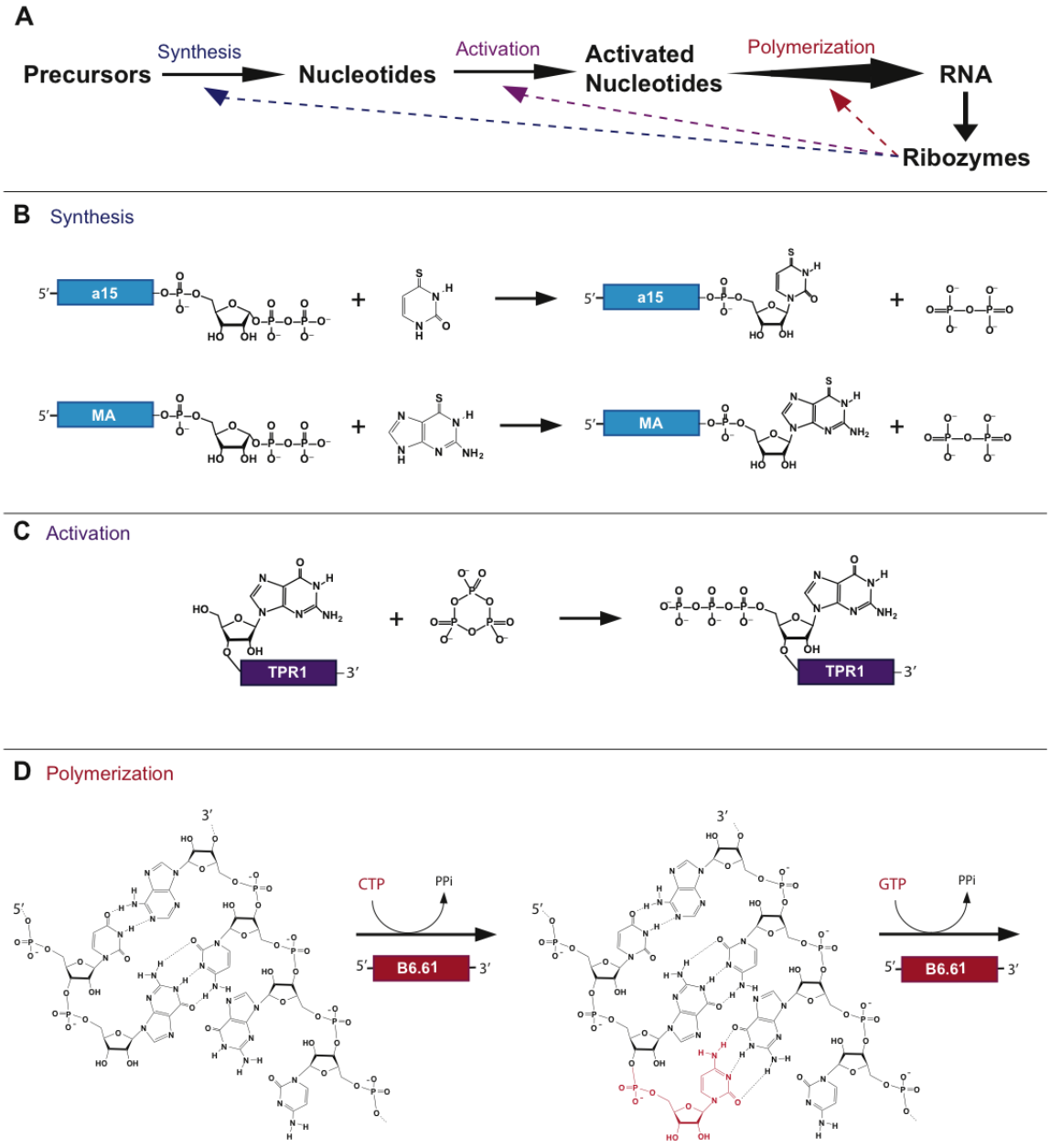

Figure 4.2. Steps in the synthesis of RNA before and during the RNA World.
A) Prior to the emergence of RNA catalysts, abiotic processes led to the generation of RNA polymers (black arrows). After the emergence of RNA catalysts some or all of these reactions were catalyzed by ribozymes (dotted lines). **B)** Reactions catalyzed by nucleotide synthase ribozymes. The a15 ribozyme promotes the synthesis of ^{4S}UMP from free ^{4S}Ura and PRPP tethered to its 3′ terminus. Similarly the MA ribozyme catalyses the synthesis of ^{6S}GMP from free ^{6S}Gua and tethered PRPP. **C)** The TPR1 ribozyme promotes the triphosphorylation of its 5′ terminus with trimetaphosphate. **D)** Template dependent RNA polymerization is catalyzed by the B6.61 RNAP ribozyme.

The takeover of prebiotic chemistry by RNA world ribozyme catalysis increased the rate and specificity of RNA synthesis, thereby allowing faster replication of the RNA world organisms (Figure 4.2). To recapitulate an RNA World organism in the lab, three types of ribozymes have been identified by *in vitro* selection methods, and characterized: Ribozymes that synthesize nucleotides from activated ribose and nucleobases ((Unrau and Bartel 1998), (Lau et al. 2004)) (Figure 4.2), ribozymes that convert RNA 5′-hydroxyl groups to 5′-triphosphates (Moretti and Muller 2014) (Figure 4.2), and ribozymes that catalyze the template-dependent polymerization of nucleoside 5′-triphosphates to RNA polymers (Ekland and Bartel 1996), (Johnston et al. 2001) (Figure 4.2). These three types of ribozymes are the focus of this review. All of these ribozymes use phosphate activation groups to mediate steps in a pathway for RNA synthesis that may have existed in the RNA world.

4.2.1. Nucleotide synthesis ribozymes

A nucleotide is a fusion of three discrete components: a phosphate group, a ribose sugar, and a nucleobase. These elements combine to give RNA its information encoding and catalytic properties. The preferred conformations of ribose together with the 5′-3′ phosphodiester linkages formed by polymerization stabilize purine and pyrimidine nucleobase pairs. The chemistry, size, and geometry of the resulting polymers is fine-tuned for the storage and transmission of genetic information via base pairing (Krishnamurthy 2012). Nucleoside triphosphates, the activated monomers used in biology for RNA polymerization, are also used metabolically, as central sources of cellular free energy. Together with canonical base pairs, abundant alternative purine and pyrimidine hydrogen bonding patterns (Leontis et al. 2002), and sugar-base conformations make possible a broad range of alternative RNA structures which are required for the formation of complex catalytic RNAs (Cate et al. 1996), (Krasilnikov et al. 2003), (Guo et al. 2004). The discrete elements of a nucleotide beg the question: did the synthesis of nucleotides in an RNA World reflect the three-fold modularity found in modern metabolism?

In modern organisms, 5-phosphoribosyl 1-pyrophosphate (PRPP) is essential for all nucleotide synthesis. PRPP, which is obtained from ribose 5-phosphate (R5P) and ATP, brings together phosphate based energy metabolism with ribose carbohydrate

chemistry. PRPP then reacts with nucleobases that are either scavenged or synthesized by nitrogen dependent metabolic pathways to yield a nucleotide product linking R5P and a nucleobase. Both the information containing and catalytic cofactor type nucleotides are synthesized by this strategy. Purine nucleobases such as adenine, guanine, xanthine, hypoxanthine, and 6-thioguanine can be reacted in a single step with PRPP to produce their corresponding nucleotide monophosphates (NMPs). Alternatively, the first step in *de novo* purine nucleotide synthesis aminates the 1-position of ribose 5-phosphate (R5P) using PRPP as substrate (Kanehisa et al. 2014). Pyrimidine nucleotides such as UMP can be produced directly by reacting PRPP with uracil, but are primarily produced via the orotate dependent pathway, where orotodine is produced in a single concerted reaction. Subsequent decarboxylation at the 5 position then yields UMP, and can be followed by amination at the 4 position to produce CMP. Likewise pyrimidine nucleobases such as nicotinamide and nicotinate are used to synthesize NAD^+, an essential redox cofactor. PRPP considerably simplifies nucleotide synthesis in modern metabolism by providing two of the three essential elements of a nucleotide in the form of activated R5P.

Modern biology′s dependence on PRPP, and the advantages of modular RNA nucleotide synthesis with PRPP suggest that this molecule might have played an important role in a potential RNA World. However, promoting PRPP dependent nucleotide synthesis chemistry with RNA presents a potential problem: both PRPP and the nucleobases used to make nucleotides are small relative to the size of the catalytic RNA itself, making it unclear if RNA is up to the task of nucleotide synthesis. Specifically, it is a challenge for RNA to position its functional groups necessary for specific ligand binding and catalysis of a reaction into the tight space of a binding pocket for a small molecule. While riboswitches and RNA aptamers that specifically bind and recognize adenine, guanine, and derivatives are now well-understood (McCown et al. 2014), (Batey et al. 2004), (Serganov et al. 2004) few naturally occurring RNAs are known that chemically manipulate small substrates. Among these few is the GlmS glucosamine-6-phosphate (G6P) ribozyme, which utilizes bound G6P to self-cleave by promoting an acid-base mediated cleavage reaction (Klein and Ferré-D′Amaré 2006). The combination of small substrate binding and catalysis may be possible for the GlmS ribozyme because self-cleavage is intrinsic to all RNAs: It only requires a mechanism to deprotonate and position a 2′ hydroxyl in line with its adjacent phosphodiester bond; an event that proceeds at

measurable rates at most positions in almost all RNA molecules. Given the limited chemical repertoire of the ribozymes that manipulate small molecules, the isolation of ribozymes in the laboratory provides one of the few tools to study the potential of RNA to perform reactions like nucleotide synthesis that could have played an essential role in an RNA World.

Encouragingly *in vitro* selections for pyrimidine and purine nucleotide synthesis have demonstrated the potential for RNA to promote nucleotide synthesis (Figure 4.2). Selections were designed where large pools (~10^{15} distinct sequences) of random sequence RNA were ligated to PRPP via their 3′ termini. These pools were then incubated with either 4-thiouracil (Unrau and Bartel 1998), (Wang and Unrau 2005) or 6-thioguanine (Lau et al. 2004). Those RNAs able to promote glycosidic bond formation between tethered PRPP and thiolated nucleobase were purified using polyacrylamide gels derivatized with *N*-acryloylaminophenylmercuric acetate (Igloi 1988). Purine nucleotide synthase pool populations were found to be some 50 to 100 times more efficient than the equivalent pyrimidine nucleotide synthase ribozymes (Lau et al. 2004). Consistent with the bulk pool reaction rates, the three most common pyrimidine nucleotide synthase families named A, B and C had efficiencies of 4.3, 1.3, and 0.7 $M^{-1}min^{-1}$, while the two most abundant purine nucleotide synthase families named RA and MA had efficiencies of 230, and 284 $M^{-1}min^{-1}$ respectively. Only the pyrimidine family A and purine family RA had measurable K_m values with the remaining families having reaction rates directly proportional to nucleobase concentration. This suggested that mechanistically, ribozymes isolated from purely random sequence can invest in either substrate binding or in chemical rate enhancement but have difficulty optimizing both binding and rate acceleration simultaneously. Consistent with this hypothesis reselection and truncation of the pyrimidine nucleotide synthase family A ribozyme improved chemical efficiency from 4.3 to 150 $M^{-1}min^{-1}$ (Chapple et al. 2003) and it is possible that similar gains could be obtained for purine nucleotide synthase ribozymes should such reselections be attempted in the future.

Whether or not a nucleotide synthase ribozyme had measurable nucleobase binding affinity, all families of nucleotide synthetase ribozymes were extremely sensitive to nucleobase modification and showed surprising parallels to the substrate preferences

of highly evolved protein nucleotide synthase enzymes. For example, the family A pyrimidine nucleotide synthetase ribozyme was ~10,000 times slower with uracil than with 4-thiouracil and reactivity with 2-thiocytosine, 2-thiopyrimidine, 2-thiopyridine and 5-carboxy-2-thiouracil could not be detected (Figure 4.3). In contrast, the MA purine nucleotide synthetase ribozyme was 600 to 3,000 times slower when incubated with 6-thiopurine than with 6-thioguanine; while the RA purine nucleotide synthase ribozyme was 5 to 10 times slower still. As these purine nucleotide synthetase ribozymes were much slower with other purine modifications this indicates a lack of substrate discrimination at the 2 position. A similar situation is seen with the protein enzyme HGPTase, which is also unable to discriminate between guanine and hypoxanthine (Craig and Eakin 2000). Similar patterns are seen in guanine binding riboswitches (Mandal et al. 2003). That two purine nucleotide synthase ribozymes share patterns in common with these naturally selected protein and RNA systems suggests that a single optimal strategy to specifically recognize guanine by hydrogen bonding and stacking interactions exists in nature.

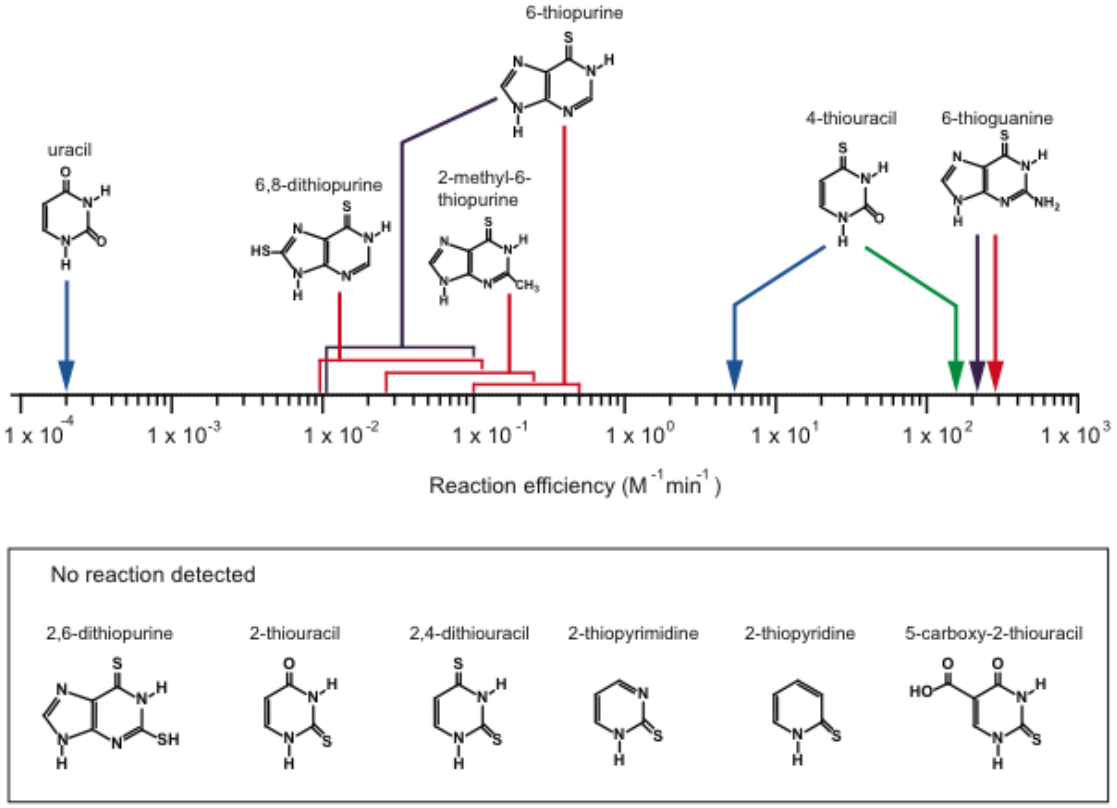

Figure 4.3. Pyrimidine and purine nucleotide synthase ribozymes show good substrate discrimination.

The arrows point to reaction efficiencies for ribozymes with the indicated substrate molecule: red arrows indicate the MA purine nucleotide synthetgase ribozyme reaction efficiencies, purple indicates the RA purine nucleotide synthethase ribozyme, blue indicates the Family A pyrimidine nucleotide synthethase ribozyme and green indicates the improved a15 ribozyme derived from the Family A ribozyme by reselection. In the box nucleobases for which no activity was detected are shown.

The differences in ribozyme efficiencies strongly suggest that pyrimidine nucleotide synthesis is the more challenging chemistry compared to purine nucleotide synthesis. This is consistent with the disfavoured kinetics and thermodynamics of pyrimidine glycosidic bond formation relative to that of purine nucleotide synthesis (Lau et al. 2004). Additionally, the large difference in K_m between pyrimidine and purine nucleotide synthase ribozymes (28 +/- 4 mM for family A and, 78 μM for family RA) suggests that 6-thioguanine is considerably easier to bind (presumably via stacking interactions to the purine base) than 4-thiouracil. Consistent with these barriers to pyrimidine glycosidic bond

formation, kinetic isotope measurements performed on the family A pyrimidine nucleotide synthase ribozyme point to an unusually dissociative reaction mechanism, where a positively charged oxocarbenium-ion is stabilized presumably by the negatively charged phosphodiester backbone of the ribozyme (Unrau and Bartel 2003).

Given that ribozymes can easily be found that promote nucleotide synthesis chemistry, how might new small molecule chemistry have evolved in an RNA World? If *in vitro* selected ribozymes are modular in nature, then it might be possible to convert a pyrimidine nucleotide synthetase into a purine nucleotide synthase by varying one subdomain, the substrate recognition domain, and conserving the other, catalytic domain. A high diversity set of RNAs whose sequences where derived from that of the family A pyrimidine nucleotide synthetase was constructed to test this hypothesis explicitly, but found no evidence for functional modularity (Lau et al. 2004). Instead the secondary structure of the family A ribozyme had been disrupted by a series of point mutations able to form completely new RNA folds. This complete change in RNA structure provides additional support for the theory of ′neutral networks′ where high densities of distinct RNA folds (and hence functionalities) are highly proximal in sequence for RNAs of sufficiently small size space. This allows the efficient evolution of RNA functionalities with domains that are sufficiently small (Schultes and Bartel 2000). Later in evolution these domains can be combined to generate multidomain ribozymes such as the ribosome or, as we will see later, convert an RNA ligase into an RNA polymerase (Lehman and Unrau 2005).

Ribozyme substrate promiscuity may provide an additional mechanism to efficiently evolve new ribozyme functions rapidly. A ribozyme able to react with two alternative substrates can in principle be evolved into two distinct ribozymes, each optimal for its own substrate. Therefore, promiscuity could be exploited to rapidly populate an RNA world with activities that can be related to each other by promiscuity. Such an evolutionary mechanism is particularly interesting if ribozymes able to ′metabolize′ simpler RNA relevant substrates can, via promiscuity, naturally evolve to utilize substrates that carry a higher free energy potential and are more similar to today′s metabolites than the initial substrates. Experimental support for such a model was found unexpectedly by the selection of a ribozyme able to react tethered R5P to 6-thioguanine via an aldehyde dependent chemistry (Lau and Unrau 2009). Surprisingly, changing the R5P substrate to

PRPP allowed efficient nucleotide synthesis chemistry to occur with 6-thioguanine, even though this nucleotide synthesis chemistry had never been selected for. The directionality of this promiscuity was not general in the sense that ribozymes selected for their PRPP dependent nucleotide synthase ability failed to react with R5P. Since R5P and PRPP are fundamental to modern RNA metabolism and are likely to have been important in an RNA world, it is interesting to speculate how many other examples of such promiscuity can be found with other important biomolecules in the future.

4.2.2. Triphosphorylation ribozymes

Nucleotides require chemical activation to drive the entropically disfavorable RNA polymerization in aqueous medium. All known life forms use nucleoside 5′-triphosphates as energy currency, therefore the use of nucleoside 5′-triphosphates by RNA World organisms would fit well with their ancestry to today′s life forms. Assuming that nucleosides were synthesized on prebiotic Earth, how might they have been converted to 5′-triphosphates? Long, linear polyphosphates are ubiquitous in microbes and animals, serve multiple biological roles, and are a potential remnant of the RNA world (Kornberg et al. 1999). Polyphosphates are prebiotically plausible activation groups of nucleoside 5′-phosphates partially because polyphosphate-activated nucleotides are kinetically stable in aqueous solution (Orgel 2004). Cyclic trimetaphosphate (Tmp) is the most active polyphosphorylating reagent of all polyphosphates (Feldmann 1967) and it is one of the more abundant polyphosphates that are generated by three prebiotically plausible synthetic routes (Osterberg and Orgel 1972), (Yamagata et al. 1991), (Pasek et al. 2013), most importantly by the synthesis from the meteoritic mineral Schreibersite, or $(Fe_3,Ni_3)P$ (Pasek et al. 2008). Tmp reacts with adenosine at room temperature to produce about 5% 5′-triphosphorylated adenosine at a pH of 12 (Etaix and Orgel 1978). The phosphorylation of the 2′- and 3′-hydroxyl groups is much more efficient (Saffhill 1970) because these hydroxyl groups have pK_A values in the range of 12.3 while the 5′-hydroxyl groups has a pK_A around 15 (Izatt et al. 1966) (Stewart 1985) but 2′- or 3′-triphosphorylated nucleosides quickly decompose to form 2′,3′-cyclic phosphates and pyrophosphate (Robertson and Joyce 2012). An additional limitation of the yield in triphosphorylation reactions appears to be the hydrolysis of Tmp at high pH (Kura 1987). The biggest problem with the necessity for pH 12 in this nucleoside triphosphorylation reaction is that RNA polymers hydrolyze

rapidly at this pH. Therefore, the triphosphorylation of nucleosides in an RNA world organism would require a catalyst that facilitates the reaction near neutral pH.

To obtain ribozymes that catalyze the triphosphorylation of RNA 5′-hydroxyl groups near neutral pH (Figure 4.2), an *in vitro* selection from a diverse library (~10^{14}) of RNA sequences was performed (Moretti and Muller 2014). The RNA library was challenged to triphosphorylate their 5′-hydroxyl groups using trimetaphosphate. To facilitate this, the RNA pool molecules were prepared with a 5′-hydroxyl group with the help of a cis-cleaving hammerhead ribozyme (Milligan et al. 1987). After incubation of the RNA pool with trimetaphosphate, pool molecules that affected their own 5′-triphosphorylation were isolated by the use of a ligase ribozyme (Rogers and Joyce 2001). The ligase ribozyme linked 5′-triphosphorylated RNAs to the 3′-terminus of a biotinylated RNA oligonucleotide, which allowed the capture of active RNAs by streptavidin. Active clones were isolated after 5 cycles of the selection, and all but two of the 36 clones isolated (16 after 5 cycles, and 20 after 8 cycles of the selection) catalyzed the triphosphorylation reaction. Most of these sequences were unrelated to each other, suggesting that on the order of 100 different triphosphorylation ribozymes were obtained in this selection. Because the effective pool size in the selection was around 10^{14} this suggests that about 1 in 10^{12} random sequences with a length of 150 nucleotides is able to catalyze the triphosphorylation of their 5′-hydroxyl groups. The ribozyme-catalyzed reaction, at 1.6 M^{-1} min^{-1} under optimal conditions (100 mM Tmp, 500 mM total $MgCl_2$, pH 8.1) was about 10^7-fold faster than the uncatalyzed reaction. These results showed two important points regarding the RNA world: First, catalytic RNAs are able to use Tmp for the triphosphorylation of RNA 5′-hydroxyl groups, which demonstrated that RNA world organisms could have used Tmp as an energy source. Second, triphosphorylation ribozymes are relatively frequent (~ 1 / 10^{12} in our library), apparently about 20-fold more than ribozymes catalyzing RNA ligation (Bartel and Szostak 1993). This suggests that triphosphorylation ribozymes were at least as accessible to RNA world organisms as ligase ribozymes, which are also required in an RNA world (see next chapter).

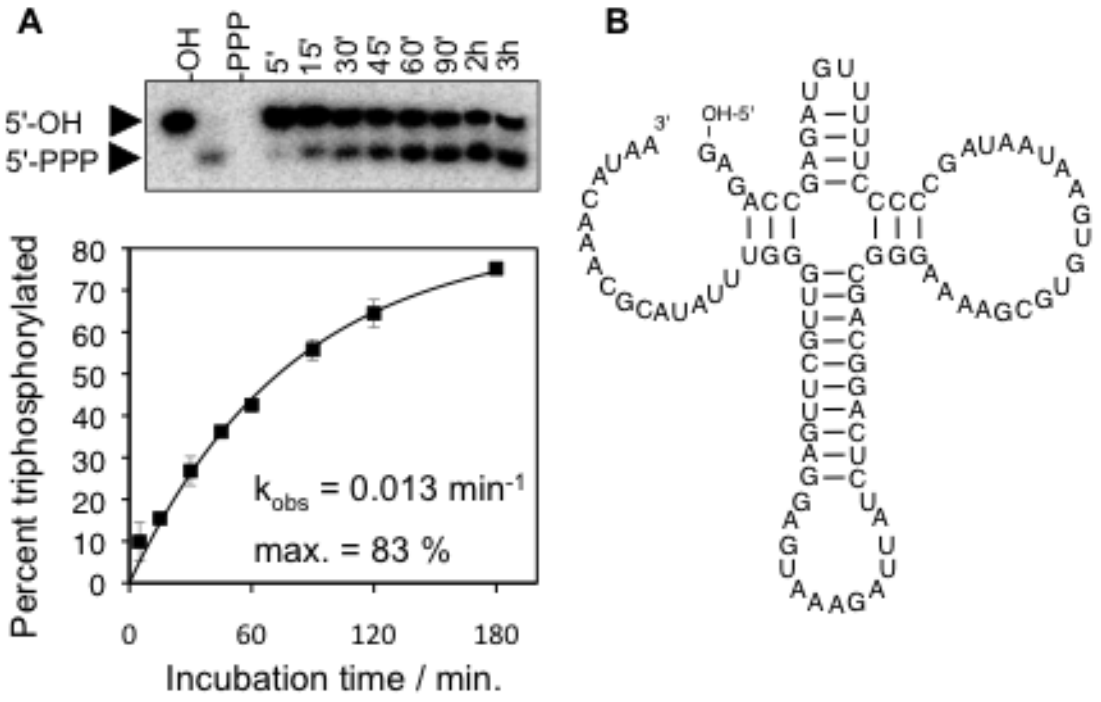

Figure 4.4. Reaction and secondary structure of a triphosphorylation ribozyme.
A) Gel shift assay of the triphosphorylation reaction at different reaction times with Tmp (5 minutes to 3 hours). An 8-nucleotide fragment is cleaved from the 5′-terminus of the ribozyme after incubation with Tmp. The short length of this fragment allows separating the fragments with a 5′-terminal hydroxyl group (5′-OH) and with a 5′-triphosphate (5′-PPP). The percent of the fragment that are triphosphorylated are plotted as a function of the incubation with trimetaphosphate, which allows determining the single-exponential reaction rate. **B)** Secondary structure of TPR1 resulting from Shape probing and base covariation experiments.

Eight of the selected triphosphorylation ribozymes were analyzed for their reaction kinetics, and had rates between 0.26 M^{-1} min^{-1} and 0.56 M^{-1} min^{-1}, under selection conditions (50 mM Tmp, 100 mM total $MgCl_2$, pH 8.3) (Moretti and Muller 2014). One of the ribozymes reacted to 83% and was analyzed in more detail. It was truncated from 182 nucleotides to 96 nucleotides while maintaining full activity. This truncated ribozyme was termed TPR1 (triphosphorylation ribozyme 1). A modification at its 5′-terminus allowed the reaction to proceed *in trans*, allowing a 14-nucleotide long RNA oligonucleotide to be triphosphorylated at its 5′-terminus by the ribozyme. This allowed the confirmation of the triphosphorylated product by mass spectroscopy. The dependence of the rate on the Mg^{2+}

concentration suggested that each Tmp molecule coordinated one Mg^{2+} ion in a bidentate fashion and a second Mg^{2+} ion with the remaining negatively charged oxygen. The pH dependence of the reaction kinetics showed that a single deprotonation step was rate-limiting for the reaction, presumably the deprotonation of the 5′-hydroxyl group that made the nucleophilic attack on the trimetaphosphate. The secondary structure of the triphosphorylation ribozyme was analyzed using SHAPE analysis (Mortimer and Weeks 2007) and base covariation analysis (Figure 4.4). The ribozyme appears to be highly compact, with the help of a 4-way helical junction. It is currently unclear how this 4-way helical junction is arranged in three dimensions, and how Tmp is bound by the ribozyme.

Six of the isolated triphosphorylation ribozymes were tested for their ability to triphosphorylate free nucleosides, instead of RNA oligomers (Moretti and Muller 2014). None of them showed activity for the triphosphorylation of free nucleosides (^{14}C-labeled guanosine). However, the triphosphorylation of free nucleosides may not have been necessary for an RNA world organism: Only if RNA polymerization proceeds in 5′- to 3′-direction is it necessary to triphosphorylate free nucleosides (Figure 4.5). If one allows for the idea to polymerize in 3′- to 5′-direction (McGinness et al. 2002), (Lutay et al. 2007) then the triphosphorylation of an RNA primer by a ribozyme would prepare the 5′-terminus for the addition of a free nucleoside. Polymerization in 3′- to 5′- direction would then occur by the alternating triphosphorylation of the RNA 5′-terminus and the addition of a nucleoside (Figure 4.5). This mechanism has not been observed in today′s life forms; in today′s biology it is probably more beneficial to utilize nucleoside triphosphates and thereby proceed in 5′-3′ direction because nucleoside triphosphates are also used as energy currency to power a large diversity of metabolic reactions. In contrast, the simpler metabolism in the earliest RNA world organisms would have made freely diffusing energy equivalents less beneficial and placed a higher reward on efficient RNA polymerization. In the RNA world, RNA polymerization without nucleoside triphosphates could even have had an advantage: Nucleoside triphosphates carry several negative charges, which cause charge-charge repulsion with the templating RNA strand. In contrast, non-phosphorylated nucleosides do not carry negative charges, mediating stronger binding of the monomer to the growing primer strand (Sigel and Griesser 2005). In today′s DNA/RNA/protein life forms the negative charges can be easily shielded by a protein′s positively charged residues, but in an RNA world such shielding would have been more difficult because RNA

does not contain positive charges at physiological pH. RNA polymerization in 3′-5′ direction with alternating RNA triphosphorylation and nucleoside addition would have avoided this problem.

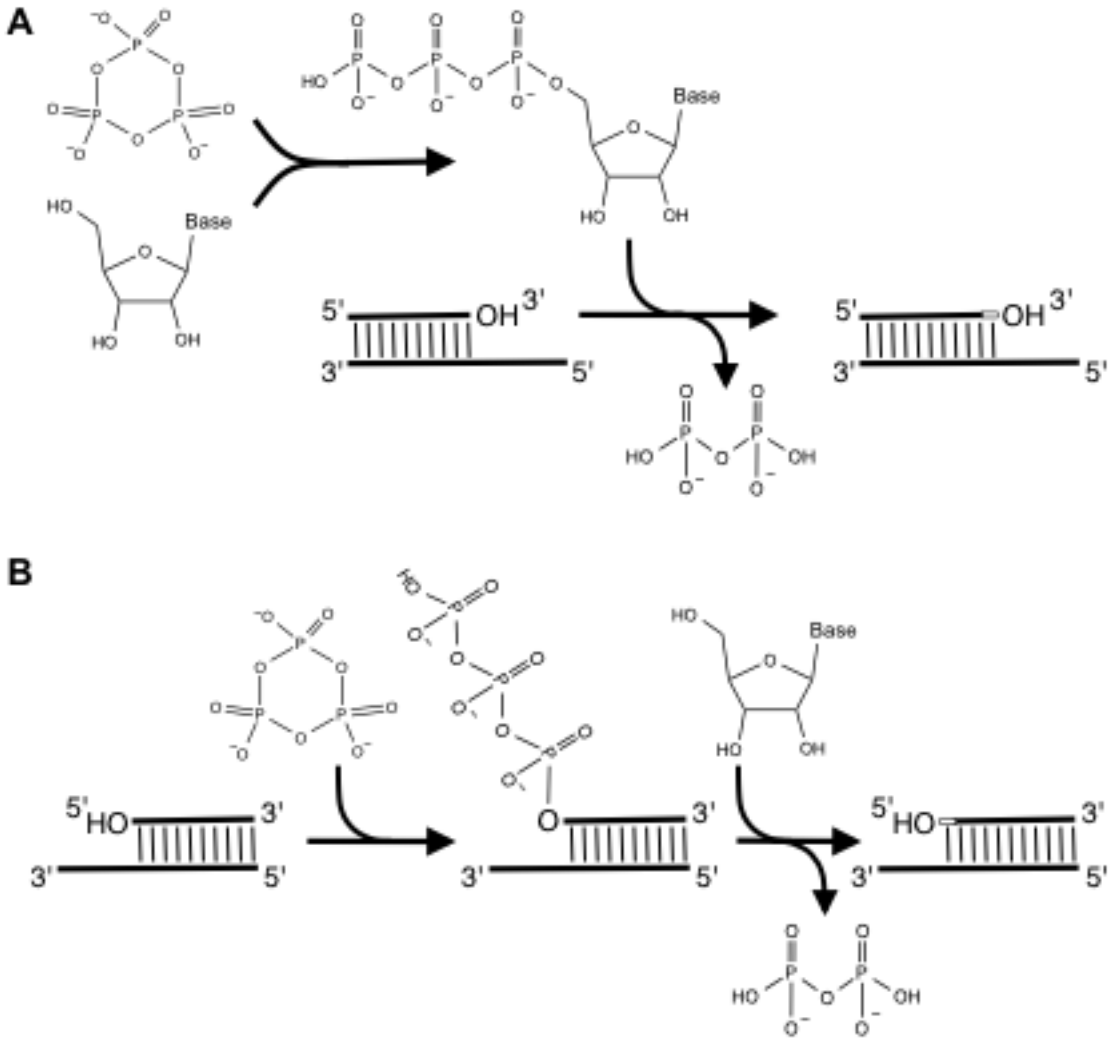

Figure 4.5. Two alternative routes to RNA polymerization in an RNA World.
A) RNA polymerization could have proceeded in a fashion analogous to modern DNA or RNA synthesis, where nucleoside triphosphates are first formed, then react with the primer 3′-hydroxyl groups to elongate the primers in 5′- to 3′- direction. **B)** Alternatively, RNA World organisms could have relied on RNA polymerization in the 3′- to 5′-direction: here the primer 5′-hydroxyl group is first triphosphorylated, then the activated primer reacts with a nucleoside 3′-hydroxyl group to extend the primer in 3′- to 5′-direction. Note that the monomer in (A) carries three negative charges whereas it is uncharged in (B), facilitating stronger binding of the nucleoside to the elongating primer 5′-terminus.

In summary, the existence of triphosphorylation ribozymes shows that ribozymes are able to utilize trimetaphosphate as an energy source, with the same chemistry that generates nucleoside triphosphate from nucleosides and trimetaphosphate. Future studies will show how these ribozymes can be integrated with ribozymes that generate RNA polymers.

4.2.3. RNA polymerase ribozymes

The central activity of an RNA World organism is RNA polymerization, to facilitate both self-replication and evolution. Self-replication has also been shown for ligase ribozymes (Hayden and Lehman 2006), (Lincoln and Joyce 2009), (Vaidya et al. 2012) but these systems are limited in their evolutionary potential because they assemble from large RNA fragments of defined sequence. In contrast, polymerase ribozymes generate RNA polymers from monomers such as nucleoside triphosphates, which allows for errors to occur at each position during replication. Therefore, polymerase ribozymes allow the evolutionary exploration of sequence space on the single-nucleotide level and have the potential to invent new activities. This open-ended evolutionary feature is crucial for early life forms to give them the potential to evolve into more complex life forms.

The most successful polymerase ribozyme to date was developed in three stages: First, a ligase ribozyme was developed by *in vitro* selection from a random sequence library containing ~10^{15} different sequences with 220 randomized nucleotides (Bartel and Szostak 1993b). This ribozyme, termed the ′Class I Ligase′ (Figure 4.6) catalyzes the nucleophilic attack of 3′-hydroxyl groups on RNA 5′-triphosphates, generating 3′-5′ phosphodiester bonds at a rate about 10^{7}-fold above that of the uncatalyzed reaction. Second, variants of this ligase ribozyme were designed to extend an RNA primer by six nucleotides, using nucleoside triphosphates (Ekland and Bartel 1996). Importantly, the fidelity of these nucleotide additions was 92%, on average. This is much higher than the fidelity estimated from the stability of Watson-Crick pairing (~40%) (Ekland and Bartel 1996), implying that the ribozyme recognizes to some extent the geometry of a Watson-Crick base pair between the template strand and the incoming nucleoside triphosphate at the catalytic site (Shechner et al. 2009). Third, an accessory domain was developed for the polymerase ribozyme by *in vitro* selection (Johnston et al. 2001). To do this, a 76-

nucleotide long randomized sequence was appended to the 3′-terminus of the ligase domain. After 18 rounds of in vitro selection this library gave rise to the R18 (round 18) polymerase ribozyme, which facilitates the templated primer extension of 14 nucleotides, with an average fidelity of 97%. The R18 ribozyme has been the starting point for reselections which have generated the closely related R18 family: notable members include the B6.61 (Zaher and Unrau 2007) (Figure 4.6) and tC19z RNA polymerase ribozymes (Wochner et al. 2011) (Figure 4.6).

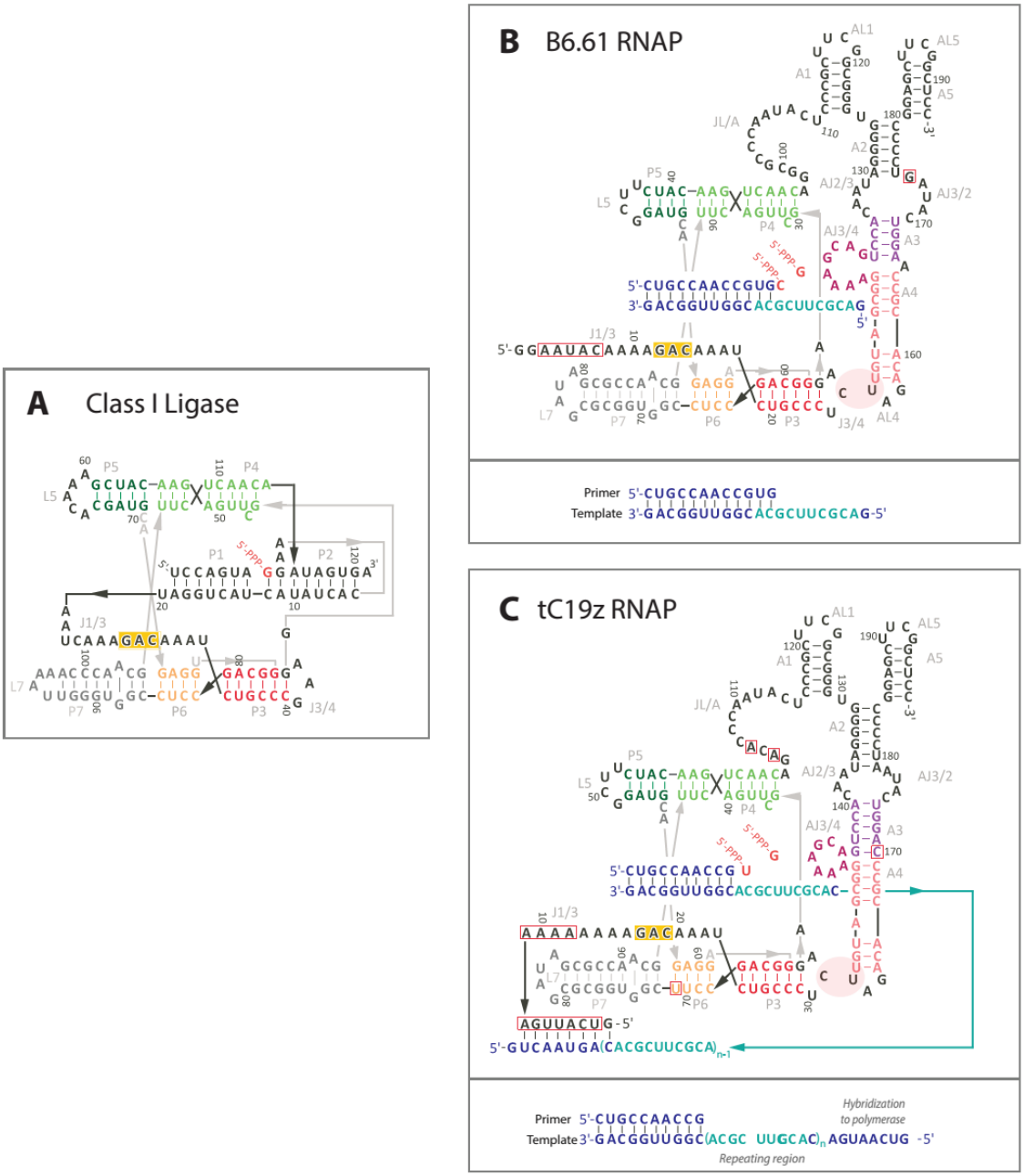

Figure 4.6. Secondary structure comparison of *in vitro* selected ligase and polymerase ribozymes.

A) The Class I ligase secondary structure. B) Proposed secondary structure of the B6.61 RNAP ribozyme in complex with its preferred primer-template (blue) and optimal template sequence shown in cyan. Mutations relative to R18 polymerase are boxed in red. C) Proposed secondary structure of the tC19z RNAP ribozyme in complex with a highly repetitive template sequence (cyan). The 5′ end of the ribozyme was engineered to hybridize to the 3′ of the template, promoting longer extension of the primer but also blocking the ribozyme from complete primer extension (Wochner et al. 2011). Mutations relative to R18 polymerase are boxed in red.

The structure of the R18 polymerase ribozyme was characterized by crystallographical and biochemical studies. Two crystal structures of the catalytic core show that the ligase domain of the polymerase ribozyme adopts a tripod-like structure, in which two of the three tripod legs are formed by stem-loops of the catalytic domain and the third leg by the primer/template duplex (Shechner et al. 2009). The legs of the tripod join at the catalytic center, where the base of the incoming nucleoside triphosphate is stacked onto the primer 3′-terminus and base paired to the first templating nucleotide (Figure 4.6). The triphosphate is positioned in a bent conformation at the catalytic site by coordination to several Mg^{2+} ions. The α-phosphate of this nucleoside triphosphate is positioned in a nearly ideal conformation for in-line nucleophilic attack by the primer 3′-hydroxyl group: The distance between the 3′-hydroxyl group and the phosphorus is 3 Å, and the angle between these groups and the oxygen in the leaving group is 176°, close to the ideal 180°. Deprotonation of the 3′-hydroxyl group appears to be mediated by an inner-sphere coordination of Mg^{2+}; the negative charge on the pyrophosphate leaving group is stabilized by hydrogen bonds to a ribozyme 2′-hydroxyl group and the exocyclic amino group of a cytosine (Shechner et al. 2009). The accessory domain of the polymerase ribozyme is draped on top of the tripod-like structure of the catalytic domain (Wang et al. 2011) (Figure 4.6). It contains a purine-rich 8-nucleotide bulge that appears to be involved in stabilizing incoming nucleoside triphosphates and is positioned by interactions between the J3/4 stem loop in the ligase core and the AL4 triloop found in the accessory domain. The primer/template duplex is bound by the ribozyme via three well-defined contacts to the 2′-hydroxyl groups of the primer and the template, two on the primer strand (position -2 and -3 relative to the 3′-terminus) and one on the template strand (position -3 relative to the primer 3′-terminus) (Müller and Bartel 2003). Additionally, several positions in the single-stranded portion of the primer/template (position +3, +4, and +5 relative to the primer 3′-terminus) appear to establish weaker, more flexible hydrogen bond(s). The 2′-hydroxyl group at the primer 3′-terminus was important for catalysis, probably by lowering the pK_A of the 3′-hydroxyl group (Izatt et al. 1966). Together, these results generate a structural picture of the polymerase ribozyme; however the precise three-dimensional structure of the accessory domain remains to be determined.

The efficiency of the polymerase ribozyme is limited by its weak affinity to its primer/template (PT) substrate, with an effective dissociation constant in the millimolar range (Lawrence and Bartel 2003). This causes a release of the primer/template duplex after most additions of a nucleotide to the primer, generating a highly distributive (as opposed to processive) polymerization mechanism. To date five different approaches have been attempted to overcome the limited efficiency of the R18 polymerase family. First, hydrophobic anchors were attached to the ribozyme and the primer/template such that both RNAs were co-localized on micelles (Müller and Bartel 2008). This approach led to a 3- to 20-fold increase of the product yield, depending on the template sequences. Second, incubation of the polymerization reaction below freezing temperature, in eutectic phase, allowed the polymerization of up to 118 nucleotides when done together with the two other approaches (see below): the use of a specific template sequence, and tethering the ribozyme′s favorite primer/template to the ribozyme by base pairing (Attwater et al. 2013). Eutectic phases benefit the polymerase ribozyme because the reactants are concentrated between the water crystals formed in the freezing process, and the life-time of the ribozyme is elongated at low temperature. Third, it was tested whether single arginine cofactors attached to the ribozyme could improve polymerization efficiency (Yao and Muller 2011). Although ten different positions on the polymerase ribozyme were tested none of them led to improved polymerization, probably due to the high Mg^{2+} ion concentrations required for polymerase ribozyme activity. Fourth, G/T-rich sequences appended to the ribozyme near the active site caused dramatic improvements of polymerization efficiency (Yao and Muller 2011). The templates used in this study were intentionally chosen to be different from the favorite sequence: Without the G/T-rich oligonucleotides the primer was extended only by 3-5 nucleotides; with G/T-rich sequences the polymerization extended to 7-14 nucleotides. The mechanism of these G/T-rich sequences appeared to be base pairing to the template strands with low sequence specificity because G and T can form base pairs to each two other bases. This principle could be exploited further by cofactors that bind RNA without sequence specificity. Fifth, tethering the PT to the ribozyme led to significant increases in product length. Tethering primer templates using a flexible PEG linker via the L5 or L7 loops or to the 5′ terminus of the polymerase resulted in a significant increase in polymerization efficiency, while tethering PT to the accessory domain at a number of sites resulted in only modest enhancements in polymerization (Wang et al. 2011). Only one specific 10-

nucleotide long template sequence is known to give efficient polymerization (Figure 4.6), repeats of which have been demonstrated to yield polymerization of 95 nucleotides (10 repeats, (Wochner et al. 2011)) and 206 nt (19 repeats (Attwater et al. 2013)) when repeats of this sequence where used and tethered to the 5′ terminus of the polymerase by direct hybridization (Figure 4.6). These reactions demonstrate that there are no steric factors preventing the polymerization of long RNA polymers, consistent with earlier findings with the R18 polymerase (Johnston et al. 2001). More importantly, this same tethered setup has yielded a functional RNA ′transcript′ of the hammerhead ribozyme as demonstrated by Holliger′s group (Wochner et al. 2011).

Despite the increases in processivity described above it may prove impossible to rid the relatives of the R18 ribozyme from its addiction to its favorite 10-nucleotide template sequence. A solution to this situation may be presented by polymerase ribozymes with different accessory domains, which were *in vitro* selected in similar fashion as the R18 ribozyme (Lawrence and Bartel 2005). These accessory domains are initial isolates from an *in vitro* selection, but they have not yet been optimized like the R18 ribozyme. Therefore, while the R18 ribozyme family may have reached its fitness peak in the latest evolution experiments, some of the different accessory domains may give rise to polymerase ribozymes that are efficient enough for self-replication. It is also possible that whole new approach may be needed to break free of the processivity barrier experienced thus far. A recent, elegant study demonstrates a solution to the problem of limited processivity: The cross-replication of two stereoisomeric ribozymes. Because the stereoisomers cannot form a stable Watson-Crick duplex, these ribozymes are able to replicate their respective stereoisomer so that two of these ribozymes can form a self-replicating system (Sczepanski and Joyce 2014). The current version of the system is promising, since it can use nucleoside triphosphates as substrates. At present it requires short oligomers in the self-replication process, but future versions of this approach may form a self-replicating system from a racemic mixture of nucleoside triphosphates.

4.3. Outlook: Outstanding obstacles to forging an RNA World

The ribozymes described in the three chapters above promote three types of reactions that might make it possible to generate an RNA World organism in the lab. What remains to be done to obtain such an organism? Three types of challenges remain: First, the three types of ribozymes are not yet efficient enough to sustain a self-replicating system of ribozymes, and second, their substrate specificities are not yet orchestrated to work with each other. For example, the ribozymes promoting nucleotide synthesis from PRPP and pyrimidine or purine nucleobases require tethering of the PRPP substrate to the ribozyme 3′-terminus (Unrau and Bartel 1998), (Lau et al. 2004). To work together with the existing triphosphorylation ribozymes (Moretti and Muller 2014) the nucleotide synthetase ribozymes would have to be modified (or new ribozymes selected) that react free PRPP with pyrimidine or purine nucleobases. Alternatively, it could be envisioned that the ribozyme-PRPP conjugates act as covalent intermediates that are later released as nucleosides (the 5′-phosphate of the newly created nucleotide could remain as 2′,3′-cyclic phosphate on the ribozyme 3′-terminus, perhaps allowing a new ribose to be conjugated). A similar problem is faced by the existing triphosphorylation ribozymes (Moretti and Muller 2014). These ribozymes triphosphorylated the 5′-terminal nucleoside of the ribozyme with trimetaphosphate but were unable to triphosphorylate free nucleosides. If ribozymes generating free nucleoside triphosphates could be generated they would work together with the existing polymerase ribozymes (Johnston et al. 2001), (Zaher and Unrau 2007), (Wochner et al. 2011). Alternatively, it would be sufficient to triphosphorylate the 5′-terminal nucleoside of RNAs if RNA polymerization could proceed in 3′-5′ direction (Figure 4.5) instead of the biological 5′-3′ direction (Fig. 5A). However, this would require that the triphosphorylation ribozymes act *in trans* on substrate RNAs, and that ligase ribozymes are identified that condensate the RNA 5′-triphosphate with incoming nucleosides. The polymerase ribozyme faces a different problem related to substrate binding: Only one specific template sequence is bound so well that concatemers of this sequence facilitate the polymerization of 206 nucleotides (Attwater et al. 2013). However, template sequences that are useful for the generation of ribozymes (and therefore self-replication) have so far led to the polymerization of only 24 nucleotides (Wochner et al. 2011). It may be possible to obtain polymerase ribozymes efficient enough for self-replication by one of

the approaches outlined in the last chapter: Different types of tethering between ribozyme and substrate (Müller and Bartel 2008), (Attwater et al. 2010), (Yao and Muller 2011), (Wang et al. 2011)), reactions under different physical conditions (Attwater et al. 2010), (Attwater et al. 2013), or the optimization of different accessory domains (Lawrence and Bartel 2005).

A third set of challenges to generating an RNA World organism in the lab stems from general chemical requirements for the components to work together (Szostak 2012). A potential stumbling block may be encountered when encapsulating ribozymes in a phospholipid membrane, since most catalytic RNAs require concentrations of Mg^{2+} that lead to the aggregation of lipid vesicles. Luckily, the coordination of Mg^{2+} with citrate was recently shown to protect lipid vesicles while allowing non-enzymatic RNA polymerization to occur (Adamala and Szostak 2013). A similar mechanism may be possible to allow ribozyme-catalyzed RNA polymerization in lipid vesicles (Müller and Tor 2014). Another issue is the relatively low fidelity of template-dependent RNA polymerization, currently in the range of 97% , which may not be sufficient for the stable propagation of genetic information. However, variants of the polymerase ribozyme can have improved fidelity (Zaher and Unrau 2007), and the effect of polymerization stalling after mismatches may be sufficient to reach the necessary fidelity (Rajamani et al. 2010). Perhaps of greatest significance is the ′strand displacement problem′: Namely, all current polymerase ribozymes generate an RNA double strand. Long RNAs are extremely thermostable, and even after heat denaturation the strands are likely to re-form. How can the individual strands be kept separate so that they can fold into functional ribozymes? While some ideas have been discussed on these topics (Cheng and Unrau 2010) no solution yet exists to these difficult problems. These questions show that a series of discoveries have to be made before it will be possible to generate an RNA World organism in the lab.

If we are successful in generating an RNA world organism, what will we learn about the origin of life? How could we use such an ′artificial′ organism to find constraints on the origin of life? First, it would show us that, indeed, ribozymes are able to generate self-replicating and evolving systems, and thereby support the RNA World hypothesis. Second, RNA World organisms in the lab would make it possible to study them in ways that are currently impossible: The evolution of these simple organisms could be studied

by sequencing their genome every few generations and following their mutations. Because these systems would probably consist of less than a dozen catalytic RNAs it would be possible to analyze and understand all mutual molecular interactions - and thereby understand, for the first time, a life-like system on the molecular level. These organisms would also evolve to become more efficient and would show us more efficient ways to construct RNA world organisms, and perhaps simpler and more likely ways how they could have originated.

While the origin of life may have coincided with the origin of the RNA World it should not be forgotten that understanding the origin of life requires understanding the prebiotic chemistry that made the RNA World possible. For example, the prebiotic synthesis of ribose, nucleobases, and nucleotides is debated with very different models (Powner et al. 2009), (Chen et al. 2013) and it may be a long time until it is possible to generate the necessary compounds for an RNA world in a prebiotically plausible scenario. Some important questions still lack satisfactory answers, such as: how the necessary regiospecificity and stereospecificity could have been achieved, and how the first RNA polymers could have been made in the absence of ribozymes (Szostak 2012). While the answers may lie many years in the future, the quest to understand how life originated will remain one of the most exciting activities of mankind.

References

Acevedo OL, Orgel LE (1987) Non-enzymatic transcription of an oligodeoxynucleotide 14 residues long. J Mol Biol 197:187–193.

Adamala K, Szostak JW (2013) Nonenzymatic template-directed RNA synthesis inside model protocells. Science 342:1098–100. doi: 342/6162/1098 [pii] 10.1126/science.1241888

Anet FA (2004) The place of metabolism in the origin of life. Curr Opin Chem Biol 8:654–9.

Attwater J, Wochner A, Holliger P (2013) In-ice evolution of RNA polymerase ribozyme activity. Nat Chem 5:1011–8. doi: nchem.1781 [pii] 10.1038/nchem.1781

Attwater J, Wochner A, Pinheiro VB, et al (2010) Ice as a protocellular medium for RNA replication. Nat Commun 1:76. doi: 10.1038/ncomms1076

Bahulayan D, Das SK, Iqbal J (2003) Montmorillonite K10 clay: an efficient catalyst for the one-pot stereoselective synthesis of beta-acetamido ketones. J Org Chem 68:5735–5738. doi: 10.1021/jo020734p

Banin A, Lawless JG, Mazzurco J, et al (1985) pH profile of the adsorption of nucleotides onto montmorillonite. Orig Life Evol Biosphere 15:89–101.

Bartel DP, Szostak JW (1993) Isolation of new ribozymes from a large pool of random sequences. Science 261:1411–8.

Bartel DP, Unrau PJ (1999) Constructing an RNA world. Trends Cell Biol 9:M9–M13.

Batey RT, Gilbert SD, Montange RK (2004) Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. Nature 432:411–415. doi: 10.1038/nature03037

Bigi F, Chesini L, Maggi R, Sartori G (1999) Montmorillonite KSF as an Inorganic, Water Stable, and Reusable Catalyst for the Knoevenagel Synthesis of Coumarin-3-carboxylic Acids. J Org Chem 64:1033–1035.

Bowler FR, Chan CKW, Duffy CD, et al (2013) Prebiotically plausible oligoribonucleotide ligation facilitated by chemoselective acetylation. Nat Chem 5:383–389. doi: 10.1038/nchem.1626

Cate JH, Gooding AR, Podell E, et al (1996) RNA tertiary structure mediation by adenosine platforms. Science 273:1696–9.

Cech TR (2000) The Ribosome Is a Ribozyme. Science 289:878 –879. doi: 10.1126/science.289.5481.878

Cernak P, Sen D (2013) A thiamin-utilizing ribozyme decarboxylates a pyruvate-like substrate. Nat Chem 5:971–977. doi: 10.1038/nchem.1777

Chapple KE, Bartel DP, Unrau PJ (2003) Combinatorial minimization and secondary structure determination of a nucleotide synthase ribozyme. RNA 9:1208–1220.

Cheng LKL, Unrau PJ (2010) Closing the circle: replicating RNA with RNA. Cold Spring Harb Perspect Biol 2:a002204. doi: 10.1101/cshperspect.a002204

Chen IA, Roberts RW, Szostak JW (2004) The emergence of competition between model protocells. Science 305:1474–6.

Chen MC, Cafferty BJ, Mamajanov I, et al (2013) Spontaneous Prebiotic Formation of a beta-Ribofuranoside That Self-Assembles with a Complementary Heterocycle. J Am Chem Soc. doi: 10.1021/ja410124v

Chen X, Li N, Ellington AD (2007) Ribozyme Catalysis of Metabolism in the RNA World. Chem Biodivers 4:633–655. doi: 10.1002/cbdv.200790055

Cockell CS (1998) Ultraviolet radiation, evolution and the π-electron system. Biol J Linn Soc 63:449–457. doi: 10.1111/j.1095-8312.1998.tb01528.x

Cooper J, Miller R, Patterson J (1986). A Trip Through Time: Principles of Historical Geology. Columbus: Merril Publishing Company.

Craig SP, Eakin AE (2000) Purine Phosphoribosyltransferases. J Biol Chem 275:20231–20234. doi: 10.1074/jbc.R000002200

Crick FH (1958) On protein synthesis. In: Symposia of the Society for Experimental Biology. p 138

Ekland EH, Bartel DP (1995) The secondary structure and sequence optimization of an RNA ligase ribozyme. Nucleic Acids Res 23:3231–8.

Ekland EH, Bartel DP (1996) RNA-catalysed RNA polymerization using nucleoside triphosphates. Nature 382:373–376. doi: 10.1038/382373a0

Eliasson R, Reichard P (1978) Primase initiates Okazaki pieces during polyoma DNA synthesis. Nature 272:184–185.

Ellington AD, Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. Nature 346:818–22.

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

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

Ferris JP (2002) Montmorillonite catalysis of 30-50 mer oligonucleotides: laboratory demonstration of potential steps in the origin of the RNA world. Orig Life Evol Biosphere 32:311–32.

Ferris JP, Ertem G (1992) Oligomerization of ribonucleotides on montmorillonite: reaction of the 5'-phosphorimidazolide of adenosine. Science 257:1387–9.

Ferris JP, Ertem G (1993) Montmorillonite catalysis of RNA oligomer formation in aqueous solution. A model for the prebiotic formation of RNA. J Am Chem Soc 115:12270–5.

Fuller WD, Sanchez RA, Orgel LE (1972) Studies in prebiotic synthesis. VI. Synthesis of purine nucleosides. J Mol Biol 67:25–33.

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

Guerrier-Takada C, Gardiner K, Marsh T, et al (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell 35:849–857. doi: 10.1016/0092-8674(83)90117-4

Guo F, Gooding AR, Cech TR (2004) Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site. Mol Cell 16:351–62.

Hayden EJ, Lehman N (2006) Self-assembly of a group I intron from inactive oligonucleotide fragments. Chem Biol 13:909–18. doi: S1074-5521(06)00224-9 [pii] 10.1016/j.chembiol.2006.06.014

Huang W, Ferris JP (2006) One-step, regioselective synthesis of up to 50-mers of RNA oligomers by montmorillonite catalysis. J Am Chem Soc 128:8914–9. doi: 10.1021/ja061782k

Igloi GL (1988) Interaction of tRNAs and of phosphorothioate-substituted nucleic acids with an organomercurial. Probing the chemical environment of thiolated residues by affinity electrophoresis. Biochemistry (Mosc) 27:3842–9.

Inoue T, Orgel LE (1982) Oligomerization of (guanosine 5'-phosphor)-2-methylimidazolide on poly(C). An RNA polymerase model. J Mol Biol 162:201–17. doi: 0022-2836(82)90169-3 [pii]

Izatt RM, Rytting JH, Hansen LD, Christensen JJ (1966) Thermodynamics of proton dissociation in dilute aqueous solution. V. An entropy titration study of adenosine, pentoses, hexoses, and related compounds. J Am Chem Soc 88:2641–5.

Johnston WK, Unrau PJ, Lawrence MS, et al (2001) RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension. Science 292:1319–1325. doi: 10.1126/science.1060786

Joshi PC, Aldersley MF, Delano JW, Ferris JP (2009) Mechanism of Montmorillonite Catalysis in the Formation of RNA Oligomers. J Am Chem Soc 131:13369–13374. doi: 10.1021/ja9036516

Joyce GF (2002) The antiquity of RNA-based evolution. Nature 418:214–21.

Kanehisa M, Goto S, Sato Y, et al (2014) Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res 42:D199–205. doi: 10.1093/nar/gkt1076

Kasting JF, Ono S (2006) Palaeoclimates: the first two billion years. Philos Trans R Soc Lond B Biol Sci 361:917–929. doi: 10.1098/rstb.2006.1839

Klein DJ, Ferré-D'Amaré AR (2006) Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. Science 313:1752–1756. doi: 10.1126/science.1129666

Kornberg A, Rao NN, Ault-Riche D (1999) Inorganic polyphosphate: a molecule of many functions. Annu Rev Biochem 68:89–125.

Krasilnikov AS, Yang X, Pan T, Mondragón A (2003) Crystal structure of the specificity domain of ribonuclease P. Nature 421:760–764. doi: 10.1038/nature01386

Krishnamurthy R (2012) Role of pK(a) of nucleobases in the origins of chemical evolution. Acc Chem Res 45:2035–44. doi: 10.1021/ar200262x

Kura G (1987) Alkaline Hydrolysis of Inorganic cyclo-Polyphosphates. Bull Chem Soc Jpn 60:2857–2860.

Lathe R (2005) Tidal chain reaction and the origin of replicating biopolymers. Int J Astrobiol 4:19–31. doi: 10.1017/S1473550405002314

Lau MW, Cadieux KE, Unrau PJ (2004) Isolation of fast purine nucleotide synthase ribozymes. J Am Chem Soc 126:15686–15693.

Lau MW, Unrau PJ (2009) A promiscuous ribozyme promotes nucleotide synthesis in addition to ribose chemistry. Chem Biol 16:815–825.

Lawrence MS, Bartel DP (2005) New ligase-derived RNA polymerase ribozymes. RNA 11:1173–80.

Lawrence MS, Bartel DP (2003) Processivity of ribozyme-catalyzed RNA polymerization. Biochemistry (Mosc) 42:8748–55.

Lehman N, Unrau PJ (2005) Recombination during in vitro evolution. J Mol Evol 61:245–252. doi: 10.1007/s00239-004-0373-4

Leontis NB, Stombaugh J, Westhof E (2002) The non-Watson–Crick base pairs and their associated isostericity matrices. Nucleic Acids Res 30:3497–3531. doi: 10.1093/nar/gkf481

Levy M, Griswold KE, Ellington AD (2005) Direct selection of trans-acting ligase ribozymes by in vitro compartmentalization. RNA 11:1555–62.

Lincoln TA, Joyce GF (2009) Self-sustained replication of an RNA enzyme. Science 323:1229–32.

Lohrmann R, Orgel LE (1968) Prebiotic synthesis: phosphorylation in aqueous solution. Science 161:64–6.

Lohse PA, Szostak JW (1996) Ribozyme-catalysed amino-acid transfer reactions. Nature 381:442–444.

Lutay AV, Zenkova MA, Vlassov VV (2007) Nonenzymatic recombination of RNA: possible mechanism for the formation of novel sequences. Chem Biodivers 4:762–767. doi: 10.1002/cbdv.200790062

MacKenzie AP (1977) Non-equilibrium freezing behaviour of aqueous systems. Philos Trans R Soc Lond B Biol Sci 278:167–189.

Mandal M, Boese B, Barrick JE, et al (2003) Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. Cell 113:577–86.

Martin LL, Unrau PJ, Müller UF (2015) RNA Synthesis by in Vitro Selected Ribozymes for Recreating an RNA World. Life 5:247–268. doi: 10.3390/life5010247

Martin W, Russell MJ (2007) On the origin of biochemistry at an alkaline hydrothermal vent. Philos Trans R Soc B Biol Sci 362:1887–1926. doi: 10.1098/rstb.2006.1881

McCown PJ, Liang JJ, Weinberg Z, Breaker RR (2014) Structural, functional, and taxonomic diversity of three preQ1 riboswitch classes. Chem Biol 21:880–889. doi: 10.1016/j.chembiol.2014.05.015

McGinness KE, Wright MC, Joyce GF (2002) Continuous in vitro evolution of a ribozyme that catalyzes three successive nucleotidyl addition reactions. Chem Biol 9:585–96. doi: S1074552102001369 [pii]

Meunier A, Petit S, Cockell CS, et al (2010) The Fe-rich clay microsystems in basalt-komatiite lavas: importance of Fe-smectites for pre-biotic molecule catalysis during the Hadean eon. Orig Life Evol Biosphere 40:253–272. doi: 10.1007/s11084-010-9205-2

Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. Nucleic Acids Res 15:8783–98.

Miyakawa S, Joshi PC, Gaffey MJ, et al (2006) Studies in the Mineral and Salt-Catalyzed Formation of RNA Oligomers. Orig Life Evol Biospheres 36:343–361. doi: 10.1007/s11084-006-9009-6

Monnard PA, Deamer DW (2003) Preparation of vesicles from nonphospholipid amphiphiles. Methods Enzym 372:133–51. doi: 10.1016/S0076-6879(03)72008-4 S0076687903720084 [pii]

Moretti JE, Muller UF (2014) A ribozyme that triphosphorylates RNA 5'-hydroxyl groups. Nucleic Acids Res 42:4767–78. doi: gkt1405 [pii] 10.1093/nar/gkt1405

Mortimer SA, Weeks KM (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. J Am Chem Soc 129:4144–5. doi: 10.1021/ja0704028

Müller UF, Bartel DP (2003) Substrate 2'-hydroxyl groups required for ribozyme-catalyzed polymerization. Chem Biol 10:799–806.

Müller UF, Bartel DP (2008) Improved polymerase ribozyme efficiency on hydrophobic assemblies. RNA 14:552–562. doi: 10.1261/rna.494508

Müller UF, Tor Y (2014) Citric acid and the RNA world. Angew Chem Int Ed Engl 53:5245–5247. doi: 10.1002/anie.201400847

Mutschler H, Wochner A, Holliger P (2015) Freeze–thaw cycles as drivers of complex ribozyme assembly. Nat Chem 7:502–508. doi: 10.1038/nchem.2251

Nielsen PE, Egholm M, Berg RH, Buchardt O (1991) Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. Science 254:1497–1500.

Oberbeck VR, Fogleman G (1989) Estimates of the maximum time required to originate life. Orig Life Evol Biosphere 19:549–560.

Orgel LE (2004) Prebiotic chemistry and the origin of the RNA world. Crit Rev Biochem Mol Biol 39:99–123.

Osterberg R, Orgel LE (1972) Polyphosphate and trimetaphosphate formation under potentially prebiotic conditions. J Mol Evol 1:241–8.

Paige RA (1970) Stalactite Growth beneath Sea Ice. Science 167:171–172. doi: 10.1126/science.167.3915.171-a

Pasek MA, Harnmeijer JP, Buick R, et al (2013) Evidence for reactive reduced phosphorus species in the early Archean ocean. Proc Natl Acad Sci USA 110:10089–94. doi: 1303904110 [pii] 10.1073/pnas.1303904110

Pasek MA, Kee TP, Bryant DE, et al (2008) Production of potentially prebiotic condensed phosphates by phosphorus redox chemistry. Angew Chem Int Ed Engl 47:7918–20. doi: 10.1002/anie.200802145

Patel BH, Percivalle C, Ritson DJ, et al (2015) Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. Nat Chem 7:301–307. doi: 10.1038/nchem.2202

Patel RM, Hurwitz A (1972) Eutectic temperature determination of preformulation systems and evaluation by controlled freeze drying. J Pharm Sci 61:1806–1810.

Paul N, Joyce GF (2002) A self-replicating ligase ribozyme. Proc Natl Acad Sci USA 99:12733–12740.

Powner MW, Gerland B, Sutherland JD (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. Nature 459:239–242. doi: 10.1038/nature08013

Powner MW, Sutherland JD, Szostak JW (2010) Chemoselective Multicomponent One-Pot Assembly of Purine Precursors in Water. J Am Chem Soc 132:16677–16688. doi: 10.1021/ja108197s

Prabahar KJ, Ferris JP (1997) Adenine derivatives as phosphate-activating groups for the regioselective formation of 3’,5’-linked oligoadenylates on montmorillonite: possible phosphate-activating groups for the prebiotic synthesis of RNA. J Am Chem Soc 119:4330–7.

Rajamani S, Ichida JK, Antal T, et al (2010) Effect of stalling after mismatches on the error catastrophe in nonenzymatic nucleic acid replication. J Am Chem Soc 132:5880–5. doi: 10.1021/ja100780p

Rajamani S, Vlassov A, Benner S, et al (2008) Lipid-assisted synthesis of RNA-like polymers from mononucleotides. Orig Life Evol Biosphere 38:57–74. doi: 10.1007/s11084-007-9113-2

Reichard P (1955) Utilization of deoxyuridine and 5-methyluridine for the biosynthesis of thymine by the rat.

Renz M, Lohrmann R, Orgel LE (1971) Catalysts for the polymerization of adenosine cyclic 2', 3'-phosphate on a poly(U) template. Biochim Biophys Acta 240:463–471.

Robertson MP, Joyce GF (2012) The origins of the RNA world. Cold Spring Harb Perspect Biol. doi: cshperspect.a003608 [pii] 10.1101/cshperspect.a003608

Rogers J, Joyce GF (2001) The effect of cytidine on the structure and function of an RNA ligase ribozyme. RNA 7:395–404.

Saffhill R (1970) Selective phosphorylation of the cis-2', 3'-diol of unprotected ribonucleosides with trimetaphosphate in aqueous solution. J Org Chem 35:2881–2883. doi: 10.1021/jo00834a004

Sanchez R, Ferris J, Orgel LE (1966) Conditions for purine synthesis: did prebiotic synthesis occur at low temperatures? Science 153:72–3.

Sawai H, Orgel LE (1975) Letter: Oligonucleotide synthesis catalyzed by the Zn-2+ ion. J Am Chem Soc 97:3532–3.

Schultes EA, Bartel DP (2000) One sequence, two ribozymes: implications for the emergence of new ribozyme folds. Science 289:448–52.

Schwartz AW (2006) Phosphorus in prebiotic chemistry. Philos Trans R Soc B Biol Sci 361:1743–1749. doi: 10.1098/rstb.2006.1901

Sczepanski JT, Joyce GF (2014) A cross-chiral RNA polymerase ribozyme. Nature 515:440–442. doi: 10.1038/nature13900

Seelig B, Jäschke A (1999) A small catalytic RNA motif with Diels-Alderase activity. Chem Biol 6:167–176. doi: 10.1016/S1074-5521(99)89008-5

Seelig B, Szostak JW (2007) Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. Nature 448:828–31.

Serganov A, Yuan Y-R, Pikovskaya O, et al (2004) Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. Chem Biol 11:1729–1741. doi: 10.1016/j.chembiol.2004.11.018

Shechner DM, Grant RA, Bagby SC, et al (2009) Crystal structure of the catalytic core of an RNA-polymerase ribozyme. Science 326:1271–1275. doi: 10.1126/science.1174676

Sigel H, Griesser R (2005) Nucleoside 5’-triphosphates: self-association, acid-base, and metal ion-binding properties in solution. Chem Soc Rev 34:875–900. doi: 10.1039/b505986k

Sprengel G, Follmann H (1981) Evidence for the reductive pathway of deoxyribonucleotide synthesis in an archaebacterium. FEBS Lett 132:207–209.

Srinivas KVNS, Das B (2003) A highly convenient, efficient, and selective process for preparation of esters and amides from carboxylic acids using Fe(3+)-K-10 montmorillonite clay. J Org Chem 68:1165–1167. doi: 10.1021/jo0204202

Stewart R (1985) The proton: applications to organic chemistry. In: Organic Chemistry. Academic Press, pp 42–44

Szostak JW (2012) The eightfold path to non-enzymatic RNA replication. J Syst Chem 3:2.

Tsukiji S, Pattnaik SB, Suga H (2003) An alcohol dehydrogenase ribozyme. Nat Struct Biol 10:713–7.

Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249:505–10.

Unrau PJ, Bartel DP (1998) RNA-catalysed nucleotide synthesis. Nature 395:260–263.

Unrau PJ, Bartel DP (2003) An oxocarbenium-ion intermediate of a ribozyme reaction indicated by kinetic isotope effects. Proc Natl Acad Sci USA 100:15393–7.

Vaidya N, Manapat ML, Chen IA, et al (2012) Spontaneous network formation among cooperative RNA replicators. Nature 491:72–77. doi: 10.1038/nature11549

Voytek SB, Joyce GF (2007) Emergence of a fast-reacting ribozyme that is capable of undergoing continuous evolution. Proc Natl Acad Sci USA 104:15288–93. doi: 0707490104 [pii] 10.1073/pnas.0707490104

Wang QS, Cheng LKL, Unrau PJ (2011) Characterization of the B6.61 polymerase ribozyme accessory domain. RNA 17:469–477. doi: 10.1261/rna.2495011

Wang Q, Unrau P (2005) Ribozyme motif structure mapped using random recombination and selection. RNA 11:404–411. doi: 10.1261/rna.7238705

White HB (1976) Coenzymes as fossils of an earlier metabolic state. J Mol Evol 7:101–4.

Wickramasinghe C (2004) The universe: a cryogenic habitat for microbial life. Cryobiology 48:113–125. doi: 10.1016/j.cryobiol.2004.03.001

Wochner A, Attwater J, Coulson A, Holliger P (2011) Ribozyme-catalyzed transcription of an active ribozyme. Science 332:209–212. doi: 10.1126/science.1200752

Wolf ET, Toon OB (2013) Hospitable archean climates simulated by a general circulation model. Astrobiology 13:656–673. doi: 10.1089/ast.2012.0936

Yamagata Y, Watanabe H, Saitoh M, Namba T (1991) Volcanic production of polyphosphates and its relevance to prebiotic evolution. Nature 352:516–9.

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

Zaher HS, Unrau PJ (2007) Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. RNA 13:1017–1026. doi: 10.1261/rna.548807

Zhang B, Cech TR (1997) Peptide bond formation by in vitro selected ribozymes. Nature 390:96–100. doi: 10.1038/36375

(2011) Deathly ice finger caught on film. BBC