Mechanistic studies of cation catalysis in nonenzymatic RNA primer extension

Citation

Pazienza, Lydia Trinidad. 2022. Mechanistic studies of cation catalysis in nonenzymatic RNA primer extension. Doctoral dissertation, Harvard University Graduate School of Arts and Sciences.

Permanent link

https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37372002

Terms of Use

Share Your Story

HARVARD UNIVERSITY
Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Department of Chemistry & Chemical Biology

have examined a dissertation entitled:

Mechanistic studies of cation catalysis in nonenzymatic RNA primer extension

presented by: Lydia Trinidad Pazienza

candidate for the degree of Doctor of Philosophy and hereby
certify that it is worthy of acceptance.

Signature _____
Typed name: Professor Jack Szostak

Signature _____
Typed name: Professor Brian Liau

Signature _____
Typed name: Professor David Bartel

Date: 21 April 2022

Mechanistic studies of cation catalysis in nonenzymatic RNA primer extension

A dissertation presented by

Lydia Trinidad Pazienza

to

The Department of Chemistry and Chemical Biology

In partial fulfillment of the

requirements for

the degree of

Doctor of Philosophy

in the subject of

Chemistry

Harvard University

Cambridge, Massachusetts

April 2022

Mechanistic studies of cation catalysis in nonenzymatic RNA primer extension

Abstract

The origin of life is believed to have centered around RNA acting as both genetic information and catalyst. Several decades into research around the RNA world hypothesis, however, life has yet to be created synthetically. This is in part due to the difficulty of transitioning from the prebiotic synthesis of RNA monomers to RNA oligonucleotides, and the subsequent replication of RNA. Only once RNA can undergo replication nonenzymatically and react to the lengths of ribozymes can the full potential of this theory be evaluated. While much has been learned mechanistically about nonenzymatic RNA copying in the last several years, a particular gap in knowledge surrounds the role of metal ion catalysis in nonenzymatic RNA primer extension, due to the weak and transient nature of this catalytic interaction.

This dissertation aims to characterize the nature of the interactions of cations in primer extension. In Chapter 1, the history of prebiotic chemistry is reviewed, as is what is known about RNA-cation interactions. In Chapter 2, I describe the metal ion, kinetics, and binding studies that support the model of 3′-OH nucleophilic activation from an inner-sphere contact with a catalytic cation. I additionally show the sequence, structural, and cation dependent effects of metal ion interactions in primer extension. I further support the previously published evidence shown indicating the 3′-OH is deprotonated before the rate-limiting step through kinetic isotope effects, and exclude the role of a metal-bound hydroxide species. Lastly, I measure the relative binding of metal ions as affected by the distance of the reacting bridged dinucleotide, suggesting an interaction with an oxygen of the bridged dinucleotide and the catalytic metal ion. In Chapter 3, I further validate this metal-bridged dinucleotide interaction as being relevant for primer extension through thiophosphoroimidazolide studies and metal rescue. We show that only one product is

observed in the presence of Mg^{2+} *in crystallo*, regardless of starting material identity. Due to racemization of the starting material over the time scale of crystallography, this observation supports that only one diastereomer is reactive for primer extension, supporting a metal ion interaction with one specific prochiral oxygen on phosphorous, and an S_N2-like mechanism for primer extension. Through this mechanistic understanding, the stereochemistry of the diastereomers can be assigned, and the coordination geometry of the catalytic metal ion is proposed, allowing chelation development to be performed intelligently. In Chapter 4, we address the possibility of Mn^{2+} as a prebiotically plausible catalyst in the RNA world, and find it improves ligation, primer extension rates (homo-polymeric and mixed template), and reaction yields, despite an increase in hydrolysis of reacting bridged dinucleotide. However, the increase in reactivity in the presence of Mn^{2+} results in a loss in fidelity at comparable concentrations of metal ion used for Mg^{2+}, likely due to increased competition of activated monomers (which react with lower fidelity), supporting pre-existing hypotheses on fidelity for nonenzymatic primer extension.

In sum, the results in this dissertation show the importance of metal ions in nonenzymatic primer extension, provide experimental results to support a model of cation catalysis, and provide a scaffold for designing co-catalysts to better enable metal ion-RNA interactions to improve nonenzymatic primer extension.

Table of Contents

Acknowledgements

During one of the most difficult times of my life, I sent a cold email to Jack asking if he would be interested in helping me write an NSF GRFP application and expressed interest in working in his lab during my graduate studies. His prompt response and encouragement of me during this time gave me purpose, and changed the course of my life for the last 6 years. His support in letting me explore a future as an educator is something I am immensely grateful for, and the fact that I got to teach for one course every year during my PhD kept me sane and ignited a passion in me which I hope to continue in my future career. I find it incredibly funny that I started my interactions with Jack during my time at UChicago, and I'm ending my PhD with him moving to UChicago- the world works in mysterious ways.

 I'm also incredibly grateful to Fanny, for her friendship and guidance on day-to-day things in the lab, in swapping baked goods, and for her support when I was lost- you are the heart of the Szostak lab. I am thankful to all the Szostak lab members for their help on methodology and instrumentation, and for their creation of a friendly space to work. I am so grateful to all the collaborators I've had across the country and world, whom without I would not have achieved a full picture of my research.

These acknowledgements wouldn't be complete without thanking my vast support network at Harvard and MIT who I could commiserate with and talk science to, in addition to having fun:

To Connor and Jo - even though you're inorganic chemists, you're somehow my best friends. I am so thankful for your support through all the ups and downs during our PhDs, and all the silly things we have done together. I'll always remember our times chairing, watching bad movies, and meat cupcakes. I'm so proud of where you two are headed, and cherish you greatly.

Thanks for helping enable me to attend (nearly) every year of the MIT Chemistry Christmas Party.

To my Pandemic crew- Bryan, Will, Sophia, (and also Jo)- we have finally curtailed the pandemic, Jumanji-style. Granted, that also means we *started* the pandemic…

To Hope- As my main contact back into Harvard Chemistry, I've been so thankful for your friendship, your desire to explore Boston with me, and our shared love of baking and fiber arts.

To Katherine and Jessica- Our weekly hangouts "watching" the Bachelor, or essentially just talking about our lives, hopes and dreams, have meant so much to me over the years. You both are such great scientists and I'm so proud of your contributions to your community and for your candor in your research difficulties and successes.

To Michelle- I've always felt grateful for our friendship, and our honest ability to share our struggles since our time together at UChicago.

To Zoe- I am so grateful for your friendship, guidance, and support during lab. Also, our shared love of (redacted). I admire you and look up to you so much. Your generosity led me to having the great experiences I had with the Harvard Pre College and Summer School, and this truly changed my career.

To the Origins Consortium grad students- Our bonding through traveling the world to try and understand the past is something I will never forget. Whether smashing rocks in Norway, treading off the beaten path to Octopus Springs in Yellowstone, or climbing cliffs in Iceland (and broadly trying not to die), we were able to not only get to learn more about each other's fields, but about each other's lives.

Lastly, I'd like to thank my support network outside of Chemistry and Harvard/MIT:

To Megan- you're my all-around best friend, and I am so grateful to have met you in high school. You've probably seen me undergo the most changes, and your friendship and support through the roller coaster that has been the last decade means the world to me.

To Ryan- Your support, empathy, and thoughtfulness the last year and a half have been critical to me getting through the pandemic and finishing this PhD.

To my therapist- when you spend an hour with someone every week for 6 years, it's hard to not have a substantial relationship. Going from where I was during my first year of graduate school with my daily panic attacks, to where I am now could not have been achieved without you.

To my sister, Grace- I am so glad we get to become doctors at the same time- albeit different kinds. You've been a great friend to me the last few years, and I'm so proud of us and our achievements.

To my parents- I could not have gotten to where I am today without your support and guidance. Everything I do is to make you proud.

This thesis is dedicated to survivors.

List of Figures and Tables

Chapter 1

From the early Earth to the RNA world: a tale of cations and nucleic acids in the origin of life

Past and current approaches to studying the origin of life

Throughout human history, society, culture, and religion have attempted to answer the question of how life began. Only since the Scientific Revolution has the scientific community begun to address how life may have begun with rigor. The discovery that compounds made by life, such as urea, can be made by abiotic processes arguably resulted in the downfall of vitalism and spurred the beginning of the search to create life *de novo*.[1-3] It took another century before Miller and Urey discovered that amino acids could be made from simple compounds and electrical discharge, and although the atmospheric composition they assumed for the early Earth was likely incorrect, it inspired substantial subsequent research that brings us to modern prebiotic chemistry.[4,5] Additionally, the discovery of prebiotically relevant compounds in space, on meteorites, and formed through relatively simple chemical feedstocks supports that the formation of some of the necessary compounds for life is possible in a number of different conditions.[6] While much is still unknown about the early Earth and what conditions definitively led to life, constraints exist to favor certain hypotheses over others. This dissertation will address chemical reactions of the RNA world and the benefits of cationic catalysts, which can be reconciled with the likely geochemistry of the early Earth. The compatibility of past chemical environments and relevant chemical reactions for life is one means to narrow down the many theories of how life was created and to approach creating life in the laboratory.

It is often debated if defining life is necessary in order to create or identify life elsewhere.[7-9] The full analysis of the definitions of life is outside the scope of this thesis, but considering life as having a series of specific characteristics is helpful for understanding the purpose of studying nonenzymatic RNA primer extension, the focus of this thesis. It is also important to consider that definitions of life tend to be biased towards the fields of those who

write the definitions, whether focusing on the physical, evolutionary, or chemical properties we have observed in Earth life. In terms of viewing life as a transition between chemistry and biology,[7] NASA, through the guidance of Gerald Joyce, had life defined as a self-sustaining chemical system capable of Darwinian evolution,[10] but when this was achieved with an ribozyme catalyzing its own replication, it did not satisfy the notion for life for the scientific community.[11,12] Chemical systems can evolve and not be much like known life— there are more characteristics to meet in order to be considered "alive". Living organisms are characterized by having information systems that chemically or physically encode the chemical reactions the organisms are capable of (metabolism), are capable of responding to their environments in the short-term through metabolism or long-term through Darwinian evolution, and are somehow separated from their environment physically so that the benefits of their reactions and variability through evolution can provide fitness to the individual. For the earliest forms of life, this competition for resources in the prebiotic milieu likely drove much of the earliest evolution of chemical reactions. A final consideration, as a chemist, is that life is out of equilibrium, and maintains homeostasis through a series of chemical reactions (and thus is a chemical system).[11] An important caveat to all these considerations is that life on other planets may look different than life on Earth, in addition to the fact that life when it first began may look different than life does now. As we delve further into the origin of life in this chapter, we will primarily focus on the information storage and transfer as was necessary at the advent of life, and how that interrelates to the types of chemical reactions preceding life and would have been advantageous for fitness of a protocell.

The study of the origin of life is approached in two main fashions: "top-down", and "bottom-up".[13,14] The "top-down" approach aims to narrow conditions for the origin of life and

the basic reactions of early life using information from biology and evolution, particularly by understanding the biochemistry and genetics of the last universal common ancestor (LUCA) shared between all domains of extant life.[15,16] However, LUCA is a product of extensive evolution from the earliest instances of life— LUCA is a complex single cellular organism that does most of the chemistry that modern cells are capable of today, and evolution would have honed this sophisticated biomolecular machinery. The top-down approach also includes the synthetic biology community, which aims to create a "minimal" cell containing only genes essential for life.[17] However, whether this minimal cell is contingent on the particular form of life that exists on Earth and if it is contingent on the evolutionary history of the cells studied act as caveats in this path of study, and is unlikely to inform on Earth's earliest cells. The "bottom-up" approach, on the other hand, relies on looking at the conditions that existed on the early Earth and tries to reconcile the types of chemistry that would have been possible in the Hadean environment with reactions the field believes are necessary for life: compartmentalization, information storage, energy production, and catalytic capacity.[18] Here, we consider the "bottom-up" approach to the origin of life, and rely on chemistry and Earth history to inform our understanding of the reactions relevant for creating life.

Earth history and life history

Earth accreted from the planetary disk of our solar system approximately 4.5 bya,[19-21] and the conditions of its cooling and progression towards stable oceans and possible habitats for life are integral for narrowing the possible date of the origin of life. Our Earth and Sun were dominated by different chemical conditions 4.5 bya, and understanding these differences is important for determining conditions that are prebiotically plausible. Evidence from zircons suggests oceanic bodies proliferated at least 4.2 bya,[22-25] but the period of extraterrestrial impacts

known as the Late Heavy Bombardment (3.8-4.0 bya) may have affected the presence of these oceans and the stability of different habitats in which life could have formed.[26-29] However, these impacts also could have served as a means to deliver chemical nutrients to the young planet.[22,30,31] The presence of a stable hydrosphere is likely necessary for early life on Earth, although the extent to which this stability needed to be global or local is unknown.

The atmospheric composition is also an important consideration for early Earth chemistry. Volcanic outgassing would be the main source of early volatiles, with our current understanding supporting a mildly-reducing atmosphere, although the extent of the presence of hydrogen is frequently debated.[32-34] Earth's early atmosphere was CO_2 and N_2 rich, with the most notable difference from the modern atmosphere being the low concentration of oxygen.[22,35-37] Oxygen did not proliferate until after photosynthetic organisms outproduced the oxygen sinks of the early Earth, and the lack of oxygen, while an important molecule in multicellular life, was likely important to the chemistry of early life, as oxygen can poison many reaction mechanisms with its reactivity.[38,39] Another consequence of the lower concentration of atmospheric oxygen is that more UV light would reach the surface of the Earth due to the lack of an ozone layer, which would notably affect the stability of compounds and the reactions possible on the surface.[40] In sum, life would have been unable to begin too early into the accretion and cooling process of Earth. The role of impactors could either prevent life or be necessary for prebiotic chemistry, depending on the models considered. Most members of the field would agree that life began between 4.4-3.7 bya,[41,42] but this does not narrow geochemical conditions substantially. This is a large period of time, with great variation of geological conditions—if we understand what reactivity is and is not necessary for life, we can perhaps reconcile the past with our hypotheses for how life began.

Another constraint on the date of the origin of life is the existence of micro/nanofossils, however, these are often highly contested and debated evidence of early life. The oldest potential fossil evidence of cellular life dates to ~3.8 bya, with more conservative fossil evidence dating to 3.5bya.[43-47] However, the dates determined with the fossil record are conservative estimates towards the origin of life— the cellular forms found in these fossils are likely closer to in composition and functionality to LUCA than the original wisps of cellular life. The lack of many unaltered samples of Earth's oldest rocks also means that life could have existed earlier and not been recorded or preserved— the lack of fossil evidence in a time period does not confirm or deny any suppositions about life. Assuming this earlier date in the fossil record is convincing evidence, that means life would have been evolving throughout the late Hadean eon. The Late Heavy Bombardment is believed to have persisted through the beginning of the Archaean eon, which would imply impacts are not incompatible with life and did not sterilize the Earth at this point.[22,48] Since life today is much more complex than what is imagined to be the earliest life on earth, or even LUCA, and the rate of early evolution of life is unknown, this results in more than half a billion year range of time in which life could have possibly arisen. Nevertheless, there were some constants during this period, in terms of planetary abundances, on the atmospheric level, and with microenvironments that would be present. Depending on when in the Late Heavy Bombardment life arose, differing amounts of volatiles and chemicals would have been delivered to the planet, and the requirements for this early chemistry is one question prebiotic chemists aim to address. Ultimately, neither the top-down or bottom-up approaches towards studying the origin of life are mutually exclusive. In fact, a major effort of the top-down approach is to reconcile gaps in the bottom-up approach by studying LUCA and the history of current life on Earth through evolution.

In considering the environment that would lead to life, looking at the gene contents of the last universal common ancestor provides a possible source of constraints. For example, the temperature that earliest life evolved in would be incredibly useful for prebiotic chemists. The temperature range for the origin of life has serious implications for the types of chemistry involved in early life, the time-scales of that chemistry, and the environments that are compatible with the temperature constraints. Whether early life was psychrophilic, mesophilic, or thermophilic can be assessed in a few ways from our tree of life. A few things are not debated about LUCA, mainly being that it is a DNA- and protein-based organism, with complex regulation of its genome, and an anaerobic metabolism.[16] The nucleic acid and amino acid content of early life would likely reflect the temperature in which LUCA existed due to the intrinsic intermolecular interactions that differ between A, C, G, T/U, and the ability of different amino acids to support stable protein structures. However, all of this information about LUCA is obfuscated by years of evolution and statistical analysis. While projected G/C content, the evolution of the protein reverse gyrase, and inferred amino acid content of LUCA support a mesophilic environment for the proliferation of life,[15] other evolutionary analyses support a thermophilic LUCA.[16] These analyses, regardless of which is correct, may still not inform on the environment of earliest life, depending on the time and extent of diversification from the origin of life until LUCA. Trying to reconcile top-down and bottom-up approaches ultimately leaves many questions on the environment of early life, so much of prebiotic chemistry relies on assumptions and ultimately will not be further constrained until life is created in the laboratory or discovered elsewhere in the universe.

Earth today is filled with numerous microcosms of environments of different pH, temperature, and mineral/chemical composition. Due to this diversity of geochemical scenarios,

it is hard to use them as means to rule out any prebiotic chemistry. However, different scenarios have been proposed and can be assessed according to their likelihoods and tested empirically with regards to the chemistry believed to be necessary for life (depending on which theory is prescribed). Ultimately temperature, pH, inorganic species concentration and presence of organics are the main considerations for each possibility as a creche for life. For example, hydrothermal vents are often proposed as a possible location for the creation of life— there is often a carbon source, other nutrients, and possible concentration and charge gradients to be used as a source of energy.[49,50] On the other hand, hydrothermal vents exist at temperature, pressure and pH extremes, which leads to the degradation of many biomolecules.[51,52] Darwin's "warm little ponds" are another commonly assessed environ for life's creation, allowing for periodic concentration and dilution of reactants, with relatively mild temperature and pressure ranges.[53] Whether this provides enough of an energetic and chemical input for early life to be formed and sustained upon is unknown, however. More exotic scenarios can also be proposed, such as life forming in the compartments/pockets of minerals, which could be made compatible with either a hydrothermal or mesophilic geochemical scenario, or life forming in the pockets of liquid between ice in its eutectic phase.[50,54-56] Considering these environments in the context of the origin of life is only useful when a particular series of chemical reactions are being assessed. The list of possible chemical reactions that could be of interest for the origin of life is varied and extensive, but for the purpose of this thesis will be limited to RNA-forming and RNA propagation reactions, which are necessary for information transfer and metabolism in the RNA world theory in the origin of life.

RNA world hypothesis

All extant life follows the "central dogma", with information being stored in DNA, transcribed into RNA, and then translated into proteins that do the work of life.[57] When considering how life started, it is paramount to consider the synthesis and interactions of these biomolecules and how that intersects with conditions on the early Earth. Making nucleotides (DNA/RNA) prebiotically is possible, albeit limited.[58] Amino acid monomers, on the other hand, are synthesized readily in prebiotic conditions.[4,5,59] The next key consideration is if these biomolecules can become bio*polymers*, as this is what their function hinges on in cells (either for information storage or catalysis), and if these polymers are heritable and can be reproduced. While proteins do much of the catalysis in extant cells, they cannot be synthesized without the aid of DNA/RNA in cells, and their information is not heritable without DNA/RNA, suggesting nucleic acids were central in early life. With the discovery that RNA can act as a catalyst in addition to an information carrier,[60,61] RNA became a strong candidate as the main biomolecule of interest at the origin of life— the genetic carrier and catalyst for early life, independent of DNA and protein.[62] The RNA world hypothesis was contributed to by Crick, Gilbert, Orgel, and Woese,[63] and is further supported by alleged remnants of a time where RNA was more important in biological history that remain in cells today. For example, much of cellular metabolism relies on cofactors with RNA-based scaffolds that do not play a catalytic role. Cofactors such as SAM, ATP, coenzyme A, NAD, and FAD all have a nucleotide as a part of their chemical structure, with the "functional" moiety for catalysis elsewhere in the molecule.[64,65] A number of ribozymes are found throughout biology, such as the self-splicing introns, the spliceosome, and perhaps most importantly, the catalytic core of the ribosome.[60] Synthetic ribozymes have also been evolved through laboratory selection to perform many types of reactions, even those that are outside of biology (Diels-Alder) and to bind many small molecules (aptamers).[66,67] The potential

of RNA sequence-space is promising for the RNA world hypothesis, but this hypothesis is not without flaws. Without protein enzymes around, the RNA world, with polymers made only of four nucleobases with limited chemical reactivity, has a limited reaction scope. No native nucleobases are redox-active, and the nucleobase pK_a range is fairly limited for acid-base catalysis, especially in comparison to peptides. Thus, the RNA world likely necessitated the use of a number of environmental cofactors, such as minerals, metal ions, and small peptides. This is a major consideration for the types of environments the RNA world could have occurred in. Additionally, of the main biomolecules in cells, RNA is the least stable. The presence of the 2′-OH in RNA makes it incredibly prone to hydrolysis, which is worse at high temperature and pH extremes.[51] When considering the RNA world in the context of LUCA, since LUCA was a fully DNA-RNA-protein based organism, it must have preceded LUCA in a sufficient amount of time for the evolution of the chemistry that would allow the synthesis of DNA and protein for information carrying and catalytic roles, respectively, and the corresponding loss of much of the RNA-based catalytic functions and genomic role. It is not possible to extrapolate an RNA-based genome from LUCA, as there are no appropriate outgroups alive or preserved to base the phylogeny on. Therefore, much of the history between LUCA and the RNA world, if it existed, is lost.

With the RNA world as a promising lead for creating life, environmental conditions that favor RNA synthesis, reactivity, and stability are integral constraints for prebiotic chemists. While it took decades for a prebiotically plausible synthetic route to RNA to be elucidated, part of this issue was due to the assumption that the nucleobase would be added to the ribose through nucleophilic attack, which is essentially negligible due to the low nucleophilicity of the aromatic nucleobases of RNA/DNA and poor leaving group ability of hydroxide.[68-70] Additionally,

10

synthesizing RNA exclusively, and not any other stereoisomers of the sugar/base is difficult in a prebiotic chemical milieu. The stereoselective synthesis of single diastereomers of sugars remains a difficulty on the early Earth. However, the synthesis of complex mixtures is not as much of an issue as anticipated,[71] and can be honed by chemical selection to favor the propagation of RNA.[16,17,73,74] Nevertheless, while the cyanosulfidic synthesis of nucleotides proposed by Sutherland et al. brings the prebiotic synthesis of RNA into the realm of possibility, there are still a number of caveats. Currently, the only RNA species that can be made through this pathway are pyrimidines— ribocytidine and ribouridine—with purine synthesis limited to DNA, specifically deoxyriboadenine and deoxyriboinosine.[58,75] Considering the nucleic acid alphabet utilized in biology of A, C, T/U, and G, there are clearly still some gaps to be figured out in what the RNA world alphabet may have consisted of. Having a reduced or different alphabet raises issues in what led to the subsequent selection of different nucleobases, or what happened to any alternate nucleobases before the evolution of LUCA. The nucleobases are incredibly important to the subsequent copying chemistry exhibited by RNA monomers into polymers, which we will delve into in Chapter 2, as is the sugar composition of ribose versus an isomer of it. The UV stability of nucleobases, polymers of nucleic acids, and their synthetic precursors is also another consideration that is being currently studied, as the early Earth would have been broadly exposed to the solar spectrum, including UV, due to the lack of ozone layer.[15,22,76] At times this can be a useful fact for prebiotic chemistry, allowing C and U to be interconverted,[75,77] but ultimately the interaction of UV with nucleotides is often detrimental through the creation of photo-radicals. The interaction of nucleotides with UV leads to mutation and cancer in extant cells, and would likely play a considerable role in the mutation rate of early RNA copying. Too high of a mutation rate, and information cannot persist. This necessitates a

certain amount of protection against UV in a prebiotic environment for successful information propagation.[76-79]

One means of protection of nucleotides from their environment is through compartmentalization. Compartmentalization also ensures that when sequences are capable of creating some fitness benefit, that only the sequence that is responsible for said benefit reaps it. However, RNA, a polyanionic molecule, poses some difficulties in encapsulation and in ensuring appropriate exchange of chemical feedstocks from what is being synthesized environmentally and what is encapsulated. Membranes will need to have sufficient permeability for some exchange with the environment, but also to retain competitive sequences.[13] The most traditional form of compartmentalization imagined prebiotically is a lipid bilayer of some sort, mimicking that seen in modern life, known as a protocell.[80] Alternate compartmentalization theories involve coacervates, eutectic spaces between ice, and mineral surfaces/compartments.[13,15] However, these alternate forms of compartmentalization then leave the issue of how our current form of encapsulation evolved. Thus, short, single-chain fatty acids are proposed as sufficiently prebiotically plausible membrane components, due to their increased permeability compared to glycerol-based lipids of current life, and are found in the interstellar medium.[13] Another important issue to consider with these fatty acids is their tendency to interact with positively charged ions, especially metal ions. These interactions can cause precipitation and aggregation, and membranes leak in the presence of high concentrations of metal ions, which becomes an issue with RNA in terms of both its structure and catalysis.[81] This is a major motivation for the work described in the subsequent chapters of this thesis. Oligomers of RNA need cation stabilization to be neutralized, and cations are also integral in the catalysis of RNA polymerization. Without a stable protocell, the benefit of encapsulation is irrelevant.

In order for the RNA world to be a valid hypothesis, there must be a period of time in which, once RNA is synthesized in the prebiotic milieu, that it is able to polymerize into longer strands, and then be copied in a template-directed (information-encoding) manner with enough fidelity to retain information in absence of any enzymatic scaffolds. Even though RNA can act catalytically, it needs to achieve a long enough length in order to have a potentially useful three-dimensional fold. One of the largest differences in prebiotic RNA copying is that it may have occurred with different leaving groups than observed in modern cells. Nucleoside triphosphates (NTPs) are incredibly kinetically stable (the half-life of hydrolysis of alkyl phosphate dianions is 1.1×10^{12} years, and NTP reactivity for nonenzymatic ligation is on the scale of decades),[82,83] but hydrolytically exothermic. Activation of phosphates is necessary to drive the endothermic formation of phosphodiester bonds.[84] This balance of thermodynamic favorability and kinetic stability is important for modern metabolism, and is exploited to drive endothermic reactions. In absence of enzymes, however, NTPs hardly react with each other to form oligonucleotides. Therefore, it has been proposed that early life may have used alternate activation to triphosphates.[70] In the context of the cyanosulfidic synthesis of nucleotides, nucleo-5′-phosphoroimidazolides are prebiotically plausible.[85,86] In fact, the formation of 2-aminoimidazole activated nucleotides requires simple chemical feedstocks that could be delivered through impacts on the early Earth in the Late Heavy Bombardment.[22,30] However, even with this improved activation chemistry, it is difficult to make native RNA oligomers through templated polymerization of substantial length.[70,87] In order to improve this reaction, it has been important to study the mechanism of RNA oligomerization, as has been done over the last sixty years, with recent substantial progress.

Review of the mechanism of nonenzymatic primer extension

For the context of this thesis, the pre-RNA world is referring to the period of time after RNA monomer synthesis but before the chemical capability to copy the length of ribozymes consistently and before RNA replication.[22] Replication refers specifically to when an RNA sequence is able to be copied, and then the daughter strand is also able to be copied, making a copy of the original parent sequence, and not just the complement. Current progress in nonenzymatic RNA replication is limited by poor RNA copying chemistry.[87]

The use of phosphoroimidazolides as an alternative to NTPs was first proposed by Orgel, as a means of better studying nonenzymatic RNA reactions at a reasonable time scale compared to NTPs.[88] 2-methylimidazole activated nucleotides (2MeImpN) were used as the primary phosphate activation form until 2015.[87,89] While the use of nucleotide-phosphoroimidazolides improved nonenzymatic RNA copying chemistry compared to polymerization of NTPs, there are several issues that prevent pre-RNA world reactions from occurring efficiently in our current understanding of RNA phosphoroimidazolide chemistry. One instance is the formation of RNA primers and templates through random oligomerization of RNA monomers.[90-94] Without a diverse set of initial oligomers, templated sequencing is biased from the start, and accessing the full sequence space of RNA is impossible. One notable observation is that there is a metal dependence of nonenzymatic untemplated polymerization, and that the regioselectivity of the polymerization often favors 2′-5′ phosphodiester bonds,[94] which are non-native. While these can be isomerized over time, it takes energetic input.[95] For the purpose of this thesis we assume diverse, primarily 3′-5′ connected oligomers are accessible for templated extension known as nonenzymatic RNA primer extension.

Once oligomers are formed, templated nonenzymatic primer extension is a key hurdle in the progress towards the origin of life. Having a template helps the polymerization of

complement oligomers substantially, increasing the rate 20-fold compared to untemplated extension.[96,97] Some of the best copying chemistry observed is 2MeImpG on a poly-C template, leading to oligomers from monomers in an 80% yield with primarily 3′-5′ phosphodiester bonds, and the maximum length of a 50mer.[89] While promising, homo-polymeric sequences on the early Earth are not only unlikely, but also not functionally useful. Additionally, in attempting RNA replication of this sequence, 2MeImpC does not copy as well on a poly-G template, and thus replication cannot occur in this system.[98] Semi-conservative replication requires that no one nucleotide be more than 50% of the template, and with this data, it is difficult to reconcile how RNA could achieve Darwinian evolution in this model system.[99] Sequence effects show that C and G act as the best copying nucleobases, with A and U much worse,[100] making diverse sequences even more difficult to synthesize nonenzymatically. Mixed sequences are incredibly difficult to extend— on a G template, A and T can inhibit copying, and even a C in the otherwise homo-polymeric template make the reaction less efficient.[101] Side reactions abound, with the formation of pyrophosphate occurring at low pH with the activated monomers,[89] and these side products can act as inhibitors of primer extension. Nevertheless, much was learned about what benefits nonenzymatic primer extension through these model systems, and primer extension has been improved through careful study.

Many of the outstanding issues surrounding the RNA world as imagined at the origin of life have been addressed in the last decade.[102] Regiospecificity refers to the extension of the native 3′-5′ linkage over the nonnative 2′-5′ linkage, which is a notable issue with RNA, as it has two adjacent nucleophilic hydroxyls available for reactivity. This connectivity can be biased by ambient metal ions, with Pb^{2+} resulting in 2′-5′ linkages, while Zn^{2+} predominantly 3′-5′.[103] Lower temperature conditions favor more of 3′-5′ formation , and efficient primer extension also

favors the correct connectivity.[104] The connectivity in this position affects the three-dimensional structure of RNA,[105] and is also external to heritability. While this was a major concern, it was discovered that the heterogeneity of templates can be tolerated, and even the 3′-5′ linkage can be generated from a 2′-5′ linkage through dissipative energy-cycling.[95,104,106] This requires iterative degradation and re-reaction, and requires both a phosphate activating agent and acylating agent as an energy source.[95] The actual extent of a problem these heterogenous templates and primers have is not clear— it is possible the heterogeneity actually addresses another issue had during nonenzymatic primer extension, of the high melting temperature of RNA duplexes once fully copied.[107] This thermodynamic trap becomes a difficulty in the ability of RNA to undergo replication, but is not insurmountable. This also can be circumvented through strand displacement and occasional temperature cycling.[107] Strand-reannealing does remain faster than copying chemistry, but with sufficient splint sequences, this may be sufficient to allow replication.[107]

The prominent issue of A and T/U stymying primer extension served as a severe limitation to the RNA world for decades.[108] TT, GT, and TG serve as partial sequence barriers,[109] and AT, TA, AA, GA, and AG as total barriers of primer extension with methylimidazolides.[109] However, it turns out that this can be overcome with better phosphate activation.[99] In fact, this sequence issue can also be worked through with cold temperature reactions favoring better binding of the reactants, multiple monomer additions, and helper displacement.[99] The discovery of better phosphate activation helped move the field forward by enabling more thorough mechanistic probing of primer extension through kinetics.[87,110-112] The current best performing leaving group is 2-aminoimidazole, which has the additional benefit of having a prebiotically plausible synthesis, and can facilitate mixed-template copying up to seven nucleotides.[87] This is

partly facilitated by activated helper oligomers (even as small as trimers), which improve primer

extension substantially with high enough fidelity to encode functioning ribozymes.[66,113] This

does require that all the relevant trimers for a given sequence exist in solution and are also

activated, however. The benefit of helper oligomers is in part due to increased formation of the

imidazolium-bridged dinucleotide, which we will discuss shortly,[114] as well as pre-organizing the

bridged dinucleotide for primer extension, bringing the phosphate closer to the 3′-OH of the

reacting primer, and in aiding binding of the reactants to the template.[115-117]

Re-contextualizing mechanistic knowledge of nonenzymatic primer extension with the

imidazolium-bridged dinucleotide
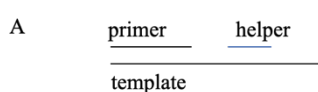
Scheme 1. Nonenzymatic primer extension and relevant compounds. (A) Nonenzymatic primer extension is the extension of a primer in a template directed manner, one nucleotide at a time. A downstream helper oligomer, which may be activated, improves the rates of primer extension[113] (B) Nonenzymatic primer extension requires an activated phosphate in order to occur. In biological systems, the leaving group (LG) is a pyrophosphate. (C) The activated monomer relevant for this dissertation: 2-aminoimidazole activated nucleotides, abbreviated as either 2AIpN or *pN. (D) The imidazolium-bridged dinucleotide is the reactive species in nonenzymatic primer extension with imidazole activation. This compound is abbreviated as Np*pN.
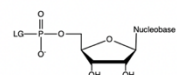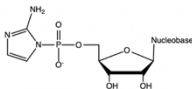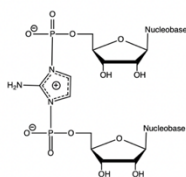
Through the detailed kinetic studies, the understanding of the mechanism of

nonenzymatic primer extension with phosphoroimidazolides was recently revolutionized. It was

discovered that instead of the activated monomer being the functional species in nonenzymatic

17

primer extension, it was instead an imidazolium-bridged dinucleotide that forms in equilibrium with the activated monomer (Scheme 1).[110,118,119] The bridged dinucleotide forms both in solution and on-template,[110,117,118] and it has been determined that the bridged dinucleotide reacts much faster than the activated monomer, accounting for the majority of the observed rate.[120] In fact, the activated monomer inhibits primer extension and increases error rate.[120,121] In addition to studying the role of the bridged dinucleotide, efforts have focused on thoroughly characterizing the structural and electronic requirements for efficient nonenzymatic RNA primer extension.

Structurally, A-form helices have been known to be better for the rates of nonenzymatic primer extension,[122] and recent mechanistic studies have ascertained this is due to the fact that C3′-endo conformation is a better reaction geometry for nonenzymatic primer extension.[74,123] This accounts for some of the poor reactivity of DNA and ANA, which gives a prebiotic scenario for the selection of RNA out of other nucleotide species.[72-74] Additionally, upon binding, RNA activated nucleotides demonstrate conformational switches from C2′-endo to C3′-endo, an important benefit of templated extension.[123] In addition to these conformational effects, systems that have higher kinetics also correlate with the native phosphodiester linkage formation (3′-5′), but for inefficient extension of nucleobases such as A and U, 2′-5′ linkage formation is a more significant issue.[104] The reason behind the slower reaction with A and U at saturating bridged dinucleotide concentrations is not fully understood, although is likely contributed to by the fact A/U has one fewer hydrogen bond than G/C Watson-Crick base pairs. However, the ability of a bridged dinucleotide to react does not necessarily indicate that it extended correctly (although once an incorrect base has added, subsequent extension is often slower).[124]

Electronically, the 3′-OH requires activation for efficient primer extension. This explains why the presence of Mg^{2+} is essential for primer extension.[125] Additionally, solvent kinetic isotope effect studies have shown that the 3′-OH is deprotonated before the rate-limiting step of primer extension, likely due to lowering of the pK_a of the 3′-OH by an associated Mg^{2+}.[74,117] There additionally appears to be a steric benefit to the presence of Mg^{2+}, as it has been shown that the presence of Mg^{2+} favors extension over hydrolysis for template-bound reactants, likely through exclusion of water.[117] In spite of this knowledge, much remains to be ascertained in the interactions of metal ion catalysts in primer extension, particularly the necessary contacts between the catalysts and the reacting RNA, and is the primary focus of this dissertation, particularly in Chapter 2 and Chapter 3. Unfortunately, the level of divalent cations for efficient primer extension exceeds the concentration compatible with stable fatty acid vesicles.[126] Many outstanding issues prevent efficient primer extension of RNA, which we will discuss first by considering nonenzymatic primer extension of alternate genetic systems.

The term "XNA" refers to alternate nucleic acid systems outside of RNA or canonical nucleobases observed in life today. Some of the origins field believes there was a "pre-RNA" world of another sort, where instead of RNA undergoing nonenzymatic chemistry in order to reach the RNA world, another nucleic acid or genetic material existed before RNA and acted as a genetic material and catalyst to form RNA for an RNA world, or bypassed it altogether towards the creation of life as we know it.[68,98] This raises many questions on the evolutionary history of life, and defies the logic of Occam's razor. Nevertheless, much has been learned through studying alternate genetic systems. For example, NP-DNA, where instead of a 3′-OH, a 3′-NH_2 group is present, outperforms RNA in nonenzymatic primer extension universally.[127,128] NP-DNA has the benefit of having a more nucleophilic nucleophile (-NH_2 versus -OH), indicating

the reactivity in nonenzymatic primer extension relies substantially on the nucleophilic activation at the 3′ position. In fact, the presence of other better nucleophiles, such as (prebiotically plausible) phosphate, can react more quickly than the native 3′-OH, forming pyrophosphates instead of phosphodiester linkages.[129] If free phosphates are available to react at the ends of oligomers, they act as inhibitors, however their pyrophosphate products are tolerable to an extent in the template for subsequent copying.[130] Not all XNA is problematic for prebiotic study— the non-biological NP-DNA outperforms the prebiotically plausible reactions, and NP-primer extension of mixed activated nucleotides can be performed for sequences up to 25mers, inside vesicles. There is no metal dependence for NP-DNA due to the innate nucleophilicity of amines compared to hydroxyls.[127] The stepwise yield is 96-97% in NP- systems, something that RNA has yet to approach non-enzymatically.[127] However, NP-DNA is not observed in any biological systems, and the difficulty of evolving enzymes to work with these alternate nucleic acids makes developing selections more labor-intensive, although not impossible. Studying NP-DNA in this way may allow us to learn more about the RNA world by circumventing the low reactivity that RNA has, but ultimately, if RNA were to be a major component in the origin of life, it needs to be determined what conditions lead to its efficient copying and Darwinian evolution. Studies of alternate nucleobases in RNA also fall under this type of work, because it is uncertain yet what the prebiotic cadre of nucleobases was, and how it transitioned to those present in life today. The inability to synthesize ribo-G (or deoxyribo-G), prebiotically poses a major issue, as G is one of the most efficient copying nucleobases in the arsenal of nonenzymatic primer extension.[58] While inosine can act as an alternative, there are no other 8-oxo purines that function well in primer extension.[131] Issues with A and U extension can be diminished with the substitution of 2-thio-

U,[131] but ultimately the greater deviations from life as we know it creates more questions than answers on what the origin of life looked like.

Another prominent issue that remains in the understanding of nonenzymatic RNA primer extension as leading to the RNA world is that of fidelity. It is imperative for sequences to be correctly copied to maintain information, although some mutation is also integral for evolution.[98,121,132] Currently, error rates of primer extension as we understand it to have occurred are too high for conservative replication.[121] Bridged dinucleotides increase fidelity in primer extension compared to activated monomers, likely in part due to the fact two nucleobases need to bind the template in order to react instead of just one.[121,132] As observed in the case of rates, the use of 2-thio-U can help with fidelity by avoiding some wobble error,[13] but the wobble base pairing between G and U remain a significant issue with RNA.[133] Ultimately, error rates are a complex consideration in the evolution of early functional sequences. The extent of tolerance of error will depend not only on the individual sequence and its structural robustness, but also in how much relative fitness it holds compared to other sequences present.[98] Nonetheless, this is not the most pressing issue, as copying long sequences is still a rate-limiting aspect of the pre-RNA world models as we work with them today.

A major issue of RNA copying chemistry and its inability to copy sequences longer than seven mixed nucleotides is the fact that much of copying chemistry is on the same timescale as substrate degradation.[87] For example, the "limiting concentration" of reactant, in which the initial monomer concentration equally undergoes oligomerization and hydrolysis for 2MeImpG is 1.7 mM at 37 °C, 0.36 mM 1 °C with 50 mM poly-C template (expressed as monomer equivalents).[134] These concentrations may even be difficult to reach depending on the geochemical scenario, however, there are a few means for which this issue could have been

worked around prebiotically. One possibility is that long sequences were not actually necessary to be copied, that instead shorter oligomers were be synthesized and ligated together to functional sequences.[66,135] However, ligation chemistry is even worse than primer extension,[136] although it has been successfully used to make functional ribozyme sequences.[66] Having the correct sequences made and associated together to become functional ribozymes is statistically difficult, and the sequestration for this correct combination of oligomers becomes a problem for the prebiotic soup. Perhaps this localization issue is a means of early selection, but this is yet to be experimentally exhibited. Another alternative along this vein is that the genome of early life could be many overlapping short oligonucleotides. No particular oligomer would contain all the information, but together, information would be encoded (this is known as the virtual circular genome aka VCG).[137] The plausibility of this hypothesis is currently being experimentally evaluated.

Another consideration to favor primer reactivity over degradation by hydrolysis is through better catalysis of primer extension through exclusion of water through directed reactivity. Being in an aqueous solution makes it hard for a low concentration, weak nucleophile such as the 3′-OH to react with the reactant, compared to the 55M solvent H_2O.[122] The presence of a template helps in raising the local concentration of the 3′-OH, and Mg^{2+} helps favor extension over hydrolysis.[117] RNA outcompetes other nucleic acids by increasing preorganization of the reactant for primer extension on the template,[74,122] but further activation of the nucleophile to outcompete water would be another means for improving primer extension fitness in the prebiotic soup.[122] An alternate means to counteract the issue of the intrinsic rates of reaction and hydrolysis is to counter hydrolysis of activated nucleotides by continuous reactivation.[86] Sutherland and coworkers have proposed a prebiotically-plausible activating

agent, methyl isocyanide[86] as a solution to this issue, and it is compatible with vesicles, primer extension, and can be used to favor bridged dinucleotide formation.[138]

Progress towards the RNA world is ongoing, although limited by these aforementioned difficulties in making the best RNA copying chemistry compatible with vesicles for protocells. Even with the best RNA copying chemistry, the formation of functional ribozymes capable of self-replication and Darwinian evolution starting from only nonenzymatic primer extension has not been exhibited. However, we believe our system of phosphate activation is consistent with downstream RNA world activities. RNA ligases have also been evolved with 2-aminoimidazole activated oligomers, showing the compatibility of 2-aminoimidazole with the RNA world as well.[139] The compatibility of these ribozymes and their reactivity with protocells is important, because this will allow for some protocells to have increased fitness compared to others and lead to Darwinian evolution.[13] A major goal in nonenzymatic primer extension is to find ways to chemically minimize the concentrations of metal ions required for primer extension, which may ultimately involve an RNA polymerase capable of utilizing 2-aminoimidazole activated monomers. The focus of this dissertation is to understand the role of divalent metal ions with RNA chemistry, particularly nonenzymatic RNA primer extension, with this ultimate goal in mind.

RNA polymerization in biology

Nonenzymatic primer extension and its efficiency remains an issue in the big picture of transitioning to the RNA world. The formation of a phosphodiester bond requires the reaction between the 3′-OH of RNA and the phosphate of an activated nucleotide, expelling its leaving group. The 3′-OH gets deprotonated at some point along its reaction coordinate. Catalysis of phosphodiester bond formation could involve activating the 3′-OH nucleophile through

deprotonation or stabilization of the deprotonation product, by stabilizing the leaving group of the reactant, and by organizing the various reactants for in-line attack. This type of catalysis can occur enzymatically and nonenzymatically, and by understanding the enzymatic mechanism, we can learn about the possible forms of chemistry that were harnessed by enzymatic scaffolds in evolutionary history. RNA polymerases are the proteins evolved to do this chemistry. The predominant theory of DNA and RNA polymerase catalysis centers on a two-metal ion mechanism.[140,141] This involves nucleophilic activation of the 3′-OH and charge stabilization of the pyrophosphate leaving group.[142] Since the leaving group in nonenzymatic primer extension is differently charged than pyrophosphate, this already informs us that two metal ions may not be necessary nonenzymatically in our understanding of the pre-RNA world. The amount of metal ions involved enzymatically is somewhat contested, with some models involving additional outer- sphere metal ions involved in the mechanism. At its core, however, Mg^{2+} is key in the phosphodiester bond formation, and is used to orient the 3′-OH for in-line nucleophilic attack of dNTPs.[143] This is key for an S_N2 mechanism, having the nucleophile and leaving group oxygen at apical positions in a trigonal bipyramidal transition state.[144] The Mg^{2+} ions involved in this catalysis have inner-sphere interactions with carboxylate amino acids side groups.[145] These metal ions also help stabilize the electronegative transition state, as RNA is not capable of stabilizing this on its own, due to its moieties primarily being neutral and negatively charge at neutral pH.[144] This is an objective benefit of protein catalysis, as a full range of acid and base moieties are present in enzymes, but ribozymes would be dependent heavily on metal ion co-catalysts and cofactors.[146,147] Ribozymes, being long, negatively-charged oligomers, require metal ions such as Mg^{2+} in order to overcome intramolecular repulsion and have three-dimensional folds. In the Group I intron, mutations that eliminate core Mg^{2+} binding eliminates domain structure[148]. These

core metal ion interactions have been proposed to be the RNA counterpart of hydrophobic core.[148] Additionally, metal ions are key for ribosome structure and function, although not sufficient for catalysis.[149] Therefore, understanding the role of metal ions nonenzymatically in primer extension is key for understanding its mechanism of reactivity and how to improve it.

Metal ion interactions with nucleic acids

In order to understand the roles of metal ions in nonenzymatic RNA primer extension, it is important to understand the possible modes of metal ion interaction with RNA. At the core of this is the type of coordination observed with the metal ion.[146] Broadly, metal ion coordination occurs on a spectrum between inner-sphere and outer-sphere interactions. Inner-sphere coordination involves the direct contact between the metal ion and a ligand, not mediated by solvent.[150] The extent of electronic interaction observed between main group cations and their ligands is less so than in transition metal inorganic complexes, but still has important effects on the chemical behavior of the coordinated species. Outer-sphere interactions are mediated through a cation-solvent sphere, and primarily contribute hydrogen bond stabilization or electrostatic stabilization. Outer-sphere specific reactions can be rescued by exchange-inert complexes and high concentrations of monovalent ions, as observed with the hammerhead ribozyme.[151,152] Cobalt hexammine mimics the same network of hydrogen bonds that a fully hydrated magnesium cation can provide,[153] albeit with a slight decrease in hydrogen-bonding capacity,[152] and if a catalytic mechanism requires inner-sphere catalysis, these exchange inert species can potently inhibit a reaction.[154]

A more detailed way of categorizing metal ion interactions with substrates along the inner-sphere/outer-sphere spectrum is through a continuum characterizing an ion as between diffuse and site-bound as defined by Bowman et al.[155] This is a key consideration in catalysis as

it determines where a metal ion is in a reaction coordinate, and its transience or continuity in the mechanism. A fully diffuse ion can be characterized as "free". A free ion has no ligands, and diffuses freely in all dimensions, and acts primarily in contributing electrostatic stabilization to an RNA duplex, but not specifically to structure.[155] These thermodynamic contributions are necessary, but the identity of the metal ion elementally is exchangeable as long as ionic strength can be maintained. Some diffusive ions can be found localized to general structures in RNA, with monovalent ions often localizing to the major groove of B-form DNA,[156] and divalent ions observable in the minor groove in DNA.[156,157] Ultimately these interactions are dominated by the electrostatic field produced by the nucleic acid. The ion atmosphere is affected by the charge density of the ion, with divalent ions having tighter associations than monovalent ions and are key in neutralizing the nucleic acid.[158]

A slightly more associated ion would be characterized as "condensed".[155] These ions can have up to six RNA ligands, and are usually sodium or potassium ions.[155] Despite their connectivity to ligands, they are not structurally relevant. "Glassy" ions have up to four RNA ligands, and are not very diffusive.[155] For Mg^{2+} ions, these would be inner-sphere ligands. A classic example of this type of ion would be a Mg^{2+} clamped in a 10-membered ring,[155] and can have some diffusivity/transience. Chelated ions are the most restricted in motion, and are very associated with their ligands, such as in a G-quadruplex. These contacts are also held in place to a specific location through electrostatics.[159] Along the continuum from diffusive to chelated, there are energetic tradeoffs to be made. It costs energy to dehydrate metal ions, and there is an entropic deficit, but this is balanced with an electrostatic benefit depending on the ligand.[159] Depending on the balance of these features, an interaction between a metal ion and a ligand will have varying transience. The ability to study these ions depends on the strength of these ligand

interactions and their dynamics. It is possible to measure binding (interactions more restricted than a free ion) using ^{25}Mg NMR, EPR, or ITC depending on experimental setup, but all require strong (sub-millimolar) interactions with a measurable thermodynamic benefit to be observed.[160,161] Nevertheless, a binding site can still exist with weaker interactions. For this dissertation, a binding site is defined as having a defined set of contacts, and saturates at integral stoichiometries, following a mass action law.[159] The binding of metal ions in nonenzymatic RNA primer extension will be addressed thoroughly in Chapters 2 and 4. With these metal-ligand interactions defined, we will now discuss more RNA-cation interactions and their possible effects in nonenzymatic primer extension.

Much of the variety in metal ion interactions with RNA depends on the elemental properties of the ion.[162] Size and charge density influence the "hardness" and "softness" of a metal ion. These are on a spectrum with one another, with hard ions being smaller and charge dense, and soft ions being larger and polarizable in their charge density.[163] Hard ions interact best with other hard charges, and the same can be said about soft ions. This helps predict which atoms an ion interacts with most strongly (Mg-O, Cd-S for example). These effects are key to the studies in Chapter 3. The hardness or softness of an ion dictates where it will interact with primarily on a nucleic acid. Harder ions prefer the electrostatic localization to phosphate oxygens, while softer ions will interact and intercalate with nucleobases, disrupting stacking and helicity of a duplex, preventing stable structures.[146] Cation-pi interactions are favored with softer ions.[164] During heating and cooling, transition metals such as Cu^{2+} and Zn^{2+} maintain interactions with the nucleobases, favoring rewinding of DNA duplexes.[165] Metals in between these on the hardness/softness scale, such as Co^{2+}, Ni^{2+}, Mn^{2+} only partially unwind/rewind helices.[165] Interestingly, Mn^{2+} acts to much improve duplex stability until it reaches a Mn^{2+}(DNA

base-pair) ratio of 1.5, and beyond this point duplex stability drops, due to interactions with the nucleobases.[165] Harder ions are better at stabilizing ribozyme folds, likely due to preferred interactions with the phosphodiester backbone.[166] Hardness and softness correlate with the level of ionic vs. covalent character a metal-ligand interaction has, which additionally correlates with how mobile vs. fixed said ion is in a structure.

These structural effects aren't the only properties of metal ions with nucleic acids. Lewis acidity is another characteristic to consider, and is reflected by the pK_a of an inner sphere water of an ion. The lower the pK_a, the more Lewis acidic, and the more metal-hydroxo species present at a given pH, possibly favoring hydrolysis of RNA oligomers and reactants.[167,168] However, metal-hydroxo species are not all equivalently soluble, which poses an issue of metal identity if a reaction is dependent on hydroxo-species for catalysis. $M(OH)^+$ species are more nucleophilic than water, but less nucleophilic than free hydroxide,[73] which could possibly be taken advantage of as a catalyst if the ion has a higher local concentration to the reactants than free hydroxide. More Lewis acidic ions can stabilize negative charges more readily, and activate nucleophiles more readily as well. The Lewis acidity likely has an optimum for a given RNA reaction, between being active enough to aid in catalysis, but not too active that it destroys the RNA and the necessary reactants.[163] In the case of observed RNA reactions, metal ions can activate water or ribose nucleophilic hydroxyl by withdrawing electrons and facilitating deprotonation through these Lewis acid effects.[144] In addition to these effects, transition metals additionally have the possibility to donate or accept electrons through redox chemistry, which is key in metabolic reactions. However, radical reactions that often correspond with single electron redox are incredibly high energy, and can be destructive to RNA.[146] All of these catalytic capacities described in this section require inner-sphere interactions, and are incredibly dependent on the

other ligands present in the inner-sphere. Chelation of an ion by certain ligands may have been integral in certain prebiotic reactions, either through improving local concentration or by mitigating the side reactivity of hydrolysis for an ion. For example, citrate mitigates RNA degradation by Mg^{2+}.[81] However, too many coordination contacts through chelation can prevent reactivity (such as the case with M^{2+}-EDTA), especially if induced intramolecularity through inner-sphere coordination of reactants is important in a reaction mechanism.[168] This is observed in phosphodiester formation in enzymes, and is a possibly important feature of catalysis in prebiotic nonenzymatic RNA primer extension.

Magnesium and RNA are intricately involved with one another in cellular life, and it is unknown if this has always been the case or is a product of evolutionary pressures after the great oxygenation event removed oxygen sensitive cations from the ocean.[155] It is possible it was prebiotically important as well, as it is the most abundant non alkali earth metal, at 2% of the crust.[155] Mg^{2+} is highly soluble at near neutral pH ranges, but also at alkaline pH, which is not the case for many transition metals.[155] Mg^{2+} exists in aqueous solutions as an octahedral hexa-aqua species,[169] and with biological molecules is primarily found with oxygen and nitrogen ligands, and never with sulfur ligands.[169] Water, carboxylates, and alcohols are the preferred inner-sphere ligands for Mg^{2+}.[170] When undergoing ligand exchange, Mg^{2+} tends to undergo a dissociative, S_N1-like intermediate, with this 5 membered species only 5-10 kcal higher in energy than its octahedral configuration.[169,171] Mg^{2+} compared to other ions undergoes relatively slow ligand exchange, as a result of its high enthalpy of hydration.[155,172] Nevertheless, the on and off rates for Mg^{2+} inner-sphere contacts is on the order of $10^8\,M^{-1}s^{-1}$, which is incredibly fast compared to the time-scale of the experiments in this thesis.[161] In general, Mg^{2+} binds weakly to proteins, with a $K_a < 10^5\,M^{-1}$, which may in part be related to the relatively high cellular concentrations of 0.5

mM.[172,173] This may have been the case in prebiotic scenarios as well, as Mg^{2+} would have had a high oceanic concentration, and therefore not have as much of a selection to bind tightly. Common analogs for Mg^{2+} include Mn^{2+}, Zn^{2+} due to similar size, hydration, and charge density, and inert Cr and Co complexes if explicitly trying to mimic the fully-hydrated magnesium species.[172, 174] In biological systems, Mn^{2+} tends to bind tighter and catalyze faster reactions,[172] likely due to both the stronger Lewis acidity of Mn^{2+} and its 100-fold faster inner-sphere exchange rates.[175] The benefits of Mn^{2+} in the RNA world will be explored in Chapter 4.

Metal ions and nonenzymatic primer extension

The key issue with metal ions in the RNA world is that in order to have fast enough primer extension rates compared to hydrolysis, dozens of millimolar of metal ion are required, if not hundreds of millimolar. Some metal ions induce precipitation of fatty acid components involved in vesicles due to strong association with carboxylates. High concentrations (>5mM) of Mg^{2+} induces vesicle leakage.[126] Citrate can stabilize fatty acid vesicles in the presence of Mg^{2+}, but there is a catalytic cost to primer extension rates.[81] It is unclear if this is due to less free Mg^{2+}, or if the tridentate-bound Mg^{2+} is less catalytically efficient for nonenzymatic primer extension. It is additionally possible to offset the leakage of vesicles due to high concentrations of metal ions by linking an enzymatic/ribozymatic activity to producing more membrane components, which could be an important early source of fitness and Darwinian evolution.[126] Ultimately, in nonenzymatic primer extension, the 3′-OH needs to be deprotonated during the reaction coordinate. The bridged dinucleotide that is the main reactant for primer extension does not need leaving group stabilization as NTPs do in polymerization, as the leaving group is a neutral species. It was recently determined that the deprotonation of primer 3′-OH occurs before rate-limiting step,[74] and it is suspected that the metal ion aids in this deprotonation by lowering

the pK$_a$ of the 3'-OH to 9.1.[74] Additional questions about the contacts of the metal ion during nonenzymatic primer extension motivated research direction in this dissertation. Chapter 2 describes the kinetic and binding studies performed to elucidate the contacts involved between metal ions and cations in nonenzymatic primer extension.

In Chapter 3, we address the mechanism nonenzymatic primer extension occurs through and further elucidate the interactions in the reaction center, as determined using 5'-thiophosphoro-2-aminoimidazolides and metal rescue experiments. Previously, the mechanism of nonenzymatic primer extension has only been studied through hydrolysis models, and often phosphoryl transfer mechanisms are very sensitive to the moieties attached to the oxygens. It is unclear if nonenzymatic primer extension occurs through a concerted, associative, or dissociative substitution mechanism, and how it compares to the enzymatic S$_N$2 mechanism of primer extension.[144,145] Previous computational studies report the hydrolysis of phosphoroimidazolides follows a loose transition state with extensive scissile bond cleavage, and inversion of configuration at the phosphate,[176] and our experiments support the hypotheses proposed from these studies.

Chapter 4 uses the application of the understanding of the mechanism of cation catalysis in nonenzymatic primer extension to study the prebiotic plausibility of Mn^{2+}as a substitute for Mg^{2+}, and its effects on ligation, mixed sequence primer extension, and fidelity.

In sum, this thesis aims to study weak and transient metal ion-nucleic acid interactions that are integral for pre-RNA world reactions, with the ultimate goal of informing of possible co-catalysts and binding scaffolds that can improve nonenzymatic primer extension to allow progression towards sequence replication and ribozyme synthesis.

Bibliography

(1)   Wohler, M. F. Artificial Formation of Urea. *The Philosophical Magazine* **1828**, *4* (22), 309–310. https://doi.org/10.1080/14786442808674820.

(2)   Ramberg, P. J. The Death of Vitalism and The Birth of Organic Chemistry: Wohler's Urea Synthesis and the Disciplinary Identity of Organic Chemistry. *Ambix* **2013**. https://doi.org/10.1179/amb.2000.47.3.170.

(3)   Lipman, T. O. Wohler's Preparation of Urea and the Fate of Vitalism. *J. Chem. Educ.* **1964**, *41* (8), 452. https://doi.org/10.1021/ed041p452.

(4)   Urey, H. C. On the Early Chemical History of the Earth and the Origin of Life. *Proc Natl Acad Sci U S A* **1952**, *38* (4), 351–363.

(5)   Miller, S. L. A Production of Amino Acids Under Possible Primitive Earth Conditions. *Science* **1953**, *117* (3046), 528–529. https://doi.org/10.1126/science.117.3046.528.

(6)   Orgel, L. E. The Origin of Life—a Review of Facts and Speculations. *Trends in Biochemical Sciences* **1998**, *23* (12), 491–495. https://doi.org/10.1016/S0968-0004(98)01300-0.

(7)   Szostak, J. W. Attempts to Define Life Do Not Help to Understand the Origin of Life. *Journal of Biomolecular Structure and Dynamics* **2012**, *29* (4), 599–600. https://doi.org/10.1080/073911012010524998.

(8)   Chyba, C. F.; McDonald, G. D. The Origin of Life in the Solar System: Current Issues. *Annual Review of Earth and Planetary Sciences* **1995**, *23* (1), 215–249. https://doi.org/10.1146/annurev.ea.23.050195.001243.

(9)   Ma, W. The Origin of Life: A Problem of History, Chemistry, and Evolution. *Chemistry & Biodiversity* **2014**, *11* (12), 1998–2010. https://doi.org/10.1002/cbdv.201400188.

(10) Zuckerman, B.; Hart, M. H. *Extraterrestrials: Where Are They?*; CUP Archive, 1995.

(11) Joyce, G. F. Evolution in an RNA World. *Cold Spring Harb Symp Quant Biol* **2009**, *74*, 17–23. https://doi.org/10.1101/sqb.2009.74.004.

(12) Chodasewicz, K. Evolution, Reproduction and Definition of Life. *Theory Biosci.* **2014**, *133* (1), 39–45. https://doi.org/10.1007/s12064-013-0184-5.

(13) Joyce, G. F.; Szostak, J. W. Protocells and RNA Self-Replication. *Cold Spring Harbor Perspectives in Biology* **2018**, *10* (9), a034801. https://doi.org/10.1101/cshperspect.a034801.

(14) Sutherland, J. D. The Origin of Life—Out of the Blue. *Angewandte Chemie International Edition* **2016**, *55* (1), 104–121. https://doi.org/10.1002/anie.201506585.

(15) Cantine, M. D.; Fournier, G. P. Environmental Adaptation from the Origin of Life to the Last Universal Common Ancestor. *Orig Life Evol Biosph* **2018**, *48* (1), 35–54. https://doi.org/10.1007/s11084-017-9542-5.

(16) Weiss, M. C.; Sousa, F. L.; Mrnjavac, N.; Neukirchen, S.; Roettger, M.; Nelson-Sathi, S.; Martin, W. F. The Physiology and Habitat of the Last Universal Common Ancestor. *Nat Microbiol* **2016**, *1* (9), 1–8. https://doi.org/10.1038/nmicrobiol.2016.116.

(17) Towards synthesis of a minimal cell https://www.embopress.org/doi/epdf/10.1038/msb4100090 (accessed 2022 -03 -09). https://doi.org/10.1038/msb4100090.

(18) Hengeveld, R. Two Approaches to the Study of the Origin of Life. *Acta Biotheor* **2007**, *55* (2), 97–131. https://doi.org/10.1007/s10441-007-9017-6.

(19) Kleine, T.; Mezger, K.; Palme, H.; Münker, C. The W Isotope Evolution of the Bulk Silicate Earth: Constraints on the Timing and Mechanisms of Core Formation and

Accretion. *Earth and Planetary Science Letters* **2004**, *228* (1), 109–123. https://doi.org/10.1016/j.epsl.2004.09.023.

(20)   Touboul, M.; Kleine, T.; Bourdon, B.; Palme, H.; Wieler, R. Late Formation and Prolonged Differentiation of the Moon Inferred from W Isotopes in Lunar Metals. *Nature* **2007**, *450* (7173), 1206–1209. https://doi.org/10.1038/nature06428.

(21)   THE Hf-W ISOTOPIC SYSTEM AND THE ORIGIN OF THE EARTH AND MOON | Annual Review of Earth and Planetary Sciences https://www-annualreviews-org.ezp-prod1.hul.harvard.edu/doi/10.1146/annurev.earth.33.092203.122614 (accessed 2022 -04 -04).

(22)   The Astrobiology Primer v2.0 | Astrobiology https://www-liebertpub-com.ezp-prod1.hul.harvard.edu/doi/10.1089/ast.2015.1460 (accessed 2022 -01 -05).

(23)   Cavosie, A. J.; Valley, J. W.; Wilde, S. A.; E.i.m.f. Magmatic Δ18O in 4400–3900 Ma Detrital Zircons: A Record of the Alteration and Recycling of Crust in the Early Archean. *Earth and Planetary Science Letters* **2005**, *235* (3), 663–681. https://doi.org/10.1016/j.epsl.2005.04.028.

(24)   Wilde, S. A.; Valley, J. W.; Peck, W. H.; Graham, C. M. Evidence from Detrital Zircons for the Existence of Continental Crust and Oceans on the Earth 4.4 Gyr Ago. *Nature* **2001**, *409* (6817), 175–178. https://doi.org/10.1038/35051550.

(25)   Mojzsis, S. J.; Harrison, T. M.; Pidgeon, R. T. Oxygen-Isotope Evidence from Ancient Zircons for Liquid Water at the Earth's Surface 4,300 Myr Ago. *Nature* **2001**, *409* (6817), 178–181. https://doi.org/10.1038/35051557.

(26) Schoenberg, R.; Kamber, B. S.; Collerson, K. D.; Moorbath, S. Tungsten Isotope Evidence from ∼3.8-Gyr Metamorphosed Sediments for Early Meteorite Bombardment of the Earth. *Nature* **2002**, *418* (6896), 403–405. https://doi.org/10.1038/nature00923.

(27) Strom, R. G.; Malhotra, R.; Ito, T.; Yoshida, F.; Kring, D. A. The Origin of Planetary Impactors in the Inner Solar System. *Science* **2005**, *309* (5742), 1847–1850. https://doi.org/10.1126/science.1113544.

(28) Wetherill, G. W. Occurrence of Giant Impacts During the Growth of the Terrestrial Planets. *Science* **1985**, *228* (4701), 877–879. https://doi.org/10.1126/science.228.4701.877.

(29) Frei, R.; Rosing, M. T. Search for Traces of the Late Heavy Bombardment on Earth— Results from High Precision Chromium Isotopes. *Earth and Planetary Science Letters* **2005**, *236* (1), 28–40. https://doi.org/10.1016/j.epsl.2005.05.024.

(30) Todd, Z. R.; Öberg, K. I. Cometary Delivery of Hydrogen Cyanide to the Early Earth. *Astrobiology* **2020**, *20* (9), 1109–1120. https://doi.org/10.1089/ast.2019.2187.

(31) Chyba, C.; Sagan, C. Endogenous Production, Exogenous Delivery and Impact-Shock Synthesis of Organic Molecules: An Inventory for the Origins of Life. *Nature* **1992**, *355*, 125–132. https://doi.org/10.1038/355125a0.

(32) Tian, F.; Toon, O. B.; Pavlov, A. A.; De Sterck, H. A Hydrogen-Rich Early Earth Atmosphere. *Science* **2005**, *308* (5724), 1014–1017. https://doi.org/10.1126/science.1106983.

(33) Comment on "A hydrogen-rich early Earth atmosphere" - PubMed https://pubmed-ncbi-nlm-nih-gov.ezp-prod1.hul.harvard.edu/16400134/ (accessed 2022 -04 -05).

(34)   Tian, F.; Toon, O. B.; Pavlov, A. A. Response to Comment on "A Hydrogen-Rich Early Earth Atmosphere." *Science* **2006**, *311* (5757), 38–38. https://doi.org/10.1126/science.1118412.

(35)   Pavlov, A. A.; Kasting, J. F.; Brown, L. L.; Rages, K. A.; Freedman, R. Greenhouse Warming by CH4 in the Atmosphere of Early Earth. *J Geophys Res* **2000**, *105* (E5), 11981–11990. https://doi.org/10.1029/1999je001134.

(36)   Haqq-Misra, J. D.; Domagal-Goldman, S. D.; Kasting, P. J.; Kasting, J. F. A Revised, Hazy Methane Greenhouse for the Archean Earth. *Astrobiology* **2008**, *8* (6), 1127–1137. https://doi.org/10.1089/ast.2007.0197.

(37)   How Earth's Atmosphere Evolved to an Oxic State: A Status Report. *Earth and Planetary Science Letters* **2005**, *237* (1–2), 1–20. https://doi.org/10.1016/j.epsl.2005.06.013.

(38)   Holland, H. D. The Oxygenation of the Atmosphere and Oceans. *Philos Trans R Soc Lond B Biol Sci* **2006**, *361* (1470), 903–915. https://doi.org/10.1098/rstb.2006.1838.

(39)   THE EARLY HISTORY OF ATMOSPHERIC OXYGEN: Homage to Robert M. Garrels | Annual Review of Earth and Planetary Sciences https://www-annualreviews-org.ezp-prod1.hul.harvard.edu/doi/10.1146/annurev.earth.33.092203.122711 (accessed 2022 -04 - 04).

(40)   Sagan, C.; Mullen, G. Earth and Mars: Evolution of Atmospheres and Surface Temperatures. *Science* **1972**, *177* (4043), 52–56. https://doi.org/10.1126/science.177.4043.52.

(41)   Betts, H. C.; Puttick, M. N.; Clark, J. W.; Williams, T. A.; Donoghue, P. C. J.; Pisani, D. Integrated Genomic and Fossil Evidence Illuminates Life's Early Evolution and Eukaryote

Origin. *Nat Ecol Evol* **2018**, *2* (10), 1556–1562. https://doi.org/10.1038/s41559-018-0644-x.

(42)   Dodd, M. S.; Papineau, D.; Grenne, T.; Slack, J. F.; Rittner, M.; Pirajno, F.; O'Neil, J.; Little, C. T. S. Evidence for Early Life in Earth's Oldest Hydrothermal Vent Precipitates. *Nature* **2017**, *543* (7643), 60–64. https://doi.org/10.1038/nature21377.

(43)   Schopf, J. W. Microfossils of the Early Archean Apex Chert: New Evidence of the Antiquity of Life. *Science* **1993**, *260*, 640–646. https://doi.org/10.1126/science.260.5108.640.

(44)   Schopf, J. W.; Kudryavtsev, A. B.; Agresti, D. G.; Wdowiak, T. J.; Czaja, A. D. Laser--Raman Imagery of Earth's Earliest Fossils. *Nature* **2002**, *416* (6876), 73–76. https://doi.org/10.1038/416073a.

(45)   Mojzsis, S. J.; Arrhenius, G.; McKeegan, K. D.; Harrison, T. M.; Nutman, A. P.; Friend, C. R. Evidence for Life on Earth before 3,800 Million Years Ago. *Nature* **1996**, *384* (6604), 55–59. https://doi.org/10.1038/384055a0.

(46)   Brasier, M. D.; Green, O. R.; Lindsay, J. F.; McLoughlin, N.; Steele, A.; Stoakes, C. Critical Testing of Earth's Oldest Putative Fossil Assemblage from the ~3.5Ga Apex Chert, Chinaman Creek, Western Australia. *Precambrian Research* **2005**, *140* (1), 55–102. https://doi.org/10.1016/j.precamres.2005.06.008.

(47)   De Gregorio, B. T.; Sharp, T. G.; Flynn, G. J.; Wirick, S.; Hervig, R. L. Biogenic Origin for Earth's Oldest Putative Microfossils. *Geology* **2009**, *37* (7), 631–634. https://doi.org/10.1130/G25683A.1.

(48)   O, A.; Sj, M. Microbial Habitability of the Hadean Earth during the Late Heavy Bombardment. *Nature* **2009**, *459* (7245). https://doi.org/10.1038/nature08015.

(49)   Baross, J. A.; Hoffman, S. E. Submarine Hydrothermal Vents and Associated Gradient Environments as Sites for the Origin and Evolution of Life. *Origins Life Evol Biosphere* **1985**, *15* (4), 327–345. https://doi.org/10.1007/BF01808177.

(50)   Martin, W.; Baross, J.; Kelley, D.; Russell, M. J. Hydrothermal Vents and the Origin of Life. *Nat Rev Microbiol* **2008**, *6* (11), 805–814. https://doi.org/10.1038/nrmicro1991.

(51)   Kanavarioti, A. Kinetics of the Hydrolysis of Guanosine 5′-Phospho-2-Methylimidazolide. *Origins Life Evol Biosphere* **1986**, *17* (1), 85–103. https://doi.org/10.1007/BF01809815.

(52)   Mojarro, A.; Jin, L.; Szostak, J. W.; Head, J. W.; Zuber, M. T. In Search of the RNA World on Mars. *Geobiology* **2021**, *19* (3), 307–321. https://doi.org/10.1111/gbi.12433.

(53)   Follmann, H.; Brownson, C. Darwin's Warm Little Pond Revisited: From Molecules to the Origin of Life. *Naturwissenschaften* **2009**, *96* (11), 1265–1292. https://doi.org/10.1007/s00114-009-0602-1.

(54)   RNA Oligomerization in Laboratory Analogues of Alkaline Hydrothermal Vent Systems http://www.liebertpub.com/doi/epub/10.1089/ast.2014.1280 (accessed 2022 -01 -05). https://doi.org/10.1089/ast.2014.1280.

(55)   Ferris, J. P. Mineral Catalysis and Prebiotic Synthesis: Montmorillonite-Catalyzed Formation of RNA. *Elements* **2005**, *1* (3), 145–149. https://doi.org/10.2113/gselements.1.3.145.

(56)   Kanavarioti, A.; Monnard, P.-A.; Deamer, D. W. Eutectic Phases in Ice Facilitate Nonenzymatic Nucleic Acid Synthesis. *Astrobiology* **2001**, *1* (3), 271–281. https://doi.org/10.1089/15311070152757465.

(57)   Crick, F. Central Dogma of Molecular Biology. **1970**, 3.

(58) Xu, J.; Chmela, V.; Green, N. J.; Russell, D. A.; Janicki, M. J.; Góra, R. W.; Szabla, R.; Bond, A. D.; Sutherland, J. D. Selective Prebiotic Formation of RNA Pyrimidine and DNA Purine Nucleosides. *Nature* **2020**, *582* (7810), 60–66. https://doi.org/10.1038/s41586-020-2330-9.

(59) Eliash, R.; Weissbuch, I.; Weygand, M. J.; Kjaer, K.; Leiserowitz, L.; Lahav, M. Structure and Reactivity in Langmuir Films of Amphiphilic Alkyl and Thio-Alkyl Esters of α-Amino Acids at the Air/Water Interface. *J. Phys. Chem. B* **2004**, *108* (22), 7228–7240. https://doi.org/10.1021/jp049381j.

(60) The Intervening Sequence RNA of Tetrahymena Is an Enzyme. *Science*.

(61) Gilbert, W. Origin of Life: The RNA World. *Nature* **1986**, *319* (6055), 618–618. https://doi.org/10.1038/319618a0.

(62) Hager, A. J.; Pollard, J. D.; Szostak, J. W. Ribozymes: Aiming at RNA Replication and Protein Synthesis. *Chemistry & Biology* **1996**, *3* (9), 717–725. https://doi.org/10.1016/S1074-5521(96)90246-X.

(63) Ferris, J. P. The Prebiotic Leslie Orgel: Some of His Contributions to Our Understanding of the Origins of Life. *Orig Life Evol Biosph* **1997**, *27* (5), 431–435. https://doi.org/10.1023/A:1006592627693.

(64) The "Strong" RNA World Hypothesis: Fifty Years Old | Astrobiology https://www-liebertpub-com.ezp-prod1.hul.harvard.edu/doi/full/10.1089/ast.2012.0868 (accessed 2022 -01 -05).

(65) Nucleic acids: function and potential for abiogenesis | Quarterly Reviews of Biophysics | Cambridge Core https://www-cambridge-org.ezp-prod1.hul.harvard.edu/core/journals/quarterly-reviews-of-biophysics/article/nucleic-acids-

function-and-potential-for-abiogenesis/842529B9BDAD6E86F7919827725C1931 (accessed 2022 -01 -05).

(66)   Adamala, K.; Engelhart, A. E.; Szostak, J. W. Generation of Functional RNAs from Inactive Oligonucleotide Complexes by Non-Enzymatic Primer Extension. *J. Am. Chem. Soc.* **2015**, *137* (1), 483–489. https://doi.org/10.1021/ja511564d.

(67)   Ellington, A. D.; Szostak, J. W. In Vitro Selection of RNA Molecules That Bind Specific Ligands. *Nature* **1990**, *346* (6287), 818–822. https://doi.org/10.1038/346818a0.

(68)   Anastasi, C.; Buchet, F. F.; Crowe, M. A.; Parkes, A. L.; Powner, M. W.; Smith, J. M.; Sutherland, J. D. RNA: Prebiotic Product, or Biotic Invention? *Chemistry & Biodiversity* **2007**, *4* (4), 721–739. https://doi.org/10.1002/cbdv.200790060.

(69)   Sutherland, J. D. Opinion: Studies on the Origin of Life — the End of the Beginning. *Nat Rev Chem* **2017**, *1* (2), 1–7. https://doi.org/10.1038/s41570-016-0012.

(70)   Orgel, L. E. Some Consequences of the RNA World Hypothesis. 8.

(71)   An optimal degree of physical and chemical heterogeneity for the origin of life? | Philosophical Transactions of the Royal Society B: Biological Sciences https://royalsocietypublishing.org/doi/full/10.1098/rstb.2011.0140 (accessed 2022 -04 - 05).

(72)   Kim, S. C.; Zhou, L.; Zhang, W.; O'Flaherty, D. K.; Rondo-Brovetto, V.; Szostak, J. W. A Model for the Emergence of RNA from a Prebiotically Plausible Mixture of Ribonucleotides, Arabinonucleotides, and 2′-Deoxynucleotides. *J. Am. Chem. Soc.* **2020**, *142* (5), 2317–2326. https://doi.org/10.1021/jacs.9b11239.

(73)   Kim, S. C.; O'Flaherty, D. K.; Giurgiu, C.; Zhou, L.; Szostak, J. W. The Emergence of RNA from the Heterogeneous Products of Prebiotic Nucleotide Synthesis. *J. Am. Chem. Soc.* **2021**, *143* (9), 3267–3279. https://doi.org/10.1021/jacs.0c12955.

(74)   Giurgiu, C.; Fang, Z.; Aitken, H. R. M.; Kim, S. C.; Pazienza, L.; Mittal, S.; Szostak, J. W. Structure–Activity Relationships in Nonenzymatic Template-Directed RNA Synthesis. *Angewandte Chemie* **2021**, *133* (42), 23107–23114. https://doi.org/10.1002/ange.202109714.

(75)   Patel, B. H.; Percivalle, C.; Ritson, D. J.; Duffy, C. D.; Sutherland, J. D. Common Origins of RNA, Protein and Lipid Precursors in a Cyanosulfidic Protometabolism. *Nature Chem* **2015**, *7* (4), 301–307. https://doi.org/10.1038/nchem.2202.

(76)   Todd, Z. R.; Szabla, R.; Szostak, J. W.; Sasselov, D. D. UV Photostability of Three 2-Aminoazoles with Key Roles in Prebiotic Chemistry on the Early Earth. *Chem. Commun.* **2019**, *55* (70), 10388–10391. https://doi.org/10.1039/C9CC05265H.

(77)   Todd, Z. R.; Fahrenbach, A. C.; Ranjan, S.; Magnani, C. J.; Szostak, J. W.; Sasselov, D. D. Ultraviolet-Driven Deamination of Cytidine Ribonucleotides Under Planetary Conditions. *Astrobiology* **2020**, *20* (7), 878–888. https://doi.org/10.1089/ast.2019.2182.

(78)   Ranjan, S.; Kufner, C. L.; Lozano, G. G.; Todd, Z. R.; Haseki, A.; Sasselov, D. D. UV Transmission in Natural Waters on Prebiotic Earth. *Astrobiology* **2022**, *22* (3), 242–262. https://doi.org/10.1089/ast.2020.2422.

(79)   Todd, Z. R.; Szostak, J. W.; Sasselov, D. D. Shielding from UV Photodamage: Implications for Surficial Origins of Life Chemistry on the Early Earth. *ACS Earth Space Chem.* **2021**, *5* (2), 239–246. https://doi.org/10.1021/acsearthspacechem.0c00270.

(80)   Chen, I. A.; Salehi-Ashtiani, K.; Szostak, J. W. RNA Catalysis in Model Protocell Vesicles. *J. Am. Chem. Soc.* **2005**, *127* (38), 13213–13219. https://doi.org/10.1021/ja051784p.

(81)   Adamala, K.; Szostak, J. W. Nonenzymatic Template-Directed RNA Synthesis Inside Model Protocells. *Science* **2013**, *342* (6162), 1098–1100. https://doi.org/10.1126/science.1241888.

(82)   Lad, C.; Williams, N. H.; Wolfenden, R. The Rate of Hydrolysis of Phosphomonoester Dianions and the Exceptional Catalytic Proficiencies of Protein and Inositol Phosphatases. *Proceedings of the National Academy of Sciences* **2003**, *100* (10), 5607–5610. https://doi.org/10.1073/pnas.0631607100.

(83)   Rohatgi, R.; Bartel, D. P.; Szostak, J. W. Kinetic and Mechanistic Analysis of Nonenzymatic, Template-Directed Oligoribonucleotide Ligation. *Journal of the American Chemical Society* **1996**, *118* (14), 3332–3339. https://doi.org/10.1021/ja953712b.

(84)   Sosson, M.; Richert, C. Enzyme-Free Genetic Copying of DNA and RNA Sequences. *Beilstein J. Org. Chem.* **2018**, *14*, 603–617. https://doi.org/10.3762/bjoc.14.47.

(85)   Fahrenbach, A. C.; Giurgiu, C.; Tam, C. P.; Li, L.; Hongo, Y.; Aono, M.; Szostak, J. W. Common and Potentially Prebiotic Origin for Precursors of Nucleotide Synthesis and Activation. *Journal of the American Chemical Society* **2017**, *139* (26), 8780–8783. https://doi.org/10.1021/jacs.7b01562.

(86)   Liu, Z.; Wu, L.-F.; Xu, J.; Bonfio, C.; Russell, D. A.; Sutherland, J. D. Harnessing Chemical Energy for the Activation and Joining of Prebiotic Building Blocks. *Nat. Chem.* **2020**, *12* (11), 1023–1028. https://doi.org/10.1038/s41557-020-00564-3.

(87)   Li, L.; Prywes, N.; Tam, C. P.; O'Flaherty, D. K.; Lelyveld, V. S.; Izgu, E. C.; Pal, A.;
Szostak, J. W. Enhanced Nonenzymatic RNA Copying with 2-Aminoimidazole Activated
Nucleotides. *Journal of the American Chemical Society* **2017**, *139* (5), 1810–1813.
https://doi.org/10.1021/jacs.6b13148.

(88)   Weimann, B. J.; Lohrmann, R.; Orgel, L. E.; Schneider-Bernloehr, H.; Sulston, J. E.
Template-Directed Synthesis with Adenosine-5'-Phosphorimidazolide. *Science* **1968**, *161*
(3839), 387–387. https://doi.org/10.1126/science.161.3839.387.

(89)   Inoue, T.; Orgel, L. E. Oligomerization of (Guanosine 5′-Phosphor)-2-Methylimidazolide
on Poly(C): An RNA Polymerase Model. *Journal of Molecular Biology* **1982**, *162* (1),
201–217. https://doi.org/10.1016/0022-2836(82)90169-3.

(90)   Kanavarioti, A.; Monnard, P. A.; Deamer, D. W. Eutectic Phases in Ice Facilitate
Nonenzymatic Nucleic Acid Synthesis. *Astrobiology* **2001**, *1* (3), 271–281.
https://doi.org/10.1089/15311070152757465.

(91)   Monnard, P.-A.; Apel, C. L.; Kanavarioti, A.; Deamer, D. W. Influence of Ionic Inorganic
Solutes on Self-Assembly and Polymerization Processes Related to Early Forms of Life:
Implications for a Prebiotic Aqueous Medium. *Astrobiology* **2002**, *2* (2), 139–152.
https://doi.org/10.1089/15311070260192237.

(92)   Miyakawa, S.; Joshi, P. C.; Gaffey, M. J.; Gonzalez-Toril, E.; Hyland, C.; Ross, T.; Rybij,
K.; Ferris, J. P. Studies in the Mineral and Salt-Catalyzed Formation of RNA Oligomers.
Origins of Life and Evolution of Biospheres **2006**, *36* (4), 343–361.
https://doi.org/10.1007/s11084-006-9009-6.

(93) Sawai, H. Oligonucleotide Formation Catalyzed by Divalent Metal Ions. The Uniqueness of the Ribosyl System. *Journal of Molecular Evolution* **1988**, *27* (3), 181–186. https://doi.org/10.1007/BF02100072.

(94) Sawai, H. Catalysis of Internucleotide Bond Formation by Divalent Metal Ions. *Journal of the American Chemical Society* **1976**, *98* (22), 7037–7039. https://doi.org/10.1021/ja00438a050.

(95) Mariani, A.; Sutherland, J. D. Non-Enzymatic RNA Backbone Proofreading through Energy-Dissipative Recycling. *Angewandte Chemie International Edition* **2017**, *56* (23), 6563–6566. https://doi.org/10.1002/anie.201703169.

(96) Kanavarioti, A.; White, D. H. Kinetic Analysis of the Template Effect in Ribooligoguanylate Elongation. *Origins Life Evol Biosphere* **1987**, *17* (3), 333–349. https://doi.org/10.1007/BF02386472.

(97) Kervio, E.; Claasen, B.; Steiner, U. E.; Richert, C. The Strength of the Template Effect Attracting Nucleotides to Naked DNA. *Nucleic Acids Research* **2014**, *42* (11), 7409–7420. https://doi.org/10.1093/nar/gku314.

(98) Joyce, G. F.; Orgel, L. E. For Conditions See Www.Cshlpress.Com/Copyright. 1 Prospects for Understanding the Origin of the RNA World.

(99) Vogel, S. R.; Richert, C. Adenosine Residues in the Template Do Not Block Spontaneous Replication Steps of RNA. *Chem. Commun.* **2007**, No. 19, 1896. https://doi.org/10.1039/b702768k.

(100) Wu, T.; Orgel, L. E. Nonenzymic template-directed synthesis on hairpin oligonucleotides. 2. Templates containing cytidine and guanosine residues

https://pubs.acs.org/doi/pdf/10.1021/ja00040a002 (accessed 2022 -01 -05).

https://doi.org/10.1021/ja00040a002.

(101) Hill, A. R.; Orgel, L. E.; Wu, T. The Limits of Template-Directed Synthesis with
Nucleoside-5′-Phosphoro(2-Methyl)Imidazolides. *Origins Life Evol Biosphere* **1993**, *23*
(5), 285–290. https://doi.org/10.1007/BF01582078.

(102) Szostak, J. W. The Eightfold Path to Non-Enzymatic RNA Replication. *J Syst Chem* **2012**,
3 (1), 2. https://doi.org/10.1186/1759-2208-3-2.

(103) Lohrmann, R.; Bridson, P.; Orgel, L. Efficient Metal-Ion Catalyzed Template-Directed
Oligonucleotide Synthesis. *Science* **1980**, *208* (4451), 1464–1465.
https://doi.org/10.1126/science.6247762.

(104) Giurgiu, C.; Li, L.; O'Flaherty, D. K.; Tam, C. P.; Szostak, J. W. A Mechanistic
Explanation for the Regioselectivity of Nonenzymatic RNA Primer Extension. *J. Am.
Chem. Soc.* **2017**, *139* (46), 16741–16747. https://doi.org/10.1021/jacs.7b08784.

(105) Wasner, M.; Arion, D.; Borkow, G.; Noronha, A.; Uddin, A. H.; Parniak, M. A.; Damha,
M. J. Physicochemical and Biochemical Properties of 2',5'-Linked RNA and 2',5'-
RNA:3',5'-RNA "Hybrid" Duplexes ,. *Biochemistry* **1998**, *37* (20), 7478–7486.
https://doi.org/10.1021/bi980160b.

(106) Engelhart, A. E.; Powner, M. W.; Szostak, J. W. Functional RNAs Exhibit Tolerance for
Non-Heritable 2′–5′ versus 3′–5′ Backbone Heterogeneity. *Nature Chem* **2013**, *5* (5), 390–
394. https://doi.org/10.1038/nchem.1623.

(107) Zhou, L.; Kim, S. C.; Ho, K. H.; O'Flaherty, D. K.; Giurgiu, C.; Wright, T. H.; Szostak, J.
W. Non-Enzymatic Primer Extension with Strand Displacement. *eLife* **2019**, *8*, e51888.
https://doi.org/10.7554/eLife.51888.

(108) Joyce, G. F.; Orgel, L. E. Non-Enzymatic Template-Directed Synthesis on RNA Random Copolymers: Poly(C,A) Templates. *Journal of Molecular Biology* **1988**, *202* (3), 677–681. https://doi.org/10.1016/0022-2836(88)90297-5.

(109) Wu, T.; Orgel, L. E. Nonenzymatic Template-Directed Synthesis on Hairpin Oligonucleotides. 3. Incorporation of Adenosine and Uridine Residues. *J. Am. Chem. Soc.* **1992**, *114* (21), 7963–7969. https://doi.org/10.1021/ja00047a001.

(110) Walton, T.; Szostak, J. W. A Highly Reactive Imidazolium-Bridged Dinucleotide Intermediate in Nonenzymatic RNA Primer Extension. *Journal of the American Chemical Society* **2016**, *138* (36), 11996–12002. https://doi.org/10.1021/jacs.6b07977.

(111) Kanavarioti, A.; Bernasconi, C. F.; Alberas, D. J.; Baird, E. E. Kinetic Dissection of Individual Steps in the Poly(C)-Directed Oligoguanylate Synthesis from Guanosine 5'-Monophosphate 2-Methylimidazolide. *J. Am. Chem. Soc.* **1993**, *115* (19), 8537–8546. https://doi.org/10.1021/ja00072a003.

(112) Kanavarioti, A.; Baird, E. E.; Hurley, T. B.; Carruthers, J. A.; Gangopadhyay, S. Unique Catalysis and Regioselectivity Observed in the Poly(C)-Directed RNA Dimer Formation from 2-MeImpG: Kinetic Analysis as a Function of Monomer and Polymer Concentration. *J. Org. Chem.* **1999**, *64* (22), 8323–8333. https://doi.org/10.1021/jo991216q.

(113) Prywes, N.; Blain, J. C.; Del Frate, F.; Szostak, J. W. Nonenzymatic Copying of RNA Templates Containing All Four Letters Is Catalyzed by Activated Oligonucleotides. *eLife* **2016**, *5*, e17756. https://doi.org/10.7554/eLife.17756.

(114) R. Vogel, S.; Deck, C.; Richert, C. Accelerating Chemical Replication Steps of RNA Involving Activated Ribonucleotides and Downstream-Binding Elements. *Chemical Communications* **2005**, *0* (39), 4922–4924. https://doi.org/10.1039/B510775J.

(115) Zhang, W.; Tam, C. P.; Zhou, L.; Oh, S. S.; Wang, J.; Szostak, J. W. Structural Rationale for the Enhanced Catalysis of Nonenzymatic RNA Primer Extension by a Downstream Oligonucleotide. *Journal of the American Chemical Society* **2018**, *140* (8), 2829–2840. https://doi.org/10.1021/jacs.7b11750.

(116) Tam, C. P.; Fahrenbach, A. C.; Björkbom, A.; Prywes, N.; Izgu, E. C.; Szostak, J. W. Downstream Oligonucleotides Strongly Enhance the Affinity of GMP to RNA Primer–Template Complexes. *J. Am. Chem. Soc.* **2017**, *139* (2), 571–574. https://doi.org/10.1021/jacs.6b09760.

(117) Walton, T.; Pazienza, L.; Szostak, J. W. Template-Directed Catalysis of a Multistep Reaction Pathway for Nonenzymatic RNA Primer Extension. *Biochemistry* **2019**, *58* (6), 755–762. https://doi.org/10.1021/acs.biochem.8b01156.

(118) Zhang, W.; Tam, C. P.; Walton, T.; Fahrenbach, A. C.; Birrane, G.; Szostak, J. W. Insight into the Mechanism of Nonenzymatic RNA Primer Extension from the Structure of an RNA-GpppG Complex. *Proc Natl Acad Sci USA* **2017**, *114* (29), 7659–7664. https://doi.org/10.1073/pnas.1704006114.

(119) Zhang, W.; Walton, T.; Li, L.; Szostak, J. W. Crystallographic Observation of Nonenzymatic RNA Primer Extension. 15.

(120) Walton, T.; Szostak, J. W. A Kinetic Model of Nonenzymatic RNA Polymerization by Cytidine-5′-Phosphoro-2-Aminoimidazolide. *Biochemistry* **2017**, *56* (43), 5739–5747. https://doi.org/10.1021/acs.biochem.7b00792.

(121) Duzdevich, D.; Carr, C. E.; Ding, D.; Zhang, S. J.; Walton, T. S.; Szostak, J. W. Competition between Bridged Dinucleotides and Activated Mononucleotides Determines the Error Frequency of Nonenzymatic RNA Primer Extension. 11.

(122) Kaiser, A.; Richert, C. Nucleotide-Based Copying of Nucleic Acid Sequences without Enzymes. *J. Org. Chem.* **2013**, *78* (3), 793–799. https://doi.org/10.1021/jo3025779.

(123) Activated Ribonucleotides Undergo a Sugar Pucker Switch upon Binding to a Single-Stranded RNA Template | Journal of the American Chemical Society https://pubs.acs.org/doi/abs/10.1021/ja212027q (accessed 2022 -04 -04).

(124) Rajamani, S.; Ichida, J. K.; Antal, T.; Treco, D. A.; Leu, K.; Nowak, M. A.; Szostak, J. W.; Chen, I. A. Effect of Stalling after Mismatches on the Error Catastrophe in Nonenzymatic Nucleic Acid Replication. *J. Am. Chem. Soc.* **2010**, *132* (16), 5880–5885. https://doi.org/10.1021/ja100780p.

(125) Kozlov, I. A.; Orgel, L. E. Nonenzymatic Template-Directed Synthesis of RNA from Monomers. **2000**, *34* (6), 9.

(126) Adamala, K. P.; Engelhart, A. E.; Szostak, J. W. Collaboration between Primitive Cell Membranes and Soluble Catalysts. *Nat Commun* **2016**, *7* (1), 11041. https://doi.org/10.1038/ncomms11041.

(127) O'Flaherty, D. K.; Zhou, L.; Szostak, J. W. Nonenzymatic Template-Directed Synthesis of Mixed-Sequence 3′-NP-DNA up to 25 Nucleotides Long Inside Model Protocells. *J. Am. Chem. Soc.* **2019**, *141* (26), 10481–10488. https://doi.org/10.1021/jacs.9b04858.

(128) Zhang, S.; Zhang, N.; Blain, J. C.; Szostak, J. W. Synthesis of N3′-P5′-Linked Phosphoramidate DNA by Nonenzymatic Template-Directed Primer Extension. *J Am Chem Soc* **2013**, *135* (2), 924–932. https://doi.org/10.1021/ja311164j.

(129) Kanavarioti, A.; Rosenbach, M. T.; Brian Hurley, T. Nucleotides as Nucleophiles: Reactions of Nucleotides with Phosphoimidazolide Activated Guanosine. *Origins Life Evol Biosphere* **1992**, *21* (4), 199–217. https://doi.org/10.1007/BF01809856.

(130) Wright, T. H.; Giurgiu, C.; Zhang, W.; Radakovic, A.; O'Flaherty, D. K.; Zhou, L.; Szostak, J. W. Prebiotically Plausible "Patching" of RNA Backbone Cleavage through a 3′–5′ Pyrophosphate Linkage. *J. Am. Chem. Soc.* **2019**, *141* (45), 18104–18112. https://doi.org/10.1021/jacs.9b08237.

(131) Kim, S. C.; O'Flaherty, D. K.; Zhou, L.; Lelyveld, V. S.; Szostak, J. W. Inosine, but None of the 8-Oxo-Purines, Is a Plausible Component of a Primordial Version of RNA. *Proc Natl Acad Sci USA* **2018**, *115* (52), 13318–13323. https://doi.org/10.1073/pnas.1814367115.

(132) Duzdevich, D.; Carr, C. E.; Szostak, J. W. Deep Sequencing of Non-Enzymatic RNA Primer Extension. 13.

(133) Zielinski, M.; Kozlov, I. A.; Orgel, L. E. A Comparison of RNA with DNA in Template-Directed Synthesis. *Helvetica Chimica Acta* **2000**, *83* (8), 1678–1684. https://doi.org/10.1002/1522-2675(20000809)83:8<1678::AID-HLCA1678>3.0.CO;2-P.

(134) Kanavarioti, A.; Chang, S.; Alberas, D. J. Limiting Concentrations of Activated Mononucleotides Necessary for Poly(C)-Directed Elongation of Oligoguanylates. *J Mol Evol* **1990**, *31* (6), 462–469. https://doi.org/10.1007/BF02102072.

(135) Zhou, L.; O'Flaherty, D. K.; Szostak, J. W. Assembly of a Ribozyme Ligase from Short Oligomers by Nonenzymatic Ligation. *J. Am. Chem. Soc.* **2020**, *142* (37), 15961–15965. https://doi.org/10.1021/jacs.0c06722.

(136) Zhou, L.; O'Flaherty, D. K.; Szostak, J. W. Template-Directed Copying of RNA by Non-enzymatic Ligation. *Angew. Chem. Int. Ed.* **2020**, *59* (36), 15682–15687. https://doi.org/10.1002/anie.202004934.

(137) Zhou, L.; Ding, D.; Szostak, J. W. The Virtual Circular Genome Model for Primordial RNA Replication. 12.

(138) Zhang, S. J.; Duzdevich, D.; Szostak, J. W. Potentially Prebiotic Activation Chemistry Compatible with Nonenzymatic RNA Copying. *J. Am. Chem. Soc.* **2020**, *142* (35), 14810–14813. https://doi.org/10.1021/jacs.0c05300.

(139) Walton, T.; DasGupta, S.; Duzdevich, D.; Oh, S. S.; Szostak, J. W. In Vitro Selection of Ribozyme Ligases That Use Prebiotically Plausible 2-Aminoimidazole–Activated Substrates. *Proc Natl Acad Sci USA* **2020**, *117* (11), 5741–5748. https://doi.org/10.1073/pnas.1914367117.

(140) Carvalho, A. T. P.; Fernandes, P. A.; Ramos, M. J. The Catalytic Mechanism of RNA Polymerase II. *Journal of Chemical Theory and Computation* **2011**, *7* (4), 1177–1188. https://doi.org/10.1021/ct100579w.

(141) Castro, C.; Smidansky, E.; Maksimchuk, K. R.; Arnold, J. J.; Korneeva, V. S.; Gotte, M.; Konigsberg, W.; Cameron, C. E. Two Proton Transfers in the Transition State for Nucleotidyl Transfer Catalyzed by RNA- and DNA-Dependent RNA and DNA Polymerases. *Proceedings of the National Academy of Sciences* **2007**, *104* (11), 4267–4272. https://doi.org/10.1073/pnas.0608952104.

(142) Horton, N. C.; Perona, J. J. Making the Most of Metal Ions. *nature structural biology* **2001**, *8* (4), 4.

(143) Batra, V. K.; Beard, W. A.; Shock, D. D.; Krahn, J. M.; Pedersen, L. C.; Wilson, S. H. Magnesium-Induced Assembly of a Complete DNA Polymerase Catalytic Complex. *Structure* **2006**, *14* (4), 757–766. https://doi.org/10.1016/j.str.2006.01.011.

(144) Fedor, M. J. The Role of Metal Ions in RNA Catalysis. *Current Opinion in Structural Biology* **2002**, *12* (3), 289–295. https://doi.org/10.1016/S0959-440X(02)00324-X.

(145) Cleland, W. W.; Hengge, A. C. Enzymatic Mechanisms of Phosphate and Sulfate Transfer. *Chemical Reviews* **2006**, *106* (8), 3252–3278. https://doi.org/10.1021/cr050287o.

(146) *Bioinorganic Chemistry*; Bertini, I., Ed.; University Science Books: Mill Valley, Calif, 1994.

(147) Giacobelli, V. G.; Fujishima, K.; Lepšík, M.; Tretyachenko, V.; Kadavá, T.; Bednárová, L.; Novák, P.; Hlouchová, K. In Vitro Evolution Reveals Primordial RNA-Protein Interaction Mediated by Metal Cations. bioRxiv August 1, 2021, p 2021.08.01.454623. https://doi.org/10.1101/2021.08.01.454623.

(148) Cate, J. H.; Hanna, R. L.; Doudna, J. A. A Magnesium Ion Core at the Heart of a Ribozyme Domain. *Nat Struct Mol Biol* **1997**, *4* (7), 553–558. https://doi.org/10.1038/nsb0797-553.

(149) K. Lenz, T.; M. Norris, A.; V. Hud, N.; Dean Williams, L. Protein-Free Ribosomal RNA Folds to a near-Native State in the Presence of Mg 2+. *RSC Advances* **2017**, *7* (86), 54674–54681. https://doi.org/10.1039/C7RA08696B.

(150) Black, C. B.; Huang, H.-W.; Cowan, J. A. Biological Coordination Chemistry of Magnesium, Sodium, and Potassium Ions. Protein and Nucleotide Binding Sites. *Coordination Chemistry Reviews* **1994**, *135–136*, 165–202. https://doi.org/10.1016/0010-8545(94)80068-5.

(151) Curtis, E. A.; Bartel, D. P. The Hammerhead Cleavage Reaction in Monovalent Cations. *RNA* **2001**, *7* (4), 546–552. https://doi.org/10.1017/S1355838201002357.

(152) Cowan, J. A. Magnesium Activation of Nuckase Enzymes - the Importance of Water. 4.

(153) Jou, R.; Cowan, J. A. Ribonuclease H Activation by Inert Transition-Metal Complexes. Mechanistic Probes for Metallocofactors: Insights on the Metallobiochemistry of Divalent Magnesium Ion. *Journal of the American Chemical Society* **1991**, *113* (17), 6685–6686. https://doi.org/10.1021/ja00017a056.

(154) Kucharski, L. M.; Lubbe, W. J.; Maguire, M. E. Cation Hexaammines Are Selective and Potent Inhibitors of the CorA Magnesium Transport System. *Journal of Biological Chemistry* **2000**, *275* (22), 16767–16773. https://doi.org/10.1074/jbc.M001507200.

(155) Bowman, J. C.; Lenz, T. K.; Hud, N. V.; Williams, L. D. Cations in Charge: Magnesium Ions in RNA Folding and Catalysis. *Current Opinion in Structural Biology* **2012**, *22* (3), 262–272. https://doi.org/10.1016/j.sbi.2012.04.006.

(156) Hud, N. V.; Polak, M. DNA–Cation Interactions: The Major and Minor Grooves Are Flexible Ionophores. *Current Opinion in Structural Biology* **2001**, *11* (3), 293–301. https://doi.org/10.1016/S0959-440X(00)00205-0.

(157) Hud, N. V.; Feigon, J. Localization of Divalent Metal Ions in the Minor Groove of DNA A-Tracts. *Journal of the American Chemical Society* **1997**, *119* (24), 5756–5757. https://doi.org/10.1021/ja9704085.

(158) Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D. Understanding Nucleic Acid–Ion Interactions. *Annual Review of Biochemistry* **2014**, *83* (1), 813–841. https://doi.org/10.1146/annurev-biochem-060409-092720.

(159) Draper, D. E. A Guide to Ions and RNA Structure. *RNA* **2004**, *10* (3), 335–343. https://doi.org/10.1261/rna.5205404.

(160) Cowan, J. A. Coordination Chemistry of Magnesium Ions and 5S RRNA (Escherichia Coli): Binding Parameters, Ligand Symmetry, and Implications for Activity. *Journal of*

the American Chemical Society **1991**, *113* (2), 675–676.
https://doi.org/10.1021/ja00002a046.

(161) Huang, H.-W.; Cowan, J. A. Metallobiochemistry of the Magnesium Ion. *European Journal of Biochemistry* **1994**, *219* (1–2), 253–260. https://doi.org/10.1111/j.1432-1033.1994.tb19936.x.

(162) Cammack, R.; Hughes, M. Considerations for the Specification of Enzyme Assays Involving Metal Ions. **2022**.

(163) Frederiksen, J. K.; Fong, R.; Piccirilli, J. A. Chapter 8. Metal Ions in RNA Catalysis. In *Nucleic Acid-Metal Ion Interactions*; Royal Society of Chemistry: Cambridge, 2008; pp 260–306. https://doi.org/10.1039/9781847558763-00260.

(164) McFail-Isom, L.; Shui, X.; Williams, L. D. Divalent Cations Stabilize Unstacked Conformations of DNA and RNA by Interacting with Base Π Systems. 7.

(165) Eichhorn, G. L.; Shin, Y. Ae. Interaction of Metal Ions with Polynucleotides and Related Compounds. XII. The Relative Effect of Various Metal Ions on DNA Helicity. *Journal of the American Chemical Society* **1968**, *90* (26), 7323–7328.
https://doi.org/10.1021/ja01028a024.

(166) Koculi, E.; Hyeon, C.; Thirumalai, D.; Woodson, S. A. Charge Density of Divalent Metal Cations Determines RNA Stability. *Journal of the American Chemical Society* **2007**, *129* (9), 2676–2682. https://doi.org/10.1021/ja068027r.

(167) Breslow, R.; Huang, D. L. Effects of Metal Ions, Including Mg2+ and Lanthanides, on the Cleavage of Ribonucleotides and RNA Model Compounds. *Proceedings of the National Academy of Sciences* **1991**, *88* (10), 4080–4083. https://doi.org/10.1073/pnas.88.10.4080.

(168) Herschlag, D.; Jencks, W. P. Catalysis of the Hydrolysis of Phosphorylated Pyridines by Mg(OH)+: A Possible Model for Enzymic Phosphoryl Transfer. *Biochemistry* **1990**, *29* (21), 5172–5179. https://doi.org/10.1021/bi00473a025.

(169) Bock, C. W.; Kaufman, A.; Glusker, J. P. Coordination of Water to Magnesium Cations. *Inorganic Chemistry* **1994**, *33* (3), 419–427. https://doi.org/10.1021/ic00081a007.

(170) Harding, M. M. The Geometry of Metal–Ligand Interactions Relevant to Proteins. *Acta Cryst D, Acta Cryst Sect D, Acta Crystallogr D, Acta Crystallogr Sect D, Acta Crystallogr D Biol Crystallogr, Acta Crystallogr Sect D Biol Crystallogr* **1999**, *55* (8), 1432–1443. https://doi.org/10.1107/S0907444999007374.

(171) Banyasz, J. L.; Stuehr, J. E. Interactions of Divalent Metal Ions with Inorganic and Nucleoside Phosphates. III. Temperature Dependence of the Magnesium(II)--Adenosine 5'-Triphosphate, --Adenosine 5'-Diphosphate, and --Cytidine 5'-Diphosphate Systems. *Journal of the American Chemical Society* **1973**, *95* (22), 7226–7231. https://doi.org/10.1021/ja00803a007.

(172) Cowan, J. A. Metal Activation of Enzymes in Nucleic Acid Biochemistry. *Chemical Reviews* **1998**, *98* (3), 1067–1088. https://doi.org/10.1021/cr960436q.

(173) Cowan, J. A. Structural and Catalytic Chemistry of Magnesium-Dependent Enzymes. 11.

(174) Kim, S.; Cowan, J. A. Inert Cobalt Complexes as Mechanistic Probes of the Biochemistry of Magnesium Cofactors. Application to Topoisomerase I. *Inorganic Chemistry* **1992**, *31* (17), 3495–3496. https://doi.org/10.1021/ic00043a001.

(175) Katz, A. K.; Glusker, J. P.; Beebe, S. A.; Bock, C. W. Calcium Ion Coordination: A Comparison with That of Beryllium, Magnesium, and Zinc. *Journal of the American Chemical Society* **1996**, *118* (24), 5752–5763. https://doi.org/10.1021/ja953943i.

(176) Li, L.; Lelyveld, V. S.; Prywes, N.; Szostak, J. W. Experimental and Computational Evidence for a Loose Transition State in Phosphoroimidazolide Hydrolysis. *Journal of the American Chemical Society* **2016**, *138* (12), 3986–3989. https://doi.org/10.1021/jacs.6b00784.

Chapter 2

Mechanism of divalent ion catalysis in nonenzymatic RNA primer extension

Mechanism of divalent ion catalysis in nonenzymatic RNA primer extension

Pazienza, LT; Walton T, and JW Szostak

Abstract

Nonenzymatic template-directed RNA primer extension is a crucial component of the RNA world hypothesis of the origin of life. However, nonenzymatic primer extension is fairly slow, often within an order of magnitude of hydrolysis of reactants, preventing formation of long RNA sequences. The presence of cations improves the rate of primer extension, but at the concentrations for optimal primer extension, is incompatible with encapsulation by protocells. It is of interest to understand the mechanism of metal ion catalysis in nonenzymatic primer extension to enable future development of co-catalysts to improve primer extension and minimize total metal ion concentration. We find through kinetic analysis of primer extension in a variety of sequence and cation conditions that divalent ions activate the 3′-OH through inner-sphere contacts. We additionally show a metal bound-hydroxide does not aid substantially in catalyzing nonenzymatic primer extension. Lastly, we find that the binding of magnesium ions is improved when the 3′-OH and the reacting phosphate are closer together, suggesting the metal ion interacts with the phosphate oxygen of the reacting imidazolium-bridged dinucleotide. The sum of these contacts provides a potential scaffold for prebiotic co-catalysts for nonenzymatic primer extension.

Introduction

The RNA world hypothesis, which relies on the dual capacity of RNA as a carrier of genetic information and as a catalyst, is a strongly supported early stage in the evolution of life.[1] However, the evolution of ribozymes must have been preceded by an earlier period during which functional RNA polymers were assembled chemically, without the aid of enzymatic machinery. Small RNA oligonucleotides can be made non-specifically in the absence of enzymes,[2–7] but for information to be copied, such oligonucleotides must act as templates for copying chemistry. Based on our current understanding, nonenzymatic primer extension with native RNA cannot extend accurately beyond seven nucleotides on mixed sequence templates.[8] However, while the smallest ligase ribozyme has a catalytic core of 18 nucleotides, its full-length scaffold is closer to 40 nucleotides.[9] A better understanding of the process of nonenzymatic RNA primer extension is needed to explain how genomic sequences long enough to encode useful ribozymes could have been replicated prior to the evolution of RNA replicases.

Ribozyme polymerases that utilize nucleotide triphosphates (NTPs) as reactive building blocks have been evolved in laboratory evolution experiments[10] but NTPs are effectively unreactive in the context of nonenzymatic reactions due to their kinetic stability. Therefore, more highly activated nucleotide species would have been required for prebiotic reactions; Orgel first suggested the use of nucleoside-5′-phosphoroimidazolides (*pN, "*" = the imidazole moiety, "N" is the nucleoside) as reactive monomer species in the late 1960s, ultimately determining 5′-phosphoro-2-methylimidazolides were an optimal active species.[11-13] Recent studies have identified 2-aminoimidazole (2AI) as a prebiotically-plausible activation group for ribonucleotides that improves the rate and yield of nonenzymatic RNA primer extension.[8,14] Mechanistic studies of primer extension have demonstrated that 2AI-activated monomers react to

form a 5′-5′ imidazolium-bridged dinucleotide (Np*pN, **Scheme 1**) which subsequently binds to

the template via two Watson-Crick base-pairs, and reacts with the primer, which

then becomes extended by one nucleotide.[15,16] The imidazolium-bridged dinucleotide can

form both in solution and on the template, and its formation has been shown to be critical for

efficient primer extension by 2AI-activated monomers.[15–17] A recent crystallographic study

shows that the imidazolium-bridged dinucleotide is preorganized for primer extension.[18] As in

all previous crystallographic studies for nonenzymatic primer extension, magnesium has not

been observed interacting with either the 3′-OH or the adjacent phosphate of the imidazolium

bridged dinucleotide. However, the failure to observe the catalytic metal ion by crystallography

may simply reflect its weak or transient binding to the reaction center, and does not rule out any

specific role for the ion in phosphoryl transfer.[18,19]

Divalent cations are known to be important co-catalysts for RNA and enzyme-catalyzed

phosphoryl transfer reactions.[5,6] Metal ions have been screened for their catalytic effects in

nonenzymatic RNA polymerization,[3,19–22] RNA-activated monomer and oligomer hydrolysis, and

RNA ligation.[23] Optimal nonenzymatic RNA primer extension with 5′-phosphoro-2-

methylimidazolides requires a high concentration of magnesium ions.[24,25] However, these

previous studies assumed that the primer reacted directly with an activated monomer, whereas

we now know that the primer reacts with a template-bound imidazolium-bridged dinucleotide.

This new understanding of the mechanism of primer extension requires a re-examination of the

role of divalent cations in catalysis. In principle, metal ions could affect the equilibrium between

activated monomers and the bridged dinucleotide, either by modulating rates of formation of the

bridged dinucleotide or its hydrolysis, or even through differential binding effects of the

activated monomer or bridged dinucleotide to the template (**Scheme 1**).[15,26] Here we seek to

eliminate these confounding factors, and to focus our mechanistic studies directly on the reaction of the primer with the adjacent bridged dinucleotide. We have recently shown that Mg^{2+} ions promote primer extension (i.e. attack of the primer 3′-OH on the adjacent phosphate of the bridged dinucleotide) over hydrolysis of the bridged dinucleotide (i.e. attack of water or hydroxide on the bridged dinucleotide).[5,17,27] However, the exact mechanism by which Mg^{2+} specifically catalyzes primer extension remains unknown. While previous works and chemical logic have long suggested a role for the catalytic metal ion in facilitating the deprotonation of the primer 3′-OH,[28-30] our study specifically focuses on the role of cations on the chemical step of primer extension, i.e. the attack of the primer 3′-OH on the phosphate of the imidazolium-bridged dinucleotide.
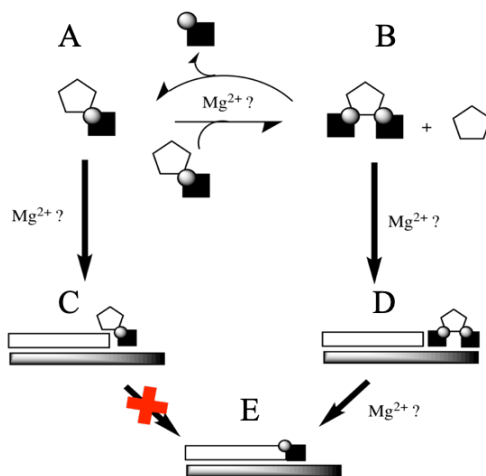
Scheme 1. Possible roles of Mg^{2+} in the activated monomer-bridged dinucleotide equilibrium and on nonenzymatic RNA primer extension. The equilibrium between states A and B is that of the formation and decay of the bridged dinucleotide from the activated monomer. C and D are the processes of the activated monomer or bridged dinucleotide binding the primer/template complex, and E is the state of the extended primer, which only the bridged dinucleotide can access on the timescales of study.[31]
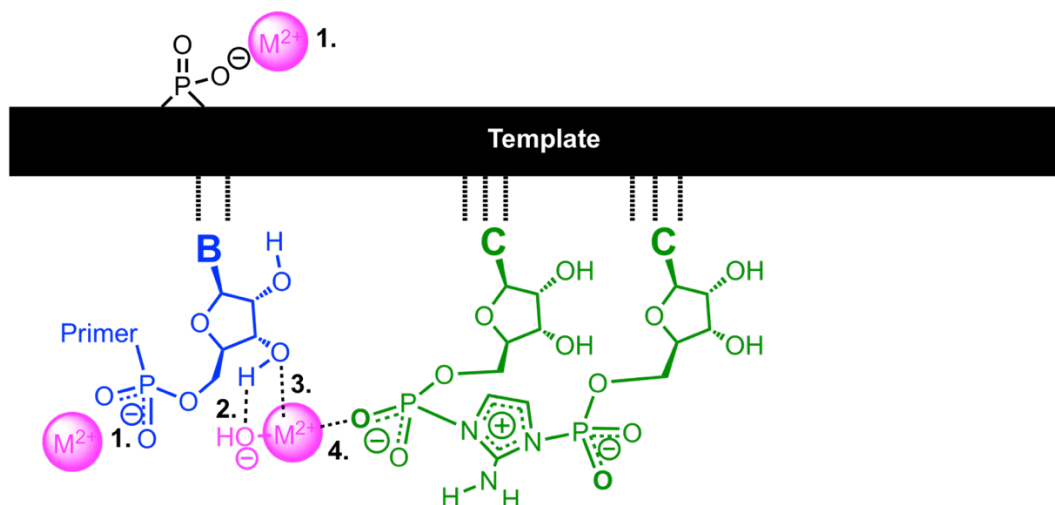
Scheme 2. Potential modes of metal ion catalysis in nonenzymatic primer extension (1) Electrostatic stabilization of primer/template duplex (2) Bound-hydroxide deprotonation of 3′-OH or stabilization of reaction geometry by hydrogen bonding (3) Inner-sphere Lewis acid interaction of a metal ion with O3′ with facilitation of deprotonation/ nucleophilic activation of the 3′-OH (4) Dual-coordination of both O3′ and a non-bridging phosphate oxygen optimizing the distance and angle of attack, or electrophilic activation of the phosphorus. Other catalytic possibilities not visually depicted could involve outer-sphere catalysis through hydrogen bonds, or local electrostatic stabilization of negative charge build-up in the transition state.

There are a number of possible mechanisms by which Mg^{2+} and other cations could potentially catalyze primer extension (**Scheme 2**). Determining whether inner-sphere or outer-sphere interactions dominate the role of metal ions as catalysts for primer extension is an important mechanistic distinction. For example, the waters in the innermost coordination sphere must exchange to gain an inner-sphere contact with a ligand such as the primer 3′-OH,[32] while outer-sphere contacts would not require these dynamics. Inner-sphere interactions of an ion not only restrict its motion (**Scheme 2**, interaction 3 and 4), but also can affect the nucleophilicity of the atoms in contact with the metal ion (**Scheme 2**, interaction 3). Divalent cations are Lewis acids, which results in the polarization of water and other acids in their inner-sphere, making

protons easier to remove from the associated compounds.[32] Additionally, more Lewis acidic metals have more bound-hydroxide present in solution, which could act as a base to deprotonate the 3′-OH, or stabilize binding of the metal ion through hydrogen bonding to directly facilitate 3′-OH deprotonation (**Scheme 2**, interactions 2 and 3). However, distinguishing between each of these interactions is difficult. We expect Lewis acidity and some form of acid/base catalysis to be important in primer extension, since the deprotonation of the 3′-OH is necessary during phosphoryl transfer. This could be conferred through an inner-sphere contact. Inner-sphere coordination of Mg^{2+} with the primer 3′-OH occurs in polymerases,[33] and we aim to address if it also occurs in nonenzymatic primer extension.

Outer-sphere contacts are mediated by metal-coordinated water molecules and rely on hydrogen bonding and electrostatic interactions. Cations that interact with a ligand solely by outer-sphere contacts can diffuse readily. Catalysis of primer extension by ionic stabilization of the duplex conformation via outer sphere contacts (**Scheme 2**, interaction 1) would not be expected to exhibit strong metal ion specificity, and any ion with similar or increased charge density should have similar effects.[34–36] Whether the interactions between RNA and Mg^{2+} (or similar metal ions) are responsible for the cation-dependent rate acceleration is unknown, but the simplest and least unique way that Mg^{2+} could facilitate nonenzymatic primer extension would be via stabilizing duplex formation or by stabilizing the binding of the bridged dinucleotide to the primer/template complex.[32,37] Cobalt hexammine has long been used to probe if an interaction is outer-sphere, due to its exchange-inert nature.[38-42] We utilize cobalt hexammine in our studies to probe if the benefit of a metal ion catalyst is indeed through direct chelation or through a second-shell solvent interaction.

Here we use kinetic studies in different sequence and conformational contexts to explore the role of the catalytic metal ion in nonenzymatic primer extension. We additionally probe a number of different cations and pH conditions to ascertain precisely the mode of catalysis. The rate of primer extension is enhanced by two orders of magnitude at saturating concentrations of Mg^{2+}. We focused primarily on Mg^{2+} in these studies due to its prevalence in use for primer extension reactions, and its prebiotic plausibility is invariant to the geochemical conditions studied as it is soluble across a wide range of pH. However, the high concentrations for divalent metal ions (tens to hundreds of millimolar) required suggests that the catalytic interactions are very weak and transient. The exchange-inert cobalt hexammine only enhances the rate of primer extension by two-fold, suggesting that the majority of the rate acceleration for primer extension is through inner-sphere contacts. We observed an increase in cation binding affinity for the reaction center with increased pH, which supports the role of Mg^{2+} in lowering the pK_a of the 3′-OH. Furthermore, compaction of the geometry of the primer extension reaction center by a downstream "helper" oligonucleotide leads to increased affinity for the metal ion, supporting the hypothesis of an interaction between the bridged dinucleotide and the metal ion.[19] Our results are consistent with the hypothesis that the catalytic metal ion interacts with both the primer 3′-hydroxyl and the adjacent reactive phosphate, and aids primer extension through inner-sphere contact mediated deprotonation of the primer 3′-hydroxyl.

Results and discussion

<u>Effect of cations on the affinity of the reactive bridged dinucleotide Cp*pC for the primer/template complex</u>

To distinguish between specific metal ion catalysis of primer extension and indirect effects on the affinity of the reactive bridged dinucleotide Cp*pC for the primer/template

complex (**Scheme 1**), we measured the rate of primer extension at different Cp*pC

concentrations under a range of metal ion conditions.

In absence of any added divalent metal cations, primer extension rates saturate at $0.29 \pm$

0.02 hr^{-1}, with a K_M for Cp*pC of 7.4 ± 1.3 mM (**Figure 1.0A**). Upon addition of 5 mM Mg^{2+},

the K_M for Cp*pC decreases to 5.0 ± 1.7 mM, and the V_{max} increases to 3.0 ± 0.28 hr^{-1}. Further

increasing the Mg^{2+} concentration to 50 mM decreases the Cp*pC K_M further to 2.1 ± 1.7 mM

(**Figure 1.0B**). This trend continues at 90 mM Mg^{2+}, with the published Cp*pC K_M as $1.06 \pm$

0.12 mM.[31] This increase in binding affinity across Mg^{2+} concentrations is likely due to

neutralizing charges in the primer/template complex, as oligomers have been desalted and the

Cp*pC is synthesized as the triethylamine salt, and there are no free metal ions in the buffer

solution (Tris HCl).

Given the beneficial effect of Mg^{2+} on the K_M of Cp*pC in the primer extension reaction,

we asked whether the increase in affinity was due to the increased ionic strength of the solution

or due to the inner-sphere binding capability of Mg^{2+}. We therefore measured the effect of two

additional cations, neither of which are able to be involved in inner-sphere interactions, on the

K_M of Cp*pC (**Figure 1.0C**). In the presence of 1M NaCl, the K_M for Cp*pC decreased 5-fold

from 7.4 ± 1.3 mM (no added Na$^+$) to 1.3 ± 0.6 mM (1M added Na$^+$). Similarly, when 50 mM

cobalt hexammine was added to the primer extension reaction mixture, the K_M for Cp*pC

dropped to 1.8 ± 0.6 mM. As expected, the affinity of Cp*pC for the primer/template complex is

improved by increasing the ionic strength of the solution, independently of the ability of cation to

engage in inner-sphere interactions with either the primer or the Cp*pC bridged dinucleotide.
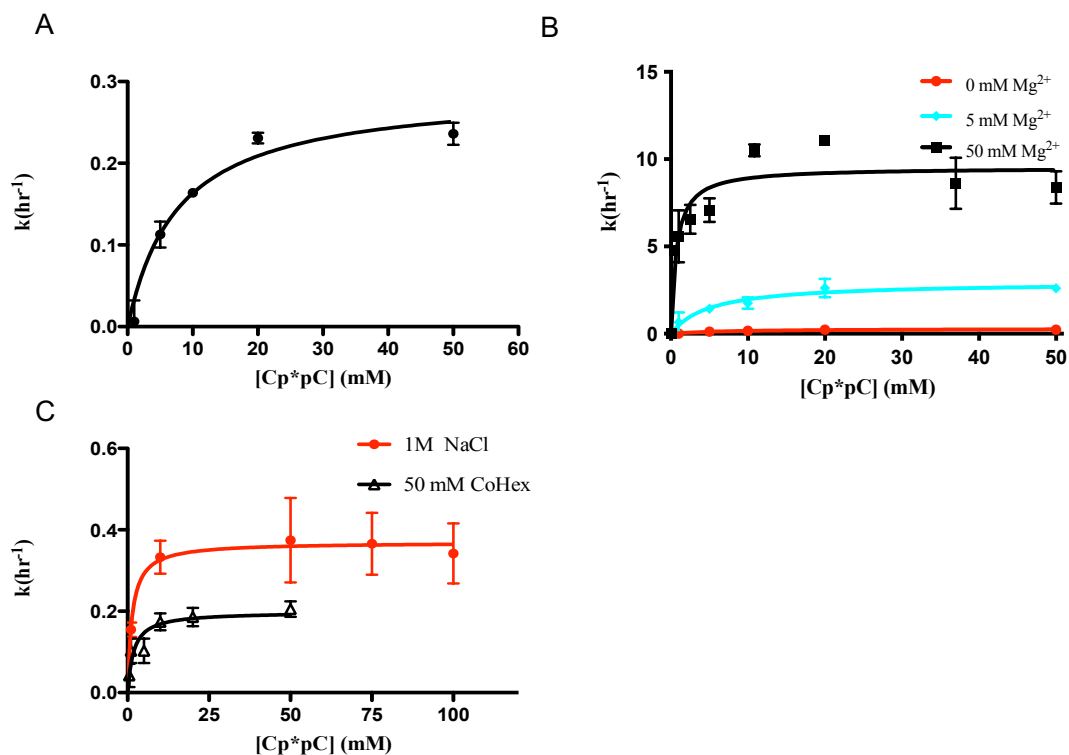
Figure 1.0. Cp*pC binding to primer/template complex is improved in presence of cations (A) Plot of primer extension rate at different concentrations of Cp*pC in absence of added metal ions. Line represents the saturation binding curve fit to empirical determinations of k and effective K_M. Error bars indicate ± SD (n≥3, independent experiments). Cp*pC binding in presence of 1μM primer, 1.5 μM template, 100 mM Tris HCl pH=8. V_{max}= 0.29 ± 0.02 hr^{-1}, with a K_M for Cp*pC of 7.4 ± 1.3 mM. (B) Cp*pC primer extension saturation curve, in the presence of 0, 5, and 50 mM Mg^{2+}. Line represents the saturation binding curve fit to empirical determinations of k and effective K_M. Error bars indicate ± SD (n≥3, independent experiments). Data for 0 mM Mg^{2+} is the same as in part A. At 5 mM Mg^{2+}: Cp*pC K_M= 5.0 ± 1.7 mM, and V_{max}= 3.0 ± 0.28 hr^{-1}. At 50 mM Mg^{2+}: Cp*pC K_M= 2.1 ± 1.7 mM and V_{max}= 13.55 ± 1.82 (C) Cp*pC primer extension saturation curve in presence of 1M NaCl or 50 mM cobalt hexammine. With 1M NaCl: K_M= 1.3 ± 0.6 mM, and V_{max}= 0.37 ± 0.02 hr^{-1}. With 50 mM cobalt hexammine K_M= 1.8 ± 0.6 mM, and V_{max}= 0.2 ± 0.01 hr^{-1}.

Nevertheless, in order to avoid confounding effects on occupancy of the primer/template complex by the reactive bridged dinucleotide, we therefore used saturating concentrations of Cp*pC (20 mM) in all subsequent primer extension experiments.

To ensure the benefits of cobalt hexammine were not due to contamination of free Co(III) ions due to degradation of the complex over time (the solution was observed to form a film and some precipitate on long time scales), we additionally performed Cp*pC binding curves with different concentrations of cobalt hexammine with and without EDTA, which would bind any ions present that are capable of inner-sphere interactions (**Figure 1.1, Table 1.1**). At low concentrations of cobalt hexammine, minute amounts of contamination improved the maximum rates and binding affinity of Cp*pC to the primer/template. This effect diminishes as the concentration of cobalt hexammine increases, with the rates and binding converging in the 10 mM cobalt hexammine samples with and without EDTA. This effect is expected to hold in **Figure 1.0C**, and contaminants of cobalt hexammine in experiments are diminished with fresh stock solutions and added 100 mM EDTA. Outside of the impurity effect, by 10 mM cobalt hexammine the binding of Cp*pC to the primer/template is comparable with the results in **Figure 1.0C**, showing that the binding affinity plateaus with differing ionic concentrations between Mg^{2+}, NaCl, and cobalt hexammine, but they all converge upon the same tightest binding value for Cp*pC and the primer/template of approximately 1 mM, likely corresponding to full electrostatic stabilization of the primer/template/bridged dinucleotide system. In order to ensure the effects with Cp*pC and metal ions was not unique to the G/C Watson Crick pair, we additionally tested the effects of cations on aiding the association of Ap*pA with the primer/template.
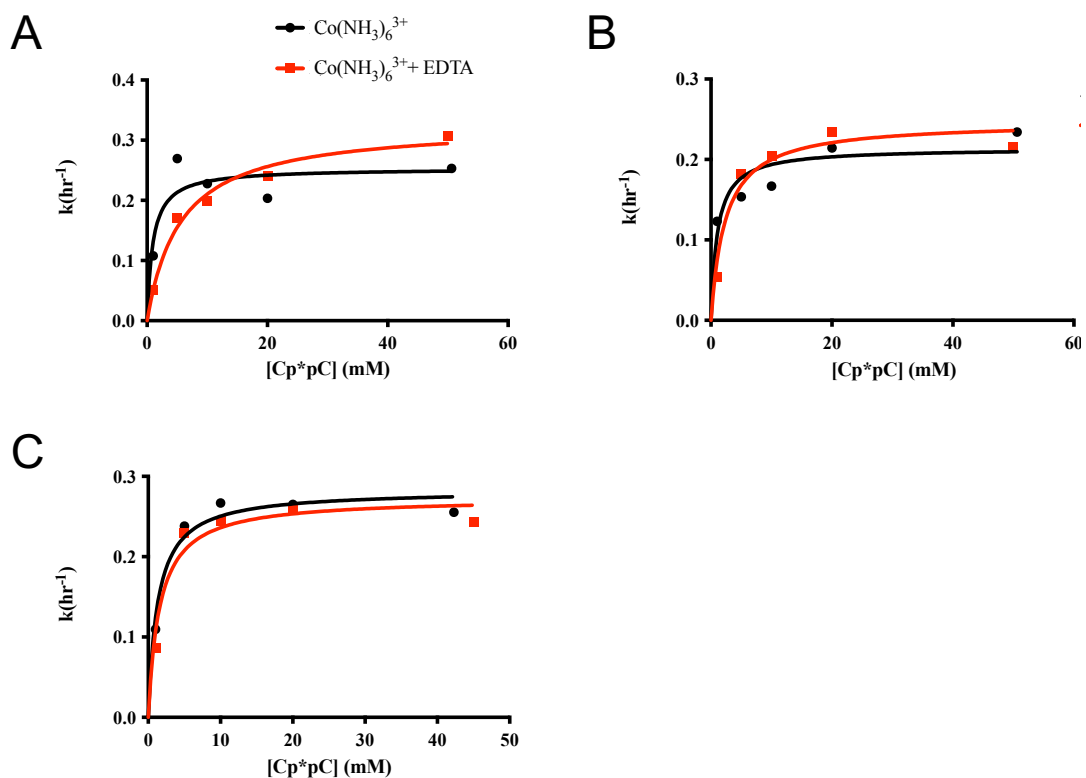
Figure 1.1. Cp*pC binding to primer/template complex with cobalt hexammine affected by impurities at low concentrations, but stabilizes in presence of EDTA. Lines represents fit to a saturation binding curve. Error bars indicate ± S.D. (A) Rates of primer extension vs. Cp*pC concentration in presence of 100 μM cobalt hexammine with and without 100 mM EDTA (B) Rates of primer extension vs. Cp*pC concentration in presence of 1 mM cobalt hexammine with and without 100 mM EDTA (C) Rates of primer extension vs. Cp*pC concentration in presence of 10 mM cobalt hexammine with and without 100 mM EDTA.

Conditions	V_{max} (hr^{-1})	K_M (mM)	Fit (R^2)
100 μM Co(NH$_3$)$_6{}^{3+}$	0.25 ± 0.03	0.94 ± 0.7	0.7
100 μM Co(NH$_3$)$_6{}^{3+}$ + EDTA	0.33 ± 0.02	5.5 ± 1.2	0.72
1 mM Co(NH$_3$)$_6{}^{3+}$	0.21 ± 0.02	1.05 ± 0.6	0.94
1 mM Co(NH$_3$)$_6{}^{3+}$ + EDTA	0.25 ± 0.02	2.4 ± 0.8	0.98
10 mM Co(NH$_3$)$_6{}^{3+}$	0.28 ± 0.01	1.3 ± 0.4	0.94
10 mM Co(NH$_3$)$_6{}^{3+}$ +EDTA	0.27 ± 0.02	1.6 ± 0.5	0.93

Table 1.1. Maximum rates and binding affinity of Cp*pC to primer/template complex from saturation binding curve fit of data in the presence of cobalt hexammine with and without 100 mM EDTA, corresponding to data in Figure 1.1.

Cation-dependent binding effects not observable in Ap*pA in the metal concentrations sampled

We additionally asked if the effect observed on Cp*pC binding affinity with metal ions was due to its identity, and performed similar experiments with Ap*pA (**Figure 1.2**). In addition to being a purine instead of a pyrimidine, which would have different stacking effects, we cannot rule out that the effects observed in **Figure 1.0** are independent of the G/C Watson-Crick base-pairing process. Ap*pA does not react appreciably in absence of Mg^{2+}, so we were unable to measure the K_M of Ap*pA for the primer/template in those conditions. We determined that the binding affinity for Ap*pA and the primer/template complex remains the same between 5 mM Mg^{2+} and 50 mM Mg^{2+}. This could be due to the fact that the benefit of Mg^{2+} has been saturated by 5 mM for Ap*pA, and because of the lack of reactivity in absence of Mg^{2+} for much of the

Ap*pA concentration range, we cannot ascertain the benefit in binding with and without Mg^{2+}. We propose the lack of reactivity with Ap*pA in absence of Mg^{2+} is due to its much weaker binding with the primer/template complex (which we expect compared to G/C) and not due to electrostatic differences since the charge satiation of the negatively charged primer/template and the net-negative Np*pN is the same regardless of nucleobase. Nevertheless, the saturation of binding between 5 and 50 mM Mg^{2+} with Ap*pA suggests that G/C base pairs are more sensitive to the presence of ions in the context of Np*pN binding. This is consistent with N7 interactions that have been observed *in crystallo* with G,[43] although not specifically with the bridged dinucleotide. Nevertheless, these binding effects are controlled for in our following studies by utilizing saturating concentrations of Ap*pA and Cp*pC that hold across the full concentration ranges of Mg^{2+}. With binding of the bridged dinucleotide accounted for, we proceeded to study the effects of metal ions on the bridged dinucleotide and activated monomer equilibrium.
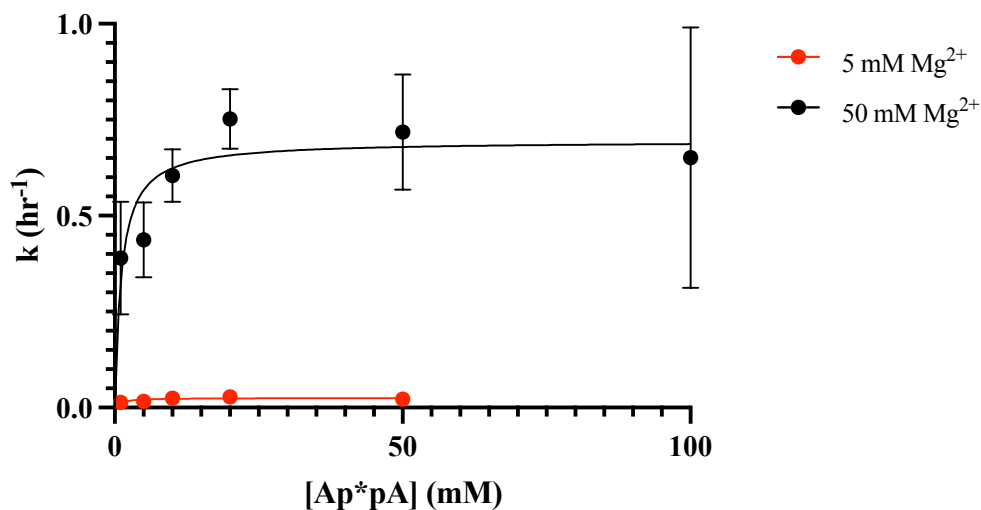
Figure 1.2. Ap*pA binding does not change with increasing Mg^{2+} in concentration range sampled. Plot of primer extension rate at different concentrations of Ap*pA at 5mM Mg^{2+} and 50 mM Mg^{2+}. Error bars indicate ± SD (n≥3, independent experiments). Line represents fit to a saturation binding curve equation of Ap*pA in presence of 1μM primer, 1.5 μM UUU template,

100 (**Figure 1.2 continued**) mM Tris HCl pH=8. For the 5mM Mg^{2+} experiment, V_{max}= 0.025 ±

0.002 hr^{-1}, with a K_M for Ap*pA of 1.05 ± 0.55 mM. For the 50 Mg^{2+} mM experiment, V_{max}=

0.70 ± 0.06 hr^{-1}, with a K_M for Ap*pA of 1.14 ± 0.65 mM.

<u>Effects of Mg^{2+} on the rates of formation and hydrolysis of Cp*pC</u>

We also studied whether Mg^{2+} could affect the free Cp*pC concentration by influencing

either the formation Cp*pC, or its decay into *pC (**Scheme 1**). We observed no significant effect

of Mg^{2+} on the rate of formation of Cp*pC from *pC as monitored by ^{31}P NMR (**Figure 1.3,**

Table 1.2). The reaction of one net negative activated monomer and one net neutral activated

monomer to the net negative bridged dinucleotide is not expected to benefit by metal ion

stabilization, so this is not surprising. At much higher metal ion concentrations, it is possible that

the rate of hydrolysis of *pC increases in competition with forming Cp*pC, decreasing the

overall formation of Cp*pC, but this would not provide any benefit to nonenzymatic primer

extension. In fact, if this was the case, we would anticipate a decrease in rate of primer extension

at high metal ion concentrations, but the opposite is observed. Therefore, the effect on bridged

dinucleotide formation is negligible to the observed rate enhancement of Mg^{2+} to primer

extension.
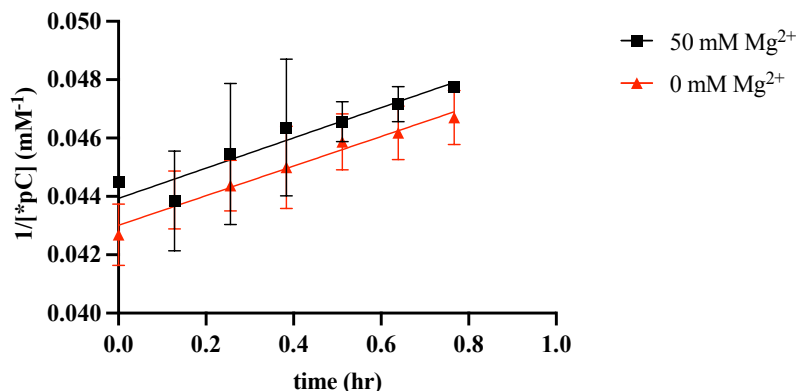
Figure 1.3. Bridged dinucleotide formation does not increase in presence of Mg^{2+}. Analysis of

*pC signal decay by ^{31}P NMR to form Cp*pC at pH=8, 25 mM initial *pC, with and without

(**Figure 1.3 continued**) $MgCl_2$. The reaction is done in 100% H_2O with a D_2O insert. The second order rate plot is shown, and fit to a linear regression as expected by a second order reaction for the formation of Cp*pC from *pC. Refer to Table 1.2 for values of k. Error bars indicate ± SD (n≥3, independent experiments).

	0 mM Mg^{2+}	*50 mM Mg^{2+}*
k ($hr^{-1}mM^{-1}$)	0.0026 ± 0.00073	0.0025 ± 0.00077

Table 1.2. Cp*pC formation under different Mg^{2+} conditions (from 1/[*pC] vs. time, slope=2k)

In contrast to bridged dinucleotide formation, increasing concentrations of Mg^{2+} linearly increases the rate of hydrolysis of Cp*pC (**Figure 1.4A**). There is slight deviation from linearity due to pyrophosphate formation in addition to hydrolysis. The observed effects of Mg^{2+} on hydrolysis are expected and have been studied with phosphoroimidazolides extensively, but not in the context of the bridged dinucleotide that is relevant for the observed rates of nonenzymatic primer extension.[44] The hydrolysis of Cp*pC is slower than primer extension across the range of Mg^{2+} sampled, but the half-lives of extension and hydrolysis begin to converge at high concentrations of Mg^{2+} (**Figure 1.4B**). To avoid the issue of Cp*pC hydrolysis in the extension of multiple nucleotides, we adjusted the time points taken for kinetics across Mg^{2+} concentrations to better suit the first-order linear regime for rates, and avoid significant hydrolysis of Cp*pC (**Table A.2, Appendix A**). Nevertheless, the role of metal ions in hydrolyzing the reactants for primer extension would not explain the observed *benefits* of Mg^{2+} in primer extension.
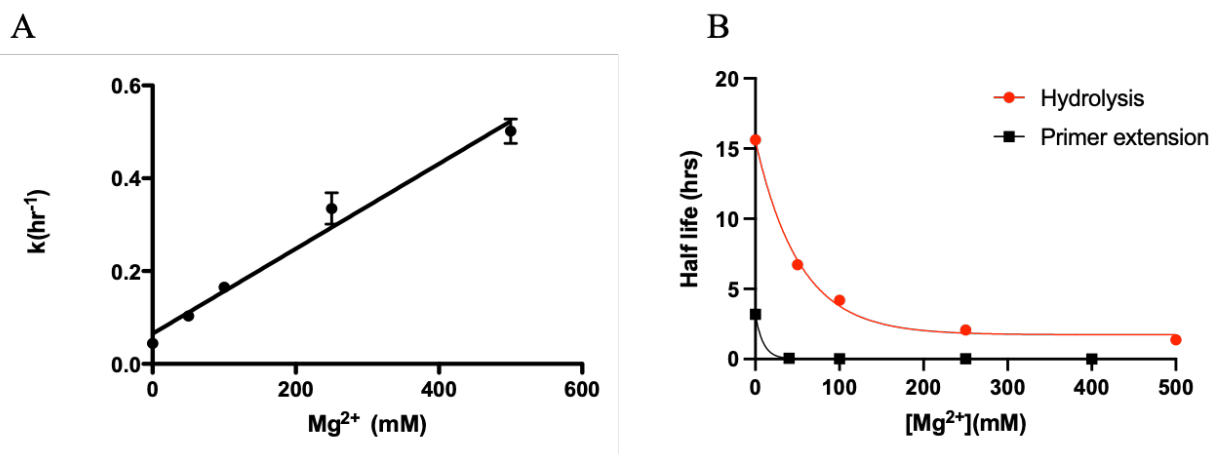
Figure 1.4. Dimer hydrolysis increases linearly with increasing Mg^{2+} concentration and approaches the half-life of primer extension at high Mg^{2+} concentrations (A) k_{obs} of Cp*pC hydrolysis fit to a line (R^2=0.98). k_{obs} determined monitoring diminishment of peak at -12.94 ppm in ^{31}P NMR (Cp*pC), with an internal standard of $OP(OMe)_3$ at pH=8, 10% D_2O and various Mg^{2+} conditions. Bars shown are SD ± (n=3 independent experiments). Quantification of hydrolysis by first-order kinetic plot of intensity of peak integration normalized to total peak integration intensity. NMR peak assignments at 0 mM Mg^{2+}: +3.03 ppm, +0.49 ppm (CMP), -2.97 ppm (phosphodiester linkages), -11.37 ppm (*pC), -12.94 ppm (Cp*pC). (B) Half-lives of hydrolysis and primer extension as a function of Mg^{2+} concentration and fit to a first order nonlinear decay curve using GraphPad Prism.

<u>Saturable binding of Mg^{2+} to the reaction center observed at high concentrations</u>

To assess the affinity of Mg^{2+} for the reaction center, we measured the rate of primer extension at saturating Cp*pC on a GGG template, as a function of Mg^{2+} concentration (**Figure 2A**). We find that Mg^{2+} binds weakly, but saturably, with a K_M of 181 ± 13 mM (**Figure 2B**). The V_{max}= 74 ± 2 hr^{-1}, which is ~300 times faster than in absence of added divalent cations (k= 0.21 ± 0.004 hr^{-1}). We also observe the binding of Mg^{2+} in the Ap*pA primer extension system is comparable, if not slightly better (**Figure 2.1A**, K_M for Mg^{2+} is 117 ± 17 mM with saturating

Ap*pA), although the maximum rate of extension is much lower when extending with Ap*pA than with Cp*pC (V_{max}=1.02 ± 0.06 hr^{-1}). Surprisingly, changing the identity of the last nucleotide of the primer from G to A has a large effect on the K_M of Mg^{2+} for primer extension, which increases 4-fold from 181 ± 13 mM to 725 ± 122 mM (**Figure 2.1B**). This is likely due to the difference in stability between G/C and A/U base pairing, and increased fraying of the end of the primer with the weaker A/U base pair, especially at lower concentrations of added metal ions. All subsequent experiments were performed using the G-terminated primer to improve the affinity of the primer/template complex for Mg^{2+}.

We additionally studied Mg^{2+} binding in the Gp*pG extension system (**Figure 2.1C**), as we were interested in if the N7 interaction with G could potentially recruit metal ions to the reaction center, or acts as an additional binding site stabilizing the interaction of Gp*pG with the primer/template. Due to the inability to fit the binding curve to a higher hill slope, it is unlikely that if there is an additional metal ion binding at the N7 that it is catalytically relevant. Oddly, while Gp*pG confers a similar maximum rate in primer extension to Cp*pC, Mg^{2+} binds in the Gp*pG system much less tightly. We suspect this is due in part to the vast amount of precipitation observed in the reactions above 100 mM Mg^{2+}. Gp*pG is much more limited in solubility than other Np*pN, which is not surprising due to the ability of G to form higher order complexes, especially with metal ions.[45] We suspect this increase in K_M for Mg^{2+} is due to Mg^{2+}-Gp*pG complexes being removed from solution, and due to the vast excess of Gp*pG to primer/template we still have a saturating amount interacting with the primer/template (20 mM compared to 1μM) but not an accurate representation of Mg^{2+} concentration. Therefore, the saturation observed is likely at a lower Mg^{2+} concentration, due to the *in-situ* precipitation of Mg^{2+} decreasing the effective concentration. When using MgSO$_4$, the issue of solubility with

Gp*pG is mitigated, but the rate data is within error of the MgCl$_2$ at low concentrations of Mg^{2+}.

Due to the lower absolute solubility of MgSO$_4$ in water compared to MgCl$_2$, a sufficiently

concentrated stock solution to enable a full binding curve at high Mg^{2+} is not accessible and the

binding curve cannot be performed. We cannot fully ascertain the binding of Mg^{2+} accurately in

this system due to these experimental constraints. Precipitation between Gp*pG and Mg^{2+} may

be an important consideration for prebiotic chemistry, although it is unlikely that the

concentrations in which this effect occurs would be reached on the early Earth. Nevertheless,

regardless of Np*pN, Mg^{2+} vastly increases primer extension rates and saturably binds to the
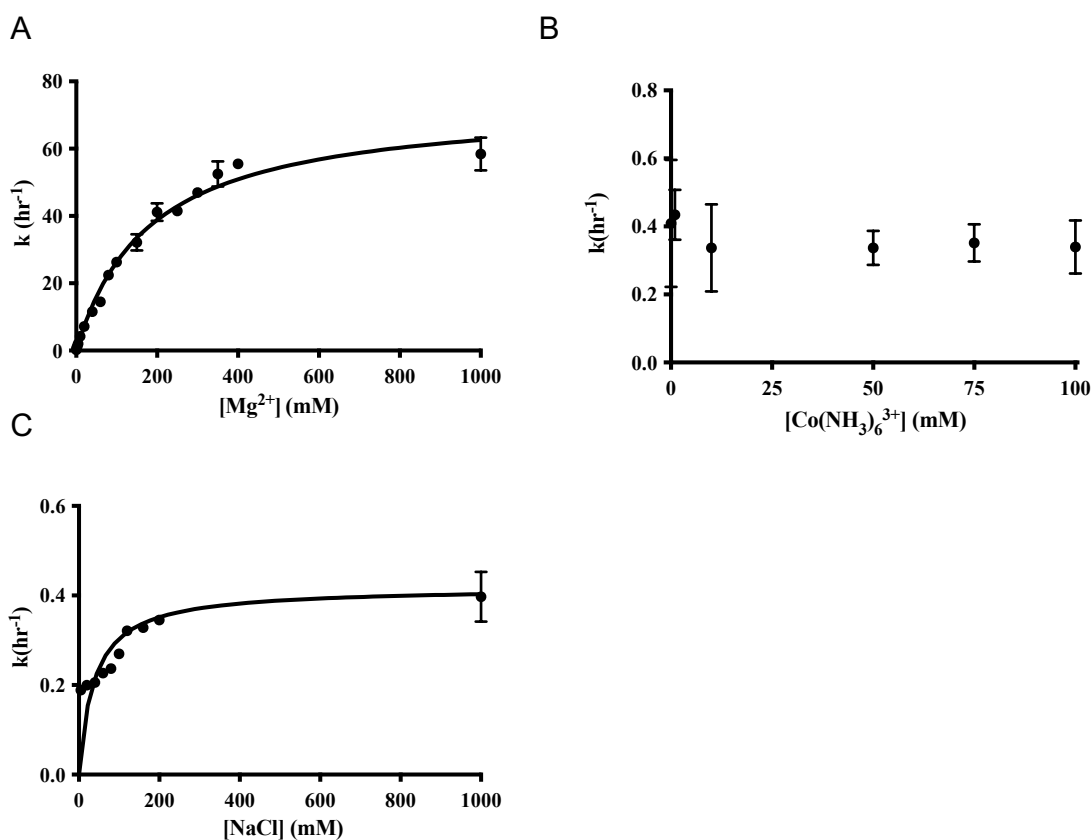
primer/template complex.

Figure 2. Metal ion dependence for the rate of primer extension at saturating Cp*pC

concentration (A) Rate of primer extension as a function of Mg^{2+} concentration, at constant 20 mM Cp*pC and pH 8. Line represents fit to a saturation binding curve for Mg^{2+} binding to primer/template complex. The V_{max}= 74 ± 2 hr^{-1} and K_M = 181 ± 13 mM. Error bars indicate ± (**Figure 2 continued**) SD (n≥3, independent experiments). (B) Rate of primer extension vs. concentration of cobalt hexammine at pH 8, without any superimposed linear or nonlinear equations. (C) Rate of primer extension as a function of NaCl concentration at pH 8. Line represents fit to a saturation binding curve for Na^+ binding to primer/template complex. Error bars indicate ± SD (n≥3, independent experiments). V_{max}= 0.42 ± 0.02 hr^{-1}, K_M= 38 ± 10 mM.
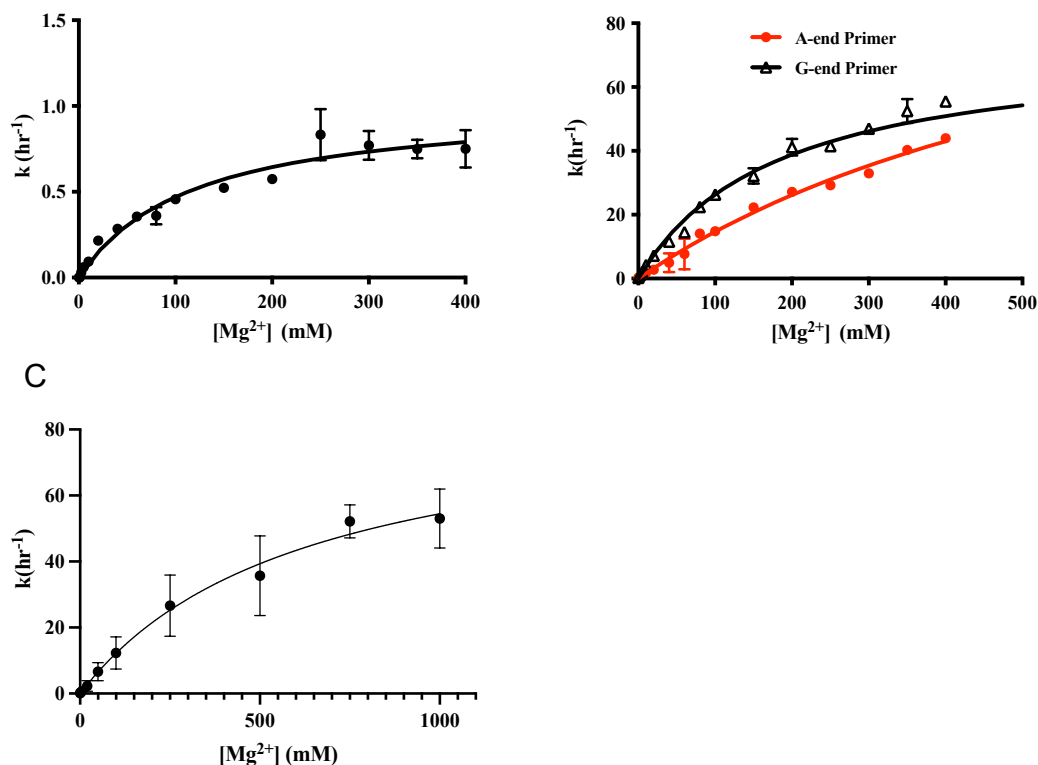
Figure 2.1. Primer extension in other sequence contexts is also sensitive to Mg^{2+} (A) Primer extension on UUU template, using 27.5 mM Ap*pA dimer. Line represents fit to a saturation binding curve for Mg^{2+} binding to primer/template complex. Error bars indicate ± SD (n≥3, independent experiments). V_{max} = 1.02 ± 0.06 hr^{-1} and K_M= 117 ± 17 mM. (B) Primer and template as shown in Table I in Materials and Methods, the only sequence difference being the

last nucleobase of the primer, and the corresponding nucleotide on the complement template. Line represents fit to a saturation binding curve for Mg^{2+} binding to primer/template complex. Error bars indicate ± SD (n≥3, independent experiments). G-end primer is a reproduction of the **(Figure 2.1 continued)** curve from Figure 2A, A-end primer has $V_{max}= 121 \pm 15$ hr^{-1} and $K_M= 725 \pm 122$ mM. (C) Primer extension on CCC template using 20 mM Gp*pG. Line represents fit to saturation binding curve for Mg^{2+} binding to primer/template complex. Error bars indicate ± SD (n=6, independent experiments). $V_{max} = 86.4 \pm 9.6$ hr^{-1} and $K_M= 600.4 \pm 137$ mM. Large amounts of white precipitate were observed above 100 mM Mg^{2+}.

With it established that Mg^{2+} confers a large increase in the rate of primer extension and saturably binds the primer/template in a variety of sequence contexts, we now aim to disentangle the effects of inner-sphere coordination from outer-sphere or ionic strength effects through studying additional metal ions and complexes.

Ionic strength and outer-sphere effects do not account for divalent metal ion catalysis of primer extension

As noted in Chapter 1, cobalt hexammine is incapable of making inner-sphere contacts because its six ammonia ligands are not exchangeable on the timescale of our experiments. This metal-ligand complex is therefore commonly used to distinguish between inner- and outer-sphere interactions in catalysis. Here we used cobalt hexammine in an effort to distinguish between nucleophilic activation of the 3′-OH or electrophilic activation of the reactive phosphate by direct inner-sphere contacts, and indirect interactions through secondary or outer-sphere coordination, as well as general effects on ionic strength. NaCl functions similarly for our determinations, as Na$^+$ ions are strongly hydrated, and the inner shell of bound waters is less likely less to exchange and allow for direct inner-sphere interactions with either the 3′-OH or the

reactive phosphate. In addition, Na^+ is much less Lewis acidic.[46] We therefore measured the rate of primer extension in the presence of saturating Cp*pC, as a function of NaCl or cobalt hexammine. For both NaCl (**Figure 2.2A**) and cobalt hexammine (**Figure 2.2B**), we find a maximum increase in reaction rate of only ~2 fold compared to the background rate, as opposed to the 300-fold increase in rate conferred by saturating concentrations of Mg^{2+}. The simplest interpretation of these observations is that Mg^{2+}-catalysis is mediated by inner-sphere contacts with the primer and/or reacting bridged dinucleotide; however, we will discuss alternative possibilities as well. The small rate increase observed with outer-sphere cations is likely due to electrostatic stabilization of the primer/template complex.

While ionic strength of the solution is unlikely to be the sole determinant of primer extension rate in light of our cobalt hexammine or NaCl results, we performed an additional experiment to explicitly address this concern (**Figure 2.2A**). The ionic strength of the reaction mixture was controlled by adding NaCl while varying the Mg^{2+}. While the maximum reaction rate was not statistically significantly affected, the K_M for Mg^{2+} increased from 181 ± 13 mM to 386 ± 57 mM (**Figure 2.2A**). This increase in K_M is likely due to competition for the reaction center between the catalytic Mg^{2+} and the fully hydrated sodium, which would be incapable of aiding in deprotonation or stabilization of deprotonated 3′-OH. We verified that a vast excess of sodium (1M) can interfere with primer extension conferred by Mg^{2+} (**Figure 2.2C**)— both the V_{max} and binding affinity for Mg^{2+} decreased. The increase in observed K_M for Mg^{2+} in the presence of 1M NaCl agrees with our hypothesis that an inner-sphere Lewis acid is necessary for primer extension to occur, and Na^+ can compete for that site. The decrease in V_{max} may be due to a conformational effect on the duplex, with the 1M NaCl leading to a conformation of the RNA

less conducive for primer extension[47]. Smaller amounts of NaCl added to primer extension did not have a substantial effect (100 mM, *not shown*).

Since cobalt hexammine was observed to stabilize binding of the bridged dinucleotide to the template as well as Mg^{2+} at fairly low concentrations, we decided we could circumvent possible structural effects in studying ion competition by using small amounts of cobalt hexammine in primer extension in addition to Mg^{2+}. Cobalt hexammine is more charge dense than Mg^{2+}, and is expected to interact more strongly through electrostatics. Primer extension the presence of 100 μM cobalt hexammine increased the K_M for Mg^{2+} to 368 ± 57 mM (**Figure 2.2B**), without significantly changing V_{max}, which directly suggests competitive inhibition. Using this data, we are able to calculate the extent of inhibition of primer extension by the presence of cobalt hexammine using Lineweaver-Burk analysis (**Figure 2.3, Table 3**). The calculated K_i of 96 μM is close to the concentration we chose to sample, and cobalt hexammine proves to be a potent inhibitor of Mg^{2+}-catalyzed primer extension. These observations support that both Na^+ and cobalt hexammine can occupy the same site as the catalytic Mg^{2+} ion, but that neither Na^+ nor cobalt hexamine have a significant catalytic effect when present in the reaction center. The inability of these ions to support catalysis of primer extension strongly supports the hypothesis of primer extension being catalyzed through inner-sphere interactions with a cation. Therefore, the role of these inner-sphere interactions was further probed to determine if the catalysis is through deprotonation of the 3′-OH through a bound hydroxide or solvent, stabilization of the negatively charged and deprotonated 3′-O, or another mechanism.
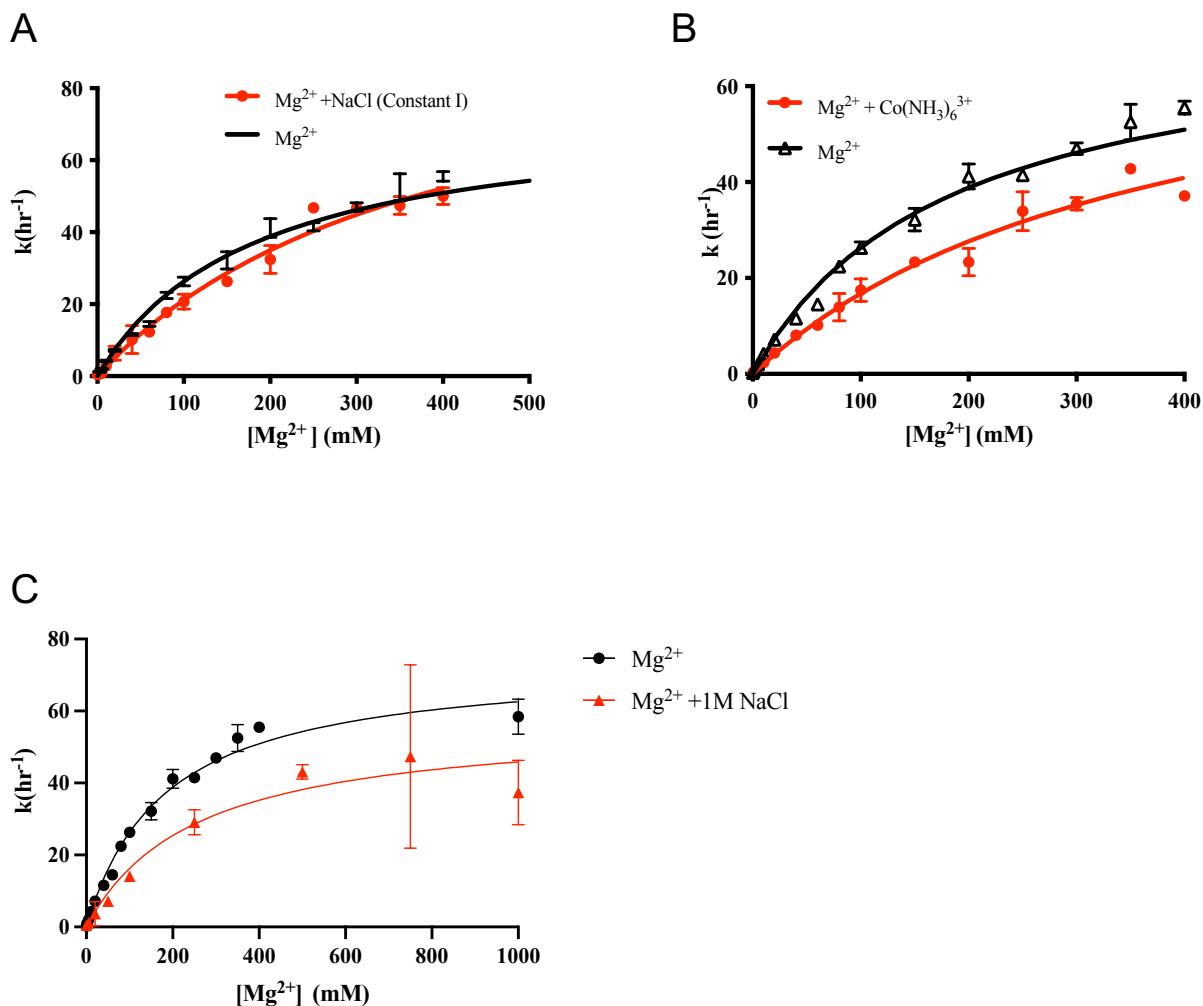
Figure 2.2. Outer-sphere ions compete with Mg^{2+} in primer extension (A) Primer extension conditions are the same as in Figure 2A, except in the constant ionic strength curve, I=1200 is maintained by a combination of $MgCl_2$ and NaCl. Lines represent fit to a saturation binding curve equation for Mg^{2+} binding to primer/template complex. Error bars indicate ± SD (n≥3, independent experiments). The Mg^{2+} curve from Figure 2.0A is reproduced here. Constant ionic strength: V_{max}= 102.5 ± 9 hr^{-1}, K_M= 368 ± 57 mM (B) Mg^{2+} curve reproduced from Figure 2A. Mg^{2+} + cobalt hexammine experiments are the same except with addition of 100 μM $Co(NH_3)_6Cl_3$ in all samples. Line represents fit to saturation binding curve for Mg^{2+} binding to primer/template complex. Error bars indicate ± SD (n≥3, independent experiments). In (**Figure**

2.2 cont.) presence of cobalt hexammine, $V_{max}= 78.5 \pm 7$ hr^{-1}, $K_M= 368 \pm 57$ mM. (C) The Mg^{2+} curve from Figure 2.0A is reproduced here for comparison. Mg^{2+} + 1M NaCl experiments are the same except with addition of 1M NaCl in all samples. Line represents fit to saturation binding curve for Mg^{2+} binding to primer/template complex. Error bars indicate $\pm$ SD (n$\geq$3, independent experiments). $V_{max}= 57.5 \pm 6.6$ hr^{-1}, $K_M= 252 \pm 93$ mM.
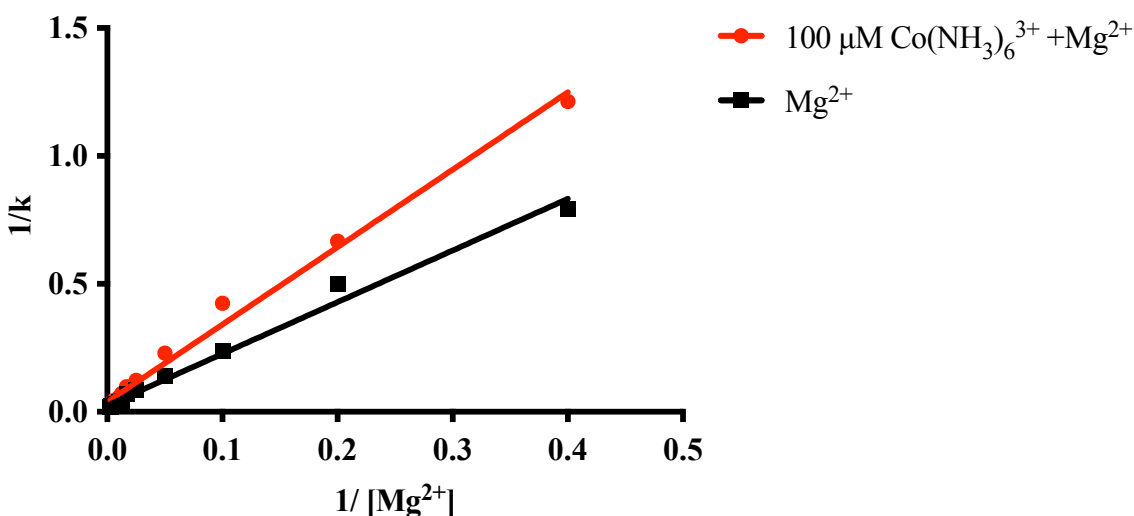
Figure 2.3. Lineweaver-Burk plot supports competitive inhibition of Mg^{2+}-catalyzed primer extension by cobalt hexammine. Data from Figure 2.2B is reanalyzed as a double reciprocal plot. Error was propagated by the following equation: $\delta(1/x) = \sqrt{(\delta x/x)^2}$ The K_i calculated assuming competitive inhibition is 96 μM ($K_i= K_M[I]/ (K_{M, apparent}-K_M)$, where K_M is the K_M of the uninhibited reaction).

Condition	Slope	Intercept
Mg^{2+}	2.02 $\pm$ 0.06	0.038 $\pm$ 0.01
$Mg^{2+}+ 100\ \mu M\ Co(NH_3)_6^{3+}$	3.03 $\pm$0.08	0.02 $\pm$ 0.01

Table 3. Lineweaver-Burk linear regression data used to calculate K_i (Fits for Figure 2.3).

The catalytic effects of Mg^{2+}, Mn^{2+} and Ca^{2+} correlate with their Lewis acidity

If an interaction between the catalytic metal ion and the 3′-OH of the primer is important for primer extension, we would expect the rate of primer extension to be sensitive to the Lewis acidity of the metal ion. The Lewis acidities of Mn^{2+}, Mg^{2+} and Ca^{2+} are reflected in the pK_a of their bound water, which are 10.6 , 11.42 and 12.7 respectively (**Table 4**).[46] Thus direct coordination of these ions with the primer 3′-OH could facilitate its deprotonation, leading to a catalytic effect that correlates with the metal ion Lewis acidity. Lewis acid trends would also hold if the 3′-OH is deprotonated before the rate-limiting step,[48] as it correlates with the ability of an ion to stabilize charge. Due to the decreasing pK_a of bound waters, a more Lewis acidic cation would also have more bound-hydroxide at a given pH. Similarly, inner-sphere coordination with one of the non-bridging oxygens of the reactive phosphate of the Cp*pC bridged dinucleotide should withdraw electron density from the phosphate oxygen leading to electrophilic activation of the phosphorus, again correlating with the metal ion Lewis acidity. Distinguishing between these subtle, but different causes for Lewis acid trends is difficult, but first requires that a Lewis acid trend is observed. Thus, we repeated our measurements of the rate of primer extension as a function of the concentration of Mn^{2+} and Ca^{2+} (**Figure 3**). The maximum rates measured for three metal species (**Table 4**) correlate with their Lewis acidity ($Mn^{2+} > Mg^{2+} > Ca^{2+}$), while their binding affinities show no apparent trend ($Mn^{2+} \gg Ca^{2+} > Mg^{2+}$). The Lewis acidity trend also holds in a comparison with the catalytic efficiency (V_{max} / K_M), with Mn^{2+} having by far the highest catalytic efficiency and Ca^{2+} the lowest (**Table 4**). These trends are also not exclusive to Cp*pC, and extend to Ap*pA (**Figure 3.1**). These observations support the hypothesis that the catalytic metal ion interacts directly with either the primer 3′-OH and/or one of the non-bridging oxygen atoms of the adjacent reactive phosphate.

In order to address if the reaction proceeds due to a bound-hydroxide interaction, we would expect the differing rates of V_{max} to correspond to the relative amount of bound-hydroxide found for each ion in the reaction conditions. This can be complicated by size and coordination differences inherent to these ions, but this method has been previously successful for other nonenzymatic RNA reactions.[49] At pH=7.5, there is 6-fold more $Mn(OH)^+$ than $Mg(OH)^+$, however we observe only a 2-fold difference in maximum rate. Considering binding affinities, Mn^{2+} is 24.5 times better at catalyzing primer extension than Mg^{2+} in these conditions. Additionally, we would expect Mg^{2+} to perform 19 times better than Ca^{2+} if primer extension catalysis is due to bound-hydroxide, yet we observe a ~4-fold difference in their maximum rates and 2-fold difference in catalytic efficiencies. The inability to correspond the difference in rates does not conclusively rule out the possibility of bound-hydroxide catalysis, but it also does not support this hypothesis, and instead suggests inner-sphere coordination with the 3′-OH is the cause.
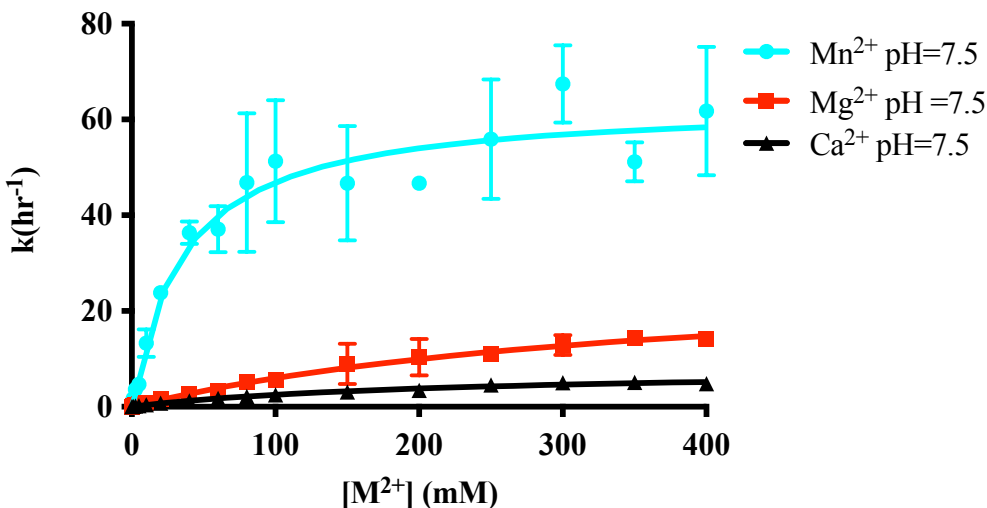
Figure 3. Maximum rate of primer extension with different metal ions correlate with Lewis acidity. Line represents fit to saturation binding curve for M^{2+} binding to primer/template complex. Error bars indicate ± SD (n≥3, independent experiments). Due to the insolubility of

(**Figure 3 continued**) $Mn(OH)_2$ species which are generated at higher pH, and increased

oxidation at high concentrations, binding curves for this assay were are all performed at pH=7.5

in an anoxic environment. See Table 4 for V_{max} and K_M values.

Metal ion	V_{max} (hr⁻¹)	K_M (mM)	Catalytic Efficiency (M⁻¹s⁻¹)	First pK$_a$ of [M(H₂O)$_x$][46]
Ca^{2+}	8 ± 0.5	217 ± 29	1.0×10^4	12.7
Mg^{2+}	31 ± 2	417 ± 54	2.0×10^4	11.42
Mn^{2+}	64 ± 2.8	36 ± 6.5	4.9×10^5	10.6

Table 4. Kinetic values of primer extension of Cp*pC on GGG template at pH 7.5 at saturating

concentrations of Ca^{2+}, Mg^{2+} and Mn^{2+} corresponding saturation binding curve fits from Figure

3.
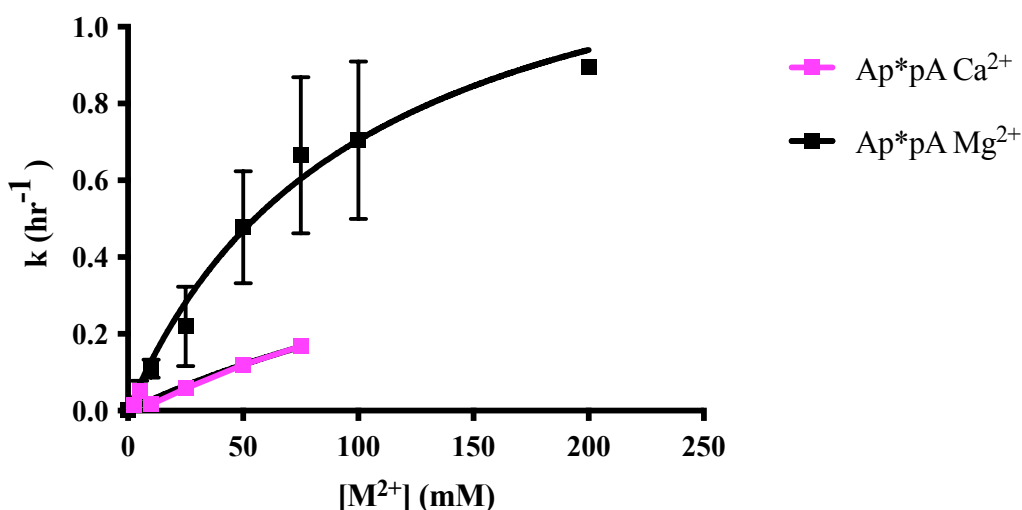
Figure 3.1. Maximum rate of primer extension trends with Lewis acidity with Ap*pA system.

Rates of primer extension with Ap*pA and varying concentrations of Mg^{2+} and Ca^{2+}. Line

represents fit to saturation binding curve for Mg^{2+} binding to primer/template complex. Error

bars indicate $\pm$ SD (n$\geq$3, independent experiments). Primer extension was performed on a UUU

template with Ap*pA at pH=8 with Ca^{2+} and Mg^{2+}. Mn^{2+} was not studied because the hydrolysis

of the bridged dinucleotide with Mn^{2+} is on the same timescale as primer extension with Ap*pA.

The Mg^{2+} curve is reproduced from Figure 2.1A. With Ca^{2+}, V$_{max}$= 0.74 ± 1.1 hr^{-1}, K$_M$= 257 ± 499 mM.

<u>Primer extension rates with other highly Lewis acidic transition metals depends on the balance between extension and hydrolysis of Np*pN and duplex stability</u>

Increasing Lewis acidity does not increase primer extension rates universally. The simplest reason for this is that there are two competing reactions in our system of nonenzymatic primer extension: extension, which is desired, and hydrolysis of the reactants, which is not. Metal hydroxo-species are known to increase hydrolysis of RNA, so this is not surprising these effects would extend to the bridged dinucleotide.[51-55] We initially found Co^{2+} a promising cation candidate, as at 20 mM it outperforms Mg^{2+} (**Table 4.1**). Co^{2+} is ten times more acidic than Mn^{2+}, yet has a maximum observed primer extension rate almost 5x slower than Mg^{2+} at saturating amounts of metal ion (**Figure 3.2, Figure 2A**). At higher concentrations of Co^{2+}, clearly hydrolysis outperforms extension, leading to this lower maximum rate. This is not the sole contributor to why Co^{2+} performs poorly (see Chapter 4), and is also likely due to increased interaction with the nucleobases and destabilization of the primer/template complex at higher Co^{2+} concentrations.[56] The observed primer extension rate would likely begin to decrease at high enough Co^{2+} concentrations if hydrolysis of Cp*pC is the cause of low V$_{max}$, but due to solubility issues of Co^{2+}, we were unable to verify this. Ni^{2+} performs poorly as well (**Table 4.1**), and this is independent of Np*pN composition. We tested our hypothesis of increased Np*pN hydrolysis with Co^{2+}, and this proves to be true (**Figure 3.3**). Co^{2+} hydrolyzes Cp*pC 3.6 times faster than Mg^{2+} in the same conditions. This is likely not the only effect that causes the lower maximum rates of Co^{2+} and Ni^{2+} compared to Mg^{2+}— comparative hardness and softness of the ions leads to different association with the phosphates of the backbone versus the nucleobases, and can

cause different structural conformations for duplexes.[57,58] Co^{2+} and Ni^{2+} are very prone to interacting with nucleobases compared to Mg^{2+}, and at higher concentrations could increase duplex fraying and other conformational changes unfavorable for primer extension.

We additionally studied Eu^{3+}, another highly Lewis acidic ion, which we would expect to perhaps similarly hydrolyze Np*pN. However, Eu^{3+} was primarily more active than Mg^{2+} in the pH ranges sampled, although there was an odd pH dependence to this, with Mg^{2+} outperforming Eu^{3+} at pH=7, and Eu^{3+} performing better at pH=6 and 8 (**Table 4.2**). Eu^{3+} Lewis acidity is incredibly sensitive to first and second coordination shells, and it is possible during this pH range it is interacting either with *pN impurities or Np*pN in addition to the primer/template which could explain these results. Nevertheless, Eu^{3+} is very low in abundance in the crust, and is unlikely to be relevant prebiotically. The increase in rate was not enough to sustain longer extension, which would have been a reason to warrant further study, therefore we abandoned Eu^{3+} based primer extension.
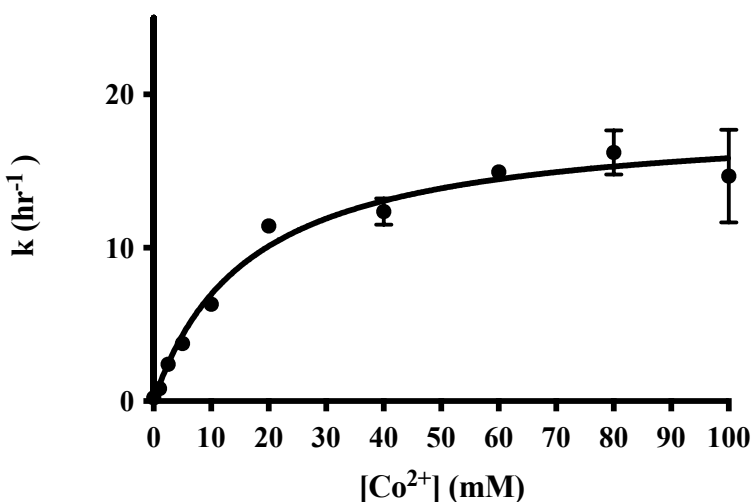
Figure 3.2. Dependence of Cp*pC primer extension on $Co(II)Cl_2$. Primer extension performed under standard conditions and fit to a saturation binding curve (R^2=0.98) Error bars indicate ± SD (n≥3, independent experiments). V_{max}= 18.5 ± 1 hr^{-1}, K_M= 16.6 ± 3 mM.

Bridged dinucleotide	$k_{Mg}/k_{Co(II)}$	k_{Mg}/k_{Ni}
Cp*pC	0.63	5.75
Ap*pA	5.5	12.7

Table 4.1. Relative primer extension rates of 20 mM Mg^{2+}: 20 mM M^{2+} in both Cp*pC and Ap*pA extension (9 timepoints)
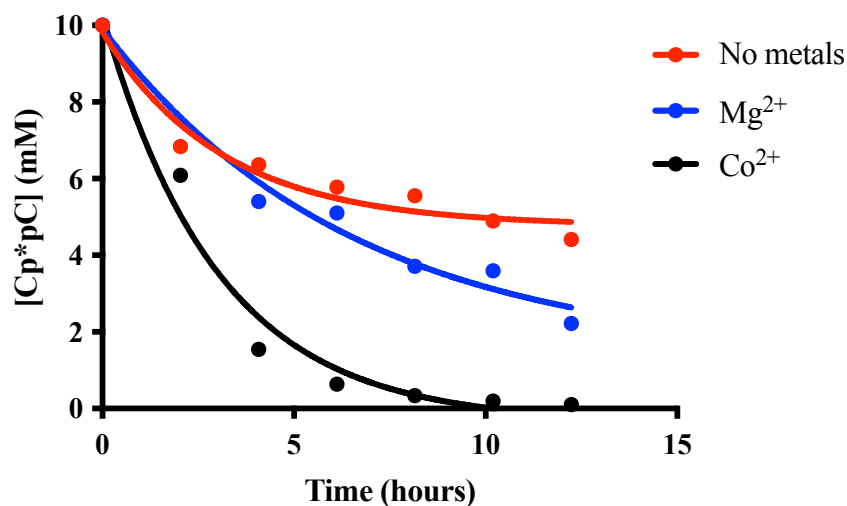
Figure 3.3. Cp*pC hydrolyzes faster in the presence of Co^{2+} than Mg^{2+}. Hydrolysis of Cp*pC was measured using analytical HPLC as described in the methods in the presence of 10 mM Co(II)Cl₂, MgCl₂ and in absence of added metal ions .The rate of hydrolysis was determined monitoring diminishment of the peak that corresponds with Cp*pC elution at 17 minutes. Quantification of hydrolysis by first-order kinetic plot of intensity of peak integration normalized to total peak integration intensity, and depicted relative to initial concentration of solution determined by A_{260} measured by NanoDrop. The rates of hydrolysis are as follows: k_{Co} is 0.3790 $\pm$ 0.036 hr^{-1}, with a $t^{Co}_{1/2}$=1.83 hrs. The k_{Mg} is 0.104 $\pm$ 0.02 hr^{-1}, with a $t^{Mg}_{1/2}$=6.64 hrs.

	pH 6	pH 7	pH 8
k_{Eu}/k_{Mg}	2.58	0.50	4.9

Table 4.2. 5 mM Europium chloride primer extension rates relative to Mg^{2+} in the same conditions across pH 6-8

<u>Different cations may have different requirements for catalysis in primer extension</u>

The cations sampled, in addition to different Lewis acidities, also have different sizes and charge densities, all of which are interrelated. However, the reactive center for primer extension has a particular geometry and defined optimal distances between the 3′-OH and the reactive phosphate of Np*pN bound to the template. Mn^{2+} is slightly larger than Mg^{2+}, with more flexible coordination geometry, yet it binds more tightly to the primer/template complex. Calcium is larger than both Mn^{2+} and Mg^{2+} but binds intermediately to the two. In attempting to measure the inhibition of calcium catalyzed primer extension in presence of cobalt hexammine (**Figure 3.4**), it is evident that cobalt hexammine inhibition is more complicated than simple competitive inhibition in this system. The V_{max} in presence of cobalt hexammine nearly doubles, instead of decreasing or remaining the same. However, the K_M also increases for Ca^{2+} substantially, supporting the two ions competing for the same site. While this is not within the scope of this chapter, these differences in catalytic requirements depending on the metal ion are important to consider in selection of metal catalysts on the early earth.
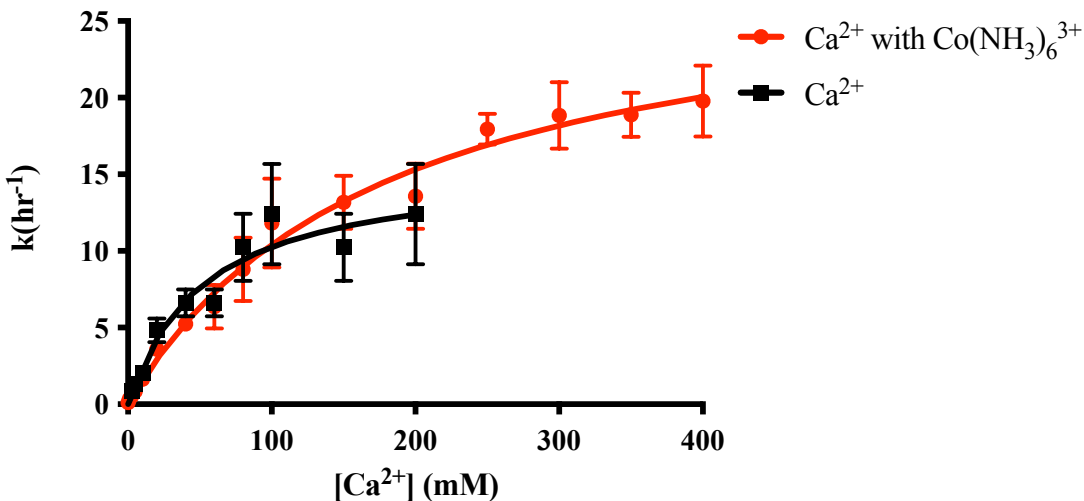
Figure 3.4. Cobalt hexammine and calcium catalyzed primer extension. Primer extension was performed with calcium chloride at pH=8, with and without 100 μM $Co(NH_3)_6^{3+}$. Lines represent the fits to a saturation binding curve for calcium binding the primer/template complex. Error bars indicate ± SD (n≥3, independent experiments). The red curve, corresponding to calcium in the presence of cobalt hexammine: V_{max}= 29 ± 1.9 hr^{-1}, K_M= 178 ± 26.4 mM. The black curve depicts primer extension in the presence of only calcium chloride, V_{max}= 15.6 ± 1.8 hr^{-1}, K_M= 52 ± 15.2 mM.

Miscellaneous attempts in metal chelation and probing of the inner-sphere contacts in primer extension

All of the metal ions sampled thus far have had their entire coordination spheres either available (all waters) or blocked for use (cobalt hexammine, as well as any ion in complex with EDTA). A future goal of this work is to understand the coordination geometry required for metal catalysis of primer extension in order to search for prebiotically plausible chelators or co-catalysts to localize metal ions in favorable conformations for primer extension, increase compatibility with protocells and prevent hydrolysis of Np*pN. Magnesium citrate has been explored previously,[59] and has been found to improve Mg^{2+} compatibility with protocells, but it

also decreases the rate of Mg^{2+}-catalysis of primer extension. It is unclear if this is due to steric effects of the Mg-citrate complex with the primer/template, if the Mg-citrate geometry is not the favored geometry for the metal ion to interact with the primer 3′-OH and phosphate oxygen of Np*pN, or an electronic effect of the citrate bound to the Mg^{2+} making Mg^{2+} less Lewis acidic. One way to look into these coordination requirements is through probing primer extension systematically with metal complexes with different exchangeable sites open in their first solvation shell.[38] This requires a fair amount of inorganic synthesis due to lack of commercial availability, but we were able to synthesize $Co(NH_3)_5OH_2^{3+}$ (**Figure B.1, Appendix B**) and probe its ability to catalyze primer extension.
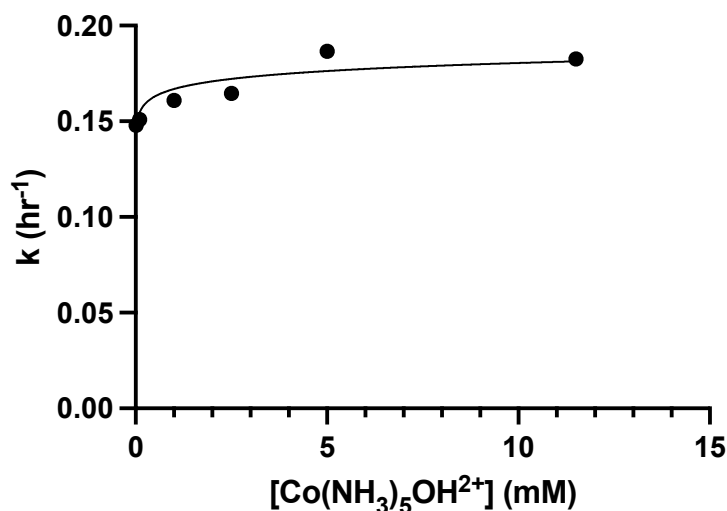
Figure 3.5. Effect of $Co(NH_3)_5OH^{2+}$ on Cp*pC primer extension. $Co(NH_3)_5OH^{2+}$ was made from $Co(NH_3)_5Cl^{2+}$ using established literature protocols and verified by spectroscopy (See **Appendix B**). Due to limited solubility, this was the concentration range that was able to be sampled for primer extension. Lines represent the fits to a saturation binding curve(R^2=0.37) for the metal complex the primer/template. Error bars indicate $\pm$ SD ($n \geq 3$, independent experiments). V_{max}= 0.17 ± 0.01 hr^{-1}, K_M= 0.0017 ± 0.001 mM.

We observed this complex has no benefit to primer extension across the ranges with which it is soluble in water (**Figure 3.5**). The ion itself should be sufficiently Lewis acidic to improve primer extension if it was able to form the correct coordination contacts. Therefore, one open coordination site is likely not sufficient for metal ion catalysis of primer extension. The inner-sphere water of the cobalto- species should be deprotonated at pH=8 (pK_a=6.2), and despite a possible bound-hydroxide, without inner-sphere coordination this is not sufficient for primer extension catalysis.[60] This species is also known to effect DNA conformation at low concentrations, through tight binding to the DNA, which may also account for the observed lack of benefit.[61] Looking into complexes with 2 open coordination sites in cis- and trans- positions would be of interest for future study (**Scheme 3**), but is beyond the scope of this dissertation. Examining similar tridentate complexes to Mg-citrate would also be of interest, but unfortunately Mg^{2+} affinity for different complexes is limited in scope. Therefore, transition metal complexes would be of interest for future study, perhaps with Mn^{2+} due to its improved catalysis of primer extension and facile complex formation.
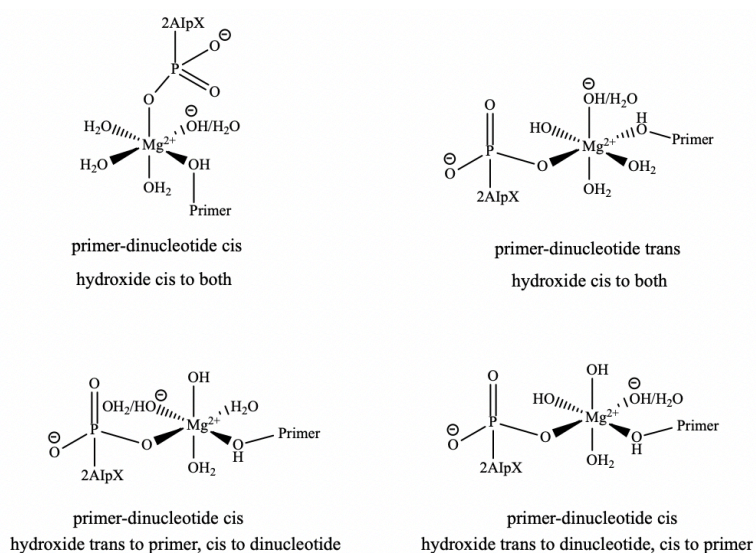
Scheme 3. Alternate inner-sphere geometries possible for Mg^{2+} with 3′-OH of primer, reacting bridged dinucleotide, and possible bound-hydroxide

Due to the weak binding affinity of Mg^{2+} for the primer/template compared to enzymatic systems of polymerization, one can imagine a selective benefit for protocells that synthesize or sequester compounds that can selectively localize Mg^{2+} for primer extension. One possible means for this would be the reversible complexation of the 2′- or 3′-OH with an aldehyde that has possible ligands for a metal ion (**Scheme 4**). The idea is similar to SHAPE chemistry but instead of acylation, hemiacetals are formed, and the process would be reversible.[62] The equilibrium between hemiacetal and free aldehyde as opposed to the hydrate is a substantial barrier to the formation of these complexes, and is also very pH dependent, however.
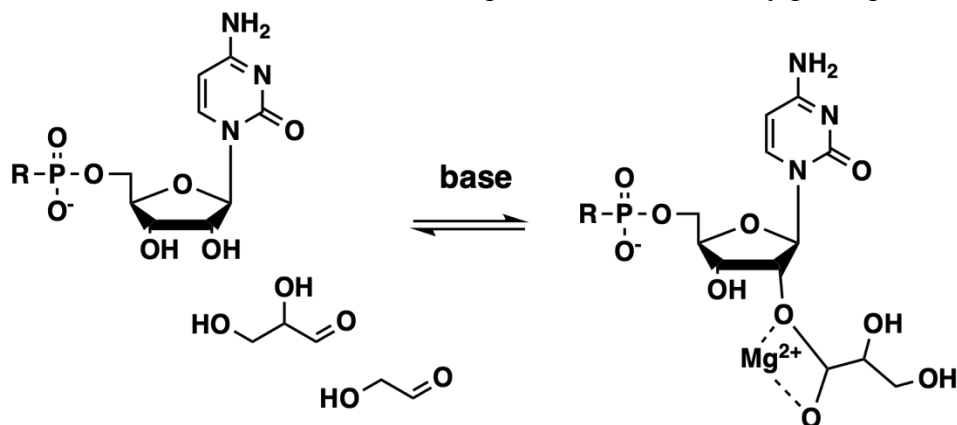
Scheme 4. Proposed transient hemiacetal formation with RNA as a possible Mg^{2+} chelator

We chose to look into glycolaldehyde and glyceraldehyde and measure if there was any benefit to their presence with low concentrations of metal ions. These aldehydes are also important in reactivation chemistry, so their compatibility with primer extension is additionally relevant to establish. What we found is that there is no benefit in the catalysis of primer extension with the presence of these aldehydes, even in concentrations that should favor the formation of the hemiacetals (**Table 4.3 and 4.4**). This was independent of Mg^{2+} concentration, although these complexes would mainly be of interest if they improved rates at low concentrations of Mg^{2+}. This is not incredibly surprising given the poor affinity of Mg^{2+} for alcohols, or the issues of ion geometry that may arise with a covalently linked chelator. The lack

of selectivity for the 2′- vs. 3′-hydroxyl would also prove to be an issue for this form of localization, and would possibly harm the regioselectivity of primer extension if the reversible covalent chelators were successful. Although this attempt at *in situ* metal ion localization failed, this form of increasing local concentration of metal ions at the reaction center for primer extension was likely an early focus of evolution in ribozyme polymerases. Optimistically for the compatibility with reactivation chemistry, the presence of these aldehydes also does not substantially harm primer extension.

Relative concentration of glycolaldehyde to primer	*Relative rate of reaction ($k_{gly, Mg}/k_{Mg}$)*
10x	0.81
100x	0.78
1000x	0.76
10,000x	0.71

Table 4.3. Comparison of rates of varying concentrations of glycolaldehyde with 5 mM Mg^{2+} in primer extension with Mg^{2+} alone.

Relative concentration of glyceraldehyde to primer	*Relative rate of reaction ($k_{gly, Mg}/k_{Mg}$)*
10x	0.79
100x	0.77
1000x	0.77
10,000x	0.66

Table 4.4. Comparison of rates of varying concentrations of glyceraldehyde with 5 mM Mg^{2+} in primer extension with Mg^{2+} alone. Rates shown are relative to the same conditions (5mM Mg^{2+})

(**Table 4.4 cont**.) in absence of the aldehyde. Experiments were performed to ensure that preincubation of the aldehyde with the primer/template had no effect (*not shown*).

Solvent effects on metal ion interactions

It has been previously shown in ribozymes that the presence of ethanol can improve structural stability and A-form helix formation, as well as improve electrostatic interactions by lowering the dielectric constant of the solution.[63] In the case of nonenzymatic primer extension, since there is no specific structural core, however, the addition of ethanol does not enhance primer extension, and instead mildly decreases the maximum rate (**Figure 3.6**) . Additionally, with the amount of ethanol used (10%), it was not sufficient to increase Mg^{2+} association. It is possible with these weak binding affinities a substantial change in dielectric constant would be required to specifically increase the local concentration of Mg^{2+} at the positions relevant for primer extension.
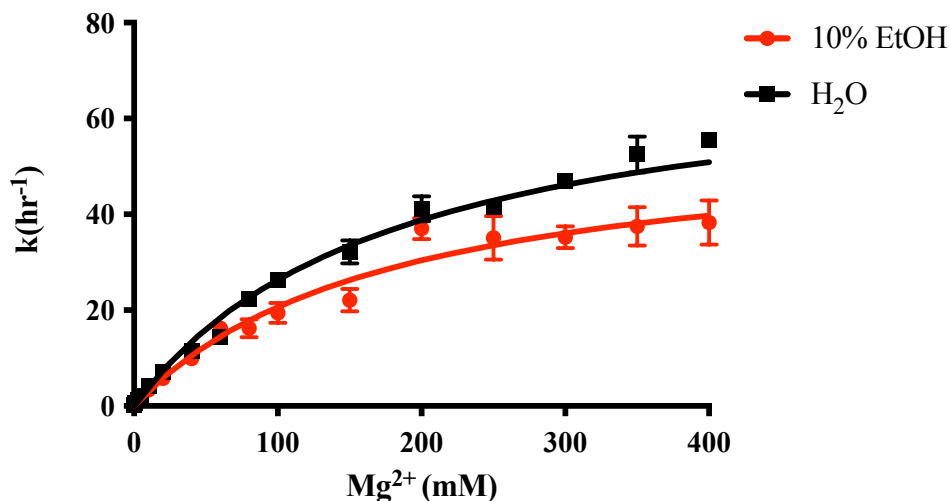
Figure 3.6. Mg^{2+} binding unaffected in alternate solvent composition- 10% EtOH. The H_2O curve is reproduced from Figure 2A, and the 10% EtOH curve is at pH=8 with 10% v/v ethanol in the final volume of the reaction. Lines depict the fit of the data to a saturation binding curve for Mg^{2+} binding to the primer/template complex. Error bars indicate $\pm$ SD (n$\geq$3,

(**Figure 3.6 continued**) independent experiments). For the 10% EtOH curve, $V_{max} = 57.5 \pm 5.1$ hr^{-1} and $K_M = 178.0 \pm 34.8$ mM.

<u>Apparent pK_a measured through kinetics supports the stabilization of the negatively charged 3′-OH being key for catalysis of primer extension</u>

The pH dependence of primer extension can illuminate mechanistic information regarding rate-limiting deprotonations for the reaction, and can also depict if the metal ion catalyst plays a role in the deprotonation of the 3′-OH. Regardless of its position on the reaction coordinate, the 3′-OH needs to lose a proton at some point during the reaction. It has been established that this deprotonation is not rate-limiting, and the following data supports this hypothesis.[28] Primer extension rates increase with increasing pH, and fit to a sigmoidal curve representing the deprotonation of the 3′-OH (**Figure 3.7**).We observed that with increasing concentrations of Mg^{2+}, apparent pK_a decreased. This effect supports a number of our proposed hypotheses regarding primer extension, which would be dependent on occupancy of Mg^{2+} at the reaction site. This effect of decreasing pK_a in presence of increasing Mg^{2+} concentration holds in presence of free cobalt hexammine as well, although the initial pK_a is much lower with the nonspecific ionic stabilization (**Figure 3.8**).

The pK_a of the 3′-OH is sensitive to the Lewis acidity of the cation present (**Figure 3.9**), with the pK_a trending $Ca^{2+} > Mg^{2+} > Mn^{2+}$, with Mn^{2+} substantially decreasing the pK_a to 7.5 from 9.32 in absence of metal ions. While the fact that the slope is near one in the pH dependence of log(k) supports the proton inventory of one deprotonation step occurring *en route* to the transition state (**Figure 3.10**),[49] there are notable limitations to measuring pK_a utilizing this kinetically rooted method. Rates above pH 9.5 are not reliable due to the deprotonation of G and destabilization of the primer/template duplex that results from this. Unless much of the pH curve

plateaus before this pH, it can be difficult to reliably determine the pK_a, which is the case for measuring the pK_a at 0 mM Mg^{2+} (literature values suggest the pK_a of the 3′-OH is 12.6).[28] This clear discrepancy in pK_a determination led us to examine other means of looking into the pH dependent effects of metal ion catalysis in primer extension.
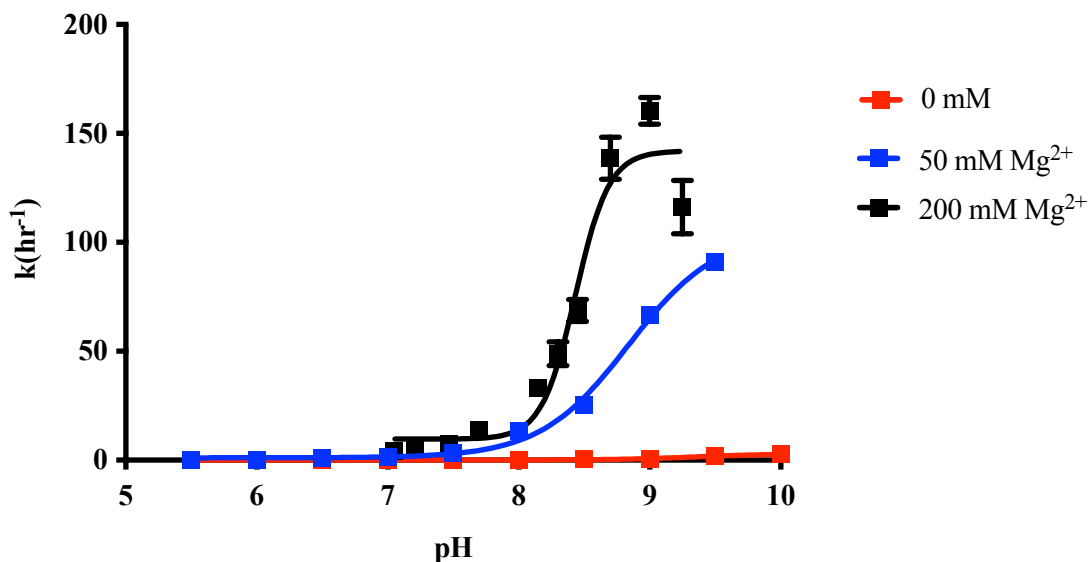
Figure 3.7. Mg^{2+} concentration affects apparent pK_a in primer extension system. The rate of primer extension was measured across different pH using MES (5.5-7), Tris (7-9), CHES (9.5-10) as described in Appendix B. Rates were measured across the above pH range and fit to an allosteric sigmoidal curve, the logEC50 corresponding to predicted pK_a of the rate limiting deprotonation. For 0 mM Mg^{2+} this value corresponds to 9.32 ± 0.3, the pKa for 50 mM Mg^2 $=8.8 \pm 0.07$, and for 200 mM $Mg^{2+} = 8.5 \pm 0.03$. Error bars indicate $\pm$ SD ($n \geq 3$, independent experiments).
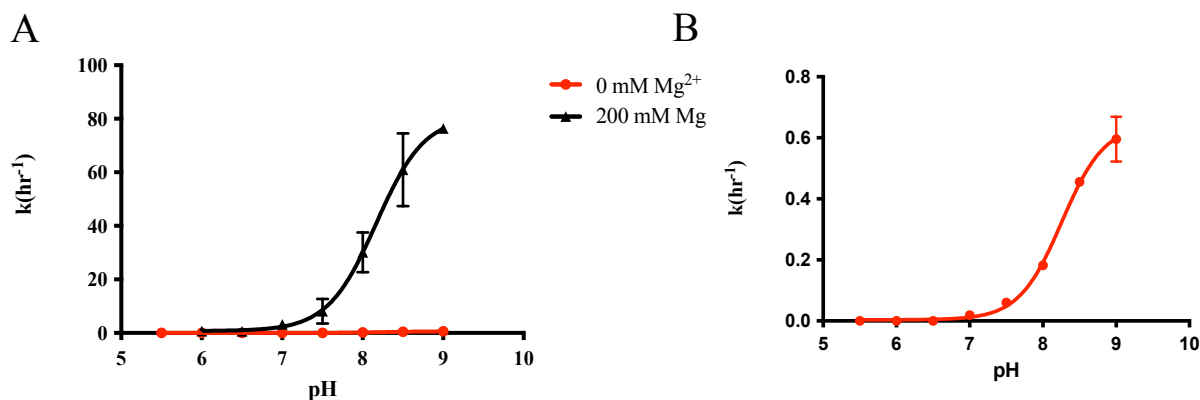
Figure 3.8. Effect of pH on primer extension at constant Mg^{2+} with 100 μM $Co(NH_3)_6^{3+}$. Rates were measured across the above pH range (using buffers as described in Appendix B) and fit to an allosteric sigmoidal curve, the logEC50 corresponding to predicted pK_a of the rate limiting deprotonation. The predicted pKa for 0mM Mg^{2+} in presence of cobalt hexammine is 8.3 ± 0.04, and in the presence of 200 mM Mg^{2+} is 8.17 ± 0.09. (B) is the closeup of the 0 mM pH curve. Error bars indicate $\pm$ SD (n≥3, independent experiments).
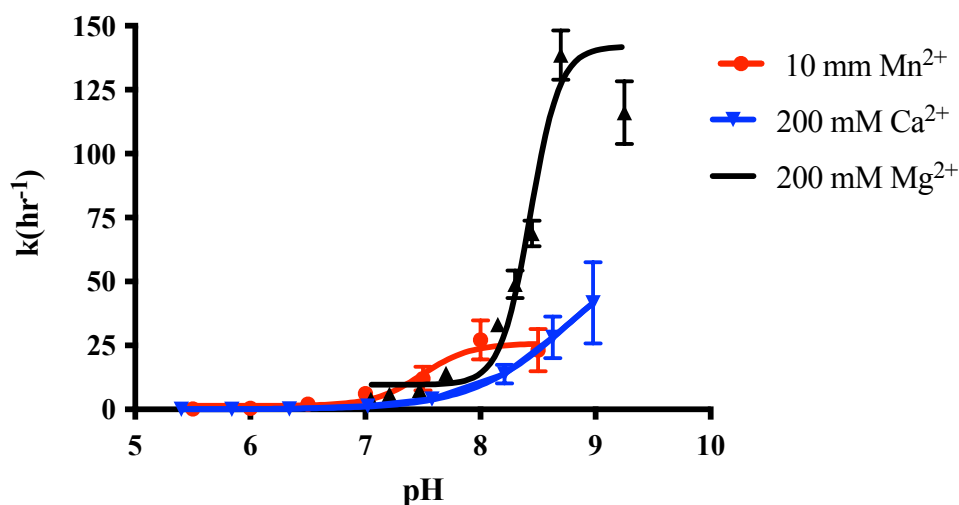
Figure 3.9. Apparent pK_a of primer extension decreases with increasingly Lewis acidic cations. The rate of primer extension was measure across different pH using buffers as described in Appendix B. Rates above 9.5 are not included due to extensive off-template extension and known dissociation of the helix due to G deprotonation above this pH. Primer

(**Figure 3.9 continued**) extension rates were measured across the above pH range and fit to an allosteric sigmoidal curve, the logEC50 corresponding to predicted pK_a of the rate limiting deprotonation. The Mg^{2+} curve is reproduced from Figure 3.6. The Mn^{2+} concentration was chosen to minimize precipitation across the pH range. The apparent pK_a in the presence of 10 mM $Mn^{2+} = 7.5 \pm 0.13$, and in the presence of 200 mM Ca^{2+} $pK_a = 8.8 \pm 0.04$. Error bars indicate $\pm$ SD (n$\geq$3, independent experiments).
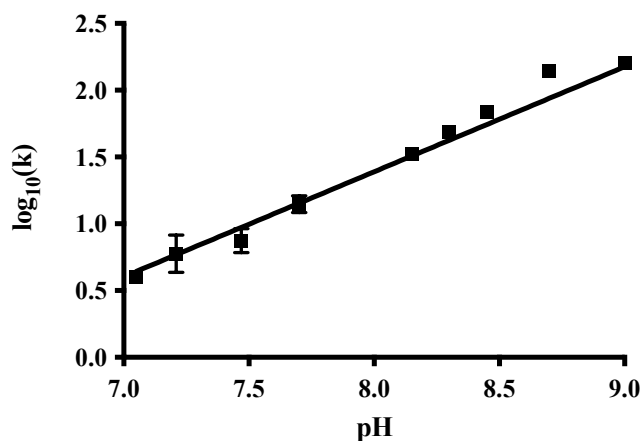
Figure 3.10. pH dependence of $\log(k_{obs})$, 200 mM Mg^{2+}. Analysis of $\log(k)$ vs. pH from Fig 3.7. The data is fit to a line with a slope of 0.76 $\pm$0.03 and R^2=0.988. Reactions were performed at room temperature using 0 mM Mg^{2+} and Tris buffer. Error bars indicate $\pm$ SD (n$\geq$3, independent experiments).

Solvent isotope effects in primer extension

Solvent isotope effects inform on rate limiting steps of reactions, and have been studied in detail in a recent publication from our group for nonenzymatic primer extension.[28] The following experiments were performed independently from this publication. Our results agree with the published conclusions, and support the hypothesis that the deprotonation of the 3′-OH is not the rate-limiting step for primer extension. For example, the apparent pK_a differences between the pD and pH experiments (**Figure 3.7 and Figure 4**) can be accounted for simply by

the change in the solvent acidity between H_2O and D_2O. This hypothesis is further supported by the comparison in rates in the linear-regime of the pL curve. At pL=8, the rate difference of primer extension in H_2O and D_2O is not significant (2.9 without Mg^{2+}, 1.3 in presence of Mg^{2+}, **Figure 4.1**).[64,65] Due to the lack of deprotonation in the transition state, we cannot use the solvent kinetic isotope effect to assess the presence or absence of a catalytically relevant metal hydroxo-species.[66]
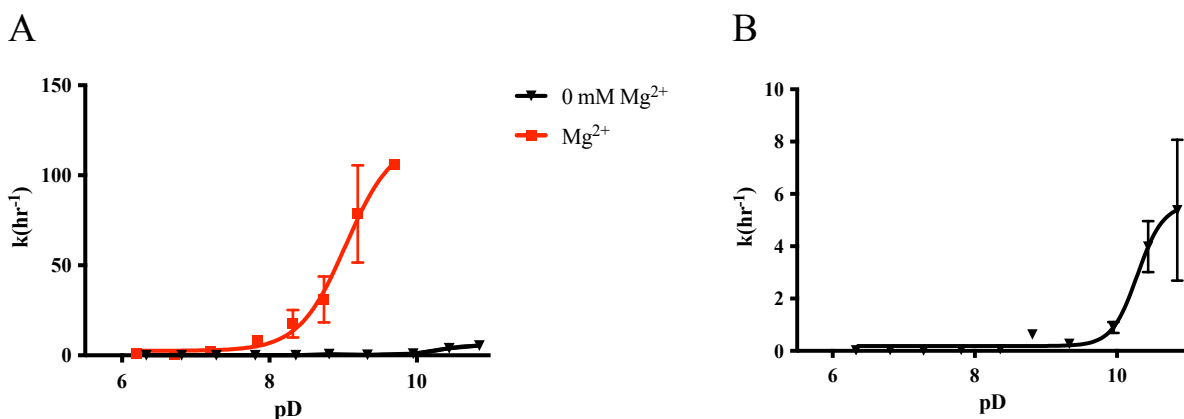
Figure 4. pD of primer extension with and without 200 mM Mg^{2+}. Primer extension reactions were performed in entirely D_2O solutions, with deuterated buffers, and anhydrous magnesium chloride dissolved in D_2O, and pH adjusted using DCl and NaOD. (B) Inset of the 0 mM $MgCl_2$ curve. The apparent pK_a in absence of Mg^{2+} was pD=10.3 ± 0.11, and in presence of 200 mM Mg^{2+} =9.04 ± 0.08. Error bars indicate ± SD (n≥3, independent experiments).
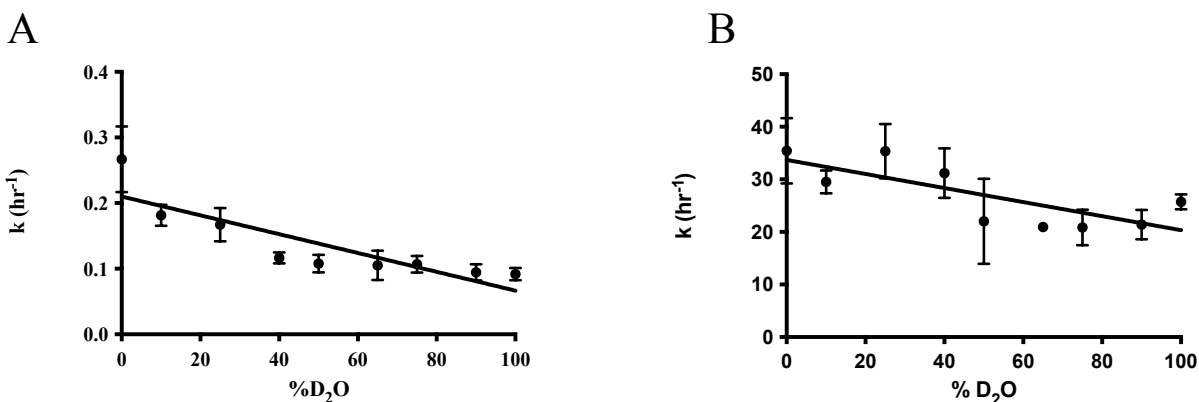
Figure 4.1. Solvent kinetic isotope effect in primer extension with and without 200 mM Mg^{2+} across different D_2O: H2O ratios (A)The kinetic isotope effect over different % of D_2O in absence of Mg^{2+}. $k^{H2O}/k^{D2O}= 2.9± 0.2$ (B) The kinetic isotope effect over different % of D_2O in presence of 200 mM Mg^{2+} $k^{H2O}_{Mg}/k^{D2O}_{Mg} = 1.38 ±0.18$. Error bars indicate ± SD (n≥3, independent experiments).

<u>Mg^{2+} binds the reaction center more tightly as pH increases</u>

As we attempted further experimentation across different pH ranges with different cations, it arose that Mg^{2+} may not be saturably bound to the primer/template across the pH ranges sampled for the pK_a curves. Therefore, we decided to measure the binding affinity of Mg^{2+} at pH=6, 7.5, 8, and 9 to address this concern (**Figure 5**). We found this Mg^{2+} is not equivalently associated with the primer/template across this pH range, adding another layer of difficulty to the pK_a curves, as there is no constant concentration of Mg^{2+} that would be saturating across the full pH range measured (pH=6 would require >1M Mg^{2+}, which would result in too much degradation and too fast of rates at higher pH). This would result in an underestimation of rates at lower pH in our constant Mg^{2+} pK_a measurements, which in worst case the pK_a determination of the metal-bound 3′-OH would also be underestimated. This necessitates for us subsequent measurement of the pK_a through non-kinetic means, although this will not be addressed in this thesis.

We asked if we could explain if this increase in Mg^{2+} binding affinity with the primer/template. There are two notable differences across the pH range. First is the substantial change in V_{max}— this is not surprising, as deprotonation is required for primer extension to occur, and the rate of primer extension is proportional to the amount of free hydroxide concentration. At a given pH the concentration of hydroxide is much larger than the amount of primer/template, and is ignored in our pseudo-first order approximation. The uncoordinated 3′-OH has a pK_a of approximately 12— assuming no metal ion interaction we would expect a negligible change in charge density and therefore no change in binding affinity. There would also be negligible amounts of magnesium-hydroxide species in the pH range we examined (up to pH 9.5), which has a pK_a of 11.42 (**Table 4**). However, a Mg^{2+} ion directly coordinated with the primer 3′-OH could potentially lower the pKa of the 3′-OH substantially (**Figures 3.7, 3.8, 3.9**).[28] Our results show that that Mg^{2+} binds more tightly at higher pH: at pH=9, the K_M for Mg^{2+} is 107 mM, while at pH=6, it is 940 mM (**Figure 5, Table 5**). As Mg^{2+} interacts with the 3′-OH, if it lowers the pK_a of the 3′-OH, we would expect in the pH range sampled that there would be a greater amount of negative charge build-up as pH increases. If the catalytic metal interacts directly with the 3′-OH, the electrostatic interaction of the +2 charge of the metal ion with the partial negative charge on the 3′-OH should become stronger as the 3′-OH becomes increasingly negatively charged. Given our previous results with the maximum rate contributions of differently Lewis acidic metals (**Figure 3**), we believe these results are not consistent with increase in metal bound hydroxide and instead are consistent with expectations for a transient inner-sphere coordination with the 3′-OH. Taken together with the failure of cobalt hexamine to exhibit any catalytic effect beyond that attributable to ionic strength, the effect of pH on binding of the catalytic metal ion supports a model in which the catalytic metal ion stabilizes the

deprotonated 3′-OH through a direct inner-sphere interaction throughout the reaction coordinate, with Mg^{2+} interacting more strongly with the deprotonated 3′-OH. However, we first wanted to ensure there weren't other notable deprotonations that could be occurring in the primer extension reaction across the pH range sampled.

A

B
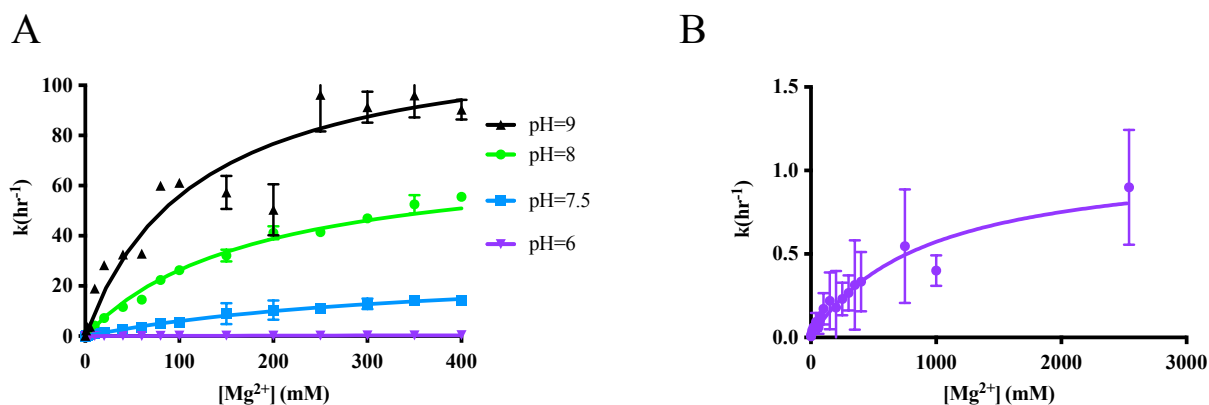
Figure 5. Mg^{2+} binding affinity improves with increasing pH. Primer extension performed under different buffer conditions: pH 9, 100mM CHES; pH 8 and 7.5, 100 mM Tris-Cl; pH 6, 100 mM MES. pH 7.5 and 8 curves are reproduced from plots in Figure 2A and 3. Lines depict the fit of the data to a saturation binding curve for Mg^{2+} binding to the primer/template complex. Error bars indicate ± SD (n≥3, independent experiments). (B) pH=6 at a scale where the curvature is visible (0-2.5M $MgCl_2$).

pH	V_{max} (hr^{-1})	K_M (mM)
6	1.1 ± 0.18	940 ± 284
7.5	31 ± 2	417 ± 54
8	74 ± 2	181 ± 13
9	116.7 ± 8.4	107 ± 21

Table 5. Mg^{2+} binding and rate parameters at different pH for primer extension of saturating Cp*pC on GGG template (**Figure 5**).

pH effects are not due to deprotonation of Cp*pC

        To ensure that the previous pH curves and increase in metal binding affinity are not due to another deprotonation event outside of the 3′-OH, we examined the other acidic protons present in the primer extension reaction. The only other proton that may be relevant in affecting the reaction is that of the exocyclic amine of the imidazole in Cp*pC, where the negative charge can be stabilized through symmetric hydrogen bonding. We determined the pK_a of this by measuring the ^{31}P NMR signal of Cp*pC across a wide range of pD (**Figure 5.1**). The phosphates are very acidic and deprotonated in the pH ranges sampled,[67,68] so any change to the phosphorus signal (which would be affected by any change in electron density in the imidazole), can be correlated to the deprotonation of the exocyclic amine. While we observed this deprotonation occurring, it did not occur until above pH*=11.8 or pD=12.2,[69,70] outside of any pH range sampled experimentally for primer extension. The deprotonated amine would not exist appreciably in solution in our experiments. It has not been measured if this is affected by the presence of metal ions, or through association with the primer/template complex, however. Mg^{2+} has fairly weak association with nitrogen without significant structural organization and we have not observed any association of Mg^{2+} with the bridged dinucleotide by NMR at pH ranges relevant for primer extension. Another possibility to rule out is that the observed lower binding affinity of Mg^{2+} at lower pH could be due to other effects on the rate of primer extension, such as increased hydrolysis of Cp*pC at lower pH.
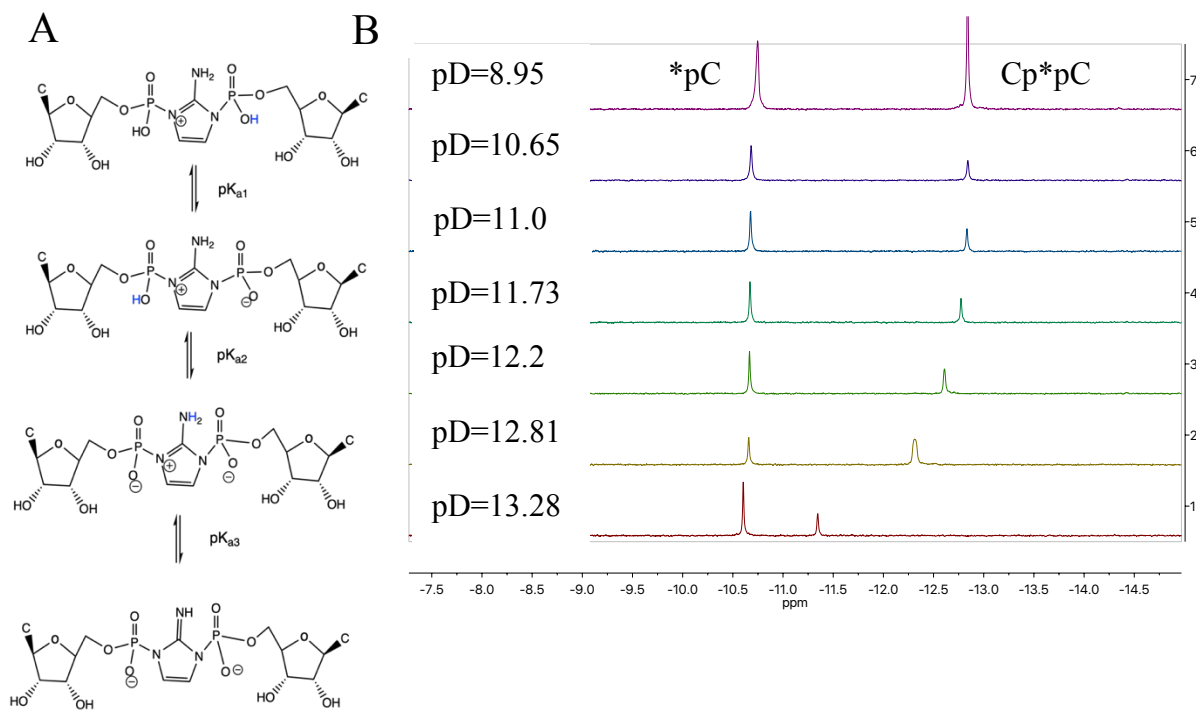
Figure 5.1. pD titration of Cp*pC studied by ^{31}P NMR shows a high pD deprotonation (A) Scheme of possible deprotonations visible through ^{31}P NMR that would be of interest to primer extension. Blue protons represent the proton being removed in the subsequent deprotonation step. (B) In pure D_2O, Cp*pC was adjusted from high pD to low pD using DCl and the ^{31}P NMR measured, no buffers present. pD is calculated from pH*+ 0.4, with pH* being the pH meter measurement of the sample in pure D_2O. Spectra shown is zoomed into relevant region for phosphorus in the bridged dinucleotide. External P(OMe)₃ standard was utilized to normalize the measurements.

Observed pH effects independent of Cp*pC hydrolysis

Another consideration for interpreting the increase in rate and Mg^{2+} binding with increasing pH is to ensure that Cp*pC is not being stabilized substantially from hydrolysis in the sampled ranges, or increasingly hydrolyzed. We find that while Cp*pC is increasingly hydrolyzed at low and high pH, with greatest stability in neutral to mildly alkaline pH (**Figure 5.2**), this difference in rate is too small to account for the previously observed effects in **Figure 5**. The half-life of Cp*pC is substantially longer than the experimental times sampled regardless of the pH (**Table 5.1**), and therefore has no effect on the previously discussed results, although the kinetics of hydrolysis are measured in absence of added Mg^{2+}.
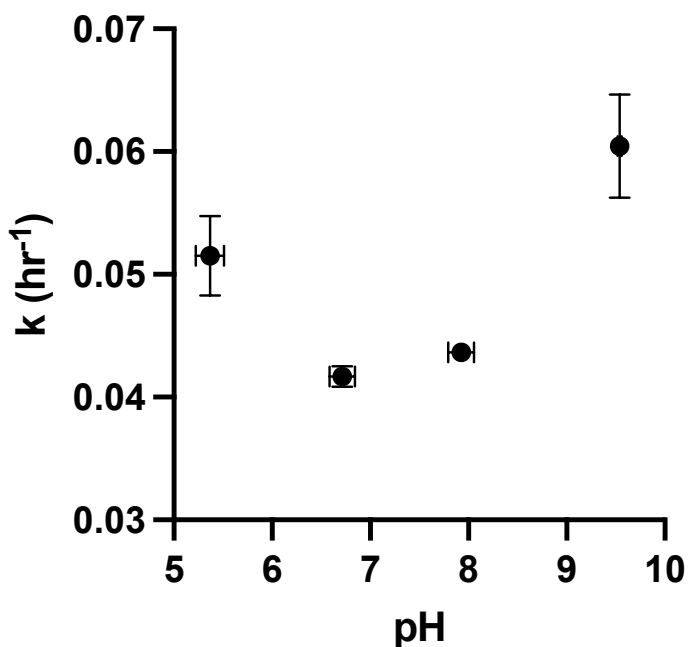
Figure 5.2. pH dependence of hydrolysis of Cp*pC. Hydrolysis of Cp*pC was measured using analytical HPLC as described in the methods. The rate of hydrolysis was determined monitoring diminishment of the peak that corresponds with Cp*pC elution at 17 minutes. Error (**Figure 5.2**

continued) bars shown are ± SD (n=3). Quantification of hydrolysis by first-order kinetic plot of intensity of peak integration normalized to total peak integration intensity.

pH	Cp*pC half-life (hr.)
5.37 ± 0.15	13.5 ± 0.86
6.71 ± 0.13	16.6 ± 0.34
7.93 ± 0.13	15.9 ±0.28
9.54 ± 0.02	11.5 ±0.77

Table 5.1. Half-life of Cp*pC at different pH as measured by analytical HPLC (**Figure 5.2**).

Mg^{2+} binds the reaction center more tightly when 3′-OH-phosphate distance is decreased

An additional potentially catalytic role for Mg^{2+} involves coordination with one of the non-bridging oxygens of the reactive phosphate of the imidazolium-bridged dinucleotide. Thus, the catalytic metal ion could potentially interact simultaneously with both the 3′-OH and a phosphate oxygen (**Scheme 2**). Such a dual interaction could bring the two reactive groups closer together, and/or could serve to correctly orient the 3′-OH for in-line attack on the phosphate. In previous crystallographic studies, we have found that the presence of a downstream helper oligonucleotide diminishes the distance between the 3′-OH and the phosphorus of the adjacent phosphate of the imidazolium-bridged dinucleotide.[19] We hypothesized that if the catalytic metal ion does indeed interact with both the primer 3′-OH and the adjacent phosphate, the diminished 3′-OH to P distance might facilitate binding of the catalytic metal ion. We therefore measured the K_M for Mg^{2+} in a primer extension reaction in which the reactive Cp*pC bridged dinucleotide was sandwiched between the primer and a downstream helper oligonucleotide (**Figure 6A**). We find that the K_M for Mg^{2+} decreased from 366 ± 33 mM in the absence of the downstream oligonucleotide to 195 ± 14 mM in the presence of the downstream oligonucleotide (**Figure 6B**).

This is consistent with the hypothesis that Mg^{2+} interacts with the 3′-OH and the phosphate oxygen of the bridged dinucleotide— however it does not give us information on if this interaction with the bridged dinucleotide is inner-sphere or outer-sphere in nature. Given the previous results with the cobalt pentaamine hydroxide, it is likely this interaction is also inner-sphere (**Figure 3.5**). The presence of the downstream oligonucleotide does not appear to affect the maximum rate of primer extension, although this is difficult to fully verify due to difficulty measuring high Mg^{2+} concentrations in the helper system due to the concentration of oligonucleotides used and their subsequent solubility. Thus, a decreased 3′-OH to P distance facilitates binding of the catalytic Mg^{2+}, but does not appear to affect the geometry of the reaction center in a catalytically significant manner for this particular system. This is surprising, as we would anticipate that the presence of a downstream helper should improve the overall conformation of the RNA helix enough to increase the overall rate of primer extension compared to the freely rotating template sans helper.[4] However, the short length of helper sampled (6mer) is likely too transient/flexible in its binding to substantially order the template in the RNA-only system.
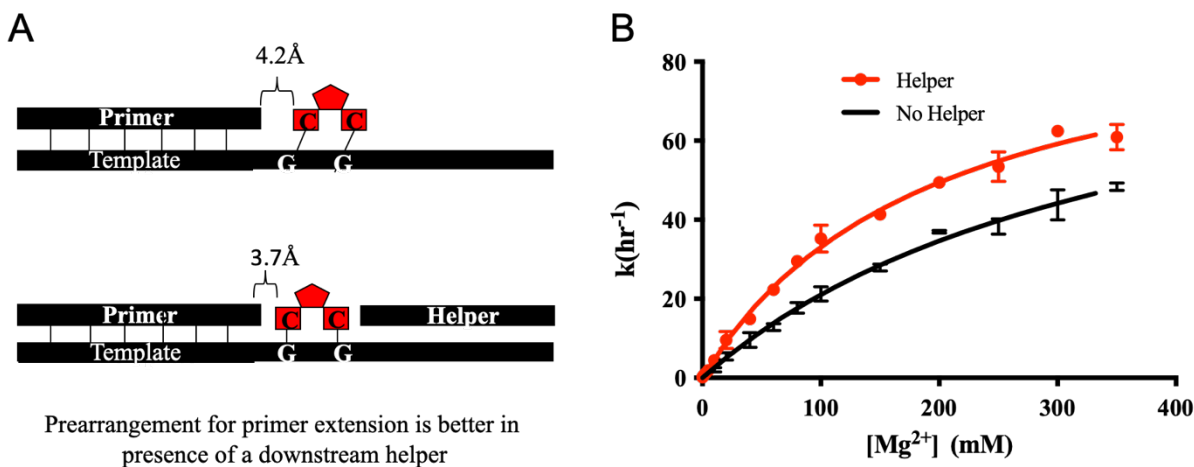
Figure 6. Presence of a downstream helper oligonucleotide increases Mg^{2+} binding affinity. Primer extension reactions were performed using the same primer, but a longer template was

(**Figure 6 cont**.) used to accommodate the downstream helper oligonucleotide (see **Table A.1 in Appendix A** for oligonucleotide sequences). Primer extension reactions were performed with 2 μM primer, 5 μM template, 20 mM Cp*pC, and when present, 10 μM helper oligomer. 100 μM CoHex was used to stabilize primer and helper binding to the template. (A) Schematic of primer/template/Cp*pC complexes with and without helper oligonucleotide. The primer and helper oligonucleotides flank the Cp*pC which binds in a two-nucleotide gap (B)The lines shown are fits of Mg^{2+} binding to the primer/template complex from a saturation binding curve. Error bars indicate ± SD (n≥3, independent experiments). In the absence of the helper, V_{max}=98 ± 5.4 hr^{-1} and K_M=366 ± 33 mM. In the presence of the helper, V_{max} is 99± 3.4 hr^{-1}, and K_M=195 ± 14 mM.

Additional conformational effects on Mg^{2+} binding

Reactivity in primer extension is a function of the conformation of the duplex, in addition to the intrinsic reactivity of the 3′-OH as a result of said conformation. RNA and A-form helices are ideal for primer extension catalysis in terms of geometry of the 3′-OH relative to the reacting phosphate.[28] Locked nucleic acids are trapped in A-form conformation. We found LNA, when present in the template beyond the primer, produced comparable maximum rates with saturating Mg^{2+} to RNA, but exhibited much tighter metal ion binding, likely due to this conformational trapping (**Figure 6.1, Table 6**). If the full primer and template were constructed of LNA we likely would have observed a rate enhancement in addition to this increase in Mg^{2+} binding affinity. We were also interested in the effects of helical conformation combined with the difference in reactivity when the 2′-OH is absent. We utilized an RNA primer and template, but the last base of the primer was a deoxyribonucleotide ("DNA-ended RNA"). When DNA is the reactant in an otherwise A-form helix (DNA-ended RNA), we found metal binding matched that

of RNA-only complexes, which suggests there is not a detriment to losing the 2′-OH for metal interactions, verifying its lack of involvement in Mg^{2+} binding as necessary for primer extension. We tested the converse, with a DNA primer and template, with the last base of the primer a ribonucleotide ("RNA ended DNA"). With the reacting 3′-OH of the RNA in an otherwise B-form helix, Mg^{2+} binding is worse, and overall reactivity is worse. The conformation of the helix is important in ordering the primer/template in an optimal geometry for primer extension, especially in the case of Mg^{2+} binding. In A-form like complexes, Mg^{2+} binds better, and rates
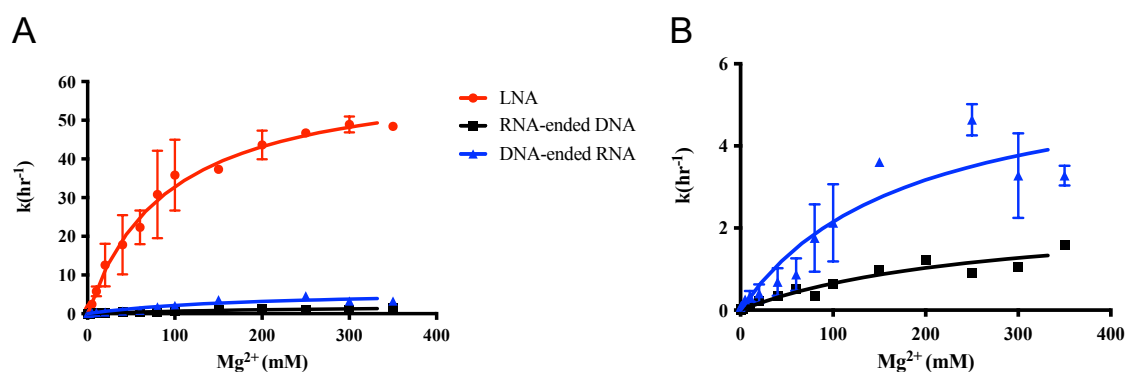
are improved.

Figure 6.1. Mg^{2+} binding is sensitive to conformation of nucleotide duplex (A) Mg^{2+} binding experiments with the same sequence, but using an LNA template construct (with LNA template in the nucleobases beyond the primer/template binding), all RNA construct with one DNA at the end of the primer ("DNA ended RNA") , all DNA construct with one RNA at the end of the primer ("RNA-ended DNA") (see **Table A.1 in Appendix A**). Lines depict the fit of the data to a saturation binding curve for Mg^{2+} binding to the primer/template complex. Error bars indicate ± SD (n≥3, independent experiments). (B) Inset of DNA-ended RNA and RNA-ended DNA constructs (k= 0-6 hr^{-1}).

Construct	V_{max} (hr^{-1})	K_M (mM)
LNA	62.97 ± 2.56	91.79 ± 10.85
DNA-ended RNA	6.02 ± 1.16	179.5 ± 71.95
RNA-ended DNA	2.41 ± 0.61	268.2 ± 122.7

Table 6. Saturation binding curve fit for Figure 6.1

Helper oligomer improves binding and rate in LNA-containing systems

While not a pure LNA construct, the "LNA helper system" contains a longer stretch of LNA in the template (all the nucleobases past the end of the primer), overlapping not just the area of extension but also the helper binding region. The sequence composition is otherwise the same as in Figure 6. The primer and helper in this system are still made of RNA, but the helper should interact more strongly with this template.[71-74] In contrast to what is observed in Figure 6, adding a downstream helper does increase the maximum rate of primer extension with LNA in addition to aiding in Mg^{2+} binding (**Figure 6.2**). In the case of the rigid LNA template, the presence of a downstream helper is less transient and more ordered, which is likely why the rate benefit is observable.[71-74] Additionally, the decrease in K_M of Mg^{2+} observed in this system further supports the interaction between Mg^{2+} and the oxygen of the reacting phosphate of Np*pN.
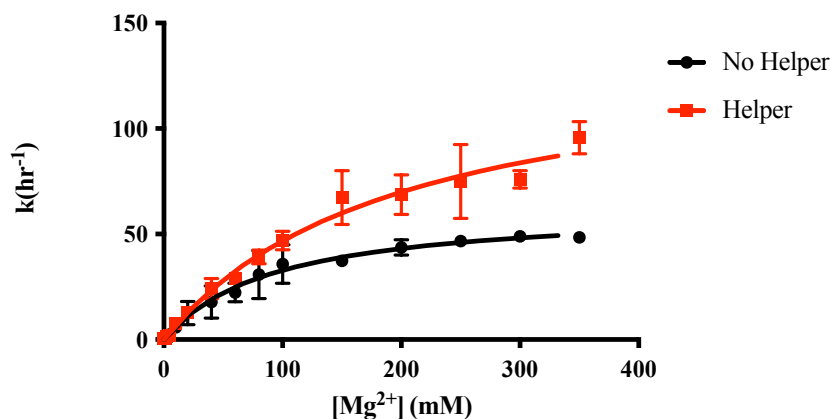
Figure 6.2. Presence of downstream helper in an LNA system aids in Mg^{2+} binding and rate. The same oligomer system as in Figure 6A was utilized for primer extension but instead was entirely constructed of LNA. Lines depict the fit of the data to a saturation binding curve for Mg^{2+} binding to the primer/template complex. Error bars indicate $\pm$ SD ($n \geq 3$, independent experiments). In absence of helper, $V_{max}= 82.5 \pm 11.9$ hr^{-1}, $K_M= 336.8 \pm 79.8$ mM. In presence of helper, $V_{max}= 138.8 \pm 12.1$ hr^{-1}, $K_M= 197 \pm 33.5$ mM.

Conclusions

We have measured the effects of a series of metal ions on the kinetics of nonenzymatic template-directed RNA primer extension reactions, under different pH and ionic conditions, and in the presence and absence of a downstream RNA helper oligonucleotide. Our observations support a model in which the catalytic metal ion interacts with both the primer 3′-OH and one of the non-bridging oxygens of the adjacent phosphate of the reactive imidazolium-bridged dinucleotide (**Figures 3, 5, 6, 6.2**). In contrast, nonspecific ionic effects confer at most a two-fold rate enhancement attributable to outer-sphere contacts (**Figure 2C**), and no interactions with the 2′-OH are necessary for primer extension (**Figure 6.1A**).

The correlation between the Lewis acidity of the catalytic metal ion is consistent with an inner-sphere interaction with either the primer 3′-OH or the reactive phosphate, or both. The pK$_a$

of the inner-sphere waters are each about one pH unit apart with $Mn^{2+} < Mg^{2+} < Ca^{2+}$. While

naively would lead us to expect rate differences of about 10-fold different between each, which

is greater than the V_{max} differences observed, since the deprotonation step is not rate-limiting for

nonenzymatic primer extension (**Figure 4.1**),[28] these will not correlate in such a straightforward

manner. Additionally, a bound-hydroxide is unlikely to play a major role in the catalysis of

primer extension based on the relative rates expected from expected $[M^{2+}(OH)_2]$ at pH=7.5

(**Figure 3**) and the observed solvent kinetic isotopic effects (**Figure 4.1**). While this may not

account for the majority of observed catalysis, due to the overall disorder of the reaction center

and weak affinity for metal ions, multiple modes of catalysis may contribute to the observed

rates.

Another aspect of evidence for an inner-sphere interaction of the catalytic metal ion with

the 3′-OH is our observation that increasing pH leads to a decreased K_M for Mg^{2+} in the primer

extension reaction. Our rationale for this observation is as follows: The pK_a of the Mg^{2+}

coordinated inner shell water molecules is 12.3,[50] which is 5.4 pH units lower than the pK_a of

unbound bulk water. Corrected for the statistical effect due to the fact that six waters are bound,

Mg^{2+} coordination lowers the pK_a of a bound water by 4.8 pH units. If Mg^{2+} coordination to the

primer 3′-OH had the same effect, the pK_a should decrease from ~ 12.3 to roughly 7.5 when

Mg^{2+} is bound. This would require a stable interaction throughout the reaction coordinate, so it is

likely this effect is more modest, and Mg^{2+}-3′-OH is predicted to have the pK_a of 9.1.[28] Even a

more modest effect should lead to substantial deprotonation when Mg^{2+} is bound, and the

resulting partial negative charge on O-3′ should therefore lead to stronger Mg^{2+} binding to the

reaction center because of the favorable electrostatic interaction with the positive charge on the

metal ion. Our observation of a lower K_M for Mg^{2+} at higher pH is consistent with this

expectation, and provides support for a direct interaction of the catalytic Mg^{2+} ion with the primer 3′-OH. However, the final proton acceptor in primer extension remains undefined, and could be bulk water, one of the non-bridging oxygens of the adjacent reactive phosphate, or even an ionized water bound to the Mg^{2+} ion itself; in addition, the 3′-OH proton could be transferred stepwise via the 2′-OH to a different final acceptor.

The observation of an improvement in Mg^{2+} binding affinity in the presence of a downstream helper additionally strongly supports an interaction with the reacting bridged dinucleotide and the metal ion (**Figures 6**, **6.2**). With the simultaneous interaction with the 3′-OH, a clear prebiotic reactive site is present, and one can imagine how prebiotic catalytic cofactors could aid in supporting this reaction geometry.

Despite the major mechanistic differences between nonenzymatic and enzyme-catalyzed primer extension, i.e. reaction of the primer with an imidazolium-bridged dinucleotide vs. reaction with a monomeric nucleoside triphosphate, our experiments suggest a common mechanistic role for the catalytic metal ion resulting from interactions with both the primer 3′-OH and the adjacent reactive phosphate. We know the existing interactions within modern polymerases often use aspartate for the orientation of Mg^{2+} in their active sites.[75] The synergy between small molecules and RNA catalysis would provide a means for early selective pressure in the prebiotic soup. Additionally, the relatively low intrinsic affinity of Mg^{2+} for the primer/template is beneficial to an extent, as too high of an affinity and these interactions would be difficult for processive primer extension, would possibly favor RNA hydrolysis over extension, and could lead to insoluble RNA aggregates.[27,76] The extremely weak binding of the catalytic metal ion to the reaction center makes it difficult to detect and define the relevant interactions by physical methods. However, it may be possible to monitor the deprotonation of

the primer 3′-OH in the presence of Mg^{2+} by NMR methods, and to define catalytically relevant interactions with the reactive phosphate by classical phosphorothioate/thiophilic metal rescue experiments.[77]

In the context of the prebiotic RNA world, the weak binding of the catalytic metal ion to the reaction center is a major difficulty. The need for high concentrations of divalent cations in order to achieve reasonable rates of RNA template copying creates many problems, including the hydrolysis of activated nucleotides, the degradation of RNA, and the destruction of fatty acid vesicles.[2] The use of citrate as a Mg^{2+} chelator mitigates these issues, but the availability of high concentrations of citrate does not seem to be prebiotically plausible. One potential solution to this conundrum would be the identification of a simple small molecule, oligonucleotide or peptide that could chelate Mg^{2+} bring it to the reaction center much as larger ribozymes and protein enzymes do,[2,75] enabling efficient primer extension at much lower metal ion concentrations. Any protocells containing RNAs capable of catalyzing the synthesis of such molecules would have had a strong selective advantage in low Mg^{2+} environments. Indeed, the ability to bind and make use of divalent cations may have been one of the strongest selective pressures acting on primordial protocells during the origin of life.

Bibliography

(1)   Gilbert, W. Origin of Life: The RNA World. *Nature* **1986**, *319* (6055), 618–618. https://doi.org/10.1038/319618a0.

(2)   Adamala, K.; Szostak, J. W. Nonenzymatic Template-Directed RNA Synthesis Inside Model Protocells. *Science* **2013**, *342* (6162), 1098–1100. https://doi.org/10.1126/science.1241888.

(3)   Lohrmann, R.; Bridson, P.; Orgel, L. Efficient Metal-Ion Catalyzed Template-Directed Oligonucleotide Synthesis. *Science* **1980**, *208* (4451), 1464–1465. https://doi.org/10.1126/science.6247762.

(4)   Prywes, N.; Blain, J. C.; Del Frate, F.; Szostak, J. W. Nonenzymatic Copying of RNA Templates Containing All Four Letters Is Catalyzed by Activated Oligonucleotides. *eLife* **2016**, *5*, e17756. https://doi.org/10.7554/eLife.17756.

(5)   Sawai, H. Catalysis of Internucleotide Bond Formation by Divalent Metal Ions. *Journal of the American Chemical Society* **1976**, *98* (22), 7037–7039. https://doi.org/10.1021/ja00438a050.

(6)   Sawai, H. Oligonucleotide Formation Catalyzed by Divalent Metal Ions. The Uniqueness of the Ribosyl System. *Journal of Molecular Evolution* **1988**, *27* (3), 181–186. https://doi.org/10.1007/BF02100072.

(7)   Sawai, H.; Orgel, L. E. Oligonucleotide Synthesis Catalyzed by the Zinc(2+) Ion. *Journal of the American Chemical Society* **1975**, *97* (12), 3532–3533. https://doi.org/10.1021/ja00845a050.

(8)   Li, L.; Prywes, N.; Tam, C. P.; O'Flaherty, D. K.; Lelyveld, V. S.; Izgu, E. C.; Pal, A.; Szostak, J. W. Enhanced Nonenzymatic RNA Copying with 2-Aminoimidazole Activated Nucleotides. *Journal of the American Chemical Society* **2017**, *139* (5), 1810–1813. https://doi.org/10.1021/jacs.6b13148.

(9)   Nomura, Y.; Yokobayashi, Y. Systematic Minimization of RNA Ligase Ribozyme through Large-Scale Design-Synthesis-Sequence Cycles. *Nucleic Acids Res* **2019**, *47* (17), 8950–8960. https://doi.org/10.1093/nar/gkz729.

(10)   Johnston, W. K.; Unrau, P. J.; Lawrence, M. S.; Glasner, M. E.; Bartel, D. P. RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension. *Science* **2001**, *292* (5520), 1319–1325. https://doi.org/10.1126/science.1060786.

(11)   Orgel, L. E.; Lohrmann, R. Prebiotic Chemistry and Nucleic Acid Replication. *Acc. Chem. Res.* **1974**, *7* (11), 368–377. https://doi.org/10.1021/ar50083a002.

(12)   Sleeper, H. L.; Lohrmann, R.; Orgel, L. E. Formation of the Imidazolides of Dinucleotides under Potentially Prebiotic Conditions. *J Mol Evol* **1978**, *11* (2), 87–93. https://doi.org/10.1007/BF01733884.

(13)   Inoue, T.; Orgel, L. E. Oligomerization of (Guanosine 5′-Phosphor)-2-Methylimidazolide on Poly(C): An RNA Polymerase Model. *Journal of Molecular Biology* **1982**, *162* (1), 201–217. https://doi.org/10.1016/0022-2836(82)90169-3.

(14)   Fahrenbach, A. C.; Giurgiu, C.; Tam, C. P.; Li, L.; Hongo, Y.; Aono, M.; Szostak, J. W. Common and Potentially Prebiotic Origin for Precursors of Nucleotide Synthesis and Activation. *Journal of the American Chemical Society* **2017**, *139* (26), 8780–8783. https://doi.org/10.1021/jacs.7b01562.

(15)   Walton, T.; Szostak, J. W. A Kinetic Model of Nonenzymatic RNA Polymerization by Cytidine-5′-Phosphoro-2-Aminoimidazolide. *Biochemistry* **2017**, *56* (43), 5739–5747. https://doi.org/10.1021/acs.biochem.7b00792.

(16)   Walton, T.; Szostak, J. W. A Highly Reactive Imidazolium-Bridged Dinucleotide Intermediate in Nonenzymatic RNA Primer Extension. *Journal of the American Chemical Society* **2016**, *138* (36), 11996–12002. https://doi.org/10.1021/jacs.6b07977.

(17)   Walton, T.; Pazienza, L.; Szostak, J. W. Template-Directed Catalysis of a Multistep Reaction Pathway for Nonenzymatic RNA Primer Extension. *Biochemistry* **2019**, *58* (6), 755–762. https://doi.org/10.1021/acs.biochem.8b01156.

(18)   Zhang, W.; Walton, T.; Li, L.; Szostak, J. W. Crystallographic Observation of Nonenzymatic RNA Primer Extension. 15.

(19)   Zhang, W.; Tam, C. P.; Zhou, L.; Oh, S. S.; Wang, J.; Szostak, J. W. Structural Rationale for the Enhanced Catalysis of Nonenzymatic RNA Primer Extension by a Downstream Oligonucleotide. *Journal of the American Chemical Society* **2018**, *140* (8), 2829–2840. https://doi.org/10.1021/jacs.7b11750.

(20)   Miyakawa, S.; Joshi, P. C.; Gaffey, M. J.; Gonzalez-Toril, E.; Hyland, C.; Ross, T.; Rybij, K.; Ferris, J. P. Studies in the Mineral and Salt-Catalyzed Formation of RNA Oligomers. *Origins of Life and Evolution of Biospheres* **2006**, *36* (4), 343–361. https://doi.org/10.1007/s11084-006-9009-6.

(21)   Breslow, R.; Huang, D. L. Effects of Metal Ions, Including Mg2+ and Lanthanides, on the Cleavage of Ribonucleotides and RNA Model Compounds. *Proceedings of the National Academy of Sciences* **1991**, *88* (10), 4080–4083. https://doi.org/10.1073/pnas.88.10.4080.

(22)   Kanavarioti, A.; Bernasconi, C. F.; Doodokyan, D. L.; Alberas, D. J. Magnesium Ion Catalyzed Phosphorus-Nitrogen Bond Hydrolysis in Imidazolide-Activated Nucleotides. Relevance to Template-Directed Synthesis of Polynucleotides. *Journal of the American Chemical Society* **1989**, *111* (18), 7247–7257. https://doi.org/10.1021/ja00200a053.

(23)   Rohatgi, R.; Bartel, D. P.; Szostak, J. W. Kinetic and Mechanistic Analysis of Nonenzymatic, Template-Directed Oligoribonucleotide Ligation. *Journal of the American Chemical Society* **1996**, *118* (14), 3332–3339. https://doi.org/10.1021/ja953712b.

(24) Sawai, H. Catalysis of Internucleotide Bond Formation by Divalent Metal Ions. *Journal of the American Chemical Society* **1976**, *98* (22), 7037–7039. https://doi.org/10.1021/ja00438a050.

(25) Jonathan Craig Blain. Non-enzymatic copying of nucleic acid templates - ProQuest https://www.proquest.com/docview/1502876441?pq-origsite=gscholar&fromopenview=true (accessed 2022 -01 -05).

(26) Walton, T.; DasGupta, S.; Duzdevich, D.; Oh, S. S.; Szostak, J. W. In Vitro Selection of Ribozyme Ligases That Use Prebiotically Plausible 2-Aminoimidazole–Activated Substrates. *Proc Natl Acad Sci USA* **2020**, *117* (11), 5741–5748. https://doi.org/10.1073/pnas.1914367117.

(27) Lin Jin, Aaron E. Engelhart, Weicheng Zhang, Katarzyna Adamala, and Jack W. Szostak. Catalysis of Template-Directed Nonenzymatic RNA Copying by Iron(II). *JACS 140* (44), 15016–15021. https://doi.org/10.1021/jacs.8b09617.

(28) Giurgiu, C.; Fang, Z.; Aitken, H. R. M.; Kim, S. C.; Pazienza, L.; Mittal, S.; Szostak, J. W. Structure–Activity Relationships in Nonenzymatic Template-Directed RNA Synthesis. *Angewandte Chemie* **2021**, *133* (42), 23107–23114. https://doi.org/10.1002/ange.202109714.

(29) Walton, T.; Pazienza, L.; Szostak, J. W. Template-Directed Catalysis of a Multistep Reaction Pathway for Nonenzymatic RNA Primer Extension. *Biochemistry* **2019**, *58* (6), 755–762. https://doi.org/10.1021/acs.biochem.8b01156.

(30) Kaiser, A.; Richert, C. Nucleotide-Based Copying of Nucleic Acid Sequences without Enzymes. *J. Org. Chem.* **2013**, *78* (3), 793–799. https://doi.org/10.1021/jo3025779.

(31)   Walton, T.; Szostak, J. W. A Kinetic Model of Nonenzymatic RNA Polymerization by Cytidine-5′-Phosphoro-2-Aminoimidazolide. *Biochemistry* **2017**, *56* (43), 5739–5747. https://doi.org/10.1021/acs.biochem.7b00792.

(32)   Bowman, J. C.; Lenz, T. K.; Hud, N. V.; Williams, L. D. Cations in Charge: Magnesium Ions in RNA Folding and Catalysis. *Current Opinion in Structural Biology* **2012**, *22* (3), 262–272. https://doi.org/10.1016/j.sbi.2012.04.006.

(33)   Batra, V. K.; Beard, W. A.; Shock, D. D.; Krahn, J. M.; Pedersen, L. C.; Wilson, S. H. Magnesium-Induced Assembly of a Complete DNA Polymerase Catalytic Complex. *Structure* **2006**, *14* (4), 757–766. https://doi.org/10.1016/j.str.2006.01.011.

(34)   Draper, D. E. A Guide to Ions and RNA Structure. *RNA* **2004**, *10* (3), 335–343. https://doi.org/10.1261/rna.5205404.

(35)   Curtis, E. A.; Bartel, D. P. The Hammerhead Cleavage Reaction in Monovalent Cations. *RNA* **2001**, *7* (4), 546–552. https://doi.org/10.1017/S1355838201002357.

(36)   Young, K. J.; Gill, F.; Grasby, J. A. Metal Ions Play a Passive Role in the Hairpin Ribozyme Catalysed Reaction. *Nucleic Acids Research* **1997**, *25* (19), 3760–3760. https://doi.org/10.1093/nar/25.19.3760.

(37)   Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D. Understanding Nucleic Acid–Ion Interactions. *Annual Review of Biochemistry* **2014**, *83* (1), 813–841. https://doi.org/10.1146/annurev-biochem-060409-092720.

(38)   Kim, S.; Cowan, J. A. Inert Cobalt Complexes as Mechanistic Probes of the Biochemistry of Magnesium Cofactors. Application to Topoisomerase I. *Inorganic Chemistry* **1992**, *31* (17), 3495–3496. https://doi.org/10.1021/ic00043a001.

(39)   Curtis, E. A.; Bartel, D. P. The Hammerhead Cleavage Reaction in Monovalent Cations. *RNA* **2001**, *7* (4), 546–552. https://doi.org/10.1017/S1355838201002357.

(40)   Fedor, M. J. The Role of Metal Ions in RNA Catalysis. *Current Opinion in Structural Biology* **2002**, *12* (3), 289–295. https://doi.org/10.1016/S0959-440X(02)00324-X.

(41)   Jou, R.; Cowan, J. A. Ribonuclease H Activation by Inert Transition-Metal Complexes. Mechanistic Probes for Metallocofactors: Insights on the Metallobiochemistry of Divalent Magnesium Ion. *Journal of the American Chemical Society* **1991**, *113* (17), 6685–6686. https://doi.org/10.1021/ja00017a056.

(42)   Kucharski, L. M.; Lubbe, W. J.; Maguire, M. E. Cation Hexaammines Are Selective and Potent Inhibitors of the CorA Magnesium Transport System. *Journal of Biological Chemistry* **2000**, *275* (22), 16767–16773. https://doi.org/10.1074/jbc.M001507200.

(43)   Zhang, W.; Tam, C. P.; Walton, T.; Fahrenbach, A. C.; Birrane, G.; Szostak, J. W. Insight into the Mechanism of Nonenzymatic RNA Primer Extension from the Structure of an RNA-GpppG Complex. *Proc Natl Acad Sci USA* **2017**, *114* (29), 7659–7664. https://doi.org/10.1073/pnas.1704006114.

(44)   Kanavarioti, A. Kinetics of the Hydrolysis of Guanosine 5′-Phospho-2-Methylimidazolide. *Origins Life Evol Biosphere* **1986**, *17* (1), 85–103. https://doi.org/10.1007/BF01809815.

(45)   Bowman, J. C.; Lenz, T. K.; Hud, N. V.; Williams, L. D. Cations in Charge: Magnesium Ions in RNA Folding and Catalysis. *Current Opinion in Structural Biology* **2012**, *22* (3), 262–272. https://doi.org/10.1016/j.sbi.2012.04.006.

(46)   Frederiksen, J. K.; Fong, R.; Piccirilli, J. A. Chapter 8. Metal Ions in RNA Catalysis. In *Nucleic Acid-Metal Ion Interactions*; Royal Society of Chemistry: Cambridge, 2008; pp 260–306. https://doi.org/10.1039/9781847558763-00260.

(47)   Mikulecky, P. J.; Feig, A. L. Heat Capacity Changes Associated with DNA Duplex

         Formation: Salt- and Sequence-Dependent Effects. *Biochemistry* **2006**, *45* (2), 604–616.

         https://doi.org/10.1021/bi0517178.

(48)   Giurgiu, C.; Li, L.; O'Flaherty, D. K.; Tam, C. P.; Szostak, J. W. A Mechanistic

         Explanation for the Regioselectivity of Nonenzymatic RNA Primer Extension. *J. Am.*

         Chem. Soc. **2017**, *139* (46), 16741–16747. https://doi.org/10.1021/jacs.7b08784.

(49)   Rohatgi, R.; Bartel, D. P.; Szostak, J. W. Kinetic and Mechanistic Analysis of

         Nonenzymatic, Template-Directed Oligoribonucleotide Ligation. *Journal of the American*

         Chemical Society **1996**, *118* (14), 3332–3339. https://doi.org/10.1021/ja953712b.

(50)   Frederiksen, J. K.; Fong, R.; Piccirilli, J. A. Chapter 8. Metal Ions in RNA Catalysis. In

         Nucleic Acid-Metal Ion Interactions; Royal Society of Chemistry: Cambridge, 2008; pp

         260–306. https://doi.org/10.1039/9781847558763-00260.

(51)   Breslow, R.; Huang, D. L. Effects of Metal Ions, Including Mg2+ and Lanthanides, on the

         Cleavage of Ribonucleotides and RNA Model Compounds. *Proceedings of the National*

         Academy of Sciences **1991**, *88* (10), 4080–4083. https://doi.org/10.1073/pnas.88.10.4080.

(52)   Cassano, A. G.; Anderson, V. E.; Harris, M. E. Evidence for Direct Attack by Hydroxide

         in Phosphodiester Hydrolysis. *J. Am. Chem. Soc.* **2002**, *124* (37), 10964–10965.

         https://doi.org/10.1021/ja020823j.

(53)   Hendry, P.; Sargeson, A. M. Metal Ion Promoted Phosphate Ester Hydrolysis.

         Intramolecular Attack of Coordinated Hydroxide Ion. *Journal of the American Chemical*

         Society **1989**, *111* (7), 2521–2527. https://doi.org/10.1021/ja00189a025.

(54)   Stern, M. K.; Bashkin, J. K.; Sall, E. D. Hydrolysis of RNA by Transition Metal
         Complexes. *Journal of the American Chemical Society* **1990**, *112* (13), 5357–5359.
         https://doi.org/10.1021/ja00169a057.

(55)   Herschlag, D.; Jencks, W. P. Catalysis of the Hydrolysis of Phosphorylated Pyridines by
         Mg(OH)+: A Possible Model for Enzymic Phosphoryl Transfer. *Biochemistry* **1990**, *29*
         (21), 5172–5179. https://doi.org/10.1021/bi00473a025.

(56)   McFail-Isom, L.; Shui, X.; Williams, L. D. Divalent Cations Stabilize Unstacked
         Conformations of DNA and RNA by Interacting with Base Π Systems. 7.

(57)   *Bioinorganic Chemistry*; Bertini, I., Ed.; University Science Books: Mill Valley, Calif,
         1994.

(58)   Schaffer, M.; Peng, G.; Spingler, B.; Schnabl, J.; Wang, M.; Olieric, V.; Sigel, R.;
         Schaffer, M. F.; Peng, G.; Spingler, B.; Schnabl, J.; Wang, M.; Olieric, V.; Sigel, R. K. O.
         The X-Ray Structures of Six Octameric RNA Duplexes in the Presence of Different Di-
         and Trivalent Cations. *International Journal of Molecular Sciences* **2016**, *17* (7), 988.
         https://doi.org/10.3390/ijms17070988.

(59)   Adamala, K.; Szostak, J. W. Nonenzymatic Template-Directed RNA Synthesis Inside
         Model Protocells. *Science* **2013**, *342* (6162), 1098–1100.
         https://doi.org/10.1126/science.1241888.

(60)   Norman, P. R.; Cornelius, R. D. Mechanism of Cobalt(III)-Promoted Hydrolysis of
         Triphosphate Ion. *Journal of the American Chemical Society* **1982**, *104* (9), 2356–2361.
         https://doi.org/10.1021/ja00373a004.

(61)   Calderone, D. M.; Mantilla, E. J.; Hicks, M.; Huchital, D. H.; Murphy, W. R.; Sheardy, R.
         D. Binding of Co(III) to a DNA Oligomer via Reaction of [Co(NH3)5(OH2)]3+ with

(5MedC-DG)4. *Biochemistry* **1995**, *34* (42), 13841–13846. https://doi.org/10.1021/bi00042a016.

(62)   Weeks, K. M.; Mauger, D. M. Exploring RNA Structural Codes with SHAPE Chemistry. *Acc. Chem. Res.* **2011**, *44* (12), 1280–1291. https://doi.org/10.1021/ar200051h.

(63)   Hanna, M.; Szostak, J. W. Suppression of Mutations in the Core of the Tetrahymena Ribozyme by Spermidine, Ethanol and by Substrate Stabilization. *Nucleic Acids Research* **1994**, *22* (24), 5326–5331. https://doi.org/10.1093/nar/22.24.5326.

(64)   Zhou, D.-M.; Zhang, L.-H.; Taira, K. Explanation by the Double-Metal-Ion Mechanism of Catalysis for the Differential Metal Ion Effects on the Cleavage Rates of 5'-Oxy and 5'-Thio Substrates by a Hammerhead Ribozyme. *Proceedings of the National Academy of Sciences* **1997**, *94* (26), 14343–14348. https://doi.org/10.1073/pnas.94.26.14343.

(65)   Takagi, Y.; Taira, K. Analyses of Kinetic Solvent Isotope Effects of a Hammerhead Ribozyme Reaction in NH4+ and Li+ Ions. *Nucleic Acids Symposium Series* **2002**, *2* (1), 273–274. https://doi.org/10.1093/nass/2.1.273.

(66)   Sawata, S.; Komiyama, M.; Taira, K. Kinetic Evidence Based on Solvent Isotope Effects for the Nonexistence of a Proton-Transfer Process in Reactions Catalyzed by a Hammerhead Ribozyme: Implication to the Double-Metal-Ion Mechanism of Catalysis. *Journal of the American Chemical Society* **1995**, *117* (8), 2357–2358. https://doi.org/10.1021/ja00113a028.

(67)   Pka-Compilation-Williams.Pdf.

(68)   Thaplyal, P.; Bevilacqua, P. C. Experimental Approaches for Measuring PKa's in RNA and DNA. *Methods Enzymol* **2014**, *549*, 189–219. https://doi.org/10.1016/B978-0-12-801122-5.00009-X.

(69)   Krężel, A.; Bal, W. A Formula for Correlating p K a Values Determined in D 2 O and H 2 O. *Journal of Inorganic Biochemistry* **2004**, *98* (1), 161–166. https://doi.org/10.1016/j.jinorgbio.2003.10.001.

(70)   Miyagi, M.; Wan, Q.; Ahmad, Md. F.; Gokulrangan, G.; Tomechko, S. E.; Bennett, B.; Dealwis, C. Histidine Hydrogen-Deuterium Exchange Mass Spectrometry for Probing the Microenvironment of Histidine Residues in Dihydrofolate Reductase. *PLoS One* **2011**, *6* (2), e17055. https://doi.org/10.1371/journal.pone.0017055.

(71)   Braasch, D. A.; Corey, D. R. Locked Nucleic Acid (LNA): ¢ne-Tuning the Recognition of DNA and RNA. 7.

(72)   Vester, B.; Wengel, J. LNA (Locked Nucleic Acid): High-Affinity Targeting of Complementary RNA and DNA. *Biochemistry* **2004**, *43* (42), 13233–13241. https://doi.org/10.1021/bi0485732.

(73)   Petersen, M.; Bondensgaard, K.; Wengel, J.; Jacobsen, J. P. Locked Nucleic Acid (LNA) Recognition of RNA: NMR Solution Structures of LNA:RNA Hybrids. *J. Am. Chem. Soc.* **2002**, *124* (21), 5974–5982. https://doi.org/10.1021/ja012288d.

(74)   Wengel, J.; Koshkin, A.; Singh, S. K.; Nielsen, P.; Meldgaard, M.; Rajwanshi, V. K.; Kumar, R.; Skouv, J.; Nielsen, C. B.; Jacobsen, J. P.; Jacobsen, N.; Olsen, C. E. Lna (Locked Nucleic Acid). *Nucleosides and Nucleotides* **1999**, *18* (6–7), 1365–1370. https://doi.org/10.1080/07328319908044718.

(75)   Harding, M. M. The Geometry of Metal–Ligand Interactions Relevant to Proteins. *Acta Cryst D, Acta Cryst Sect D, Acta Crystallogr D, Acta Crystallogr Sect D, Acta Crystallogr D Biol Crystallogr, Acta Crystallogr Sect D Biol Crystallogr* **1999**, *55* (8), 1432–1443. https://doi.org/10.1107/S0907444999007374.

(76)   Koculi, E.; Hyeon, C.; Thirumalai, D.; Woodson, S. A. Charge Density of Divalent Metal

        Cations Determines RNA Stability. *Journal of the American Chemical Society* **2007**, *129*

        (9), 2676–2682. https://doi.org/10.1021/ja068027r.

(77)   Frederiksen, J. K.; Piccirilli, J. A. Identification of Catalytic Metal Ion Ligands in

        Ribozymes. *Methods* **2009**, *49* (2), 148–166. https://doi.org/10.1016/j.ymeth.2009.07.005.

Chapter 3

Further elucidation of the mechanism of nonenzymatic primer extension through phosphorothioate substitution and metal rescue

Further elucidation of the mechanism of nonenzymatic primer extension through phosphorothioate substitution and metal rescue

Pazienza, L; Fang, Z; Zheng, W; Kim, LJ; Tam, CP; Kim, C; and JW Szostak.

[*LTP wrote the text, synthesized materials, and performed experiments; ZF, WZ, and LJK did crystallographic studies, CPT developed the synthesis route, CK did computational studies, and experimental planning was shared with LP, JWS, CPT, and ZF. This manuscript will be submitted in an extended form for peer review and open-access publication*]

Abstract

Nonenzymatic primer extension is believed to be an important step in progressing from RNA monomers to RNA oligomers of functional lengths in the RNA world hypothesis. However, the rates of nonenzymatic primer extension are often competitive with hydrolysis of reactants, and when maximally catalyzed by metal ions, incompatible with vesicles. Understanding the role of metal ions in primer extension remains a hurdle in the optimization of primer extension, although it has been previously established that inner-sphere contacts with metal ions are integral for catalysis. Here we utilize sulfur substitutions in 2-aminophosphoroimidazolides to better probe metal ion contacts during primer extension, and to probe the stereochemical requirements of metal ion catalysis. We further verify the interaction with the reacting bridged dinucleotide and the catalyzing metal ion. We find evidence that primer extension proceeds through a stereospecific pathway, with only one phosphorothioate reactant being reactive in the presence of oxophilic cations suggesting a particular catalytic geometry is required for metal ion mediated inner-sphere catalysis. The opposite diastereomer is active in the presence of thiophilic cations, which further supports this hypothesis. The constraints ascertained through these studies further aid our understanding of catalysis and mechanism in nonenzymatic

primer extension, and enable future development of chelators or co-catalysts to better enhance primer extension rates.

Introduction

On the early earth, a milieu of chemicals and elements would have been available to seed life. However, which elements and compounds and precisely how they would be combined to create life is unknown. One of the major theories for how life began, the RNA world, hypothesizes that life began with RNA acting as an informational and catalytic polymer.[1] An important consideration in this hypothesis is that RNA, a polyanionic species, would be intimately associated with cations.[2] Before RNA was able to replicate itself, RNA would need to have been assembled and propagated nonenzymatically through native chemical reactions, and metal ions are integral in catalyzing these reactions nonenzymatically.[3,4] Nonenzymatic primer extension is the mechanism for templated RNA polymerization off of a short primer oligomer.[5] We believe this would have proceeded using RNA activated with a different leaving group than pyrophosphate, instead with an imidazolium species in which the monomers form bridged dinucleotides, increasing fidelity and rates for this nonenzymatic reaction.[5,6,7]

The use of the bridged dinucleotide and its affects with metal ions have been previously studied, establishing the extent of catalysis (Chapter 2), that Mg^{2+} affects the partitioning of hydrolysis vs. extension for the bridged dinucleotide,[8] and that the catalytic metal ion associates through an inner-sphere contact with the 3′-OH. What is known thus far is only a fraction of the overall mechanistic picture, however, and while there is experimental evidence of an interaction of the catalytic cation with the reacting bridged dinucleotide (**Chapter 2, Figure 6, 6.2**), it is unclear if this is through an inner-sphere or outer-sphere contact. Understanding the inner-sphere vs. outer-sphere contacts of the catalytic cation helps narrow down chelation possibilities for

ribozymes and co-catalysts that may enhance reactivity of primer extension and diminish undesirable side effects of membrane leakiness and hydrolysis of the reacting bridged dinucleotide and subsequent oligomers. In order to determine the exact interactions with the bridged dinucleotide, we chemically modified the bridged dinucleotide to probe which contacts are necessary and if there are inner-sphere contacts.

The chemical probing in focus for this chapter has been thoroughly utilized in the field of biochemistry, and involves substituting an oxygen for a sulfur in the phosphorus of an RNA compound.[9] Our study is unique in doing this substitution in RNA phosphoroimidazolides instead of native triphosphates, monophosphates or phosphodiester linkages. Due to different elemental properties of polarizability and charge density, oxygen and sulfur interact preferentially with other atoms with matching hardness/softness. Oxygen is charge dense and hard, and interacts most strongly with other small, charge dense atoms, while sulfur with its d orbitals is more polarizable and diffuse, and able to interact with higher order transition metals more strongly.[10-12] In order for this atomic substitution to be informative, it needs to be involved near the chemical step of interest, but not perturb the reaction coordinate substantially.[9] The same ground to transition state pathway must be maintained.[10] We hypothesize the catalytic metal ion for primer extension has an inner-sphere interaction with only one of the reacting phosphate oxygens, stabilizing the geometry for the reaction, and stabilizing charge buildup on the phosphate.

The addition of the sulfur to the phosphorus makes the phosphorus chiral, and we are interested in the steric requirements for primer extension, if there are any, and if the phosphate stereochemistry is inverted or retained upon attack with the 3′-OH. Traditionally, phosphoester transfer is considered to go through either a dissociative (proceeding through a metaphosphate

intermediate) transition state, or associative mechanism (proceeding through a phosphorane intermediate), or a concerted S_N2 like mechanism ($S_N2(P)$) and phosphoesters are incredibly sensitive to electronic conditions of their substituents in terms of which mechanism proceeds.[13,14] Racemization of products is predicted with a purely dissociative, S_N1-like mechanism.[13] S_N2-like reactions at a phosphorus center occur much like S_N2 reactions at a carbon center, with inversion of configuration observed.[13] Previous computational and experimental studies of the hydrolysis of the phosphoroimidazolide monomer suggest that hydrolysis occurs with extensive P-N bond breaking in the transition state, but not to the extent of metaphosphate formation, and is therefore S_N2-like,[15] but this has not been verified experimentally for nonenzymatic primer extension. The difference of the nucleophile as the 3′-OH and the leaving group being not just the imidazole group, but a phosphoroimidazolide, may affect the mechanistic progression in nonenzymatic primer extension compared to the hydrolysis computational system. The recent determination that the 3′-OH is deprotonated before the rate-limiting step of primer extension must also be considered— a less dissociative transition state is predicted with increasing basicity of the nucleophile through Hammond effects.[16] The mechanism of phosphoester exchange has been heavily studied and debated, and further elucidation and validation of this in our system is important for fully understanding the reactivity of nonenzymatic primer extension.[17]

We propose that only one of the phosphate prochiral oxygens is important for catalysis, and by substituting oxygen with sulfur, a loss of catalysis will be observed with this stereoisomer with an oxophilic metal ion, and a restoration of reactivity will be observed with a thiophilic metal ion such as Cd^{2+} or Mn^{2+}. This effect would be expected only with an inner-sphere contact with the catalytic cation. Additionally, the different chemical nature of phosphorothioates is ultimately informative to our understanding of nonenzymatic primer extension. The presence of

the sulfur in the reacting bridged dinucleotide affects the electrophilicity of the phosphate, and the effect of this substitution on the overall rates of the system will allow us to further assess the dissociative or associative nature of the transition state of primer extension. Thio-substitutions have been shown to increase rates in phosphate ester reactions with dissociative transition states and decrease rates in reactions with associative transition states.[16] However, with phosphoroimidazolides, the P-N bond is weaker and the phosphorus is more electrophilic, making direct comparison of these previous studies less definitive.[18] We predict that the presence of a sulfur in a phosphoroimidazolide would stabilize the phosphorus, making it less electrophilic and decrease reactivity rates. If the reaction coordinate does not involve a metaphosphate, this stabilization would instead make the thiophosphoroimidazolide less reactive for nucleophilic attack. Additionally, we do not expect the presence or absence of metal ion to affect the associative/dissociative/concerted nature of the transition state.[16] We also would expect inversion of configuration at the phosphate if the reaction indeed proceeds through a concerted mechanism, which can be ascertained through structural determination of our compounds and their reaction products. By further understanding how nonenzymatic primer extension occurs, we can better determine compatible conditions and possible prebiotic catalysts that may enable more efficient primer extension.

Metal-rescue studies have been used extensively to study ribozyme and RNA reaction mechanisms.[9-11,19-25] One caveat to this technique is that phosphorothioate substitution can significantly change the reaction coordinate and structure of RNAs.[19,26] In our particular system, there is no significant tertiary structure, just an A-form helix between the primer and template, and the sulfur substitution is in the substrate, not these oligomers. However, the sulfur in the bridged dinucleotide may affect the stable conformations most frequently sampled, especially

due to possible interactions with the exocyclic amine in the imidazole hydrogen bonding to the phosphate oxygens. Additionally, there is likely some flexibility in compensating for loss of contacts in nonenzymatic primer extension, as the reaction site geometry is not strongly maintained by an enzymatic scaffold. The interaction with the 3′-OH is very important for catalysis, which will likely result in an observed diminishment of rate with thiophilic cations. P-S bonds are also longer than P-O bonds (1.9 vs 1.5 Å), which will likely loosen some of the geometric constraints of the metal ion coordination.[26] We will assess if this is significant enough to interfere with our mechanistic studies. It has been shown that the addition of a sulfur isn't sufficient to create a new metal ion contact,[19,27] so the substitution of S in the bridged dinucleotide is unlikely to recruit a metal ion not previously involved in the reaction coordinate.

Another consideration is the effect of thiophilic metal ions on RNA structure. Cd^{2+} is known to aggregate DNA due to interactions with the nucleobases, particularly with cytidine.[28] Cd^{2+} is still able to interact with oxygen, it just has a lower preference for oxygen compared to sulfur, which will allow the 3′-OH to still interact with it for primer extension.[28] Mg prefers O vs. S by a factor of 31,000,[9] so the partitioning of reactivity with the different diastereomers will be very telling mechanistically.

In this chapter, we will describe the synthesis and purification of diastereomerically pure thiophosphoroimidazolides, their use to study metal ion interactions in nonenzymatic primer extension, and their structural characterization using X-ray crystallography.[29] We find a stereospecific coordination of metal ions for the catalysis of nonenzymatic primer extension, and that the observed data is consistent with the reaction proceeding through $S_N2(P)$.

Results and Discussion
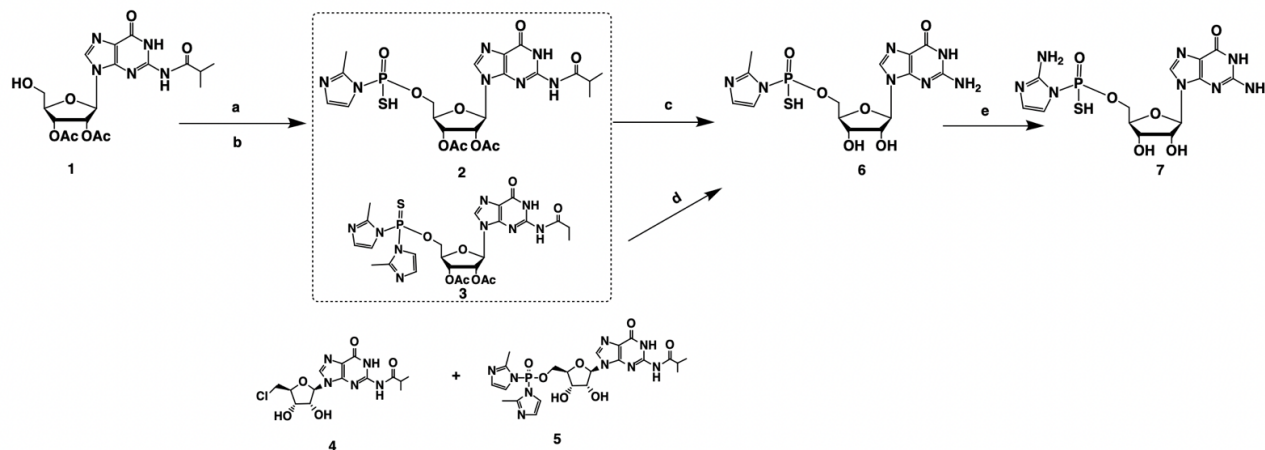
Scheme 1. Synthesis of guanosine thiophosphoroimidazolide (AIp$_s$G (**7**)). Reaction conditions: (a) Acetonitrile, PSCl$_3$, 2,6-lutidine; (b) methylimidazole 0 °C then 500 mM TEAB; (c) NH$_4$OH, MeOH, 65°C for 3 hours; (d) Hydrolysis for 1 day at 4°C and then (c); (e) 2-aminoimidazole in water, overnight.

<u>Synthesis of guanosine thiophosphoro-2aminoimidazolides (AIp$_s$G)</u>

Our synthetic strategy (**Scheme 1**) began with N2-isobutyryl-2′,3′-diacetylguanosine (**1**), which we then thiophosphorylate with PSCl$_3$ (**a**), and nucleophilically exchange the chlorides with methylimidazole to form the labeled guanosine-5′-thiophosphoromethylimidazolide (**2**). We subsequently deprotect the ribose under high temperature basic conditions (**c**), and then exchange the methylimidazole with 2-aminoimidazole in water (**e**) (See **Appendix C, Tables C.1-C.3, Figures C.1-C.12** for characterization of compounds). The subsequent separation of the diastereomers in the racemic product took considerable optimization and only the monomers of these species are separable by chromatography- the bridged dinucleotides are not. The bridged dinucleotides, which are established as the reactive species for nonenzymatic primer extension, were then selectively synthesized as a mixed thiophosphoro/phosphoro-imidazolium-bridged dinucleotide with different nucleobases to selectively localize the sulfur on the phosphorus to the

132

+1 position of the primer (Gp$_s$*pC- **Scheme 2**). This was done through a stereo-retentive reaction

as verified through NMR (**Appendix C, Figures C.13-C.24, Tables C.4, C.5**, no change in

NMR signal at phosphorus), in which the thiophosphoro-2-aminoimidazolide monomer acts as

the nucleophile and reacts to form the bridged dinucleotide with an OAt-activated cytidine

monomer, which can only react as an electrophile. The maintenance of the stereochemistry at the

phosphorus at this step is critical, for significant racemization would cause subsequent

purification difficulties.
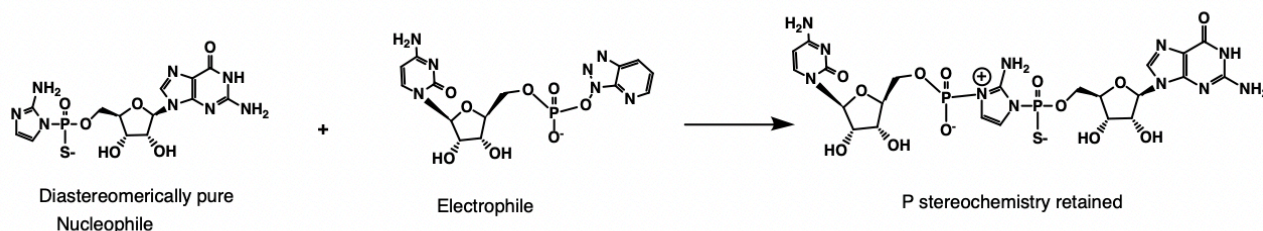
**Scheme 2**. Mixed thiophosphoro/phosphoro-2-aminoimidazolium-bridged dinucleotide

synthesis.
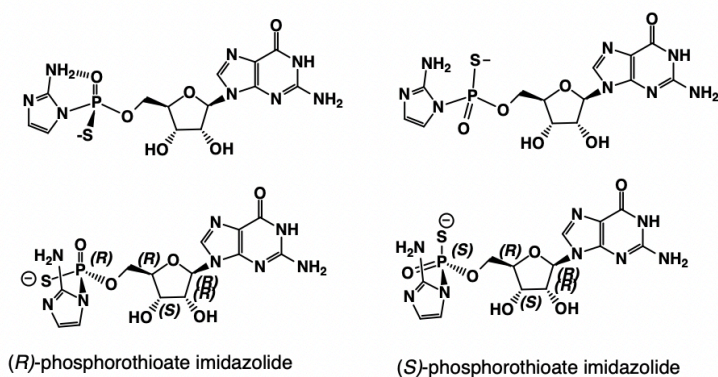
Scheme 3. *(R)*- and *(S)*- thiophosphoro-2-aminoimidazolides depicted to show potential for

intramolecular hydrogen bonding with exocyclic amine of 2-aminoimidazole. Hydrogen bonding

expected only in *(R)*-diastereomer.

Computational predictions of stable conformations for phosphorothioate dinucleotides

The stereochemistry of the *(R)*- and (*S*)- diastereomers of the thiophosphoro-2-aminoimidazolide are shown in **Scheme 3** (the formation of the bridged dinucleotide does not affect the absolute stereochemistry of the mixed bridged dinucleotide). The differing diastereomers are predicted to have different ground state energies due to the variation in possibilities for intramolecular hydrogen bonding as predicted chemically (**Scheme 3**) and through DFT simulations with and without dispersion corrections of Gp_s*pG (**Figure 1.1 and 1.2**). We do not anticipate the nature of the nucleobases to substantially affect the energetic differences observed in the molecules as calculated, although naturally there is less pi stacking between GC and GG, since the comparisons computationally are within GG, these effects should cancel out, and while the absolute energies may be incorrect, the trends will remain. The possibility for two intramolecular hydrogen bonds confers greater stability to the *(R)*-diastereomer than the *(S)*- diastereomer, due to the differing abilities to hydrogen bond with N-H between oxygen and sulfur. However, this effect is not highly significant, only contributing to a 1.39 or 1.17 kcal/mol benefit to the *(R)*- diastereomer at room temperature. We do not expect the majority of rate difference between the two diastereomers to be due to the difference in stability of the native conformation of the bridged dinucleotides, although there will likely be a small rate difference from this effect. Thus, if a substantial rate difference is observed, it will be due to the difference in reactivity as a result of thio-rescue. These simulations neglect to measure how these conformational differences are impacted by metal ions, although we expect the main kinetic difference with the metal ion interactions to be due to conformational organization of the reactive site, possible electrophilic activation of the phosphorus, and nucleophilic activation of the 3′-OH.
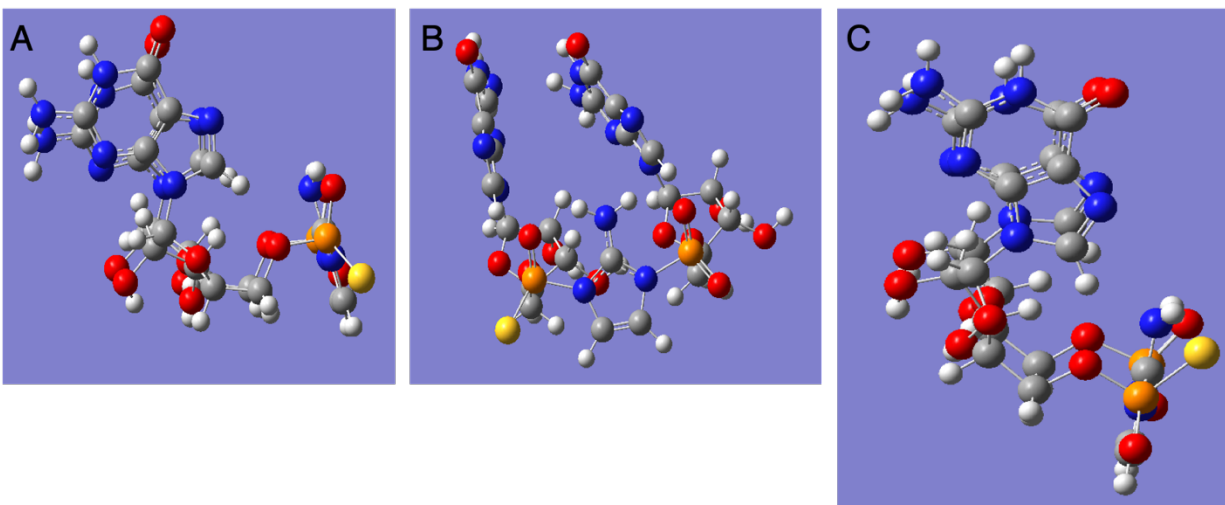
Figure 1.1. Computational predictions of stable conformations for mixed Gp$_s$*pG without dispersion corrections. See Appendix C for conditions of simulation (A) O-NH Energy=-3664.289723 J, ground state, *(R)*-diastereomer (B) alternate view of *(R)*-diastereomer (C) S-NH E= -3664.287508 J (+1.39 kcal/mol above ground state), *(S)*- diastereomer
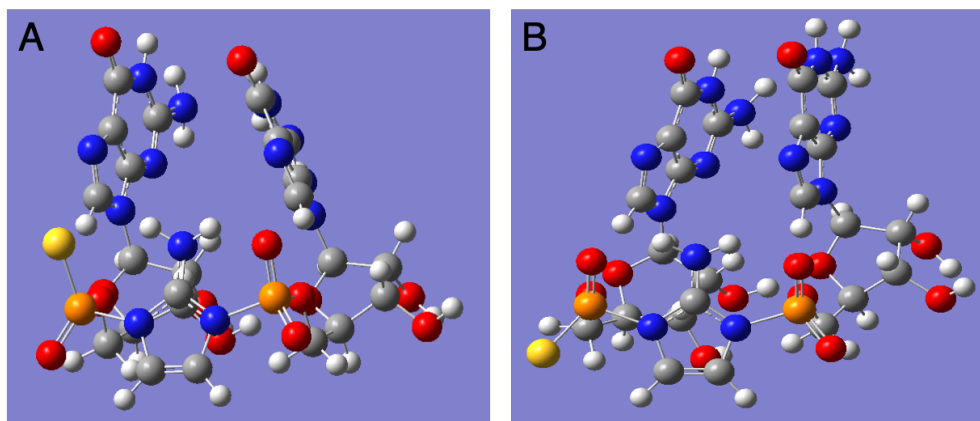
Figure 1.2. Computational predictions of stable conformations for mixed Gp$_s$*pG with dispersion corrections (A) S-NH E= -3664.506886 (1.17 kcal/mol above GS), *(S)*- diastereomer (B) O-NH E= -3664.508749 J, ground-state, *(R)*-diastereomer
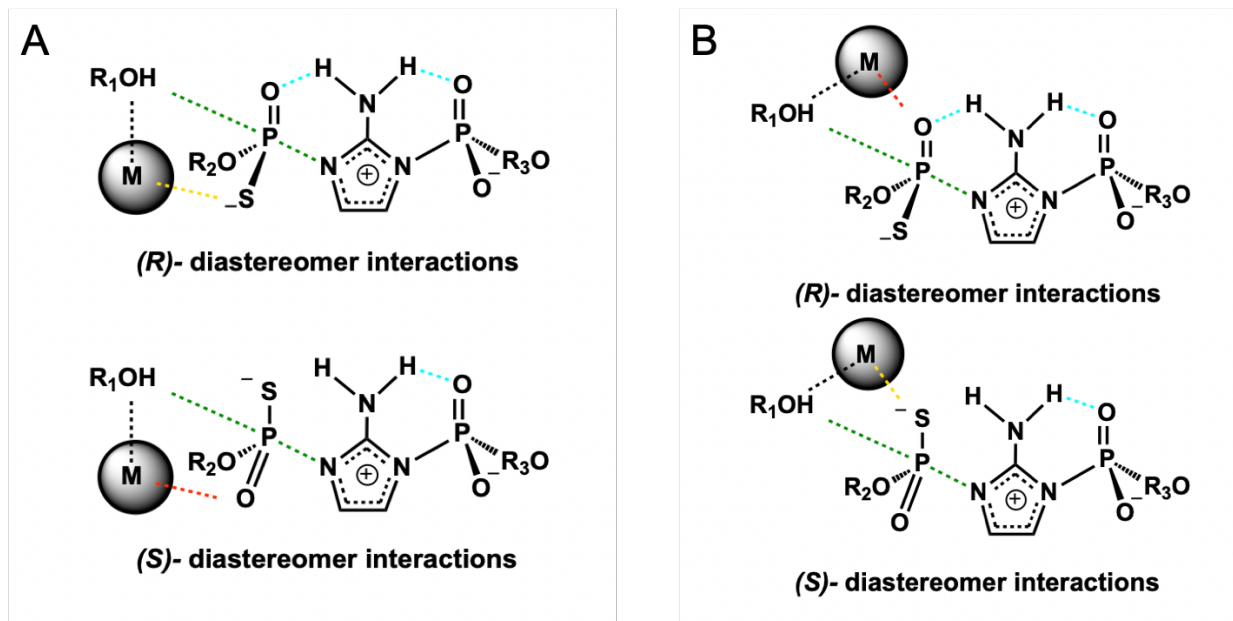
Scheme 4. Predicted interactions of the mixed thiophosphoro/phosphoro-2-aminoimidazolium-bridged dinucleotide diastereomers and thio-reactivity. R_1= primer ribose, R_2= guanosine nucleoside, R_3= cytidine nucleoside. The green dashed lines indicate the bonds forming and breaking during the transition state of nonenzymatic primer extension, the blue dashed lines indicate hydrogen bonding, and the dashed lines coming out of the metal ion sphere indicate possible inner-sphere contacts. The yellow dashed lines indicate predicted thio-reactivity, and the red dashed lines indicate predicted oxophilic reactivity. A and B depict alternate reactivity predictions depending on metal binding mode. (a) equatorial binding (b) axial binding.

Given the stable conformations predicted computationally, we expect the *(R)*-diastereomer to have more native intramolecular hydrogen bonds stabilizing it than the *(S)*-diastereomer. Additionally, given the expected conformation, we hypothesize it would be more sterically favorable for the catalytic metal ion to interact with the position coming out of the page (equatorial) in **Scheme 2** than the position in line with the exophilic amine (axial) (**Scheme 4A**). Therefore, we predict that the *(R)*- diastereomer would be reactive with the thiophilic metals, and the *(S)*-diastereomer with the oxophilic metal ions. However, if the metal ion interacts with the

136

primer and the ordered oxygen interacting with the exophilic amine, we expect the opposite reactivity (**Scheme 4B**). Previous work on phosphorothioate oligonucleotide purification suggests that the *(R)*- diastereomer tends to elute first,[30] which assuming the same trends hold with thiophosphoryl-2-aminoimidazolides would suggest diastereomer 1 (D1) is the *(R)*-diastereomer. Structurally, these compounds both contain charged sulfurs and P(V) phosphorus, with the major difference in connectivity in the fourth position of the phosphorus being the nitrogen of an imidazole instead of the oxygen of another nucleotide ribose. However, due to these chemical differences, we additionally were interested in verifying this structural assignment through crystallographic methods.

Making enough purified bridged-dinucleotide for dozens of kinetics experiments was limiting for our studies, so for our magnesium-binding experiments we instead formed the bridged-dinucleotide *in situ* on template with the AIp_sG diastereomers reacting to form Gp_s*pC in the reaction. The formation of the bridged-dinucleotide under these conditions is fairly rapid (**Figure A.13-A.15**) and unaffected by the phosphorothioate stereochemistry. We observed that AIp_sG diastereomer 2 was more innately reactive with Mg^{2+} than AIp_sG diastereomer 1 (**Figure 2.0**), which would match expectations for the aforementioned stereochemical assignments of *(S)*- and *(R)*- respectively in **Scheme 4A.** Diastereomer 2 has a higher reaction rate and binds the metal ion more tightly, suggesting that the phosphorus oxygen is in the correct position to have an inner-sphere contact with diastereomer 2 with an oxophilic metal ion.
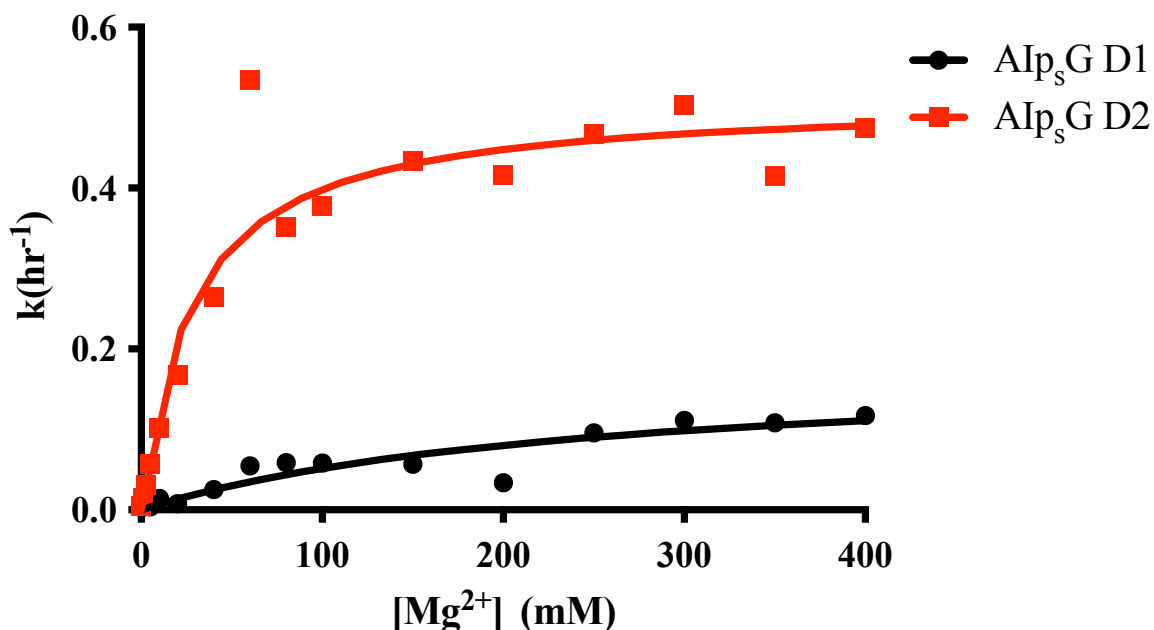
Figure 2.0. Effect of Mg^{2+} on primer extension with different phosphorothioate diastereomers. Primer extension reaction performed with *in situ* formation of Gps*pC through 20 mM AIp_SG and 20 mM CpOAt at pH=8. Diastereomer 1 (D1, black): $V_{max}= 0.181 \pm 0.05$ hr^{-1}, $K_M= 250 \pm 132.7$ mM. Diastereomer 2 (D2, red): $V_{max}= 0.51 \pm 0.04$ hr^{-1}, $K_M= 28 \pm 8.8$ mM. Error is $\pm 95\%$ CI.

Despite the fact the bridged dinucleotide should not form differently in solution depending on the starting diastereomer, we additionally wanted to study primer extension with an authentic standard of each Gp_s*pC diastereomer (**Table A.4, Figures A.16-A.19, A.22**). We selectively ensured the phosphorothioate was the reacting phosphate of the bridged dinucleotide through the sequence setup in **Figure 3A**, and then performed the reaction in absence of metal ions, and in presence of oxophilic metal ions (Mg^{2+}, Mn^{2+}) and thiophilic metal ions (Cd^{2+}). We additionally measure the rates in these conditions for the purely phosphate Gp*pC as a control. We observe that both phosphorothioate species are natively ten times slower than the oxygen system in absence of metal ions. This decrease in rate is as predicted earlier. The results from the

in-situ experiment are corroborated in **Figure 3B**, as D1 remains less active than D2 in presence of Mg^{2+}. While Mn^{2+} is sometimes used for metal rescue experiments as a thiophilic ion, we observe in our system that it is still sufficiently oxophilic so as to interact with D2, but D1 also increases in rate and is in the same order of magnitude of rate as D2. Mn^{2+} interacts with both O and S, so these results are not surprising, especially in a system where geometry for the reaction site is not carefully constrained by a chemical or enzymatic scaffold.[28,31,32] In the presence of Cd^{2+}, a much softer, more thiophilic ion, we observe a metal rescue, and D1 increases in rate compared to D2 by an order of magnitude. We calculated the thio-effect subsequently, which accounts for the relative rate due to the substitution of the oxygen for sulfur in addition to the metal ion swap. We observe D1 has a thio effect of 582, compared to 5 for D2, a case of substantial rescue and favorability for D1 over D2. This supports our geometric hypothesis of the catalytic ion interacting with the equatorial position of the phosphate, which is not hydrogen bonding simultaneously to the exocyclic amine of the imidazole through an inner-sphere contact of the metal ion. In conjunction with the inner-sphere interaction with the 3′-OH, the cation is ordering the active site for effective primer extension, in addition to the electronic activation of the nucleophile and electrophile.

A

B

M^{2+}	Diastereomer 1 (hr^{-1})	Diastereomer 2 (hr^{-1})	Gp*pC (hr^{-1})
No M^{2+}	0.0098 ± 0.004	0.0069 ± 0.001	0.025 ± 0.002
Mg^{2+}	0.013 ± 0.002	0.17 ± 0.006	2.5 ± 0.2
Mn^{2+}	0.23 ± 0.04	0.23 ± 0.04	10.3 ± 0.3
Cd^{2+}	1.1 ± 0.14	0.13 ± 0.02	0.36 ± 0.07

C

Thio effect	Diastereomer 1	Diastereomer 2
$\dfrac{\frac{k_{Cd}}{k_{Mg}}^{S}}{\frac{k_{Cd}}{k_{Mg}}^{O}}$	582	5
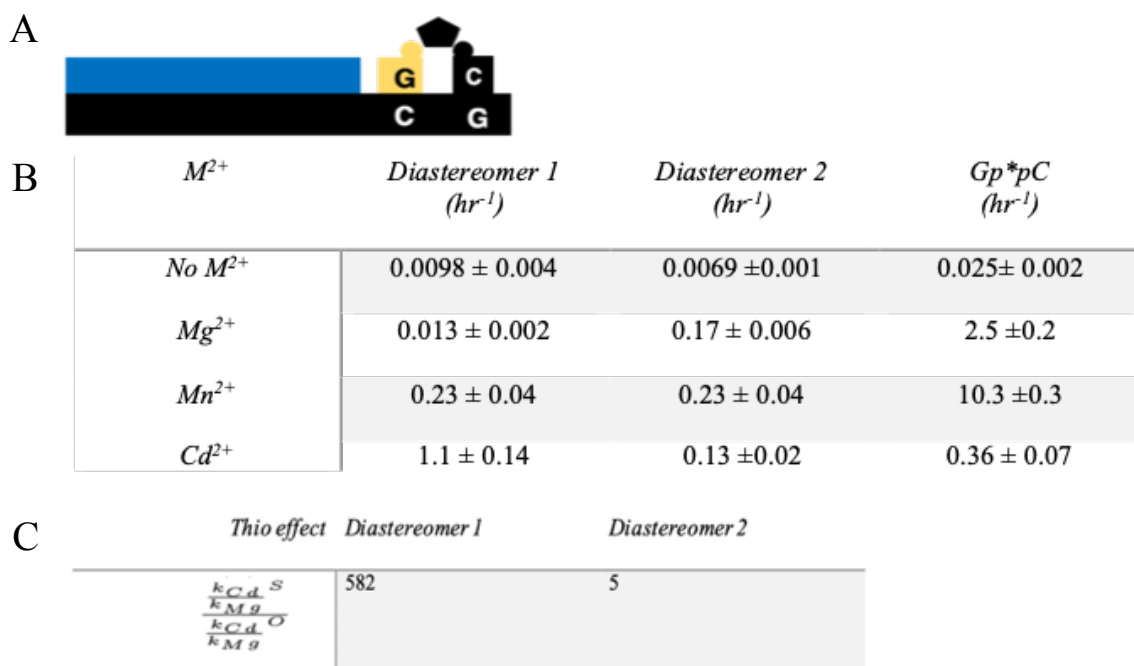
Figure 3.0. Metal rescue of primer extension with phosphorothioate analogs. (a) Sequence setup for primer extension of phosphorothioate dimer. Blue=template, black=primer and regular phosphate connections. Yellow=sulfur substituted nucleotide connectivity. (b) Kinetic results of primer extension with 1 μM P, 1.5 μM T, 20 mM intermediate, 100 mM Tris pH=7.5, 50 mM $M^{2+}Cl_2$. Gp*pC is the control, and is the bridged dinucleotide for regular oxophilic extension. Error is ± standard error (c) Thio-effect calculations for each diastereomer.

Structural determination of stereochemistry for phosphorothioate diastereomers

The observation and assignment of the activated monomer and bridged dinucleotide stereochemistry, and subsequent product structures, are critical for not only understanding the appropriate metal ion binding geometry between the primer 3′-OH and the phosphate oxygen of the reacting bridged dinucleotide, but also in verifying the nature of the mechanism of nonenzymatic primer extension (associative or dissociative). We obtained the product structures from reactions with both diastereomers of 2MeImp$_s$G (**Figure 5.1, 5.2**), 2AmImp$_s$G (**Figure 5.3, 5.4**) and D1 of Gp$_s$*pA (**Figure 5.5**). The Gp$_s$*pA dinucleotide system was chosen due to its

complementary hairpin templates having existing crystallization protocols, and issues with crystallization with the CG hairpin template. Surprisingly, regardless of starting reactant stereochemistry (>85% pure diastereomer monomer), the same product is observed in all 5 structures. This is less surprising in the case of the activated monomer reactions, as the bridged dinucleotide needs to form *in situ* in order to react. In this case, there will always be one side of the bridged dinucleotide of one stereochemistry, and the opposite stereochemistry on the other side. It appears that regardless of the conditions, the same stereoisomer is the one reacting in presence of Mg^{2+}, which is what is present in the product structure crystallization conditions. Cd^{2+} was attempted for crystallization but has not yet produced successful product structures as a result of gel and precipitate formation during the primer extension reaction. We would anticipate the opposite stereochemistry product to be observed in the presence of Cd^{2+} conditions. This effect is independent of the imidazole leaving group utilized in the complex. Additionally, when using a Gp_s*pA, a dimer preformed and purified for the reacting complex, we observe the same product structure. However, the primary impurity observed in Gp_s*pA D1 samples is the activated monomer (D1), which on the time scale of crystallization (but not for primer extension kinetics) would lead to racemization (**Figures C.13-C.15, C.20- C.22, C.24**).

We need to consider several hypotheses for the cause of the same stereochemistry in the product structures of all collected crystal structures. If the mechanism was dissociative, we would expect that the products would be racemized, not solely one stereochemical product— in fact we would not anticipate to see a notable electron density difference in the positions on the phosphate, and the bond length would be intermediate between P-O and P-S. An $S_N2(P)$ reaction, which primer extension is expected to undergo as predicted computationally,[18] would produce products of the opposite stereochemistry, but should not have a preference innately for the

141

starting reactant stereochemistry if the metal ion is not important for catalysis. We propose that in the presence of Mg^{2+}, only one diastereomer is reactive at an appreciable rate, and the same is true for the opposite diastereomer with Cd^{2+}. Small impurities and racemization over time would lead to the buildup of the reactive species, regardless of the starting material. Given the single stereoisomeric product, we propose that primer extension with the thiophosphoroimidazolides proceeds through a concerted $S_N2(P)$ mechanism, with a trigonal bipyramidal transition state with extensive bond breakage for the leaving group. Given that the product observed in the presence of Mg^{2+} is the *(R)*-diastereomer, and inversion of configuration occurs in the $S_N2(P)$ mechanism, the reacting diastereomer is predicted to be the (*S*)-diastereomer (D2), which matches our geometric predictions in **Scheme 4A** and our experimental observations in **Figures 2 and 3** for the oxophilic diastereomer.

Table 1. Parameters for X-ray data collection and structural refinement

Entry	1	2	3	4	5
PDB code	7U87	7U88	7U89	7U8A	7U8B
RNA duplex per asymmetric unit	1	1	1	1	1
Resolution range (Å)	41.67-1.70	42.95-2.14	28.56-1.65	41.42-2.10	34.16-1.73
Number of reflections	8558	4411	9696	5254	8832
R_{work} (%)	18.8	21.4	21.6	19.9	19.4
R_{free} (%)	24.0	29.2	28.2	28.2	24.8
Bond length R.M.S. (Å)	0.033	0.018	0.021	0.038	0.024
Bond angle R.M.S. (°)	3.26	2.60	2.98	2.71	2.96
Average B-factors ($Å^2$)	15.89	22.40	15.20	23.63	22.65
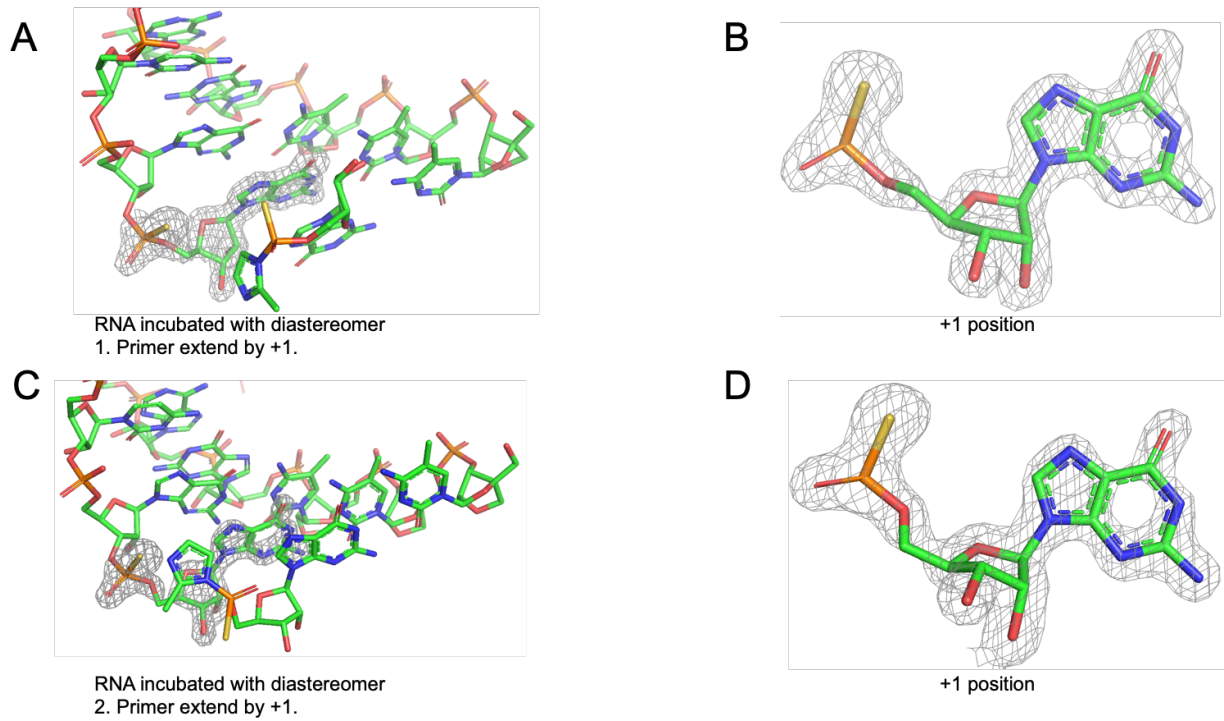
Figure 5.1. Product structures from reactions with 2MeImp$_s$G, 14mer (a) +1 extension with MeImp$_S$G D1 in the 14mer self-complementary primer complex (b) Density map and model of +1 position from (a). (c) 1 extension with MeImp$_S$G D2. (d) Density map and model of +1 position from (c).

A RNA incubated with diastereomer 1. Primer extend by +1 and +2.

B +1 position

C +2 position

D RNA incubated with diastereomer 2. Primer extend by +1 and +2.

E +1 position
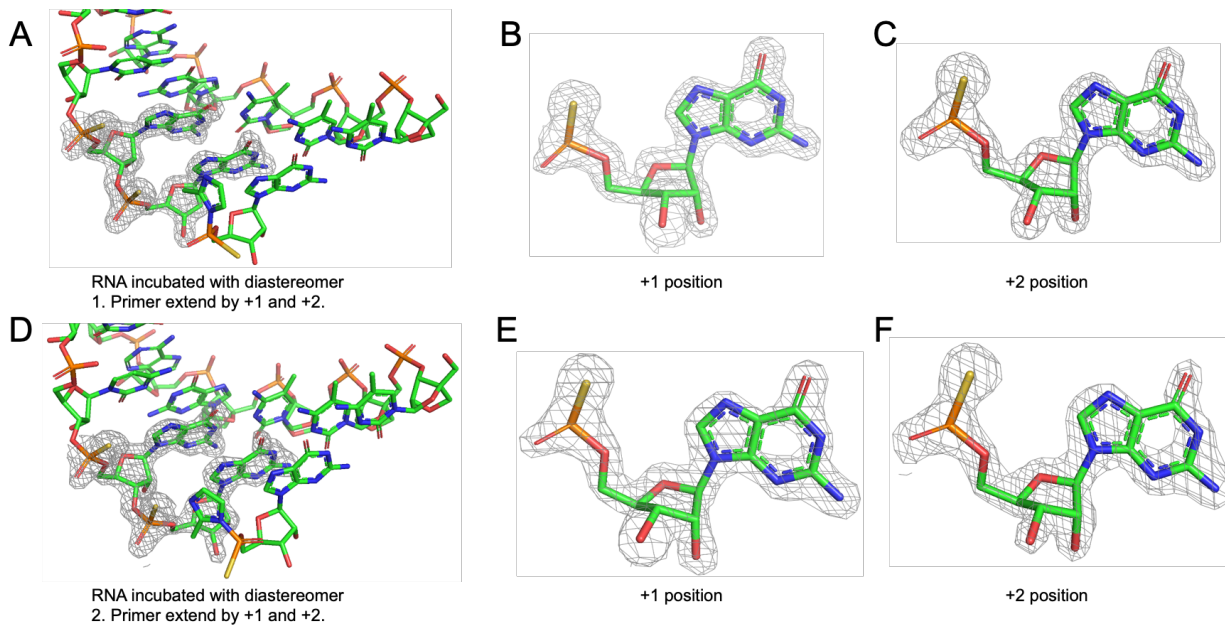
F +2 position

Figure 5.2. Product structures from reactions with 2MeImp$_s$G, 13mer (a) +1 extension with MeImp$_s$G D1 in the 13mer self-complementary primer complex (b) Density map and model of +1 position from (a). (c) +2 extension with MeImp$_s$G D1 in the 13mer self-complementary primer complex (d) +1 extension with MeImp$_s$G D2. (e) Density map and model of +1 position from (d). (f) +2 extension with MeImp$_s$G D2 in the 13mer self-complementary primer complex

A

RNA incubated with diastereomer
1. Primer extend by +1.

B

+1 position

C

RNA incubated with diastereomer
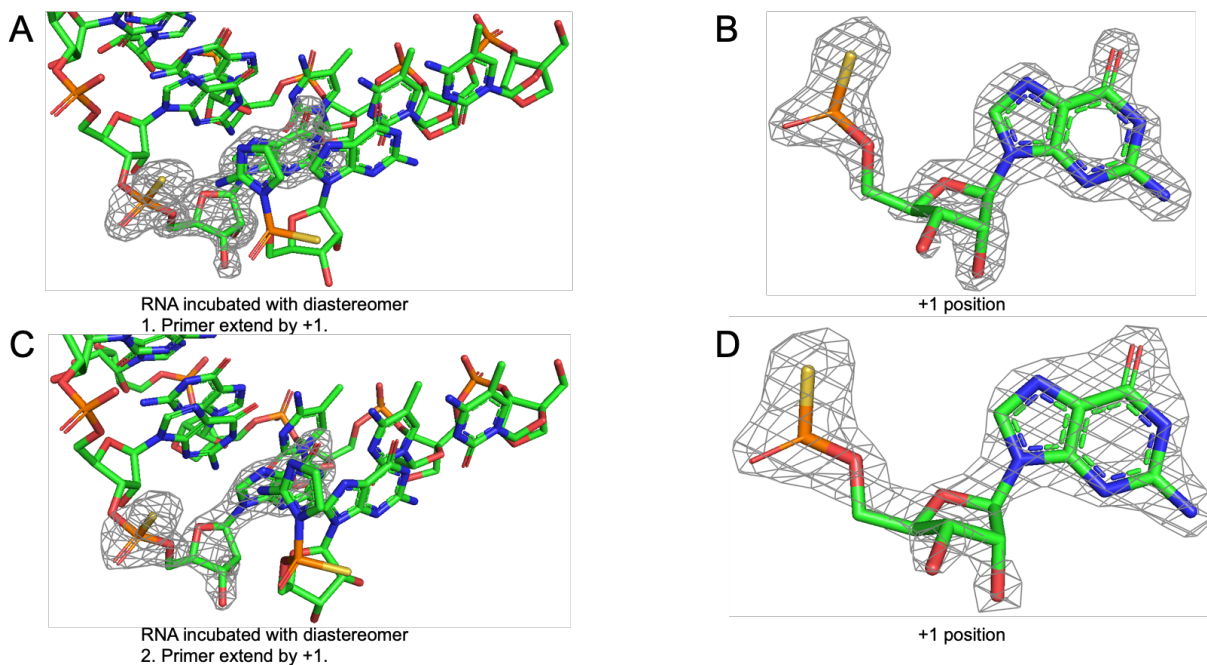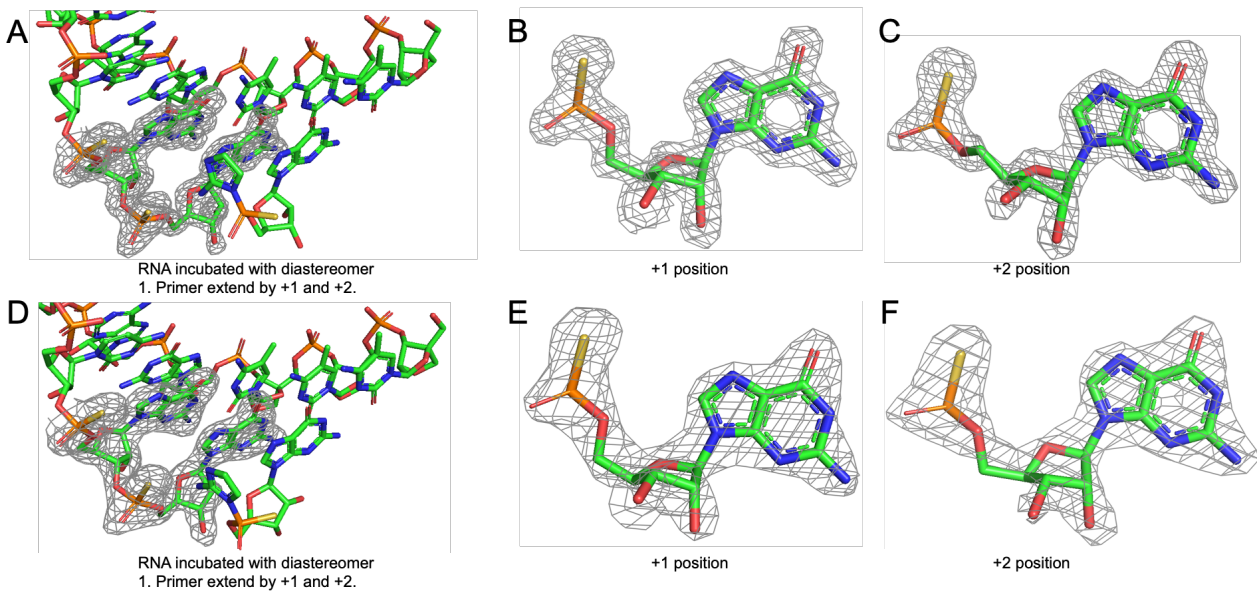2. Primer extend by +1.

D

+1 position

Figure 5.3. Product structures from reactions with 2AmImp$_s$G, 14mer (a) +1 extension with 2AIp$_S$G D1 in the 14mer self-complementary primer complex, PDB ID 7U89 (b) Density map and model of +1 position from (a). (c) 1 extension with 2AIp$_S$G D2, PDB 7U8A. (d) Density map and model of +1 position from (c).
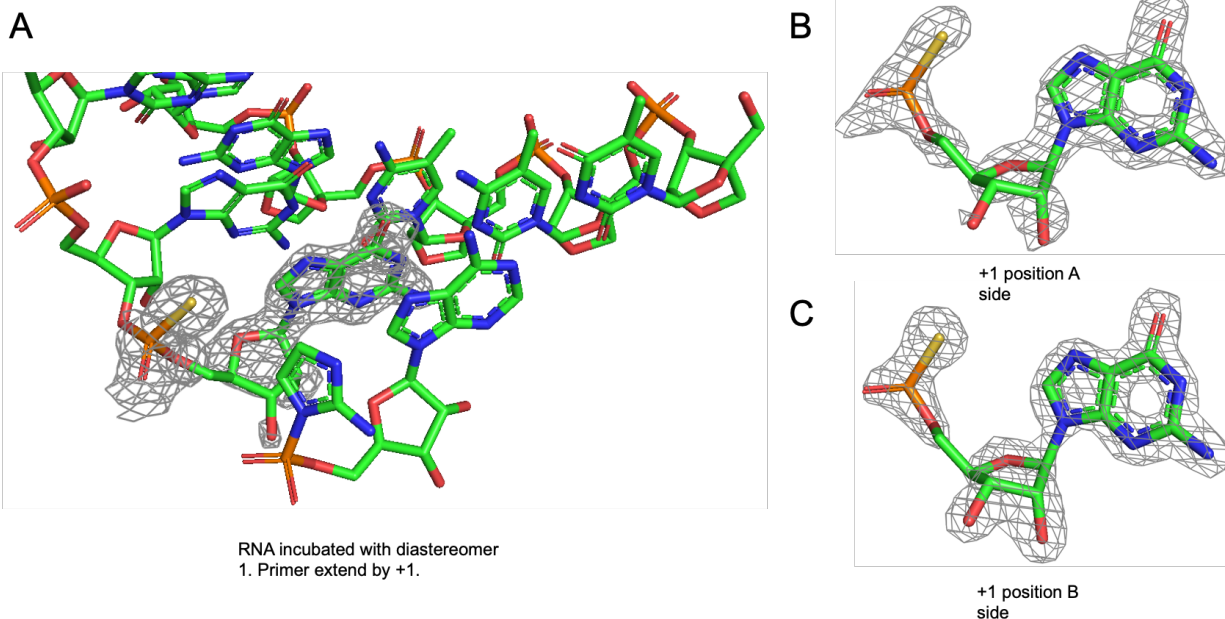
Figure 5.4. Product structures from reactions with 2AIp$_s$G, 13mer (a) +1 extension with 2AIp$_s$G D1 in the 13mer self-complementary primer complex, PDB ID 7U87 (b) Density map and model of +1 position from (a). (c) +2 extension with 2AIp$_s$G D1 in the 13mer self-complementary primer complex (d) +1 extension with 2AIp$_s$G D2, PDB ID 7U88(e) Density map and model of +1 position from (d). (f) +2 extension with 2AIp$_s$G D2 in the 13mer self-complementary primer complex

Figure 5.5. Product structures from reactions with Gp_s*pA D1, 14mer (a) +1 extension with Gp_s*pA D1 in the 14mer self-complementary primer complex PDB ID 7U8B (b) Density map and model of +1 position from (a), one side of the self-complementary complex (d) Density map and model of +1 position from (a), opposite side of self-complementary complex.

Primer extension occurs through an $S_N2(P)$ reaction as supported by the reactivity of thiophosphoroimidazolides as phosphoroimidazolide analogs. Only one stereoisomeric product is observed regardless of starting materials, but given the difficulty of separating the diastereomers of the activated monomers and bridged dinucleotides from one another, the presence of contaminating reactive diastereomer serves to produce product in all conditions. The mechanistic predictions are consistent with the metal binding modes posed in Scheme 4A, and with previously published computational studies.[18] With our understanding of the mechanism of primer extension expanding, we are able to better design scaffolds that favor the reactive geometry of metal ion binding and improve the activity of primer extension and its compatibility with protocellular membranes.

Bibliography

(1)   Gilbert, W. Origin of Life: The RNA World. *Nature* **1986**, *319* (6055), 618–618. https://doi.org/10.1038/319618a0.

(2)   Bowman, J. C.; Lenz, T. K.; Hud, N. V.; Williams, L. D. Cations in Charge: Magnesium Ions in RNA Folding and Catalysis. *Current Opinion in Structural Biology* **2012**, *22* (3), 262–272. https://doi.org/10.1016/j.sbi.2012.04.006.

(3)   Sawai, H. Oligonucleotide Formation Catalyzed by Divalent Metal Ions. The Uniqueness of the Ribosyl System. *Journal of Molecular Evolution* **1988**, *27* (3), 181–186. https://doi.org/10.1007/BF02100072.

(4)   Jonathan Craig Blain. Non-enzymatic copying of nucleic acid templates - ProQuest https://www.proquest.com/docview/1502876441?pq-origsite=gscholar&fromopenview=true (accessed 2022 -01 -05).

(5)   Weimann, B. J.; Lohrmann, R.; Orgel, L. E.; Schneider-Bernloehr, H.; Sulston, J. E. Template-Directed Synthesis with Adenosine-5'-Phosphorimidazolide. *Science* **1968**, *161* (3839), 387–387. https://doi.org/10.1126/science.161.3839.387.

(6)   Walton, T.; Szostak, J. W. A Highly Reactive Imidazolium-Bridged Dinucleotide Intermediate in Nonenzymatic RNA Primer Extension. *Journal of the American Chemical Society* **2016**, *138* (36), 11996–12002. https://doi.org/10.1021/jacs.6b07977.

(7)   Duzdevich, D.; Carr, C. E.; Ding, D.; Zhang, S. J.; Walton, T. S.; Szostak, J. W. Competition between Bridged Dinucleotides and Activated Mononucleotides Determines the Error Frequency of Nonenzymatic RNA Primer Extension. 11.

(8)   Walton, T.; Pazienza, L.; Szostak, J. W. Template-Directed Catalysis of a Multistep

Reaction Pathway for Nonenzymatic RNA Primer Extension. *Biochemistry* **2019**, *58* (6), 755–

762. https://doi.org/10.1021/acs.biochem.8b01156.

(9)   Forconi, M.; Herschlag, D. Use of Phosphorothioates to Identify Sites of Metal-Ion

Binding in RNA. In *Methods in Enzymology*; Elsevier, 2009; Vol. 468, pp 311–333.

https://doi.org/10.1016/S0076-6879(09)68015-0.

(10)   Frederiksen, J. K.; Piccirilli, J. A. Identification of Catalytic Metal Ion Ligands in

Ribozymes. *Methods* **2009**, *49* (2), 148–166. https://doi.org/10.1016/j.ymeth.2009.07.005.

(11)   Garmer, D. R.; Gresh, N. A Comprehensive Energy Component Analysis of the

Interaction of Hard and Soft Dications with Biological Ligands. *Journal of the American

Chemical Society* **1994**, *116* (8), 3556–3567. https://doi.org/10.1021/ja00087a049.

(12)   Frederiksen, J. K.; Fong, R.; Piccirilli, J. A. Chapter 8. Metal Ions in RNA Catalysis. In

Nucleic Acid-Metal Ion Interactions; Royal Society of Chemistry: Cambridge, 2008; pp 260–

306. https://doi.org/10.1039/9781847558763-00260.

(13)   Kolodiazhnyi, O. I.; Kolodiazhna, A. Nucleophilic Substitution at Phosphorus:

Stereochemistry and Mechanisms. *Tetrahedron: Asymmetry* **2017**, *28* (12), 1651–1674.

https://doi.org/10.1016/j.tetasy.2017.10.022.

(14)   Lowe, G.; Cullis, P. M.; Jarvest, R. L.; Potter, B. V. L.; Sproat, B. S. Stereochemistry of

Phosphoryl Transfer. *Philosophical Transactions of the Royal Society of London. Series B,

Biological Sciences* **1981**, *293* (1063), 75–92.

(15)   Li, L.; Lelyveld, V. S.; Prywes, N.; Szostak, J. W. Experimental and Computational

Evidence for a Loose Transition State in Phosphoroimidazolide Hydrolysis. *Journal of the

American Chemical Society* **2016**, *138* (12), 3986–3989. https://doi.org/10.1021/jacs.6b00784.

(16)   Herschlag, D.; Jencks, W. P. Phosphoryl Transfer to Anionic Oxygen Nucleophiles. Nature of the Transition State and Electrostatic Repulsion. *J. Am. Chem. Soc.* **1989**, *111* (19), 7587–7596. https://doi.org/10.1021/ja00201a048.

(17)   Humphry, T.; Forconi, M.; Williams, N. H.; Hengge, A. C. An Altered Mechanism of Hydrolysis for a Metal-Complexed Phosphate Diester. *Journal of the American Chemical Society* **2002**, *124* (50), 14860–14861. https://doi.org/10.1021/ja027671c.

(18)   Li, L.; Lelyveld, V. S.; Prywes, N.; Szostak, J. W. Experimental and Computational Evidence for a Loose Transition State in Phosphoroimidazolide Hydrolysis. *Journal of the American Chemical Society* **2016**, *138* (12), 3986–3989. https://doi.org/10.1021/jacs.6b00784.

(19)   DeRose, V. J. Metal Ion Binding to Catalytic RNA Molecules. *Current Opinion in Structural Biology* **2003**, *13* (3), 317–324. https://doi.org/10.1016/S0959-440X(03)00077-0.

(20)   Frederiksen, J. K.; Li, N.-S.; Das, R.; Herschlag, D.; Piccirilli, J. A. Metal-Ion Rescue Revisited: Biochemical Detection of Site-Bound Metal Ions Important for RNA Folding. *RNA* **2012**. https://doi.org/10.1261/rna.028738.111.

(21)   Herschlag, D.; Jencks, W. P. The Effect of Divalent Metal Ions on the Rate and Transition-State Structure of Phosphoryl-Transfer Reactions. *Journal of the American Chemical Society* **1987**, *109* (15), 4665–4674. https://doi.org/10.1021/ja00249a033.

(22)   Saunders, A. M.; DeRose, V. J. Beyond Mg 2+ : Functional Interactions between RNA and Transition Metals. *Current Opinion in Chemical Biology* **2016**, *31*, 153–159. https://doi.org/10.1016/j.cbpa.2016.02.015.

(23)   Shan, S.; Kravchuk, A. V.; Piccirilli, J. A.; Herschlag, D. Defining the Catalytic Metal Ion Interactions in the *Tetrahymena* Ribozyme Reaction. *Biochemistry* **2001**, *40* (17), 5161–5171. https://doi.org/10.1021/bi002887h.

(24)   Wang, S.; Karbstein, K.; Peracchi, A.; Beigelman, L.; Herschlag, D. Identification of the Hammerhead Ribozyme Metal Ion Binding Site Responsible for Rescue of the Deleterious Effect of a Cleavage Site Phosphorothioate. *Biochemistry* **1999**, *38* (43), 14363–14378. https://doi.org/10.1021/bi9913202.

(25)   Zhou, D.-M.; Zhang, L.-H.; Taira, K. Explanation by the Double-Metal-Ion Mechanism of Catalysis for the Differential Metal Ion Effects on the Cleavage Rates of 5'-Oxy and 5'-Thio Substrates by a Hammerhead Ribozyme. *Proceedings of the National Academy of Sciences* **1997**, *94* (26), 14343–14348. https://doi.org/10.1073/pnas.94.26.14343.

(26)   Smith, J. S.; Nikonowicz, E. P. Phosphorothioate Substitution Can Substantially Alter RNA Conformation. *Biochemistry* **2000**, *39* (19), 5642–5652. https://doi.org/10.1021/bi992712b.

(27)   Maderia, M.; Hunsicker, L. M.; DeRose, V. J. Metal−Phosphate Interactions in the Hammerhead Ribozyme Observed by [31] P NMR and Phosphorothioate Substitutions. *Biochemistry* **2000**, *39* (40), 12113–12120. https://doi.org/10.1021/bi001249w.

(28)   Jacobson, K. B.; Turner, J. E. The Interaction of Cadmium and Certain Other Metal Ions with Proteins and Nucleic Acids. *Toxicology* **1980**, *16* (1), 1–37. https://doi.org/10.1016/0300-483X(80)90107-9.

(29)   Shi, D.; Nannenga, B. L.; Iadanza, M. G.; Gonen, T. Three-Dimensional Electron Crystallography of Protein Microcrystals. *eLife* **2013**, *2*, e01345. https://doi.org/10.7554/eLife.01345.

(30)   Frederiksen, J. K.; Piccirilli, J. A. Separation of RNA Phosphorothioate Oligonucleotides by HPLC. In *Methods in Enzymology*; Elsevier, 2009; Vol. 468, pp 289–309. https://doi.org/10.1016/S0076-6879(09)68014-9.

(31)   Cammack, R.; Hughes, M. Considerations for the Specification of Enzyme Assays Involving Metal Ions. **2022**.

(32)   Harding, M. M. The Geometry of Metal–Ligand Interactions Relevant to Proteins. *Acta Cryst D, Acta Cryst Sect D, Acta Crystallogr D, Acta Crystallogr Sect D, Acta Crystallogr D Biol Crystallogr, Acta Crystallogr Sect D Biol Crystallogr* **1999**, *55* (8), 1432–1443. https://doi.org/10.1107/S0907444999007374.

Chapter 4

Evaluating Mn^{2+} as a prebiotically plausible catalyst for RNA reactions

Evaluating Mn^{2+} as a prebiotically plausible catalyst for RNA reactions

Pazienza, L; Duzdevich, D; Zhou, L; Ding, D; and JW Szostak

[LTP wrote the text, performed experiments; LZ performed ligation experiments; Duzdevich performed the sequencing experiments, DD provided experimental materials; planning was shared with all authors. A version of this text will be submitted for peer review and open-access publication]

Abstract

Metal ion catalysis is integral in nonenzymatic primer extension in the RNA world. However, not all prebiotically relevant catalysts have been thoroughly assessed in context of the RNA world. We find that Mn^{2+} is a prebiotically plausible cation that increases nonenzymatic primer extension and ligation rates compared to Mg^{2+}. We additionally find Mn^{2+} improves yields of random sequence primer extension compared to Mg^{2+}. Additionally, Mn^{2+} maximally catalyzes primer extension and ligation at much lower concentrations than Mg^{2+}, and interacts with oligomers more effectively for catalysis. However, Mn^{2+} increases errors in random-sequence primer extension. We find that the decrease in fidelity due to Mn^{2+} correlates with the increase in hydrolysis of bridged dinucleotides to activated monomers, supporting previous theories of error in nonenzymatic primer extension. We hypothesize error due to hydrolysis could be mitigated in a system of continuous re-activation of reactants in future studies.

Introduction

In the RNA world hypothesis, RNA is proposed to be central to the origin of life due to its ability to store information and act as a catalyst[1]. In this hypothesis, however, RNA must first be assembled chemically into oligomers and through template-dependent methods before it can reach lengths in which it would be able to catalyze its own polymerization and other metabolic reactions[2]. This nonenzymatic copying chemistry in RNA is dependent on metal ion catalysis, as was discussed in Chapters 2 and 3. The identity of the metal ions affects rates of extension and hydrolysis, and affects the regioselectivity of extension—therefore knowing how different ions could function in the RNA world is a relevant consideration for the evolution of RNA into life.[3,4]

In comparison to bodies of water in our present atmosphere, we anticipate a different solution environment would have existed at the advent of life. Oxygen was not a major component of the atmosphere until after oxygenic photosynthesis evolved, so on the early Earth, reduced forms of transition metals would be more commonly found.[5] Thus far, Mg^{2+} and Fe^{2+} have dominated studies as prebiotically relevant cations.[6,7] Fe^{2+} is often considered as a progenitor prebiotic cation that could have preceded the dominance of Mg^{2+}, which is proposed to have grown in prevalence due to the Great Oxygenation Event (GOE) diminishing the Fe^{2+} accessibility for life.[8,9] Whether iron is more prebiotically plausible than magnesium remains up to debate, however. Magnesium is an order of magnitude more abundant than iron in our crust, and would have been a substantial competitor for interactions with RNA, although its solubility in the presence of phosphates acts as an issue in prebiotic systems.[8,10,11] Arguments in favor of iron as a prebiotic catalyst focus on the fact that Fe^{2+} is capable of enabling native RNA folding and often improves reaction rates when substituting for Mg^{2+} and has been shown to function better in nonenzymatic primer extension than Mg^{2+} at pH 7.[5,8] Of note is that these primer

extension experiments have only been successfully completed with Fe^{2+} and 2-methylimidazolides, as precipitation is a major issue with iron and the 2-aminoimidazolides and 2-methylimidazolides are not as prebiotically plausible as 2-aminoimidazolides.[12] Additionally, in order for Fe^{2+} to be a prebiotically likely catalyst, early life reactions would have needed to occur at neutral to acidic pH. Alkaline pH, as would be the case in many hydrothermal vents, would be incompatible with iron catalysis. In fact, at high pH, the presence of Fe^{2+} inhibits Mg^{2+} catalysis of primer extension.[6] Under alkaline conditions, Fe^{2+}-hydroxide species would dominate, and while Fe^{2+}-hydroxide species adsorb polynucleotides, they do not positively or negatively affect template-directed polymerization.[13] There may be a benefit from the interaction of these species in terms of longer-term protection/ local concentration of nucleotides, but this effect has yet to be observed.

One substantial argument favoring iron as a prebiotic catalyst is its ability to substitute for Mg^{2+} in enzymes and ribozymes, but it is not unique in this ability— other cations, such as Mn^{2+} also improve reactivity compared to Mg^{2+}.[7,14] Because the ionic composition for the prebiotic soup would depend heavily on the pH of the solution, and abundance in the crust is not the only consideration for how these solutions would be distributed, we must consider additional cations for their prebiotic relevance. In modern sea water, the concentration of Mn^{2+} is 0.0001M, and has been available in its +2 form throughout Earth's history, unlike Fe^{2+}, which is much less stable in solution in the modern atmosphere.[10,15] While Mn^{2+} is less abundant than Fe^{2+} in the crust and universe,[11] depending on the pH, it can be more prevalent in solution due to the differing solubilities of their respective metal-hydroxo species.[8,16] For example, at pH 7 and 100 μM metal ion, $Mn(OH)_2$ will be primarily dissolved, while $Fe(OH)_2$ is near its solubility constant.[11] If the bodies of water had more sulfide dissolved in them, MnS would be sparingly

soluble under the same conditions, while FeS would be completely insoluble.[11] Although FeS complexes are key in redox catalysis for life, they are not known to be substantial catalysts of RNA nonenzymatic primer extension or ligation. Soluble catalysts are believed to be key in nonenzymatic primer extension, although some catalysis through adsorption onto solids has been recorded.[17,18] An additional consideration is that although Fe^{2+} is more Lewis acidic than Mn^{2+}, there is a fine balance between activating the primer for primer extension or ligation and in activating hydrolysis, which destroys the reactants.[19]

In order to consider the role of Mn^{2+} as a prebiotically relevant cation, we need to consider its elemental properties. Mn^{2+} differs from Mg^{2+} in a number of chemically relevant ways— it has a larger ionic radius, is more Lewis acidic, binds water less tightly and has a more flexible coordination geometry, in addition to being capable of redox reactions.[19,20] Mn^{2+} and Fe^{2+} are able to interact well with O, N, and S, while Mg^{2+} interacts most strongly with O (and in some instances with N).[21,22] The balance between hardness and softness for diverse chemical interactions is why transition metals may have been useful in the origin of life even though they were less abundant than alkaline earth cations. Like Fe^{2+}, Mn^{2+} can substitute for Mg^{2+} in enzymes and ribozymes, and often increases reaction rates.[14,20,23,24] However, Mn^{2+} negatively affects fidelity of polymerization in DNA polymerases,[25,26] which may additionally be the case in the nonenzymatic reaction. For example, fidelity goes from 1 error out of every 1400 nucleotides to 1 error every 600 nucleotides when swapping Mn^{2+} for Mg^{2+} in enzymatic systems.[26,27] Increasing reaction rates may not be the only goal in optimizing primer extension, but also ensuring the right conditions for high fidelity information transfer. While it has been shown nonenzymatic primer extension proceeds much more slowly after a mismatch, it is unclear if this trend holds with different catalytic metal ions.[28,29] In addition to changes in fidelity, a change in

158

regioselectivity is also a possibility in the presence of "noncanonical" metal ions.[30] Nevertheless, a mix of 2′ and 3′ linkages in RNA can be tolerated to an extent by ribozymes and in primer extension, so this may not result in as much of an issue as anticipated.[31] Ions can substantially change the structure of oligonucleotides, however. Mn^{2+} has been shown to effect the conformational state of DNA helices in the molar ratio of 0.1 to 1.5 Mn^{2+}/DNA-phosphates, through hydrogen bond breaking between base pairs and through DNA aggregation.[32] Additionally, Mn^{2+} can induce H bonds between the N7 atoms of G and phosphates.[33,34] These structural affects may not necessarily have negative consequences, however. Mn^{2+} is known to lower the melting temperature of oligonucleotides,[35] and considering that a major hurdle in achieving RNA replication is the separation of fully-extended RNA duplexes, perhaps this melting temperature depression could be of use in prebiotic systems.

From our understanding of nonenzymatic primer extension through the studies in Chapters 2 and 3, we expect cations relevant for catalysis of primer extension to interact with both the 3′-OH and the equatorial oxygen of the phosphate in the reacting bridged dinucleotide. Mn^{2+} was observed in Chapter 2 (Figure 3) to vastly improve the rate of homo-polymeric primer extension of Cp*pC bridged dinucleotides at pH=7.5, and to interact the most tightly with the primer/template complex of the metal ions sampled (36 ± 6.5 mM compared to several hundreds of millimolar for Mg^{2+}). We expect that Mn^{2+} will improve the rates of other nonenzymatic reactions that benefit from Lewis acid catalysis, and that its lower saturation concentration could be beneficial for vesicle compatibility. In this chapter, we will explore the reactivity of Mn^{2+} in nonenzymatic ligation and mixed template primer extension, and study the effects of Mn^{2+} on the fidelity of primer extension. We additionally look at the effect of Mn^{2+} on bridged dinucleotide hydrolysis and formation to evaluate the cause of the high observed error rate in Mn^{2+}-catalyzed

primer extension. While high concentrations of Mn^{2+} are likely to be detrimental in prebiotic systems, we propose that at more prebiotically relevant lower concentrations of Mn^{2+} that the cation would be beneficial in improving yields of pre-RNA world reactions, and aid in complex systems by lowering oligonucleotide melting temperatures.

Results and Discussion

Nonenzymatic ligation with Mn^{2+} is faster than with Mg^{2+}

In addition to nonenzymatic primer extension, ligation serves as an important reaction during the RNA world, enabling several shorter templates to be connected into longer templates, as opposed to extending a primer one nucleotide at a time. However, ligation is much slower than primer extension under the same conditions, with hydrolysis as a significant competing reaction[36]. Therefore, improving the rates of nonenzymatic ligation, especially at lower concentrations of metal ion, would be beneficial to reactions in the RNA world. In the presence of Mn^{2+}, we observe a rate increase in ligation compared to when using Mg^{2+} (**Figure 1**). Ligation rates with Mg^{2+} under the same conditions saturate at approximately 0.1 hr^{-1} compared to the maximum rate with Mn^{2+} saturating at ~0.25 hr^{-1} at much lower concentrations.[36] This increase in rate (~2.5 fold) is modest, but Mn^{2+} ligation approaches saturation around 100 mM versus 1.6 M in the case of Mg^{2+}, which is a much more accessible prebiotic concentration. The mechanistic aspects of ligation are outside of the scope of this chapter, but it is notable that this catalytic benefit is less than one would expect if ligation was mediated through a metal bound-hydroxide,[37] although the maximum rates observed in this system may be hindered by $Mn(OH)_2$ solubility and oxidation at pH=8. However, the rates of ligation at lower pH favorable for Mn^{2+} solubility are substantially slower and make comparison of the systems difficult with kinetic studies. In addition, we observe that Mg^{2+} competes with Mn^{2+}, decreasing the binding affinity of

160

Mn^{2+} for the ligation reaction center while maintaining a comparable maximum ligation rate. In conjunction with the results from Chapter 2, Figure 3, it is apparent that Mn^{2+} poses a rate benefit to both primer extension and ligation, and is of interest as a prebiotic catalyst.
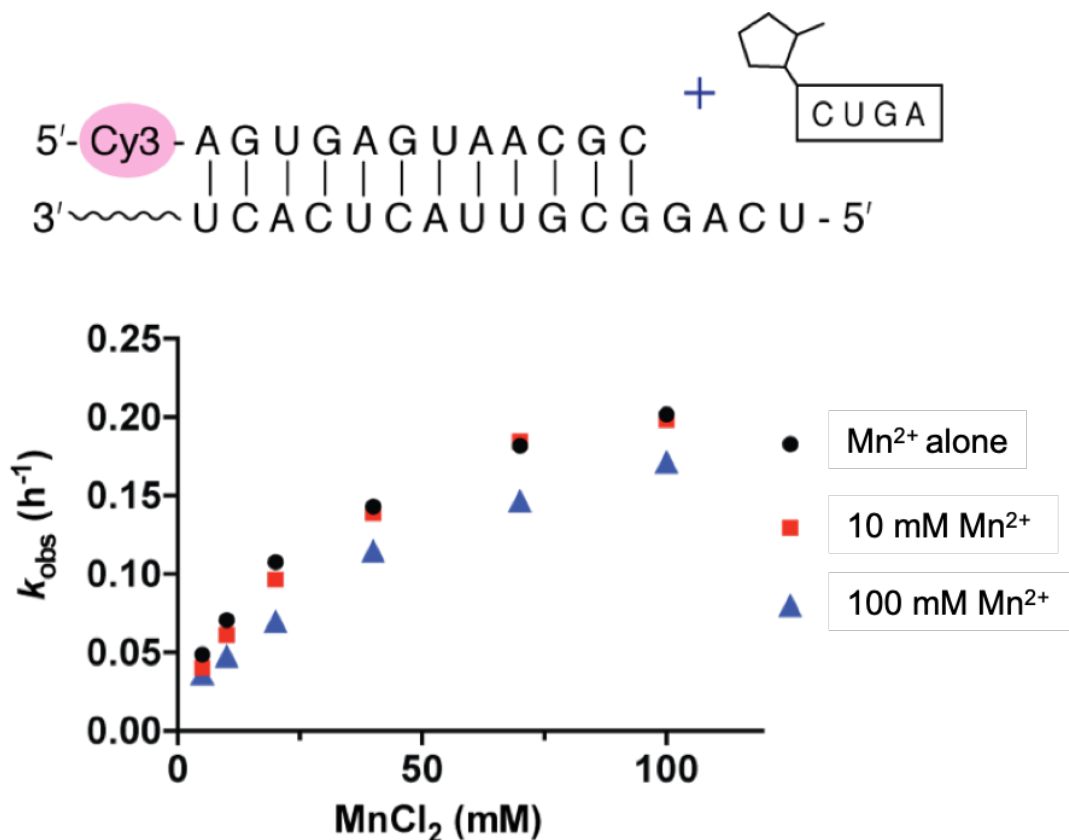
Figure 1. Ligation in the presence of varying Mn^{2+} and constant background Mg^{2+} conditions. Ligation at pH=8, in conditions described in Appendix D, with sequence shown- note the activation is 2-methyl imidazole, not 2-aminoimidazole, as it is the best activating group for nonenzymatic ligation. In the presence of Mn^{2+} alone (No Mg^{2+}), V_{max}=0.249 hr^{-1}, K_M= 25.9 mM. In the presence of Mn^{2+} and 10 mM Mg^{2+}, V_{max}= 0.268 hr^{-1}, K_M= 34.5 mM. In the presence of Mn^{2+} and 100 mM Mg^{2+}, V_{max}= 0.2420 hr^{-1}, K_M= 43.6 mM. Error bars indicate ± SD (n≥3, independent experiments).

Since previous studies (Chapter 2) of primer extension with Mn^{2+} were only of homo-polymeric extension, which serves as an ideal model system for primer extension, it was also of interest to study Mn^{2+} in catalyzing mixed template extension of longer oligomers. The longest extension performed with 2-aminoimidazole activated nucleotides and trimers is utilizing the sequence system from Li et al, and allows propagation out to +7.[38] We were interested in comparing Mn^{2+} to these results, but due to the pH sensitivity of primer extension, we performed the experiments at pH=7.5 and additionally replicated the Mg^{2+} conditions under anoxic pH=7.5 as well (**Figure 2**). Both Mg^{2+} and Mn^{2+} produce the desired product (+7) within 5 hours. Despite the faster rates with Mn^{2+} in a homopolymer, Mn^{2+} catalyzes less formation of the desired +7 product than Mg^{2+} (13% +7, Mn^{2+} vs 31% +7 Mg^{2+} or 39% primer remaining, Mn^{2+} vs. 21% primer remaining, Mg^{2+}). The lower percentage of desired product with Mn^{2+} is in part due to overextension off-template compared to Mg^{2+}. Nevertheless, Mn^{2+} tends to saturate in reaction rate at lower concentrations (**Chapter 2, Figure 3)**, and perhaps could be just as efficacious in long extension at lower metal ion concentrations compared to Mg^{2+}. Additionally, the reaction performed in Figure 2 begins with activated monomers and trimers, and requires the formation of the reacting bridged dinucleotides to occur *in situ*. The effect of Mn^{2+} on the equilibrium between activated monomers and bridged dinucleotides has not been previously studied. While it does not vastly outperform Mg^{2+} in terms of yield under these conditions, Mn^{2+} is certainly a plausible cation for successful primer extension in these conditions. The over-extension of the primer is of interest, as the amount of off-templated extension observed is surprising— off-templated extension is often much slower than templated extension.[39] Given the issue with extension at the end of long templates, there could be benefit to the presence of Mn^{2+}

in catalyzing last base addition. With this, we asked if Mn^{2+} has the same mechanistic constraints as Mg^{2+} holds in primer extension

Previously, a downstream helper was used to ascertain interactions between the catalytic metal ion and the reacting bridged dinucleotide (**Chapter 2, Figure 6**). In Mg^{2+}-catalyzed extension, a statistically significant difference in binding affinity was observed when the bridged dinucleotide was pushed closer to the reacting 3′-OH with the aid of a downstream helper oligonucleotide. Given the coordination flexibility and slightly larger size of Mn^{2+}, we wondered if this same effect would be observed with Mn^{2+} (**Figure 3**). Unlike in the Mg^{2+} system, there is no substantial change in binding affinity of Mn^{2+} with the presence of a downstream helper. This does not rule out any contacts between Mn^{2+} and the reacting bridged dinucleotide, but does suggest that the benefit of pre-organization of the primer-template and distance between 3′-OH and metal ion is less important for Mn^{2+}-catalyzed primer extension. This could be due in part to the coordinative flexibility and larger size of Mn^{2+}, which can interact similarly even in conditions with the bridged dinucleotide 0.5 angstrom further from the 3′-OH. Another possibility is since Mn^{2+} is more Lewis acidic than Mg^{2+}, and is more activating of the 3′-OH at the same concentrations, a more reactive nucleophile could have more lenient conformational standards than a less reactive nucleophile. There is a benefit observed to having the downstream helper in the presence of Mn^{2+}, however, notably increasing the maximum rate of primer extension— clearly the Mn^{2+}-catalyzed primer extension reaction benefits from improved structural organization of the duplex. These results are compatible with our understanding of the metal ion catalysis of nonenzymatic primer extension. With what we have observed thus far, a few questions remain: does Mn^{2+} based primer extension have comparable fidelity to Mg^{2+}-catalyzed primer extension? Are yields of primer extension in the presence of Mn^{2+} substantially

affected due to a change in the equilibrium of activated monomers and bridged dinucleotides?

We address this subsequently.
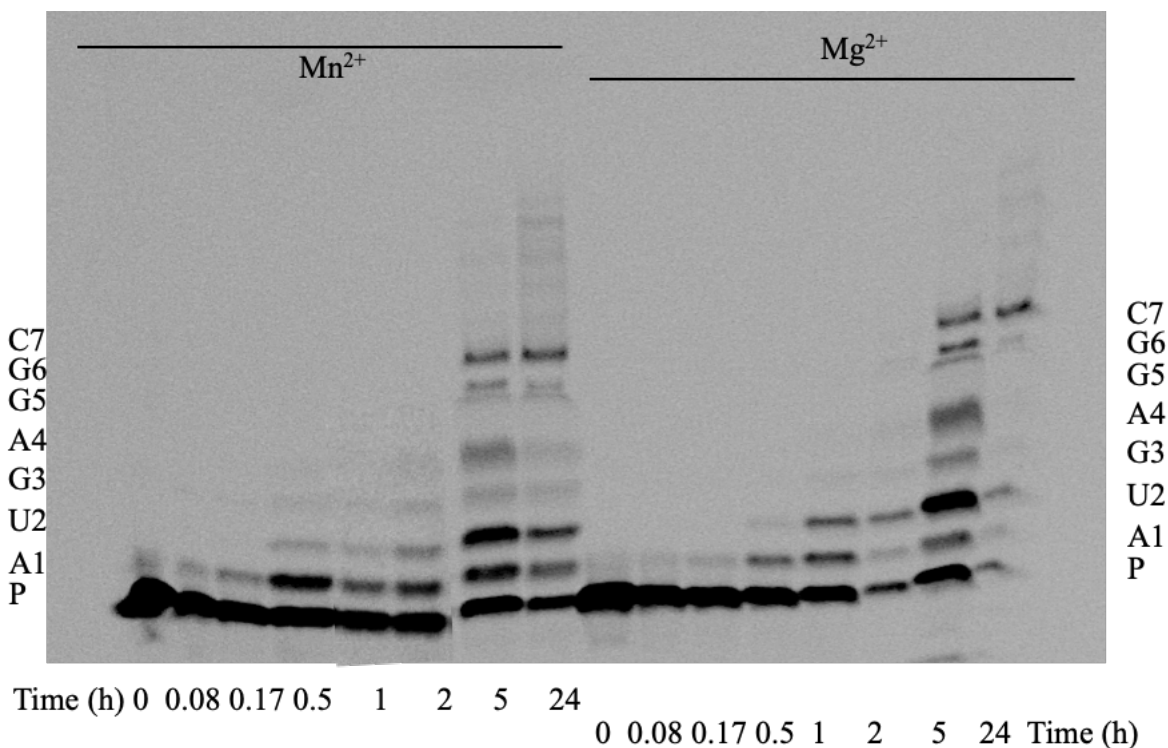
Figure 2. Mixed-template primer extension at pH=7.5 over 24 hours. Primer extension performed as described in Appendices A and D. Image is a composite of 2 gels, with t=1, 2 hours for Mn^{2+} from another replicate due to low sample concentration in quench mixture. Mixed template extension with 50 mM $MnCl_2$ and $MgCl_2$ shown. The template ends after C7- any additional higher molecular weight bands are a result of over-extension. Reactions were performed in triplicate, with representative primer extension shown in this gel.
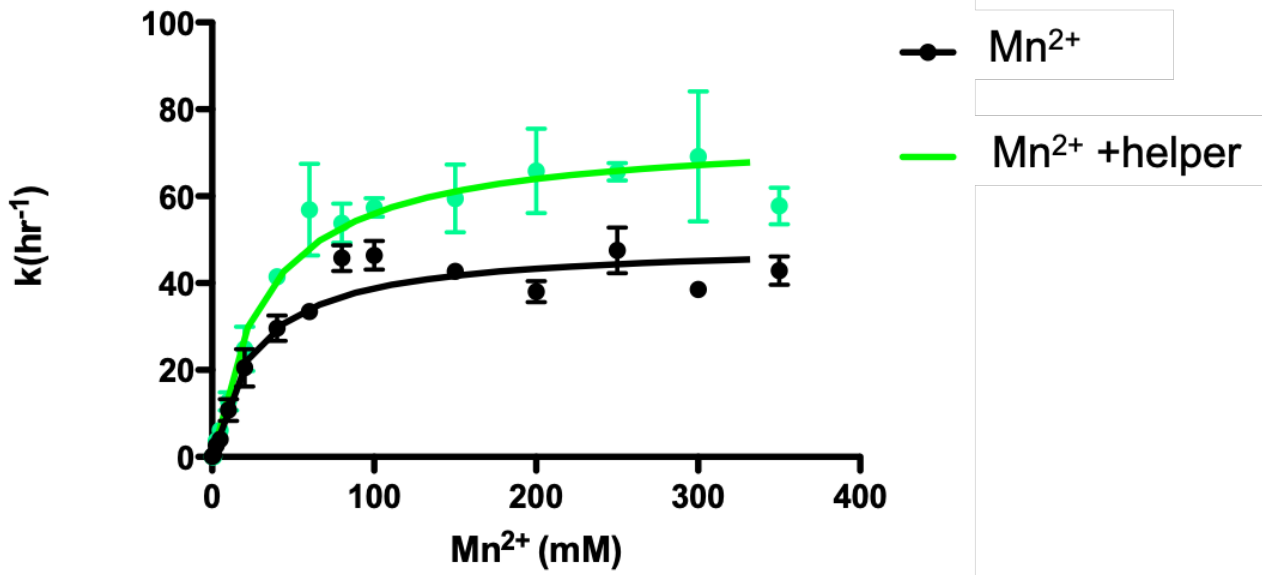
Figure 3. Mn^{2+} binding and primer extension reaction rates in presence of downstream helper oligomer. Primer extension was performed at pH 7.5 in an anoxic environment. Without the helper, V_{max}=49 $\pm$ 2 hr^{-1}, K_M= 26.5 $\pm$ 4 mM. In the presence of helper, V_{max}=74.5 $\pm$ 3 hr^{-1} , K_M= 33 $\pm$ 5 mM. Error bars indicate $\pm$ SD (n$\geq$3, independent experiments).

Using previously developed methodology for nonenzymatic primer extension sequencing (NERPE-Seq),[40] we attempted primer extension of random sequences over the course of 24 hours and subsequently sequenced the products (**Figures. 4.1, 4.2**). Compared to previously published conditions of 50 mM Mg^{2+} at pH=8, we observe in the presence of 50 mM Mn^{2+} at pH=7.5 that more products are observed across the length distribution of products at 24 hours, supporting the previous kinetic results of Mn^{2+} catalyzing faster primer extension. In addition, this system begins with the mix of activated monomers, and shows Mn^{2+} has increased extension products compared to Mg^{2+} in random sequences, so the effect shown in **Figure 2** may be to an extent sequence-dependent. However, we also observe that the error-frequency of primer extension doubles in the Mn^{2+} primer extension conditions (**Figure 4.1**). This is not likely to be

due to the pH change (hydrolysis at pH=7.5 of the bridged dinucleotide is at its slowest, **Chapter 2 Figure 5.2**), so the error frequency must in part be due to Mn^{2+}. Whether the error frequency is unique to Mn^{2+} catalyzed primer extension, or due to the increase in activated monomer to bridged dinucleotide ratio as previously proposed is of interest for us to ascertain.[41]
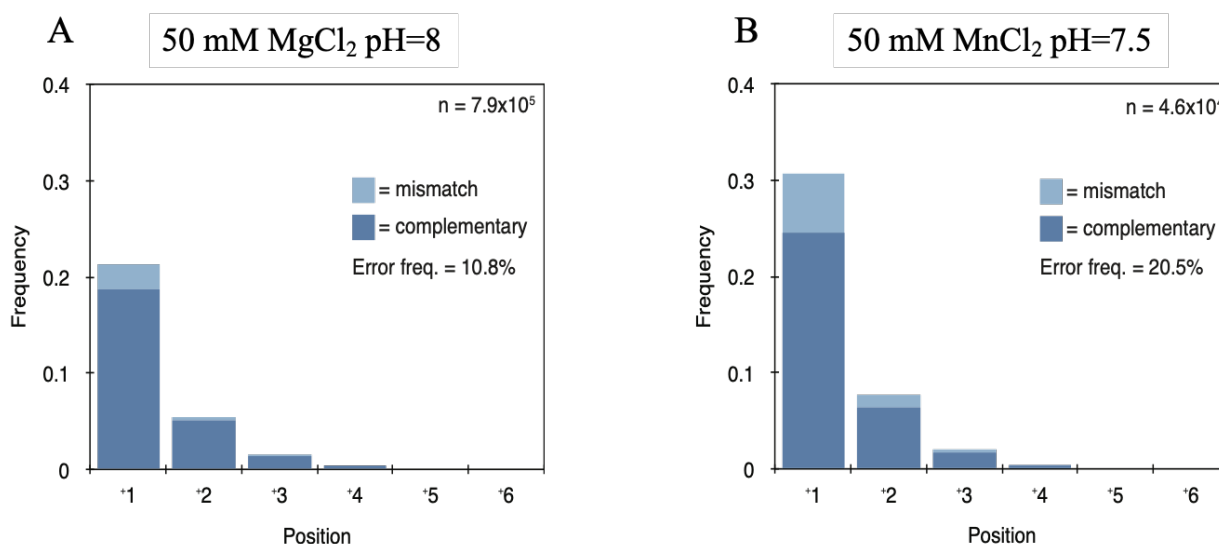
Figure 4.1. Primer extension in the presence of Mn^{2+} produces more products but is also more error prone. Frequencies of complementary and mismatched nucleotide incorporations (20 mM 2AIrN, 24 h; n = unextended hairpins + total nucleotide incorporation events). Note precipitation was observed in the Mn^{2+} catalyzed system, and the products were redissolved in the presence of EDTA for subsequent processing (A) 50 mM $MgCl_2$, pH=8 reproduced from Duzdevich et al[41]. (B) 50 mM $MnCl_2$, pH=7.5 (ambient oxygen conditions).

While an increase in error frequency would be greatly problematic during the RNA world, Mn^{2+}—catalyzed primer extension vastly flattens out the sampled sequence landscape in the oligomers it forms (**Figure 4.2**). While Mg^{2+} holds a bias for G/C containing products, Mn^{2+} diminishes this bias substantially, especially over increasing extension position. This is further

supported by the results with the inferred bridged dinucleotide extension (**Figure 4.2C and D**).

Except for Up*pU, and Ap*pU, there is greater sampling of possible sequence combinations

across all bridged dinucleotides. Lower concentrations or chelation of Mn^{2+} may confer similar

yield benefits and bias reduction due to the stronger interaction with the reaction center of primer

extension, but possibly reduce error. Since Mn^{2+} is less abundant than Mg^{2+} anyways, the

synergy between these two ions could be of benefit in the RNA world. First, we will attempt to

narrow the source of error in Mn^{2+}-catalyzed primer extension before studying Mn^{2+}-catalyzed

extension further.
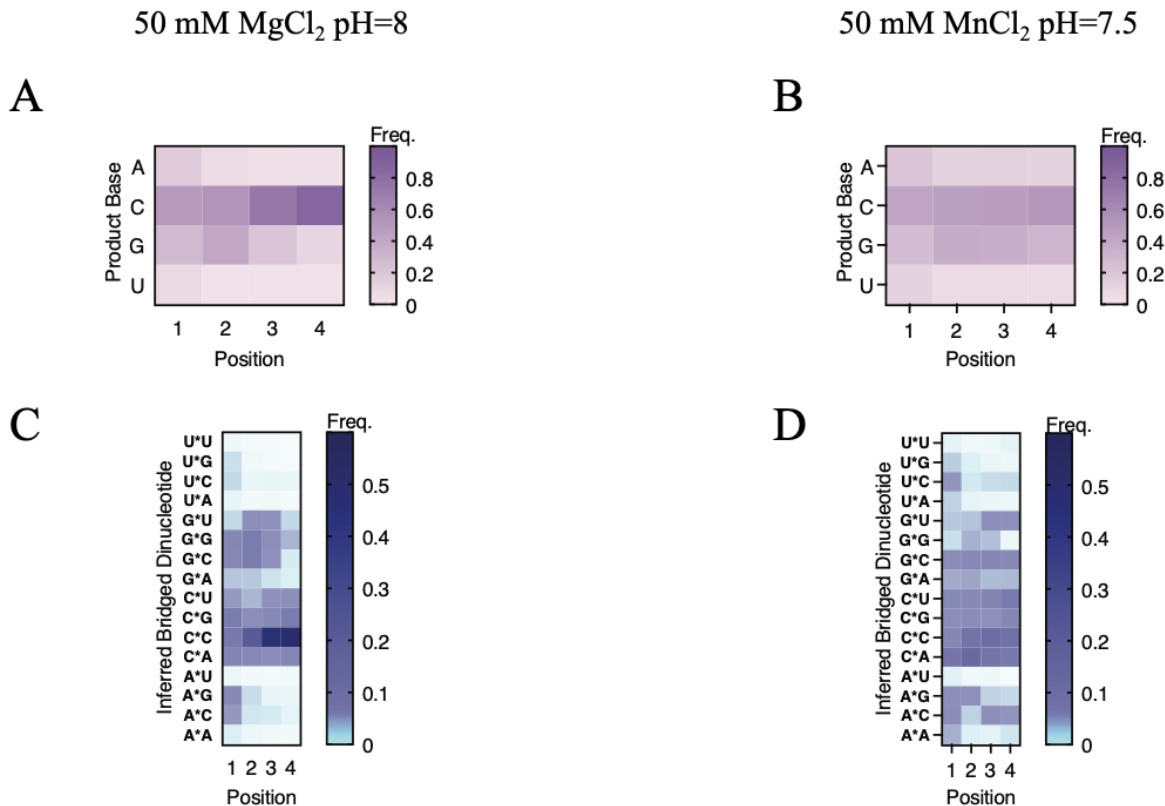
Figure 4.2. Primer extension in the presence of Mn^{2+} flattens the sequence landscape to a more

even distribution of extension products (A) Position-dependent base frequencies of

(**Figure 4.2 continued**) complementary products, 50 mM Mg^{2+}, pH=8, reproduced from[41] (B) Position-dependent base frequencies of complementary products, 50 mM Mn^{2+}, pH=7.5 (C) Position-dependent frequencies of inferred bridged dinucleotides generating complementary products, 50 mM Mg^{2+}, pH=8, reproduced from[41] (D) Position-dependent frequencies of inferred bridged dinucleotides generating complementary products, 50 mM Mn^{2+}, pH=7.5

Mn^{2+} is more Lewis acidic than Mg^{2+} by about 1 pKa unit, which would lead us to expect increased hydrolysis of reactants in addition to increased extension. We observe extension is not as fast as would be expected from the $Mn(OH)^+$ species being the active catalyst (**Chapter 2, Figure 3**). While the kinetics experiments were performed on a timescale much faster than the hydrolysis of the reactant, the experiments for sequencing and mixed template extension were not, and would be subjected to a buildup of inhibitors of primer extension over time. Although this allows us to consider Mn^{2+} in its full context as a catalyst in the prebiotic milieu, it creates difficulty in determining the source of error in Mn^{2+}-catalyzed primer extension. Therefore, we aim to characterize the effect Mn^{2+} has on hydrolysis and formation of the bridged dinucleotide to better parse the source of the error rate. Because Mn^{2+} interacts strongly with G and large concentrations of activated monomers and dinucleotides, we limited these studies to a model system with Cp*pC and *pC, which were soluble with Mn^{2+} at pH=7.5 across the concentrations sampled. We do not anticipate the effects of hydrolysis and bridged dinucleotide formation to be dependent on nucleobase substantially, and although we could imagine that G would be hindered by the precipitation in the presence of Mn^{2+}, this does not appear to reflect itself in the sequencing results (**Figure 4.2**).

Beginning with 24 mM *pC, it is evident that hydrolysis substantially occurs in the presence of Mn^{2+}. By the time the reaction is analyzed, the amount of *pC has more than halved

(**Figure 5.1**). Over time, Cp*pC builds up in concentration, and not much more CMP forms over the course of 2 hours (the time period expected to reach the maximum ratio of *pC :Cp*pC from previous studies).[41] At the point of maximum bridged dinucleotide concentration, there is 6 mM *pC and 0.6 mM Cp*pC (with 17 mM CMP) (**Figure 5.1**). Previously, the ratio of bridged dinucleotide to activated monomer correlates with the error rate observed in nonenzymatic primer extension. The presence of 10x excess activated monomer is likely a major contributor to the error observed in Mn^{2+}-catalyzed primer extension. The K_M of monophosphates for the primer/template is near 20 mM, which in these conditions would be competitive for binding.[42,43] The role of monophosphate in error is not as clear, although this competition for binding and inhibition of primer extension rate is likely to be a contributor to error similarly to *pC in the sequencing conditions.
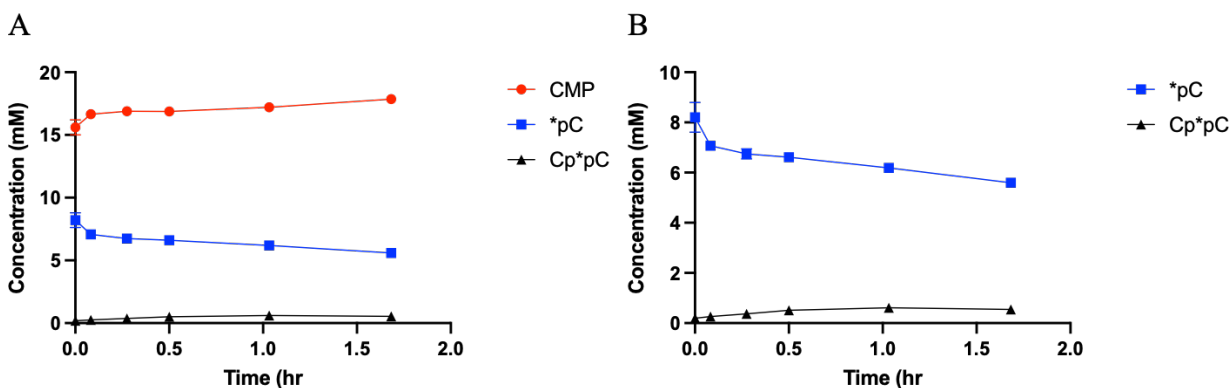
Figure 5.1. Cp*pC formation from *pC in the presence of 50 mM Mn^{2+}, pH= 7.5, measured using analytical HPLC (B) Y axis scale without CMP, same conditions as (A). Error bars indicate ± SD (n≥3, independent experiments). Concentration range for these reactions was chosen for comparison to published kinetics experiments.[43]

The difference in amounts of activated monomer and bridged dinucleotide are additionally clear when comparing the bridged dinucleotide formation with 50 mM Mg^{2+} at pH=8 (**Figure 5.2**). While beginning with the same amount of *pC, there is much less hydrolysis

of the activated monomer, and more bridged dinucleotide is able to form (although still ~10 fold less in absolute concentration). However, given the K_M of the bridged dinucleotide for the primer/template ranging between hundreds of micromolar to low millimolar, there is much more occupation of the primer/template in the Mg^{2+} system with the reactive bridged dinucleotide. The K_M for the activated monomer is competitive in these conditions, and would contribute to inhibition.[44] The ratios of activated monomer: bridged dinucleotide may be relevant assuming enough reactant is present, but since the amount of possible reactants is about half of what is possible in the Mg^{2+} system in the presence of Mn^{2+}, it is still impressive the sheer amount of primer extension is able to occur over 24 hours.
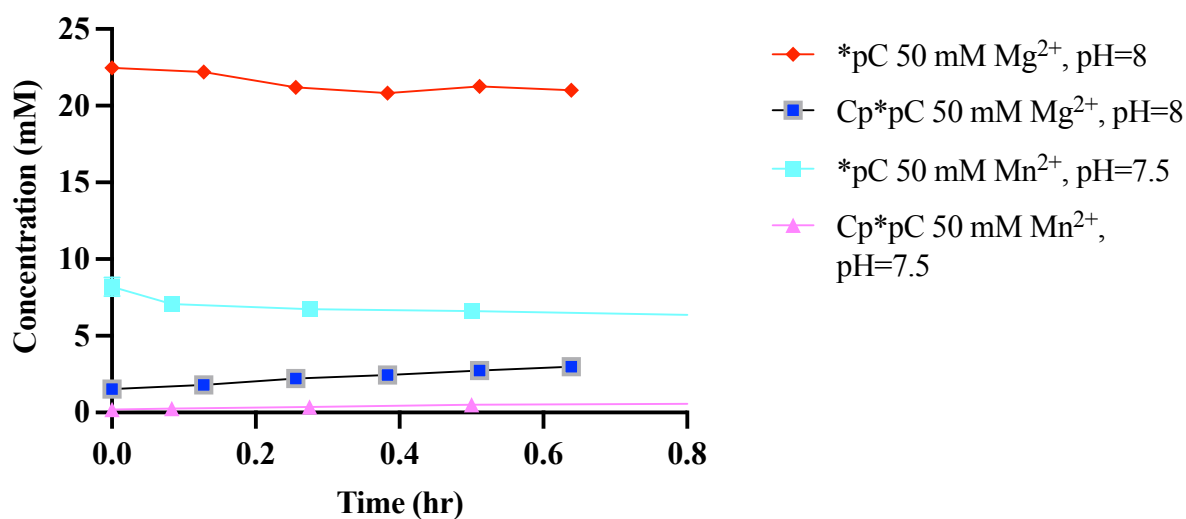
Figure 5.2. Comparison of *pC, Cp*pC concentrations under sequencing conditions. Mn^{2+} experiment performed with analytical HPLC as described in Appendix D, Mg^{2+} experiments using NMR as described in Appendix D. All *pC experiments were performed with the same starting aliquot of *pC from the same synthetic batch. Error bars indicate ± SD (n≥3, independent experiments).

Hydrolysis of the bridged dinucleotide can be elucidated more readily than that of the hydrolysis of *pC, since *pC is always forming Cp*pC at pH=7.5 and 8 as a competing reaction. Therefore, we can have a better sense of the half-life of the system by looking at the hydrolysis of Cp*pC. We find the rate of hydrolysis is 0.64 hr^{-1} ± 0.07 hr^{-1}, which is ~6x faster than Cp*pC hydrolysis at pH=8, 50 mM Mg^{2+} (the other sequencing condition, **Chapter 2, Figure 1.4**). With this k$_{hydrolysis}$, we expect the half-life of bridged dinucleotides to be approximately an hour, compared to nearly 7 hours in the Mg^{2+}-sequencing conditions. Mn^{2+} must additionally increase the reactivity of the activated monomer in addition to the bridged dinucleotide to account for the increased overall yield of extension under these conditions.
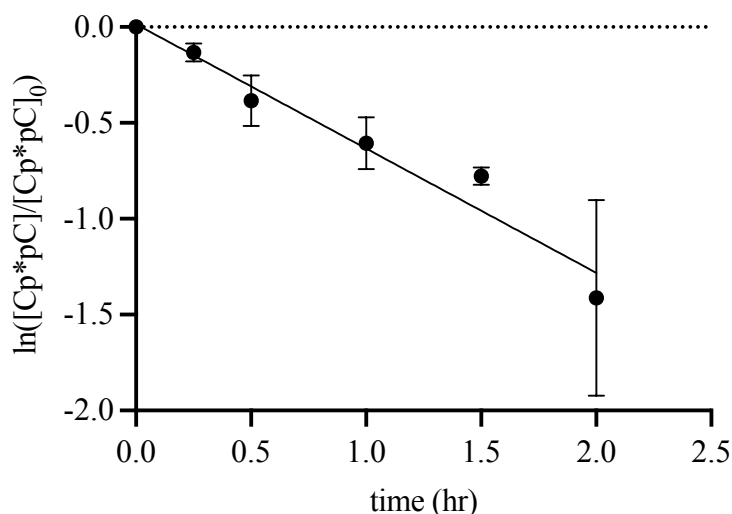
Figure 6.1. First order hydrolysis of Cp*pC with 50 mM Mn^{2+}, pH=7.5. Log of normalized concentrations over time measured by analytical HPLC. Error bars indicate ± SD (n≥3, independent experiments).

It is possible to measure the rate of primer extension with the activated monomer by using excess free 2-aminoimidazole to push the equilibrium of bridged dinucleotide and activated monomer towards the activated monomer side.[43] Beginning with concentrated aliquots

of activated monomer means likely contamination with the bridged dinucleotide. In **Figure 6.2**, we look at the effect of differing Mn^{2+} concentration on the rate of extension of inferred activated monomer. The presence of free 2-aminoimidazole forms precipitates with increasing concentrations of Mn^{2+}, although precipitate is not incompatible with subsequent analysis of the primer extension products. While the rates observed are much slower compared to the primer extension rates at the same concentration of Mn^{2+} with bridged dinucleotide (**Chapter 2, Figure 3**), these are measurable rates , which are not achievable with Mg^{2+}-catalyzed primer extension.[43] At pH=8 with 90 mM Mg^{2+}, but the same amount of 2AI, the rate of extension observed is only 0.38 hr^{-1}, which is accounted for by bridged dinucleotide formation in those conditions.[43] Bridged dinucleotide formation in the Mn^{2+} conditions at pH=7.5 cannot account fully for the observed maximum rate shown in Figure 6.2, and supports increased reactivity of activated monomers in presence of Mn^{2+}.
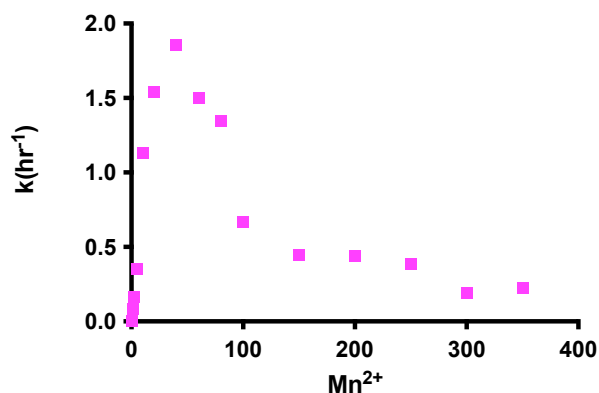
Figure 6.2. Mn^{2+} dependence of primer extension of *pC, pH=7.5 on a GGG template . Primer extension is performed in presence of 100 mM 2-aminoimidazole, which minimizes concentration of free Cp*pC. However, copious precipitation was observed across the concentrations sampled, likely due to complexes between Mn^{2+} and free 2-aminoimidazole. Error bars indicate ± SE.

While the bridged dinucleotide will outcompete the reaction of the activated monomer on short timescales, the competition between the two reactants is substantial for affecting the fidelity of primer extension on longer timescales. Mismatches are less likely to happen with the bridged dinucleotide, as both nucleotide positions need to be well ordered enough to interact with the primer/template for primer extension. This increase in activated monomer reactivity likely accounts for the observed error, and this is all a product of the increased hydrolysis of the bridged dinucleotide in the Mn^{2+}-catalyzed extension conditions. While the hydrolysis is an important influence when the source of bridged dinucleotide is from discrete influx of chemical nutrients, when there is chemical re-activation to favor a constant amount of bridged dinucleotides, we hypothesize the issue of fidelity in Mn^{2+} catalyzed primer extension would be less significant. The benefits of Mn^{2+}-catalyzed primer extension in increasing sequence space sampling and yield show it could be a beneficial prebiotic catalyst, even more so if it's hydrolytic ability is attenuated through chelation (as hydrolysis is conferred by metal-bound hydroxides). Mn^{2+} forms a number of stable coordination complexes, more so than Mg^{2+} due to the fact it's a transition metal, and is a promising candidate for searching for prebiotic scaffolds to aid in primer extension catalysis. Mn^{2+} remains an important catalyst in life today, and it is likely this arose early in the evolution of life, and is independent of the presence of oxygen in our atmosphere. Similarly, Mn^{2+} is actually compatible with the 2-aminoimidazolide system, which Fe^{2+} primarily results in precipitation and degradation. Since 2-aminoimidazole is more prebiotically plausible than 2-methyl imidazole, this could favor the use of Mn^{2+} compared to Fe^{2+}, although in actuality the prebiotic milieu likely was a mixture of a number of different cations all contributing to the necessary reactions in proportion to their concentration. Only upon

subsequent fitness benefit from the cation interacting with a compound which can specifically bind it and localize it for reactions would selection begin to occur for these cations.

Bibliography

(1)   Gilbert, W. Origin of Life: The RNA World. *Nature* **1986**, *319* (6055), 618–618. https://doi.org/10.1038/319618a0.

(2)   Joyce, G. F.; Szostak, J. W. Protocells and RNA Self-Replication. *Cold Spring Harbor Perspectives in Biology* **2018**, *10* (9), a034801. https://doi.org/10.1101/cshperspect.a034801.

(3)   Sawai, H. Oligonucleotide Formation Catalyzed by Divalent Metal Ions. The Uniqueness of the Ribosyl System. *Journal of Molecular Evolution* **1988**, *27* (3), 181–186. https://doi.org/10.1007/BF02100072.

(4)   Sawai, H. Catalysis of Internucleotide Bond Formation by Divalent Metal Ions. *Journal of the American Chemical Society* **1976**, *98* (22), 7037–7039. https://doi.org/10.1021/ja00438a050.

(5)   Holland, H. D. The Oxygenation of the Atmosphere and Oceans. *Philos Trans R Soc Lond B Biol Sci* **2006**, *361* (1470), 903–915. https://doi.org/10.1098/rstb.2006.1838.

(6)   Lin Jin, Aaron E. Engelhart, Weicheng Zhang, Katarzyna Adamala, and Jack W. Szostak. Catalysis of Template-Directed Nonenzymatic RNA Copying by Iron(II). *JACS 140* (44), 15016–15021. https://doi.org/10.1021/jacs.8b09617.

(7)   Okafor, C. D.; Lanier, K. A.; Petrov, A. S.; Athavale, S. S.; Bowman, J. C.; Hud, N. V.; Williams, L. D. Iron Mediates Catalysis of Nucleic Acid Processing Enzymes: Support for Fe(II) as a Cofactor before the Great Oxidation Event. *Nucleic Acids Research* **2017**, *45* (7), 3634–3642. https://doi.org/10.1093/nar/gkx171.

(8) Athavale, S. S.; Petrov, A. S.; Hsiao, C.; Watkins, D.; Prickett, C. D.; Gossett, J. J.; Lie, L.; Bowman, J. C.; O'Neill, E.; Bernier, C. R.; Hud, N. V.; Wartell, R. M.; Harvey, S. C.; Williams, L. D. RNA Folding and Catalysis Mediated by Iron (II). *PLoS ONE* **2012**, *7* (5), e38024. https://doi.org/10.1371/journal.pone.0038024.

(9) Saunders, A. M.; DeRose, V. J. Beyond Mg 2+ : Functional Interactions between RNA and Transition Metals. *Current Opinion in Chemical Biology* **2016**, *31*, 153–159. https://doi.org/10.1016/j.cbpa.2016.02.015.

(10) Wood, J. M. Biological Cycles for Elements in the Environment. *Naturwissenschaften* **1975**, *62* (8), 357–364. https://doi.org/10.1007/BF00625342.

(11) Williams, R. J. P. The Natural Selection of the Chemical Elements. **1997**, 14.

(12) Fahrenbach, A. C.; Giurgiu, C.; Tam, C. P.; Li, L.; Hongo, Y.; Aono, M.; Szostak, J. W. Common and Potentially Prebiotic Origin for Precursors of Nucleotide Synthesis and Activation. *Journal of the American Chemical Society* **2017**, *139* (26), 8780–8783. https://doi.org/10.1021/jacs.7b01562.

(13) Schwartz, A. W.; Orgel, L. E. Template-Directed Polynucleotide Synthesis on Mineral Surfaces. *J Mol Evol* **1985**, *21* (3), 299–300. https://doi.org/10.1007/BF02102362.

(14) Bray, M. S.; Lenz, T. K.; Haynes, J. W.; Bowman, J. C.; Petrov, A. S.; Reddi, A. R.; Hud, N. V.; Williams, L. D.; Glass, J. B. Multiple Prebiotic Metals Mediate Translation. *PNAS* **2018**, *115* (48), 12164–12169. https://doi.org/10.1073/pnas.1803636115.

(15) Kobayashi, K.; Ponnamperuma, C. Trace Elements in Chemical Evolution, I. *Origins Life Evol Biosphere* **1985**, *16* (1), 41–55. https://doi.org/10.1007/BF01808048.

(16) Johnson, J. E.; Webb, S. M.; Ma, C.; Fischer, W. W. Manganese Mineralogy and Diagenesis in the Sedimentary Rock Record. *Geochimica et Cosmochimica Acta* **2016**, *173*, 210–231. https://doi.org/10.1016/j.gca.2015.10.027.

(17) Miyakawa, S.; Joshi, P. C.; Gaffey, M. J.; Gonzalez-Toril, E.; Hyland, C.; Ross, T.; Rybij, K.; Ferris, J. P. Studies in the Mineral and Salt-Catalyzed Formation of RNA Oligomers. *Origins of Life and Evolution of Biospheres* **2006**, *36* (4), 343–361. https://doi.org/10.1007/s11084-006-9009-6.

(18) Zhao, Z.-R.; Wang, X. A Plausible Prebiotic Selection of Ribose for RNA - Formation, Dynamic Isolation, and Nucleotide Synthesis Based on Metal-Doped-Clays. *Chem* **2021**, *7* (12), 3292–3308. https://doi.org/10.1016/j.chempr.2021.09.002.

(19) Frederiksen, J. K.; Fong, R.; Piccirilli, J. A. Chapter 8. Metal Ions in RNA Catalysis. In *Nucleic Acid-Metal Ion Interactions*; Royal Society of Chemistry: Cambridge, 2008; pp 260–306. https://doi.org/10.1039/9781847558763-00260.

(20) Bock, C. W.; Katz, A. K.; Markham, G. D.; Glusker, J. P. Manganese as a Replacement for Magnesium and Zinc: Functional Comparison of the Divalent Ions. *Journal of the American Chemical Society* **1999**, *121* (32), 7360–7372. https://doi.org/10.1021/ja9906960.

(21) Poonia, N. S.; Bajaj, A. V. Coordination Chemistry of Alkali and Alkaline Earth Cations. *Chemical Reviews* **1979**, *79* (5), 389–445. https://doi.org/10.1021/cr60321a002.

(22) Harding, M. M. The Geometry of Metal–Ligand Interactions Relevant to Proteins. *Acta Cryst D, Acta Cryst Sect D, Acta Crystallogr D, Acta Crystallogr Sect D, Acta Crystallogr D Biol Crystallogr, Acta Crystallogr Sect D Biol Crystallogr* **1999**, *55* (8), 1432–1443. https://doi.org/10.1107/S0907444999007374.

(23)   Cowan, J. A. Metal Activation of Enzymes in Nucleic Acid Biochemistry. *Chemical Reviews* **1998**, *98* (3), 1067–1088. https://doi.org/10.1021/cr960436q.

(24)   Schramm, V. L. *Manganese in Metabolism and Enzyme Function*; Elsevier, 2012.

(25)   Hartwig, A. Role of Magnesium in Genomic Stability. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **2001**, *475* (1–2), 113–121. https://doi.org/10.1016/S0027-5107(01)00074-4.

(26)   Sirover, M. A.; Loeb, L. A. On the Fidelity of DNA Replication. Effect of Metal Activators during Synthesis with Avian Myeloblastosis Virus DNA Polymerase. *Journal of Biological Chemistry* **1977**, *252* (11), 3605–3610. https://doi.org/10.1016/S0021-9258(17)40295-X.

(27)   Sirover, M. A.; Loeb, L. A. Metal Activation of DNA Synthesis. *Biochemical and Biophysical Research Communications* **1976**, *70* (3), 812–817. https://doi.org/10.1016/0006-291X(76)90664-1.

(28)   Leu, K.; Kervio, E.; Obermayer, B.; Turk-MacLeod, R. M.; Yuan, C.; Luevano, J.-M.; Chen, E.; Gerland, U.; Richert, C.; Chen, I. A. Cascade of Reduced Speed and Accuracy after Errors in Enzyme-Free Copying of Nucleic Acid Sequences. *J. Am. Chem. Soc.* **2013**, *135* (1), 354–366. https://doi.org/10.1021/ja3095558.

(29)   Rajamani, S.; Ichida, J. K.; Antal, T.; Treco, D. A.; Leu, K.; Nowak, M. A.; Szostak, J. W.; Chen, I. A. Effect of Stalling after Mismatches on the Error Catastrophe in Nonenzymatic Nucleic Acid Replication. *J. Am. Chem. Soc.* **2010**, *132* (16), 5880–5885. https://doi.org/10.1021/ja100780p.

(30) Sawai, H.; Orgel, L. E. Oligonucleotide Synthesis Catalyzed by the Zinc(2+) Ion. *Journal of the American Chemical Society* **1975**, *97* (12), 3532–3533. https://doi.org/10.1021/ja00845a050.

(31) Sheng, J.; Li, L.; Engelhart, A. E.; Gan, J.; Wang, J.; Szostak, J. W. Structural Insights into the Effects of 2'-5' Linkages on the RNA Duplex. *Proceedings of the National Academy of Sciences* **2014**, *111* (8), 3050–3055. https://doi.org/10.1073/pnas.1317799111.

(32) Polyanichko, A. M. The Effect of Manganese(II) on DNA Structure: Electronic and Vibrational Circular Dichroism Studies. *Nucleic Acids Research* **2004**, *32* (3), 989–996. https://doi.org/10.1093/nar/gkh242.

(33) Millonig, H.; Pous, J.; Gouyette, C.; Subirana, J. A.; Campos, J. L. The Interaction of Manganese Ions with DNA. *Journal of Inorganic Biochemistry* **2009**, *103* (6), 876–880. https://doi.org/10.1016/j.jinorgbio.2009.03.004.

(34) Avacovici, A.; Gârban, Z. Circular Dichroism – Theory and Applications. **2006**, No. 2, 6.

(35) Eichhorn, G. L.; Shin, Y. Ae. Interaction of Metal Ions with Polynucleotides and Related Compounds. XII. The Relative Effect of Various Metal Ions on DNA Helicity. *Journal of the American Chemical Society* **1968**, *90* (26), 7323–7328. https://doi.org/10.1021/ja01028a024.

(36) Zhou, L.; O'Flaherty, D. K.; Szostak, J. W. Template-Directed Copying of RNA by Non-enzymatic Ligation. *Angew. Chem. Int. Ed.* **2020**, *59* (36), 15682–15687. https://doi.org/10.1002/anie.202004934.

(37) Rohatgi, R.; Bartel, D. P.; Szostak, J. W. Kinetic and Mechanistic Analysis of Nonenzymatic, Template-Directed Oligoribonucleotide Ligation. *Journal of the American Chemical Society* **1996**, *118* (14), 3332–3339. https://doi.org/10.1021/ja953712b.

(38) Li, L.; Prywes, N.; Tam, C. P.; O'Flaherty, D. K.; Lelyveld, V. S.; Izgu, E. C.; Pal, A.;
Szostak, J. W. Enhanced Nonenzymatic RNA Copying with 2-Aminoimidazole Activated
Nucleotides. *Journal of the American Chemical Society* **2017**, *139* (5), 1810–1813.
https://doi.org/10.1021/jacs.6b13148.

(39) Kanavarioti, A.; White, D. H. Kinetic Analysis of the Template Effect in
Ribooligoguanylate Elongation. *Origins Life Evol Biosphere* **1987**, *17* (3), 333–349.
https://doi.org/10.1007/BF02386472.

(40) Duzdevich, D.; Carr, C. E.; Szostak, J. W. Deep Sequencing of Non-Enzymatic RNA
Primer Extension. 13.

(41) Duzdevich, D.; Carr, C. E.; Ding, D.; Zhang, S. J.; Walton, T. S.; Szostak, J. W.
Competition between Bridged Dinucleotides and Activated Mononucleotides Determines
the Error Frequency of Nonenzymatic RNA Primer Extension. 11.

(42) Izgu, E. C.; Fahrenbach, A. C.; Zhang, N.; Li, L.; Zhang, W.; Larsen, A. T.; Blain, J. C.;
Szostak, J. W. Uncovering the Thermodynamics of Monomer Binding for RNA
Replication. *J. Am. Chem. Soc.* **2015**, *137* (19), 6373–6382.
https://doi.org/10.1021/jacs.5b02707.

(43) Walton, T.; Szostak, J. W. A Kinetic Model of Nonenzymatic RNA Polymerization by
Cytidine-5′-Phosphoro-2-Aminoimidazolide. *Biochemistry* **2017**, *56* (43), 5739–5747.
https://doi.org/10.1021/acs.biochem.7b00792.

(44) Tam, C. P.; Fahrenbach, A. C.; Björkbom, A.; Prywes, N.; Izgu, E. C.; Szostak, J. W.
Downstream Oligonucleotides Strongly Enhance the Affinity of GMP to RNA Primer–
Template Complexes. *J. Am. Chem. Soc.* **2017**, *139* (2), 571–574.
https://doi.org/10.1021/jacs.6b09760.

Appendix A

General Materials and Methods

All reagents were purchased from Sigma-Aldrich unless otherwise specified and were used without further purification. 2-Aminoimidazole HCl was purchased from Combi-Blocks, Inc. Deuterium oxide was purchased from Cambridge Isotope Laboratories, Inc. Bis-cyanoethyl N-,N-diisopropyl CED phosphoramidite was purchased from Chemgenes. Acetone and diethyl ether were from ThermoFisher Scientific, 2,2′-dipyridyl disulfide was from Chem-Impex International. Tris-HCl solutions (pH =7.0, 7.5, 8.0) and nuclease-free water were purchased from Fisher Scientific/Ambion. RNA oligonucleotides were purchased from Integrated DNA Technologies (IDT), RNA/LNA oligonucleotide was purchased from Exiqon/Qiagen. Cy3-labeled IDT oligonucleotides are desalted and HPLC purified (triethylamine as counter-cation). No inorganic salts are present in the oligonucleotide stock samples used for primer extension reactions.

All reactions were completed in oven-dried round-bottomed flasks with rubber septa and anhydrous Aldrich SureSeal™ solvents to minimize the presence of water. Reaction progress was monitored by low-resolution electrospray ionization ion-trap mass spectrometry on a Bruker Daltonics Esquire 6000 mass spectrometer (Billerica, MA). Reverse phase flash chromatography post-synthesis was performed on a Teledyne Isco CombiFlash Rf system (Lincoln, NE) on a C18 Aq column (30g or 50g depending on scale of synthesis).

Concentrations for all RNA/DNA-based compounds were calculated using predicted extinction coefficients at 260 nm as determined by OligoAnalyzer software provided by Integrated DNA Technologies (version 3.1, Coralville, IA)[1,2] and measured absorbance across

serial dilutions at 260 nm using Thermo Scientific Nanodrop 2000c spectrophotometer (Waltham, MA).

A.2 Synthesis of activated nucleotides (2-aminoimidazole/ hydroxy-7azabenzotriazole activated)

Synthesis of activated nucleotides was performed using existing protocols.[3,4] Activated monomer synthesis was favored in the presence of 2-aminoimidazole through excess 2-aminoimidazole compared to NMP, and larger reaction volumes. Due to the inability of HOAt to form bridged dinucleotides, these reaction considerations were not applicable for HOAt reactions. Note that HOAt activated monomers are fluorescent yellow depending on lyophilization pH, compared to the white solids produced from imidazole activation. If activated 2-aminoimidazolides are colored in their product form, it is due to free 2-aminoimidazole contamination and were be re-purified or discarded. Activated monomers were stored at -80 degrees Celsius in their solid form, and prepared into fresh stock solutions for each experiment.

A.3 Synthesis, purification, and characterization of 2-aminoimidazolium bridged dinucleotides

Synthesis and purification of 2-aminoimidazolium-bridged dinucleotides was performed as previously described,[5–7] but using 2-aminoimidazole HCl instead of the hemisulfate salt. Mixed dinucleotides ($N_1p^*pN_2$) were synthesized using published methods.[4] Post precipitation in $NaClO_4$ saturated 2:1 acetone: ether and subsequent desiccation, solid samples were dissolved in 10-20mL H_2O and subjected to flash chromatography (0-10% acetonitrile 0-10 min, 10-50% 10-14 min, 50-100% 14-16 minutes). 20 mM triethylamine bicarbonate buffer pH=7.5 was utilized as solvent A only when purifying compounds containing G, otherwise MilliQ water was utilized. Purified products were characterized by NMR (Varian INOVA 400 MHz NMR spectrometer) and mass-spectrometry (Esquire 6000 mass spectrometer from Bruker Daltonics, operated in negative ion mode). Sample purity was greater than 85% bridged dinucleotide for all primer

extension experiments, with residual 2-aminoimidazole activated monomer present at less than 15% as determined by ^{31}P NMR. Samples were additionally checked for presence of free 2-aminoimidazole, and if present, are discarded from use (ESI, ^{1}H NMR, visual inspection). The extinction coefficients for the bridged dinucleotide are double the theoretical extinction coefficient of the monomer as determined by OligoAnalyzer software provided by Integrated DNA Technologies (version 3.1, Coralville, IA).[1,2] Samples were stored as either solids or concentrated stock solutions (>300 mM), and were flash-frozen using N_2(l) if being saved for future use. As long as samples are stored on ice and flash-frozen, minimal degradation to activated monomers from bridged dinucleotides was observed over time.

A.4 Nonenzymatic primer extension reactions and analysis

We carried out primer extension experiments with previously reported methods using the oligonucleotides in Table A.1.[7,8] All primer extension reactions contained a final concentration of 100 mM buffer (Tris-HCl for pH=7-8.5; CHES for pH 9+, MES for below pH 7). MES and CHES buffers were made from the solid compound and adjusted using minimal HCl and NaOH. Buffers were chosen to minimize presence of external Na^+ ions. The primer extension solutions had a final concentration of Cp*pC and Gp*pG dimer at 20 mM, and Ap*pA dimer, 29.7 mM in accord with binding curves (**Chapter 2: Figure 1.0A, Figure 1.2**). Other added compounds are specified in the figure legends. Cobalt hexammine solutions were pH adjusted to 7.5 before use. Time points for kinetics varied depending on the concentration of metal ion, and on whether imidazolium-bridged dinucleotide used was Ap*pA or Cp*pC due to substantial differences in rate in Cp*pC- versus Ap*pA- driven primer extension. In Cp*pC-driven primer extension, it is necessary to sample more frequently to accurately represent the initial rate (**Table A.2**). We measured the pH before and after all primer extension reactions using an Orion™ 9863BN

needle-tip Micro pH Electrode at 25 °C to verify that there was no effect on the pH due to the addition of bridged dinucleotide stock or concentrated metal ion stocks. Primer extension was initiated by Cp*pC/Ap*pA dimer addition, and all reactions were at 25 °C. Quench conditions for reaction are as previously reported.[9] We performed $MnCl_2$ reactions in a Coy Laboratory vinyl anaerobic chamber(N_2 atmosphere with 2.5 to 3% H_2) due to oxidation and formation of manganese oxide precipitates in the pH range sampled. Reaction mixtures were prepared using degassed nuclease free water, with quench solutions run through the vacuum purge-fill cycle with the caps open (because these solutions were less than 10 μL). Reactions were not run at greater than 100 ppm oxygen as measured by a CAM-12 detector.

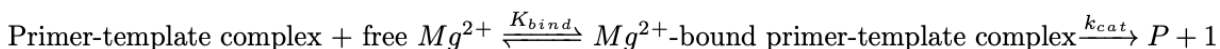| Sequence Name	Sequence (5' → 3')	Source
Primer	/5Cy3/rGrCrGrUrArGrArCrUrGrArCrUrG	IDT
GGG Template	rArGrGrGrCrArGrUrCrArGrUrCrUrArCrGrC	IDT
UUU Template	rArUrUrUrCrArGrUrCrArGrUrCrUrArCrGrC	IDT
A-end Primer	/5Cy3/rGrCrGrUrArGrArCrUrGrArCrUrA	IDT
2 Gap Template	rGrArGrUrUrArGrGrGrCrArGrUrCrArGrUrCrUrArCrGrC	IDT
Helper	rCrUrCrArArUrG	IDT
1 Gap Template	rGrArGrUrUrArGrGrCrArGrUrCrArGrUrCrUrArCrGrC	IDT
A-end template	rArGrGrGrUrArGrUrCrArGrUrCrUrArCrGrC	IDT
LNA template	A+G+G+G+CrArGrUrCrArGrUrCrUrArCrGrC	Exiqon/Qiagen
LNA 2 gap template	G+A+G+T+T+A+G+G+G+CrArGrUrCrArGrUrCrUrArCrGrC	
RNA pre-primer	rCrUrCrCrArG	In-house
^{13}C primer	rCrUrCrCrArGr^{13}C	In-house
DNA template	GAGGTCGCCGCTGGAG	In-house
DNA helper	CGACCTC	In-house
GC extension template	rArArArGrCrCrArGrUrCrArGrUrCrUrArCrGrC	IDT
Li primer	/56-FAM/CrGrCrUrCrGrArCrUrG	IDT
Li template	rGrCrGrCrCrUrCrArUrCrArGrUrCrGrArGrCrG	IDT
Li complement	CGCTCGACTGATGAGGCGC	IDT
Ligation primer	/5Cy3/rArGrUrGrArGrUrArArCrGrC	IDT
Ligation template	rUrCrArGrGrCrGrUrUrArCrUrCrArCrU	IDT
Ligator	2MeImprCrUrGrA	In-house
Activated trinucleotides for Li template	All 2AI activated: GAG, AGG, GGC, GCG, CGC, UGA	In-house

Table A.1. Oligonucleotide sequences used for experiments. (+) LNA, (r) RNA, otherwise DNA.

All quenched primer extension reaction samples were run on TBE-Urea 20% PAGE and analyzed with an Amersham Typhoon gel imager and quantified using the accompanying ImageQuant[TL] software. Bands were chosen using the peak picking tool and verified by eye, and we used the "Rolling Ball" subtraction to normalize each lane. Bands that migrated further than the primer are due to RNA degradation under the metal ion conditions, and were not included in quantification of the primer decay. Each lane was normalized to itself for calculating the percentage of primer remaining.

Kinetic Analysis

We determined the pseudo-first order reaction rate for each condition from the slope of linear fits of the natural logarithm of the ratio of the amount of unreacted primer remaining at a given time point to the initial amount of primer, plotted versus time. Each rate is calculated from the slope of three to five timepoints from one experiment. Only rates with R^2 over 0.8 for the linear fit were accepted for subsequent analysis, otherwise the reaction was repeated. Each experimental rate was then averaged together with three or more experimental replicates. Manual sampling at high metal ion concentrations was limited by human ability, with timepoints in the linear region of the burst phase at less than 10 seconds apart. These binding curves were fit using Prism non-linear fit (Michaelis-Menten type, single binding site) using entered mean, SD, and n. Standard deviations and means were calculated from experimental replicates, which were fit using Microsoft Excel.

The following kinetic scheme was used

$$\text{Primer-template complex} + \text{free } Mg^{2+} \xrightleftharpoons{K_{bind}} Mg^{2+}\text{-bound primer-template complex} \xrightarrow{k_{cat}} P+1$$

The primer-template complex is the RNA primer bound to the template saturated with imidazolium-bridged dinucleotide.

$$K_{bind} = \frac{k_{Mg,on}}{k_{Mg,off}} \textbf{ Equation 1}$$

The on/off rate for Mg^{2+} (10^{-8} s^{-1})[10] is much faster than the reaction step for primer extension, allowing us to fit to the following saturation binding curve.

$$k_{obs} = \text{primer extension rate} = [\text{primer/template complex}]_0 \frac{k_{cat}[Mg^{2+}]}{K_M + [Mg^{2+}]} \textbf{ Equation 2}$$

$$K_M = \frac{k_{Mg,off} + k_{cat}}{k_{Mg,on}} \textbf{ Equation 3}$$

A.5 NMR characterization of small molecules and oligonucleotides

All spectra were obtained using a Varian INOVA 400 MHz NMR spectrometer at 25 °C. When spectra were collected to determine purity, samples were dissolved in pure D_2O, and referenced to external 85% H_3PO_4 or internal OP(OMe)$_3$ for ^{31}P NMR, and HOD for ^1H NMR, and MeOD, acetone, or triethylamine for ^{13}C NMR. All spectra were analyzed using the MNova software.

A.6 High resolution liquid chromatography-mass spectrometry (LC-MS) characterization of small molecules and oligonucleotides

The monoisotopic mass (m/z) and purity for all synthesized RNA oligonucleotides and small molecules were assessed using an Agilent 1200 HPLC system equipped with a Waters 100 mm XBridge C18 column connected to an Agilent 6230 time of flight mass spectrometer as previously described.[3,11] This LC-MS system was equipped with a solvent degasser, auto-sampler, column oven, diode-array detector, and dual electrospray ionization source. All runs were performed with 50-100 pmole of samples. Negative ion mode was exclusively used due to the negatively charged nature of all the RNA compounds synthesized. The column utilized was a Waters 100 mm XBridge C18 column, 1 mm i.d., 3.5 μm particle size (Waters Corporation, Milford, MA) with the following solvent conditions: 200 mM HFIP, and 1.25 mM TEAA

at pH 7.0. Solvent B was methanol, and was ramped from 2.5% to 20% over 30 min at a flowrate of 0.1 mL/min.[12]

A.7 Solid phase RNA oligonucleotide synthesis, deprotection, and purification

RNA oligonucleotides were synthesized at the 50 μmol scale by solid-phase phosphoramidite polymerization chemistry on a MerMade 6(BioAutomation) RNA/DNA oligonucleotide synthesizer as previously described.[11] Purification using both C18 and anion-exchange preparative HPLC was also performed as previously described,[11] and then samples were washed with acetone and desiccated until pellet stopped shrinking to remove excess $NaClO_4$. Purity of oligonucleotides was assessed liquid chromatography-mass spectrometry (LC-MS) in negative mode as described in A.6.

A.8 Metal stock solution preparation

For metal concentration series, stock solutions of 4M and 2.5 M $MgCl_2$ were made from the solid anhydrous salt (note: these reactions are exothermic and care must be taken), and serial dilutions were performed to prepare the necessary concentrations for the experiments. Serially diluted working stocks were prepared fresh from the high concentration stocks for every experiment. $MnCl_2$ and $CaCl_2$ (Fisher/Alfa Aesar) were both purchased as 1M stock solutions and subsequently serially diluted. A 250 mM stock solution of $Co(NH_3)_6Cl_3$ was made from the solid and pH adjusted. Cobalt chloride hexahydrate (Alfa Aesar) was made into a 250 mM stock solution from the solid and pH adjusted to 7.0 by addition of NaOH.

Table A.2: Quench times

Concentrations (mM)	Mg pH=8	Mn pH=7.5	Ca pH=7.5	Mg pH=7.5	Mg pH=6	Mg pH=9	Helper	No Helper	NaCl	CoHex	Constant Ionic Strength	AA	A end	LNA helper	Mn no helper	Mn helper
0-5	3,15,30,51 m	3,15,30,45m	3m,30m,1.5h,2h	3,30m,1,1.5h	3m,1,2,3 h	3,15,30,45m	3,15,30,60m	1,2,3,15m	35m,1,1.5,2h	3m,1,2,3h	3,15,45,90m	1,2,3,4 hr	3,30m,1,1.5h	3,15,45,60m	3,14,30,45.5 m	3,10,20,30m
10 thru 60	1,2,3,5 m	1,2,3,5 m	3,30m,1,1.5h	3,15,30,45 m	1,2,3,10m	1,2,3,5m	1,2,3,15m	1,2,3,5m	35m,1,1.5,2h	3m,1,2,3h	1,2,3,5m	1,1.5,2.5,3.5	1,2,3, 5m	1,2,3,5 m	1,2,3,5 m	30,60,90,120s
80- 200	30,60,90,120s	30,60,90,120s	3,15,30,45 m	1,2,3,10m	3,10,30,60m	3,10,30,60m	30, 60, 90, 120s	0.5,1,1.5, 2m	35m,1,1.5,2,3h	3m,1,2,3h	30,60,90,120s	0.75, 1.5,2,3	30,60,90,120s	30,60,90,120s	30,60,90,120s	30,60,90,120s
250-400	30,60,90,120s	30,60,90,120s	1,2,3,5m	1,2,3,5m	1,2,3,30m	1,2,3,30m	0.5,1,1.5, 2m	0.5,1,1.5, 2m	35m,1,1.5,2,3h	3m,1,2,3h	30,60,90,120s	0.75,1,1.5,2	30,60,90,120s	30,60,90,120s	30,60,90,120s	30,60,90,120s
500+	30, 60, 90, 120s			30,60,90,120s	30, 60, 90, 120s	1, 2, 3,32 min	15,25, 35, 60s		3,30m,1.5,2h							

Reaction times ultimately chosen after long time point reaction series were performed to have more than 50% of primer left in first time point if possible (not the case in 500+ mM), so as to obtain accurate initial rates. Rates have been obtained in other time frames as well and verified to be the same as those in the final figures.

Bibliography

(1) Cavaluzzi, M. J.; Borer, P. N. Revised UV Extinction Coefficients for Nucleoside-5′-monophosphates and Unpaired DNA and RNA. *Nucleic Acids Research* **2004**, *32* (1), e13. https://doi.org/10.1093/nar/gnh015.

(2) Owczarzy, R.; Tataurov, A. V.; Wu, Y.; Manthey, J. A.; McQuisten, K. A.; Almabrazi, H. G.; Pedersen, K. F.; Lin, Y.; Garretson, J.; McEntaggart, N. O.; Sailor, C. A.; Dawson, R. B.; Peek, A. S. IDT SciTools: A Suite for Analysis and Design of Nucleic Acid Oligomers. *Nucleic Acids Research* **2008**, *36* (suppl_2), W163–W169. https://doi.org/10.1093/nar/gkn198.

(3) Walton, T.; Pazienza, L.; Szostak, J. W. Template-Directed Catalysis of a Multistep Reaction Pathway for Nonenzymatic RNA Primer Extension. *Biochemistry* **2019**, *58* (6), 755–762. https://doi.org/10.1021/acs.biochem.8b01156.

(4) Ding, D.; Zhou, L.; Giurgiu, C.; Szostak, J. W. Kinetic Explanations for the Sequence Biases Observed in the Nonenzymatic Copying of RNA Templates. *Nucleic Acids Res* **2021**, *50* (1), 35–45. https://doi.org/10.1093/nar/gkab1202.

(5) Walton, T.; Szostak, J. W. A Kinetic Model of Nonenzymatic RNA Polymerization by Cytidine-5′-Phosphoro-2-Aminoimidazolide. *Biochemistry* **2017**, *56* (43), 5739–5747. https://doi.org/10.1021/acs.biochem.7b00792.

(6) Walton, T.; Szostak, J. W. A Highly Reactive Imidazolium-Bridged Dinucleotide Intermediate in Nonenzymatic RNA Primer Extension. *Journal of the American Chemical Society* **2016**, *138* (36), 11996–12002. https://doi.org/10.1021/jacs.6b07977.

(7)   Walton, T.; Pazienza, L.; Szostak, J. W. Template-Directed Catalysis of a Multistep Reaction Pathway for Nonenzymatic RNA Primer Extension. *Biochemistry* **2019**, *58* (6), 755–762. https://doi.org/10.1021/acs.biochem.8b01156.

(8)   Lin Jin, Aaron E. Engelhart, Weicheng Zhang, Katarzyna Adamala, and Jack W. Szostak. Catalysis of Template-Directed Nonenzymatic RNA Copying by Iron(II). *JACS 140* (44), 15016–15021. https://doi.org/10.1021/jacs.8b09617.

(9)   Walton, T.; Szostak, J. W. A Kinetic Model of Nonenzymatic RNA Polymerization by Cytidine-5′-Phosphoro-2-Aminoimidazolide. *Biochemistry* **2017**, *56* (43), 5739–5747. https://doi.org/10.1021/acs.biochem.7b00792.

(10)   Neely, James.; Connick, Robert. Rate of Water Exchange from Hydrated Magnesium Ion. *Journal of the American Chemical Society* **1970**, *92* (11), 3476–3478. https://doi.org/10.1021/ja00714a048.

(11)   Tam, C. P.; Zhou, L.; Fahrenbach, A. C.; Zhang, W.; Walton, T.; Szostak, J. W. Synthesis of a Nonhydrolyzable Nucleotide Phosphoroimidazolide Analogue That Catalyzes Nonenzymatic RNA Primer Extension. *J. Am. Chem. Soc.* **2018**, *140* (2), 783–792. https://doi.org/10.1021/jacs.7b11623.

(12)   Apffel, A.; Chakel, J. A.; Fischer, S.; Lichtenwalter, K.; Hancock, W. S. Analysis of Oligonucleotides by HPLC−Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **1997**, *69* (7), 1320–1325. https://doi.org/10.1021/ac960916h.

Appendix B

Methods

Refer to Appendix A for general methods involved in these studies. Specific methods used in Chapter 2 are further elaborated here.

Primer extension with varying pH

For pH curves, the following buffer conditions were used: MES (5.5-6.5), Tris (7-9), CHES (9.5-10). At the edge of buffer conditions, the two buffers that could reach that pH were tested in primer extension at the same pH and the reaction pH was measured before and after the experiments. The buffer that led to the least change in pH throughout primer extension was utilized at that pH.

Solvent kinetic isotope effects in primer extension

Primer extension is performed as described in Appendix A, but all buffers, salts and oligomers are dissolved in D_2O and lyophilized down three times to ensure full deuterium exchange of exchangeable protons.

NMR determination of Cp*pC formation and hydrolysis

All hydrolysis spectra were obtained using a Varian INOVA 400 MHz NMR spectrometer at 25 °C in 90% H_2O, 10% D_2O. Bridged dinucleotide formation spectra were collected in 100% H_2O in a Shigemi tube with a D_2O insert. All spectra were analyzed using the MNova software. Quantifying the decay and formation of the bridged dinucleotide in presence of metal ions performed in triplicate by taking ^{31}P spectra with 256 scans every 7.5 min. Each sample contained external 85% H_3PO_4 (δ = -3.3 ppm) for ^{31}P (161 MHz) reference. For kinetics, the concentration of the molecule of interest was normalized over the total area under the curve of all phosphate peaks. Concentration was then determined using calculated starting concentration

from Nanodrop and calculated extinction coefficients, and multiplied by the fractional normalized amount of the molecule of interest.[1,2]

Analytical HPLC analysis of bridged dinucleotide hydrolysis

Samples were analyzed using an Agilent 1100 series HPLC system with a 4.6mm x 250 mm, 6 μm Varitide C18 column, autosampler, solvent degasser and automatic fraction collector. Absorbance of samples were monitored at 254 nm, with a reference at 360 nm under a gradient of 20 mM triethylamine bicarbonate buffer (pH 7.5) versus acetonitrile (3% ACN from 0-7 minutes, 3-50% from 7-27 minutes, 50-100% ACN from 27-29 min, 100% from 29-33 min, 100-3% from 33-34 min, 3% from 34-40 min). To monitor the decay of Cp*pC over time, elution profiles for standards of CMP (7 min), the pyrophosphate CppC (7.5-8 min), Cp*pC (17 minutes), *pC (10 minutes) were established under the gradient conditions. The absorbance at 254 nm was then utilized to calculate the concentration through the following procedure. The area under each peak from a given run was calculated using Agilent ChemStation software. Then, each peak was normalized to the total measured area, and the first order decay kinetics were calculated using the normalized decay of Cp*pC with timepoints taken roughly every 2 hours by the autosampler (between each data run was a blank run).

Synthesis and characterization of $[Co(NH_3)_5OH_2]^{3+}Cl_3$

Cobalt pentaamine chloride was obtained from Sigma Aldrich and dissolved in water for three weeks at room temperature until a color change (purple to red) was observed.[3] Although the species exist in equilibrium, the vast excess of water favored the formation of the cobalt pentaamine hydroxo species over several weeks. The solution was then characterized in a 1cm cuvette using a Nanodrop 2000c Spectrophotometer. Peaks were observed at 344 and 490 nm, corresponding to the published values of the complex (500 nm). The chloride complex absorbs at

535 nm, and was not observed to have substantially changed the profile of the spectra.
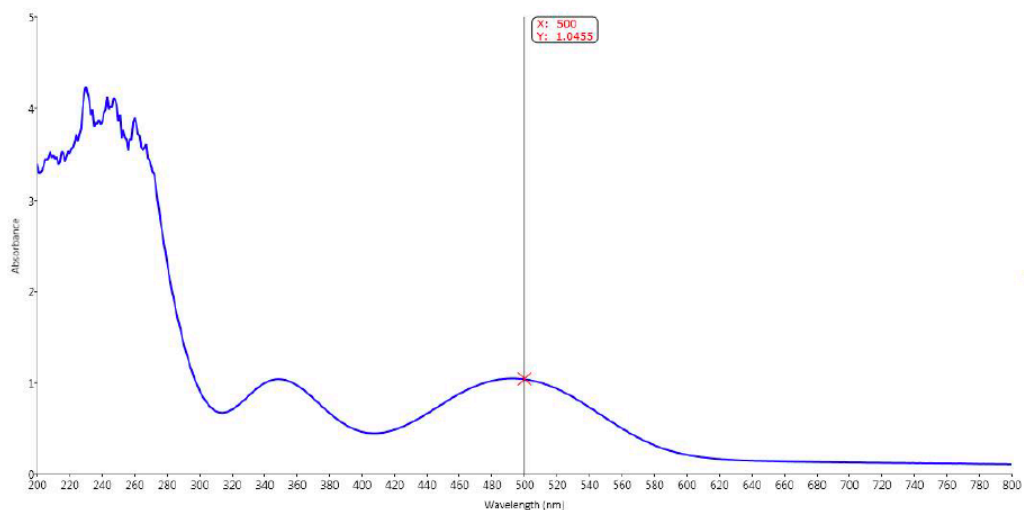
Figure B.1. UV-VIS Spectrum of $[Co(NH_3)_5OH_2]^{3+}Cl_3$ solution.

Bibliography

(1)   Owczarzy, R.; Tataurov, A. V.; Wu, Y.; Manthey, J. A.; McQuisten, K. A.; Almabrazi, H.

         G.; Pedersen, K. F.; Lin, Y.; Garretson, J.; McEntaggart, N. O.; Sailor, C. A.; Dawson, R.

         B.; Peek, A. S. IDT SciTools: A Suite for Analysis and Design of Nucleic Acid

         Oligomers. *Nucleic Acids Research* **2008**, *36* (suppl_2), W163–W169.

         https://doi.org/10.1093/nar/gkn198.

(2)   Cavaluzzi, M. J.; Borer, P. N. Revised UV Extinction Coefficients for Nucleoside-5′-

         monophosphates and Unpaired DNA and RNA. *Nucleic Acids Research* **2004**, *32* (1), e13.

         https://doi.org/10.1093/nar/gnh015.

(3)   Langford, C. H.; Muir, W. R. Thermodynamics of the Formation of Co (NH3)5Cl2+ from

         Co (NH3)5OH23+: Separation of Environmental Effects. *J. Phys. Chem.* **1967**, *71* (8),

         2602–2605. https://doi.org/10.1021/j100867a032.

Appendix C

Supplemental Materials for Chapter 3

Methods

Synthesis and purification of *p$_S$G and Gps*pC

1 mmol of N2-isobutyryl-2′,3′-diacetylguanosine was charged into a flask with positive

pressure Ar (g) in a Schlenk line setup. Anhydrous acetonitrile was used to dissolve the white

solid completely. 288 μL of 2,6 lutidine was subsequently added to the mixture, and then a slow,

drop-wise addition of 101 μL PSCl$_3$ (caution: TOXIC) with an Ar(g) headspace in the needle.

White vapor formation was observed in this process. This process was repeated until the solution

when measured by ESI-MS of reaction aliquots showed the masses for the product (2 Cl

addition, m/z= 569, one OH/one Cl m/z=551) and reactant depletion (m/z=437). This usually

took between 3 and 6 PSCl$_3$/ 2,6-lutidine additions. The reaction mixture was then cooled to 0°C

using an ice bath and solid 2-methyl imidazole was added in excess (985.2 mg). The reaction

was then stirred at room temperature for 15 minutes, forming some white solid, and re-cooled

with the ice bath to add 10 mL 500 mM triethylammonium bicarbonate solution and minimal

acetonitrile to re-dissolve any solid, bubbling observed. The solution was then re-exposed to air

and degassed via sonification. The sample was split into several injections and diluted with water

for flash chromatography. The mobile phase was 20 mM triethylammonium bicarbonate at

pH=7.5, and the organic phase was acetonitrile. The organic phase gradient was 0-40% over 20

column volumes. The fractions that corresponded to the masses for compounds **2** and **3** (Scheme

1) (m/z=596 and 661 respectively). The fractions of **2** and **3** were rotovapped separately to form

beige-white solids. **3** was subjected to hydrolysis in a refrigerator for one day to remove the

additional methylimidazole added to the phosphorus and form **2. 2** was dissolved with minimal

methanol in a heavy-walled round bottom pressure flask, and 20 mL NH$_4$OH was added. The vessel was closed tightly and reacted for 3 hours in a 65 °C oil bath. After the 3 hours, the reaction vessel was equilibrated to room temperature for 30 minutes to an hour. The reaction was then carefully rotary evaporated, and resuspended in 10 mL 250 mM TEAB, and filtered through a cotton plug. This solution was flash purified using reverse phase flash chromatography over 20 CV with an organic gradient 0-15%. The desired product was collected around 10CV (m/z= 442), and lyophilized at room temperature overnight. The sample is then re-dissolved in water and exchanged with excess 2-aminoimidazole hydrochloride for 1 day at pH=7.5, and purified using the previous flash chromatography conditions. The weight of desired product **7** is 444. This synthetic protocol also applies to creating cytidine 2-aminothiophosphoroimidazolides, however the diastereomers of the cytidine species are unable to be resolved by chromatography, and therefore were not of further interest for our studies. All products at this stage were racemic, and diastereomers were subsequently purified of **7** using preparative HPLC. Separation of the diastereomers was near-baseline, but often took 2-3 purification cycles to completely remove all of the opposing diastereomer from the samples.

Due to the different diastereomers of *p$_S$G having the same mass, purity is measured by ^{31}P NMR exclusively, with peak assignments (Tables C.1, C.2, C.3). Diastereomer 1 and 2 are denoted as such due to their order in elution through preparative chromatography- Diastereomer 1 elutes 1st, diastereomer 2 elutes 2nd, and this is robust to elution conditions as verified through NMR.

Gps*pC (D1 and D2) syntheses were performed with a 1:4 molar ratio of the diastereomerically pure AIp$_S$G: CpOAt in water at pH=8 (as adjusted using minute amounts of NaOH and HCl) for 1 hour at room temperature.[1] This procedure was also utilized to make

195

Gp*pC standards. The reaction was purified by preparatory HPLC on an Agilent Eclipse XDB C18 column (250 mm, 21.2 mm i.d., 7 μm particle size). The aqueous mobile phase was 20 mM aqueous triethylammonium bicarbonate buffer at pH 7.5, and the organic mobile phase was acetonitrile. The gradient of the organic mobile phase was 5% to 5.5% 0-18 min, 5.5-6% 18-24 min, 6-25% 24-55 min at 10 mL/min flow rate. Gp_S*pC eluted at 7-8% acetonitrile, 35 min, and flash-frozen using liquid nitrogen and placed on lyophilizer until a dry, fluffy white solid. Products were verified by ESI-MS before lyophilization in each purification.

Characterization of thiophosphoroimidazolides

Samples post lyophilization are dissolved in 400μL D_2O, and put in a 5mm Shigemi tube with relevant external reference inserts. [1]H and proton-decoupled ^{13}C and ^{31}P NMR spectrum were collected on a Varian Inova 400 MHz spectrometry equipped with a broadband PFG (z-gradient) probe (161 MHz for ^{31}P). Proton, carbon, and phosphorus chemical shifts are reported in parts per million (ppm) on the δ scale. The proton spectrum is referenced to DHO (δ=4.79 ppm).[2] Phosphorus spectra are referenced to external phosphoric acid (85%), carbon to internal MeOD or triethylamine shifts. All NMR spectra were collected at 25° C. Data are reported as follows: chemical shift, multiplicity (s=singlet, d=doublet, t=triplet, q=quartet, m=multiplet, br=broad), and integration. Refer to tables for assignment of peaks to compounds.

Computational studies $GpAIp_sG$

Models were initiated from crystal structure (PDB 6C8E), both diastereomeric oxygens were substituted for S, and conformational searching performed, followed by DFT on stationary points at the B3LYP/6-31G(d) level of theory in PCM (water) with and without D3BJ dispersion correction.[3]

Structural characterization of AIp_sG through crystallization

With limited solubility (and immediate precipitate formation) in organic and polar solvents, AIp_sG posed a difficulty for conventional small molecule crystallography. Therefore, less conventional methods were sought in order to characterize their stereochemistry at the phosphate. MicroED was attempted extensively with solid applied to grids in a number of different ways. While some crystals resulted in solutions from direct methods, none resulted in structures that made chemical sense. Additionally, for crystallography, soaking conditions did not lead to ordered structures of the monomers, so only product structures were collected.

1 mM RNA sequences (Sequences are listed in Table C.6) were mixed with 50 mM phosphorothioate G monomer or 20 mM phosphorothioate GA dimer at 1:1 ratio. The samples were heated up to 90 °C for 2 min and then slowly cooled to room temperature. Then magnesium chloride was added at the final concentration of 20 mM. The sample was incubated at room temperature for 72 hours before screening. Crystal Screen HT, Index HT, Natrix HT (Hampton Research, Aliso Viejo, CA) and Nuc-Pro HTS (Jena Bioscience, Jena, Germany) were used to screen crystallization conditions at 20 °C using the sitting-drop vapor diffusion method. An NT8 robotic system and Rock Imager (Formulatrix, Waltham, MA) were used for crystallization screening and monitoring the crystallization process. Optimal crystallization conditions are listed in Table C.7.

<u>Data collection, structure determination and refinement</u>

Diffraction data were collected at a wavelength of close to 1 Å (exact values are in Table A.8) under a liquid nitrogen stream at 98 K on Beamline 822 at the Advanced Light Source in the Lawrence Berkeley National Laboratory (USA) or Beamline 23-ID-B at the Advanced Photon Source in the Argonne National Laboratory (USA). The crystals were exposed for 1s per image with a 1 Å oscillation angle (CCD) or 0.2 s per image with a 0.2 Å oscillation angle

(Pilatus). The distances between detector and the crystal were set to 180–300 mm. The data were processed by HKL2000. The structures were solved by molecular replacement by PHASER using structure of 6C8O as the searching model. All structures were refined by Refmac. After several cycles of refinement, some water molecules and metal atoms were added in Coot. Data collection, phasing, and refinement statistics of the determined structures are listed in Table C.8 and Table 1 in Chapter 3.

Table C.1. 2MeImₛG racemate

¹H	¹H NMR (400 MHz, d₂o) δ 7.27 (dt, J = 6.7, 2.0 Hz, 1H), 6.74 (q, J = 1.7 Hz, 1H), 5.85 (dd, J = 6.0, 2.8 Hz, 1H), 4.68 (dt, J = 17.5, 5.6 Hz, 1H), 4.39 (dd, J = 6.1, 2.9 Hz, 1H), 4.25 (td, J = 3.8, 2.1 Hz, 1H), 4.03 (dtd, J = 10.1, 6.7, 3.3 Hz, 2H), 2.89 (q, J = 7.3 Hz, 16H), 2.41 (s, 3H), 1.09 (t, J = 7.3 Hz, 24H).
¹³C	¹³C NMR (101 MHz, d₂o) δ 167.96, 152.63, 136.37, 136.29, 126.46, 126.35, 123.33, 123.26, 118.48, 87.39, 87.36, 74.51, 74.44, 71.70, 71.54, 47.25(**TEAH+**), 15.49, 15.46, 9.53(**TEAH+**).
³¹P	³¹P NMR (162 MHz, d₂o) δ 44.1, 44.04, -3.3 H₃PO₄ 85% external reference).
ESI-MS	Exact mass (442.07), Observed mass (442)
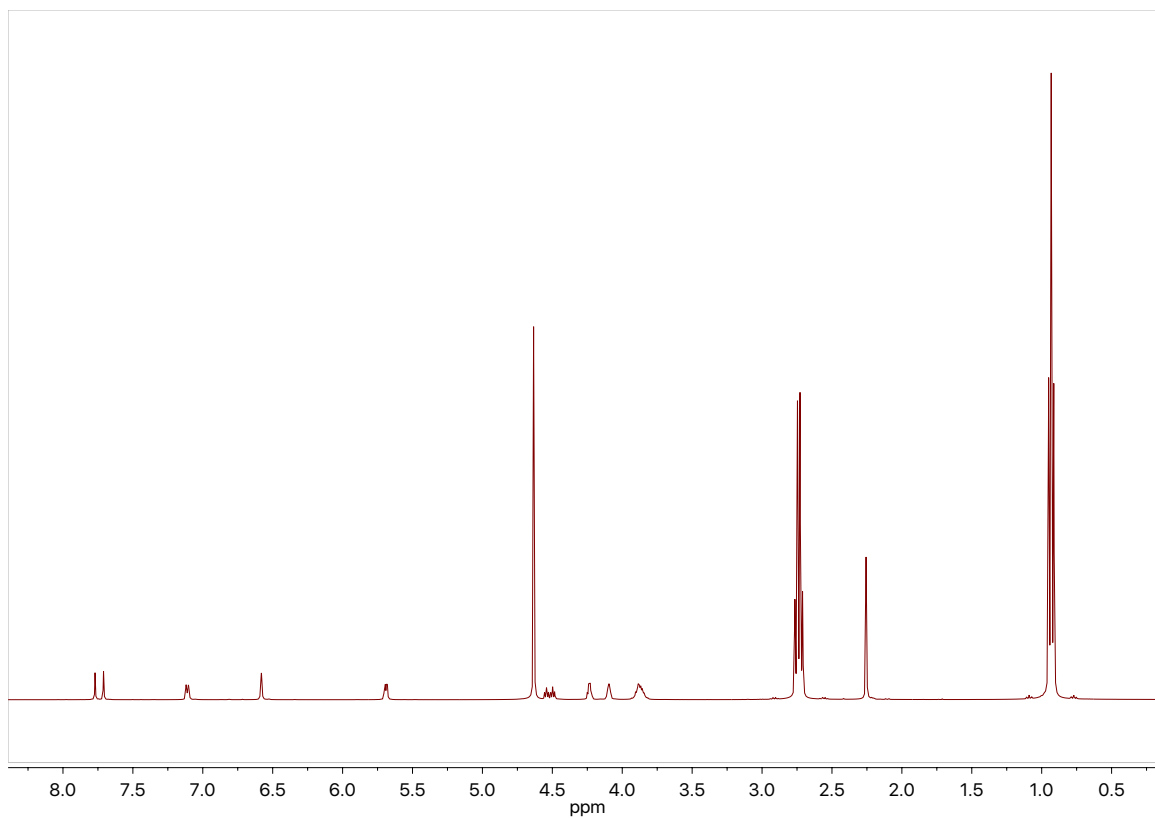
Figure C.1. ^{1}H NMR spectrum of 2MeImp$_s$G racemate. Excess triethylamine is observed from purification procedure. Refer to Table C.1 for peak assignments.
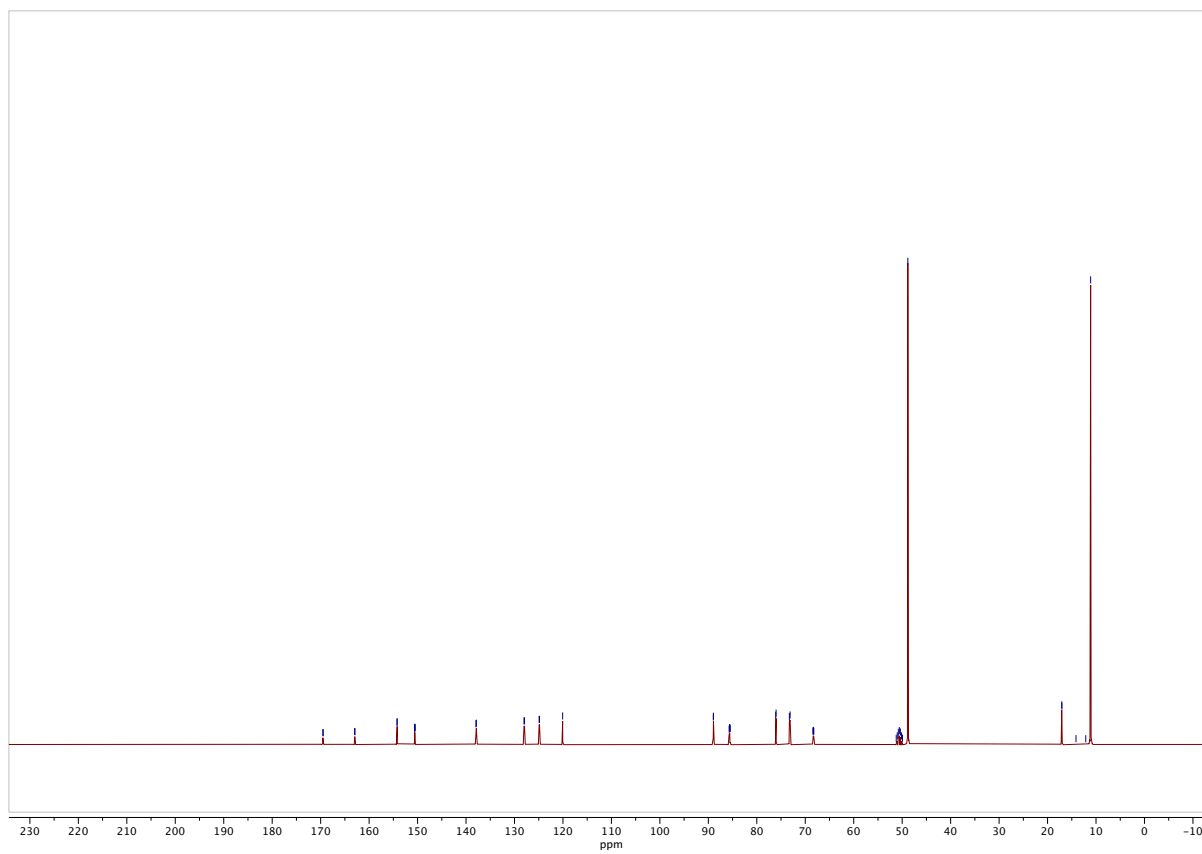
Figure C.2. ^{13}C NMR spectrum of 2MeImp$_s$G racemate. Excess triethylamine is observed from purification procedure. Refer to Table C.1 for peak assignments.
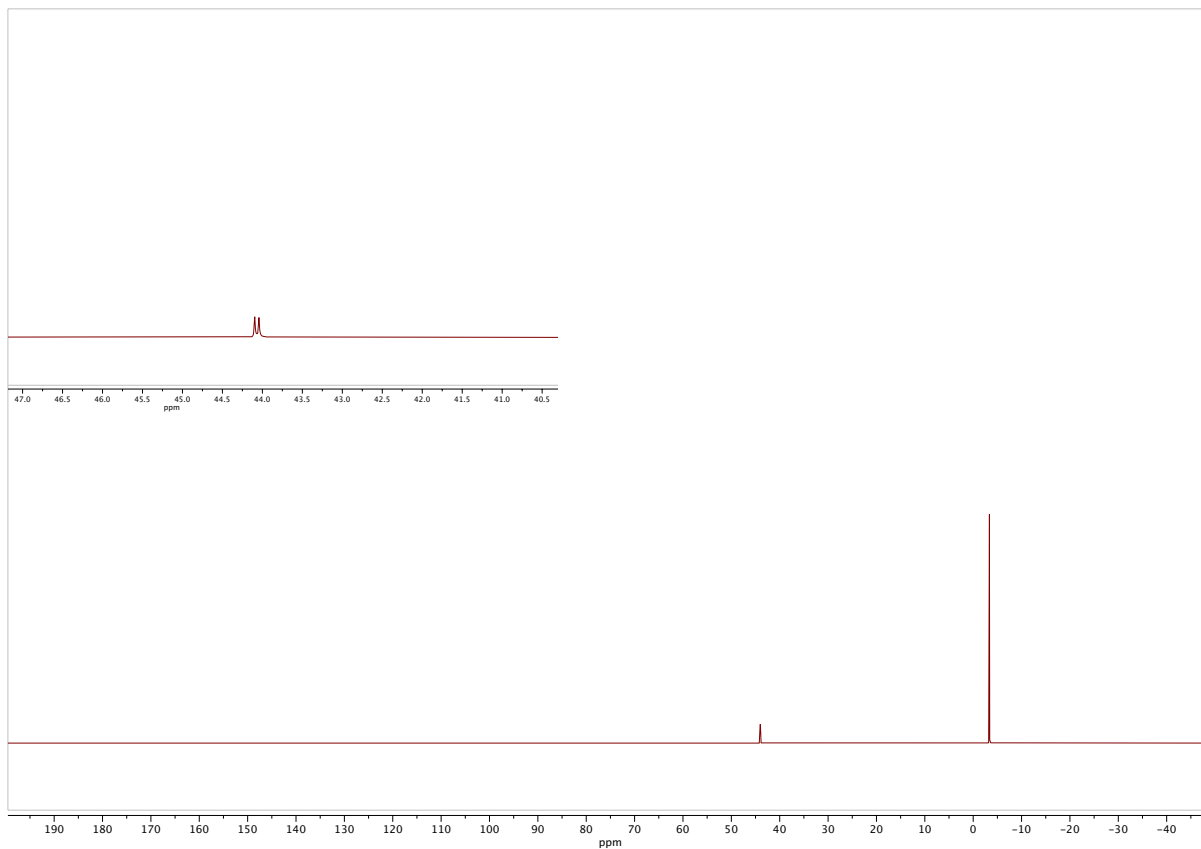
Figure C.3. ^{31}P NMR spectrum of 2MeImp$_s$G racemate. Spectrum of 2MeImp$_s$G racemate post flash chromatography purification with full spectrum and inset displaying the two peaks of the different diastereomers. Refer to Table C.1 for peak assignments.

Table C.2. 2AmImpsG Diastereomer 1

	^1H	^{1}H NMR (400 MHz, d$_2$o) δ 7.96, 6.81, 6.80, 6.51, 6.50, 5.90, 5.89, 4.83, 4.79, 4.53, 4.52, 4.51, 4.50, 4.32, 4.21, 3.21 (qd, $J = 7.4$, 2.1 Hz, 1H), 1.29 (td, $J = 7.4$, 2.1 Hz, 1H).
	^{13}C	^{13}C NMR (101 MHz, d$_2$o) δ 218.02(acetone), 162.18, 140.11, 118.71, 118.64, 89.84, 86.00, 85.90, 76.03, 73.12, 68.22, 68.15, 49.24(TEAB+), 32.81, 10.81(TEAB+).
	^{31}P	^{31}P NMR (162 MHz, d$_2$o) δ 43.84, -3.3 (H$_3$PO$_4$ external reference)
	ESI-MS	Exact mass (443.07), Observed mass (443)
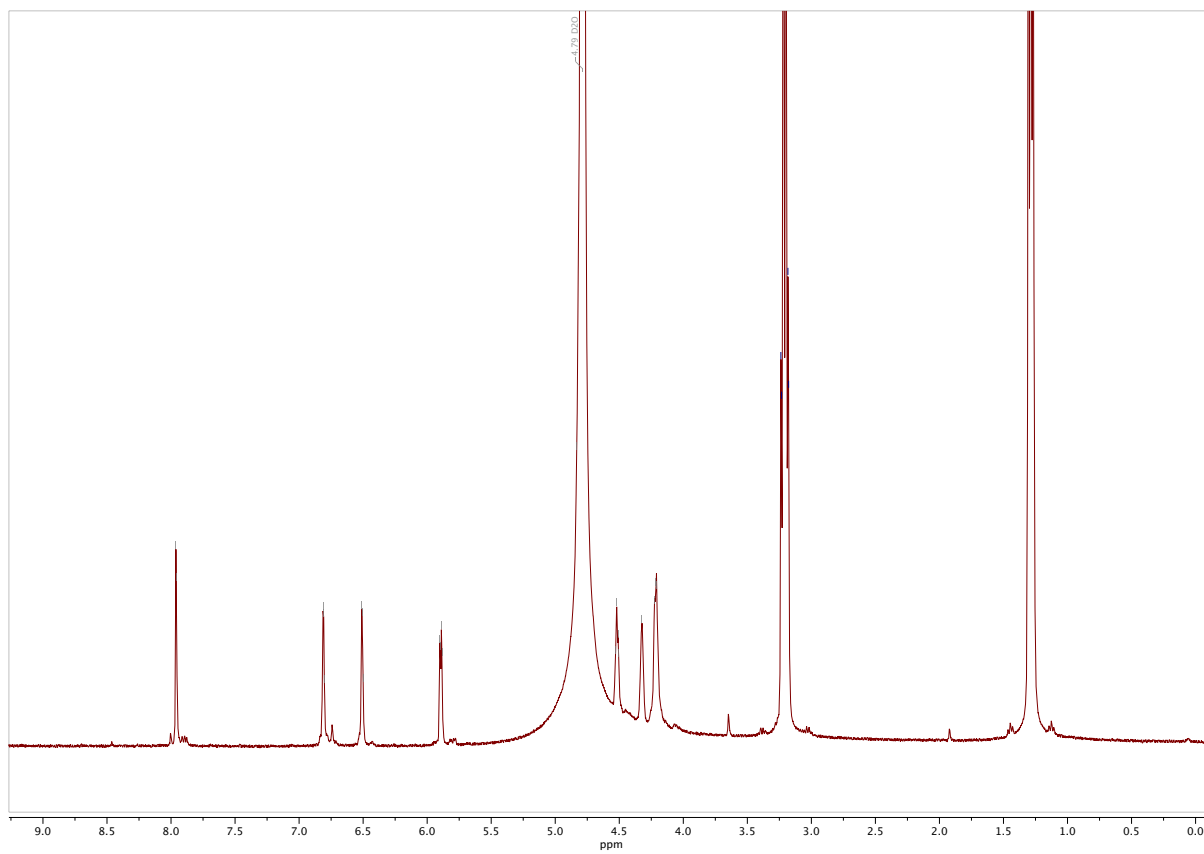
Figure C.4. ^{1}H NMR spectrum of 2AmImp$_s$G diastereomer 1. Refer to Table C.2 for peak assignments.
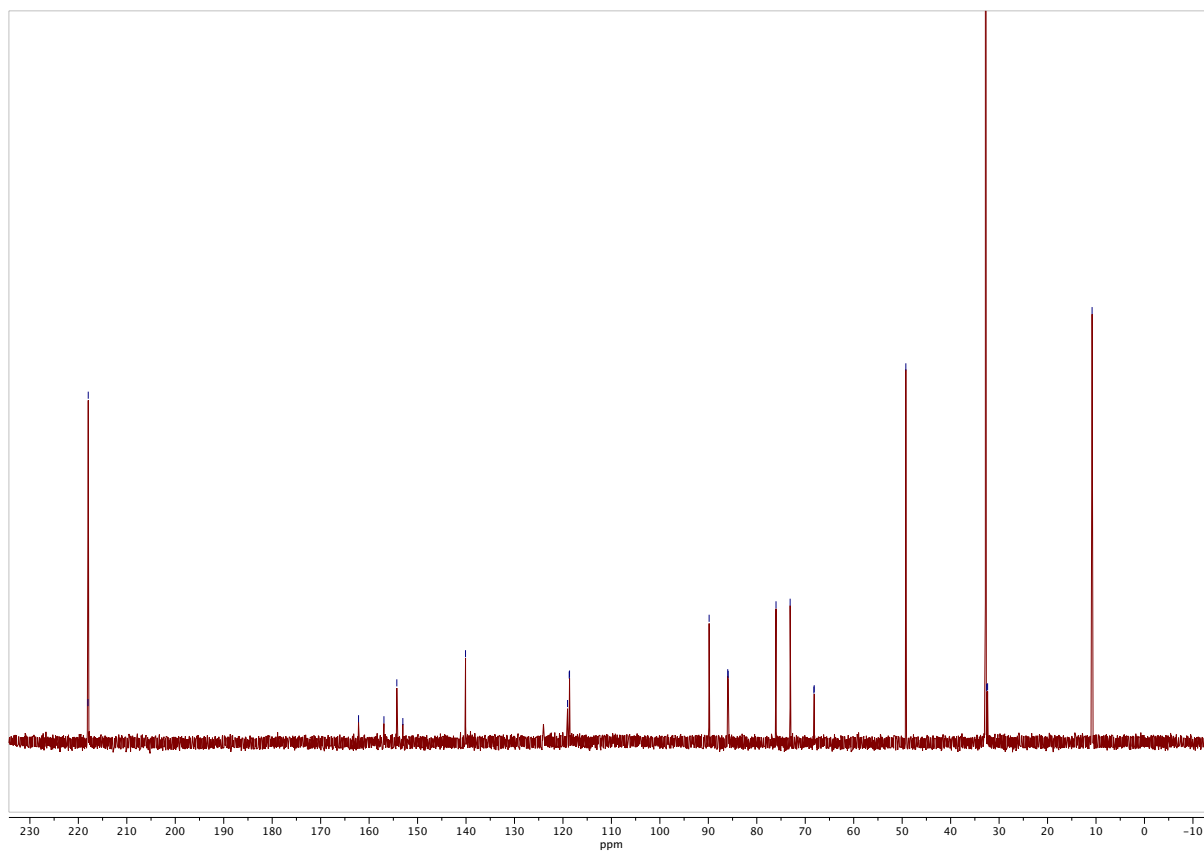
**Figure C.5**. ^{13}C NMR spectrum of 2AmImp$_s$G diastereomer 1. Refer to Table C.2 for peak assignments.
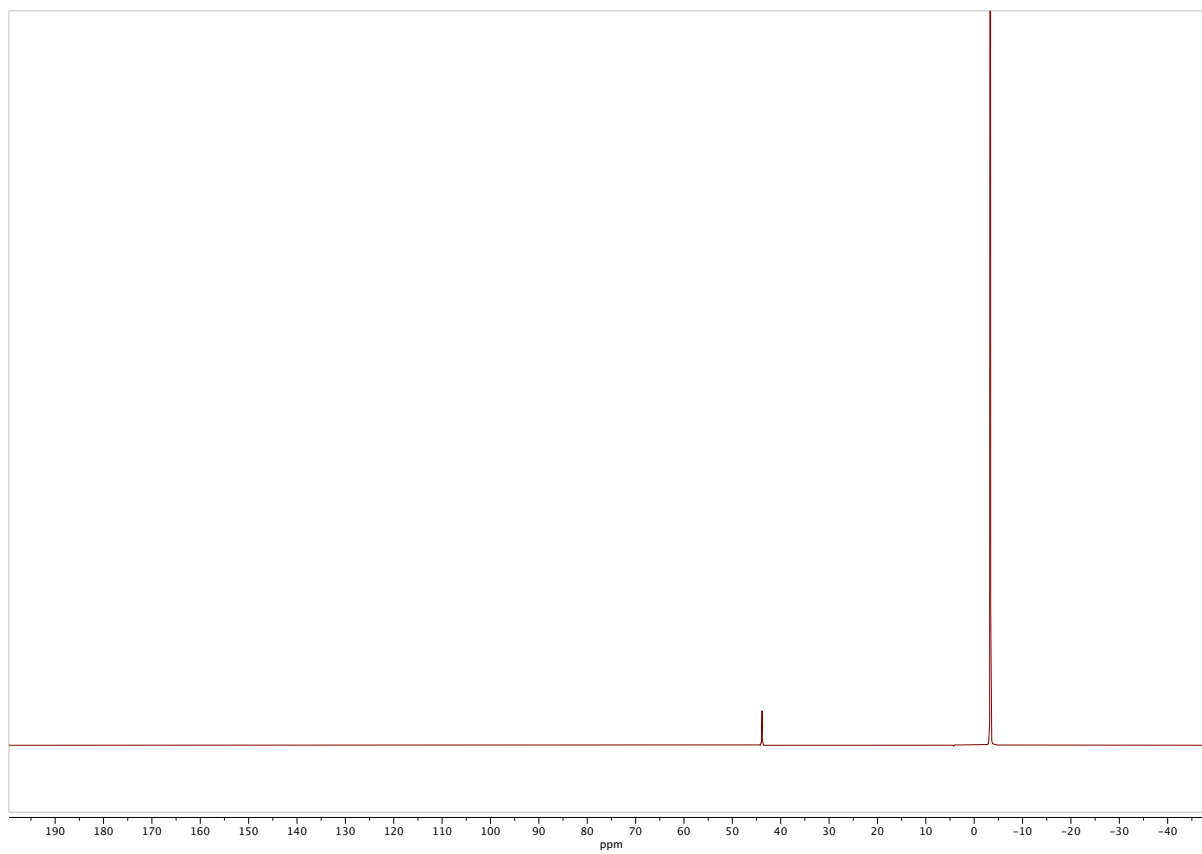
**Figure C.6.** ^{31}P NMR spectrum of 2AmImp$_s$G diastereomer 1. Refer to Table C.2 for peak assignments.

Table C.3. 2AmImpsG Diastereomer 2

^1H	^{1}H NMR (400 MHz, d$_2$o) δ 7.96 (d, J = 2.1 Hz), 6.81 (q, J = 2.2 Hz), 6.51 (q, J = 2.2 Hz), 5.89 (dd, J = 5.2, 2.1 Hz), 4.79 (s, 5H), 4.52 (q, J = 3.4 Hz), 4.32 (s), 4.24 – 4.16 (m), 3.21 (qd, J = 7.4, 2.1 Hz, TEA+), 1.29 (td, J = 7.4, 2.1 Hz, TEA+).
^{13}C	^{13}C NMR (101 MHz, d$_2$o) δ 160.78, 152.14, 138.35, 116.96, 116.42, 116.37, 88.16, 83.72, 83.62, 73.80, 70.87, 66.50, 47.19(TEA+), 8.74(TEA+).
^{31}P	^{31}P NMR (162 MHz, d$_2$o) δ 43.67, -3.3 ppm (H$_3$PO$_4$ external reference)
ESI-MS	Exact mass (443.07), Observed mass (443)
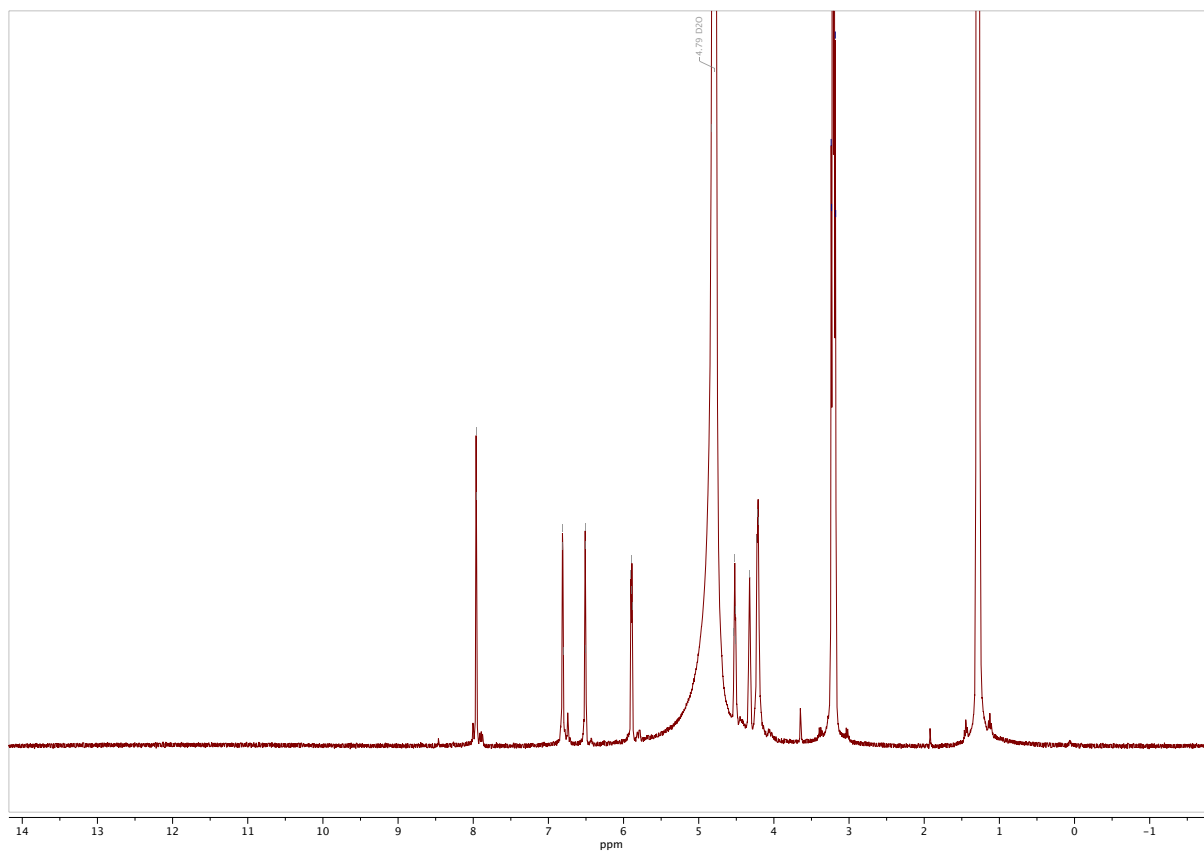
Figure C.7. ^{1}H NMR spectrum of 2AmImp$_s$G diastereomer 2. Refer to Table C.3 for peak assignments.
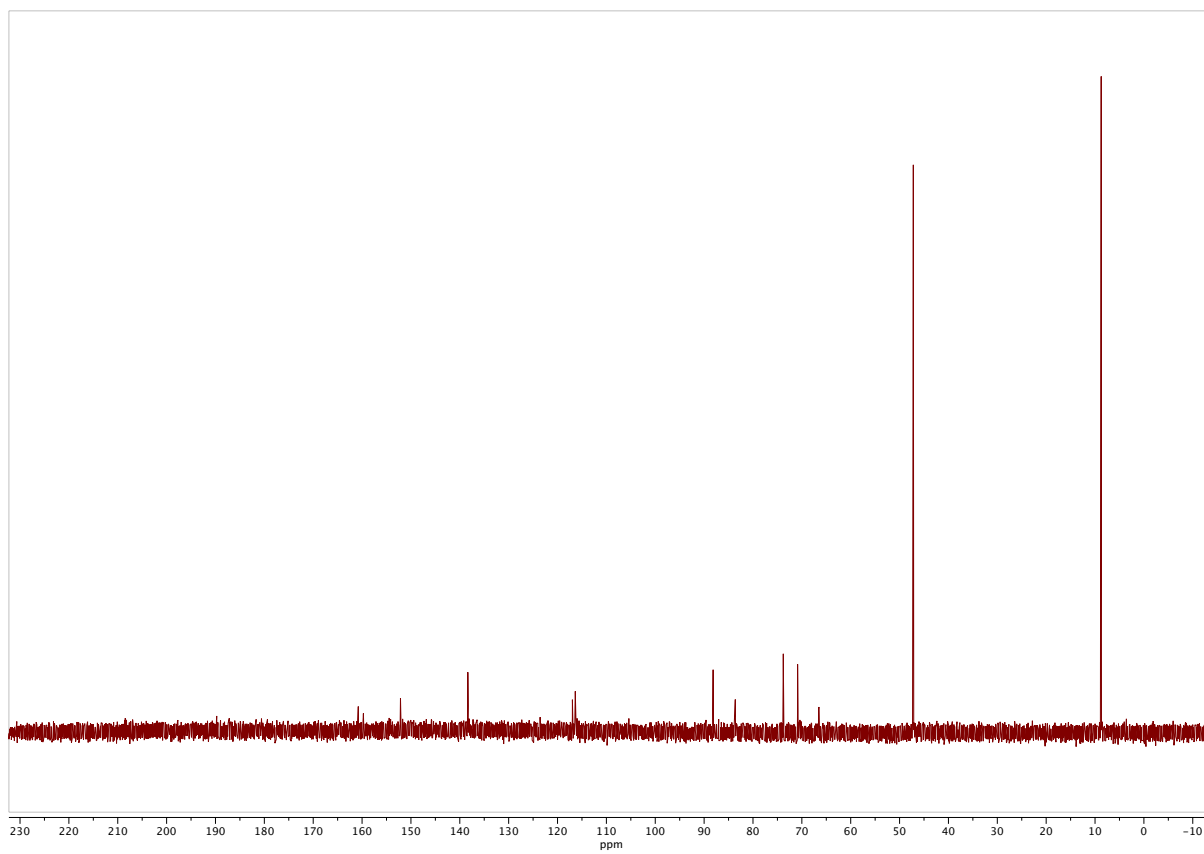
Figure C.8. ^{13}C NMR spectrum of 2AmImp$_s$G diastereomer 2. Refer to Table C.3 for peak assignments.
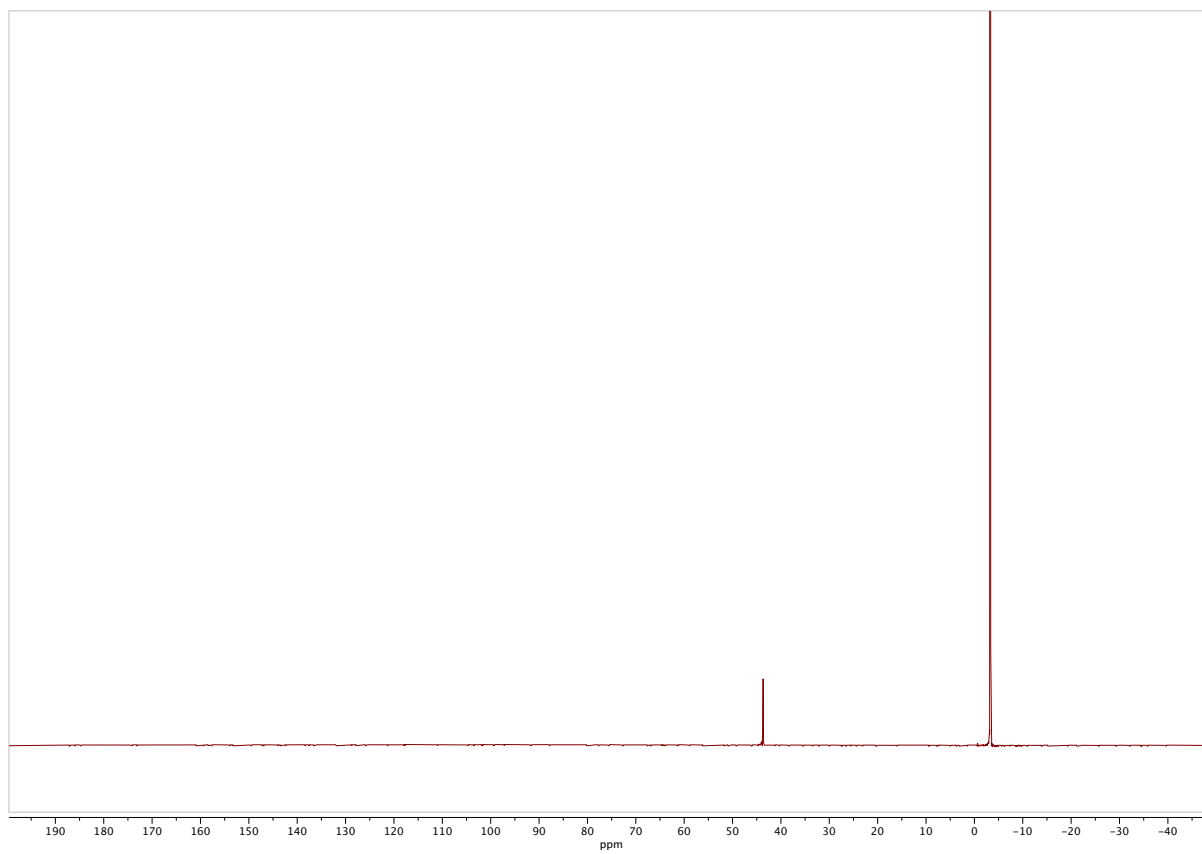
**Figure C.9.** ^{31}P NMR spectrum of 2AmImp$_s$G diastereomer 2. Refer to Table C.3 for peak assignments.
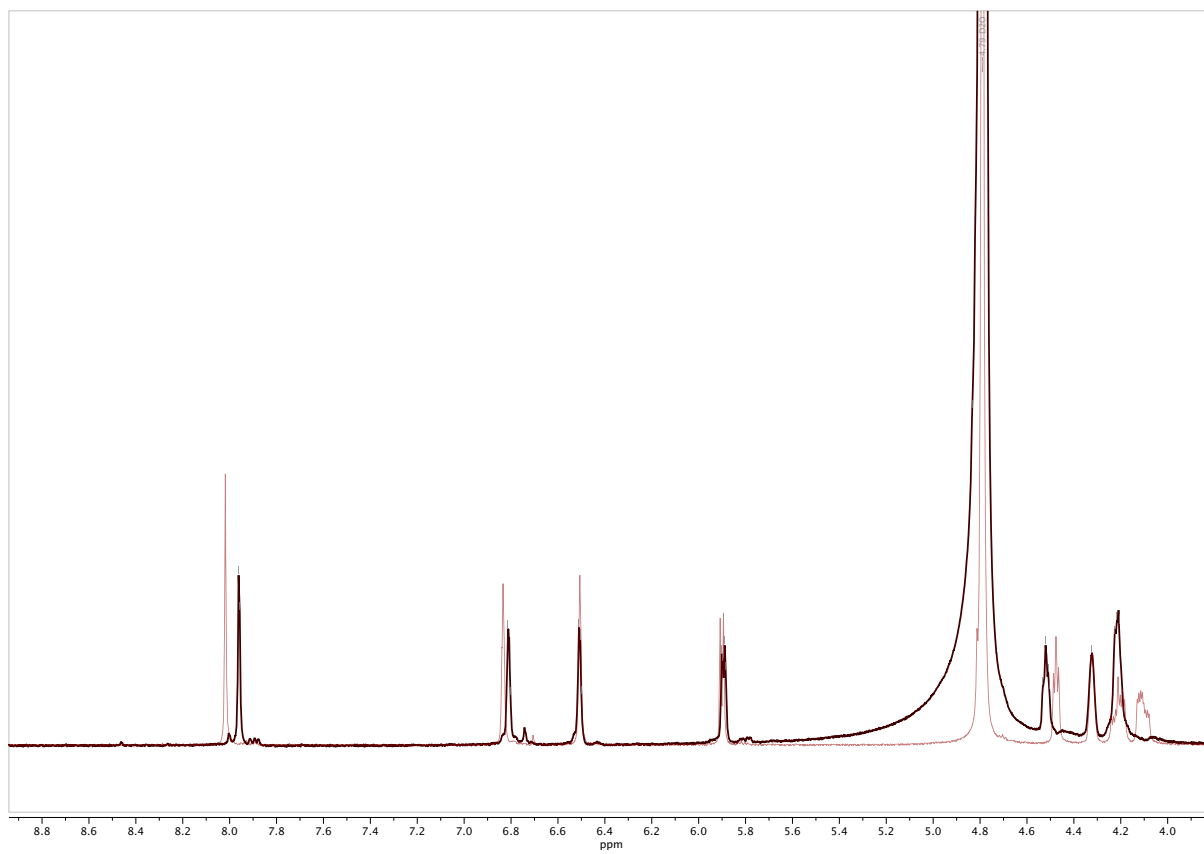
Figure C.10. Comparison of ¹H NMR spectra of diastereomer 1 and 2. Diastereomer 1 (black trace) and diastereomer 2 (red trace). Refer to Tables C.2 and C.3 for peak assignments.
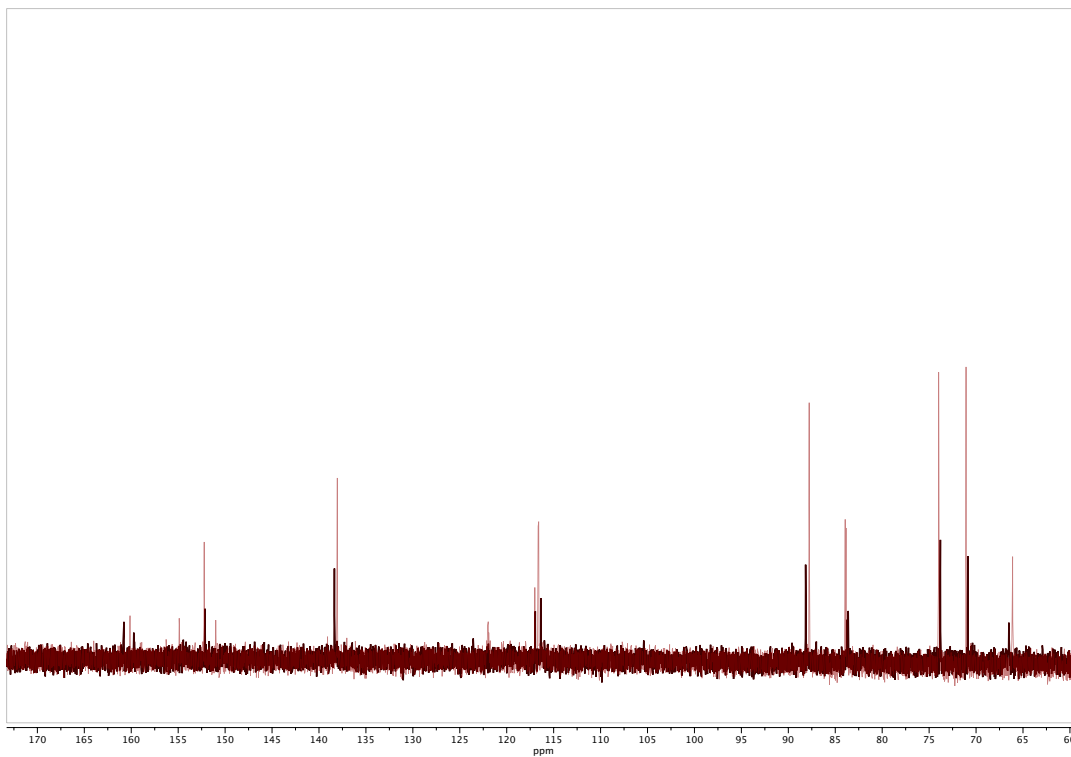
Figure C.11. Comparison of ^{13}C NMR spectra of diastereomer 1 and 2. Diastereomer 1(red trace) and diastereomer 2 (black trace). Refer to Tables C.2 and C.3 for peak assignments.
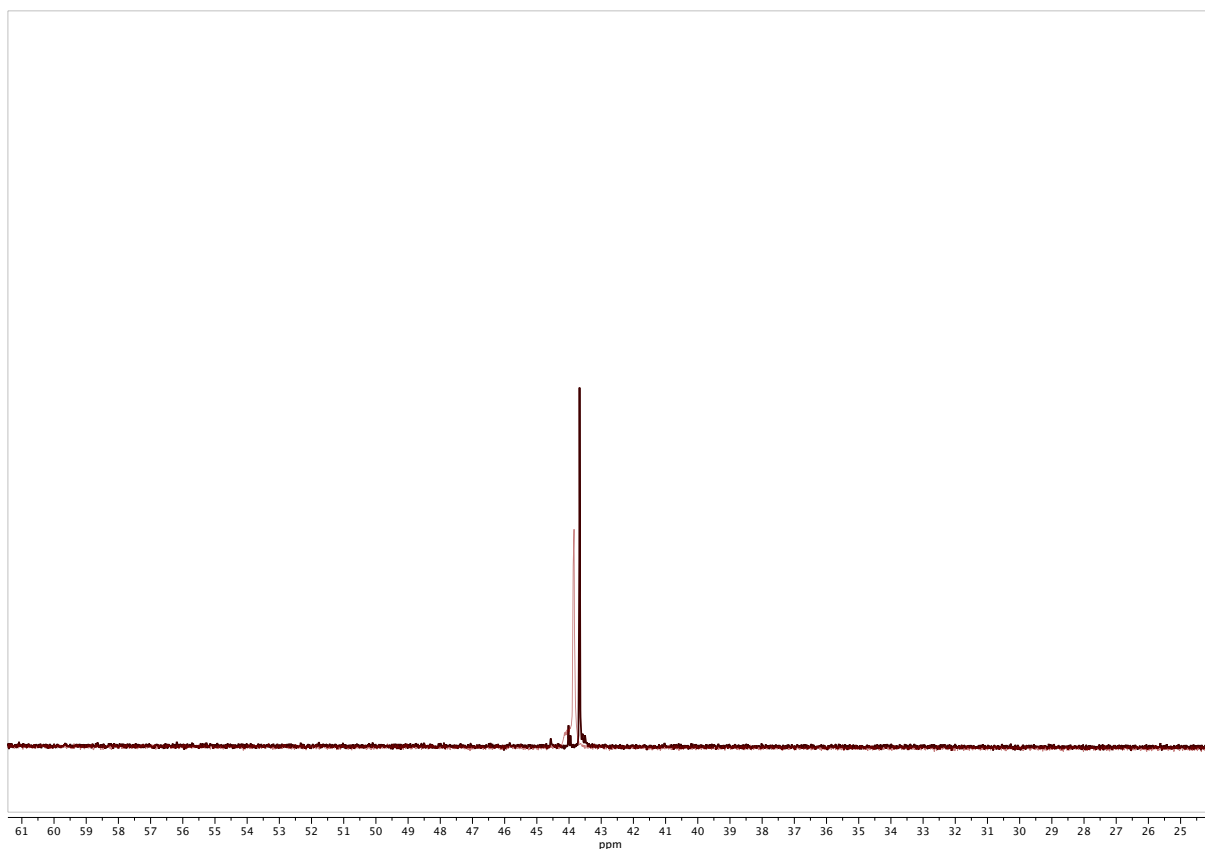
Figure C.12. Comparison of ^{31}P NMR spectra of diastereomer 1 and 2. Diastereomer 1(red trace) and diastereomer 2 (black trace). Refer to Tables C.2 and C.3 for peak assignments.

Synthesis and characterization of mixed thiophosphoro/phosphoro-2-aminoimidazolium-bridged dinucleotides

The mixed thiophosphoro/phosphoro-2-aminoimidazolium-bridged dinucleotides were synthesized in water, at pH=8 with a ratio of 1:4 AIp$_s$G: CpOAt. The reactions were stirred continuously from 1-48 hours, and analyzed by NMR for reaction products. Synthesis for pure diastereomeric products must begin with the diastereomerically pure AIp$_s$G, due to inability of the racemic Gp$_s$*pC to separate by C18 column purification.
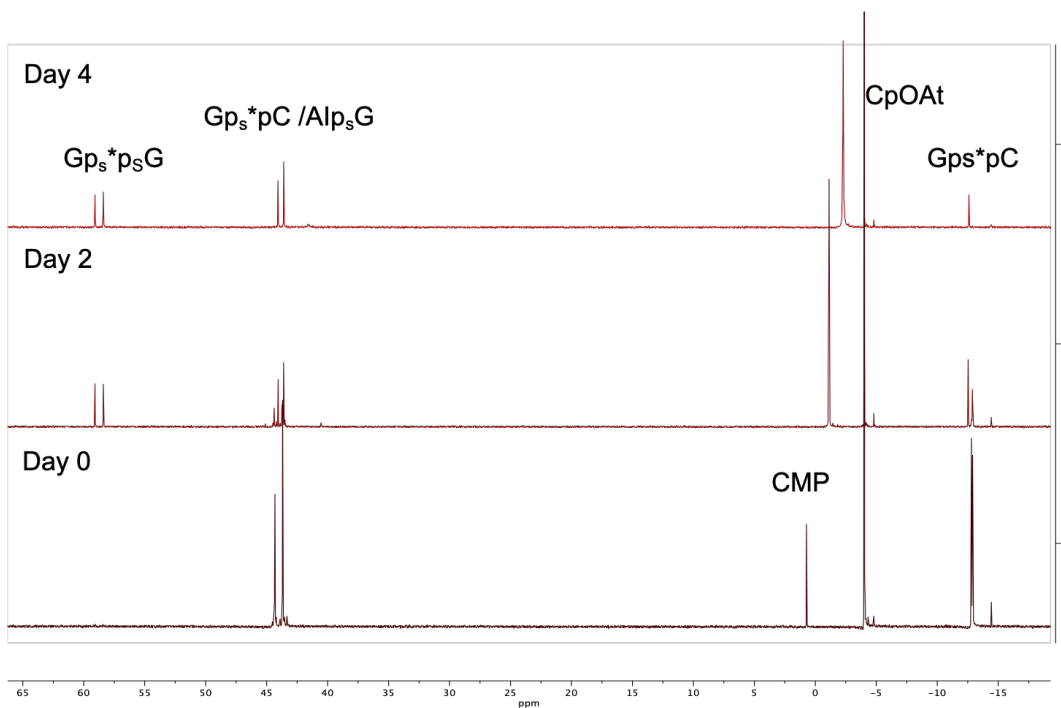
Figure C.13. Formation of racemic Gp$_s$*pC monitored by ^{31}P NMR
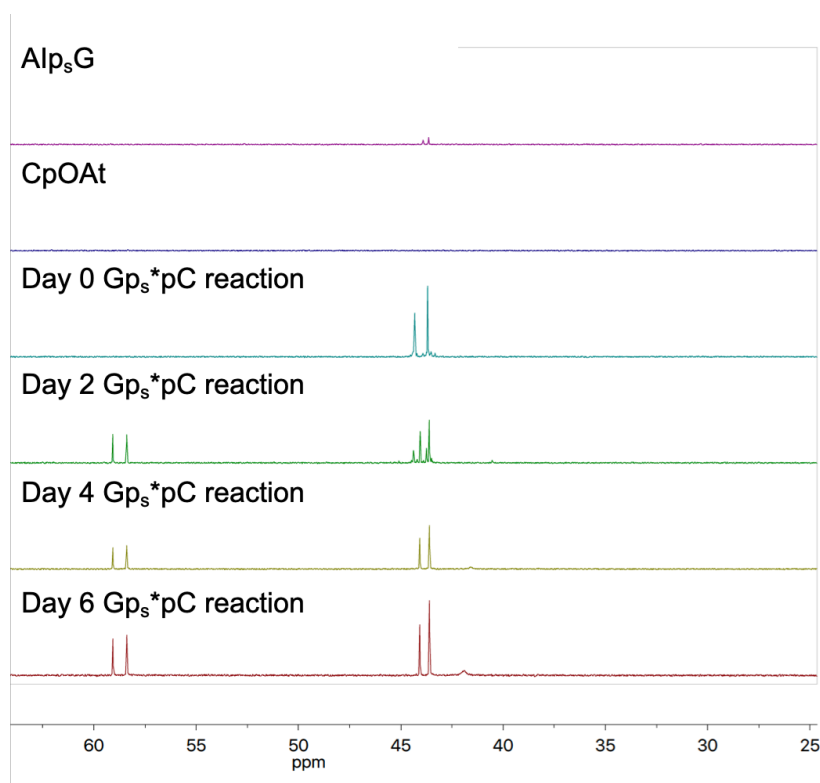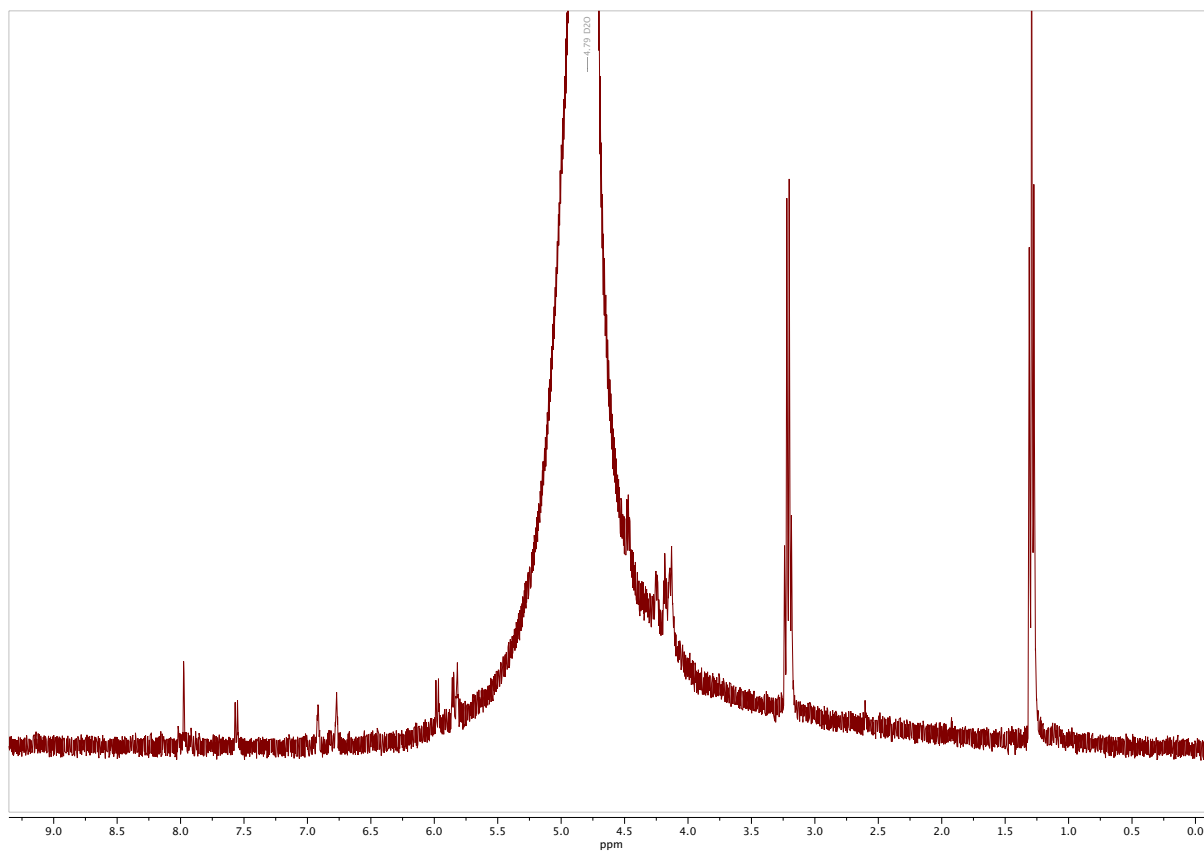

Figure C.14. Formation of racemic Gp$_s$*pC monitored by ^{31}P NMR downfield (65-25 ppm)
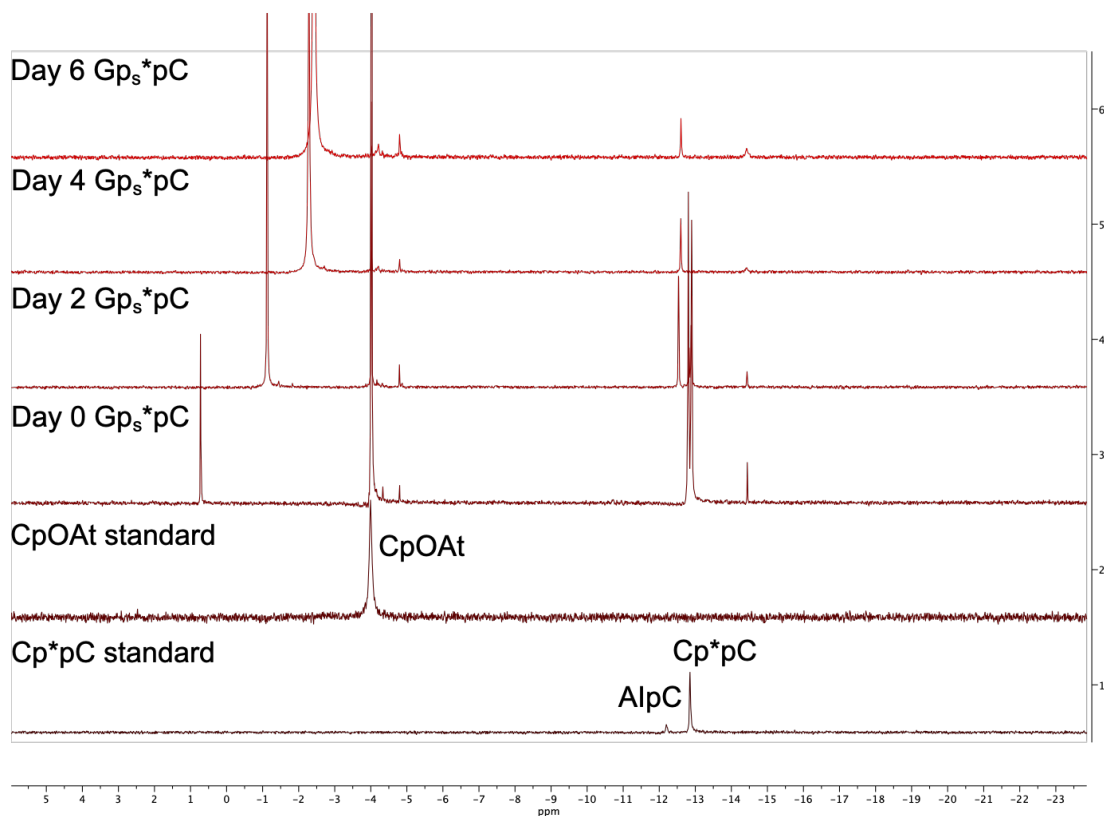
**Figure C.15**. Formation of Gp_s*pC monitored by ^{31}P NMR upfield (6 to -24 ppm)

Table C.4. Gp_s*pC diastereomer 1

^1H	^{1}H NMR (400 MHz, d$_2$o) δ 8.02, 7.98, δ 7.56 (d, J = 7.4 Hz), 6.92, 6.77, 6.02 – 5.94 (m) 5.88 – 5.84 (m), 5.82, 4.79(HOD) 4.47, 4.25, 4.18, 4.13, 3.21(m), 2.60, 1.29 (m).
^{31}P	^{31}P NMR (162 MHz, d$_2$o) δ 44.48, -12.87.
ESI-MS	Exact mass (748.11), Observed mass (748)

Figure C.16. ^{1}H NMR spectrum of C18 purified Gp$_s$*pC diastereomer 1. HOD and triethylamine are observed from purification. Refer to Table C.4 for peak assignments.
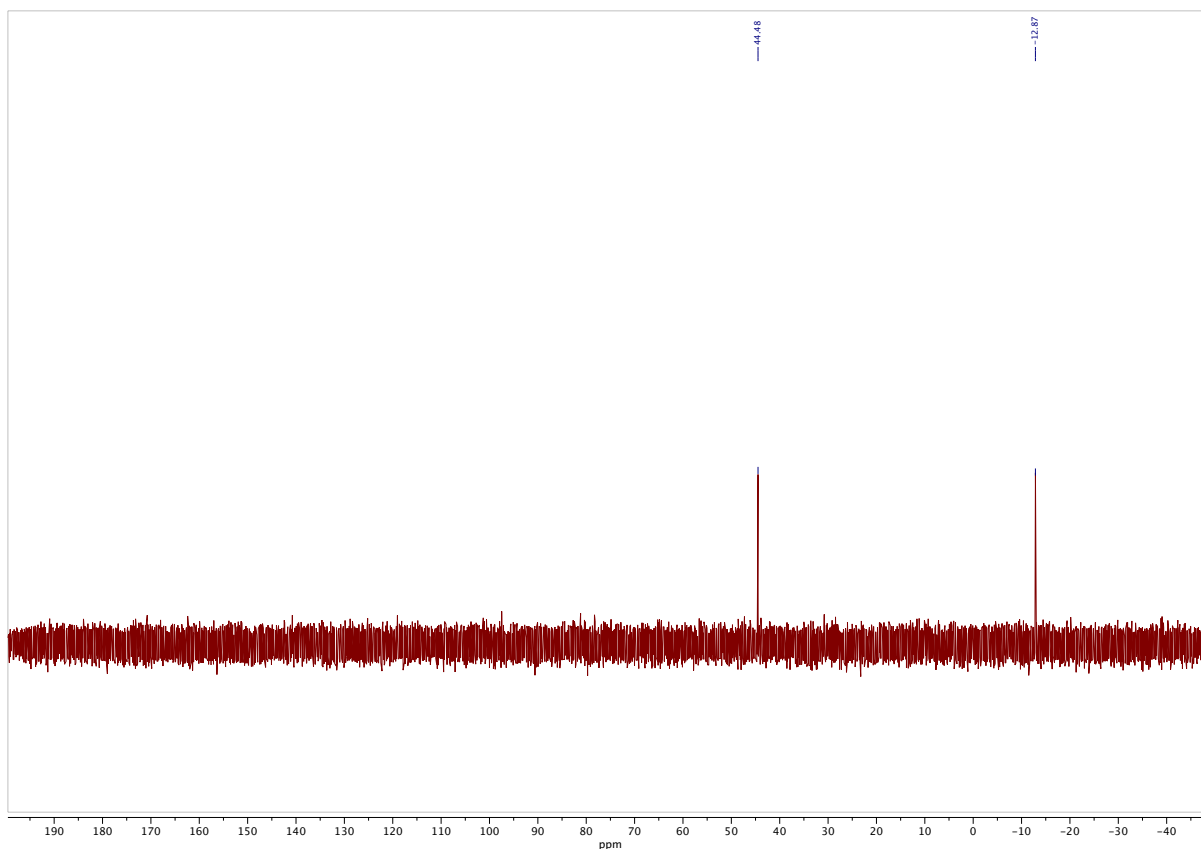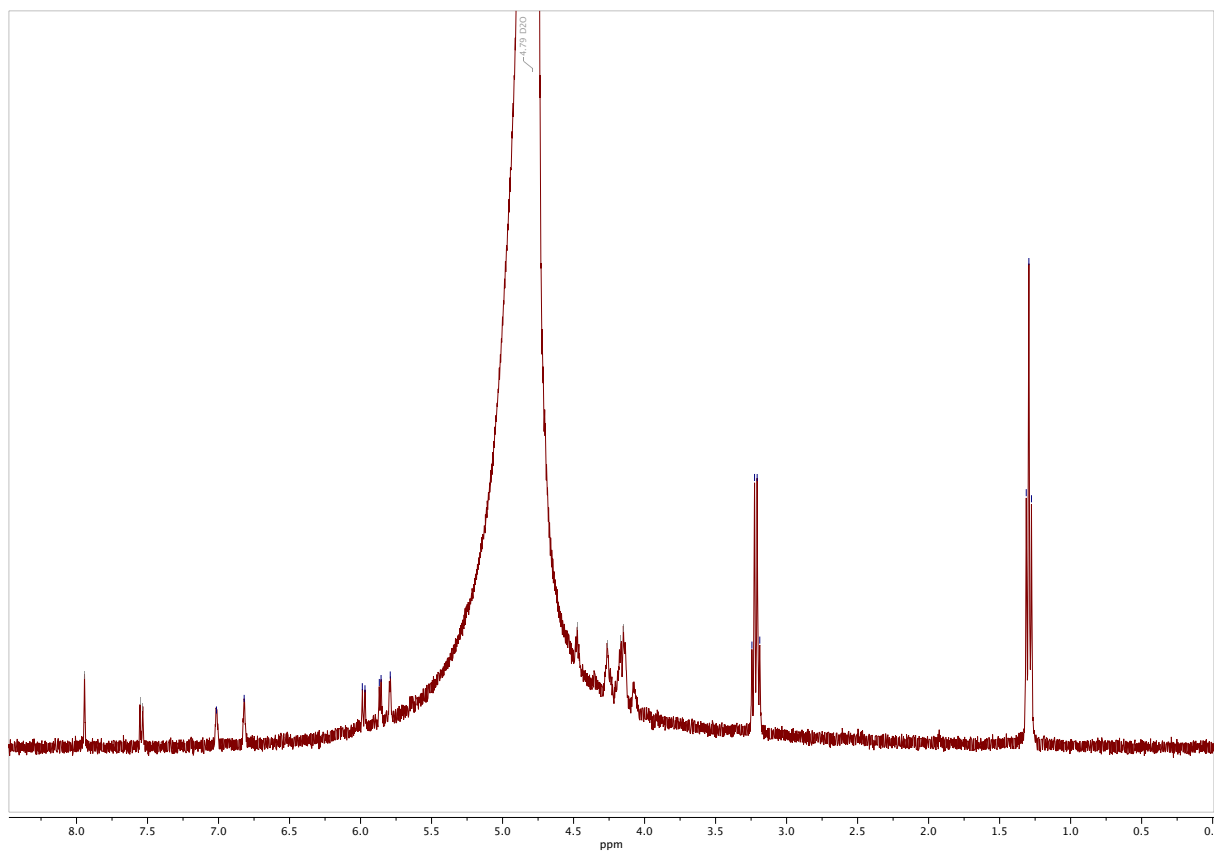
Figure C.17. ^{31}P NMR spectrum of C18 purified Gp$_s$*pC diastereomer 1. Refer to Table C.4 for peak assignments.

Table C.5. Gp$_s$*pC diastereomer 2

|		δ
^1H	^{1}H	^{1}H NMR (400 MHz, d$_2$o) δ 7.94, 7.55(d), 7.53, 7.01, 6.82, 5.98(d), 5.87(d), 5.79, 4.79(HOD) 4.47, 4.26, 4.17, 4.15, 3.22 (q)TEA, 1.29 (t) TEA.
^{31}P	^{31}P	^{31}P NMR (162 MHz, d$_2$o) δ 43.75, -12.80.
ESI-MS		Exact mass (748.11), Observed mass (748)

**Figure C.18.** ^{1}H NMR spectrum of C18 purified Gp$_s$*pC diastereomer 2. Refer to Table C.5 for peak assignments.
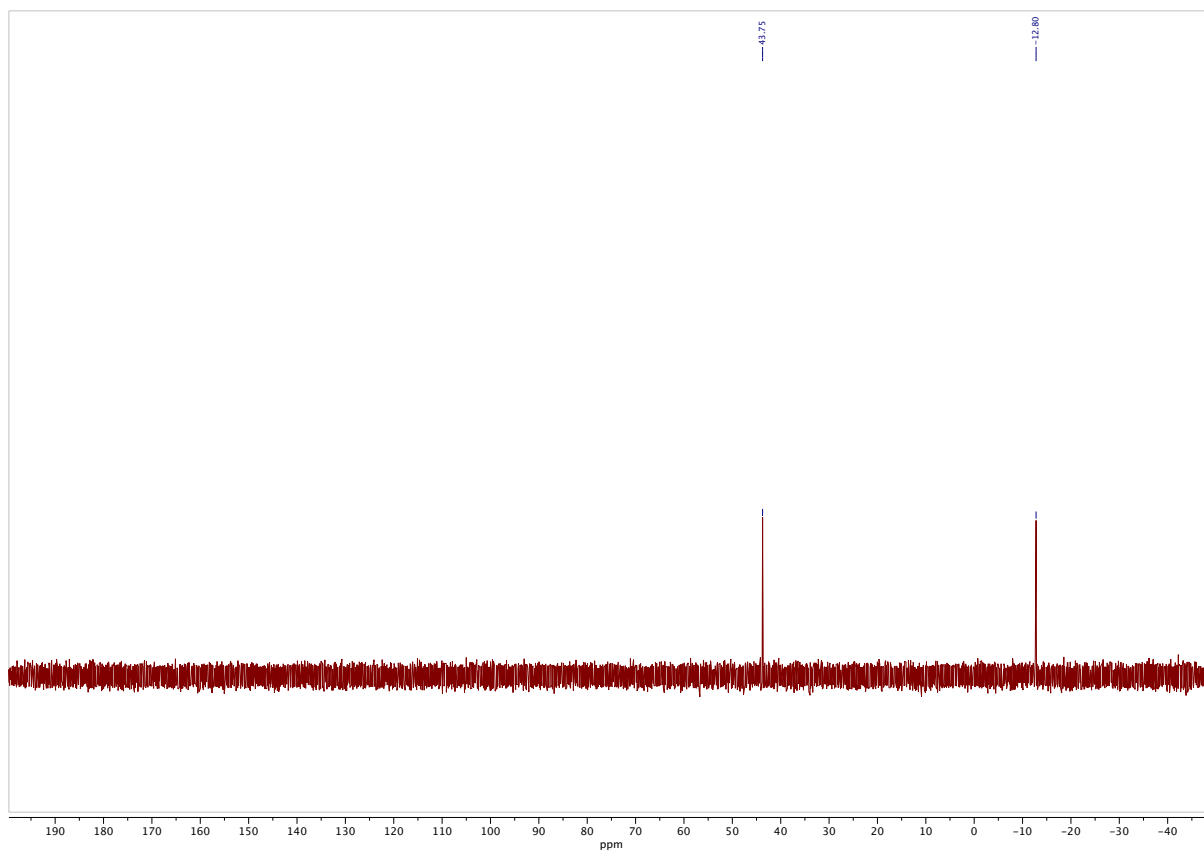
Figure C.19. ^{31}P NMR spectrum of Gp$_s$*pC diastereomer 2. Refer to Table C.5 for peak assignments.
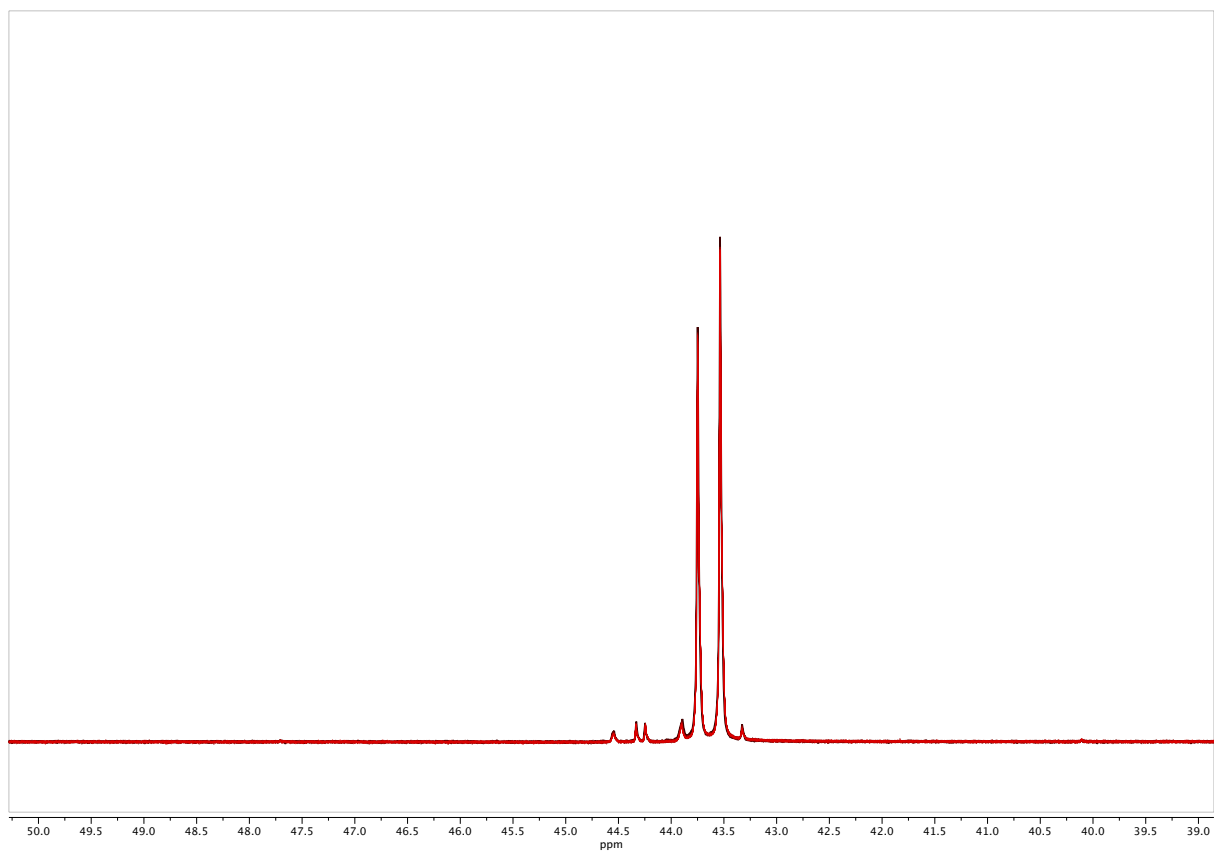
**Figure C.20.** Hydrolysis of AIpsG over the course of 16 hours ^{31}P NMR (superimposed 38.5-50 ppm). 16 hours (red trace), 0 hours (blue trace).
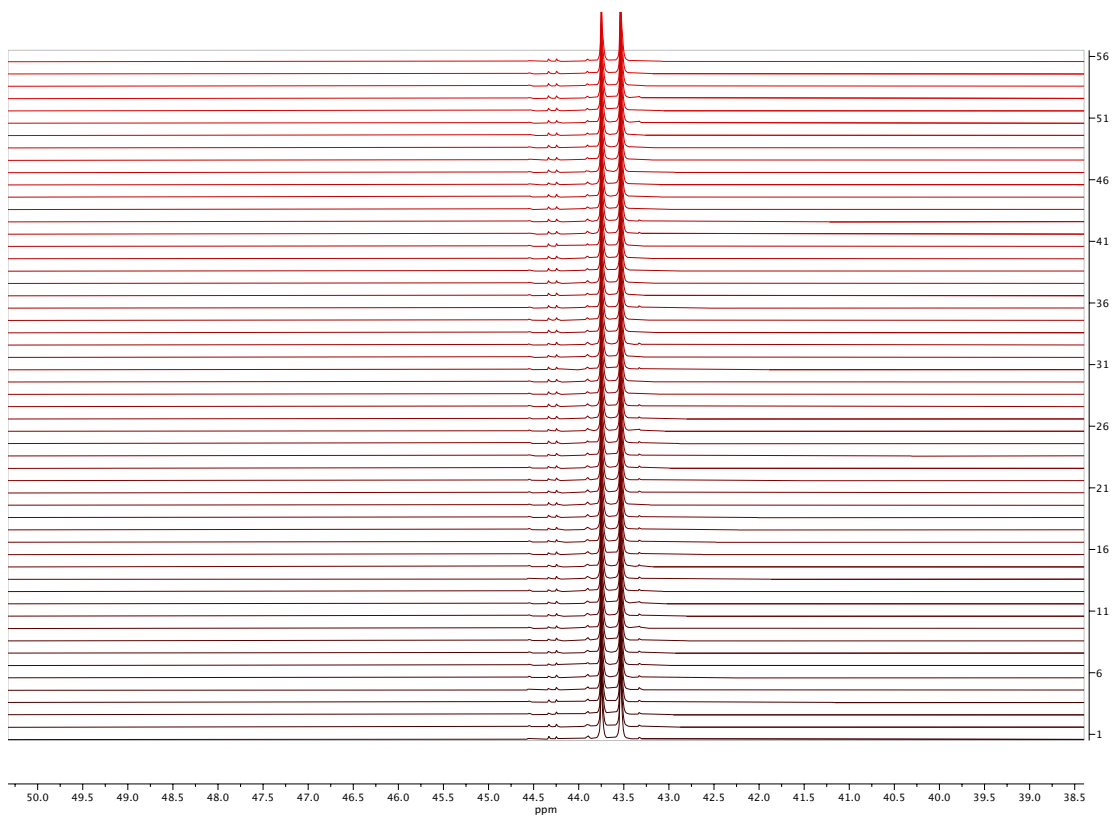
Figure C.21. Hydrolysis of AIpsG over the course of 16 hours ^{31}P NMR (stacked 38.5-50 ppm).
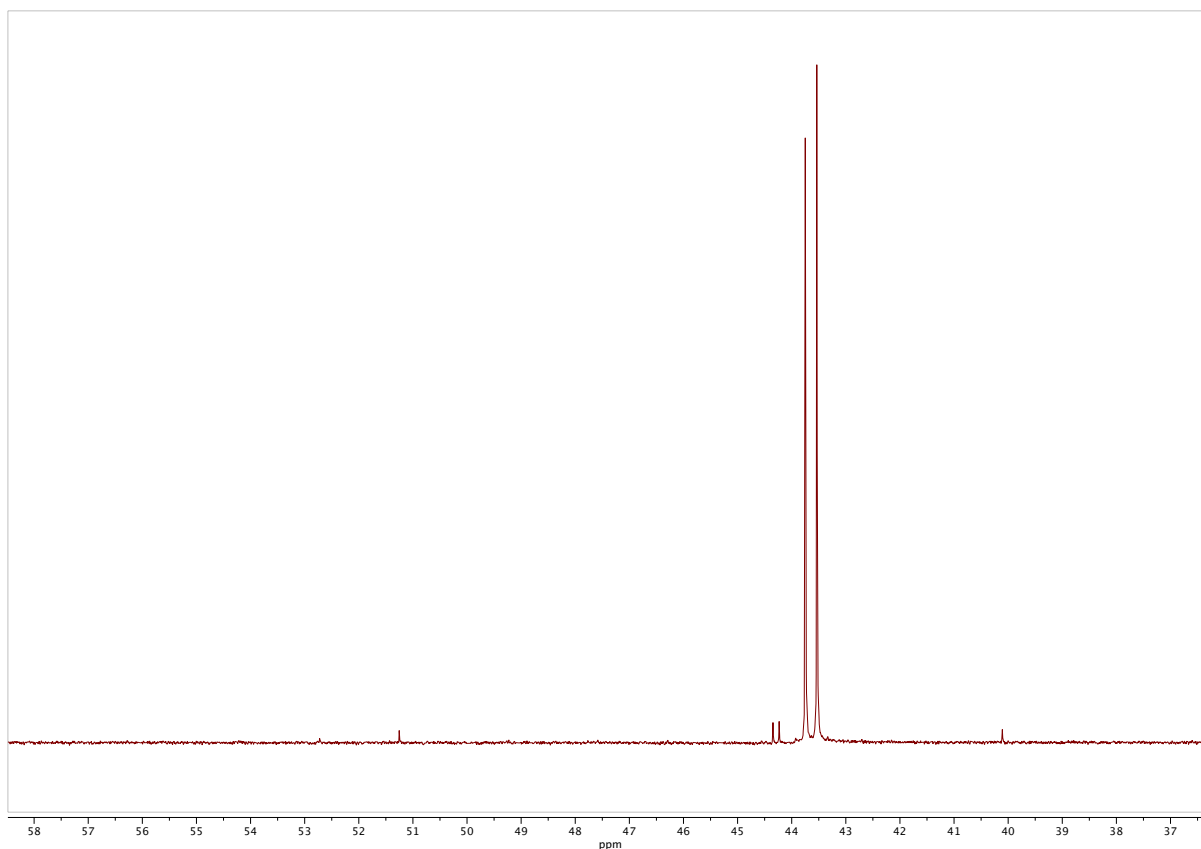
Figure C.22. AIp$_S$G hydrolysis, 3-week timepoint (37-58 ppm). ^{31}P NMR (162 MHz, d$_2$o) δ(ppm) observed include 51.25, 44.35, 44.23, 43.75(D1), 43.54 (D2), 40.11, 13.43. Identity of remaining peaks are unknown. Peaks at 44 ppm show growth from Figure C.11.
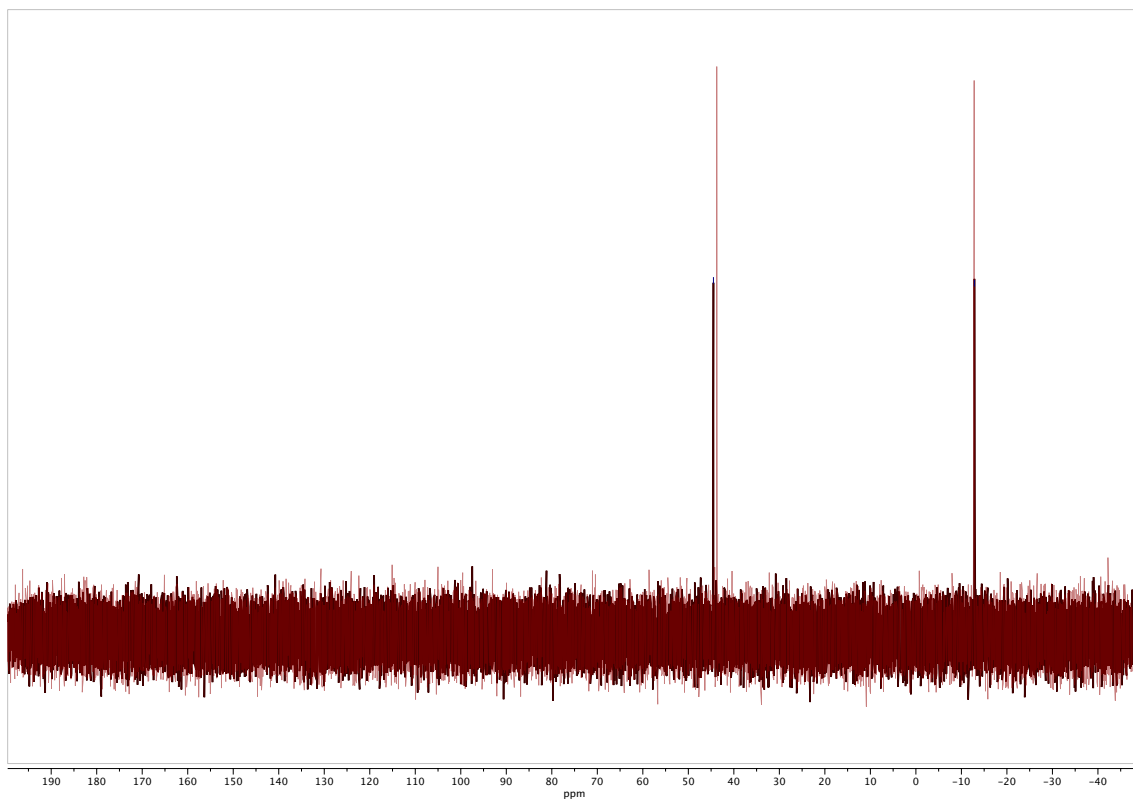
Figure C.23. Overlay of Gps*pC D1 and D2. D1 (black trace) and D2 (red trace).
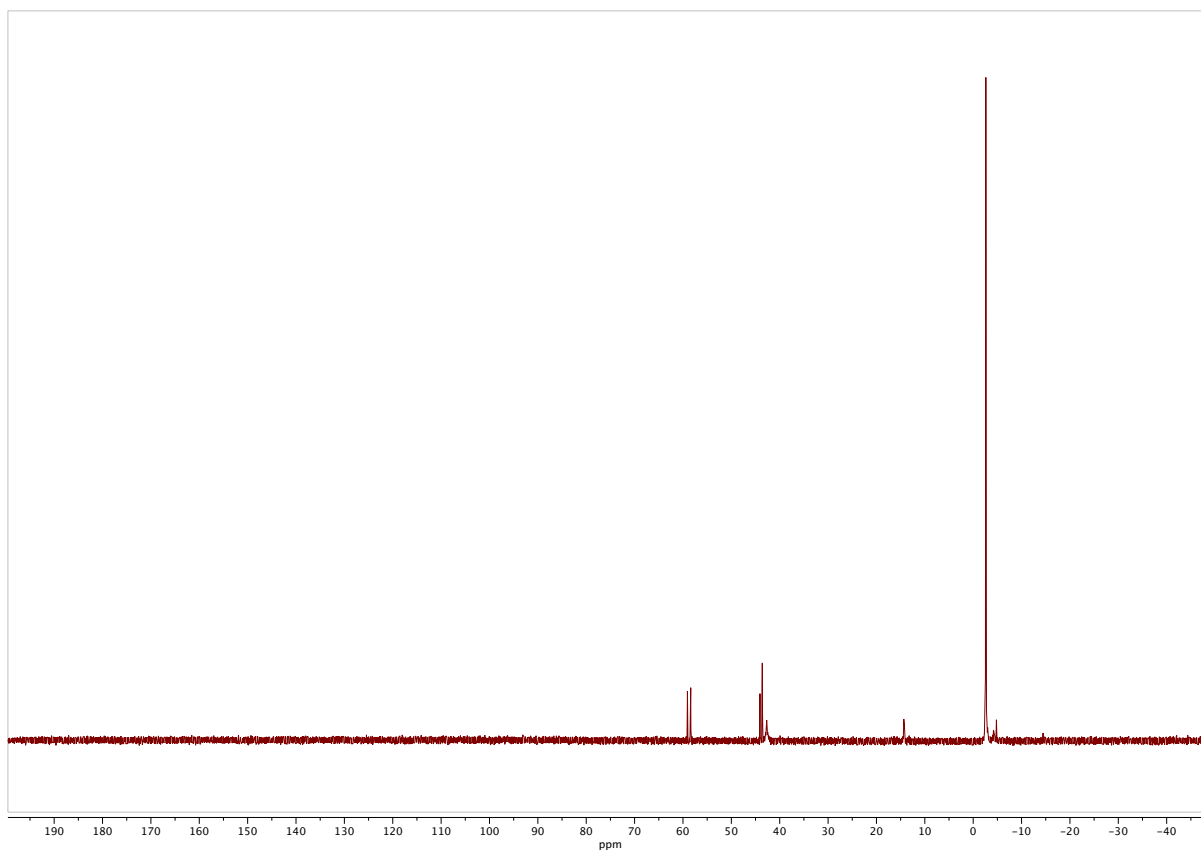
Figure C.24. Gp$_s$*pC D2 synthesis mixture, 3-week timepoint,^{31}P NMR. Very little of Gp$_s$*pC remains, hydrolysis and phosphodiester products observed. Re-formation of activated monomers and formation Gp$_s$*$_s$pG relevant on longer time scales.

X-ray crystallography

Table C.6. RNA primer sequences for crystallography.

Primer	Sequence
P1	5'-*mCmCmC*GACUUAAGUC-3'
P2	5'-*mCmCmC*GACUUAAGUCG-3'
P3	5'-*TmCmC*GACUUAAGUCG-3'

Italics: locked nucleic acid (LNA). mC:5-methyl cytosine.

Table C.7. Optimized conditions for crystallization of phosphorothioate extension products.

Entry	Primer	Phosphorothioate Substrate	Optimized crystallization conditions
1	P1	G Monomer Diastereomer 1	0.2 M Ammonium acetate, 0.1 M BIS-TRIS pH 6.5, 45% v/v (+/-)-2-Methyl-2,4-pentanediol.
2	P1	G Monomer Diastereomer 2	1.5 M Lithium sulfate, 50 mM TRIS pH 8.5, 5% w/v Glycerol.
3	P2	G Monomer Diastereomer 1	20 mM Magnesium chloride, 50 mM MOPS pH 7.0, 2.0 M Ammonium sulfate, 0.5 mM Spermine.
4	P2	G Monomer Diastereomer 2	50 mM PIPES pH 7.0, 65 mM Magnesium chloride, 1 mM Cobalt (III) Hexamine chloride, 26% v/v (+/-)-2-Methyl-2,4-pentanediol.
5	P3	GA Dimer Diastereomer 1	1.7 M Lithium sulfate, 50 mM HEPES pH 7.0, 50 mM Magnesium sulfate.

Table C.8. Data collection statistics.

Entry	1	2	3	4	5
PDB code	7U87	7U88	7U89	7U8A	7U8B
Beamline	8.2.2 (ALS)	8.2.2 (ALS)	8.2.2 (ALS)	8.2.2 (ALS)	23-ID-B (APS)
Wavelength (Å)	1.000034	1.000034	1.000034	1.000034	1.033167
Space group	P321	P321	P321	P321	P321
Unit cell parameters (Å, °)	42.74, 42.74, 83.28, 90, 90, 120	42.72, 42.72, 85.82, 90, 90, 120	43.20, 43.20, 85.64 90, 90, 120	42.87, 42.87, 82.76, 90, 90, 120	43.46, 43.46, 81.12, 90, 90, 120
Resolution range (Å)	50-1.70 (1.73-1. 70)	50-2.14 (2.18-2.14)	50-1.65 (1.68-1.65)	50-2.10 (2.14-2.10)	50-1.73 (1.76-1.73)
Unique reflections	10188 (474)	5344 (229)	11662 (553)	5553 (273)	9404 (474)
Completeness (%)	99.9 (98.8)	99.2 (88.8)	99.6 (98.9)	99.9 (100)	96.3 (98.5)
R_{merge} (%)	8.9 (34.8)	7.7 (40.2)	8.6 (26.0)	7.9 (52.7)	9.3 (49.9)
$<I/\sigma(I)>$	19.4 (2.7)	24.2 (2.2)	21.0 (5.6)	25.9 (4.2)	13.8 (4.8)
Data redundancy	9.9 (6.5)	9.5 (5.7)	10.1 (7.8)	10.2 (9.9)	5.0 (4.8)

Bibliography

(1)   Ding, D.; Zhou, L.; Giurgiu, C.; Szostak, J. W. Kinetic Explanations for the Sequence Biases Observed in the Nonenzymatic Copying of RNA Templates. *Nucleic Acids Res* **2021**, *50* (1), 35–45. https://doi.org/10.1093/nar/gkab1202.

(2)   Gottlieb, H. E.; Kotlyar, V.; Nudelman, A. NMR Chemical Shifts of Common Laboratory Solvents as Trace Impurities. *J. Org. Chem.* **1997**, *62* (21), 7512–7515. https://doi.org/10.1021/jo971176v.

(3)   Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56* (5), 2257–2261. https://doi.org/10.1063/1.1677527.

Appendix D

Methods

Primer extension: Mixed template primer extension is performed as previously described in Appendix A and using the sequence from Figure 4D in Li et al. with activated monomers and trimers,[1] except with 100 mM pH=7.5 Tris-HCl, and 50 mM $MnCl_2$ in anoxic conditions. Reaction quenching was performed with 29 μL urea-acrylamide mixture as previously described with 75 μM DNA complement.[2] Samples were heated to 90 degrees Celsius before 20% acrylamide denaturing PAGE. Subsequent analysis occurs as described in Appendix A.4.

Ligation:

Ligation experiments were performed as described in [3] except with $MnCl_2$.

Sequencing

NERPE-Seq protocol was followed as previously described except at pH=7.5 and in the presence of 50 mM $MnCl_2$.[4,5] Before spin-column purification, the samples were quenched with 100 mM EDTA.

Analysis of Cp*pC formation and hydrolysis in presence of $MnCl_2$ with analytical HPLC

Due to the paramagnetic nature of Mn^{2+}, continuous experiments measuring the concentration of activated monomers and dimers through ^{31}P NMR were not accessible as when using $MgCl_2$. Therefore, at discrete time points of the hydrolysis of Cp*pC (50 mM Tris pH=7.5, 20 mM Cp*pC, 50 mM Mn^{2+}), 1 μL was removed from the reaction mixture, and quenched with 500 mM EDTA (10 μL). Each time point was flash-frozen, and kept on ice until analysis by analytical HPLC (same gradient, protocol, and analysis as in Appendix B). First order kinetics were assumed for rate calculations for hydrolysis experiments. For Cp*pC formation from *pC, 24 mM *pC, 50 mM Mn^{2+}, 50 mM Tris pH=7.5 was kept at room temperature and discrete time

points were removed in the same manner as for hydrolysis. Second order kinetics were assumed for rate calculations for bridged dinucleotide experiments

Bibliography

(1)   Li, L.; Prywes, N.; Tam, C. P.; O'Flaherty, D. K.; Lelyveld, V. S.; Izgu, E. C.; Pal, A.; Szostak, J. W. Enhanced Nonenzymatic RNA Copying with 2-Aminoimidazole Activated Nucleotides. *Journal of the American Chemical Society* **2017**, *139* (5), 1810–1813. https://doi.org/10.1021/jacs.6b13148.

(2)   Walton, T.; Szostak, J. W. A Kinetic Model of Nonenzymatic RNA Polymerization by Cytidine-5′-Phosphoro-2-Aminoimidazolide. *Biochemistry* **2017**, *56* (43), 5739–5747. https://doi.org/10.1021/acs.biochem.7b00792.

(3)   Zhou, L.; O'Flaherty, D. K.; Szostak, J. W. Template-Directed Copying of RNA by Non-enzymatic Ligation. *Angew. Chem. Int. Ed.* **2020**, *59* (36), 15682–15687. https://doi.org/10.1002/anie.202004934.

(4)   Duzdevich, D.; Carr, C. E.; Ding, D.; Zhang, S. J.; Walton, T. S.; Szostak, J. W. Competition between Bridged Dinucleotides and Activated Mononucleotides Determines the Error Frequency of Nonenzymatic RNA Primer Extension. 11.

(5)   Duzdevich, D.; Carr, C. E.; Szostak, J. W. Deep Sequencing of Non-Enzymatic RNA Primer Extension. 13.