In Search of Primordial RNA: The Effects of Noncanonical Nucleotides on Nonenzymatic Primer Extension

In Search of Primordial RNA:

The Effects of Noncanonical Nucleotides on Nonenzymatic Primer Extension

A thesis presented by

Seohyun Chris Kim

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

August 2019

Dissertation Advisor: Jack W. Szostak *Seohyun Chris Kim*

In Search of Primordial RNA: The Effects of Noncanonical Nucleotides on Nonenzymatic Primer Extension

Abstract

The observation that the cellular machinery comprised of DNA, RNA, and proteins is more or less preserved across all terrestrial life evokes the thought that all life may stem from a unified origin. One of the many theories for biogenesis, the "RNA World" hypothesis posits that RNA was the first biopolymer that carried out all cellular function responsible for maintaining and propagating life. In early biological systems, RNA would serve a dual role in genetic information storage and transfer (allowing for inheritance) and as a catalyst of the necessary functions for sustaining life. This hypothesis is supported by RNA's diverse functions in modern biology, as a messenger of genetic information from DNA to proteins, and as a vital catalyst as the ribozyme in the ribosome, as well as by the observation that RNA serves a vestigial role in many of the most important and ubiquitous co-factors. However, there are many unanswered questions concerning the progression from a completely chemical world to the first forms of RNA life.

In Chapter 1, I discuss the key requirements for the RNA world hypothesis to hold true. The first key requirement for RNA to emerge as the primordial biopolymer is the abiotic synthesis of ribonucleotide monomers; I cover the progression from historical to modern syntheses. Satisfactory syntheses of pyrimidine ribonucleosides have been accomplished, but purine ribonucleosides remain a problem, with low-yielding routes that produce byproducts. Assuming all ribonucleotide monomers can be made and subsequently polymerize to form random sequence RNA, the next key problem is the ability for RNA to propagate through nonenzymatic copying. Thus, I then review the state of the art of nonenzymatic template-directed copying of RNA,

focusing on rates and fidelity of informational inheritance as readouts for success. Given the wealth of byproducts produced in prebiotic ribonucleotide synthesis, coupled with the lack of a purine synthesis, evaluating the behavior of non-canonical nucleotides in RNA copying is quintessential.

Recently a prebiotic pathway to 8-oxo-adenosine and 8-oxo-inosine has been described, raising the question of the suitability of 8-oxo-purines as substrates for prebiotic RNA replication. In Chapter 2, I demonstrate that 8-oxo-purine nucleotides are poor substrates for nonenzymatic RNA primer extension; both the rate and fidelity of primer extension are poor. On the other hand, inosine exhibits surprisingly rapid and accurate nonenzymatic RNA copying. Inosine, which can be derived from adenosine by deamination, exhibits copying properties consistent with the hypothesis that it could have replaced guanosine in the earliest stages of the emergence of life.

Recent prebiotic synthetic routes suggest ribonucleotide synthesis may have been accompanied by the synthesis of arabino- and 2′-deoxyribonucleotides. To determine how relatively homogeneous RNA could have emerged from complex mixtures, I discuss the properties of arabino- and 2′-deoxyribonucleotides in nonenzymatic template-directed primer extension reactions in Chapter 3. While arabino-, and to a lesser extent 2′-deoxyribonucleotides, do not possess copying profiles consistent with a main role in the primordial genetic polymer, experiments with mixtures of nucleotides suggest that the coexistence of ribo- and arabino-nucleotides does not impede the copying of RNA templates. Moreover, chimeric oligoribonucleotides containing 2′-deoxy- or arabinonucleotides act as efficient templates for RNA copying. Thus, one can envision a scenario in which the initial genetic polymers were random sequence chimeric oligonucleotides formed by untemplated polymerization, but that template copying chemistry favored RNA synthesis; multiple rounds of replication may have led to pools of oligomers composed mainly of RNA.

Table of Contents

Acknowledgments

First and foremost, I would like to thank my advisor (Prof. Jack Szostak). He has always been kind, had very high expectations, and has consistently encouraged me and all of his lab members to solve one of science's most interesting problems, while also providing his immense support, insight, and resources. His intellectual journey of tackling various challenging problems has been truly inspirational. Along with his astounding academic success, he has been very available and offered me much help. My very first project in the lab (found in Chapter 2) would not have been completed without his insights and guidance. As a traumatized fourth year graduate student with minimal accomplishments and funding, Jack had no reason to accept me into his lab. Yet, he gave me an opportunity to continue my graduate studies. I have to admit that a big part of my motivation in lab was to pay back the debt I owe to him and to prove to myself that his decision was right. I am not so sure if my work here has proven anything, but I can't thank him enough for serving as both an academic and personal role model.

I would like to also thank the many people who have supported me during my unexpected and tumultuous journey over last six years. Professor Emily Balskus has shared a wealth of experience in dealing with unexpected tragedy, such as we both encountered in 2016. Due to my abrupt graduation, I was unable to have her on my thesis committee but she has been she has served on my graduate advising committee over many years, providing scientific and life guidance. Professor Gary Ruvkun has offered a new perspective on biology that I lack and furthermore, we shared our mutual interest in and love of organic chemistry. I greatly appreciate Professor Brian Liau for his accommodation and time, and the opportunity to watch and draw inspiration from his transition from chemistry to biology to his early career. Fanny Ng has always taken care of the lab and greatly accelerated everyone's research while keeping us (mostly me) out of trouble. She was

indispensable in my ability to graduate abruptly and has made everything work out smoothly, as always. Joe Lavin has taken care of all the administrative work in the CCB department and has been very patient and supportive.

The work described below would not have been possible without many coworkers and mentors. Dr. Derek O'Flaherty taught me how to activate monomers, run experiments properly, and enjoy Golden Monkey (aka. beer with <9% ethanol is just water). Dr. Lijun Zhou's work ethic and kindness have been truly inspiring and working alongside her and getting help with oligonucleotide syntheses have been indispensable to our projects. Dr. Victor Lelyveld kindly shared his thoughts and knowledge over many late night conversations and Viva burritos. Dr. Chun Pong (James) Tam has been very kind, supportive, and entertaining. He was the first lab member I met prior to joining the lab and hearing his experience in the lab was vital to my decision to join the lab (it was a "*nail in the tombstone*"). His work ethic, passion for organic chemistry, and sense of humor were too great for me to emulate or fathom but it has been a great pleasure working with him and seeing him find success at Moderna. Additionally, intellectual conversations with Dr. Li Li, Dr. Daniel Duzdevich, and Dr. Tom Wright were always enlightening and being exposed to experts in different areas has been a great luxury in the Szostak lab. Costi Girgiu's dark sense of humor, work ethic, and intelligence are truly unparalleled to anyone I have seen and I look forward to seeing his bright future as a professor in Romania. Lydia Pazienza's work ethic, maturity, and generous donations of her amazing pastries have been very enjoyable and I am eager to see her rigorous work in the mechanistic studies of divalent metals. Furthermore, I want to thank the many collaborators I have had over the years, especially Dr. Wen Zhang and Valeria Rondo-Brovetto who have greatly helped me in our projects (found in Chapter 3).

I would like to thank two of my closest friends I met in Boston; Jayleen Bai has been a great friend who has been a continuing source of inspiration with her unwillingness to give up and her attention to details. I am delighted to see her recent success that she deserved and only wish her the best of luck!

My closest friend Dr. Andrew James Bendelsmith has been my partner in crime, going through numerous (perhaps too many) problems together. His kindness and intelligence (with self proclaimed hilarity) have made him the most smooth operator and the most likeable person that I know. It has been my greatest pleasure to work with him and I have no doubt that he will have a bright future in the Knowles lab. Thank you my friend for your support over last six years and, although in difference places, we will continue to work and complain together many years after.

I am greatly indebted to my family for their support. My sister is the closest friend of mine and she always listens to my concerns. During my graduate studies, she has faced many challenges in her career and seeing her excelling in every part of her life has been so joyful and motivational. My mother has been unconditionally supportive while pushing me with her insatiable thirst for fame and success. I doubt that I will ever meet her expectations, but I am also excited for her to surprise me. I love you both so much.

Lastly, I am thankful for while also feeling bittersweet for my friend, Dr. Daisuke Suzuki. He was a very gentle lab member, friend, great ‘prebiotic chemist', and father, so his unexpected death was a huge loss to our community. I still cannot fathom his decision, but his departure has immensely influenced me. He has taught me lessons that I will continue to ponder for the rest of my life. Rest in peace, Daisuke.

List of Abbreviations

2AIp	2-aminoimidazole activated nucleotide monophosphate
8-oxo or 8Ox	purine with a carbonyl at the 8 position
α	absolute configurations of anomeric carbon and reference atom are the same
Å	angstrom
A	adenine
Acetyl-CoA	acetyl coenzyme A
ANA	arabinonucleic acid
ATP	adenosine triphosphate
aq.	aqueous
ara	arabino-
β	absolute configurations of anomeric carbon and reference atom are opposite
C	cytosine
calc	calculated
Δ	difference or change
δ	chemical shift in ppm
°C	degree Celsius
D or *d*	deuterium
d	2'-deoxyribo-, doublet
DFT	density functional theory
DNA	deoxyribonucleic acid
E	electronic energy
Equiv.	equivalent(s)

et al.	et alia (Latin: and others)
FAD	flavin adenine dinucleotide
G	guanine
GC	gas chromatography
h or hr	hour(s)
H	concentration of proton
^{1}H	protium detected on NMR
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRMS	high resolution mass spectrometry
hν	electromagnetic irradiation (UV)
Hz	Hertz
I	inosine
IR	infrared
IMP	inosine mono-phosphate
J	scalar coupling constant (measured in Hz)
k	rate constant
K	unspecified equilibrium
K_M	Michaelis-Menten Michaelis constant
L	liter
ln	matural logarithm
μ	micro
m	milli, multiplet, messenger
M	molar

m/z	mass/charge ratio
Me	methyl
MeOH	methanol
MHz	megaHertz
min	minute(s)
MS	mass spectroscopy
Na^+	sodium ion
NADH	nicotinamide adenine dinucleotide
NASA	national aeronautics and space administration
NMR	nuclear magnetic resonance
NNN	trimer oligonucleotides
obs	observed
p	negative of the base ten logarithm
^{31}P	phosphorus isotope with molar mass 31 detected on NMR
PAGE	polyacrylamide gel electrophoresis
ppm	parts per million
r	ribo-
R	unspecified substituent
RNA	ribonucleic acid
s	second, singlet
soln.	solution
t	time
t	triplet, transfer

T	thymine, temperature
$t_{1/2}$	half life
THF	tetrahydrofuran
TLC	thin layer chromatography
TNA	threofuranosyl nucleic acid
U	uracil
V_{MAX}	Michaelis-Menten maximum velocity or rate

Chapter 1:

Prebiotic Chemistry and the RNA World Hypothesis

1.1 Introduction: Prebiotic Chemistry and the RNA World Hypothesis

Since the beginning of intelligent human thought, the existential question of how and where we, and all terrestrial life, originated has puzzled scientists, religious scholars, and the lay public. The advent of Darwinian Evolution[1] gave us the scientific groundwork, to begin to answer this question. In 1871, Darwin even went so far as to propose the landmark Miller experiment[2] demonstrating the prebiotic synthesis of proteins (vide infra):

"But if (and oh what a big if) we could conceive in some warm little pond with all sort of ammonia and phosphoric salts, light, heat, electricity present, that a protein compound was chemically formed, ready to undergo still more complex changes, at the present such matter would be instantly devoured, or absorbed, which would not have been the case before living creatures were formed […]"- Darwin[3]

[1] a) Darwin, C.; Wallace, A. On the Tendency of Species to Form Varieties; and on the Perpetuation of Varieties and Species by Natural Means of Selection. *J. Proc. Linn. Soc. London. Zool.* **1858**, *3* (9), 45–62. b) Darwin, C. The Origin of Species. **1872**.

[2] Miller, S. L. A Production of Amino Acids under Possible Primitive Earth Conditions. *Science* **1953**, *117* (3046), 528–529.

[3] Peretó, J.; Bada, J. L.; Lazcano, A. Charles Darwin and the Origin of Life. *Orig. Life Evol. Biosph.* **2009**, *39* (5), 395–406.

Before we go into the detail of the nature of the original biological polymer(s) responsible for the origin of cellular life, it is necessary to reflect on modern cellular life[4] and the function of the ubiquitous machinery that allows for life, which NASA defines as "a self-sustaining chemical system capable of Darwinian evolution."[5] Modern life relies on double stranded DNA to store genetic information as a sequence of four nucleobases, adenine (A), guanine (G), cytosine (C), and thymine (T).[6] Genetic information is then transcribed into mRNA, by assembling a complementary strand to the DNA; RNA utilizes the same nucleobases with a notable modification of uracil (U) replacing thymine (T).[7] mRNA is then translated into proteins, with the help of ribozymes and tRNA, which is able to translate a 3 letter codon RNA sequence to the next amino acid of a growing peptide.[8] The immense leaps in our understanding of DNA and cellular processes in the 1950's paved the way for being able to ascertain the requirements one would impose on a primordial genetic polymer responsible for the emergence of life.

---

[4] The following discussion of transcription and translation can be found in many textbooks such as: Lodish, H. *et al.* Molecular Cell Biology. **2000**.

[5] Deamer, D. W.; Fleischaker, G. R. *Origins of Life : The Central Concepts*; Jones and Bartlett Publishers, 1994.

[6] For lead references on DNA structure see: a) Franklin, R. E.; Gosling, R. G. Evidence for 2-Chain Helix in Crystalline Structure of Sodium Deoxyribonucleate. *Nature* **1953**, *172* (4369), 156–157. b) Franklin, R. E.; Franklin, R. E. Molecular Configuration in Sodium Thymonucleate. Franklin, R. y Gosling R. G. *Nature* **1953**, *171*, 740–741. c) Watson, J. D.; Crick. F. H. C. Molecular Structures of Nucleic Acids. *Nature*. 1953, pp 737–738. d) Wilkins, M. H. F.; Stokes, A. R.; Wilson, H. R. Molecular Structure of Nucleic Acids: Molecular Structure of Deoxypentose Nucleic Acids. *Nature* **1953**, *171* (4356), 738–740.

[7] Bladen, H. A.; Byrne, R.; Levin, J. G.; Nirenberg, M. W. An Electron Microscopic Study of a DNA-Ribosome Complex Formed in Vitro. *J. Mol. Biol.* **1965**, *11* (1), 78–83.

[8] a) For lead references in codons and translation see: Gamow, G. Possible Relation between Deoxyribonucleic Acid and Protein Structures [19]. *Nature*. 1954, p 318. b) Crick, F. H. C.; Barnett, L.; Brenner, S.; Watts-Tobin, R. J. General Nature of the Genetic Code for Proteins. *Nature* **1961**, *192* (4809), 1227–1232. c) Nirenberg, M. W.; Matthaei, J. H. The Dependence of Cell-Free Protein Synthesis in E. Coli upon Naturally Occurring or Synthetic Polyribonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **1961**, *47*, 1588–1602. d) Khorana, H. G. Synthesis in the Study of Nucleic Acids. The Fourth Jubilee Lecture. *The Biochemical journal*. 1968, pp 709–725.

In evaluating the origin of life, there are many factors to consider, such as whether metabolic function[9] pre or post-dates genetic information storage and when compartmentalization via a protocellular membrane[10] was essential.[11] However, these important questions are outside the scope of this thesis and we will instead focus on the importance of identifying a primitive biopolymer necessary for biogenesis, which would contain the ability to encode genetic information (i.e. DNA) and/or display beneficial catalytic activity for life (i.e. proteins). Two main schools of thought exist here, a "Protein first world" and a "Nucleic acid first world," as Orgel put it:

"There are only two obvious alternatives to the present genetic system which need to be considered, namely:
(1) Life based on proteins in the absence of nucleic acids
(2) Life based on nucleic acids in the absence of proteins."- Orgel[12]

Let's explore the first of Orgel's two proposals: that life originated from proteins in the absence of nucleic acids.[13] This is appealing from the standpoint that many amino acids are prebiotically available, as initially demonstrated by Miller.[2] In addition, research into this

---

[9] For lead references on metabolic function in emergence of life see: a) Keller, M. A.; Turchyn, A. V.; Ralser, M. Non-Enzymatic Glycolysis and Pentose Phosphate Pathway-like Reactions in a Plausible Archean Ocean. *Mol. Syst. Biol.* **2014**, *10* (4). b) Ralser, M. The RNA World and the Origin of Metabolic Enzymes. In *Biochemical Society Transactions*; Portland Press Ltd, 2014; Vol. 42, pp 985–988. c) Wachtershauser, G. Evolution of the First Metabolic Cycles. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87* (1), 200–204.

[10] For perspectives on cellular membranes in prebiotic chemistry see: a) Orgel, L. E. Prebiotic Chemistry and the Origin of the RNA World. *Critical Reviews in Biochemistry and Molecular Biology*. March 2004, pp 99–123. b) Szostak, J. W.; Bartel, D. P.; Luisi, P. L. Synthesizing Life. *Nature*. January 18, 2001, pp 387–390. c) Higgs, P. G.; Lehman, N. The RNA World: Molecular Cooperation at the Origins of Life. *Nature Reviews Genetics*. Nature Publishing Group January 11, 2015, pp 7–17.

[11] For work on minimal bacterial genomes see: a) Glass, J. I.; Assad-Garcia, N.; Alperovich, N.; Yooseph, S.; Lewis, M. R.; Maruf, M.; Hutchison, C. A.; Smith, H. O.; Venter, J. C. Essential Genes of a Minimal Bacterium. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (2), 425–430. b) Hutchison, C. A.; Chuang, R. Y.; Noskov, V. N.; Assad-Garcia, N.; Deerinck, T. J.; Ellisman, M. H.; Gill, J.; Kannan, K.; Karas, B. J.; Ma, L.; et al. Design and Synthesis of a Minimal Bacterial Genome. *Science* **2016**, *351* (6280).

[12] Orgel, L. E. Evolution of the Genetic Apparatus. *J. Mol. Biol.* **1968**.

[13] Oparin, A. I. (Aleksandr I.; Morgulis, S. *The Origin of Life*; Dover Publications, 2003.

hypothesis has identified a coiled-coil peptide hypercycle that can catalyze other peptide formation, with a positive feedback loop allowing for non-linear autocatalysis.[14] Certain sequence specific small peptides have catalytic activity with the ability to self-replicate.[15][16] Drawbacks to this proposal are that untemplated assembly of peptides would be necessary and is challenging due to the pKa mismatch in a peptide coupling reaction. Second, the secondary and tertiary structure ubiquitous in protein conformational space makes it hard to imagine a templating function that would allow for replication of information, and furthermore it provides no reasonable pathway to modern genetic storage via nucleic acid chemistry.[17]

On the other hand, the ability of RNA to act as a carrier of genetic information[18] and as a catalyst indispensable to cellular function[19][20] has led to the "RNA World" hypothesis, a term coined by Gilbert in 1986.[21] The RNA world hypothesis avers that RNA was the first biopolymer due to its ability to encode genetic information and carry out catalytic functions, and that cells

---

[14] Lee, D. H.; Severin, K.; Yokobayashi, Y.; Ghadiri, M. R. Emergence of Symbiosis in Peptide Self-Replication through a Hypercyclic Network. *Nature* **1997**, *390* (6660), 591–594.

[15] Li, Y.; Zhao, Y.; Hatfield, S.; Wan, R.; Zhu, Q.; Li, X.; McMills, M.; Ma, Y.; Li, J.; Brown, K. L.; et al. Dipeptide Seryl-Histidine and Related Oligopeptides Cleave DNA, Protein, and a Carboxyl Ester. *Bioorganic Med. Chem.* **2000**, *8* (12), 2675–2680.

[16] Lee, D. H.; Granja, J. R.; Martinez, J. A.; Severin, K.; Ghadiri, M. R. A Self-Replicating Peptide. *Nature* **1996**, *382* (6591), 525–528.

[17] Orgel, L. E. The Origin of Life--a Review of Facts and Speculations. *Trends Biochem. Sci.* **1998**, *23* (12), 491–495.

[18] Volkin, E.; Astrachan, L. Phosphorus Incorporation in Escherichia Coli Ribonucleic Acid after Infection with Bacteriophage T2. *Virology* **1956**, *2* (2), 149–161.

[19] a) Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R. Self-Splicing RNA: Autoexcision and Autocyclization of the Ribosomal RNA Intervening Sequence of Tetrahymena. *Cell* **1982**. b) Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. The RNA Moiety of Ribonuclease P Is the Catalytic Subunit of the Enzyme. *Cell* **1983**.

[20] Moore, P. B.; Steitz, T. A. The Involvement of RNA in Ribosome Function. *Nature*. July 11, 2002, pp 229–235.

[21] Gilbert, W. Origin of Life: The RNA World. *Nature* **1986**.

would later or co-evolve DNA to become the primary storage of genetic information, while proteins would carry out the complex cellular catalytic machinery.[22] Even before the catalytic function of RNA was known, Orgel and Crick both championed the RNA world.[12,2324] Crick noted the plethora of roles that RNA takes on, as well as its ubiquity in cellular machinery was "…giving him an overwhelming feeling that RNA was part of the primitive machinery of life."[24] Orgel went as far as to remark:

"It may be claimed, without too much exaggeration, that the problem of the origin of life is the problem of the origin of the RNA World, and that everything that followed is in the domain of natural selection"- Orgel[25]

Further secondary evidence for RNA being responsible for biogenesis is that it plays a vestigial role in a range of co-factors, including: ATP, Acetyl-CoA, cobalamin, FAD, and NADH.[26] One thought is that these are ancient remnants of the RNA world that can be found in modern cellular life. Since the initial discovery of RNA's catalytic activity, numerous other processes have been

---

[22] Joyce, G. F. RNA Evolution and the Origins of Life. *Nature*. **1989**, pp 217–224.

[23] Woese, C. R. The Genetic Code: The Molecular Basis for Genetic Expression. *New York Harp Row* **1967**.

[24] Crick, F. H. The Origin of the Genetic Code. *J. Mol. Biol.* **1968**, *38* (3), 367–379.

[25] Orgel, L. E. Prebiotic Chemistry and the Origin of the RNA World. *Critical Reviews in Biochemistry and Molecular Biology*. **2004**.

[26] For references on RNA in co-factors see: a) Neveu, M.; Kim, H.-J.; Benner, S. A. The "Strong" RNA World Hypothesis: Fifty Years Old. *Astrobiology* **2013**, *13* (4), 391–403. b) White, H. B. Coenzymes as Fossils of an Earlier Metabolic State. *J. Mol. Evol.* **1976**, *7* (2), 101–104. c) Benner, S. A.; Ellington, A. D.; Tauer, A. Modern Metabolism as a Palimpsest of the RNA World. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86* (18), 7054–7058.

found to be catalyzed by RNA including: self-splicing,[19a,27] RNA cleavage,[19b] RNA ligation,[28] phosphorylation,[29] peptide bond formation,[30] and other processes.[31]

In evaluating the RNA world hypothesis, there are a number of requirements that must be met for RNA to be a suitable progenitor to all life. Orgel elegantly lays out 4 subproblems, or criteria that must be met by any primitive biopolymer, which has been applied to RNA:[25]

"1. The nonenzymatic synthesis of nucleotides.

2. The nonenzymatic polymerization of nucleotides to give random-sequence RNA.

3. The nonenzymatic copying or replication or both, of RNA.

4. The emergence through natural selection of a set of functional RNA catalysts that together could sustain exponential growth in the prebiotic environment."

---

[27] Peebles, C. L.; Perlman, P. S.; Mecklenburg, K. L.; Petrillo, M. L.; Tabor, J. H.; Jarrell, K. A.; Cheng, H. L. A Self-Splicing RNA Excises an Intron Lariat. *Cell* **1986**, *44* (2), 213–223.

[28] a) Bartel, D. P.; Szostak, J. W. Isolation of New Ribozymes from a Large Pool of Random Sequences. *Science* **1993**, *261* (5127), 1411–1418. b) Ekland, E. H.; Szostak, J. W.; Bartel, D. P. Structurally Complex and Highly Active RNA Ligases Derived from Random RNA Sequences. **1995**, *269* (5222), 364–370.

[29] Lorsch, J. R.; Szostak, J. W. In Vitro Evolution of New Ribozymes with Polynucleotide Kinase Activity. *Nature* **1994**, *371* (6492), 31–36.

[30] Nissen, P.; Hansen, J.; Ban, N.; Moore, P. B.; Steitz, T. A. The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. *Science (80-.).* **2000**, *289* (5481), 920–930.

[31] Chen, X.; Li, N.; Ellington, A. D. Ribozyme Catalysis of Metabolism in the RNA World. *Chemistry and Biodiversity*. 2007, pp 633–655.

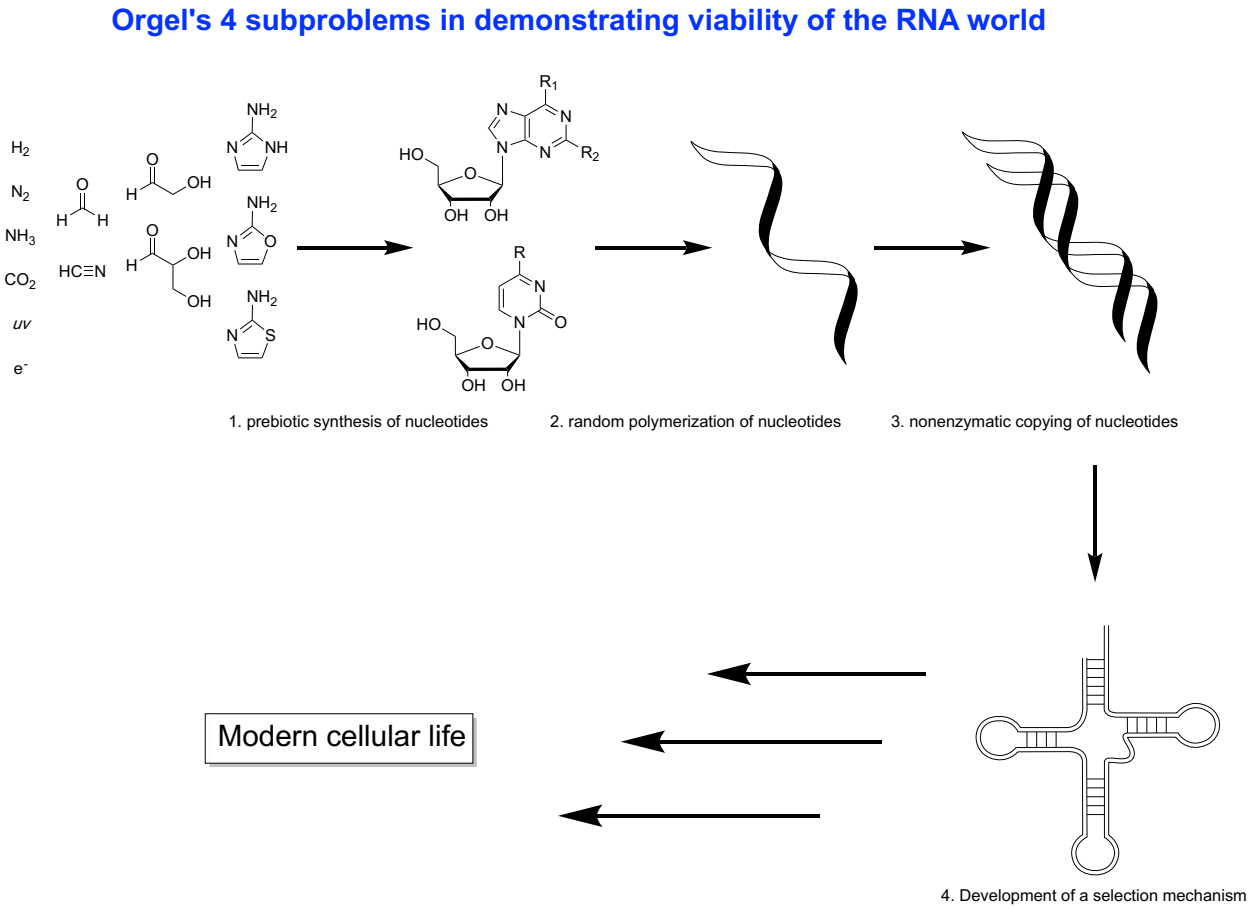


Figure 1.1: Orgel's 4 subproblems on determining the viability of RNA as the first biopolymer responsible for the emergence of life

The first of these criteria regards the nonenzymatic synthesis of nucleotides from molecules considered to be prebiotically available (Figure 1.1). This is an active area of research, and is the subject of section 1.2 of this introduction. To summarize that section, there currently exists prebiotically viable syntheses of ribopyrimidines,[32] but syntheses of ribopurines remains challenging.[33][34] The origin of homochirality is still puzzling and furthermore, the production of

[32] Powner, M. W.; Gerland, B.; Sutherland, J. D. Synthesis of Activated Pyrimidine Ribonucleotides in Prebiotically Plausible Conditions. *Nature* **2009**.

[33] Benner, S. A.; Kim, H. J.; Carrigan, M. A. Asphalt, Water, and the Prebiotic Synthesis of Ribose, Ribonucleosides, and RNA. *Acc. Chem. Res.* **2012**, *45* (12), 2025–2034.

[34] Stairs, S.; Nikmal, A.; Bŭar, D. K.; Zheng, S. L.; Szostak, J. W.; Powner, M. W. Divergent Prebiotic Synthesis of Pyrimidine and 8-Oxo-Purine Ribonucleotides. *Nat. Commun.* **2017**.

pentose stereoisomers and non-canonical nucleobases en route to ribonucleotide synthesis implies these molecules have structurally similar impurities that must be evaluated as beneficial or inhibitory to the other stages of an RNA world as laid out in Orgel's 4 subproblems above.

The second consists of the non-enzymatic uncontrolled polymerization of RNA monomers to give random RNA oligomeric sequences (Figure 1.1). In random polymerization of RNA monomers, regioselectivity in phosphate linkage remains an issue,[35] as well as the requirements of large amount of metal ions.[36] Furthermore, if RNA is not produced cleanly in nature, it may polymerize with other nucleotides, such as arabinonucleotides, as discussed in Chapter 3. We will largely not focus on the second step in this thesis.

The third criterion is the nonenzymatic copying or replication of RNA. This has been the subject of much study since the 1960's, when it was shown that in the absence of any enzymes, poly C RNA can catalyze the assembly of poly G oligonucleotide strands in the presence of activated G monomers.[37] This provides a model by which an RNA strand serves as a template for the assembly of a complementary strand oligonucleotide from activated monomers, allowing a pathway for informational inheritance.[38] The specifics of and progress made in nonenzymatic

---

[35] a) Morávek, J. Formation of Oligonucleotides during Heating of a Mixture of Uridine 2′(3′) -Phosphate and Uridine. *Tetrahedron Lett.* **1967**, *8* (18), 1707–1710. b) Prabahar, K. J.; Ferris, J. P. Adenine Derivatives as Phosphate-Activating Groups for the Regioselective Formation of 3',5'-Linked Oligoadenylates on Montmorillonite: Possible Phosphate-Activating Groups for the Prebiotic Synthesis of RNA. *J. Am. Chem. Soc.* **1997**, *119* (19), 4330–4337.

[36] a) Sawai, H.; Orgel, L. E. Oligonucleotide Synthesis Catalyzed by the Zn2+ Ion. *Journal of the American Chemical Society*. **1975**. b) Sawai, H. Catalysis of Internucleotide Bond Formation by Divalent Metal Ions. *J. Am. Chem. Soc.* **1976**, *98* (22), 7037–7039. c) Kanavarioti, A.; Monnard, P.-A.; Deamer, D. W. Eutectic Phases in Ice Facilitate Nonenzymatic Nucleic Acid Synthesis. *Astrobiology* **2003**.

[37] Weimann, B. J.; Lohrmann, R.; Orgel, L. E.; Schneider-Bernloehr, H.; Sulston, J. E. Template-Directed Synthesis with Adenosine-5′-Phosphorimidazolide. *Science* **1968**.

[38] Orgel, L. E. Some Consequences of the RNA World Hypothesis. In *Origins of Life and Evolution of the Biosphere*; **2003**; Vol. *33*, pp 211–218.

copying of oligonucleotides is the subject of section 1.3 of this intro. Furthermore, the evaluation of non-canonical prebiotically plausible nucleotides in nonenzymatic primer extension is the subject of both Chapters 2 and 3 of this thesis.

The fourth constraint concerns the emergence of a pathway for selection of functional sequences in the RNA world. Very little work has been done on this subproblem, as our understanding of the first three are not yet good enough to make meaningful progress here. In addition, secondary requirements may be needed for selection such as the emergence of metabolic pathways, encapsulation in membranes, or the co-evolution of a second biopolymer, all of which require systems biology approaches. The recent explosion of interest in directed evolution of modern bacterial life may start to reveal more about the constraints for evolving simple systems.[39]

1.2 Prebiotic Chemistry

The abiotic synthesis of the molecules essential to life is an existential problem that has captured the attention of many chemists since the early 1950s. In a seminal report from Miller and Urey, the 1950s view of the conditions of a prebiotic earth (which are somewhat more reducing than a modern understanding of prebiotic conditions) were replicated by passing an electric discharge through a gaseous mixture of water vapor, methane, ammonia, and hydrogen.[2] From these simple starting materials, each containing at most one non-hydrogenic atom, Miller was able to identify 5 amino acids. Subsequent reanalysis of their data, along with related experiments conducted in the 2000s were able to produce 32 amino acids, 12 of which are proteinogenic.[40]

---

[39] For recent reviews on directed evolution see: a) Packer, M. S.; Liu, D. R. Methods for the Directed Evolution of Proteins. *Nature Reviews Genetics* **2015**, pp 379–394. b) Arnold, F. H. Directed Evolution: Bringing New Chemistry to Life. *Angew. Chemie - Int. Ed.* **2018**, *57* (16), 4143–4148.

[40] a) Johnson, A. P.; Cleaves, H. J.; Dworkin, J. P.; Glavin, D. P.; Lazcano, A.; Bada, J. L. The Miller Volcanic Spark Discharge Experiment. *Science* **2008**, *322* (5900), 404. b) Parker, E. T.; Cleaves, H. J.; Dworkin, J. P.; Glavin, D. P.; Callahan, M.; Aubrey, A.; Lazcano, A.; Bada, J. L. Primordial Synthesis of Amines and Amino Acids in a 1958 Miller H 2S-Rich Spark Discharge Experiment. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (14), 5526–5531.

While the Miller experiment provided the first evidence that the building blocks of modern life can be produced abiotically, for the RNA world hypothesis to hold true, RNA monomeric units must exist in some abundance on the prebiotic earth. Indeed, the nonenzymatic synthesis of nucleotides is the first of Orgels 4 subproblems in investigating the origin of life, as discussed above.[25]

What exactly constitutes a plausible nonenzymatic synthesis, requires some definition. For example, does a 1% yielding reaction, or the characterization of a short-lived intermediate constitute a nonenzymatic synthesis? Fortunately, again we can turn to Orgel for an elegant definition of the requirements and restraints for a prebiotic synthesis to be considered viable:[25,41]

"(1) *It must be plausible, at least to the proposers of a prebiotic synthesis, that the starting materials for a synthesis could have been present in adequate amounts at the site of synthesis.*

(2) *Reactions must occur in water or in the absence of a solvent*.

(3) *Yield of the product must be "significant," at least in the view of the proposers of the synthesis*."

The first restraint is that required building blocks for the proposed synthesis have to be sufficiently abundant. Despite recent progress in our understanding of the early earth that gives us insights into prebiotic chemistry,[42] prebiotic availability of simple chemicals is often challenging to address. Generally, molecules derived from very simple chemical building blocks, such as cyanide, hydrogen sulfide, and formaldehyde are considered to be prebiotically plausible. Second, synthetic steps towards the desired, prebiotic product must occur in no solvent or water. Lastly, the synthesis needs to generate a sufficient quantity of the desired product and be selective towards

---

41 Orgel, L. E.; Lohrmann, R. Prebiotic Chemistry and Nucleic Acid Replication. *Acc. Chem. Res.* **1974**, *7* (11), 368–377.

42 McCollom, T. M. Miller-Urey and Beyond: What Have We Learned About Prebiotic Organic Synthesis Reactions in the Past 60 Years? *Annu. Rev. Earth Planet. Sci.* **2013**, *41* (1), 207–229.

the desired product.

Unlike enzymes that achieve remarkable selectivity via a number of noncovalent interactions, prebiotic chemical reactions rely on the reactivity of simple starting materials and, thus, often suffer from low selectivity, which leads to various side products and byproducts. Throughout this chapter, many reported syntheses produce a diverse range of products in addition to the desired target, ribonucleotides. These byproducts are often structurally very similar to ribonucleotides, with slight modifications that could potentially be beneficial or detrimental to the downstream chemistry that would be required of RNA if it were responsible for biogenesis.

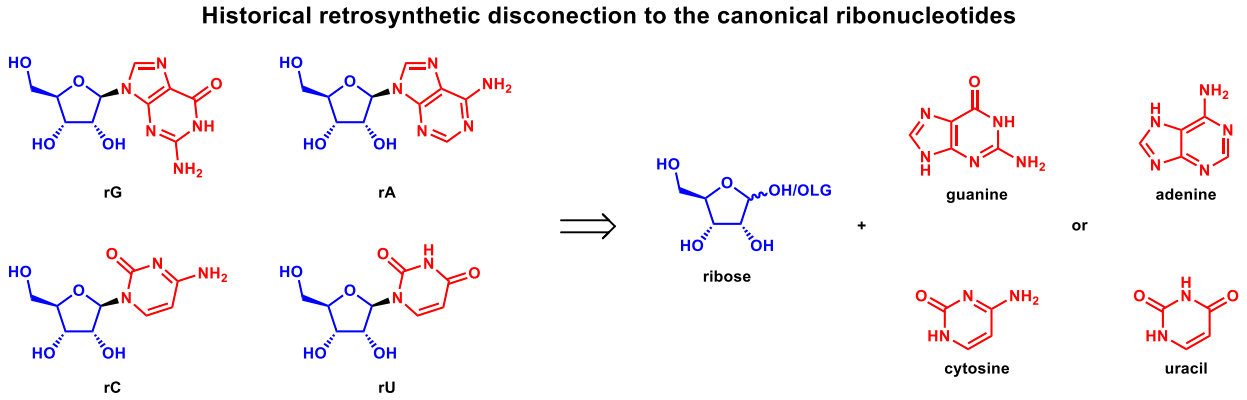


Figure 1.2: Traditional synthesis of ribonucleotides via the glycosidation of nucleobases

A nucleotide consists of a sugar and a nucleobase, so naively, the problem of ribonucleotide synthesis can be split into three subunits, the prebiotic synthesis of 1) sugars (specifically ribose), 2) nucleobases (specifically A, U, G, and C), and 3) an efficient and selective glycosidation reaction coupling the two halves (Figure 1.2). Despite the appealing simplicity of this approach, the following paragraphs will show numerous problems with this proposed route to ribonucleotides. To summarize the following paragraphs, early reports suggest very simple syntheses of some of the desired products from the polymerization of prebiotically plausible starting materials; however, the glycosidation step has been particularly challenging and later discoveries have refuted the likelihood of these pathways.

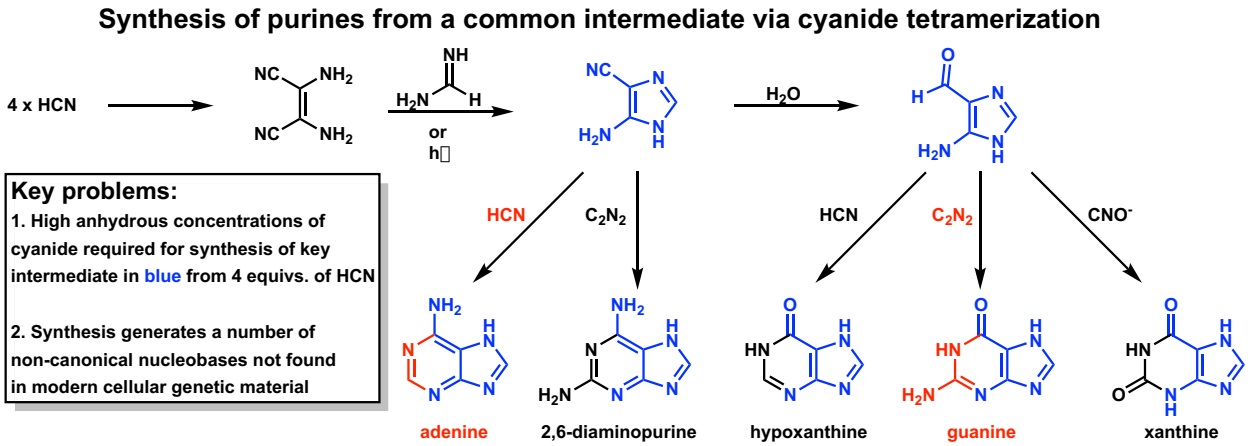


Figure 1.3: Divergent synthesis of purines from HCN tetramer intermediate (in blue)

A seminal report from Oró and Kimball in 1961 showed that purine nucleobases (such as adenine) can arise from the polymerization of hydrogen cyanide (Figure 1.3) in a refluxing solution of aqueous ammonia.[43] In this report, the yield of desired, canonical nucleobases is often very low with approximately 0.2% yield of adenine from cyanide (1L of 11.1 M HCN in aqueous ammonia yields at most 685 mgs of adenine, $C_5H_5N_5$). Despite some modifications to the reaction conditions that can improve the yield, these conditions are not prebiotically plausible (i.e. very high concentration of cyanide, and high pressure).[44] A eutectic phase reaction allows the polymerization to occur at a much more prebiotically plausible condition of 10 mM concentration of cyanide, and photoisomerization of a tetramer intermediate followed by subsequent reactions, allows for a divergent synthesis of multiple purine nucleobases.[45] However, a diverse range of purines not found in modern genetic material, such as xanthine, hypoxanthine and 2,6-diaminopurine, are

[43] a) Oró, J. Synthesis of Adenine from Ammonium Cyanide. *Biochem. Biophys. Res. Commun.* **1960**, *2* (6), 407–412. b) Oró, J.; Kimball, A. P. Synthesis of Purines under Possible Primitive Earth Conditions. I. Adenine from Hydrogen Cyanide. *Arch. Biochem. Biophys.* **1961**, *94* (2), 217–227.

[44] a) Levine, J. S.; Augustsson, T. R.; Natarajan, M. The Prebiological Paleoatmosphere: Stability and Composition. *Orig. Life* **1982**, *12* (3), 245–259. b) Pizzarello, S.; Williams, L. B. Ammonia in the Early Solar System: An Account from Carbonaceous Meteorites. *Astrophys. J.* **2012**, *749* (2). c) Shapiro, R. The Improbability of Prebiotic Nucleic Acid Synthesis. *Orig. Life* **1984**, *14* (1–4), 565–570.

[45] Schwartz, A. W.; Joosten, H.; Voet, A. B. Prebiotic Adenine Synthesis via HCN Oligomerization in Ice. *BioSystems* **1982**, *15* (3), 191–193.

generated as byproducts (Figure 1.3).[46] If a glycosidation method were found, why adenine and guanine glycosidation would be preferred over non-canonical nucleobase glycosidation would be both entirely unclear and unlikely.

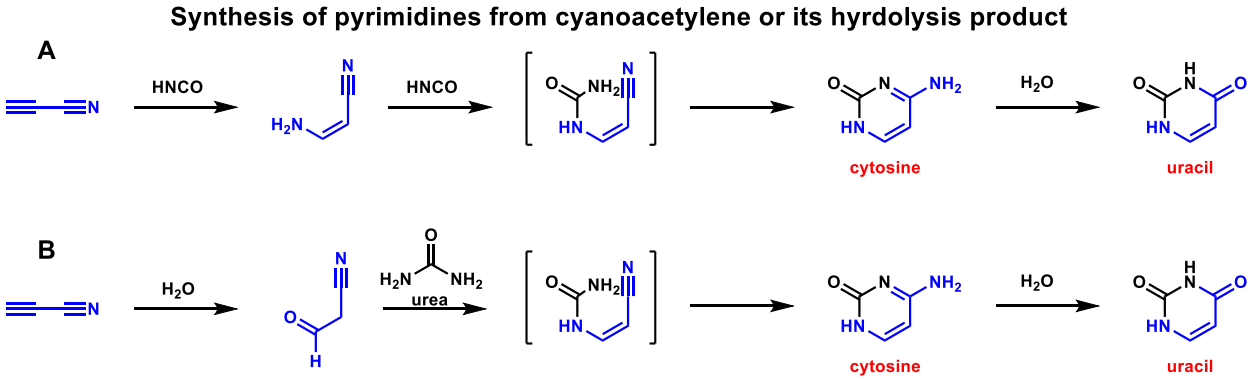


Figure 1.4: Prebiotic synthesis of pyrimidine nucleobases from cyanoacetylene or cyanoacetaldehyde, its hydrolysis product

There are two related syntheses of pyrimidine nucleotides, from cyanoacetylene. Cyanoacetylene is prebiotically plausible, as it is formed in non-negligible yields from a nitrogen and methane environment with an applied electric discharge.[47] In the first synthesis (Figure 1.4 A), cyanoacetylene reacts with 2 equivalents of hydrogen cyanate (1.0 M) to give cytosine in 5% yield, which can hydrolyze to uracil. The second synthesis (Figure 1.4 B) involves the reaction between cyanoacetaldehyde, the hydrolysis product of cyanoacetylene,[48] and urea (1.0 M) to form cytosine, also in 5% yield. In saturated urea solutions, this route can provide a 50% yield of cytosine,[49] but

---

[46] Sanchez, R. A.; Ferris, J. P.; Orgel, L. E. Studies in Prebiotic Synthesis. IV. Conversion of 4-Aminoimidazole-5-Carbonitrile Derivatives to Purines. *J. Mol. Biol.* **1968**, *38* (1), 121–128.

[47] Ferris, J. P.; Sanchez, R. A.; Orgel, L. E. Studies in Prebiotic Synthesis. III. Synthesis of Pyrimidines from Cyanoacetylene and Cyanate. *J. Mol. Biol.* **1968**, *33* (3), 693–704.

[48] Ferris, J. P.; Goldstein, G.; Beaulieu, D. J. Chemical Evolution. IV. An Evaluation of Cyanovinyl Phosphate as a Prebiotic Phosphorylating Agent. *J. Am. Chem. Soc.* **1970**, *92* (22), 6598–6603.

[49] Robertson, M. P.; Miller, S. L. An Efficient Prebiotic Synthesis of Cytosine and Uracil. *Nature* **1995**, *375* (6534), 772–774.

conditions involving saturated urea solutions have been called into question.[50] The current state of the art synthesis of purines and pyrimidines both occur in eutectic solutions, which are attractive as they would allow all four nucleobases to be present in the same location at the same time.[25] Despite progress, there are still key problems in the synthesis of both purine and pyrimidine nucleobases, not the least of which is the prohibitively low solubility of nucleobases, which further questions the validity of this approach.[51]

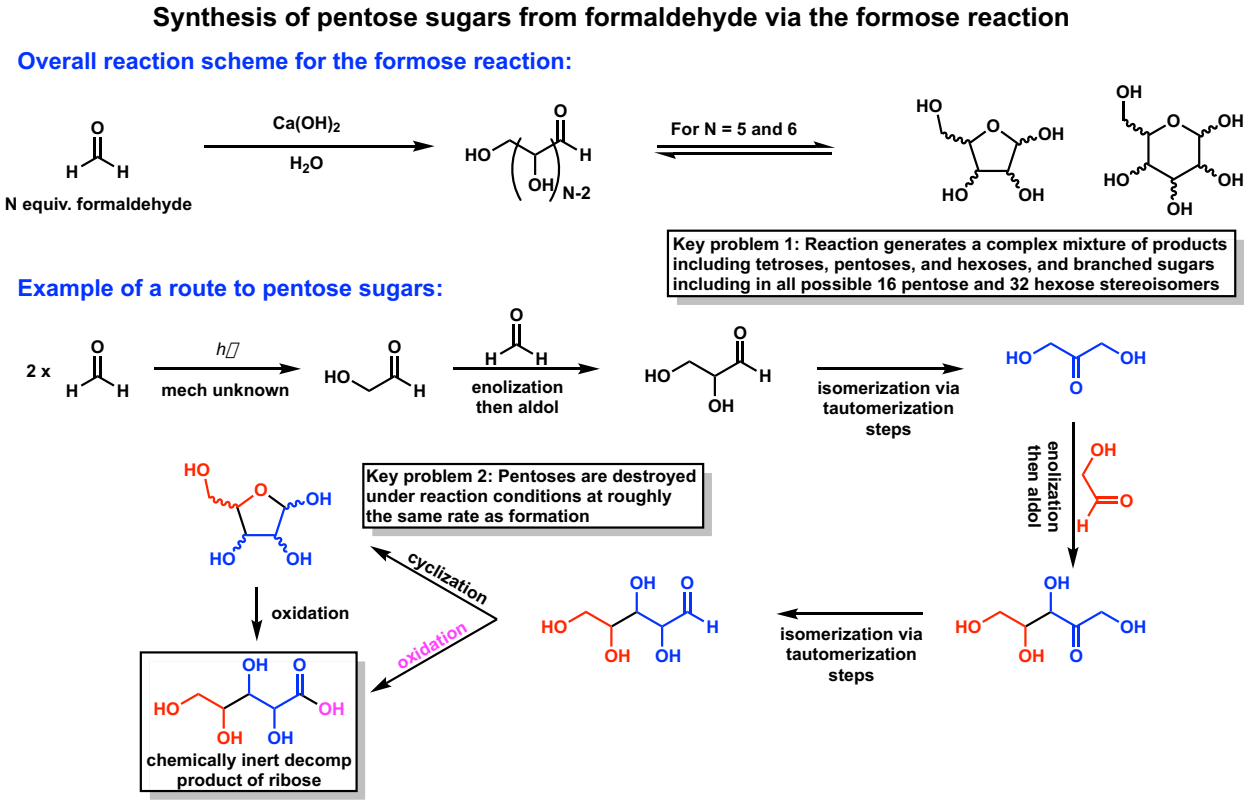


Figure 1.5: Formose reaction produces sugars for formaldehyde but is plagued with selectivity problems and destruction of pentose products at a rate similar to that of formation

The first prebiotic synthesis of ribose was proposed by Butlerov via the polymerization of

[50] a) Shapiro, R. Prebiotic Cytosine Synthesis: A Critical Analysis and Implications for the Origin of Life. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (8), 4396–4401 b) Shapiro, R. Comments on Concentration by Evaporation and the Prebiotic Synthesis of Cytosine. *Orig. Life Evol. Biosph.* **2002**, *32* (3), 275–278

[51] Albert, A.; Brown, D. J. Purine Studies. Part I. Stability to Acid and Alkali. Solubility. Ionization. Comparison with Pteridines. *J. Chem. Soc.* **1954**, 2060–2071.

formaldehyde in the presence of $Ca(OH)_2$ base (also known as the formose or Butlerov reaction) in 1861 (Figure 1.5).[52] The synthesis seemed to be very attractive for both availability of formaldehyde, which can be furnished via spark discharge of carbon dioxide, and the simplicity of the synthesis.[53] While the initial step involving the dimerization of formaldehyde to glycoaldehyde (both found in interstellar space[54] still a subject of debate, Breslow shed a great deal of light on the mechanism in 1959, proposing a series of aldol, retroaldol and tautomerization steps yielding a range of sugars.

While this transformation is prebiotically plausible, there are a couple of key problems. First, the reaction is entirely unselective and generates tetroses, hexoses, and a range of pentoses, including ribose, arabinose, xylose, and lyxose (Figure 1.5).[55] Later work with modern analytical tools have shown that the synthesis requires some forms of an initiation, likely due to the presence of a trace glycoaldehyde impurity, which may be formed via uv promoted radical initiation, in formaldehyde that catalyzes this transformation[56] and, additionally, the characterization of a different chemical pathway (Cannizaro reaction) that generates formic acid and methanol via

---

[52] Butlerov, A. Académie des sciences (France). Comptes rendus hebdomadaires des séances de l'Académie des sciences **1861**.

[53] a) Bossard, A. R.; Raulin, F.; Mourey, D.; Toupance, G. Organic Synthesis from Reducing Models of the Atmosphere of the Primitive Earth with UV Light and Electric Discharges. *J. Mol. Evol.* **1982**, *18* (3), 173–178. b) Stribling, R.; Miller, S. L. Energy Yields for Hydrogen Cyanide and Formaldehyde Syntheses: The Hcn and Amino Acid Concentrations in the Primitive Ocean. *Orig. life Evol. Biosph.* **1987**, *17* (3), 261–273. c) Schlesinger, G.; Miller, S. L. Prebiotic Synthesis in Atmospheres Containing CH4, CO, and CO2 - I. Amino Acids. *J. Mol. Evol.* **1983**, *19* (5), 376–382. d) James, H.; II, C. The Prebiotic Geochemistry of Formaldehyde. *Precambrian Res.* **2008**, *164*, 111–118.

[54] Hollis, J. M.; Vogel, S. N.; Snyder, L. E.; Jewell, P. R.; Lovas, F. J. The Spatial Scale of Glycolaldehyde in the Galactic Center. *Astrophys. J.* **2001**, *554* (1), L81–L85.

[55] Decker, P.; Schweer, H.; Pohlamnn, R. Bioids. X. Identification of Formose Sugars, Presumable Prebiotic Metabolites, Using Capillary Gas Chromatography/Gas Chromatography-Mas Spectrometry of n-Butoxime Trifluoroacetates on OV-225. *J. Chromatogr. A* **1982**, *244* (2), 281–291.

[56] Socha, R. F.; Weiss, A. H.; Sakharov, M. M. Autocatalysis in the Formose Reaction. *React. Kinet. Catal. Lett.* **1980**, *14* (2), 119–128.

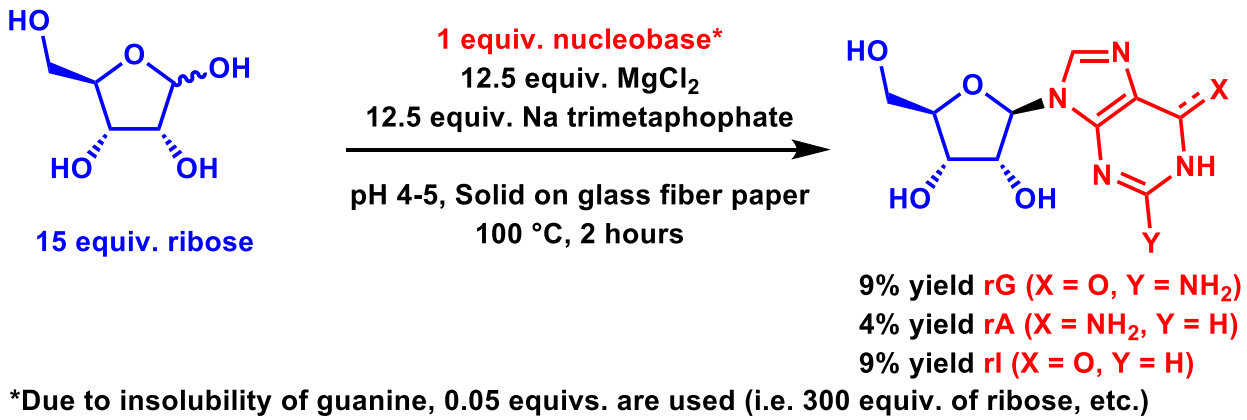
disproportionation of formaldehyde, causing low yields and a large induction period. Additionally, the stability of ribose, and all pentoses, under both its formation conditions and prebiotic conditions has been called into question (Figure 1.5).[57] Sugars are widely accepted to be unstable in strong acid or base, and Miller showed that at even at neutral pH 7.0, the half-life of ribose is 73 min at 100 °C and 44 years at 0 °C.[57a] Some modifications have shown to improve the reaction quite significantly (ie. Pb[58] or borate minerals[59] instead of Ca). Notably, Benner showed that the stability of pentoses bearing a 1,2-cis diol is greatly increased by the presence of borates (which have not been ruled out as prebiotically available), which the sugar chelates, protecting the sugars from decomposition.[58] However, even with modified syntheses, low yields of ribose with chemically similar byproducts (including its enantiomer) as a result of unselective formation of all pentoses, the instability of ribose, and incompatibility of ribose with nucleobase synthesis conditions (specifically, incompatibility with nitrogenous base) remain problematic and are unsolved problems in the prebiotic synthesis of ribose.

---

[57] a) Larralde, R.; Robertscn, M. P.; Miller, S. L. Rates of Decomposition of Ribose and Other Sugars: Implications for Chemical Evolution. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92* (18), 8158–8160. b) Shapiro, R. Prebiotic Ribose Synthesis: A Critical Analysis. *Orig. Life Evol. Biosph.* **1988**, *18* (1–2), 71–85.

[58] Zubay, G. Studies on the Lead-Catalyzed Synthesis of Aldopentoses. *Orig. Life Evol. Biosph.* **1998**, *28* (1), 13–26.

[59] Ricardo, A.; Carrigan, M. A.; Olcott, A. N.; Benner, S. A. Borate Minerals Stabilize Ribose. *Science* **2004**, *303* (5655), 196.



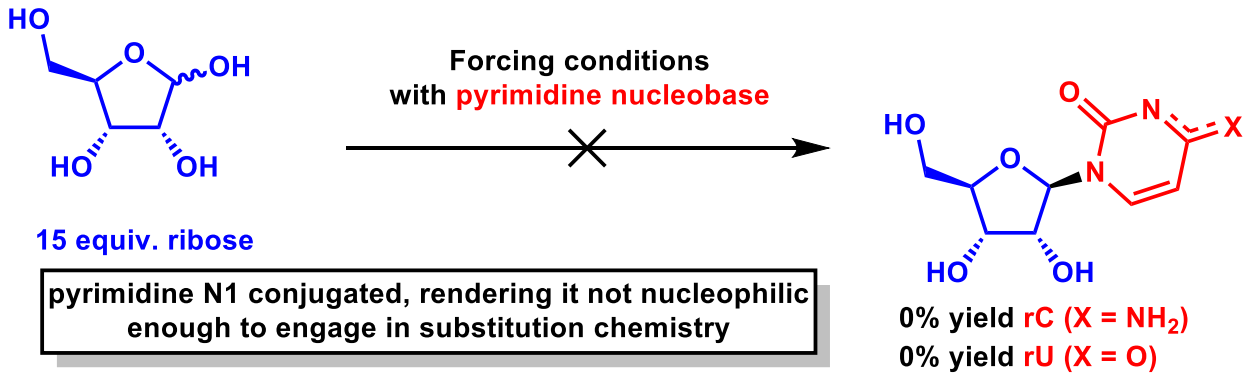


Figure 1.6: State of the art nucleobase glycosylation provides low yields of purine nucleotides and no yield of pyrimidine nucleotides

Lastly, the glycosidation of nucleobases on ribose has proven to be quite challenging. Under dry conditions, purine glycosidation using pure D-ribose generally gives a minimal yield in the case of adenine of 4% and guanine of 9% (Figure 1.6 A).[60] Instead of the desired ribonucleosides, multiple regioisomers and anomers are observed as major byproducts. On the nucleophile side, the major product arises from nucleophilic attack of the exocyclic nucleobase amine.[61] On the

[60] Fuller, W. D.; Sanchez, R. A.; Orgel, L. E. Studies in Prebiotic Synthesis. VI. Synthesis of Purine Nucleosides. *J. Mol. Biol.* **1972**, *67* (1), 25–33.

[61] Sutherland, J. D. Ribonucleotides. *Cold Spring Harbor perspectives in biology*. **2010**.

electrophile side, equilibrium conversion to the aldehyde leads to glycosidation of both furanose and pyranose forms, as well as unselective production of both α and β anomers.[62] While purine glycosidation can occur in low yields, pyrimidine glycosidation is presumed to be both kinetically and thermodynamically unfeasible (Figure 1.6 B). Conjugation of the N1 lone pair renders the nucleobase kinetically inert; furthermore, in water, the thermodynamics favor hydrolysis over nucleobase condensation.[25,32,61]

In summary, the conceptually simple idea of separate syntheses of nucleobases and ribose, followed by glycosidation suffers from so many problems that, at this point, it is considered an unlikely pathway.[63] To date, high yielding reactions that do not produce structurally and functionally similar byproducts do not exist for any nucleobase, as well as for ribose. Furthermore, glycosidation has only been demonstrated for adenine and guanine with <10% yield. Last, the conditions for the synthesis of ribose and nucleobases diverge, and ribose may decompose under many conditions. Therefore, the likelihood that a reservoir even moderately rich in ribose and nucleobases would come in contact to form an unprecedented glycosidation, is vanishingly small given the current state of the art in prebiotic synthesis. Therefore, other proposals for the synthesis of ribonucleotides must be considered for the RNA world hypothesis to hold water.

---

[62]Drew, K. N.; Zajicek, J.; Bondo, G.; Bose, B.; Serianni, A. S. 13C-Labeled Aldopentoses: Detection and Quantitation of Cyclic and Acyclic Forms by Heteronuclear 1D and 2D NMR Spectroscopy. *Carbohydr. Res.* **1998**, *307* (3–4), 199–209.

[63] Lazcano, A.; Miller, S. L. The Origin and Early Evolution of Life: Prebiotic Chemistry, the Pre-RNA World, and Time. *Cell*. **1996**, 793–798.

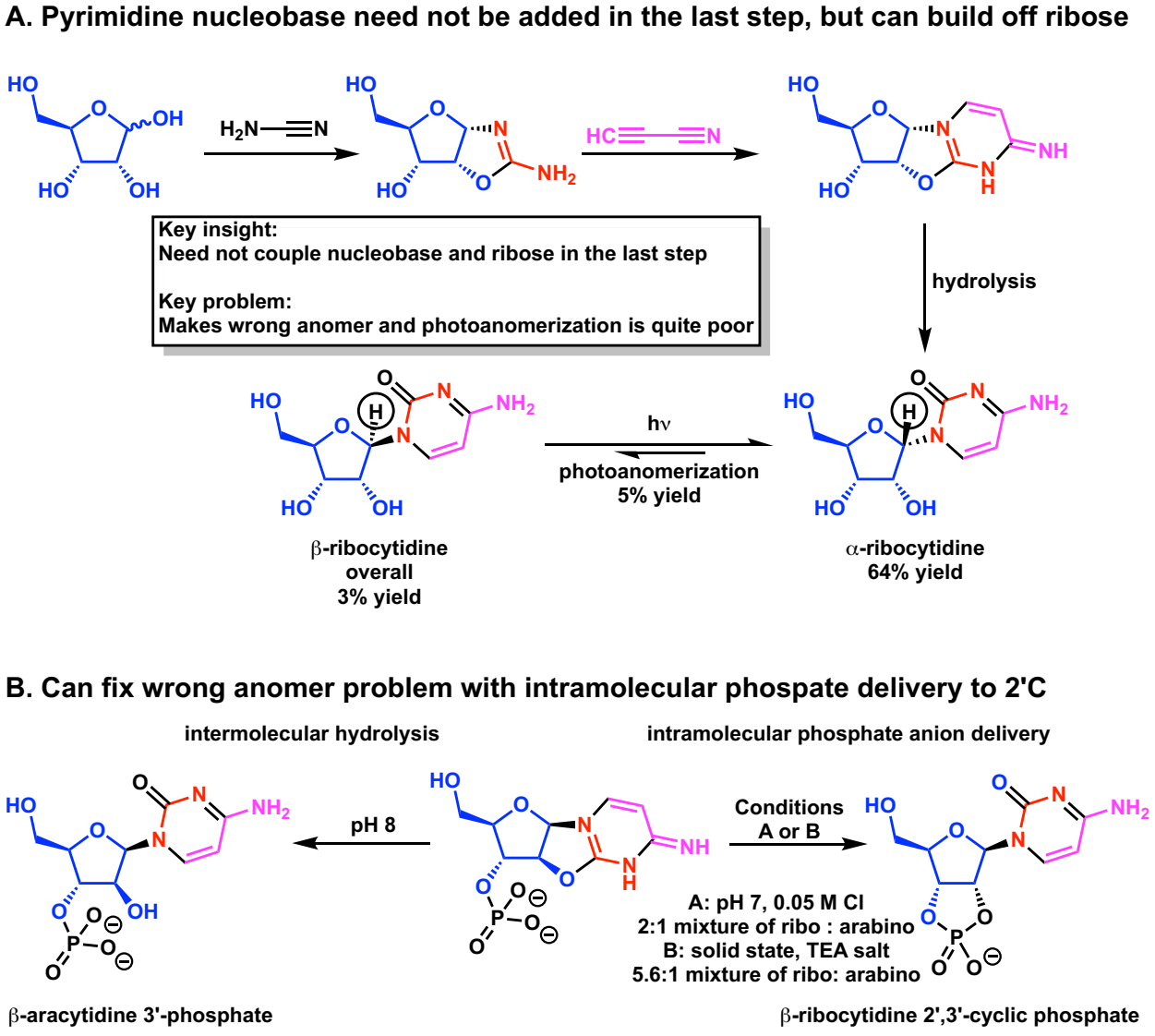


Figure 1.7: Key insights into next generation synthesis of pyrimidine ribonucleotides. A) Glycosidation of nucleobases need not be the last step, rather pyrimidines can be synthesized on ribose, but wrong anomer is made. B) The wrong anomer problem can be solved by starting with arabinose and using intramolecular delivery of the phosphate to the 2C' to form β-ribocytidine 2' : 3'-cyclic phosphate in competition with hydrolysis to β-aracytidine 3'-phosphate, suggesting coexistence on early earth.

As an alternative approach to the glycosidation condensation between ribose and nucleobases, a different approach has recently emerged as a much more reliable synthesis towards the pyrimidine ribonucleotides. The initial contribution to this synthesis was Sanchez and Orgel's attempts to circumvent the problematic glyclosidation step, in which the anomeric N arises from

the reaction between cyanamide and ribose in the presence of ammonia, yielding ribo-aminooxazoline in a remarkable yield of 87% (Figure 1.7 A).[64] Subsequent reaction with cyanoacetylene then hydrolysis, yields α-ribocytidine in 64% yield. The corresponding reaction with arabinose furnishes β-arabinocytidine. Attempts to photoisomerize the anomeric or 2' position were largely unsuccessful, with just a 5% yield of β-ribocytidine from pure α-ribocytidine. Despite the products being non-epimerizable epimers of the desired product, yield and selectivity for the novel route brought much insight. Another work that was crucial for the development of the current state of the art prebiotic synthesis is the realization that 3'-phosphate of β-arabinocytidine can facilitate the intramolecular attack on 2'-anhydrocytosine to invert the stereocenter at the 2' position to synthesize the desired β-ribocytidine (Figure 1.7 B).[65] Despite the novel route leading to the first prebiotic synthesis of a ribonucleotide, the lack of reliable prebiotic syntheses of arabinose remained problematic; additionally, nitrogenous compounds such as cyanamide, further complicate the synthesis of furanoses.[44c]

[64]Sanchez, R. A.; Orgel, L. E. Studies in Prebiotic Synthesis. V. Synthesis and Photoanomerization of Pyrimidine Nucleosides. *J. Mol. Biol.* **1970**, *47* (3), 531–543.

[65] a) Tapiero, C. M.; Nagyvary, J. Prebiotic Formation of Cytidine Nucleotides. *Nature* **1971**, *231* (5297), 42–43. b) Ingar, A.-A.; Luke, R. W. A.; Hayter, B. R.; Sutherland, J. D. Synthesis of Cytidine Ribonucleotides by Stepwise Assembly of the Heterocycle on a Sugar Phosphate. *Chembiochem* **2003**, *4* (6), 504–507.

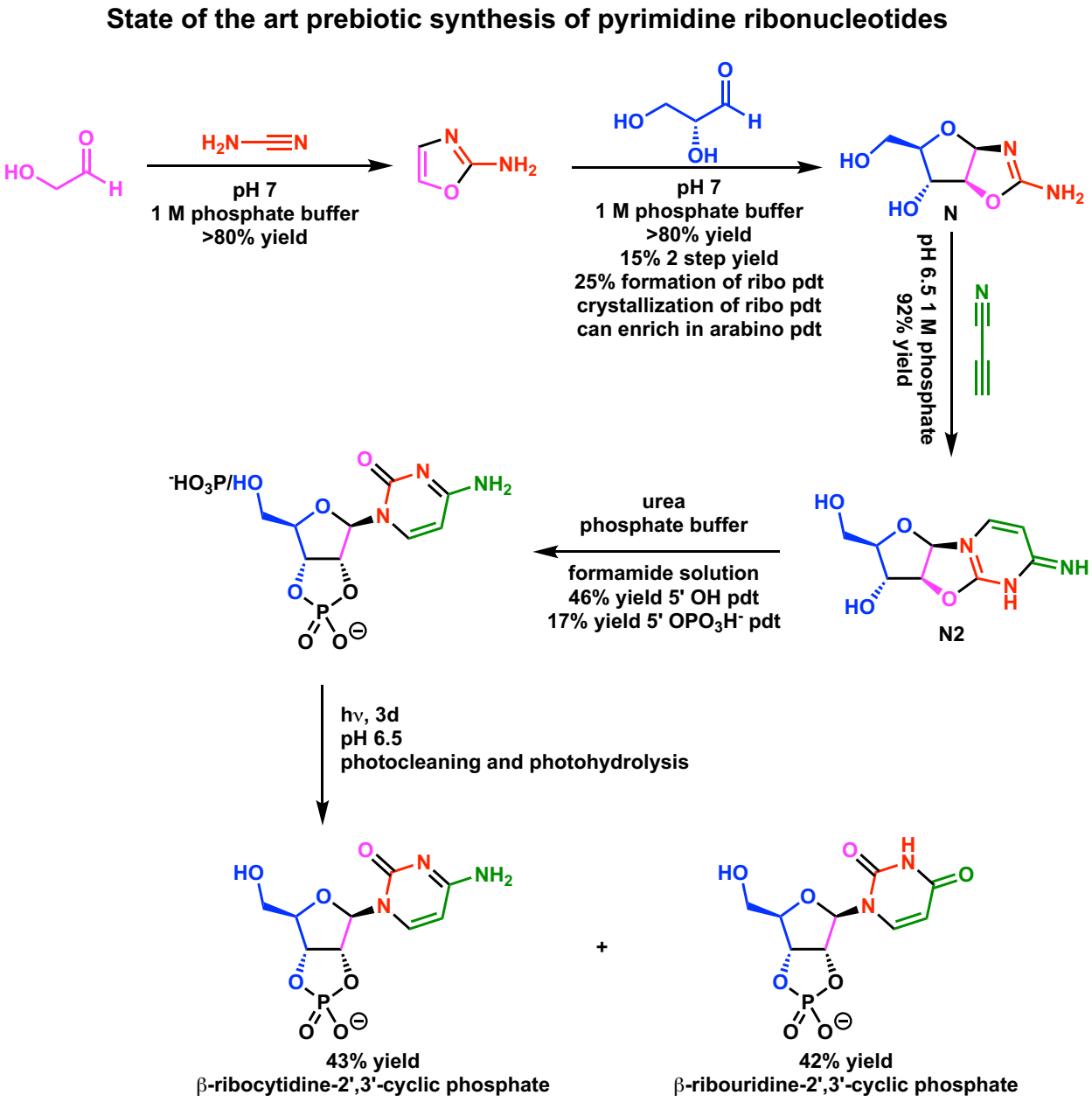


Figure 1.8: Powner and Sutherland's synthesis of pyrimidine ribonucleosides

Sutherland and Powner solved the problem regarding the availability of pentoses, by using glycoaldehyde and cyanamide to generate 2-aminooxazole (Figure 1.8).[32] 2-aminooxazole subsequently reacts with glyceraldehyde in the presence of phosphate buffer to form the desired product **N** in high yield of >80%, circumventing the need for a prebiotic synthesis of pentoses. The compound **N** serves as the key intermediate to generate the anhydroarabinonucleoside using the

same approach as Sanchez and Orgel,[64] with a key insight that the phosphate buffer maintains a pH high enough to prevent protonation of the nucleobase and premature hydrolysis, which leads to the incorrect stereochemistry at C3 **N2**. Subsequently the C3 phosphorylation strategy above was employed to invert the C2 stereocenter, giving the β-ribocytidine cyclic phosphate in up to 63% yield.[65a,b] Exposure of the product to ultraviolet light selectively destroys side products, while simultaneously converting half of the β-ribocytidine cyclic phosphate to β-ribouridine cyclic phosphate through deamination, providing access to both pyrimidine nucleotides.

This landmark synthesis by Sutherland and Powner mark a giant leap in the prebiotic availability of ribopyrimidines.[32] It elegantly starts from simple prebiotically available building blocks, circumvents the need to synthesize and build up a reservoir of ribose, and uses a phosphate buffer for all steps of the remarkably short synthesis. With that, there still exist two challenges to overcome. First, no satisfactory synthesis of purine ribonucleotides exists,[34] which will be discussed in subsequent chapters. Second, these syntheses of canonical building blocks are driven by a common intermediates, which is both a blessing and a curse. In some cases a common intermediate may provide a divergent synthesis to multiple desired building blocks. While in other cases, structurally and functionally similar byproducts (often stereoisomers) may be formed that could compete in subsequent steps or create diversity that does not resemble the 4 letter ribose system that modern biology employs. The evaluation of these byproducts as partial or complete substitutes for the canonical ribonucleotides then becomes a question that will be the focus of much of this thesis.

1.3 Nonenzymatic copying/replication of oligonucleotides

The template directed, nonenzymatic replication of RNA is one of the key subproblems in the RNA world hypothesis, as laid out by Orgel and discussed above.[25] Substantial progress has

been made in the first two subproblems 1) the abiotic synthesis of nucleotides and 2) the untemplated polymerization of the building block, which together demonstrate the prebiotic availability of long, random sequences of nucleotides. The third subproblem, is the nonenzymatic replication of random oligonucleotide templates to generate multiple copies, which then would lead to an emergence of a set of functional ribozymes via natural selection able to sustain exponential growth in the prebiotic world.[25]

Prior to the emergence of ribozymes, RNA oligonucleotides would have to rely on chemical replication to inherit genetic information. Unlike modern cellular life that utilizes nucleoside triphosphate as substrates in replication/transcription, nonenzymatic copying with nucleoside triphosphates has been shown to be extremely unlikely in the prebiotic world for several reasons (Figure 1.9 A). First, the poor permeability of nucleoside triphosphates, primarily caused by multiply charged triphosphate groups, limits the concentration of nucleotides in protocells. Second, nonenzymatic, templated directed copying with nucleoside triphosphates as substrates has been shown to proceed prohibitively slowly with a half-life of approximately 15 to 30 years in pH 7.4 and 100 mM Mg^{2+} in the case of ligation.[66]

[66] a) Rohatgi, R.; Bartel, D. P.; Szostak, J. W. Nonenzymatic, Template-Directed Ligation of Oligoribonucleotides Is Highly Regioselective for the Formation of 3′-5′ Phosphodiester Bonds. *J. Am. Chem. Soc.* **1996**, *118* (14), 3340–3344. b) Rohatgi, R.; Bartel, D. P.; Szostak, J. W. Kinetic and Mechanistic Analysis of Nonenzymatic, Template-Directed Oligoribonucleotide Ligation. *J. Am. Chem. Soc.* **1996**, *118* (14), 3332–3339.

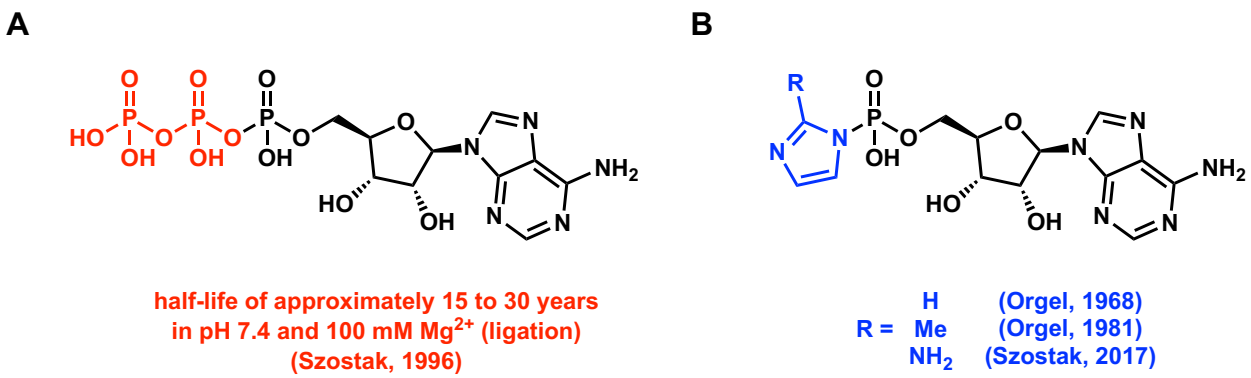


Figure 1.9: A) Nucleoside triphosphates are poor substrates for nonenzymatic copying, while B) 2-substituted imidazole activated monophosphates are substantially better substrates

One of the founding fathers of the field, Orgel proposed a great insight where, instead of using modern nucleoside triphosphates as substrates, activated nucleoside monophosphates with higher protocell permeability and higher reactivity would be more suitable for successful nonenzymatic copying and, thus, the emergence of functional RNA sequences.[25] This insight led to the discovery of imidazole activated nucleotides, and, subsequently, 2-methylimidazole activated nucleotides that show drastically superior reactivities in template directed, nonenzymatic copying (Figure 1.9 B).[37, 67] Recently, our lab has shown that 2-aminoimidazole activated nucleotides are superior to the corresponding 2-methylimidazole activated substrates, with up to 100 fold rate acceleration in nonenzymatic copying.[68] Additionally, the prebiotic synthesis of 2-aminoimidazole has been demonstrated,[69] which shares both intermediates and mechanistic

[67]a) Lohrmann, R.; Orgel, L. E.; Orgel, L. E.; Schneider-Bernloehr, H.; Sulston, J. E. Prebiotic Synthesis: Phosphorylation in Aqueous Solution. *Science* **1968**, *161* (3836), 64–66. b) Inoue, T.; Orgel, L. E. Substituent Control of the Poly(C)-Directed Oligomerization of Guanosine 5'-Phosphoroimidazolide. *J. Am. Chem. Soc.* **1981**, *103* (25), 7666–7667.

[68] Li, L.; Prywes, N.; Tam, C. P.; Oflaherty, D. K.; Lelyveld, V. S.; Izgu, E. C.; Pal, A.; Szostak, J. W. Enhanced Nonenzymatic RNA Copying with 2-Aminoimidazole Activated Nucleotides. *J. Am. Chem. Soc.* **2017**.

[69] 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. *J. Am. Chem. Soc.* **2017**.

pathways with the prebiotic synthesis of 2-aminooxazole, a building block in the synthesis of ribopyrimidine nucleotides.[32]

Along with the discovery of new activating groups, our understanding of the chemical mechanism of nonenzymatic primer extension with imidazole activated nucleotides has significantly improved. The mechanism of nonenzymatic primer extension with imidazole activated nucleotides was assumed to occur via nucleophilic attack of the 3′-hydroxyl group of the primer on the 5′-phosphate of the activated nucleotide, with the protonated imidazolium as a leaving group. However, there are several experimental observations that were not able to be explained by the proposed mechanism. For instance, much slower rates of extension were observed with the copying of the last nucleotide, referred to as "last addition problem".[70] Although the binding of the copying nucleotides is generally weaker due to the lack of structural rigidity and pi-stacking, the striking difference in the rates still couldn't be fully explained by the difference in binding affinity.[71] Another example that casts doubts on the proposed mechanism is the ligation of imidazole activated, short oligonucleotides, called nonenzymatic ligation. With the much stronger binding of oligonucleotides than a single nucleotide, the rate of nonenzymatic ligation should be much faster than that of activated monomer polymerization if the reaction undergoes a likely rate limiting step involving displacement of the protonated imidazole leaving group. However, the rate of nonenzymatic ligation has been shown to be significantly slower than the rate of polymerization of imidazole activated G:C pairs.[72] Different plausible models for explaining

---

[70] a) Taifeng, W.; 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. b) Kurz, M.; Göbel, K.; Hartel, C.; Göbel, M. W. Acridine-Labeled Primers as Tools for the Study of Nonenzymatic RNA Oligomerization. *Helv. Chim. Acta* **1998**, *81* (5–8), 1156–1180.

[71] Taifeng, W.; Orgel, L. E. Nonenzymic Template-Directed Synthesis on Oligodeoxycytidylate Sequences in Hairpin Oligonucleotides. *J. Am. Chem. Soc.* **1992**, *114* (1), 317–322.

[72] Szostak, J. W. The Eightfold Path to Non-Enzymatic RNA Replication. *Journal of Systems Chemistry*. **2012**.

these results have been proposed, such as noncovalent bonding and stacking interactions, but none fully explain all of the observation, and furthermore, detailed mechanistic studies for understanding these outcomes (last addition problem and ligation problem) had not been performed to bolster these claims.[72]

25 years after Orgel's initial insight, the true mechanism of nonenzymatic copying with imidazole activated nucleotides was revealed via rigorous kinetic analysis and structural studies. Kinetic analysis revealed a rate dependence on both the activated nucleotide solution pH and the optimal incubation period of monomers, which was found to be 15-30 minutes.[72] These observations suggest that non-covalent interactions may not be responsible for the acceleration observed with the polymerization of activated monomers as such observations disobey the nature of non-covalent interactions, but rather this points to a chemical process akin to an induction period. Furthermore, the lack of catalytic activity from pyrrole and pyrazole-phosphate analogues that mimic the activated monomer questioned the validity of non-covalent interactions as the primary source of rate acceleration.[73] Nuclear magnetic resonance spectroscopic (^{1}H and ^{31}P) characterization of 2-aminoimidaozle activated monomers has shown the build-up of a new species that has symmetrical dinucleotides,[74] which was later determined to be 5′,5′-imidazolium-bridged dinucleotides (Figure 1.10 A). This new species, is likely the reactive intermediate, as its presence and intermediacy explains many of anomalous and ambiguous experimental outcomes mentioned

[73] a) Zhang, W.; Tam, C. P.; Wang, J.; Szostak, J. W. Unusual Base-Pairing Interactions in Monomer-Template Complexes. *ACS Cent. Sci.* **2016**. b) 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**.

[74] a) Walton, T.; Szostak, J. W. A Highly Reactive Imidazolium-Bridged Dinucleotide Intermediate in Nonenzymatic RNA Primer Extension. *J. Am. Chem. Soc.* **2016**. b) 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.* **2017**.

above. Furthermore, the presence of a dimeric substrate allows for better recognition and stronger binding to the template, allowing for lower concentrations of monomers in solution. The mechanism for nonenzymatic templated directed primer extension with imidazolium bridged dimeric nucleosides is depicted in Figure 1.10 B.

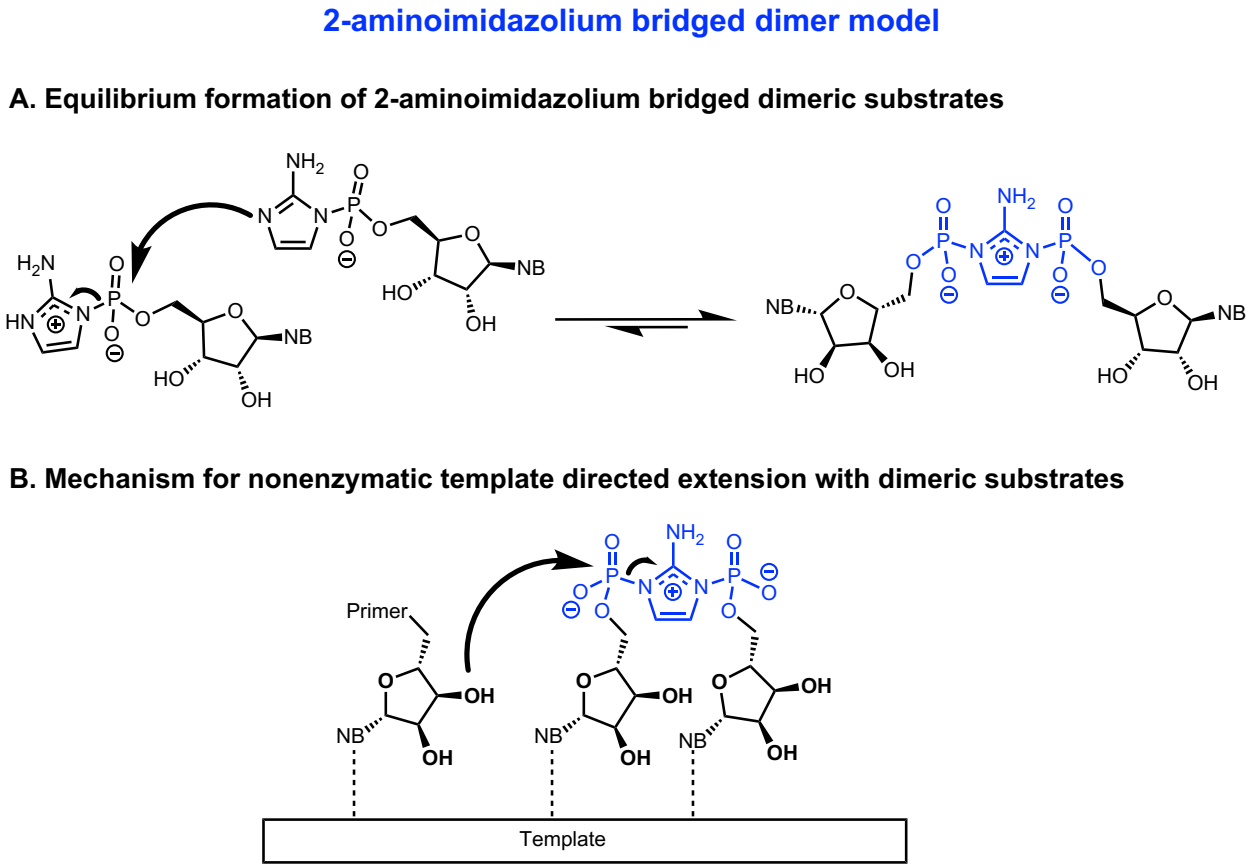


Figure 1.10: 2-aminoimidazolium bridged dimer model in template directed nonenzymatic primer extension. A) Mechanism of formation for the imidazolium bridged dimers. B) Mechanism for primer extension with imidazolium bridged dimers.

Even with decades of advances and mechanistic understanding in nonenzymatic replication/copying, there still remain significant challenges in efficiently copying template sequences of A, U, G, and C, where previous reports above rely on much stronger G:C base pairs in copying homopolymeric G or C sequences. A recent breakthrough in our laboratory has shown that 5′-activated downstream oligonucleotides significantly accelerate primer extension with activated monomers, such that an RNA template with all four canonical nucleobases can be

copied.[75] This catalytic effect is due to the reaction of the incoming activated monomer with the activated downstream oligonucleotide, to form a highly preorganized imidazolium-bridged intermediate that is poised to react with the primer.[74a] Structural stabilization of the correct helical geometry by the downstream helper oligonucleotide also contributes to the overall rate enhancement.[76]

One major remaining challenge in nonenzymatic copying is the requirement of high concentrations of divalent metal ions (normally 50-100 mM).[72,77] Divalent metal ions (such as Mg^{2+} and Mn^{2+}) are envisaged to play a dual role in both facilitating deprotonation of the primer 3′-hydroxyl via acidification upon coordination, and organizing nucleophiles (3′-hydroxyl) and electrophiles (bridged dinucleotide phosphate group) in the transition state assembly. Unfortunately, high concentrations of divalent metal ions can both catalyze the degradation of activated nucleotides, and RNA, and disrupt fatty acid vesicles.[72,78] Therefore, being able to lower the concentration of divalent metals while maintaining reactivity would be crucial for compatible, robust nonenzymatic primer extension. A deeper mechanistic understanding of the precise role of divalent metals would be necessitated to accomplish the task, by finding a more suitable prebiotically viable replacement or an alternate activation mode.

---

[75] 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**.

[76] 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. *J. Am. Chem. Soc.* **2018**.

[77] Walton, T.; Zhang, W.; Li, L.; Tam, C. P.; Szostak, J. W. The Mechanism of Nonenzymatic Template Copying with Imidazole-Activated Nucleotides. *Angew. Chemie* **2019**, *131* (32).

[78] Adamala, K.; Szostak, J. W. Nonenzymatic Template-Directed RNA Synthesis inside Model Protocells. *Science* **2013**.

1.4 Fidelity of Nonenzymatic copying

Since there were no evolved macromolecular catalysts such as enzymes or ribozymes during the origin of life, both the rate and fidelity of nonenzymatic replication are of special interest.[79][80] Because RNA is easily degraded, nonenzymatic replication cannot be arbitrarily slow.[79] Similarly, the error rate of nonenzymatic RNA replication must not exceed the Eigen error threshold (roughly the reciprocal of the number of critical bases in a sequence), above which genetic information cannot be maintained because mutations accumulate faster than they can be eliminated by selection.[81] Small ribozymes with catalytic rates on the order of 1 min^{-1} are typically at least 30–50 nucleotides in length,[82] suggesting that an error rate of less than 2–3% would be necessary to preserve functional information during repeated cycles of replication.[83]

Modern cellular life employs highly evolved polymerases and error correcting enzymatic pathways to ensure high fidelity;[84][85] primitive protocells, on the other hand, could only have relied on the intrinsic chemical properties of nonenzymatic RNA polymerization to govern the fidelity

---

[79] Szostak, J. W. The Narrow Road to the Deep Past: In Search of the Chemistry of the Origin of Life. *Angewandte Chemie - International Edition.* **2017**.

[80] Robertson, M. P.; Joyce, G. F. The Origins of the RNA World. *Cold Spring Harb. Perspect. Biol.* **2012**.

[81] 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**.

[82] Ferré-D'Amaré, A. R.; Scott, W. G. Small Self-Cleaving Ribozymes. *Cold Spring Harbor perspectives in biology*. **2010**.

[83] Kun, Á.; Santos, M.; Szathmáry, E. Real Ribozymes Suggest a Relaxed Error Threshold. *Nat. Genet.* **2005**.

[84] Garmendia, C.; Bernad, A.; Esteban, J. A.; Blanco, L.; Salas, M. The Bacteriophage Φ29 DNA Polymerase, a Proofreading Enzyme. *J. Biol. Chem.* **1992**.

[85] Natrajan, G.; Lamers, M. H.; Enzlin, J. H.; Winterwerp, H. H. K.; Perrakis, A.; Sixma, T. K. Structures of Escherichia Coli DNA Mismatch Repair Enzyme MutS in Complex with Different Mismatches: A Common Recognition Mode for Diverse Substrates. *Nucleic Acids Res.* **2003**.

of heredity prior to the advent of enzymatic mechanisms to survey and correct errors. Nonenzymatic template-directed RNA synthesis has been reported to show error rates of up to 17%, with most errors arising from G:U wobble pairing.[86] However, the significant decrease in the rate of continued RNA primer extension following a mismatched base pair by as much as 2 orders of magnitude, referred to as the stalling effect, increases the effective fidelity of the most rapidly synthesized full-length products.[81]

Elucidating the parameters that govern fidelity, rate and stalling factors is therefore critical in evaluating the prebiotic conditions that would favor the emergence of Darwinian evolution. With these criteria, namely rate and fidelity, prebiotically plausible nucleotides (Chapter 2: 8-oxo-purine and Inosine ribonucleotides and Chapter 3: arabino- and deoxyribonucleotides) are assessed while paired across canonical nucleotides (A, U, G, and C) in this thesis.

Lastly, the strand displacement of a full length copied oligonucleotides is a quintessential component of nonenzymatic replication of oligonucleotides. Ongoing projects in our lab have made much progress in addressing this issue.

1.5 Plausible models for primordial RNA

Within the RNA World hypothesis there exists a number of possible monomeric units that could compose the first biopolymer. At one extreme is the conceptually simple idea that life began with a much simpler genetic polymer, such as might be formed from the self-assembling melamine : barbituric acid nucleotides studied by Hud (Figure 1.11).[87] Efficient glycosidation of

---

[86] Leu, K.; Obermayer, B.; Rajamani, S.; Gerland, U.; Chen, I. A. The Prebiotic Evolutionary Advantage of Transferring Genetic Information from RNA to DNA. *Nucleic Acids Res.* **2011**.

[87] Cafferty, B. J.; Fialho, D. M.; Khanam, J.; Krishnamurthy, R.; Hud, N. V. Spontaneous Formation and Base Pairing of Plausible Prebiotic Nucleotides in Water. *Nat. Commun.* **2016**.

melamine and barbituric acid nucleobases by ribose-5-phosphate furnishes the corresponding products of pentose/hexose and α/β anomers with high overall yields of 55% and 82%, respectively. These compounds can form both Watson-Crick like base pairs as monomers (Figure 1.11 A) and can also form supramolecular assemblies of stacked hydrogen-bonded hexads (Figure 1.11 B) of up to tens of thousands of melamine and barbituric acid as polymers. Furthermore, crude products of glycosidation have shown to form the polymeric form with biologically relevant *β*-anomers preferred to *α*-anomers. However, informational inheritance has not yet been demonstrated in such simple systems, and how such systems could transition to more modern genetic polymers found in nature is unclear.

Melamine and Barbituric acid base pairing as the primordial genetic polymer

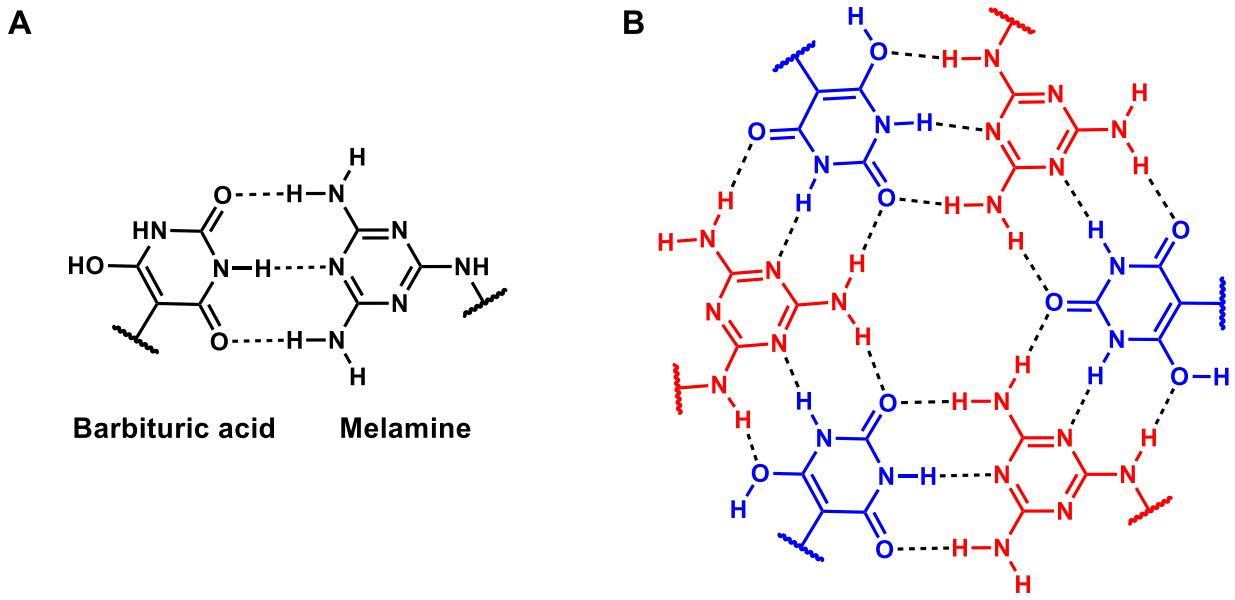


Figure 1.11: Hud's proposal that melamine and barbituric acid nucleotides may have been composed the primordial biopolymer. A) Base pairing between barbituric acid and melamine. B) Hexad assembly of melamine and barbituric acid.

At the other extreme lies the idea that life began with essentially modern RNA, synthesized directly by prebiotic chemical processes. This model, championed by Sutherland, is supported by

the demonstration of potentially prebiotic pathways to the canonical pyrimidine nucleotides, presented above.[32] As discussed in section 1.2, simple prebiotically available organic compounds like glycoaldehyde, cyanamide, and glyceraldehyde, can condense to synthesize pentose amino-oxazolines (Compound **N**, in Figure 1.8) in astonishingly high yields, which serve as the key intermediate to generate the anhydroarabinonucleoside (Compound **N2** in Figure 1.8). Phosphorylation of the 3′-hydroxyl group, followed by the intramolecular nucleophilic attack of the 3′-phosphate group on the 2′-position displaces the nucleobase to generate pyrimidine ribonucleotides.

There exist two potential pitfalls to the proposal that the primordial genetic polymer building blocks are identical to modern nucleotides. First, no satisfactory synthesis of purine ribonucleotides exists.[34] Attempts to synthesize purine ribonucleotides with the same pathway (in Figure 1.8) have been made, yet no yield of product is observed and a number of byproducts have been produced, which will be discussed in later chapters. Alternative pathways towards the syntheses of purine ribonucleotides have been proposed but these syntheses generally suffer from a lack of divergence, selectivity, or scope (guanine is particularly challenging). Second, syntheses of these canonical building blocks generate products that are similar, but not identical to, ribonucleotides that could be competent in nonenzymatic copying and subsequent selection.

In between these two extremes lie multiple models that suggest that the primordial polymer was closely related to but not identical with modern RNA. Many of these models invoke a degree of heterogeneity, either in the backbone, e.g. a mixture of 2′-5′ and 3′-5′ linkages,[88] in the sugar, e.g. a mixture of 2′-deoxyribose and ribose, or in the nucleobases,[89] which might have been

[88] Engelhart, A. E.; Powner, M. W.; Szostak, J. W. Functional RNAs Exhibit Tolerance for Non-Heritable 2′-5′ versus 3′-5′ Backbone Heterogeneity. *Nat. Chem.* **2013**.

partially or fully replaced with modified nucleobases. Notably, 2-thio-uracil has been shown to provide stronger and more accurate base-pairing when used in lieu of uracil.[90] Last, as the arsenal of prebiotically plausible nucleotides expands (while ribopurines synthesis remains elusive), and if these non-canonical nucleotides can be shown to be competent in non-enzymatic primer extension (following the requirements for success laid out in the next section), the odds that a purely RNA system would initially form from a soup of heterogeneous nucleotides is statistically unlikely and thus an initial mixed system is entropically favored.

1.6 Conclusion

The remainder of this thesis, Chapters 2 and 3, will focus on evaluating two classes of non-canonical nucleotides in nonenzymatic template directed primer extension to explore how modern RNA evolved out of a more heterogeneous primordial genetic polymer that initially had non-canonical nucleobases incorporated. As discussed in section 1.2, given the lack of prebiotic purine ribonucleotide synthetic pathways, coupled with the wealth of non-canonical nucleotides generated en route to the canonical ribonucleotides, one would envision that the initial random polymerization of nucleotides would generate a heterogeneous polymer incorporating non-canonical nucleotides. If some of these heterogeneities are either advantageous or innocuous in nonenzymatic nucleotide copying, this could provide an alternate primordial genetic polymer and/or one with a selection mechanism to modern RNA. Therefore, in Chapter 2, we will evaluate the 8-oxo-purines and inosine ribonucleotides in nonenzymatic primer extension, and in Chapter

[89] Trevino, S. G.; Zhang, N.; Elenko, M. P.; Lupták, A.; Szostak, J. W. Evolution of Functional Nucleic Acids in the Presence of Nonheritable Backbone Heterogeneity. *Proc. Natl. Acad. Sci. U. S. A.* **2011**.

[90] Heuberger, B. D.; Pal, A.; Del Frate, F.; Topkar, V. V.; Szostak, J. W. Replacing Uridine with 2-Thiouridine Enhances the Rate and Fidelity of Nonenzymatic RNA Primer Extension. *J. Am. Chem. Soc.* **2015**.

3, we will evaluate arabinonucleotides in nonenzymatic primer extension, subject to the conditions laid out in section 1.3, and using fidelity and rate as criteria for success, as laid out in section 1.4.

Chapter 2:

Inosine, but none of the 8-oxo-purines, is a plausible component of a primordial version of RNA

2.1 Introduction

As discussed above, an elegant prebiotic synthesis of the pyrimidine nucleotides was accomplished by Sutherland and Powner in 2009.[32] Using the same strategy, a chemically divergent synthesis of the purine and pyrimidine ribonucleotides was attempted.[34] Such a divergent approach is ideal in accessing both purine and pyrimidine ribonucleotides simultaneously, as entirely distinct prebiotic routes are often incompatible, and diversifying a late stage intermediate increases the odds that all nucleotides could be present in high concentrations in the same environment. However, this also suggests that any byproduct generated in the synthesis would resemble the desired product and indicate the likelihood of coexistence of byproducts,[91] some of which could be detrimental to genetic information storage or transfer.

[91] a) Becker, S.; Schneider, C.; Crisp, A.; Carell, T. Non-Canonical Nucleosides and Chemistry of the Emergence of Life. *Nat. Commun.* **2018**, *9* (1), 5174. b) Whitaker, D.; Powner, M. W. Prebiotic Nucleic Acids Need Space to Grow. *Nat. Commun.* **2018**, 9–12.

2.1.1 Prebiotic Synthesis of 8-oxo-purine and Inosine

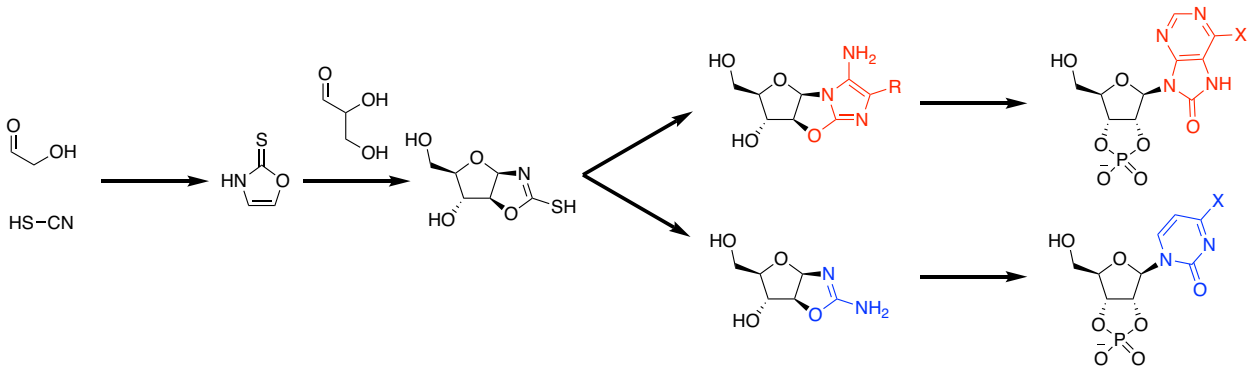


Figure 2.1: Divergent prebiotic syntheses of 8-oxo-purine ribonucleotides (in red, X= OH or NH_2) and pyrimidine ribonucleotides (in blue, X= OH or NH_2) by the Powner group.

In attempts to synthesis purine ribonucleotides, the Powner group utilizes 2-thiooxazole, structurally similar to 2-aminooxazole, along with glyceraldehye to generate pentose oxazolidinone thione (Figure 2.1).[34] This compound serves as the key intermediate in the divergent syntheses of purine and pyrimidine ribonucleotides via subsequent reaction with cyanoacetylene then ammonia to generate pentose aminooxazolidinone or aminonitrile to generate pentose aminoimidazole. Corresponding nucleobases can be built on these moieties to furnish purine (with the addition of formamidine and formamide) and pyrimidine (with the addition of cyanoacetylene) anhydroarabinonucleosides. Interestingly, the Powner group was unable to synthesize guanosine anhydroarabinonucleosides, but instead generated inosine anhydroarabinonucleosides in moderate yield via this route. Phosphorylation of the 3′-hydroxyl group and subsequent intramolecular substitution inverts the 2′ stereocenter and generates pyrimidines, 8-oxo-adenosine, and 8-oxo-inosine ribonucleotides. Chemoselective reduction of the 8-oxo-purines was attempted, but no significant progress has been made. Assuming that the reduction is possible, this report suggests the coexistence of adenosine, inosine, 8-oxo-adenosine, and 8-oxo-inosine alongside pyrimidine

ribonucleotides on the prebiotic earth. Powner proposes the idea that 8-oxo-purine ribonucleotides could be utilized as surrogates for purine ribonucleotides in non-enzymatic primer extension, which the remainder of this chapter will demonstrate is unlikely. Assuming that adenosine can be prepared, there exists an additional pathway for the synthesis of inosine via deamination. A previous report regarding the prebiotic activation of nucleotides by the Sutherland group requires nitrous acid (HNO_2),[92] which, in acidic conditions, has been demonstrated to deaminate adenosine to generate inosine.[93]

2.1.2 Properties of 8-oxo-purine and Inosine

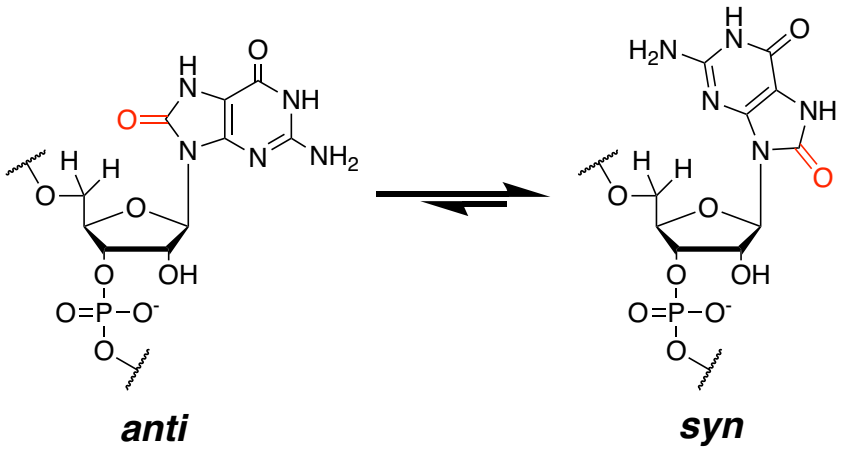


Figure 2.2: Influence of oxidation on the conformation of 8-oxo-purine ribonucleotides

Simple arene oxidation of purine nucleotides (both deoxyribo and ribo) from proton to oxygen at the 8 position has shown to drastically alter the properties of the nucleotides. The thermodynamic landscape of dihedral rotation about the glycosyl bond diverges for 8-oxo-purines and purines due to the steric clash between the 8-oxo and 5′-carbon of the pentose 8-oxo-purine

[92] Mariani, A.; Russell, D. A.; Javelle, T.; Sutherland, J. D. A Light-Releasable Potentially Prebiotic Nucleotide Activating Agent. *J. Am. Chem. Soc.* **2018**.

[93] Francom, P.; Janeba, Z.; Shibuya, S.; Robins, M. J. Nucleic Acid Related Compounds. 116. Nonaqueous Diazotization of Aminopurine Nucleosides. Mechanistic Considerations and Efficient Procedures with Tert-Butyl Nitrite or Sodium Nitrite. *J. Org. Chem.* **2002**.

nucleotides.[94] This results in the nucleobase adopting an *syn* conformation rather than its usual *anti* conformation (in the case of purine nucleotides).[95] Corollary to this conformational change is the utterly different binding properties, where Hoogsteen base pairing becomes much more prominent due to the nucleobase ground state being flipped, poising it for errant Hoogsteen binding.[96] 8-oxo-A pairs much more strongly to G and 8-oxo-G pairs with A when compared with their corresponding non-oxygenated purine nucleotides;[97] this increase in errant binding affinity occurs via Hoogsteen base pairing. In modern biological systems, 8-oxo-purine deoxyribonucleotides are found due to oxidative damage of guanosine, which is much more prevalent than oxidation of adenosine (3:1 ratio of 8-oxo-guanosine to 8-oxo-adenosine in cellular DNA),[94] with guanosine having a low oxidation potential of 1.02 V (versus NHE).[98] A number of diseases have been attributed to 8-oxo-purine deoxyribonucleotides, such as Alzeheimer disease, and muscular disorders.[99]

---

[94] Herbert, C.; Dzowo, Y. K.; Urban, A.; Kiggins, C. N.; Resendiz, M. J. E. Reactivity and Specificity of RNase T1, RNase A, and RNase H toward Oligonucleotides of RNA Containing 8-Oxo-7,8-Dihydroguanosine. *Biochemistry* **2018**.

[95] a) Uesugi, S.; Ikehara, M. Carbon-13 Magnetic Resonance Spectra of 8-Substituted Purine Nucleosides. Characteristic Shifts for the Syn Conformation. *J. Am. Chem. Soc.* **1977**. b) Leonard, G. A.; Brown, T.; Guy, A.; Téoule, R.; Hunter, W. N. Conformation of Guanine-8-Oxoadenine Base Pairs in the Crystal Structure of d(CGCGAATT(08A)GCG). *Biochemistry* **1992**, *31* (36), 8415–8420. c) Thiviyanathan, V.; Somasunderam, A.; Hazra, T. K.; Mitra, S.; Gorenstein, D. G. Solution Structure of a DNA Duplex Containing 8-Hydroxy-2′-Deoxyguanosine Opposite Deoxyguanosine. *J. Mol. Biol.* **2003**, *325* (3), 433–442. d) McAuley-Hecht, K. E.; Gibson, N. J.; Thomson, J. B.; Brown, T.; Leonard, G. A.; Hunter, W. N.; Watson, W. P. Crystal Structure of a DNA Duplex Containing 8-Hydroxydeoxyguanine-Adenine Base Pairs. *Biochemistry* **1994**, *33* (34), 10266–10270.

[96] Cheng, X.; Kelso, C.; Hornak, V.; De Los Santos, C.; Grollman, A. P.; Simmerling, C. Dynamic Behavior of DNA Base Pairs Containing 8-Oxoguanine. *J. Am. Chem. Soc.* **2005**.

[97] Choi, Y. J.; Chang, S. J.; Gibala, K. S.; Resendiz, M. J. E. 8-Oxo-7,8-Dihydroadenine and 8-Oxo-7,8-Dihydroadenosine—Chemistry, Structure, and Function in RNA and Their Presence in Natural Products and Potential Drug Derivatives. *Chemistry - A European Journal*. **2017**.

[98] Steenken, S.; Jovanovic, S. V. How Easily Oxidizable Is DNA? One-Electron Reduction Potentials of Adenosine and Guanosine Radicals in Aqueous Solution. *J. Am. Chem. Soc.* **1997**.

Inosine, bearing a carbonyl group at the 6 position of purine, could wobble pair with a number of canonical nucleotides (A, C, and U). The promiscuous nature of inosine pairing stems from the substitution of an N-H with the 6-carbonyl, flipping the donor acceptor properties as compared to A and deleting the NH_2 contact as compared to G.[100] In modern biological systems, inosine is found in tRNA and serves a critical role in translation, due to its ability to wobble pair and thus recognize multiple codons.

2.2 Results

Given the significance of prebiotic synthetic pathways as well as the unusual pairing properties of 8-oxo-purines and inosine, we sought to evaluate how these noncanonical nucleotides behave when used as a single nucleobase modification in lieu of canonical nucleotides in nonenzymatic, template-directed primer extension reactions. To summarize our findings discussed extensively below, we find that activated 8-oxo-purine monomers show extremely low rates of primer extension, and once incorporated they strongly decrease the rate of incorporation of the next nucleotide. Furthermore, 8-oxo-purine nucleotides on the template strand lower the fidelity by promoting primer extension with both Watson-Crick and Hoogsteen pairing partners. However, activated inosine monomers can be incorporated at a rate comparable to activated guanosine opposite C in the template, while minimally impacting subsequent primer extension. In addition, inosine in the template strand promotes rapid primer extension with cytidine, suggesting that

[99]a) Poulsen, H. E.; Specht, E.; Broedbaek, K.; Henriksen, T.; Ellervik, C.; Mandrup-Poulsen, T.; Tonnesen, M.; Nielsen, P. E.; Andersen, H. U.; Weimann, A. RNA Modifications by Oxidation: A Novel Disease Mechanism? *Free Radical Biology and Medicine*. **2012**. b) Tanaka, M.; Song, H.; Küpfer, P. A.; Leumann, C. J.; Sonntag, W. E. An Assay for RNA Oxidation Induced Abasic Sites Using the Aldehyde Reactive Probe. *Free Radical Research*. **2011**. c) Moreira, P. I.; Nunomura, A.; Nakamura, M.; Takeda, A.; Shenk, J. C.; Aliev, G.; Smith, M. A.; Perry, G. Nucleic Acid Oxidation in Alzheimer Disease. *Free Radical Biology and Medicine*. **2008**. d) Honda, K.; Smithi, M. A.; Zhu, X.; Baus, D.; Merrick, W. C.; Tartakoff, A. M.; Hattier, T.; Harris, P. L.; Siedlak, S. L.; Fujioka, H.; et al. Ribosomal RNA in Alzheimer Disease Is Oxidized by Bound Redox-Active Iron. *J. Biol. Chem.* **2005**.

[100] Crick, F. H. C. Codon—Anticodon Pairing: The Wobble Hypothesis. *J. Mol. Biol.* **1966**.

inosine would be an effective component of a primitive genetic polymer.

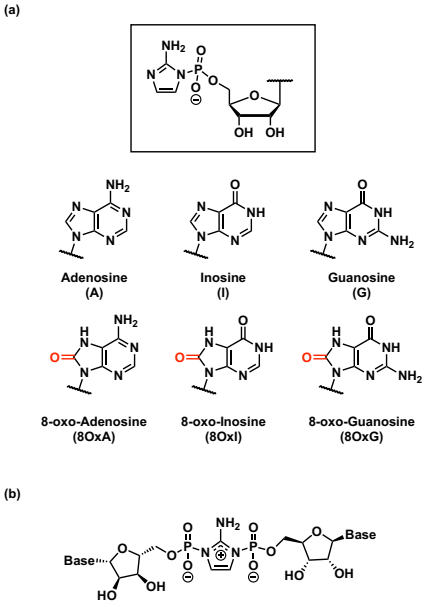
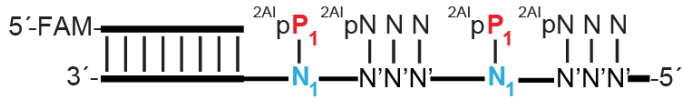


Figure 2.3: (a) Chemical structures of 2-amino-imidazole activated purine and 8-oxo-purine nucleotides. (b) Chemical structure of reactive 2-amino-imidazolium-bridged intermediate in template-directed primer extension.

(a)



(b)

Copying rate k (hr^{-1})

N_1	P_1 8OxA	8OxG	8OxI	W.C.
A	0.00075(9)	0.020(8)	0.00083(7)	6.2(2)
U	0.0036(4)	0.040(9)	0.00042(3)	3.1(1)
G	0.0010(1)	0.0052(4)	0.00089(6)	10.2(2)
C	0.0091(5)	0.0081(4)	0.0075(7)	11.7(5)

(c)

N_1	P_1 A	U	G	C	μ_N
8OxG	4.0(1)	0.045(2)	0.85(7)	4.1(1)	0.54
G	0.0025(1)	0.075(2)	0.040(9)	10.2(2)	0.011
8OxA	0.12(2)	2.1(1)	0.34(1)	0.019(4)	0.19
A	0.013(4)	6.2(2)	0.0021(1)	0.11(3)	0.020

Figure 2.4: Evaluation of 8-oxo-purine nucleotides in nonenzymatic primer extension. All reactions were performed at pH 8.0, 200 mM HEPES, 50 mM Mg^{2+}, 20 mM 2AIpP_1, 0.5 mM of activated trimer. (a) Schematic representation of a primer extension reaction. 2AIpP_1 represents 2-aminoimidazole activated monomers and 2AIpNNN represents 2-aminoimidazole activated trimer helper. (b) Rates of primer extension for 2-aminoimidazole activated 8-oxo-purines across canonical nucleotides. W.C. indicates the Watson-Crick pair monomer corresponding to each nucleobase on template (c) Rates of primer extension for 2-aminoimidazole activated canonical nucleotides across 8-oxo-purine nucleotides. Projected misincorporation rates (μ_N) for template bases N_1 were determined by dividing the sum of the incorporation rates across the incorrect pair by the sum of all the incorporation rates across represented P_1 bases.

2.2.1 Primer extension with 8-oxo-purine monomers

In order to investigate the activity of the 8-oxo-purine nucleotides in nonenzymatic primer extension, we evaluated the rate of primer extension using these noncanonical nucleotides as activated monomers across from all four canonical nucleotides on a series of templates (Figure 2.4 b). Our laboratory has previously shown that 5-activated downstream oligonucleotides significantly accelerate primer extension with activated monomers, such that an RNA template

with all four canonical nucleobases can be copied.[75] This catalytic effect is due to the reaction of the incoming activated monomer with the activated downstream oligonucleotide, to form a highly preorganized imidazolium-bridged intermediate that is poised to react with the primer.[74a] Structural stabilization of the correct helical geometry by the downstream helper oligonucleotide also contributes to the overall rate enhancement.[101] We therefore prepared 2-aminoimidazole (2AI) activated 5′-monophosphates of 8-oxo-A (8OxA), 8-oxo-G (8OxG), and 8-oxo-I (8OxI) to give the corresponding activated monomers 2AIp-8OxA, 2AIp-8OxG, and 2AIp-8OxI respectively (Figure 2.3).[68] We also generated a 2AI-activated trimer helper, and we used this downstream trimer in all experiments where we measured the kinetics of monomer incorporation in helper assisted nonenzymatic primer extension reactions (Figure 2.4 a).

The observed rate of primer extension for each activated monomer was compared with that of the canonical activated monomers (A, U, G, or C, Figure 2.4 b and c) across from their Watson-Crick pairing partners on the template (Figure 2.4). The observed rates (k_{obs}) varied over three orders-of-magnitude, indicating that, as expected, the identity of the nucleobase plays a critical role in the rate of template-directed primer extension. Activated canonical monomers that form Watson-Crick base pairs with the template exhibited rates of primer extension ranging from 3.1 to 11.7 hr^{-1}. The activated 8-oxo-purine monomers exhibited drastically slower rates of primer extension across from all the canonical nucleotides. While addition rates of the 8-oxo-purine monomers opposite Watson-Crick pairing partners were higher than their addition rates when opposite Hoogsteen pairing partners, both were at least 100-fold slower than that of the standard purine monomers with Watson-Crick pairing to the template. Overall, 2AIp-8-oxo-purines were

[101] 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**.

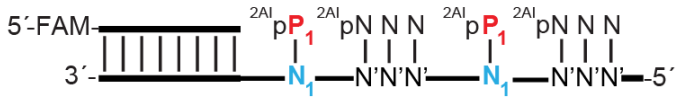
incorporated at rates similar to those of the canonical nucleotides when mismatched with the template.

2.2.2 Primer Extension Across 8-oxo-purines in the Template Strand

Although templates containing 8-oxo-purines would only rarely be generated by the copying chemistry described above, they could potentially be generated by the non-templated polymerization of activated monomers, e.g. in ice eutectic phases. Additionally, the 8-position of purines is susceptible to oxidative damage, suggesting that 8-oxopurine containing templates could be generated by radical oxidation of the parent nucleobases,[102] e.g. in environments subject to high levels of radiation. We therefore examined the effects of 8-oxo-purine nucleotides at internal positions in the template strand (Figure 2.4 c). To measure the rates of primer extension with activated canonical nucleotides (2AIpN) across from template sites consisting of 8-oxo-purine nucleotides (Table 1c), we prepared RNA templates containing 8-oxo-A and 8-oxo-G by solid phase synthesis. Intriguingly, 8-oxo-G containing templates show high rates of primer extension with both activated C and A monomers (~ 4 hr^{-1}); whereas the 2AIp-8-oxo-G incorporation rate was almost negligible across from either C or A in the template (less than 0.05 hr^{-1}). Compared with the corresponding canonical purines, 8-oxo-purine templates have lower rates of primer extension with Watson-Crick pairing monomers while they have much higher rates for Hoogsteen pairing monomers (8OxG:A and 8OxA:G) with greater than 1000-fold rate acceleration. Thus, the estimated error rates for the copying of 8-oxo-purines in a template were very high with $\boldsymbol{\mu_{8OxG}}$ = 0.54 and $\boldsymbol{\mu_{8OxA}}$ = 0.19 (Figure 2.4 c).

[102] Frelon, S.; Douki, T.; Favier, A.; Cadet, J. Comparative Study of Base Damage Induced by Gamma Radiation and Fenton Reaction in Isolated DNA. *J. Chem. Soc. Perkin 1* **2002**.

(a)



(b)

		P_1 A	U	C	I	μ_N
N_1	A	0.013(4)	6.2(2)	0.11(3)	0.024(1)	0.023
	U	3.1(1)	0.17(1)	0.034(3)	0.17(1)	0.11
	C	0.17(2)	0.065(3)	0.16(1)	6.4(7)	0.058
	I	0.05(1)	0.015(1)	8.0(5)	0.0070(1)	0.0089
					μ_{avg}	0.049

(c)

		P_1 A	U	C	G	μ_N
N_1	A	0.013(4)	6.2(2)	0.11(3)	0.0021(1)	0.020
	U	3.1(1)	0.17(1)	0.034(3)	0.57(4)	0.20
	C	0.17(2)	0.065(3)	0.16(1)	11.7(5)	0.033
	G	0.0025(1)	0.075(2)	10.2(2)	0.040(9)	0.011
					μ_{avg}	0.066

Figure 2.5: Evaluation of inosine nucleotide in nonenzymatic primer extension. (a) Schematic representation of a primer extension reaction. 2AIpP_1 represents 2-aminoimidazole activated monomers and 2AIpNNN represents 2-aminoimidazole activated trimer helper. (b) Rates of primer extension for 2AIpP_1 across A,U,C, and I. Projected misincorporation rates (μ_N) for template bases N_1 were determined by dividing the sum of the incorporation rates across the incorrect pair by the sum of all the incorporation rates across represented P_1 bases. (c) Rates of primer extension for 2AIpP_1 across A,U,C, and G. Projected misincorporation rates (μ_N) were determined analogously as in (b), with G replacing I. All reactions were performed at pH 8.0, 200 mM HEPES, 50 mM Mg^{2+}, 20 mM 2AIpP_1, 0.5 mM of activated trimer.

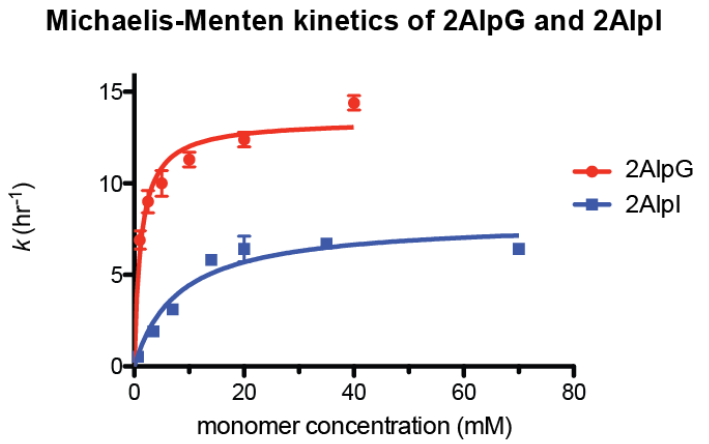
2.2.3 Primer Extension with Inosine as the Monomer and in the Template

As a control for comparison with 8-oxo-inosine, and to further examine the competency of non-canonical nucleobases in nonenzymatic primer extension, we evaluated the rates of primer extension using inosine in nonenzymatic primer extension reactions (Figure 2.5 a). We used the same reaction conditions as used previously in evaluating activated 8-oxo-purine monomers to evaluate the addition of 2AIpI across from A, U, G, C and I on the template (Figure 2.5 b). Unlike

activated 8-oxo-purine monomers, the inosine monomer showed a high rate of primer extension (6.4 hr^{-1}) when paired with cytosine in the template. Of note, this rate was faster than the rate of A addition when paired with U in the template (3.1 hr^{-1}), and was comparable to the observed addition rates for the other canonical Watson-Crick base pairs. Interestingly, the rate of addition of I across from U in the template was slower (0.17 hr^{-1}) than the rate of incorporation of G opposite U (0.58 hr^{-1}) (Figure 2.5 c). Primer extension with activated I across from A and I in the template was very slow, exhibiting similar behavior to activated G (Figure 2.5 c). These incorporation rates, taken together, demonstrate that I behaves similarly to G as an activated monomer, but might exhibit modestly higher fidelity due to the reduced rate of incorporation of I vs. G when paired with U in the template.

Given the comparable rates of primer extension with inosine and guanosine as monomers, we further evaluated inosine at internal positions in the template. Most importantly, the rate of primer extension with 2AIpC across from I in the template was similar to the rate of 2AIpI addition across from C in the template, and these rates were comparable to the rate of primer extension with 2AIpC across from G in the template (Figure 2.5 c). Based on the observed rates of primer extension with A,U,C, and I as activated monomers, across from inosine in the template, the projected error rate ($\boldsymbol{\mu_I}$) for copying of inosine within the template was 0.0089. This misincorporation rate was slightly lower than that calculated for both G ($\boldsymbol{\mu_G}$=0.011) and A ($\boldsymbol{\mu_A}$=0.020) within the template. Overall, a system with A, U, C, and I in the template and activated A, U, C, and I as monomers has an average projected error rate of 0.049, slightly lower than that of the corresponding system with A, U, C, and G, which has an estimated error rate of 0.066. The rates of primer extension with I and G, and fidelity data, all indicate that I and G behave similarly, with I acting as a slightly superior nucleotide in nonenzymatic primer extension.

(a)



(b)

	V_{max}	K_M
2AIpG	13.2 ± 0.8	1.1 ± 0.3
2AIpI	7.9 ± 0.8	8.0 ± 2.7

Fjgure 2.6: Michaelis-Menten study of the nonenzymatic RNA copying rates of 2AIpG and 2AIpI addition across C. (a) Plot of k (hr^{-1}) against the concentration of 2AIpI and 2AIpG. All reactions were performed at pH 8.0, 200 mM HEPES, 50 mM Mg^{2+}, 0.5 mM of activated trimer. (b) Michaelis-Menten paramters (V_{max} and K_M) for 2AIpG and 2AIpI.

Given the high rate of I addition across from C in the template, we evaluated the Michaelis-Menten parameters for primer extension with both G and I on a 3′-CCC-5′ template (Figure 2.6). We measured primer extension rates for a range of monomer concentrations to allow estimation of V_{max} and K_M values. The K_M for 2AIpI (~ 8.0 mM) was approximately 7-fold higher than that of 2AIpG (~ 1.1 mM), whereas the V_{max} for 2AIpI and 2AIpG differed by less than 2-fold (7.9 and 13.2 mM hr^{-1}, respectively). The catalytic efficiency (V_{max}/K_M) was an order of magnitude higher for guanosine relative to inosine across from cytidine, primarily because of the tighter binding of the activated guanosine mononucleotide.

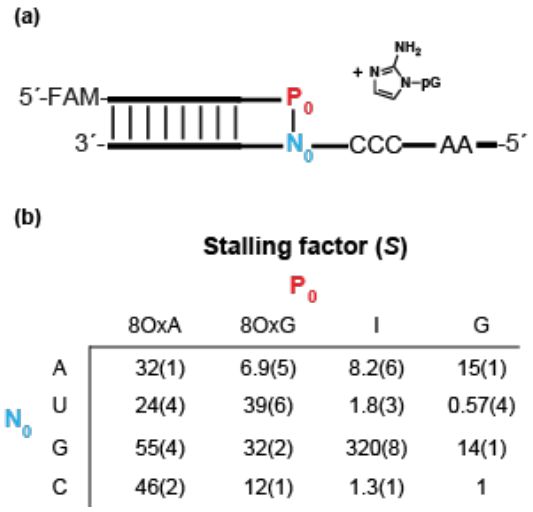


Figure 2.7: (a) Schematic representation of nonenzymatic primer extension with a primer that contains a noncanonical nucleotide (P_0) 3′ end. (b) Stalling factors ($S_{P:N}$) of nonenzymatic RNA polymerization of 2AImpG.

2.2.4 Effect of 8-oxo-purines and Inosine at the 3′-end of the Primer

Previous reports indicate that the rate of primer extension is strongly dependent on the nature of the last primer-template base pair. A factor of 30 to 200-fold rate decrease has been observed for non-Watson-Crick, non-wobble base pairs at the 3′-end of 3′-amino,2′,3′-dideoxynucleotide-terminated primers.[81] Because of this potentially large effect, we explored the influence of noncanonical bases at the 3′-end of the primer. We therefore measured the rate of nonenzymatic RNA primer extension using RNA primers containing either 8-oxo-purines or inosine at the 3′-terminus, opposite all four canonical nucleotides in the template. In each case we measured the rate of primer extension using 2AIpG as the monomer on a C_3 template. Because G is incorporated efficiently on an oligo-C template, no downstream activated helper oligonucleotide was used (Figure 2.7 a). The ratio of the rate of primer extension relative to that with a correctly matched pair was calculated as the stalling factor (*S*). For example, the stalling factor for primer extension beyond a 3′-terminal inosine in a terminal C:I pair is calculated as $S_{C:I} = k_{C:G} / k_{C:I}$, where $k_{C:G}$ is the extension rate beyond a C:G pair, and $k_{C:I}$ is the extension rate beyond the C:I pair (see

methods for additional details). A higher value corresponds to an increase in stalling due to the 3′-terminal primer-template pair. A primer terminated with inosine at the 3′ position gave a minimal stalling factor of 1.3 when the inosine could form a Watson-Crick C:I pair with C in the template.[81] Stalling factors for I and G opposite the wobble pairing partner U were similar at 1.8 and 0.55. Stalling factors for inosine were much higher for all other non-Watson-Crick pairs (Figure 2.7 b).

In contrast to the favorable properties of inosine, primer extension rates for primers with 3′-terminal 8-oxo-purines were much slower, with stalling factors ranging from 6.9 to 55. When 8-oxo-purines are paired across the corresponding Watson-Crick nucleotide in the template, the stalling factors were 12 and 24 for 8-oxo-G and 8-oxo-A respectively. When the template nucleotide could potentially form a Hoogsteen base pair with the nucleotide at the end of the primer (A:8OxG and G:8OxA, for instance), high stalling factors were observed. These results suggest that 8-oxo-purine monomers would not be functional components of nonenzymatic RNA copying system; not only are the 8-oxo-purine monomers very poor substrates for primer extension, but if they are incorporated, they strongly inhibit the next step of primer extension.

2.3 Discussion

We have found that two different classes of noncanonical nucleotides, 8-oxo-purines and inosine, behave very differently from each other in nonenzymatic primer extension. Three separate lines of evidence suggest that the 8-oxo-purines are unlikely to have been a functional component of a primitive RNA based genetic system: the incorporation of the 8-oxo-purine monomers in helper assisted primer extension is kinetically slow and, in the rare cases where an 8-oxo-purine residue is incorporated at the 3′ end of a primer, subsequent primer extension is also very slow. In addition, 8-oxo-adenosine in a template leads to slow primer extension with U, while 8-oxo-guanosine in a template results in highly error prone primer extension. Thus, oxidative damage of

the standard purines in a template strand would be highly deleterious and conditions that promote extensive purine oxidation would likely be incompatible with the emergence of functional RNAs.

In contrast to the 8-oxo-purines, inosine behaves as a reasonable monomer in nonenzymatic RNA primer extension. As an activated monomer, it is incorporated with a rate comparable to that of guanosine, and the I:C base pair at the end of a primer exhibits minimal stalling effect. In the template, I was copied with a rate and fidelity similar to that of G. The fidelity with which a template that contained I (AUIC) was copied with A, U, C, and I monomers (error frequency of 0.049) was slightly better than that with which an AUCG template would be copied with A, U, C, and G as monomers (error frequency of 0.066). Interestingly, the fidelity of the mixed system in which A, U, C, and I monomers are copying an AUGC template was similar, with an error rate of ~0.050. The mixed system where A, U, C, and G monomers are copying an AUIC template had a slightly higher misincorporation rate of ~0.075 (Supplementary Figures S25)**.** The fidelity of RNA copying with inosine in place of guanosine requires further study, as both sequence context and reaction conditions such as nucleotide concentrations are likely to affect the error rates. For example, fidelity might be significantly improved simply due to competition for binding to the template when all four monomers are present together. In addition, the stalling effect will certainly increase the fidelity of the most rapidly completed copies of a template. Assuming that purine-purine mismatches essentially permanently stall primer extension, while ignoring the minimal stalling observed for wobble-pairs, leads to an estimated overall error rate for the A, U, C, and I system of 0.024, while the effect is smaller for the A, U, C, G system (estimated error rate of 0.050) where errors are dominated by G:U wobble pairing (Supplementary Figure S2.26).

Orgel and coworkers previously studied nonenzymatic primer extension with inosine, and reported poor copying of dC-rich templates with 2-methylimidazole-activated inosine.[103] In addition, cytidine activated with 2-methylimidazole showed negligible primer extension across from inosine on a DNA template.[104] Our results differ significantly from these earlier observations. However, there are several important differences in the conditions in which template-directed synthesis experiments were carried out. Notably, as first described by Orgel and coworkers,[105] template directed polymerization of activated RNA mononucleotides is more efficient on RNA templates (e.g., our current work using an RNA primer–template duplex) relative to DNA templates (e.g., previous work on inosine by Orgel's group). Moreover, our primer extension reactions made use of activated trimer helpers to further catalyze the reaction. Last, we used a superior leaving group, 2-aminoimidazole, for our primer extension reactions, whereas previous work was carried out using the less effective 2-methylimidazole leaving group. Our optimized conditions for RNA copying may have more prebiotic relevance, given that our group has previously described a prebiotically plausible synthesis of 2-aminoimidazole.[69] Furthermore, previous work from our lab has shown that s2U or s2rT in place of U can lead to substantial improvements in fidelity. Future work should address the fidelity of template copying in systems that contain these modified nucleotides. Nevertheless, our current observations suggest that the primordial genetic polymer could have been a version of RNA in which inosine took the place of guanosine, either partially or completely.

---

[103] Rembold, H.; Robins, R. K.; Seela, F.; Orgel, L. E. Polycytidylate and Poly(7-Deazaguanylate): A Pair of Complementary Templates. *J. Mol. Evol.* **1994**.

[104] Kozlov, I. A.; Orgel, L. E. Nonenzymatic Oligomerization Reactions on Templates Containing Inosinic Acid or Diaminopurine Nucleotide Residues. *Helv. Chim. Acta* **1999**.

[105] Zielinski, M.; Kozlov, I. A.; Orgel, L. E. A Comparison of RNA with DNA in Template-Directed Synthesis. *Helv. Chim. Acta* **2000**, *83* (8), 1678–1684.

One of the key challenges in nonenzymatic RNA replication is strand separation, which is necessary to propagate RNA replication. RNA duplexes of 30-50 nucleotides are extremely thermostable, making RNA strand separation a significant challenge.[72] Thus, factors that lower the melting temperature of RNA duplexes without compromising RNA copying are of considerable interest. The weaker I:C base pairs formed as a result of template copying in the product duplex would decrease the overall melting temperature of the product duplex. Previous reports indicate that substituting guanosine to inosine decreases the stability of RNA duplexes by 3.44 kcal/mol per G:C to I:C change, primarily due to the lack of hydrogen bonding to the C2-carbonyl of cytidine.[106] With inosine as a surrogate for guanosine, long RNA duplexes will have lower melting temperatures and, thus, inosine could potentially facilitate RNA replication via thermal cycling.

2.4 Conclusion and Outlook

Our results inform the constraints we place on the prebiotic synthesis of nucleotides, because just as the pathway to C also provides U by deamination, so a pathway to A could also provide I by deamination. Although the spontaneous deamination of A to I is very slow, this conversion is greatly accelerated by the presence of nitrous acid, which in turn can be derived from atmospherically generated NO, or NO from cometary impacts. It has recently been proposed that dilute atmospherically derived NO could be concentrated by coordination with ferrocyanide complexes in the form of nitroprusside,[92] which in turn could serve as a starting material for isonitrile mediated nucleotide activation with 2-aminoimidazole. A dual role for NO in nucleotide activation chemistry and more directly in nucleotide synthesis through the deamination of adenosine to inosine would be a satisfyingly parsimonious systems approach to both the origin of

[106] Siegfried, N. A.; Kierzek, R.; Bevilacqua, P. C. Role of Unsatisfied Hydrogen Bond Acceptors in RNA Energetics and Specificity. *J. Am. Chem. Soc.* **2010**.

RNA and RNA replication. Assuming that sufficient I can be derived from A, the search for a prebiotically plausible path to the purine nucleotides need not extend beyond the identification of a path to the synthesis of adenosine.

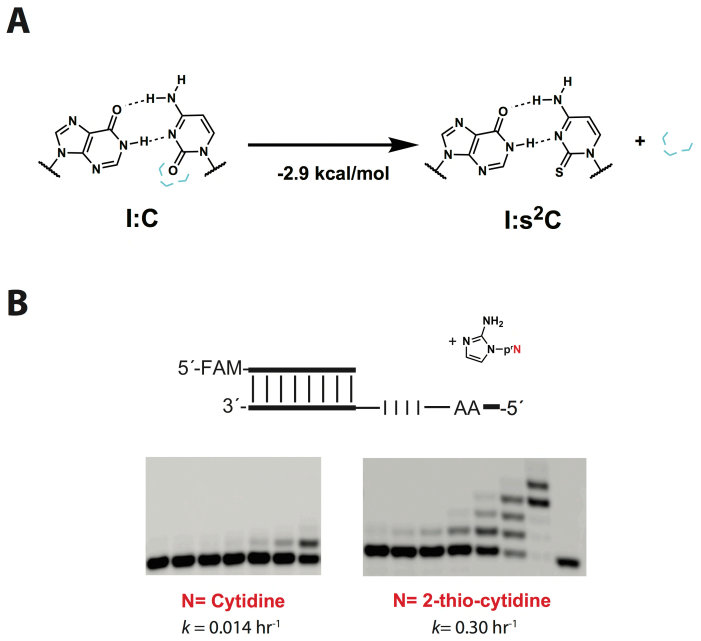


Figure 2.8: (a) Schematic representation of Inosine binding to Cytidine (C, left) and 2-thiocytidine (s^2C, right). Water molecules are represented by light blue color. (b) Gel electrophoresis images and rates of primer extension for 2-aminoimidazole activated cytidine and 2-thio-cytidine across I_4 template.

Additionally, our work has shown that a nucleotide besides the four canonical nucleotides (A, C, G, and U) can exhibit surprisingly rapid and accurate nonenzymatic RNA copying. In line with our observation, we envision that utilizing different prebiotically plausible nucleotides besides inosine could further allow us to construct alternative genetic systems. The Bevilaqua lab has rigorously studied the binding properties of guanosine and inosine and found that 2-thiocytidine (s^2C) binds much stronger to inosine than cytidine does (Figure 2.8 a).[106] The Prebiotic

synthesis of 2-thiocytidine was also completed by the Sutherland lab,[107] suggesting that this alternated nucleotide that binds much stronger to inosine could have existed. In our early evaluation of ongoing studies, we observe that 2-aminoimidazole activated 2-thiocyitidine polymerizes much faster than 2-aminoimidazole activated cytidine across a poly inosine template (Figure 2.8 b). We envision that an alternative genetic system with adenosine, uracil, inosine, and 2-thiocytidine could be both prebiotically plausible and competent to store and transfer genetic information. The landscape of prebiotically plausible nucleotides is expanding and, ultimately, possible permutations of these compounds and the construction of plausible genetic systems will further allow us to understand the RNA world hypothesis and the origin of life.

2.5 Materials and Methods

2.5.1 Oligonucleotide Synthesis

All oligonucleotides used in this study were purchased from Integrated DNA Technologies (Coralville, IA) or prepared by solid-phase synthesis using an Expedite 8909 DNA/RNA synthesizer. Synthesizer reagents and phosphoramidites were purchased from Glen Research (Sterling, VA) and Chemgenes (Wilmington, MA). In-house prepared oligonucleotides were deprotected using standard methods and subsequently purified by polyacrylamide gel electrophoresis. The oligonucleotides were analyzed by high resolution mass spectrometry (HRMS) on an Agilent 6520 QTOF LC-MS.

2.5.2 Synthesis of activated nucleotides

[107] Xu, J.; Tsanakopoulou, M.; Magnani, C. J.; Szabla, R.; Šponer, J. E.; Šponer, J.; Góra, R. W.; Sutherland, J. D. A Prebiotically Plausible Synthesis of Pyrimidine β-Ribonucleosides and Their Phosphate Derivatives Involving Photoanomerization. *Nat. Chem.* **2017**.

Adenosine 5′-phosphoro-2-aminoimidazolide, cytidine 5′-phosphoro-2- aminoimidazolide, guanosine 5′-phosphoro-2- aminoimidazolide, uridine 5′-phosphoro-2- aminoimidazolide, inosine 5′-phosphoro-2- aminoimidazolide, 8-oxo-adenosine 5′-phosphoro-2- aminoimidazolide, 8-oxo-inosine 5′-phosphoro-2- aminoimidazolide and 8-oxo-guanosine 5′-phosphoro-2-aminoimidazolide were prepared according to a previously reported procedure. 8-oxo-adenosine, 8-oxo-inosine and 8-oxo-guanosine were prepared according to reported procedures and ^{1}H NMR spectra and mass match those in the literature. GAC and AGG 5′-phosphoro-2- aminoimidazolide were prepared as described previously. All nucleotides were purified using reverse phase chromatography using a Teledyne Isco Combiflash on a RediSepRf C18Aq column with 20mM triethylammonium bicarbonate and acetonitrile as eluents.

Inosine 5′-phosphoro-2-aminoimidazolide: ^{1}H NMR (400 MHz, Deuterium Oxide) δ 8.00 (s, 1H), 6.62 (t, J = 1.7 Hz, 1H), 6.38 (t, J = 1.7 Hz, 1H), 5.81 (d, J = 4.6 Hz, 1H), 5.12 (t, J = 5.2 Hz, 1H), 4.54 (t, J = 5.1 Hz, 1H), 4.17 (m, 3H). ^{31}P NMR (162 MHz, Deuterium Oxide) δ -10.97. HRMS: calc for [$C_{13}H_{16}N_7O_7P$-H^+] 414.0927, found 414.0943.

8-oxo-adenosine 5′-phosphoro-2-aminoimidazolide: ^{1}H NMR (400 MHz, Deuterium Oxide) δ 7.94 (s, 1H), 6.64 (s, 1H), 6.40 (s, 1H), 5.80 (d, J = 5.5 Hz, 1H), 5.19 (t, J = 5.6 Hz, 1H), 4.46 (t, J = 5.6 Hz, 1H), 4.09 (m, 3H). ^{31}P NMR (162 MHz, Deuterium Oxide) δ -10.53. HRMS: calc for [$C_{13}H_{17}N_8O_7P$-H^+] 429.1036, found 429.1036.

8-oxo-inosine 5′-phosphoro-2-aminoimidazolide: ^{1}H NMR (400 MHz, Deuterium Oxide) δ 8.00 (s, 1H), 6.62 (t, J = 1.7 Hz, 1H), 6.38 (t, J = 1.7 Hz, 1H), 5.81 (d, J = 4.6 Hz, 1H), 5.12 (t, J = 5.2 Hz, 1H), 4.54 (t, J = 5.1 Hz, 1H), 4.17 (m, 3H). ^{31}P NMR (162 MHz, Deuterium Oxide) δ -10.97. HRMS: calc for [$C_{13}H_{16}N_7O_8P$-H^+] 430.0876, found 430.0869.

8-oxo-guanosine 5′-phosphoro-2-aminoimidazolide: ^{1}H NMR (400 MHz, Deuterium Oxide) δ 6.67 (s, 1H), 6.46 (s, 1H), 5.78 (d, J = 5.7 Hz, 1H), 5.16 (t, J = 5.7 Hz, 1H), 4.55 (t, J = 5.3 Hz, 1H), 4.18 (m, 2H), 4.05 (m, 1H). ^{31}P NMR (162 MHz, Deuterium Oxide) δ -11.78. HRMS: calc for [$C_{13}H_{17}N_8O_8P$-H^+] 445.0985, found 445.0980.

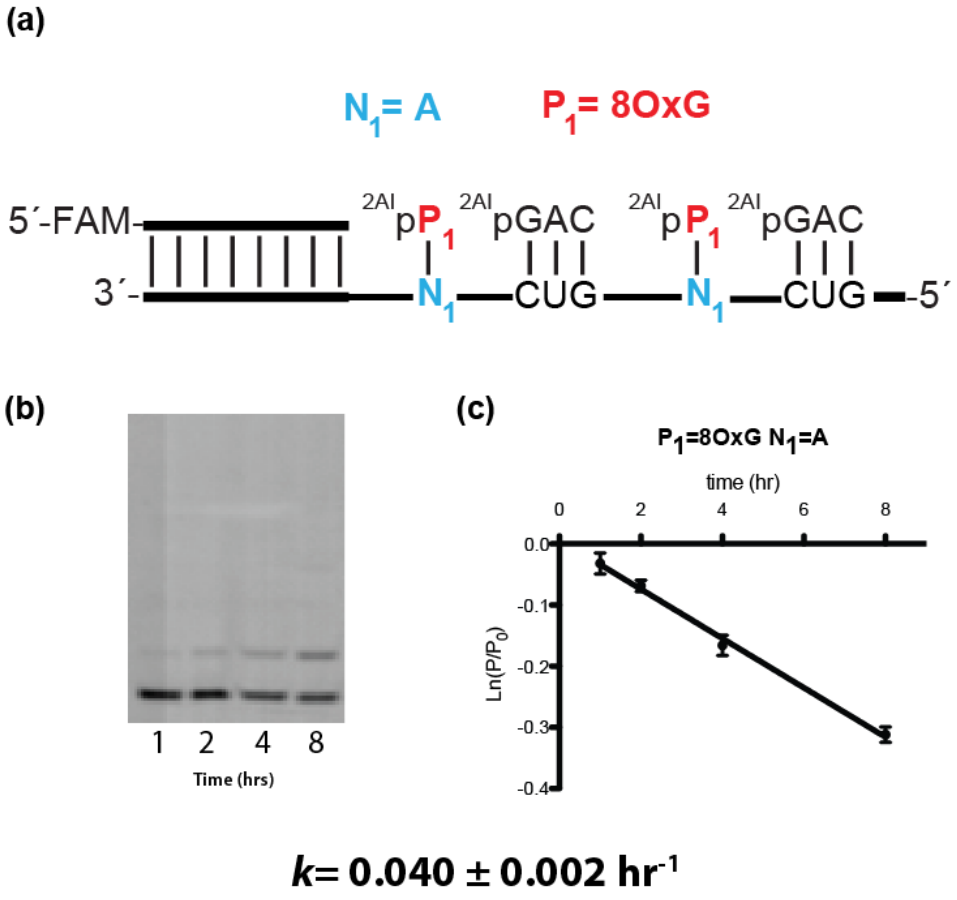


Figure S2.1: Primer extension reactions were carried out in triplicate using 20 mM 2AIp8OxG, 0.5 mM 2AIpGAC, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_1= 8OxG, N_1=A. (b) Representative PAGE analysis of result for P_1= 8OxG, N_1=A. (c) Plot of ln(P/P_0) as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

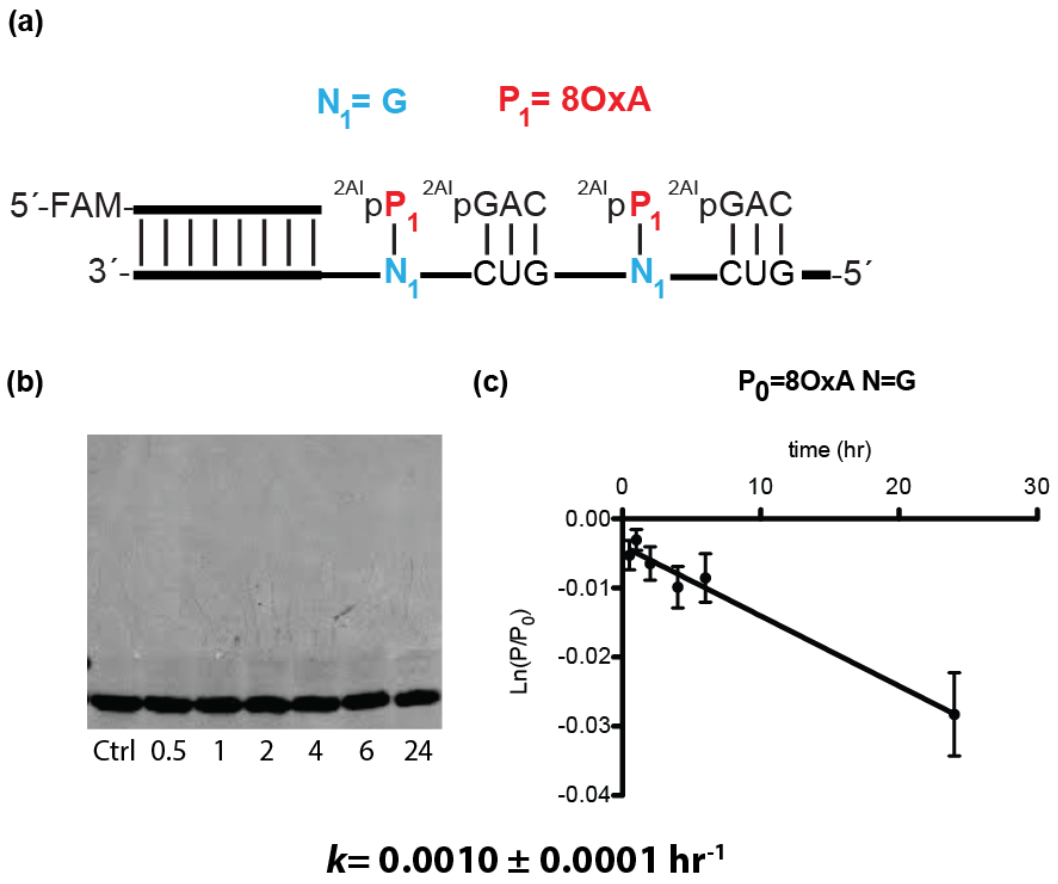


Figure S2.2: Primer extension reactions were carried out in triplicate using 20 mM 2AIp8OxA, 0.5 mM 2AIpGAC, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_1= 8OxA, N_1=G. (b) Representative PAGE analysis of result for P_1= 8OxA, N_1=G. (c) Plot of $\ln(P/P_0)$ as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

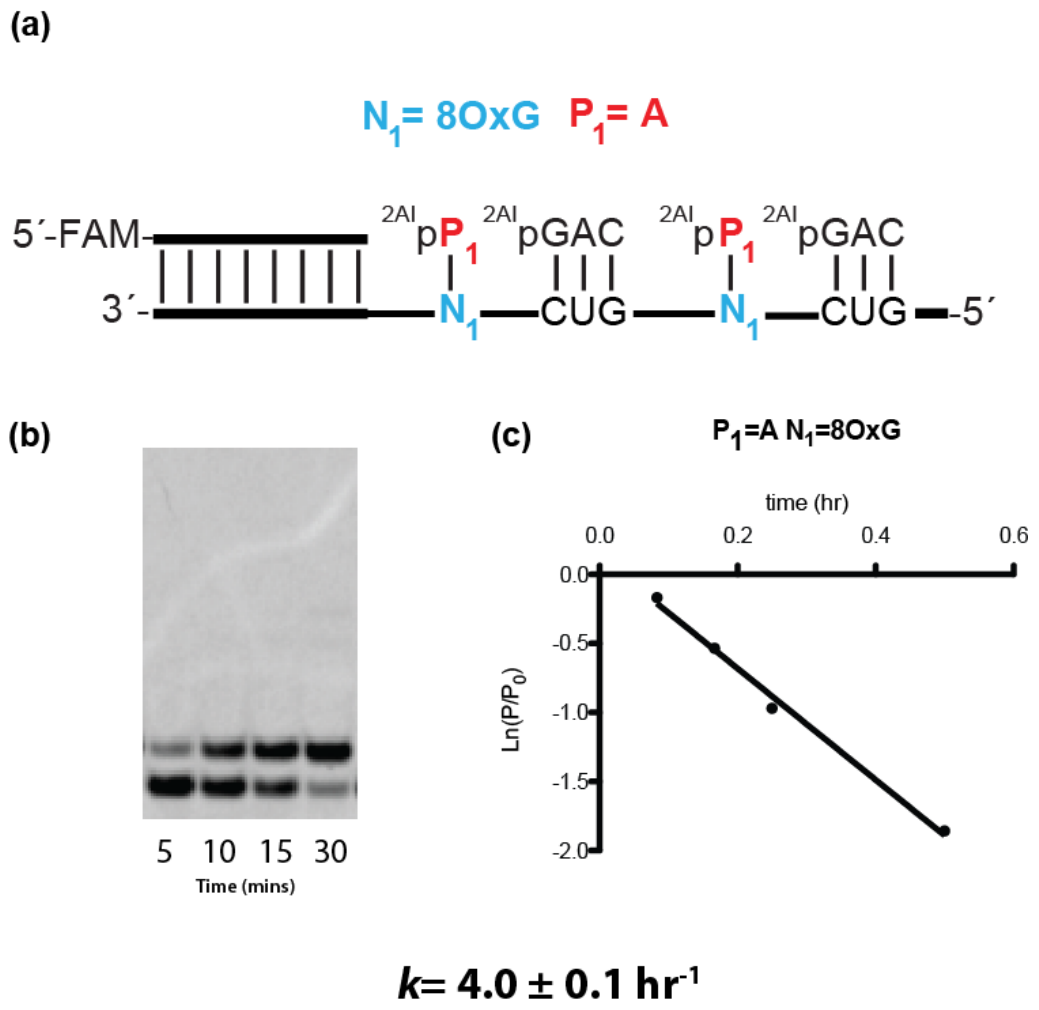


Figure S2.3: Primer extension reactions were carried out in triplicate using 20 mM 2AIpA, 0.5 mM 2AIpGAC, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_1= A, N_1=8OxG. (b) Representative PAGE analysis of result for P_1= A, N_1=8OxG.(c) Plot of ln(P/P_0)as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

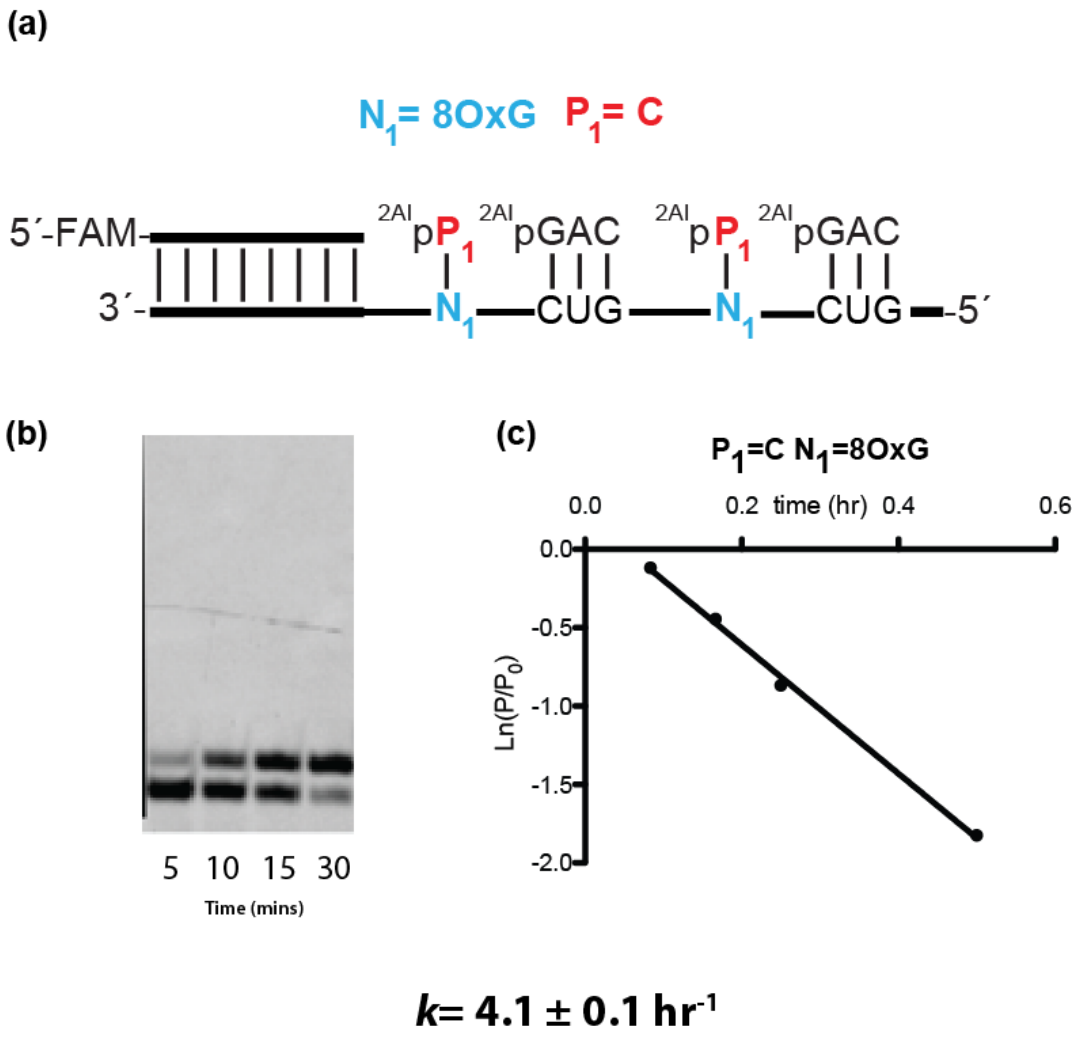


Figure S2.4: Primer extension reactions were carried out in triplicate using 20 mM 2AIpC, 0.5 mM 2AIpGAC, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_1= C, N_1=8OxG. (b) Representative PAGE analysis of result for P_1= C, N_1=8OxG.(c) Plot of ln(P/P_0)as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

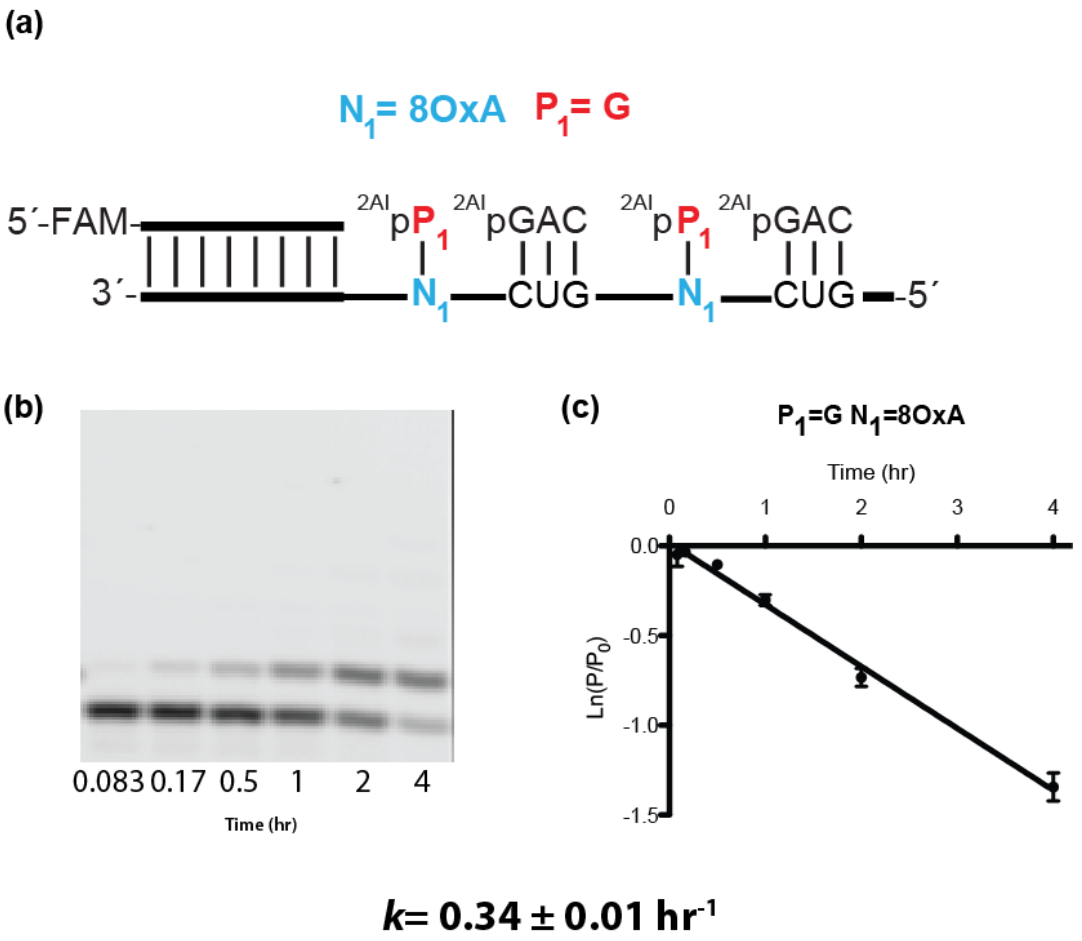


Figure S2.5: Primer extension reactions were carried out in triplicate using 20 mM 2AIpG, 0.5 mM 2AIpGAC, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_1= G, N_1=8OxA. (b) Representative PAGE analysis of result for P_1= G, N_1=8OxA.(c) Plot of ln(P/P_0)as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

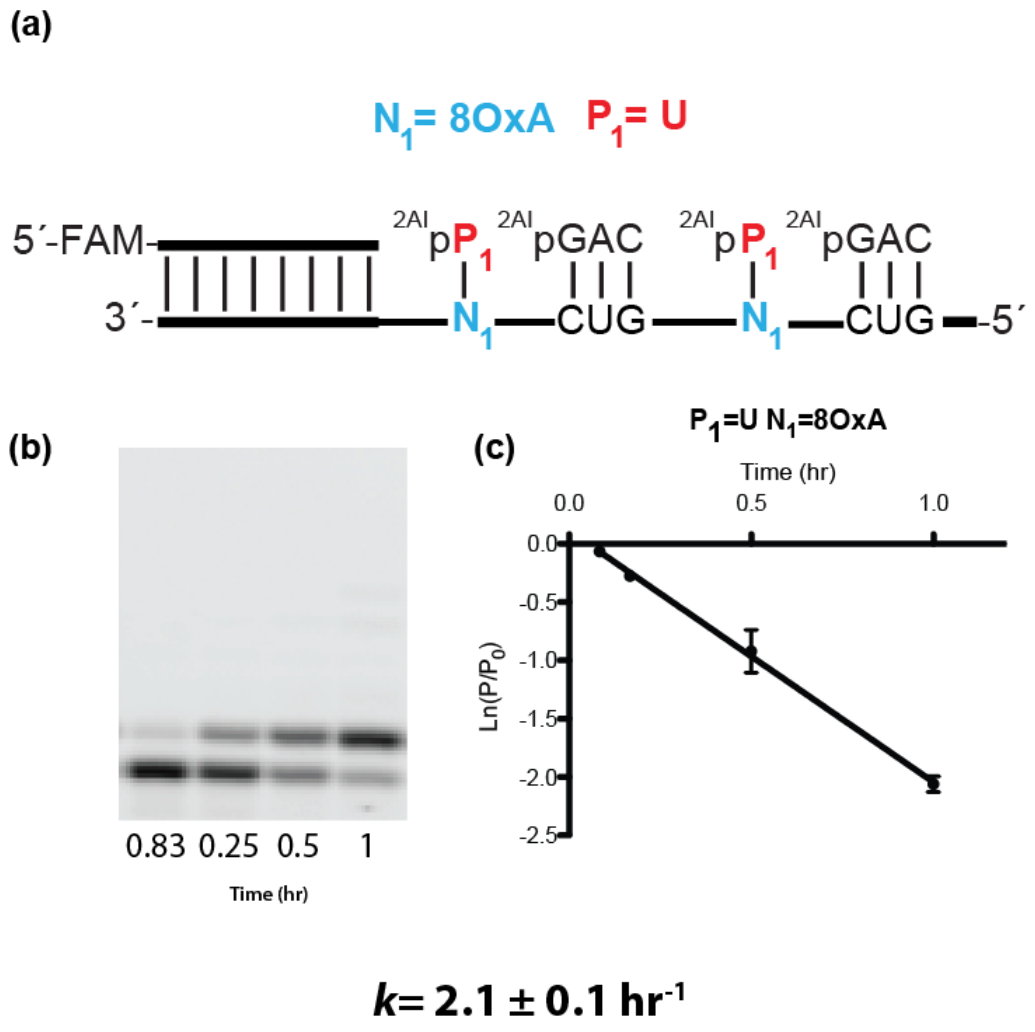


Figure S2.6: Primer extension reactions were carried out in triplicate using 20 mM 2AIpU, 0.5 mM 2AIpGAC, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_1= U, N_1=8OxA. (b) Representative PAGE analysis of result for P_1= U, N_1=8OxA. (c) Plot of ln(P/P_0) as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

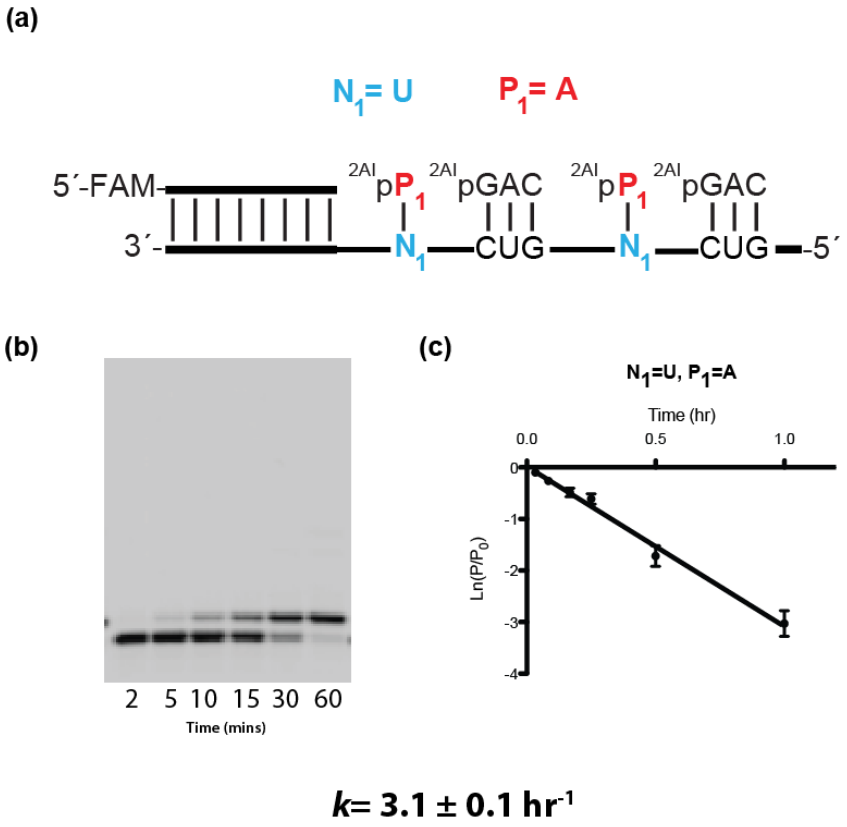


Figure S2.7: Primer extension reactions were carried out in triplicate using 20 mM 2AIpA, 0.5 mM 2AIpGAC, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_1= A, N_1=U. (b) Representative PAGE analysis of result for P_1= A, N_1=U. (c) Plot of ln(P/P_0) as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

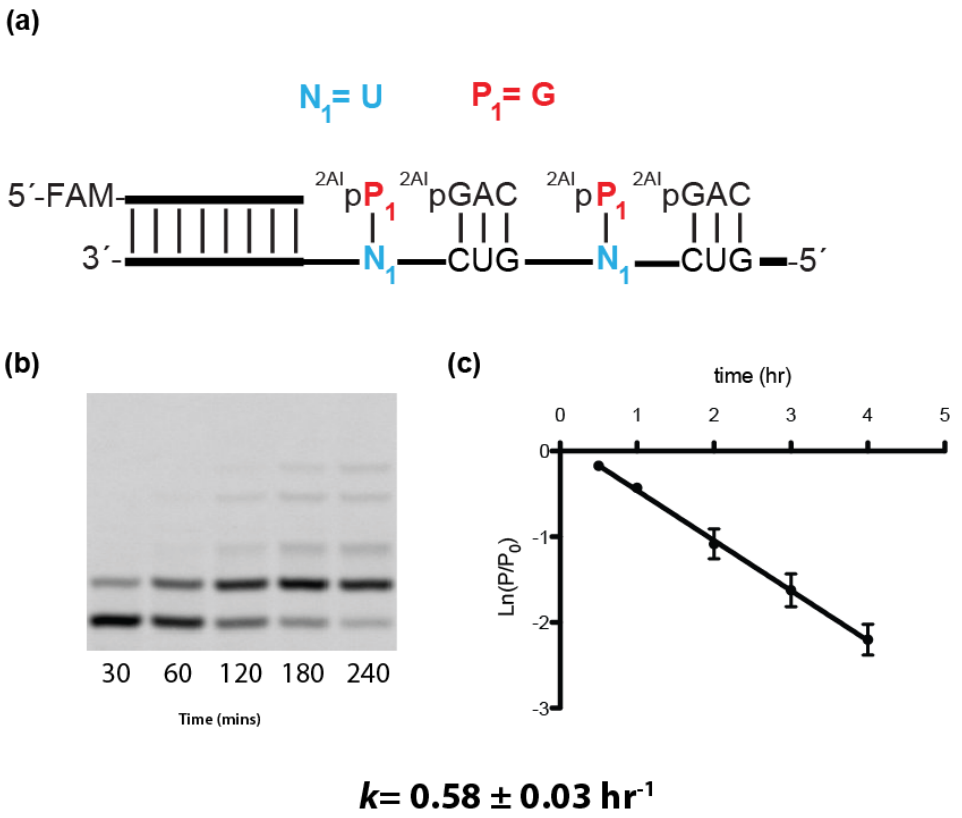


Figure S2.8: Primer extension reactions were carried out in triplicate using 20 mM 2AIpG, 0.5 mM 2AIpGAC, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_1= G, N_1=U. (b) Representative PAGE analysis of result for P_1= G, N_1=U. (c) Plot of ln(P/P_0) as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

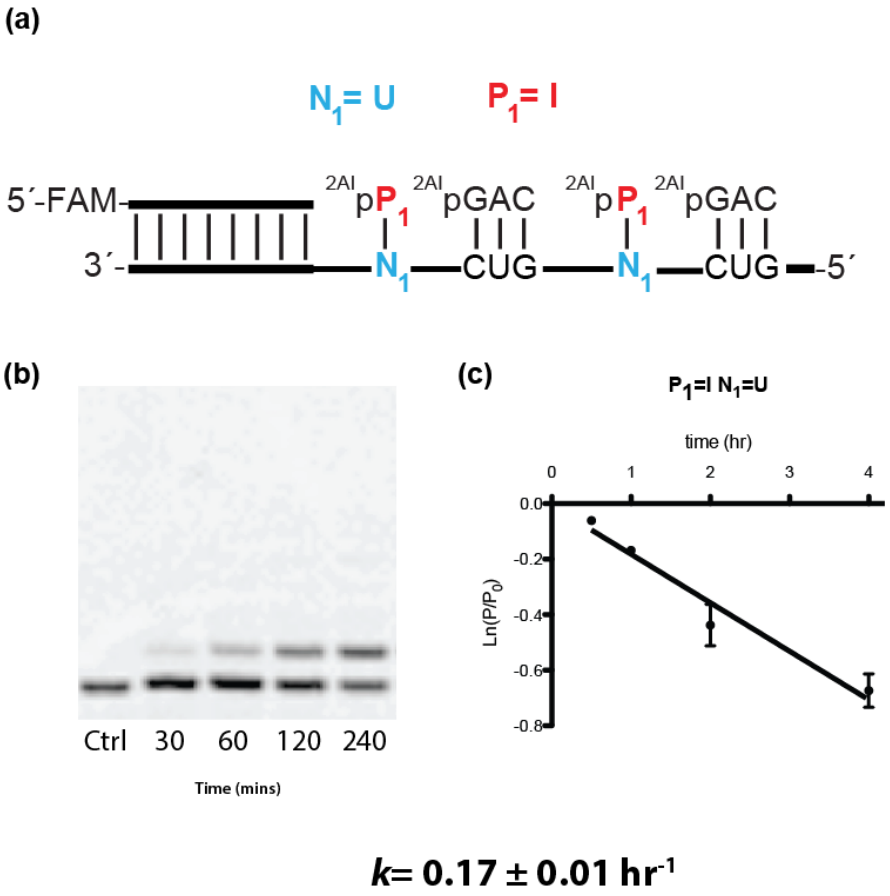


Figure S2.9: Primer extension reactions were carried out in triplicate using 20 mM 2AIpI, 0.5 mM 2AIpGAC, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_1= I, N_1=U. (b) Representative PAGE analysis of result for P_1= I, N_1=U. (c) Plot of ln(P/P_0) as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

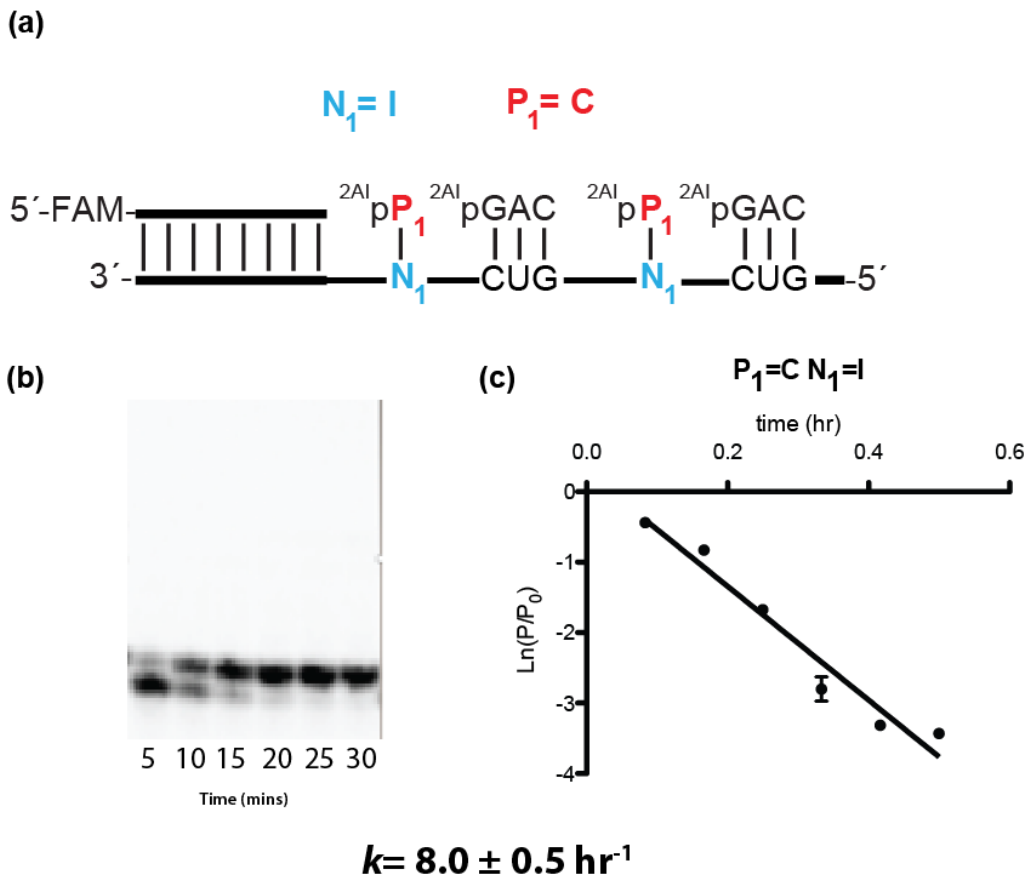
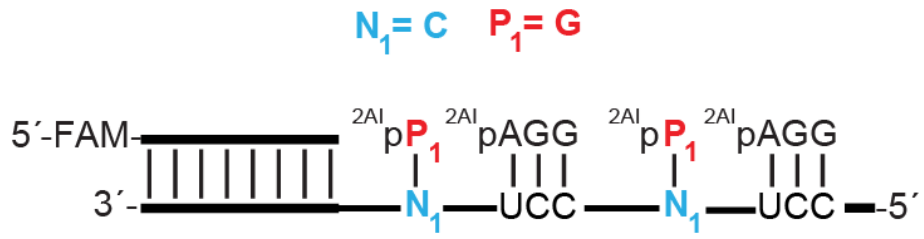
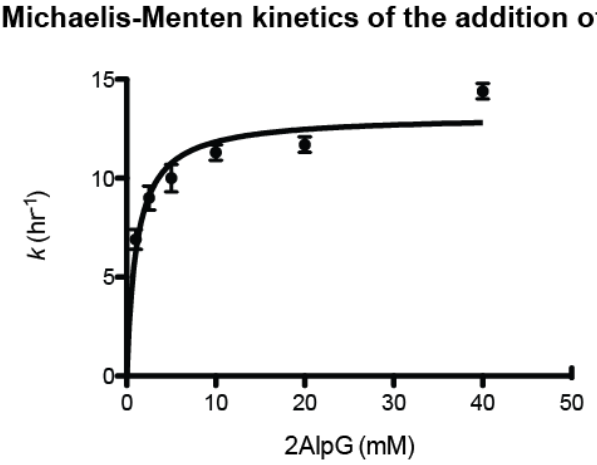


Figure S2.10: Primer extension reactions were carried out in triplicate using 20 mM 2AIpC, 0.5 mM 2AIpGAC, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_1= C, N_1=I. (b) Representative PAGE analysis of result for P_1= C, N_1=I. (c) Plot of ln(P/P_0) as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

(a)



(b)



$K_M = 1.1 \pm 0.3$ mM

$V_{MAX} = 13.2 \pm 0.8$ mM hr^{-1}

Figure S2.11: Michaelis-Menten kinetics of the addition of 2AIpG. Primer extension reactions were carried out in triplicate using various concentrations of 2AIpG, 0.5 mM 2AIpAGG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_1= G, N_1=C. (b) Plot of k (hr^{-1}) as a function of the concentration of 2AIpG. Michaelis-Menten parameters are K_M : 1.1 mM and V_{MAX} of 13.2 mM hr^{-1}.

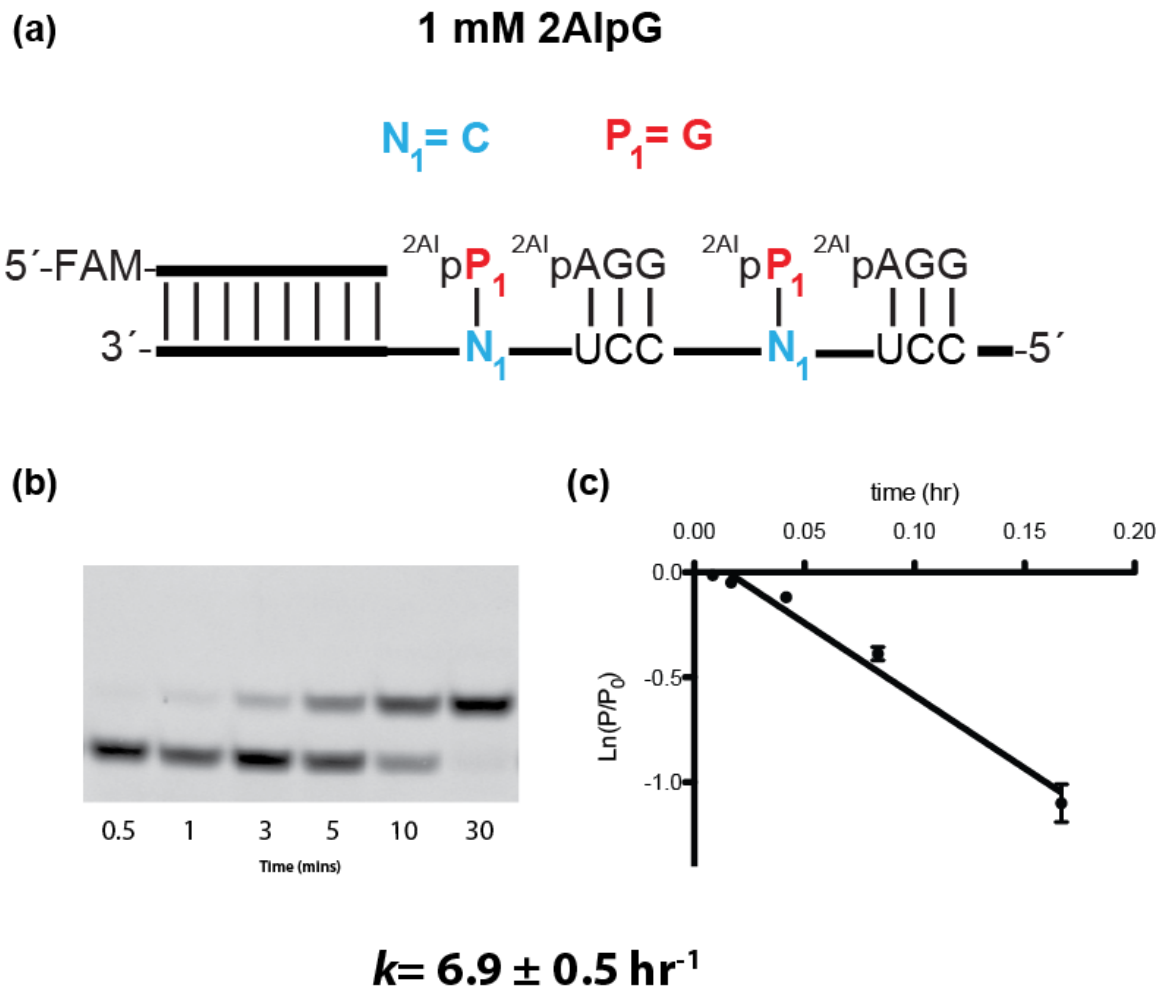


Figure S2.12: Michaelis-Menten kinetics of the addition of 2AIpG. Primer extension reactions were carried out in triplicate using 1 mM of 2AIpG, 0.5 mM 2AIpAGG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_1= G, N_1=C. (b) Plot of ln(P/P_0)as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

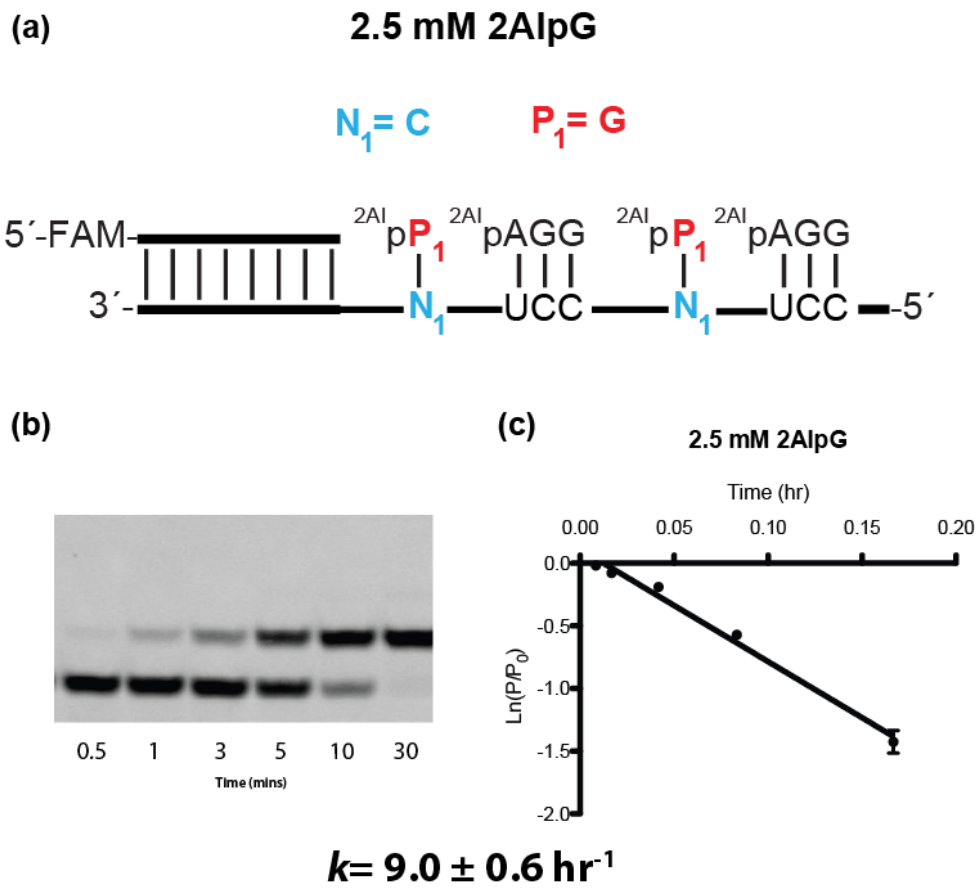


Figure S2.13: Michaelis-Menten kinetics of the addition of 2AIpG. Primer extension reactions were carried out in triplicate using 2.5 mM of 2AIpG, 0.5 mM 2AIpAGG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_1= G, N_1=C. (b) Plot of ln(P/P_0)as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

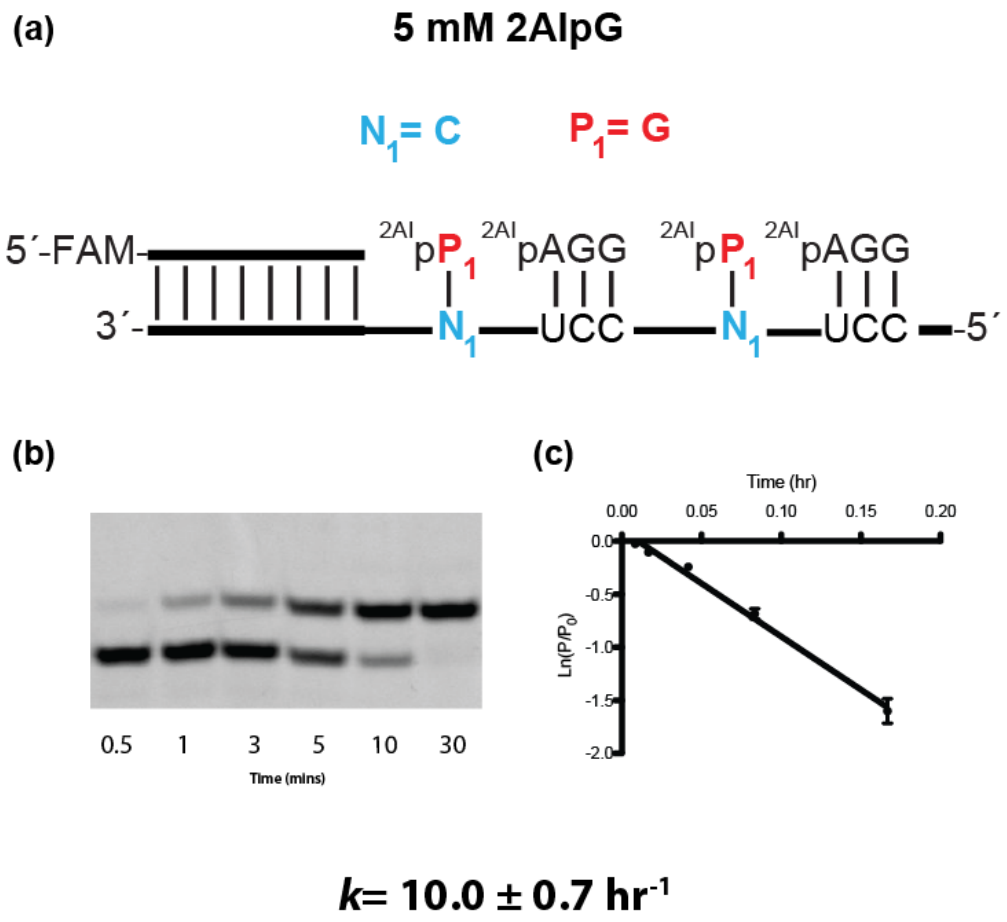


Figure S2.14: Michaelis-Menten kinetics of the addition of 2AIpG. Primer extension reactions were carried out in triplicate using 5 mM of 2AIpG, 0.5 mM 2AIpAGG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_1= G, N_1=C. (b) Plot of ln(P/P_0)as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

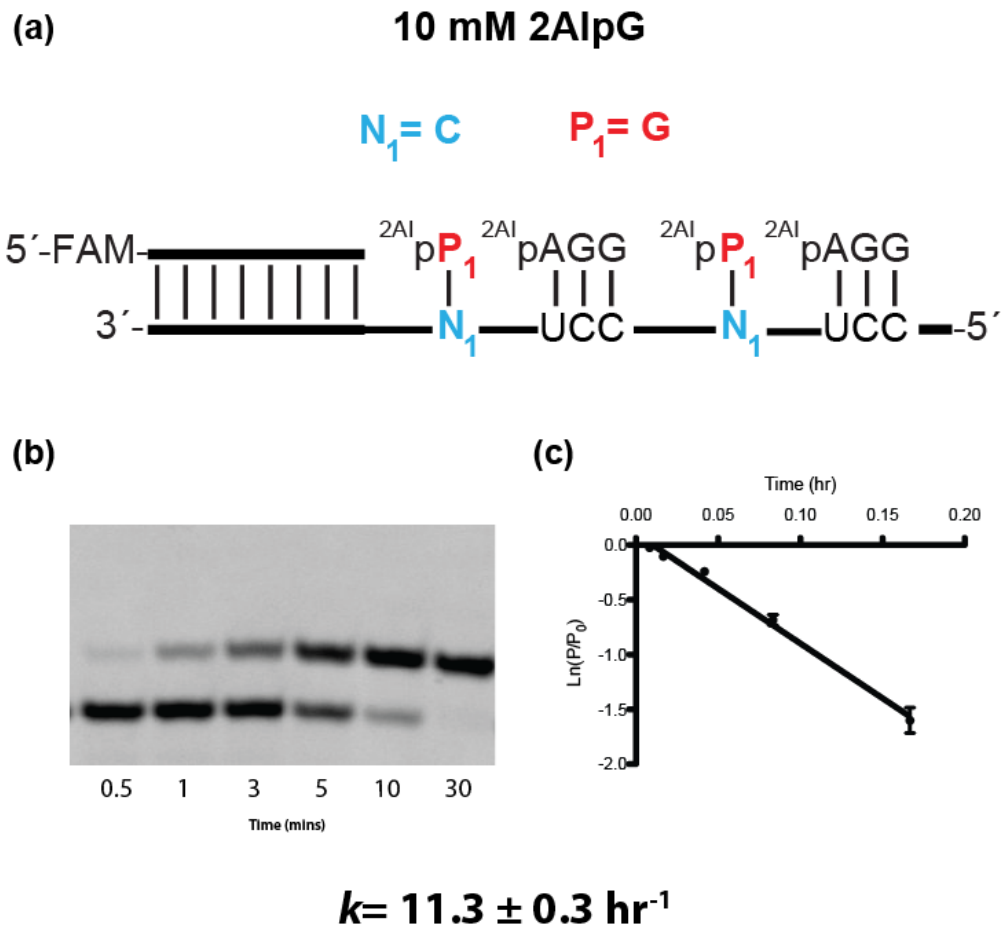
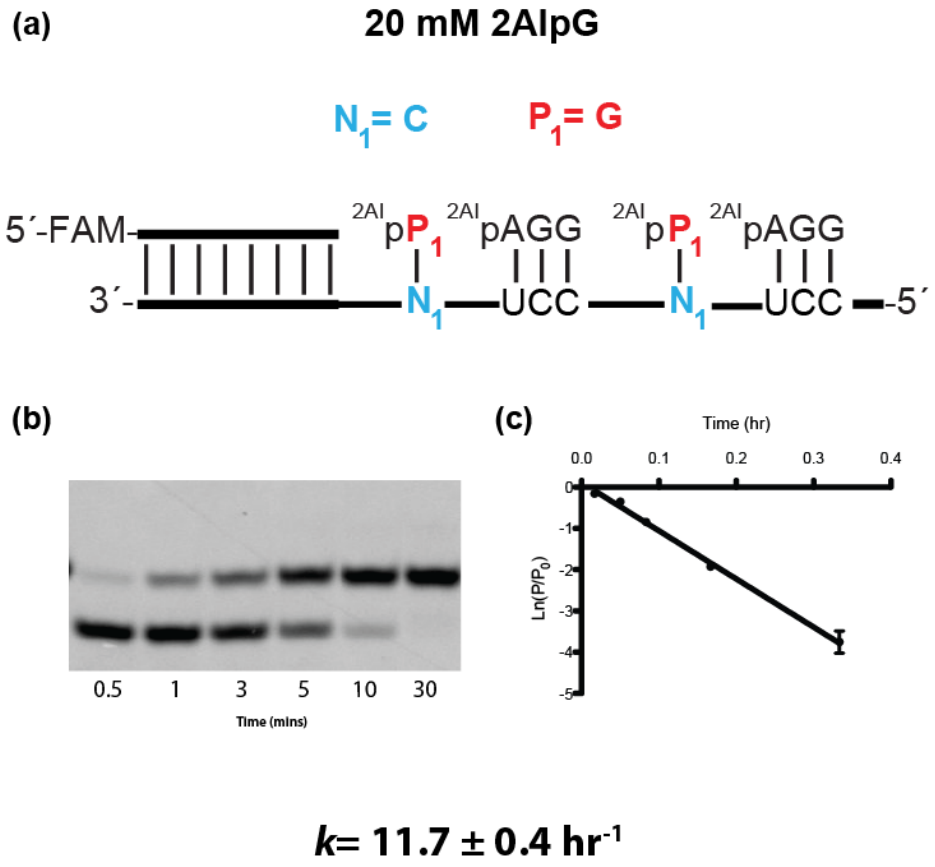


Figure S2.15: Michaelis-Menten kinetics of the addition of 2AIpG. Primer extension reactions were carried out in triplicate using 10 mM of 2AIpG, 0.5 mM 2AIpAGG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_1= G, N_1=C. (b) Plot of ln(P/P_0)as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.



$k = 11.7 \pm 0.4\ hr^{-1}$

Figure S2.16: Michaelis-Menten kinetics of the addition of 2AIpG. Primer extension reactions were carried out in triplicate using 20 mM of 2AIpG, 0.5 mM 2AIpAGG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_1= G, N_1=C. (b) Plot of ln(P/P_0)as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

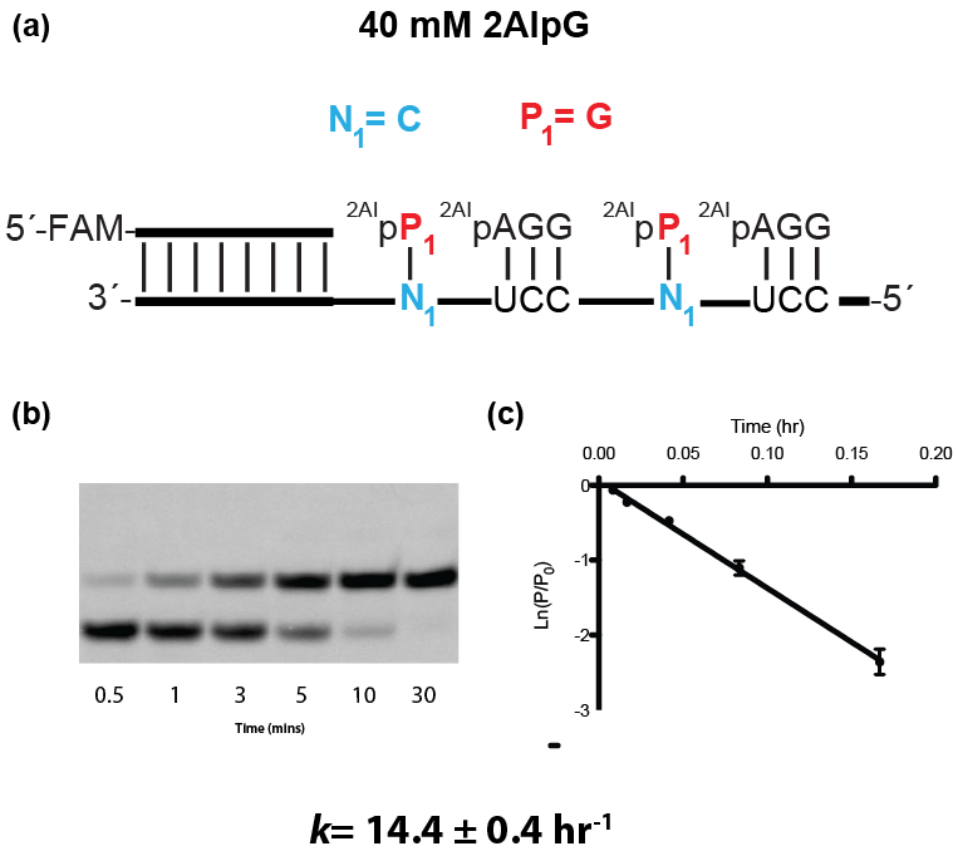


Figure S2.17: Michaelis-Menten kinetics of the addition of 2AIpG. Primer extension reactions were carried out in triplicate using 40 mM of 2AIpG, 0.5 mM 2AIpAGG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_1= G, N_1=C. (b) Plot of ln(P/P_0)as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

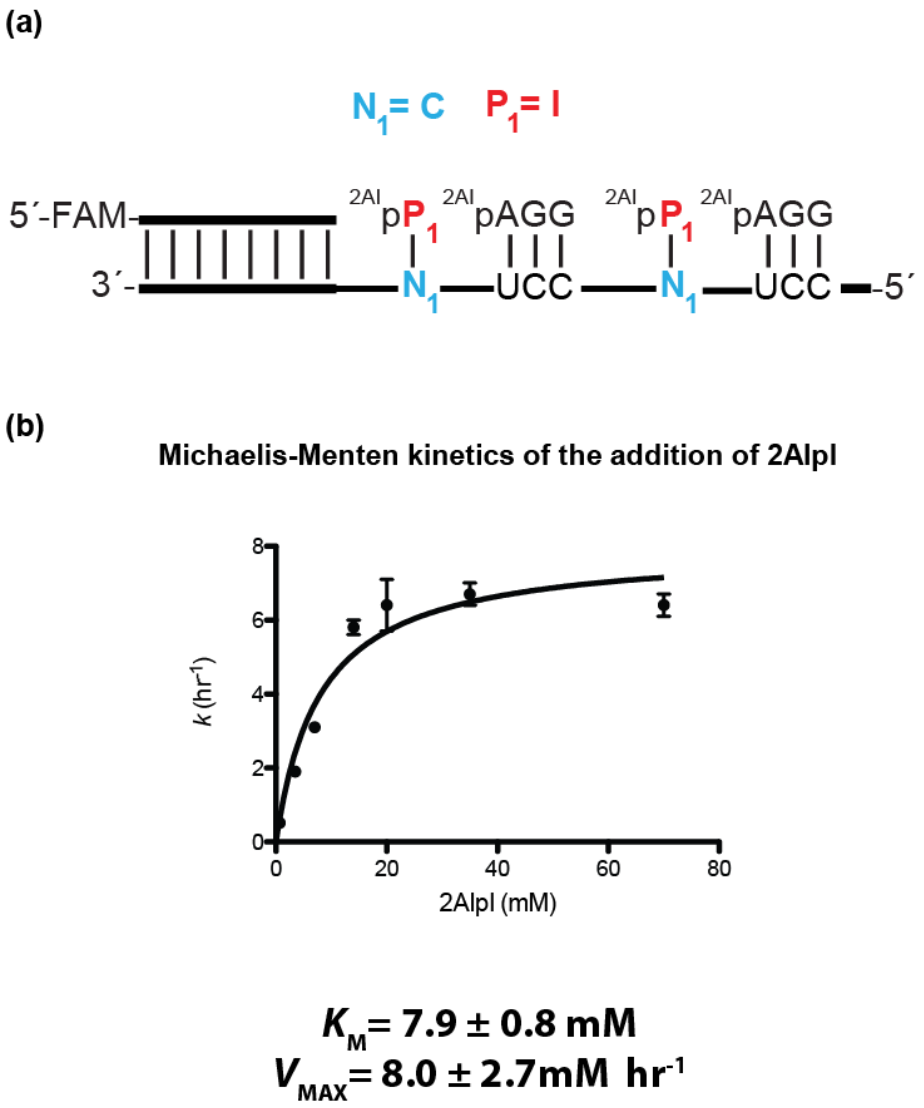


Figure S2.18: Michaelis-Menten kinetics of the addition of 2AIpI. Primer extension reactions were carried out in triplicate using various concentrations of 2AIpI, 0.5 mM 2AIpAGG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_1= I, N_1=C. (b) Plot of the k (hr^{-1}) as a function of the concentration of 2AIpI. Michaelis-Menten parameters are K_M : 7.9 mM and V_{MAX} of 8.0 mM hr^{-1}.

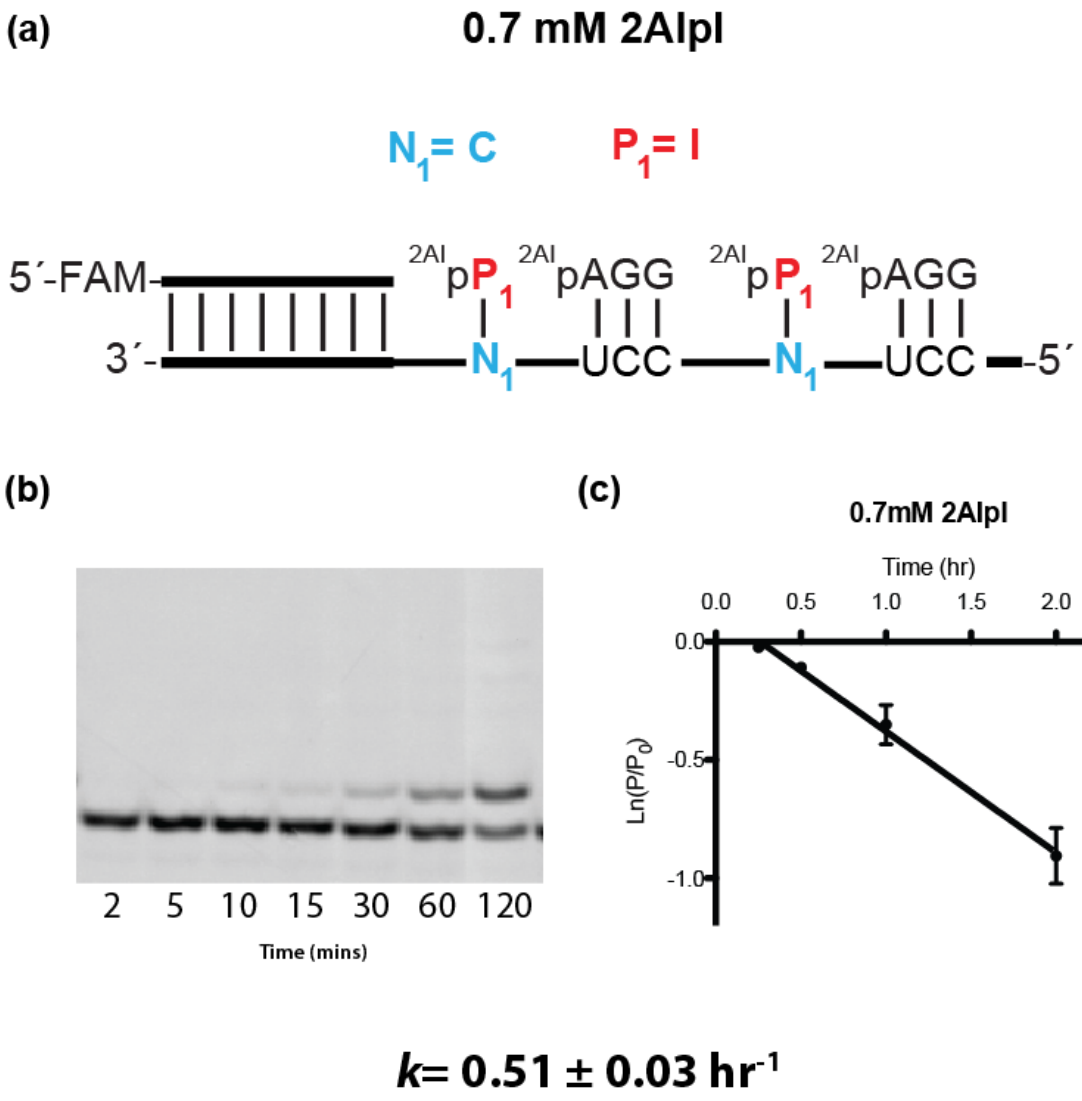


Figure S2.19: Michaelis-Menten kinetics of the addition of 2AIpI. Primer extension reactions were carried out in triplicate using 0.7 mM of 2AIpI, 0.5 mM 2AIpAGG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_1= I, N_1=C. (b) Plot of ln(P/P_0)as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

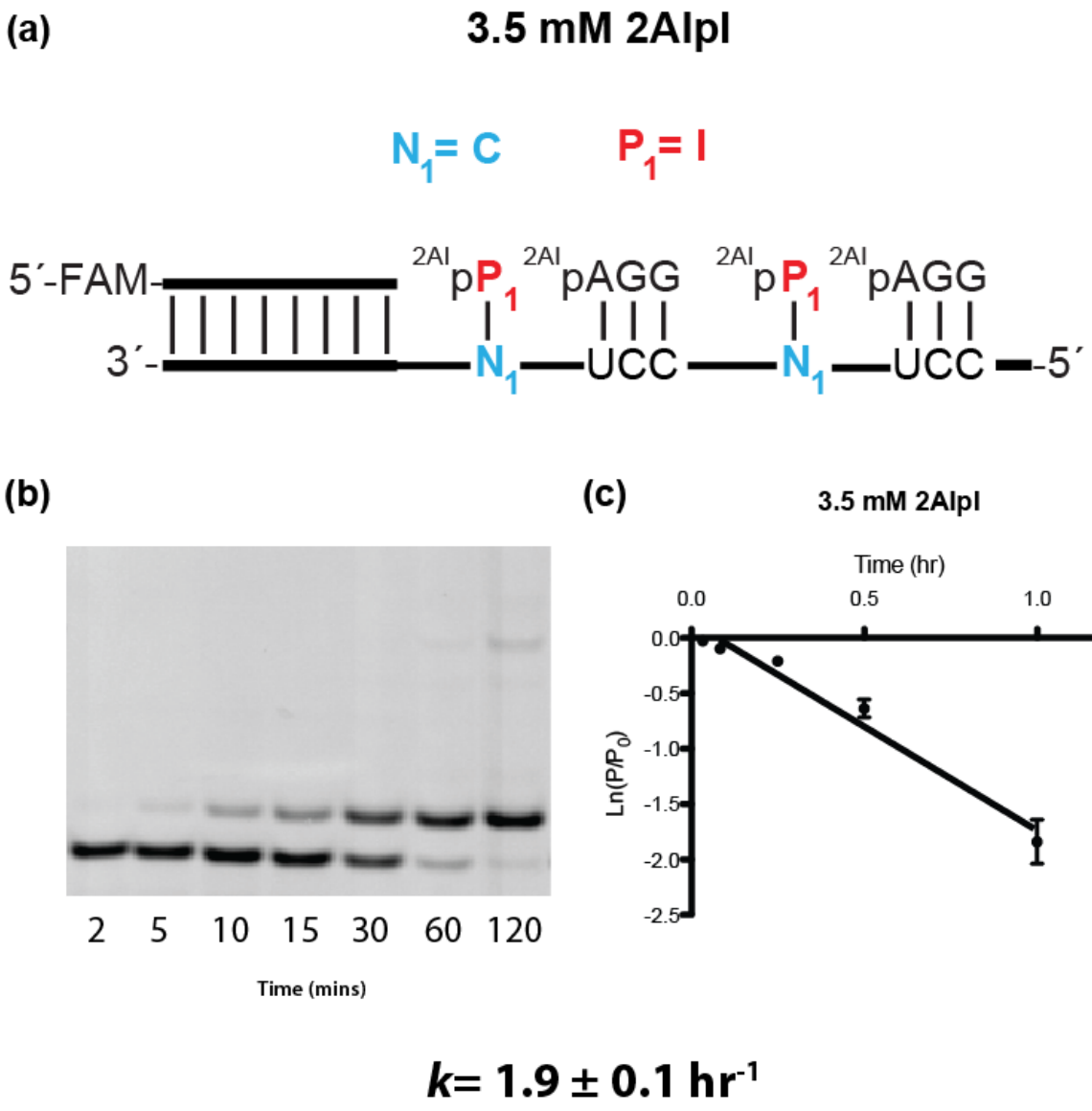


Figure S2.20: Michaelis-Menten kinetics of the addition of 2AIpI. Primer extension reactions were carried out in triplicate using 3.5 mM of 2AIpI, 0.5 mM 2AIpAGG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_1= I, N_1=C. (b) Plot of ln(P/P_0)as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

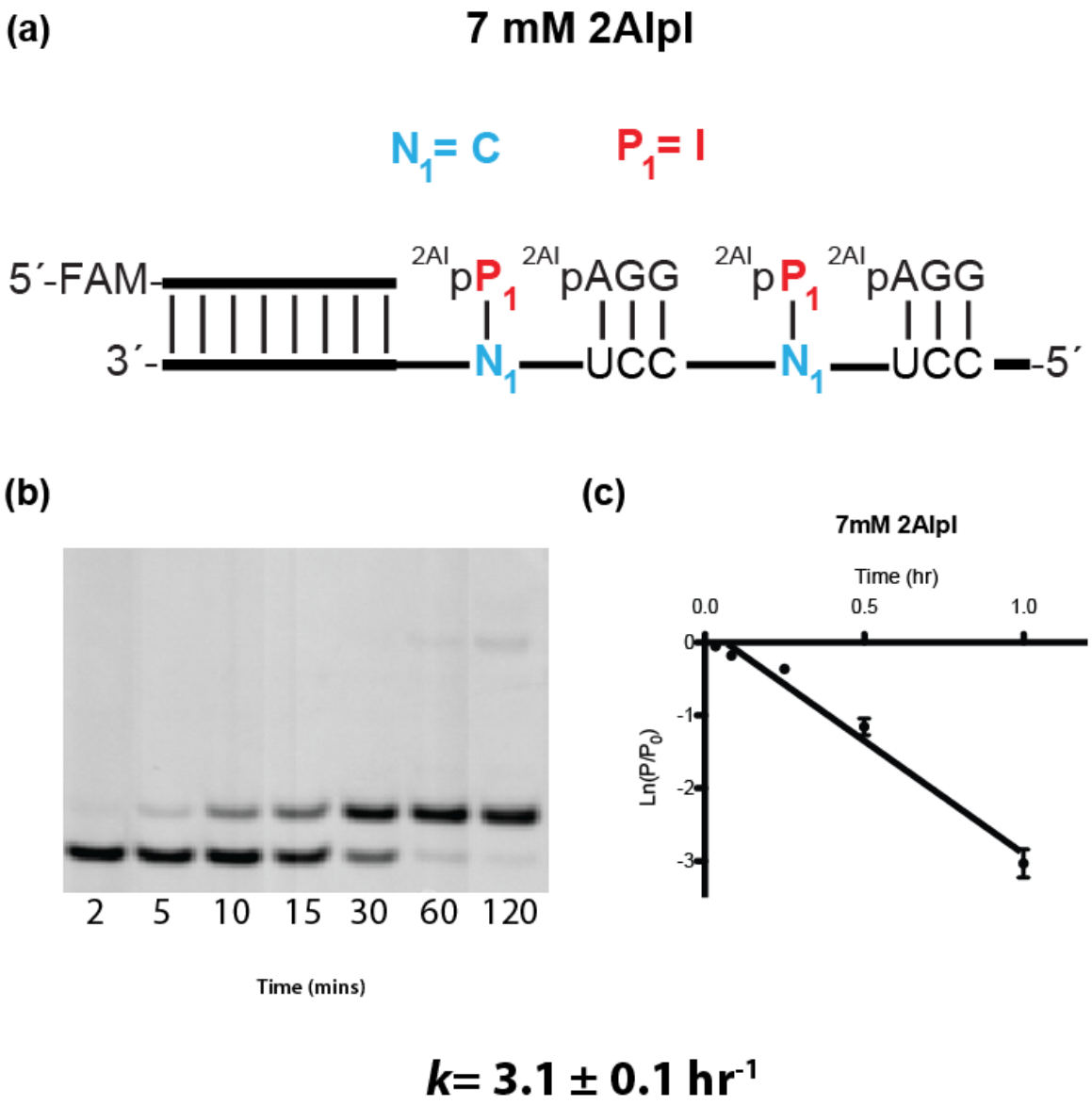


Figure S2.21: Michaelis-Menten kinetics of the addition of 2AIpI. Primer extension reactions were carried out in triplicate using 7 mM of 2AIpI, 0.5 mM 2AIpAGG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_1= I, N_1=C. (b) Plot of ln(P/P_0)as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

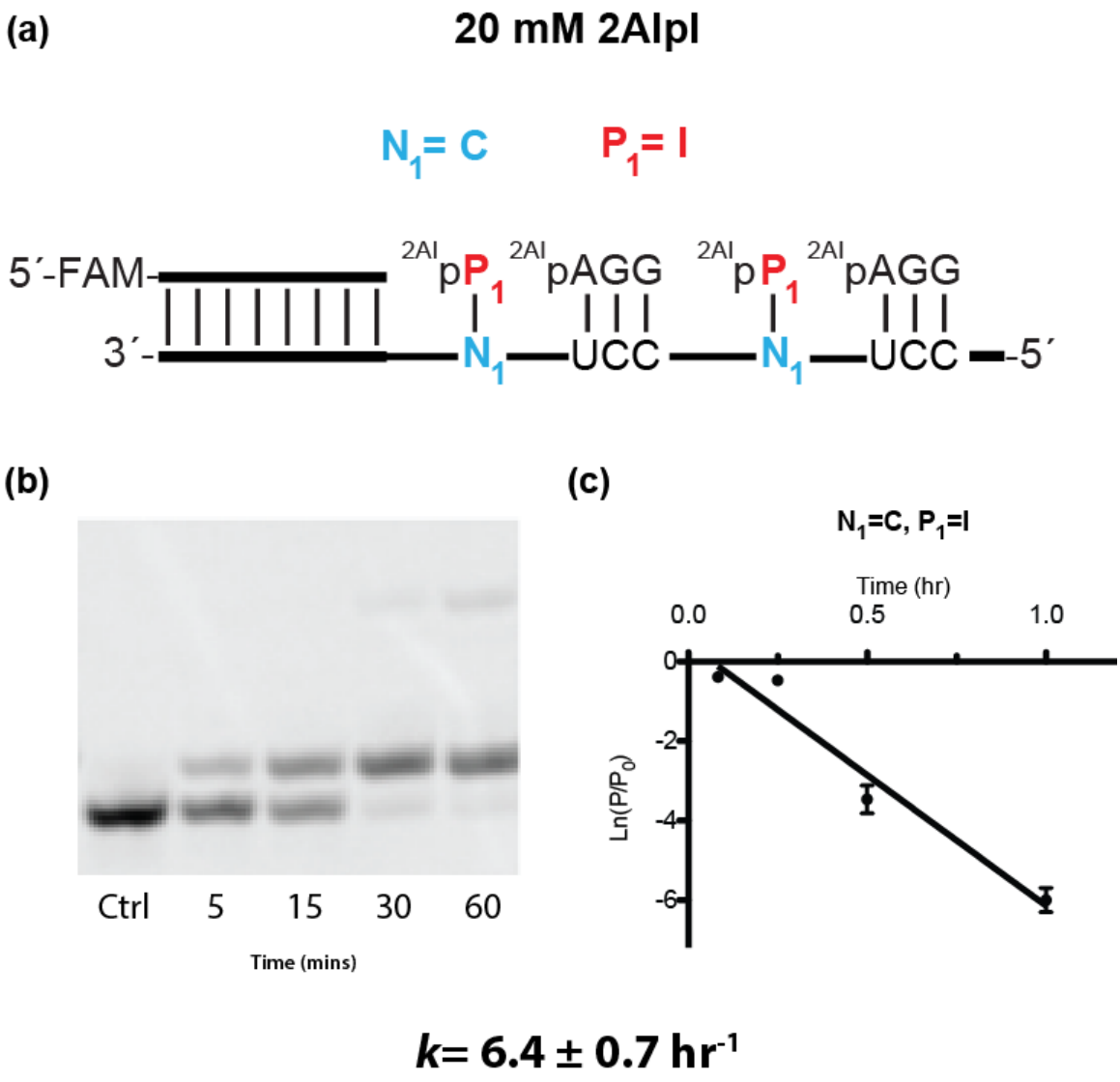


Figure S2.22: Michaelis-Menten kinetics of the addition of 2AIpI. Primer extension reactions were carried out in triplicate using 20 mM of 2AIpI, 0.5 mM 2AIpAGG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_1= I, N_1=C. (b) Plot of ln(P/P_0)as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

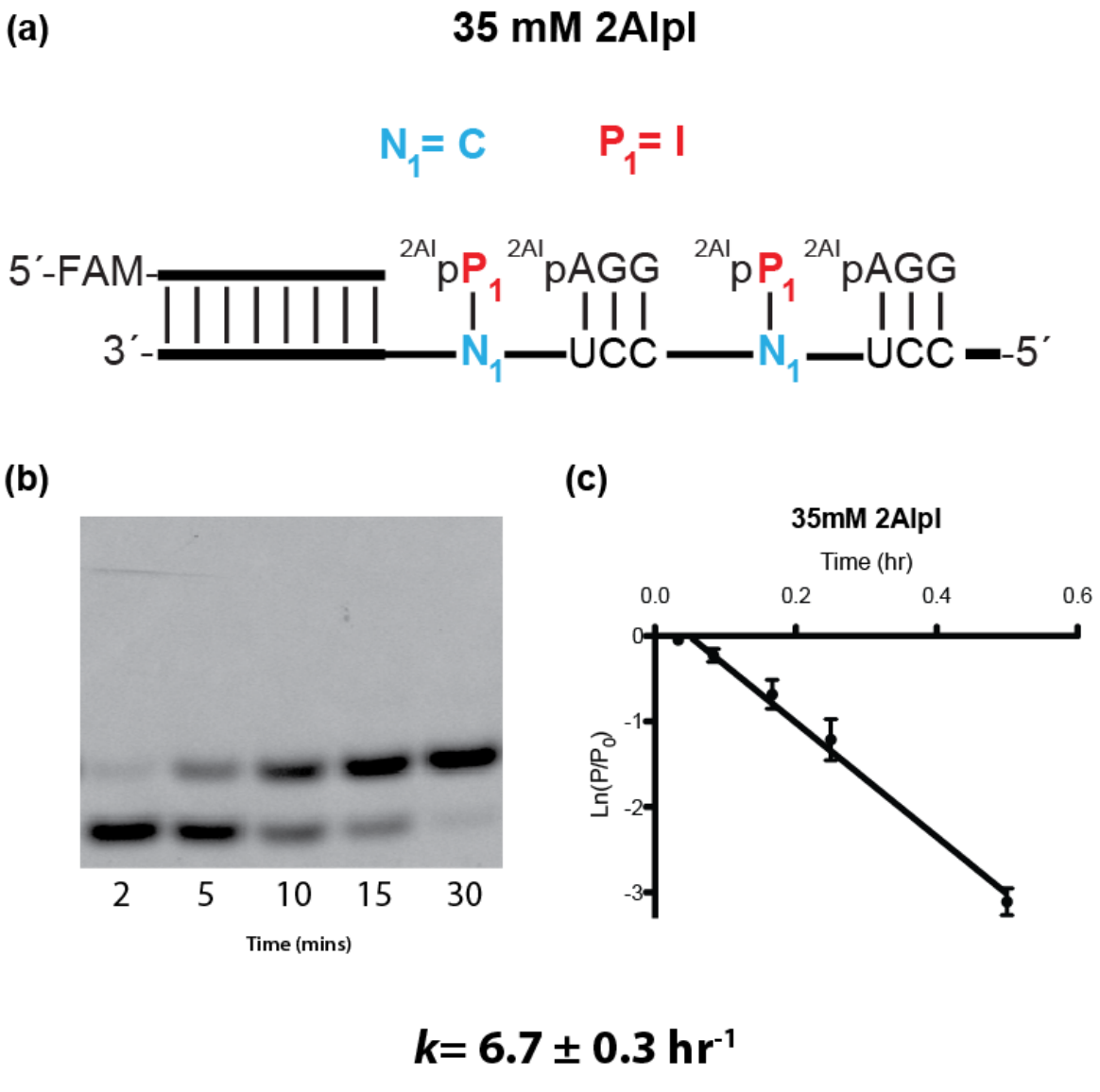
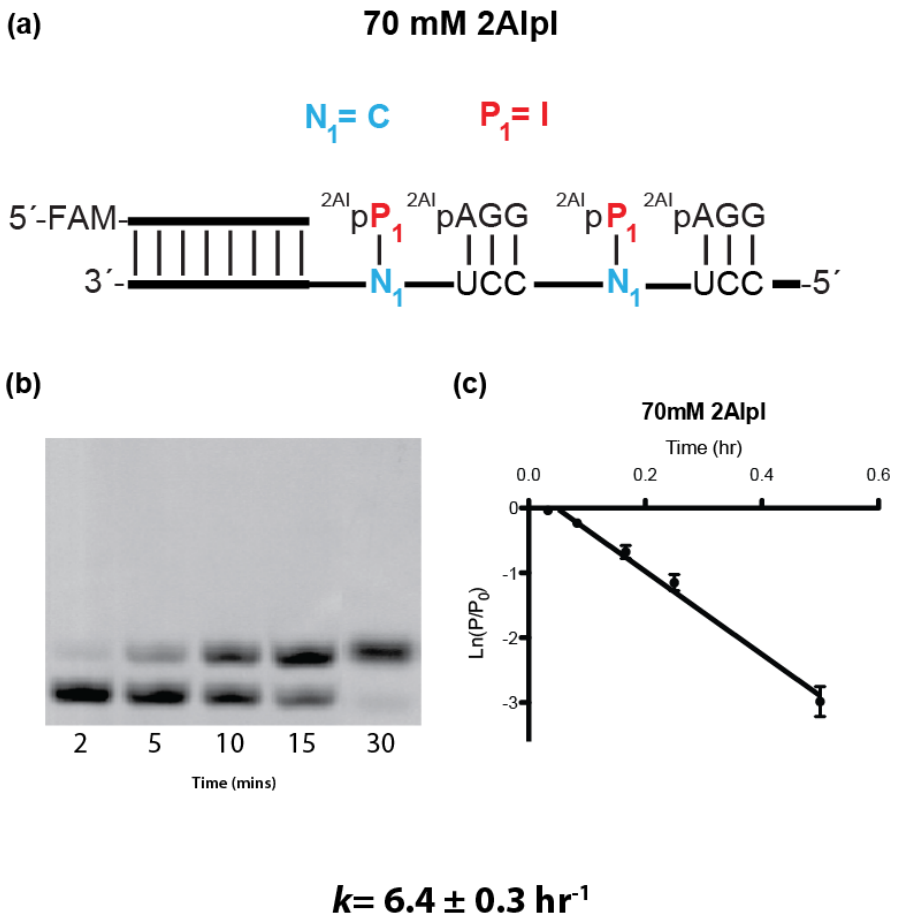


Figure S2.23: Michaelis-Menten kinetics of the addition of 2AIpI. Primer extension reactions were carried out in triplicate using 35 mM of 2AIpI, 0.5 mM 2AIpAGG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_1= I, N_1=C. (b) Plot of ln(P/P_0)as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.



$k = 6.4 \pm 0.3\ hr^{-1}$

Figure S2.24: Michaelis-Menten kinetics of the addition of 2AIpI. Primer extension reactions were carried out in triplicate using 70 mM of 2AIpI, 0.5 mM 2AIpAGG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_1= I, N_1=C. (b) Plot of ln(P/P_0)as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

(a)

N_1 \ P_1	A	U	C	G	μ_N
A	0.013(4)	6.2(2)	0.11(3)	0.0021(1)	0.020
U	3.1(1)	0.17(1)	0.034(3)	0.57(4)	0.20
C	0.17(2)	0.065(3)	0.16(1)	11.7(5)	0.033
I	0.05(1)	0.015(1)	8.0(5)	0.31(2)	0.045
				μ_{avg}	0.075

(b)

N_1 \ P_1	A	U	C	I	μ_N
A	0.013(4)	6.2(2)	0.11(3)	0.024(1)	0.023
U	3.1(1)	0.17(1)	0.034(3)	0.17(1)	0.11
C	0.17(2)	0.065(3)	0.16(1)	6.4(7)	0.058
G	0.0025(1)	0.075(2)	10.2(2)	0.0014(1)	0.0077
				μ_{avg}	0.050

Figure S2.25: Evaluation of Projected misincorporation rates ($\boldsymbol{\mu_N}$) in the mixed system of AUCG and AUCI. All reactions were performed at pH 8.0, 200 mM HEPES, 50 mM Mg^{2+}, 20 mM $2AIpP_1$, 0.5 mM of activated trimer. (a) Rates of primer extension for activated canonical monomers (AUCG) $2AIpP_1$ across A, U, C, and I. Projected misincorporation rates ($\boldsymbol{\mu_N}$) for template bases N_1 were determined by dividing the sum of the incorporation rates across the incorrect pairs by the sum of all the incorporation rates across represented P_1 bases. (b) Rates of primer extension for activated monomers (AUCI) $2AIpP_1$ across A, U, C, and G. Projected misincorporation rates ($\boldsymbol{\mu_N}$) were determined analogously as in (a).

(b)

		P_1 A	U	C	I	μ_N	μ'_N
	A	0.013(4)	6.2(2)	0.11(3)	0.024(1)	0.023	0.017
	U	3.1(1)	0.17(1)	0.034(3)	0.17(1)	0.11	0.052
N_1	C	0.17(2)	0.065(3)	0.16(1)	6.4(7)	0.058	0.026
	I	0.05(1)	0.015(1)	8.0(5)	0.31(2)	0.0089	0.0019
					μ_{avg}	0.049	0.024

(c)

		P_1 A	U	C	G	μ_N	μ'_N
	A	0.013(4)	6.2(2)	0.11(3)	0.0021(1)	0.020	0.017
	U	3.1(1)	0.17(1)	0.034(3)	0.57(4)	0.20	0.16
N_1	C	0.17(2)	0.065(3)	0.16(1)	11.7(5)	0.033	0.014
	G	0.0025(1)	0.075(2)	10.2(2)	0.040(9)	0.011	0.0072
					μ_{avg}	0.066	0.050

Figure S2.26: Evaluation of adjusted misincorporation rates (μ'_N). All reactions were performed at pH 8.0, 200 mM HEPES, 50 mM Mg^{2+}, 20 mM 2AIpP$_1$, 0.5 mM of activated trimer. (a) Rates of primer extension for activated monomers 2AIpP$_1$ across A, U, C, and I. Projected misincorporation rates (μ_N) for template bases N$_1$ were determined by dividing the sum of the incorporation rates across the incorrect pairs by the sum of all the incorporation rates across represented P$_1$ bases. Adjusted misincorporation rates (μ'_N) were determined by dividing the sum of the incorporation rates across the incorrect purine-pyrimidine base pairs by the sum of the incorporation rates across purine-pyrimidine base pairs represented P$_1$ bases. (b) Rates of primer extension for activated monomers (AUCI) 2AIpP$_1$ across A,U,C, and G. Projected misincorporation rates (μ_N) were determined analogously as in (a).

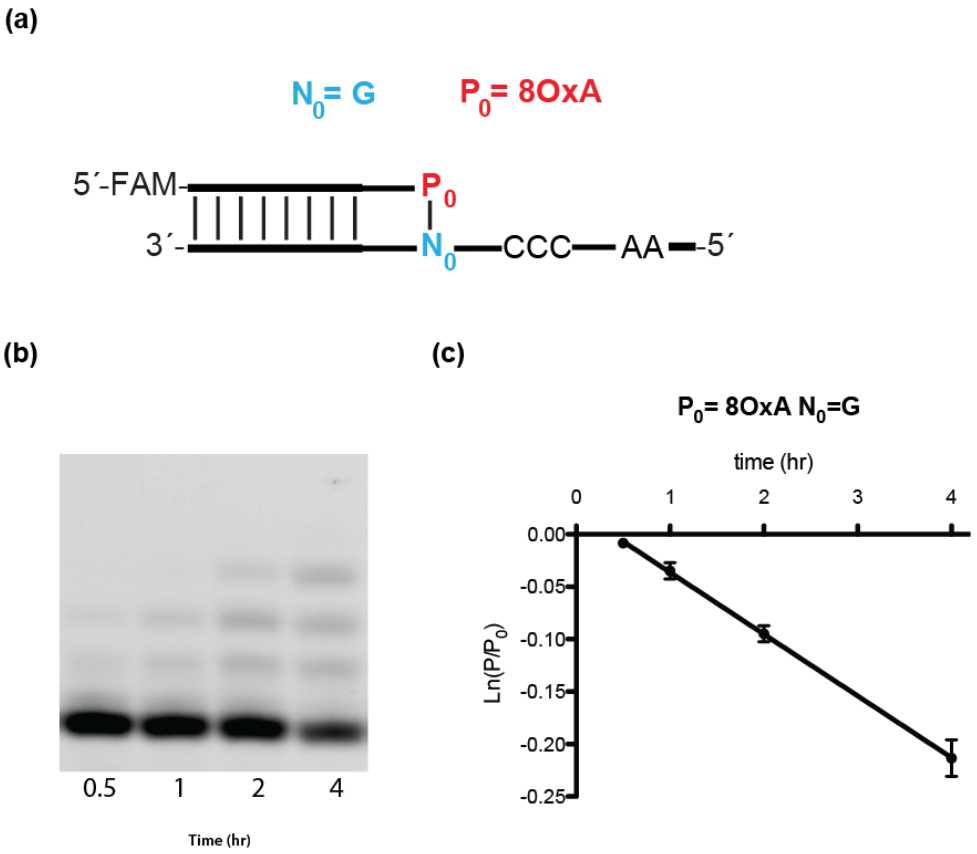


Figure S2.27: Stalling factor for terminal 8OxA:G primer template base pair. Primer extension reactions were carried out in triplicate using 20 mM 2AIpG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_0= 8OxA, N_0=G. (b) Representative PAGE analysis of result for P_0= 8OxA, N_0=G. (c) Plot of ln(P/P_0) as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments. Stalling factor (*S*) was determined by dividing the rate of 2AIpG addition for P_0=C, N_0=G by that of P_0= 8OxA, N_0=G ($k_{C:G}$/ $k_{8OxA:G}$).

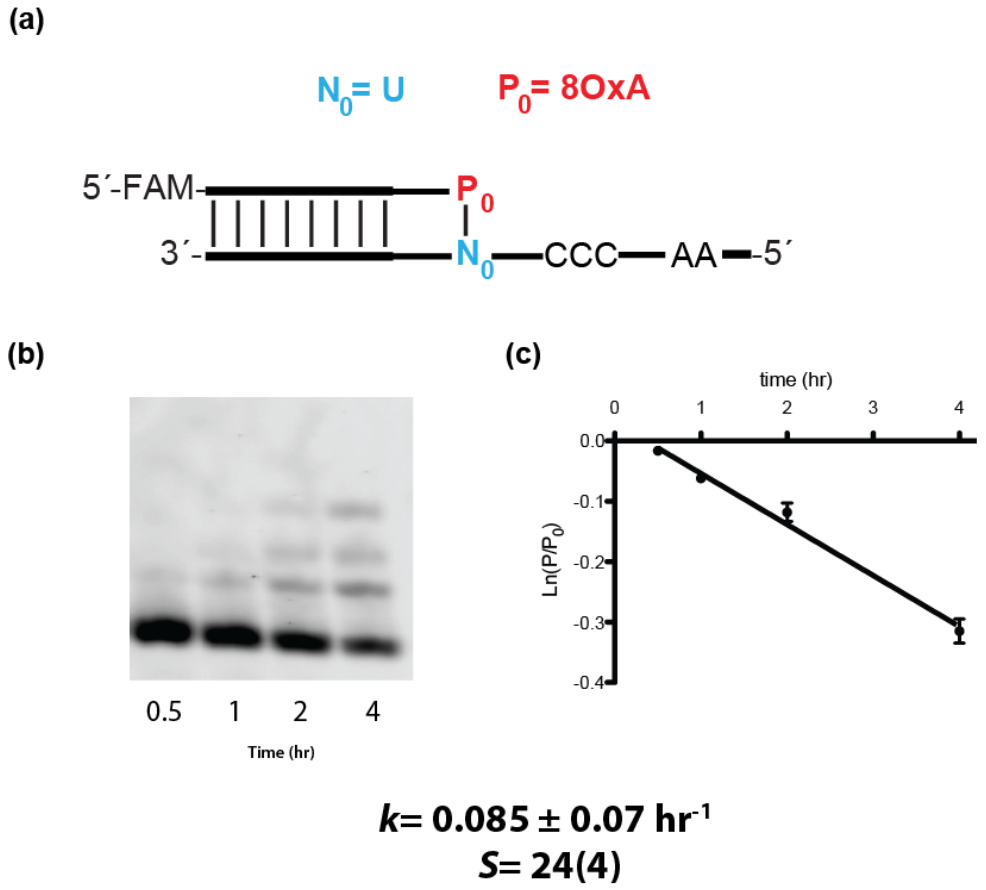


Figure S2.28: Stalling factor for terminal 8OxA:U primer template base pair. Primer extension reactions were carried out in triplicate using 20 mM 2AIpG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_0= 8OxA, N_0=U. (b) Representative PAGE analysis of result for P_0= 8OxA, N_0=U. (c) Plot of ln(P/P_0) as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments. Stalling factor (*S*) was determined by dividing the rate of 2AIpG addition for P_0=A, N_0=U by that of P_0= 8OxA, N_0=G ($k_{A:U}$/ $k_{8OxA:U}$).

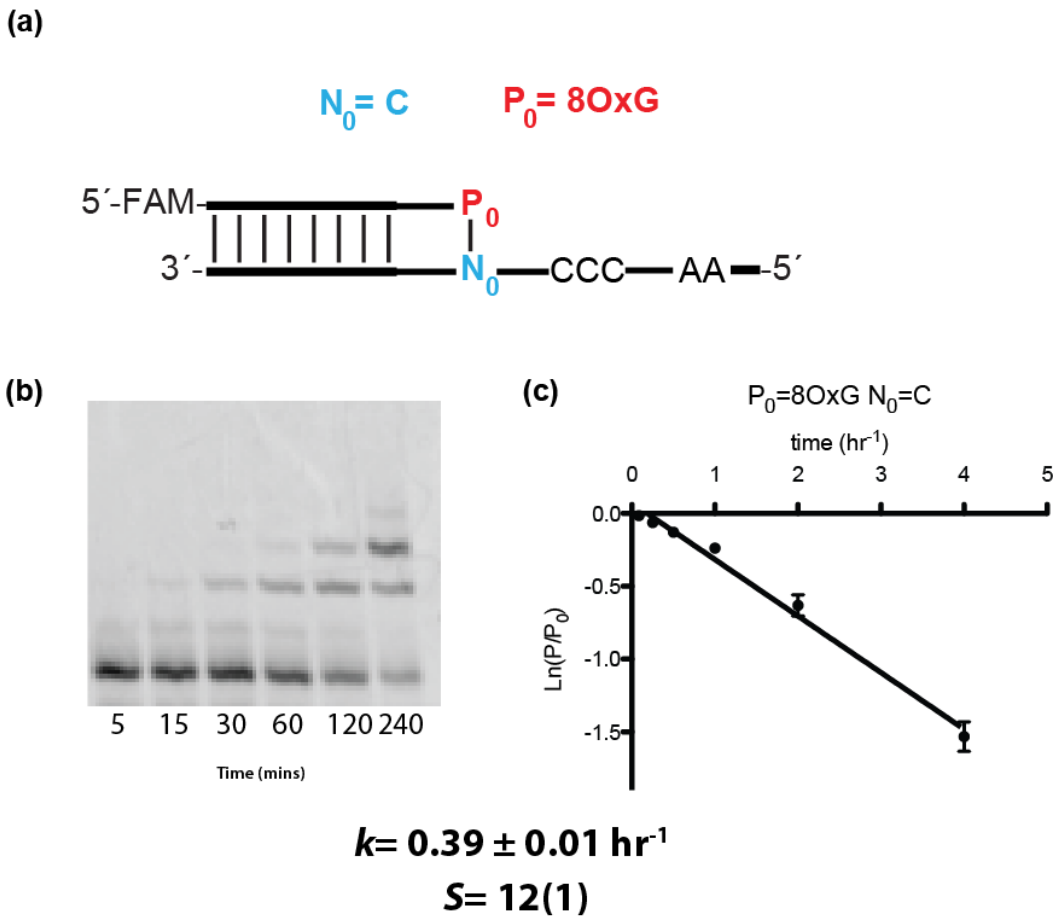


Figure S2.29: Stalling factor for terminal 8OxG:C primer template base pair. Primer extension reactions were carried out in triplicate using 20 mM 2AIpG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_0= 8OxG, N_0=C. (b) Representative PAGE analysis of result for P_0= 8OxG, N_0=C. (c) Plot of ln(P/P_0) as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments. Stalling factor (*S*) was determined by dividing the rate of 2AIpG addition for P_0=G, N_0=C by that of P_0= 8OxG, N_0=C ($k_{G:C}$/ $k_{8OxG:C}$).

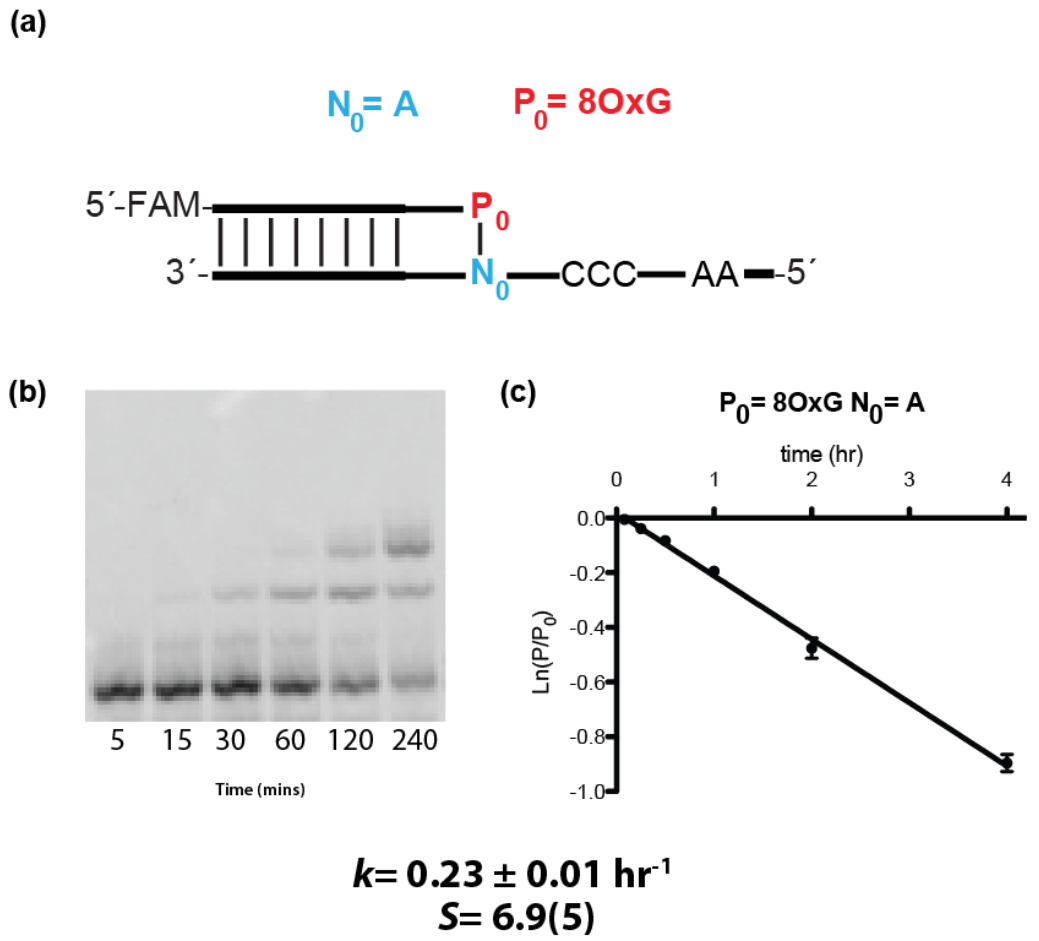


Figure S2.30: Stalling factor for terminal 8OxG:A primer template base pair. Primer extension reactions were carried out in triplicate using 20 mM 2AIpG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_0= 8OxG, N_0=A. (b) Representative PAGE analysis of result for P_0= 8OxG, N_0=A. (c) Plot of ln(P/P_0) as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments. Stalling factor (*S*) was determined by dividing the rate of 2AIpG addition for P_0=U, N_0=A by that of P_0= 8OxG, N_0=A ($k_{U:A}$/ $k_{8OxG:A}$).

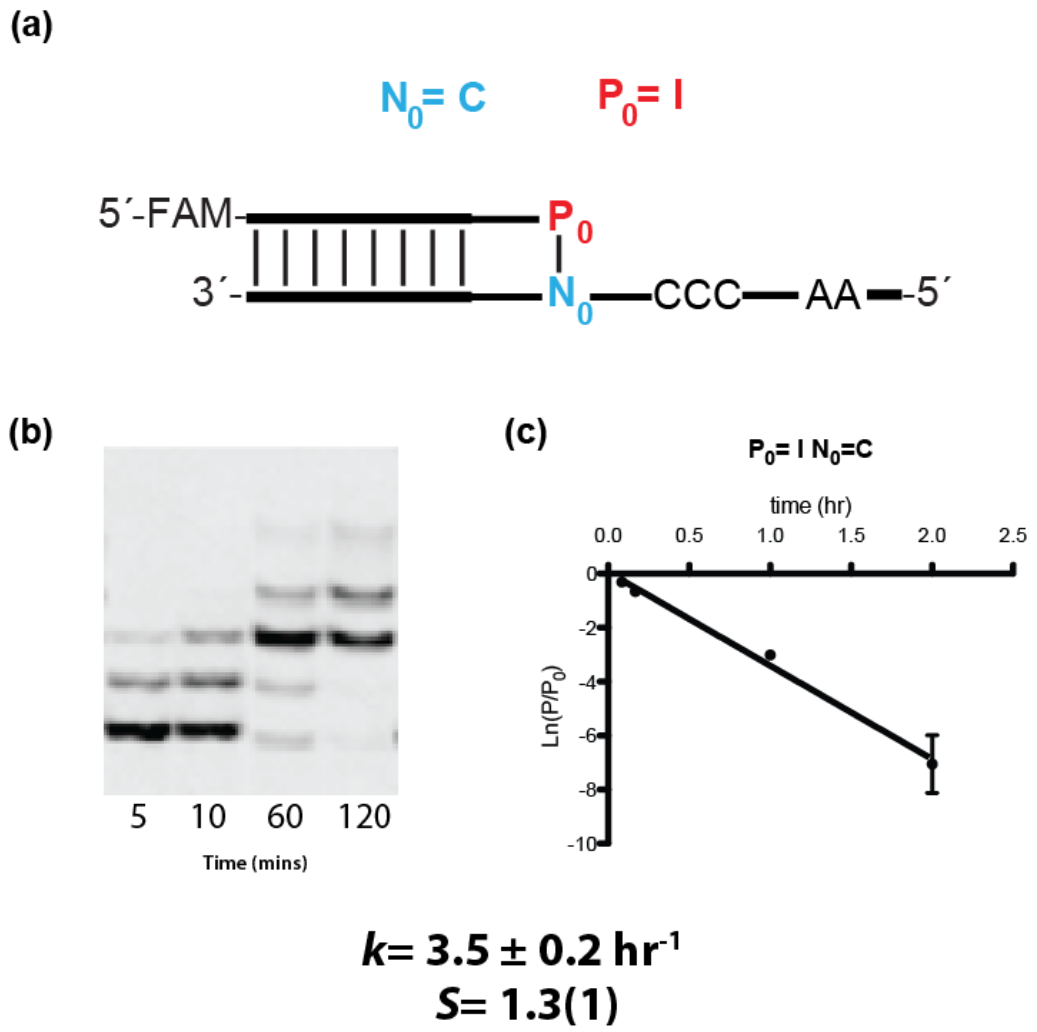


Figure S2.31: Stalling factor for terminal I:C primer template base pair. Primer extension reactions were carried out in triplicate using 20 mM 2AIpG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_0= I, N_0=C. (b) Representative PAGE analysis of result for P_0= I, N_0=C. (c) Plot of ln(P/P_0) as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments. Stalling factor (*S*) was determined by dividing the rate of 2AIpG addition for P_0=G, N_0=C by that of P_0= I, N_0=C ($k_{G:C}$/ $k_{I:C}$).

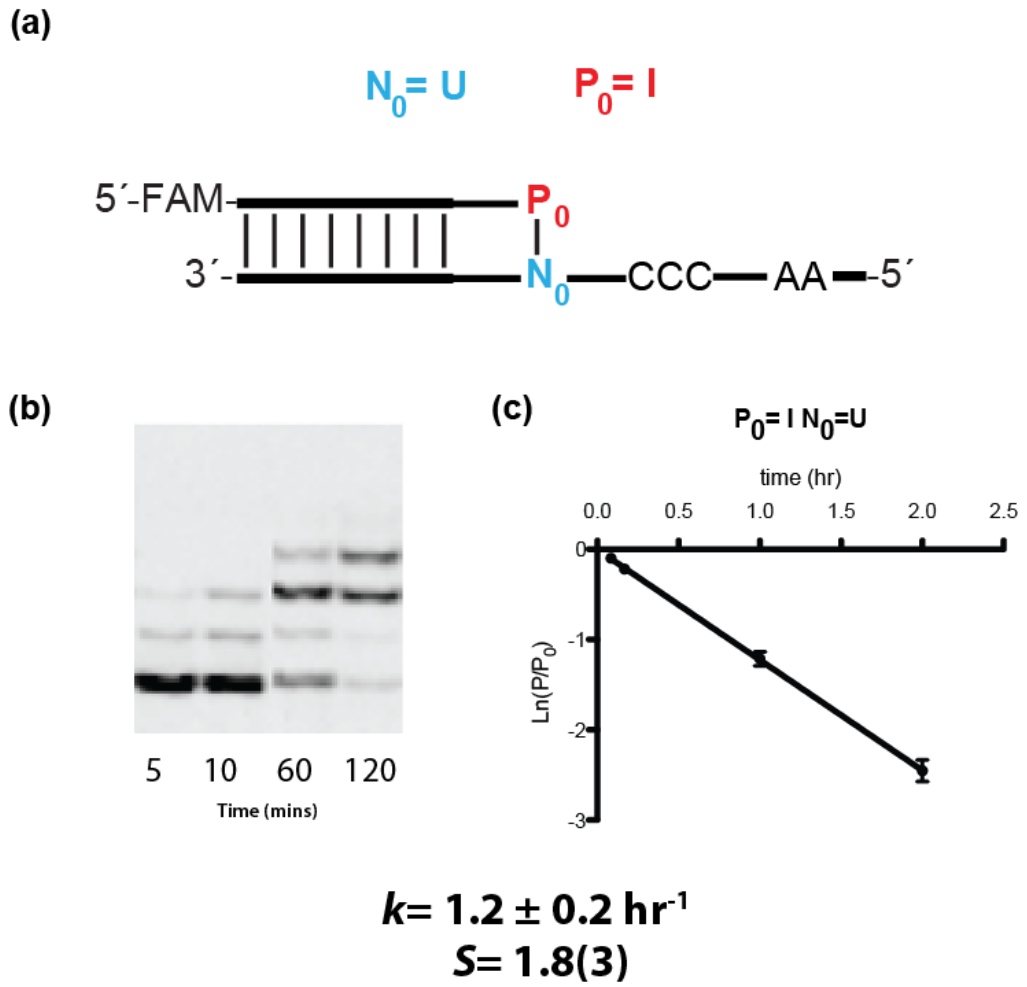


Figure S2.32: Stalling factor for terminal I:U primer template base pair. Primer extension reactions were carried out in triplicate using 20 mM 2AIpG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_0= I, N_0=U. (b) Representative PAGE analysis of result for P_0= I, N_0=U. (c) Plot of ln(P/P_0) as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments. Stalling factor (*S*) was determined by dividing the rate of 2AIpG addition for P_0=A, N_0=U by that of P_0= I, N_0=U ($k_{A:U}$/ $k_{I:U}$).

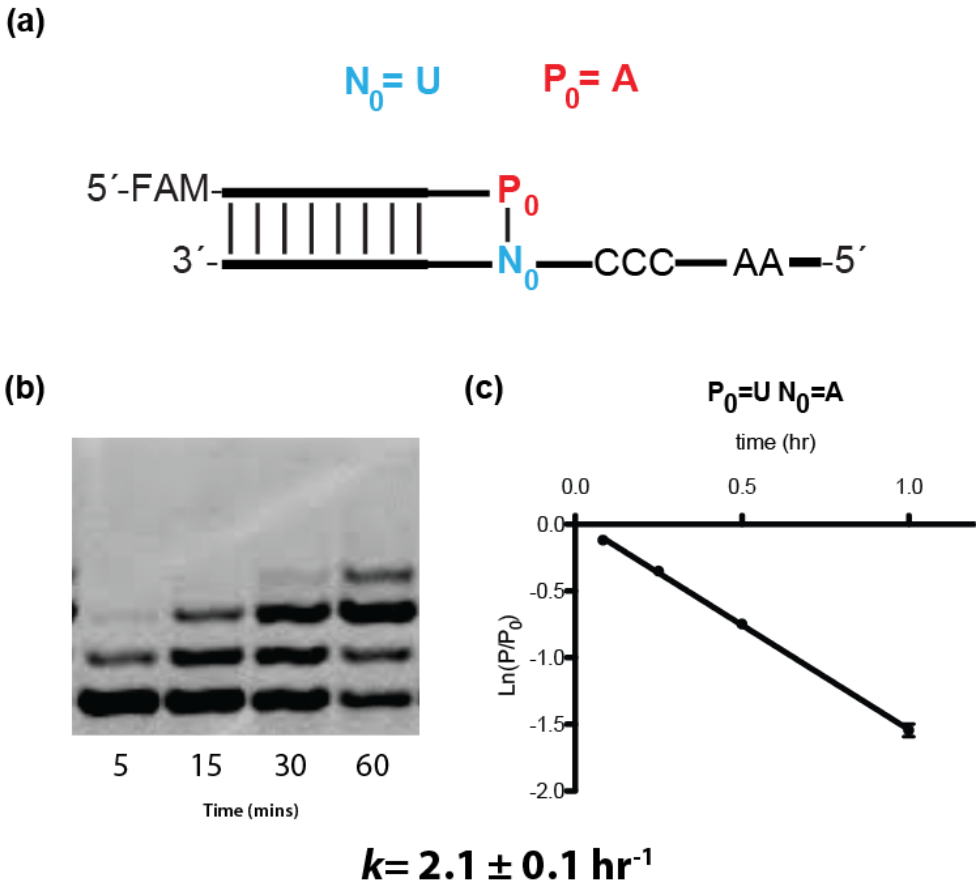


Figure S2.33: Watson-Crick control for the determination of stalling factors. Primer extension reactions were carried out in triplicate using 20 mM 2AIpG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_0= A, N_0=U. (b) Representative PAGE analysis of result for P_0= A, N_0=U. (c) Plot of $\ln(P/P_0)$ as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

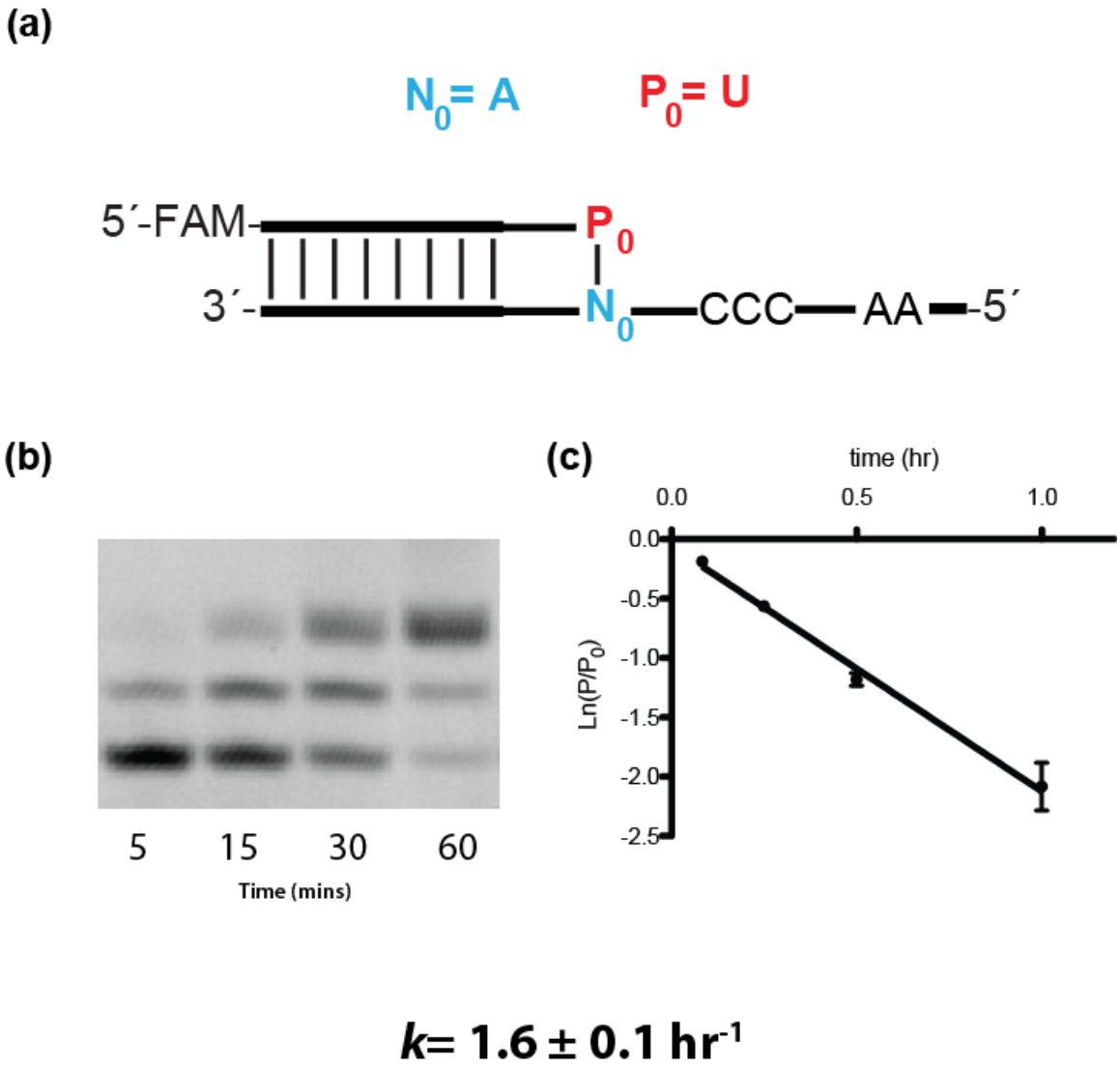


Figure S2.34: Watson-Crick control for the determination of stalling factors. Primer extension reactions were carried out in triplicate using 20 mM 2AIpG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_0= U, N_0=A. (b) Representative PAGE analysis of result for P_0= U, N_0=A. (c) Plot of ln(P/P_0) as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

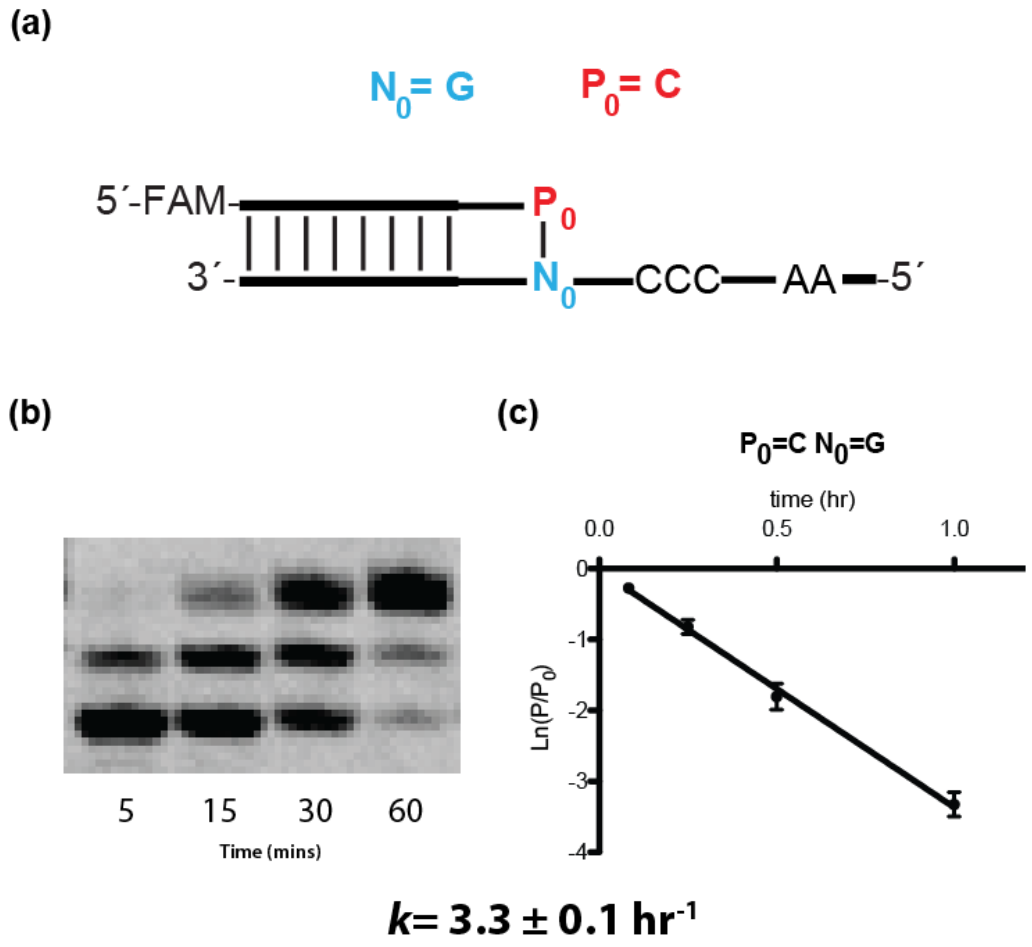


Figure S2.35: Watson-Crick control for the determination of stalling factors. Primer extension reactions were carried out in triplicate using 20 mM 2AIpG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_0= C, N_0=G. (b) Representative PAGE analysis of result for P_0= C, N_0=G. (c) Plot of ln(P/P_0) as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

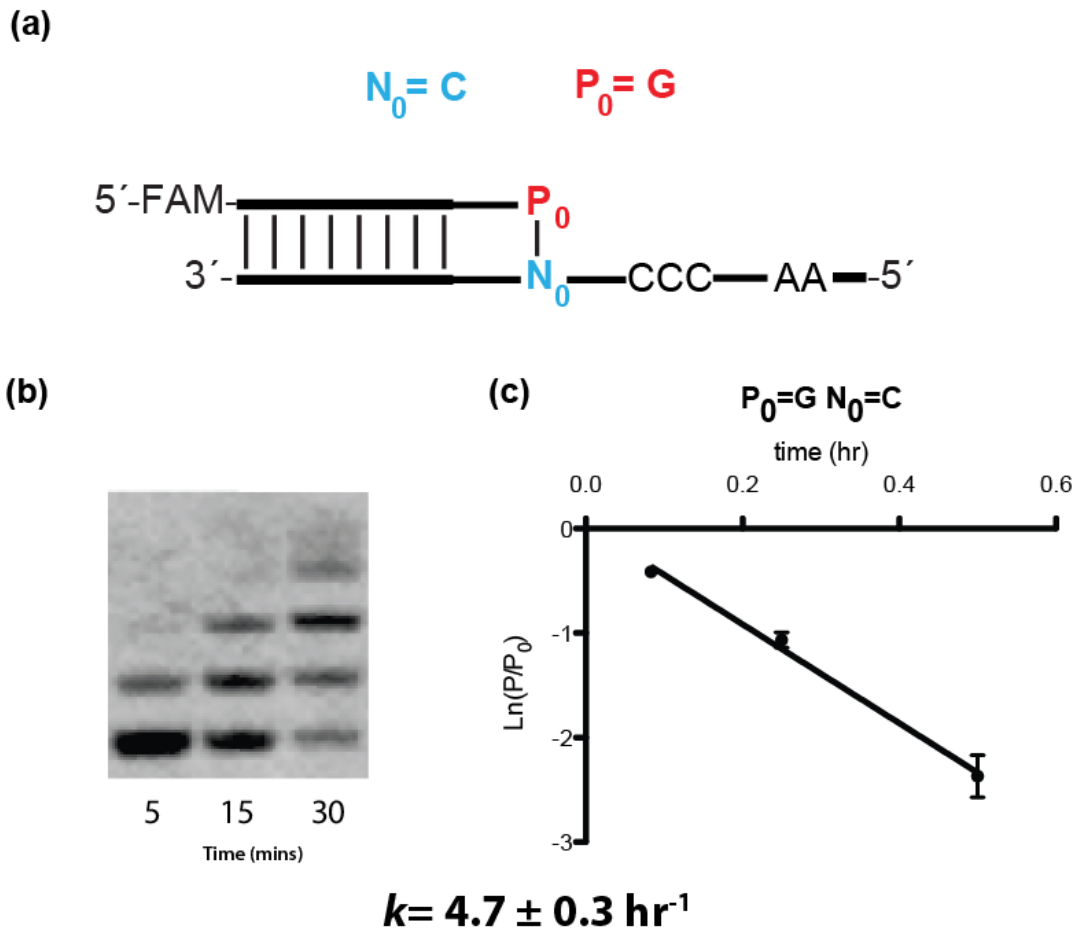


Figure S2.36: Watson-Crick control for the determination of stalling factors. Primer extension reactions were carried out in triplicate using 20 mM 2AIpG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction for P_0= G, N_0=C. (b) Representative PAGE analysis of result for P_0= G, N_0=C. (c) Plot of ln(P/P_0) as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

Chapter 3:

A model for the Emergence of RNA from a Prebiotically Plausible Mixture of arabino-, ribo- and 2′-deoxynucleotides

3.1 Introduction

The RNA world hypothesis proposes that RNA oligonucleotides served dual roles as both genetic and functional polymers during the origin and early evolution of life.[12,21,23] One of the essential requirements for the emergence of the RNA world is the prebiotic availability of ribonucleotides via abiotic synthesis.[25,108109] There has been significant progress in understanding the prebiotic synthesis of the pyrimidine ribonucleotides, led by Sutherland who has demonstrated prebiotically viable syntheses of pyrimidine ribonucleotides (U and C) and their 2-thio variants (*vide infra*).[32] Despite recent progress, equivalently direct and high yielding routes to the purine ribonucleotides remain unknown. Instead, the continued exploration of plausible routes to the canonical ribonucleotides suggests that a diverse set of non-canonical nucleotides may have been present together with the standard ribonucleotides.[91a,b] These non-canonical nucleotides include nucleobase variations such as 8-oxo-purine,[34] inosine[110] and the 2-thio-pyrimidines.[32]

---

[108] Joyce, G. F. The Antiquity of RNA-Based Evolution. *Nature*. **2002**.

[109] Powner, M. W.; Sutherland, J. D.; Szostak, J. W. The Origins of Nucleotides. *Synlett*. **2011**.

[110] 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.* **2018**. *115* (52), 13318-13323.

Intriguingly, some of these variant ribonucleotides have been shown to be compatible with or even beneficial for the copying of RNA templates. For instance, 2-thiouridine greatly improves the rate and fidelity of nonenzymatic primer extension compared with uridine,[90] due to the 2-thio substituent increasing the binding affinity to adenosine and destabilizing wobble binding to guanosine.[111][112][113] Inosine, which can arise from the deamination of adenosine, improves the fidelity of nonenzymatic primer extension relative to guanosine.[110] In contrast, the 8-oxo-purine ribonucleotides are extremely poor substrates for nonenzymatic primer extension.[110] Given the wide range of effects of non-canonical nucleotides on nonenzymatic template directed copying chemistry, a systematic study of the effects of prebiotically plausible variant nucleotides on copying chemistry is needed in order to understand if and how more homogeneous modern RNA could have arisen from complex prebiotic mixtures of nucleotides.

[111] Kumar, R. K.; Davis, D. R. Synthesis and Studies on the Effect of 2-Thiouridine and 4-Thiouridine on Sugar Conformation and RNA Duplex Stability. *Nucleic Acids Res.* **1997**.

[112] Testa, S. M.; Disney, M. D.; Turner, D. H.; Kierzek, R. Thermodynamics of RNA-RNA Duplexes with 2- or 4-Thiouridines: Implications for Antisense Design and Targeting a Group I Intron. *Biochemistry* **1999**.

[113] Larsen, A. T.; Fahrenbach, A. C.; Sheng, J.; Pian, J.; Szostak, J. W. Thermodynamic Insights into 2-Thiouridine-Enhanced RNA Hybridization. *Nucleic Acids Res.* **2015**.

3.1.2. Prebiotic Synthesis of arabino- and deoxyribo- nucleotides

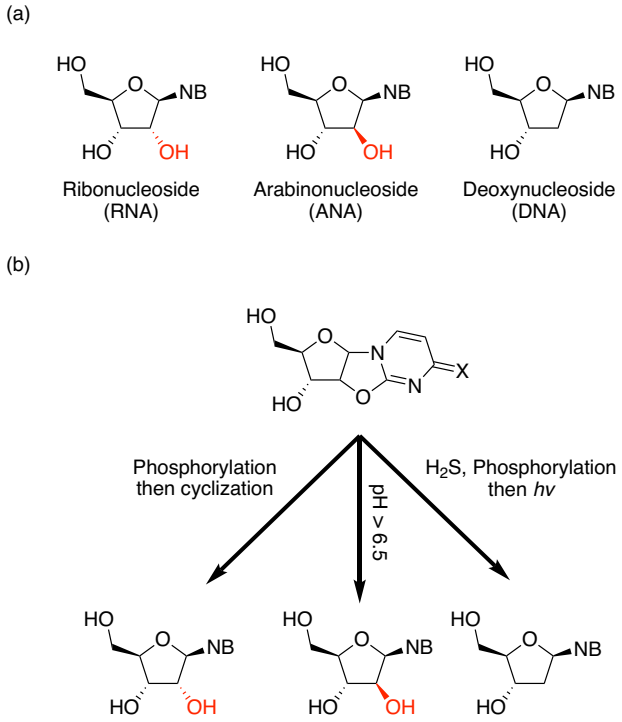


Figure 3.1: Ribo-, arabino-, and 2′-deoxy- nucleosides (a) Schematic representation of ribo- and arabino- nucleotides. NB: nucleobase. (b) Reported divergent syntheses of RNA (left), ANA (top), and DNA (right) from a common intermediate, anhydronucleoside.

Among the most plausible candidates for primordial nucleosides, arabinonucleosides, the building blocks of arabino-nucleic acid (ANA) are of particular interest (Figure 3.1 a). The Powner group has recently demonstrated the first potentially prebiotic synthesis of all four canonical arabinonucleosides.[114] In this approach, the syntheses of pyrimidine arabino- and ribo- nucleotides share the same synthetic intermediate (Figure 1.8. **N2**), with external nucleophiles, such as hydroxide, furnishing the corresponding pyrimidine arabinonucleosides, while intramolecular 3′ phosphate nucleophiles furnishes pyrimidine ribonucleotides. In the case of purine arabinonucleotides, sulfide nucleophiles are utilized, which generate 8-mercaptopurine

[114] Roberts, S. J.; Szabla, R.; Todd, Z. R.; Stairs, S.; Bučar, D.-K.; Šponer, J.; Sasselov, D. D.; Powner, M. W. Selective Prebiotic Conversion of Pyrimidine and Purine Anhydronucleosides into Watson-Crick Base-Pairing Arabino-Furanosyl Nucleosides in Water. *Nat. Commun.* **2018**, *9* (1), 4073.

arabinonucleosides. Upon exposure to ultraviolet light or oxidants such as H_2O_2, desulfurization occurs and furnishes the corresponding purine arabinonucleosides.

Prebiotic syntheses of 2′-deoxynucleotides proceed from 2-thiopyrimidine ribonucleotides,[115] which have in turn been abiotically synthesized from the common intermediate anhydropyrimidine nucleotide. To N2 is added and sulfide, which performs a nucleophilic aromatic substitution (S_NAR) to furnish the α-anomer of 2-thiocytidine ribonucleotide, which is photoanomerized to the desired β-anomer. From the β-2-thiocytidine ribonucleotide, the 2′,3′-cyclic phosphate can be formed by exposure to phosphate buffer and urea. The 2-thio group can then act as a nucleophile to generate 2′,2-thioanhydronucleosides, which then undergo photoredox driven C-S bond homolysis in the presence of HS^- and H_2S, followed by a proposed S-H abstraction to furnish 2′-deoxy-2-thiocytidine nucleotides. Key problems with this route are the relatively poor yield of 33%, for the key last step. Additionally, this synthesis has only been demonstrated with 2-thiopyrimidines, and further elaboration to the canonical purines revealed that unlike the pyrimidine 2-thioribonucleotides, nucleobase hydrolysis to 2-deoxyribose outcompetes nucleobase dethiolation. Attempts to displace the nucleotide with purines, was attempted under forcing conditions with adenine (5 equiv. solid state, 100 °C), yet only 4% is converted to deoxyadenosine.

Contemporaneous with the discovery of the route described above was a route to deoxynucleotides using nucleobases, acetaldehyde and glyceraldehyde as building blocks by the Trapp lab.[116] In this synthetic pathway, nucleobases condense on to acetaldehyde to perform

---

[115] Xu, J.; Green, N. J.; Gibard, C.; Krishnamurthy, R.; Sutherland, J. D. Prebiotic Phosphorylation of 2-Thiouridine Provides Either Nucleotides or DNA Building Blocks via Photoreduction. *Nat. Chem.* **2019**.

[116] Teichert, J. S.; Kruse, F. M.; Trapp, O. Direct Prebiotic Pathway to DNA Nucleosides. *Angew. Chemie - Int. Ed.* **2019**.

enamine aldol reactions on glyceraldehyde, which cyclizes to form the product. While the approach is simple, it suffers from the questionable prebiotic availability of acetaldehyde, as well as the inability to produce the products in high yield (yields are 0.68%, 0.45%, 2.49%, and 0.33% for dA, DG, dC, and dT respectively), while also generating many structurally and configurationally similar byproducts. In conclusion, while much progress has been made with two different syntheses of deoxyribose nucleotides in the past year, neither is satisfactory in producing purines or pyrimidine deoxynucleotides, and currently only the synthesis of 2-thiopyrimidine deoxyribonucleotides could be considered prebiotically viable.

3.1.2. Properties of arabino- and deoxyribo- nucleotides

The divergent syntheses of arabino- and ribo- nucleotides from a common intermediate suggests that the first genetic polymer may have been assembled from a heterogeneous mixture of various nucleotides.[117] Although homogeneous ANA alone cannot form stable duplexes, ANA can form a stable Watson-Crick paired duplex with the corresponding complementary RNA oligonucleotide. Reported melting temperature of the RNA:ANA duplex is significantly lower than that of the RNA:RNA duplex. Furthermore, recent studies have shown that ANA oligonucleotides can have RNA cleaving catalytic activity when paired across from an RNA oligonucleotide substrate.[118] Taken together, these results suggest that ANA oligonucleotides can serve as both genetic and functional polymers in the presence of RNA oligonucleotides.

---

[117] a) Noronha, A. M.; Wilds, C. J.; Lok, C. N.; Viazovkina, K.; Arion, D.; Parniak, M. A.; Damha, M. J. Synthesis and Biophysical Properties of Arabinonucleic Acids (ANA): Circular Dichroic Spectra, Melting Temperatures, and Ribonuclease H Susceptibility of ANA·RNA Hybrid Duplexes. *Biochemistry* **2000**. b) Li, F.; Sarkhel, S.; Wilds, C. J.; Wawrzak, Z.; Prakash, T. P.; Manoharan, M.; Egli, M. 2′-Fluoroarabino- and Arabinonucleic Acid Show Different Conformations, Resulting in Deviating RNA Affinities and Processing of Their Heteroduplexes with RNA by RNase H. *Biochemistry* **2006**.

[118] Taylor, A. I.; Pinheiro, V. B.; Smola, M. J.; Morgunov, A. S.; Peak-Chew, S.; Cozens, C.; Weeks, K. M.; Herdewijn, P.; Holliger, P. Catalysts from Synthetic Genetic Polymers. *Nature* **2015**.

Herein, we evaluate the effects of arabinonucleotides on nonenzymatic primer extension, both as activated monomers and when incorporated into primers and templates. We have also revisited the effects of 2′-deoxyribonucleotides in parallel experiments, using more modern and prebiotically plausible activation chemistry than in earlier studies. We aim to understand the behavior of heterogeneous chimeric, predominantly ribonucleotide oligomers, by systematically introducing single variant residues to the RNA system.

3.2 Results and Discussion

3.2.1 Primer Extension with Monomers

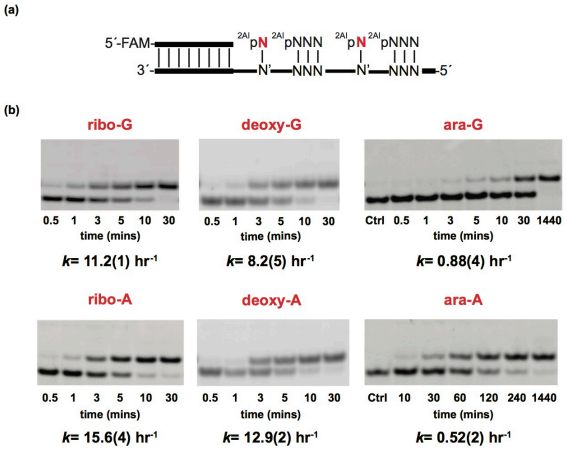


Figure 3.2: Evaluation of arabino- and 2′-deoxy- nucleotides in nonenzymatic primer extension. (a) Schematic representation of a primer extension reaction. 2AIpN represents 2-aminoimidazole activated monomers and 2AIpNNN represents 2-aminoimidazole activated trimer helpers. (b) Gel electrophoresis images and rates of primer extension for 2-aminoimidazole activated ribonucleotides and arabinonucleotides. All reactions were performed at pH 8.0, 200 mM HEPES, 50 mM Mg^{2+}, 20 mM $2AIpP_1$, 0.5 mM of activated trimer. See SI appendix for kinetic analysis of primer extension reactions.

To examine arabino- and deoxyribo- nucleotides as substrates in nonenzymatic primer extension, we measured the rate of nonenzymatic primer extension using the 2-aminoimidazole activated nucleotide monomers, 2AI-araG, 2AI-araA, 2AI-dG, and 2AI-dA, across from their respective Watson-Crick pairing partners C and 2^sU in RNA templates (Figure 3.2 a). We used 2-thio-uridine (2^sU) instead of uridine in the template because 2^sU is prebiotically plausible and

exhibits stronger pairing with A than U. Previous reports from our lab have demonstrated that activated downstream oligonucleotides can greatly accelerate nonenzymatic primer extensions[75] by both catalyzing the formation of highly reactive imidazolium-bridged intermediates[74a] and because of the structural preorganization afforded by an extended helical geometry.[76] Therefore, we prepared 2-aminoimidazole activated RNA trimer helpers and utilized these activated downstream helpers in the evaluation of arabino- and deoxyribo- nucleotides in nonenzymatic primer extension. We also prepared the corresponding ribonucleotides and directly compared the rates of nonenzymatic primer extension for all three classes of nucleotides. The rates of primer extension using activated ribonucleotides were the fastest, at 11 hr^{-1} for 2AI-rG and 16 hr^{-1} for 2AI-rA. Primer extension rates with activated 2′-deoxy- nucleotides were slightly slower than with ribonucleotide monomers (8.2 hr^{-1} and 13 hr^{-1} for 2AI-dG and 2AI-dA respectively). In contrast, primer extension with arabinonucleotide monomers was much slower, at 0.88 hr^{-1} for 2AI-araG and 0.52 hr^{-1} for 2AI-araA (Figure 3.2 b).

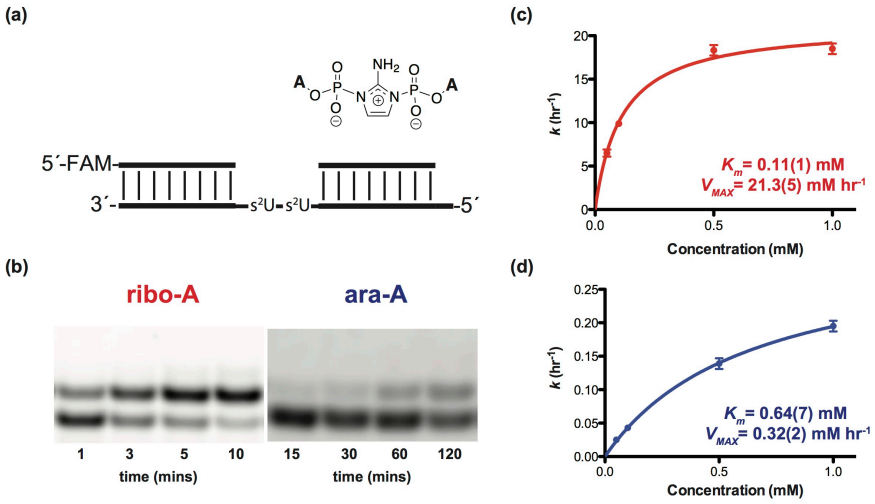


Figure 3.3: Michaelis-Menten study of the nonenzymatic copying rates of rA, dA, and araA imidazolium bridged dimers**.** All reactions were performed at pH 8.0, 200 mM HEPES, 50 mM Mg^{2+}, and various concentrations of dimers ranging from 0.05 to 1 mM. (a) Schematic representation of a primer extension reaction with imidazolium bridged dimers. (b) Gel electrophoresis images of primer extension for imidazolium bridged dimers at 1 mM

concentration. (c) Plot of k (hr^{-1}) against the concentration of rA dimer. (d) Plot of k (hr^{-1}) against the concentration of rA dimer. See SI appendix for kinetic analysis of primer extension reactions.

In order to distinguish between effects due to the formation, degradation or steady state levels of the imidazolium-bridged intermediate, as opposed to the reactivity of that intermediate once formed, we prepared the 2-aminoimidazolium-bridged dinucleotides from ribo-, 2′-deoxy- and arabino-adenosine. We measured the rates of hydrolysis of the three imidazolium bridged dimers under primer extension conditions (50 mM Mg^{2+}, pH 8 in 10% D_2O). Measured rates of hydrolysis revealed no significant difference in the stability of activated dimer species with k= 0.31 hr^{-1} and 0.33 hr^{-1} for rA dimer and araA dimer respectively, showing that the stereochemical information at the 2′ position does not impact the decomposition rates of activated species under relevant conditions.

We then evaluated the Michaelis-Menten parameters for primer extension for all three imidazolium-bridged dinucleotides. For these experiments, we prepared RNA templates containing two consecutive s^2U residues by solid phase synthesis; we also used downstream oligonucleotides to further promote the nonenzymatic copying (Figure 3.3 a). The K_m values of the imidazolium-bridged dimer intermediates are moderately different, at 0.11 mM and 0.64 mM for rA and araA respectively (Figure 3.3 c,d), in line with observations for RNA:RNA vs. ANA:RNA duplexes.[119] However, V_{max} for the two dimer intermediates differ by approximately 70 fold, at 21 mM hr^{-1} and 0.32 mM hr^{-1} for rA and araA respectively (Figure 3.3 c,d). The large difference in V_{max} is surprising, since the 2′-hydroxyl is remote from the site of the reaction between the primer 3′-hydroxyl and the phosphate of the adjacent incoming nucleotide. Furthermore, simple nucleophiles such as hydroxide or water attack the phosphate of both

[119] Watts, J. K.; Martín-Pintado, N.; Gómez-Pinto, I.; Schwartzentruber, J.; Portella, G.; Orozco, M.; González, C.; Damha, M. J. Differential Stability of 2′F-ANA•RNA and ANA•RNA Hybrid Duplexes: Roles of Structure, Pseudohydrogen Bonding, Hydration, Ion Uptake and Flexibility. *Nucleic Acids Res.* **2010**, *38* (7), 2498–2511.

imidazolium-bridged dimers with similar rates, indicating that the 2′ hydroxyl influences the rate of phosphate substitution only with stereochemically rich nucleophiles, such as the ribose at the end of the primer. Arabino-nucleotides have been shown to adopt a C2′ endo pucker with intramolecular hydrogen bonding between the 2′-hydroxyl and 5′-oxygen,[120] while ribonucleotides adopt a C3′ endo pucker. This difference in sugar conformation could affect the position of the 5′-phosphate of the nucleotide adjacent to the 3’-end of the primer, potentially explaining the observed rate differences.

3.2.2 Effect of arabino- and 2′-deoxy-nucleotides at the 3′-End of the primer

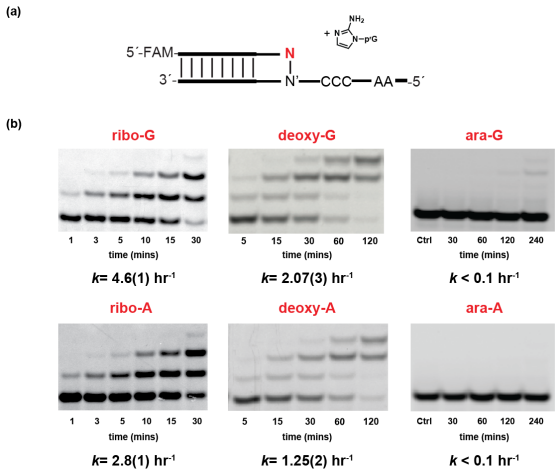
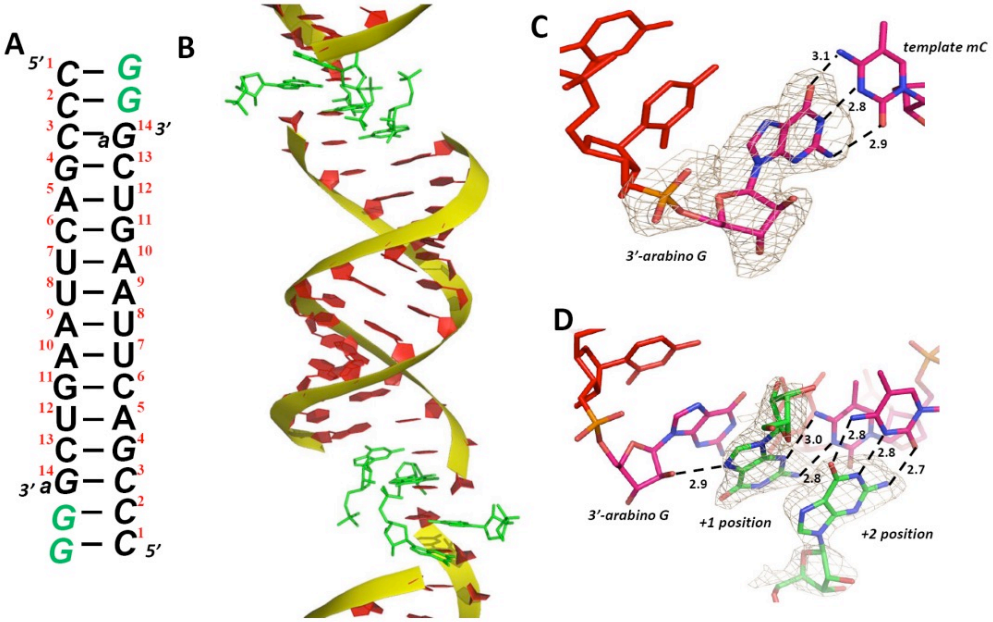


Figure 3.4. Evaluation of nonenzymatic primer extension with various pentose nucleotides terminated primers. (a) Schematic representation of nonenzymatic primer extension with a primer that contains an arabinonucleotide (P_0) at the 3′ end. (b) Gel electrophoresis images and rates of primer extension for primers containing ribose and arabinose nucleotides in the 3′ end. See SI appendix for kinetic analysis of primer extension reactions.

Previous studies indicate that the nature of the 3′-end of the primer can strongly influence the subsequent nonenzymatic primer copying; up to a 300 fold decrease in the primer extension rate has been observed for a mismatched base pair vs. a Watson-Crick base pair at the 3′-end of

[120] Venkateswarlu, D.; Ferguson, D. M. Effects of C2’-Substitution on Arabinonucleic Acid Structure and Conformation. *J. Am. Chem. Soc.* **1999**, *121* (23), 5609–5610.

the primer.[110,121] We therefore evaluated the effect on nonenzymatic primer extension of replacing the ribonucleotide at the 3′-end of the primer with either a 2′-deoxynucleotide or an arabinonucleotide. We prepared the three primers by solid phase synthesis, and then measured the rate of primer extension in the presence of 2-aminoimidazole activated guanosine monomers using templates containing the Watson-Crick paired primer binding region followed by a C_3 templating region (Figure 3.4 a). Strikingly, primers containing a single arabinonucleotide at the 3′-end showed no significant extension even after 4 hours with less than 1% yield. In contrast, primers with ribo- or 2′-deoxy-nucleotides at the 3′-end were extended at high rates under the same reaction conditions (k_{obs}= 4.6 hr^{-1}, 2.8 hr^{-1}, 2.1 hr^{-1}, and 1.3 hr^{-1} for primers containing a 3′ terminal rG, rA, dG, and dA respectively), suggesting that the change in the configuration of the 2′ hydroxyl group on the pentose at the 3′ end of primer plays a significant role in nonenzymatic primer extension.



[121] Pal, A.; Das, R. S.; Zhang, W.; Lang, M.; McLaughlin, L. W.; Szostak, J. W. Effect of Terminal 3′-Hydroxymethyl Modification of an RNA Primer on Nonenzymatic Primer Extension. *Chem. Commun.* **2016**.

Figure 3.5: Crystal structure of RNA primer/template complex with an arabinonucleotide at the 3'-end of the primer. Red: RNA and ANA backbone; green: GMP monomers. The wheat mesh indicates the $2F_o - F_c$ omit map contoured at 1.5σ. (A) Schematic of the primer/template complex. Italic letters: locked nucleotides; *aG* at the 3′ end: arabino guanosine primer. (B) Overall structure of the arabino RNA-GMP complex, stacking with the neighboring complexes. (C) The local structure of the arabino-G-terminated primer. The corresponding omit map indicates that the primer arabinose sugar is in C2′-endo conformation, and the guanine is Watson-Crick base paired with the template. (D) The GMP monomers are bound through both canonical and non-canonical G:C base pairs. The primer 2′-OH and N7 of the +1 GMP monomer are within hydrogen bonding distance.

Given the dramatic stalling effect of a terminal ANA residue at the 3′ end of the primer, we sought to better understand why ANA at the 3′ end decelerates primer extension at the atomic level. Thus, we crystallized a chimeric oligonucleotide with ANA at the 3′ end, complexed with RNA-monomers. The rGMP monomer was co-crystallized with a self-complementary 14mer RNA 5′-**^{m}C^{m}C^{m}CG**ACUUAAGUC^{ara}G-3′, which contains 4 locked nucleotides at the 5′ end (in bold) to rigidify the template and speed up crystallization (Figure 3.5 a). The residue at the 3′ end is araG (^{ara}G), to mimic the nonenzymatic primer extension shown in Fig 4. At each end, two GMP nucleotides bind to the templating locked 5-methyl C consecutively, forming a 16mer RNA-monomer helical structure. The assembled complex crystallizes in the rhombohedral R32 space group, and the structure is determined to 2.0-Å resolution. In one asymmetric unit, there is one strand of RNA complexed with 2 GMP monomers. Except for the crystallographic symmetry, this ANA-modified RNA complex structure resembles our previously determined RNA-GMP complex structures,[73a,122] including the A-form double helices, C3′-endo conformation of the RNA sugars and slip-stacking of the neighboring duplexes (Figure 3.5 b). The crystal structure reveals two key features about the araG residue: the araG residue is in the C2′-endo conformation, and forms a Watson-Crick base pair with the templating **^{m}C**. The sugar of the terminal araG is well-ordered,

[122] Zhang, W.; Walton, T.; Li, L.; Szostak, J. W. Crystallographic Observation of Nonenzymatic RNA Primer Extension. *Elife* **2018**.

and its C2′-endo conformation contrasts with reported structures of an RNA/ANA heteroduplex, in which the arabinonucleotides are in the O4′-endo (Eastern) conformation (Figure 3.5 c).[117a,119,123] The nucleobase of the arabino-G residue at the 3′-end of our primer is Watson-Crick base paired with the templating **^{m}C**, with hydrogen bonds ranging from 2.8 to 3.1 Å .

In the evaluation of various nucleotides in nonenzymatic primer extension, several prior reports have shown a strong correlation between primer extension rates and the conformation of nucleotides, where primers composed of, or ending in nucleotides (such as RNA, LNA, and hexitol nucleic acid) that prefer C3′-endo conformations (A-form) engage in faster primer extension than primers composed of, or ending in nucleotides that do not adapt C3′-endo conformations (e.g. DNA).[124] Additionally, pseudoaxial/axial nucleophiles generally exhibit lower reactivities than pseudoequatorial/equatorial nucleophiles. [125] This is due to steric encumbrance of axial/pseudoaxial substituents, especially by 1,3-diaxial interactions, which are exacerbated by an increased steric environment in the product and transition state leading to product. One plausible explanation for the significant stalling effect we observe in nonenzymatic polymerization with an arabinonucleotide terminated primer is due to the preferred C2′- conformation of the arabino-nucleotide.

---

[123] Damha, M. J.; Noronha, A. M.; Wilds, C. J.; Trempe, J. F.; Denisov, A.; Pon, R. T.; Gehring, K. Properties of Arabinonucleic Acids (ANA & 20′F-ANA): Implications for the Design of Antisense Therapeutics That Invoke RNase H Cleavage of RNA. *Nucleosides, Nucleotides and Nucleic Acids* **2001**.

[124]a) Kozlov, I. A.; Politis, P. K.; Van Aerschot, A.; Busson, R.; Herdewijn, P.; Orgel, L. E. Nonenzymatic Synthesis of RNA and DNA Oligomers on Hexitol Nucleic Acid Templates: The Importance of the a Structure. *J. Am. Chem. Soc.* **1999**. b) Schrum, J. P.; Ricardo, A.; Krishnamurthy, M.; Blain, J. C.; Szostak, J. W. Efficient and Rapid Template-Directed Nucleic Acid Copying Using 2′-Amino-2′,3′-Dideoxyribonucleoside-5′-Phosphorimidazolide Monomers. *J. Am. Chem. Soc.* **2009**. c) 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**.

[125] Winstein, S.; Holness, N. J. Neighboring Carbon and Hydrogen. XIX. t-Butylcyclohexyl Derivatives. Quantitative Conformational Analysis. *J. Am. Chem. Soc.* **1955**.

The base-pairing of the unactivated GMP monomer bound downstream of the primer is significantly different from that observed in our previously determined RNA-GMP structures. A hydrogen bonding interaction between the N7 atom of the +1 GMP and the 2′-OH group of the arabinonucleotide at the end of the primer is observed (2.9 Å, Figure 3.5 d). This flips the nucleobase of the +1 GMP into a non-canonical G:C base pair in which the guanine N3 of the +1 GMP forms a hydrogen bond with the exocyclic amine of the templating **^{m}C**, while the exocyclic amine of the guanine hydrogen bonds with N3 of the **^{m}C**. This non-canonical base pairing positions the monomer 5′-phosphate distant from the 3′-OH of the primer. The phosphate of +1 GMP is highly disordered, and the primer 3′-OH of the arabinose sugar is pointing to the minor groove instead of the incoming monomer. In previously determined RNA-GMP structures, the first bound GMP monomer was Watson-Crick paired with the template, with its binding affinity and the distance between primer 3′-OH and the incoming phosphorus improved by the downstream bound monomer at +2 position. These observed structural changes would be detrimental to nonenzymatic primer extension, in line with our experimental data; however the crystal structure may not reflect the conformation of the template bound imidazolium-bridged dinucleotide.

3.3.3 Competition Experiment between Activated rA and araA

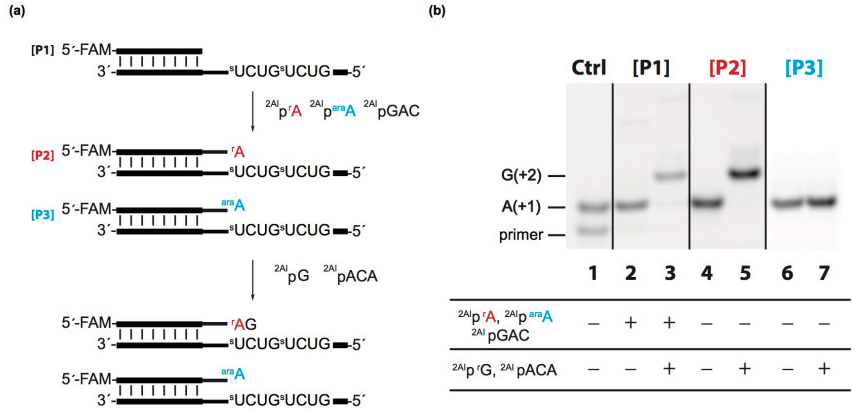


Figure 3.6: Competition experiment between 2AI-araA and 2AI-rA. (a) Schematic representation of nonenzymatic primer extension with both ribonucleotides and arabinonucleotides monomers then a subsequent nonenzymatic copying, in which we expect [P2] to extend and [P3] to not extend (shown in lanes 5 and 7 in b) to determine the incorporation ratio. (b) Gel electrophoresis images of competition experiment. Primers used for the experiment are shown on the top ([P1] for lane 2/3, [P2] for lane 4/5, and [P3] for lane 6/7) and the activated nucleotide monomers and trimer helpers added are shown below. P2 and P3 synthetic primers used as controls.

The common intermediate in the prebiotic synthesis of RNA and ANA implies that the two may have coexisted on early earth. Given the strong stalling effect when the 3′ end of the primer is an arabinonucleotide, the incorporation of an arabinonucleotide into a growing primer will essentially act as a chain terminator, potentially preventing the synthesis of oligonucleotides long enough to have catalytic function. Fortunately, arabino-nucleotide incorporation rates were 15-30 fold slower than ribonucleotide incorporation rates in the case of trimer assisted nonenzymatic copying, suggesting that ribonucleotides may outcompete arabinonucleotides well enough to minimize this concern. To test this idea, we performed a nonenzymatic copying reaction in a one pot competition experiment with a 1:1 ratio of activated araA and rA monomers, utilizing the reaction conditions in Figure 3.2. The initial primer, P1, was exposed to activated rA and araA monomers with activated trimer helper (2AI-rGAC) for one hour, to allow for incorporation of either rA (generating primer P2) or araA (generating primer P3). A subsequent nonenzymatic

copying step was then performed by adding the activated RNA monomer 2AI-rG and the trimer helper 2AI-rACA. Primers that initially incorporated araA to generate P3 will not be extended in the second step (Lane 7, Figure 3.6 b) while primers that initially incorporated rA to generate primer P2 will be further extended (Lane 5, Figure 3.6 b). After 4 hours, the reaction was quenched and PAGE gel analysis was performed to analyze the product ratios (Figure 3.6 b). Only a small fraction (approximately 7%) does not extend beyond the initial addition (Lane 3, Figure 3.6 b). Generally, iterative nonenzymatic copying of oligonucleotides suffers some level of rate deceleration and stalling,[68,75] presumably due to the presence of multiple trimer helpers, which have high binding affinity to the template and are thus inhibitory to subsequent copying. Our results therefore represent a lower bound on the true ratio of rA vs. araA incorporation, and indicate that activated rA monomers substantially outcompete activated araA in the nonenzymatic copying of oligonucleotides. Thus, the coexistence of arabino- and ribo- nucleotides should not significantly interfere with oligonucleotide copying.

3.3.4 Primer Extension Across Arabinonucleotides in the Template Strand

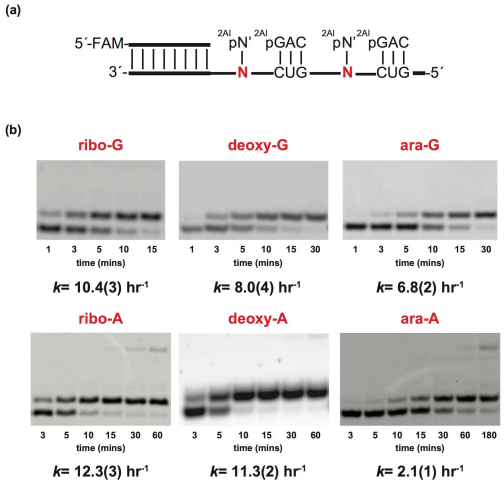


Figure 3.7: Evaluation of primer extension for 2-aminoimidzole activated pyrimidine nucleotides across arabinose nucleotides. (a) Schematic representation of a primer extension reaction with template containing arabinose nucleotides. 2AIpN represents the corresponding 2-aminoimidazole activated ribonucleotide monomers and 2AIpNNN represents 2-aminoimidazole

activated trimer helper. (b) Gel electrophoresis images and rates of primer extension for 2-aminoimidazole activated pyrimidine ribose across arabinose nucleotides. See SI appendix for kinetic analysis of primer extension reactions.

Although the above results suggest that arabinonucleotide-containing oligonucleotides would only rarely be generated by nonenzymatic primer extension, untemplated polymerization of activated monomers could lead to the synthesis of oligonucleotide templates containing arabinonucleotides in internal positions.[36a,c] Therefore, we have evaluated the effects of internal arabinonucleotides in template strands on the rate of primer extension with standard activated ribonucleotides. We prepared templates containing mixed RNA and ANA sequences by solid phase synthesis, and then performed nonenzymatic primer extensions with activated RNA monomers and trimer helpers, and compared the observed rates with the corresponding rates measured on all-RNA templates and RNA templates containing single 2′-deoxyribonucleotide residues (Fig 7a). Intriguingly, primer extension with 2-aminoimidazole activated cytidine monomer exhibited a minimal rate difference when copying across rG, dG, and araG in the template strand, with k_{obs}= 10, 8.0, and 6.8 hr^{-1} (Figure 3.7 b). Primer extension with activated 2-thiouridine monomer exhibited a minor rate difference when copying across rA, dA, and araA in the template strand, with k_{obs}= 12, 11, and 2.1 hr^{-1},respectively. For both arabino- and deoxyribo-nucleotide containing templates, we observe complete conversion of the primer to extended products in the presence of the corresponding complementary Watson-Crick RNA monomers.

The minimal effect of an internal arabinonucleotide in the template strand is in marked contrast to the much slower rate of addition of activated arabinonucleotides to the 3′ end of a primer, and the strongly detrimental effect of a terminal arabinonucleotide on continued primer extension. The large difference may stem from the fact that when arabinonucleotides are capping or extending on the primer, the sugar conformation can significantly increase the distance from the

primer 3′-hydroxyl to the adjacent reactive phosphate. In contrast, arabinonucleotides in the template are further from the reaction center, so unless the conformation of the template is drastically affected, the effect on primer extension may be negligible. We suggest that the faster rates of primer extension with ribonucleotides than with arabinonucleotides (and to a lesser extent 2′-deoxyribonucleotides) could lead to the formation of the RNA enriched oligonucleotides from initially chimeric RNA/ANA/DNA oligonucleotides formed via untemplated polymerization.

3.4 Conclusion

Our study demonstrates that arabinonucleotides behave very differently when compared with the corresponding ribonucleotides and deoxyribonucleotides in nonenzymatic primer extension. Activated arabinonucleotide monomers show up to 30-fold slower rates of copying when compared to ribo- and deoxyribo- nucleotides, and, furthermore, subsequent nonenzymatic primer extension is strongly inhibited once an arabinonucleotide is incorporated. Our kinetic and structural studies are consistent with this rate difference being due to the C2′-endo sugar conformation of arabinonucleotides. Although deoxyribonucleotides also prefer a C2′-endo sugar conformation, they have a low kinetic barrier to rapid interconversion between C2′-endo and C3′-endo. In contrast, chimeric oligonucleotides composed of mixed pentose-nucleotides can serve as effective templates for copying by primer extension with ribonucleotides. Given the common intermediate in the prebiotic syntheses of ribo-, deoxyribo-, and arabino-nucleotides, it seems likely that ANA, DNA, and RNA would all have been present on the early earth. The earliest oligonucleotides would likely have been formed via nonselective untemplated polymerization of heterogeneous nucleotides, resulting in chimeric oligonucleotides. Our results suggest that such chimeric oligonucleotides could have served as initial genetic polymers, which via chemical selection that preferentially copies deoxyribo- and ribo- nucleotides, would phase out ANA,

allowing a pathway to the modern biological system that stores and transfers genetic information in DNA and RNA.

Furthermore, the difference in observed rates and the one pot competition experiment between arabino- and ribonucleotides demonstrate a potential mechanism by which highly RNA enriched oligonucleotides could arise from a heterogeneous mixture of RNA and ANA. Once an arabinonucleotide is added to a primer, subsequent copying will be much slower, implying that arabino-terminated oligonucleotides are likely to have shorter lengths and therefore lower melting temperatures than purely RNA oligonucleotides. These short mixed oligonucleotide sequences could have beneficial roles in RNA replication, serving as downstream helper sequences or short, inert templates with low melting temperatures for RNA polymerization or ligation. Therefore, long, functional oligonucleotides are much more likely to contain RNA, and, after a few rounds of replication, RNA oligonucleotides could emerge from an initial heterogeneous mixture of ANA and RNA, providing a pathway to modern RNA.

3.5 Outlook: Progress towards the prebiotic synthesis of ribonucleotides

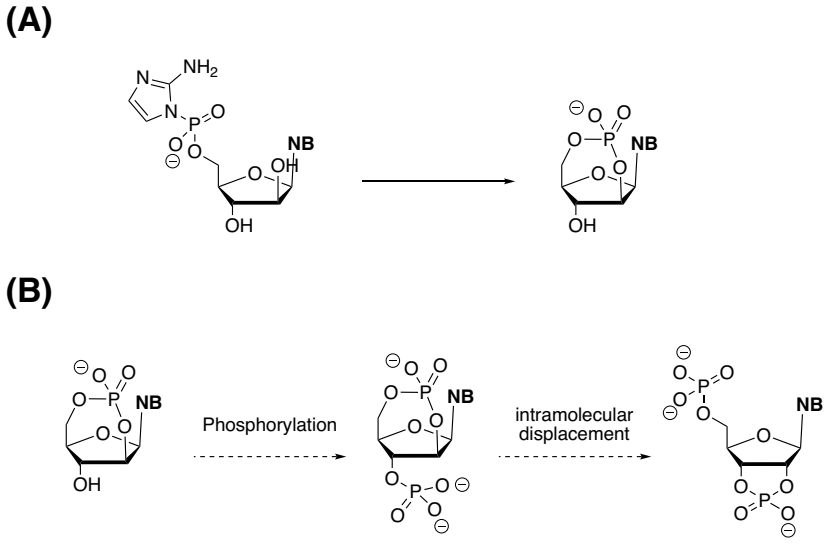


Figure 3.8: Proposed prebiotic syntheses of purine ribonucleotides from arabinonucleotides. (A) Cyclization observed during the synthesis of 2-aminoimidazole activated arabinonucleotides. (B) Plausible pathway of converting cyclic arabinonucleotide byproducts to purine ribonucleotides.

During the synthesis of activated ANA monomers, we observed formation of an arabinonucleotide cyclic phosphate (Figure 3.8 a). Due to the unique structure of arabinonucleosides, which contain trans- 2′,3′ hydroxyl groups, 2-aminoimidazole activated arabinonucleotides generate the 2′,5′- cyclic phosphate, which also has been observed by different groups previously.[126] We envision that this byproduct could serve as an intermediate en route to the synthesis of purine ribonucleotides, the key remaining challenge in the first step of RNA world hypothesis. Specifically, we envision that by employing the same strategy as Orgel, phosphorylation of the cyclic phosphate at the 3′-hydroxyl group followed by intramolecular attack of phosphate to displace the cyclic phosphate could synthesize purine ribonucleotides (Figure 3.8 b). We propose that the overall reaction would be thermodynamically favorable, due to the release of ring strain from the unusual 7-membered ring cyclic phosphate, and the increase of entropic energy from the corresponding glycosyl bond that will have much higher degree of freedom upon the removal of the adjacent, sterically hindering cyclic phosphate. Preliminary ground state enthalpic computations of our proposed product and starting material shows promising results, with the proposed energy difference of -2.87 kcal/mol (B3LYP/6-31G(d,p), polarization continuum model with water dielectric). This proposed route takes inspiration from Sutherland and Powner, by relying on previously reported prebiotically plausible compounds and routes, while also leveraging the greater insight and knowledge we have about the constraints of prebiotically plausible syntheses. Finishing the abiotic synthesis of purine ribonucleotides would be placing the final piece into a puzzle that has been slowly assembled over the last 60 years, which

[126] Harada, K.; Orgel, L. E. The Cyclization of Arabinosyladenine-5′-Phosphorimidazolide. *J. Mol. Evol.* **1991**.

would finally clear the first of Orgel's subproblems. Thus, our current efforts are devoted to completing the synthesis of purine ribonucleotides from an arabinonucleotide cyclic phosphate.

3.6. Materials and Methods

3.6.1 Oligonucleotide Synthesis

All oligonucleotides used in this study were purchased from Integrated DNA Technologies (Coralville, IA) or prepared by solid-phase synthesis using an Expedite 8909 DNA/RNA synthesizer. Synthesizer reagents and phosphoramidites were purchased from Glen Research (Sterling, VA) and Chemgenes (Wilmington, MA). In-house prepared oligonucleotides were deprotected using standard methods and subsequently purified by polyacrylamide gel electrophoresis. The oligonucleotides were analyzed by high resolution mass spectrometry (HRMS) on an Agilent 6520 QTOF LC-MS.

3.6.2 Synthesis of activated nucleotides

Adenosine 5′-phosphoro-2-aminoimidazolide, cytidine 5′-phosphoro-2- aminoimidazolide, guanosine 5′-phosphoro-2- aminoimidazolide, and 2-thio-uridine 5′-phosphoro-2-aminoimidazolide were prepared according to a previously reported procedure. 8-oxo-adenosine, 8-oxo-inosine and 8-oxo-guanosine were prepared according to reported procedures and ^{1}H NMR spectra and mass match those in the literature. GAC and AGG 5′-phosphoro-2- aminoimidazolide were prepared as described previously. All nucleotides were purified using reverse phase chromatography using a Teledyne Isco Combiflash on a RediSepRf C18Aq column with 2 mM triethylammonium bicarbonate and acetonitrile as eluents.

Arabinofuranosyladenine 5′-phosphoro-2-aminoimidazolide (2AIparaA): ^{1}H NMR (400 MHz, Deuterium Oxide) δ 8.17 (s, 1H), 8.05 (s, 1H), 6.69 (t, J = 1.7 Hz, 1H), 6.46 (t, J = 2.1 Hz, 1H),

6.37 (d, J = 6.0 Hz, 1H), 4.59 (d, J = 6.2 Hz, 1H), 4.42 (t, J = 6.6 Hz, 1H), 4.22 – 4.14 (m, 2H), 4.14 – 4.08 (m, 1H). ^{31}P NMR (162 MHz, Deuterium Oxide) δ -10.59. HRMS: calc for [$C_{13}H_{16}N_8O_6P^-$] 411.0936, found 411.0933.

Arabinofuranosylguanine 5′-phosphoro-2-aminoimidazolide (2AIparaG): 7.88 (s, 1H), 6.68 (t, *J* = 1.7 Hz, 1H), 6.45 (t, *J* = 2.2 Hz, 1H), 6.17 (d, *J* = 6.1 Hz, 1H), 4.54 (t, *J* = 6.3 Hz, 1H), 4.45 (t, *J* = 6.6 Hz, 1H), 4.26 – 4.14 (m, 2H), 4.14 – 4.00 (m, 1H). ^{31}P NMR (162 MHz, Deuterium Oxide) δ -10.71. HRMS: calc for [$C_{13}H_{17}N_8O_7P$-H^+] 429.1031, found 429.1046.

2AIpdG: 1H NMR (400 MHz, Deuterium Oxide) δ 7.93 (s, 1H), 6.61 (t, J = 1.8 Hz, 1H), 6.46 (t, J = 2.0 Hz, 1H), 6.27 (t, J = 6.7 Hz, 1H), 4.68 – 4.62 (m, 1H),4.20 (d, J = 4.0 Hz, 1H), 4.07 (t, J = 4.8 Hz, 2H), 2.86 (dt, J = 13.7, 6.5 Hz, 1H), 2.51 (ddd, J = 14.2, 6.8, 4.3 Hz, 1H). ^{31}P NMR (162 MHz, Deuterium Oxide) δ -10.32. HRMS: calc for [$C_{13}H_{17}N_8O_6P$-H^+] 413.1081, found 413.1085.

2AIdeoxyA: 1H NMR (400 MHz, Deuterium Oxide) δ 8.28 (s, 1H), 8.23 (s, 1H), 6.54 (dd, J = 2.3, 1.7 Hz, 1H), 6.48 – 6.38 (m, 2H), 4.25 (d, J = 5.4 Hz, 1H), 4.11 (dd, J = 5.5, 2.8 Hz, 2H), 2.91 (dt, J = 13.7, 6.3 Hz, 1H), 2.62 (ddd, J = 14.2, 7.4, 4.6 Hz, 1H). ^{31}P NMR (162 MHz, Deuterium Oxide) δ -10.02. HRMS: calc for [$C_{13}H_{16}N_8O_5P^-$] 395.0987, found 395.0982.

Arabinofuranosyladenine 5′-phosphoro-2-aminoimidazolide (2AIparaA-imidazolium bridged dimer): ^{1}H NMR (400 MHz, Deuterium Oxide) δ 8.00 (s, 1H), 6.62 (t, J = 1.7 Hz, 1H), 6.38 (t, J = 1.7 Hz, 1H), 5.81 (d, J = 4.6 Hz, 1H), 5.12 (t, J = 5.2 Hz, 1H), 4.54 (t, J = 5.1 Hz, 1H), 4.17 (m, 3H). ^{31}P NMR (162 MHz, Deuterium Oxide) δ -12.72. HRMS: calc for [$C_{23}H_{28}N_{13}O_{12}P_2^-$] 740.1461, found 740.1474.

Ribofuranosyladenine 5′-phosphoro-2-aminoimidazolide (2AIparaA-imidazolium bridged dimer): ^{1}H NMR (400 MHz, Deuterium Oxide) δ 8.00 (s, 1H), 6.62 (t, J = 1.7 Hz, 1H), 6.38 (t, J = 1.7 Hz, 1H), 5.81 (d, J = 4.6 Hz, 1H), 5.12 (t, J = 5.2 Hz, 1H), 4.54 (t, J = 5.1 Hz, 1H), 4.17 (m, 3H). ^{31}P NMR (162 MHz, Deuterium Oxide) δ -10.97. HRMS: calc for [$C_{23}H_{28}N_{13}O_{12}P_2^-$] 740.1461, found 740.1470.

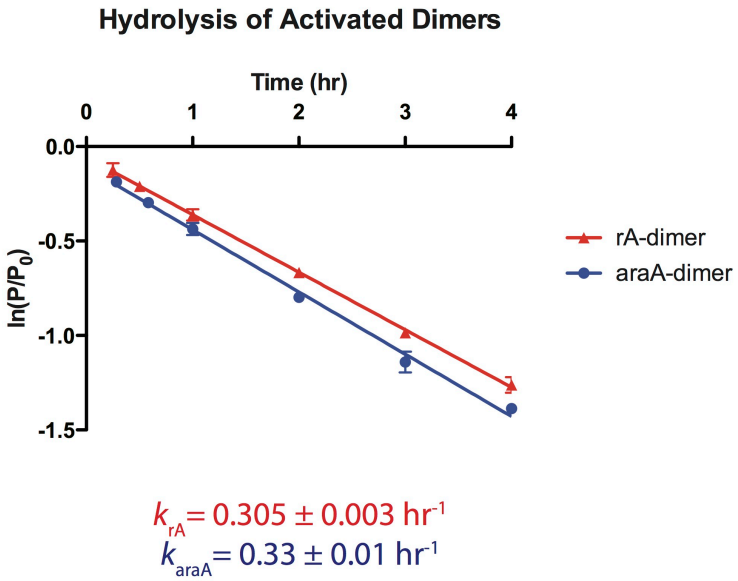


Figure S.3.1: Kinetic analysis of the hydrolysis of imidazolium bridged dimers for araA and rA was carried out in triplicate using 5 mM imidazolium bridged dimers, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. As standard for first-order kinetics, a linearized plot of $\ln(P/P_0)$ as a function of time is shown above and the rate of extension was determined from linear least-squares fits of the data from an average of three independent experiments.

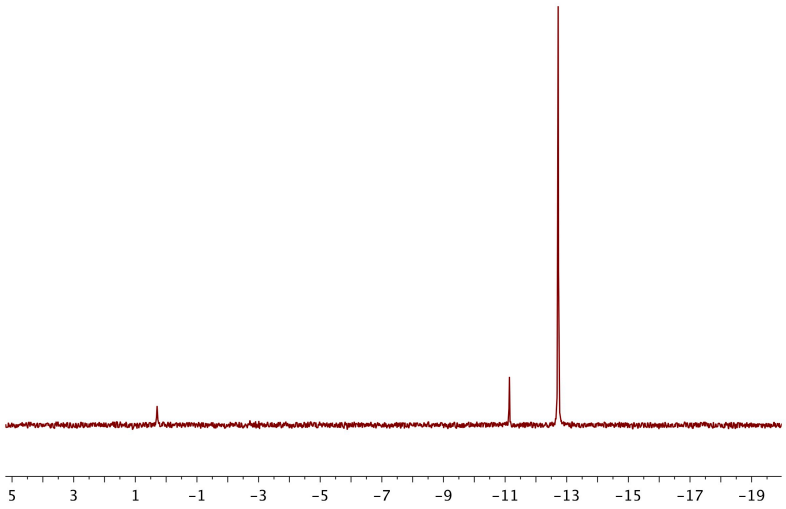


Figure S.3.2: Representative ^{31}P NMR spectrum of the hydrolysis of araA imidazolium bridged dimer under nonenzymatic primer extension conditions (5 mM imidazolium bridged dimer, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0) after 18 minutes of reaction time. Peaks are referenced to araA bridged dimer (δ = -12.72).

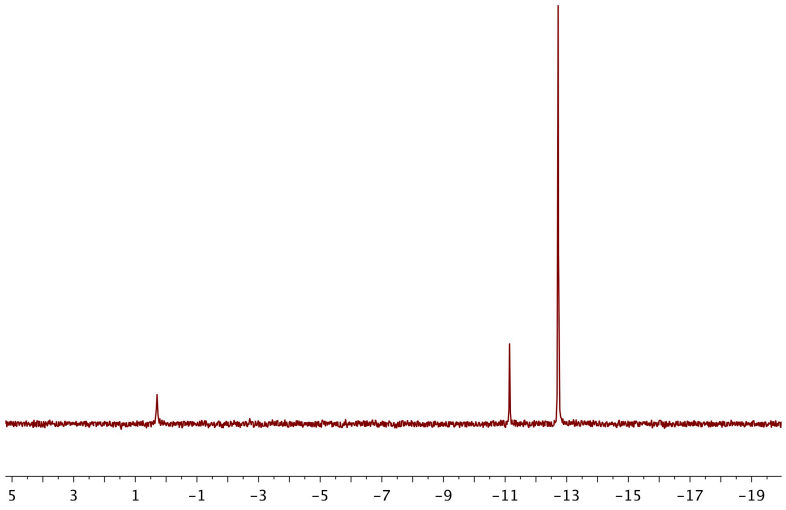


Figure S.3.3: Representative ^{31}P NMR spectrum of the hydrolysis of araA imidazolium bridged dimer under nonenzymatic primer extension conditions (5 mM imidazolium bridged dimer, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0) after 35 minutes of reaction time. Peaks are referenced to araA bridged dimer (δ = -12.72).

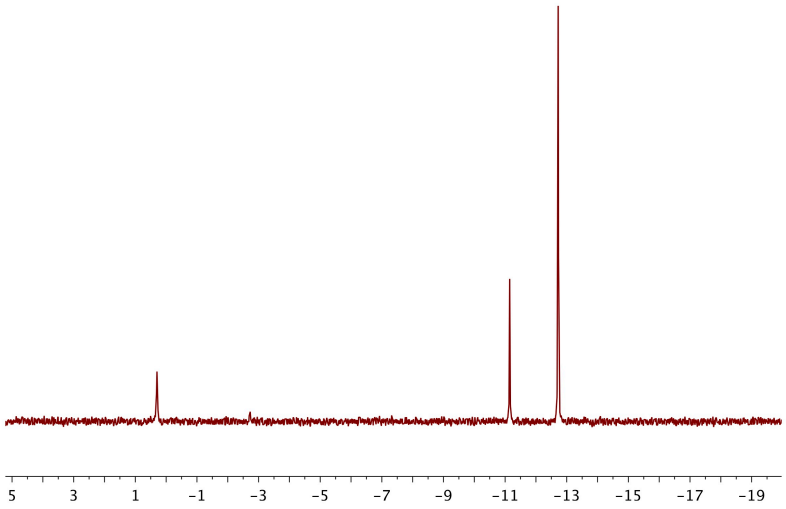


Figure S.3.4: Representative ^{31}P NMR spectrum of the hydrolysis of araA imidazolium bridged dimer under nonenzymatic primer extension conditions (5 mM imidazolium bridged dimer, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0) after 60 minutes of reaction time. Peaks are referenced to araA bridged dimer (δ = -12.72).

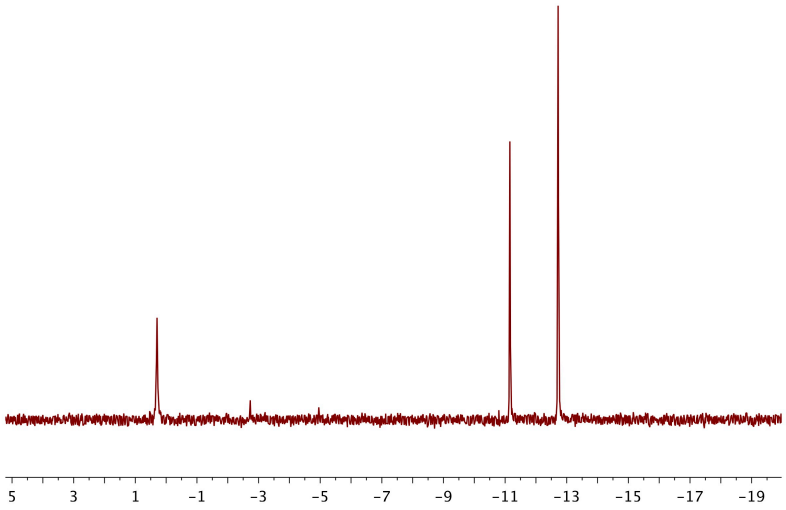


Figure S.3.5: Representative ^{31}P NMR spectrum of the hydrolysis of araA imidazolium bridged dimer under nonenzymatic primer extension conditions (5 mM imidazolium bridged dimer, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0) after 120 minutes of reaction time. Peaks are referenced to araA bridged dimer (δ = -12.72).

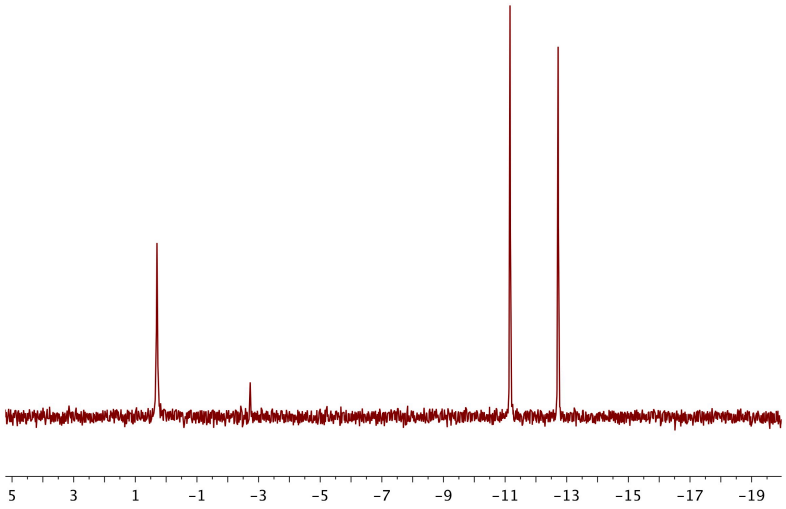


Figure S.3.6: Representative ^{31}P NMR spectrum of the hydrolysis of araA imidazolium bridged dimer under nonenzymatic primer extension conditions (5 mM imidazolium bridged dimer, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0) after 180 minutes of reaction time. Peaks are referenced to araA bridged dimer (δ = -12.72).

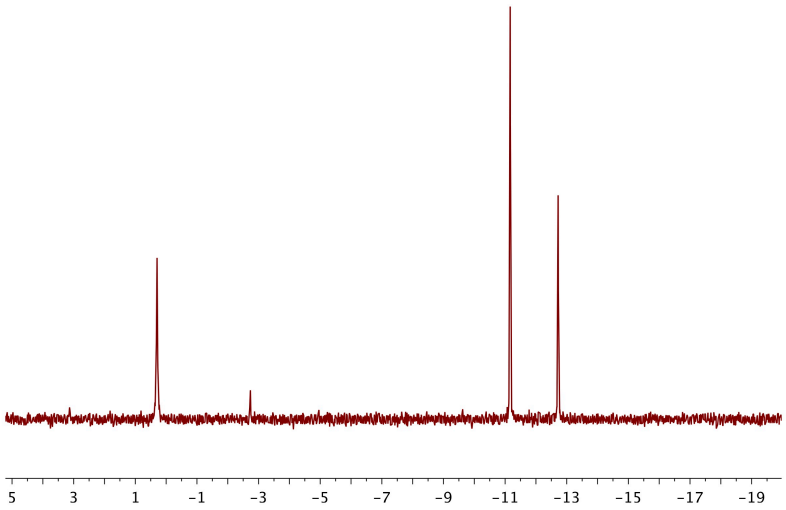


Figure S.3.7: Representative ^{31}P NMR spectrum of the hydrolysis of araA imidazolium bridged dimer under nonenzymatic primer extension conditions (5 mM imidazolium bridged dimer, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0) after 240 minutes of reaction time. Peaks are referenced to araA bridged dimer (δ = -12.72).

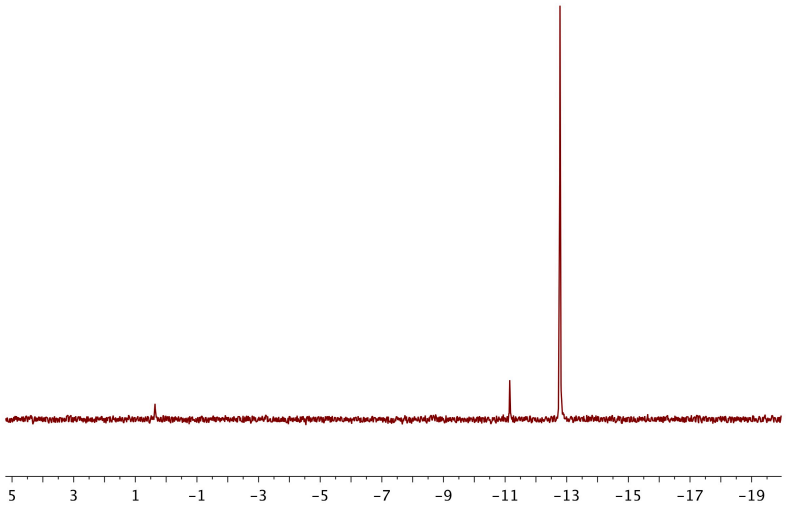


Figure S.3.8: Representative ^{31}P NMR spectrum of the hydrolysis of rA imidazolium bridged dimer under nonenzymatic primer extension conditions (5 mM imidazolium bridged dimer, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0) after 15 minutes of reaction time. Peaks are referenced to rA bridged dimer (δ = -12.72).

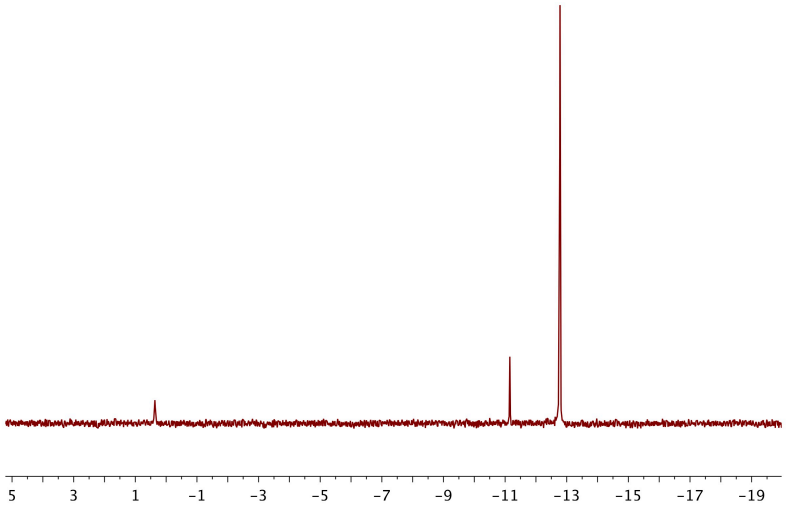


Figure S.3.9: Representative ^{31}P NMR spectrum of the hydrolysis of rA imidazolium bridged dimer under nonenzymatic primer extension conditions (5 mM imidazolium bridged dimer, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0) after 35 minutes of reaction time. Peaks are referenced to rA bridged dimer (δ = -12.72).

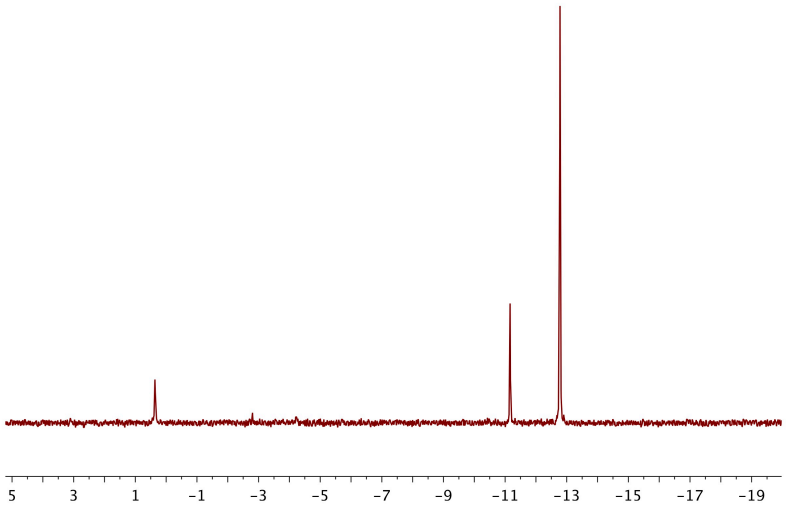


Figure S.3.10: Representative ^{31}P NMR spectrum of the hydrolysis of rA imidazolium bridged dimer under nonenzymatic primer extension conditions (5 mM imidazolium bridged dimer, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0) after 60 minutes of reaction time. Peaks are referenced to rA bridged dimer (δ = -12.72).

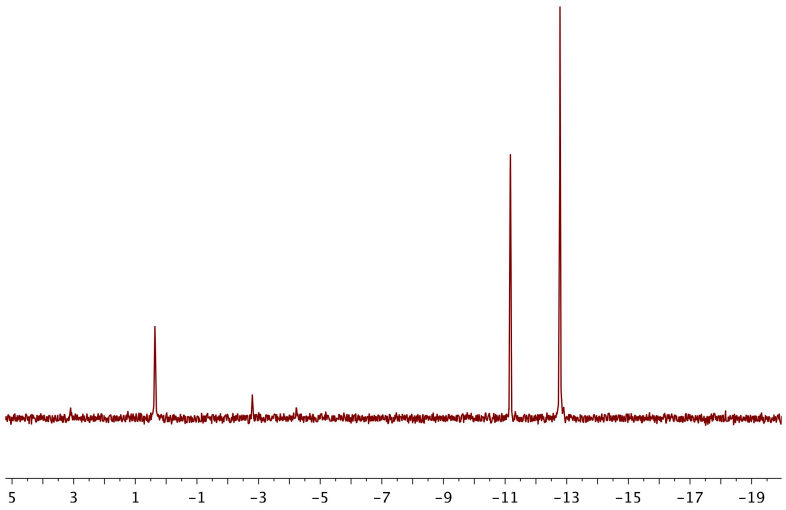


Figure S.3.11: Representative ^{31}P NMR spectrum of the hydrolysis of rA imidazolium bridged dimer under nonenzymatic primer extension conditions (5 mM imidazolium bridged dimer, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0) after 120 minutes of reaction time. Peaks are referenced to rA bridged dimer (δ = -12.72).

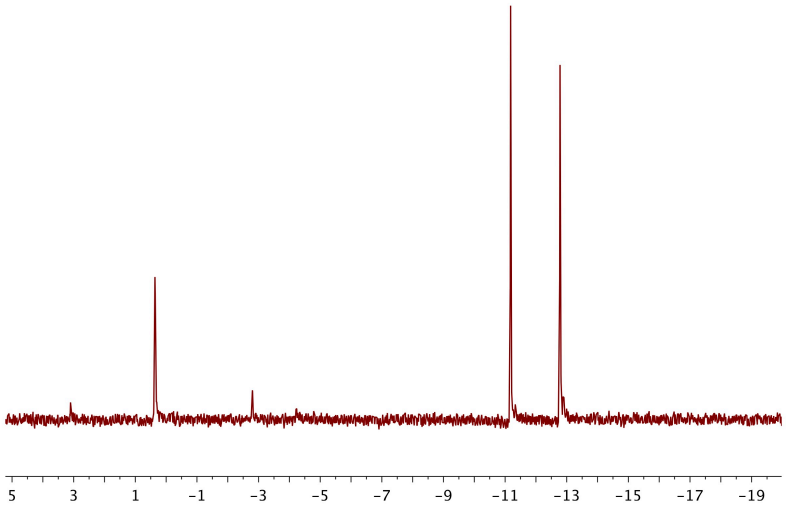


Figure S.3.12: Representative ^{31}P NMR spectrum of the hydrolysis of rA imidazolium bridged dimer under nonenzymatic primer extension conditions (5 mM imidazolium bridged dimer, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0) after 180 minutes of reaction time. Peaks are referenced to rA bridged dimer (δ = -12.72).

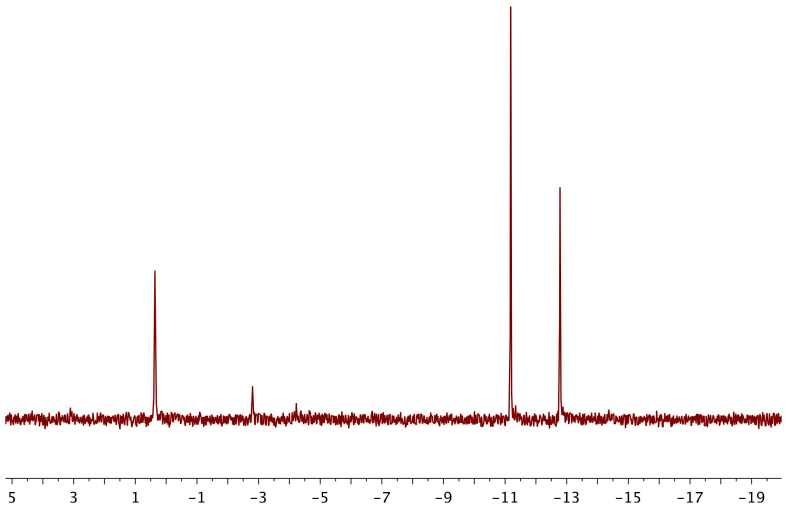
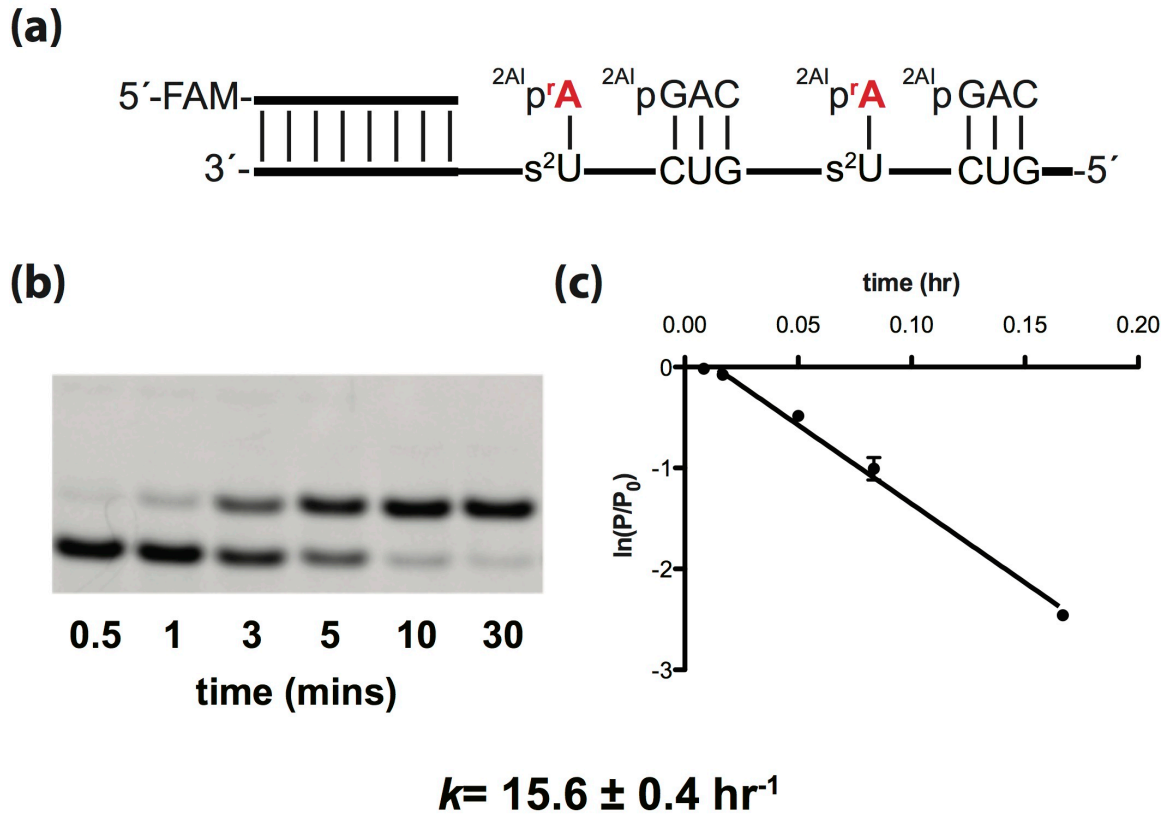
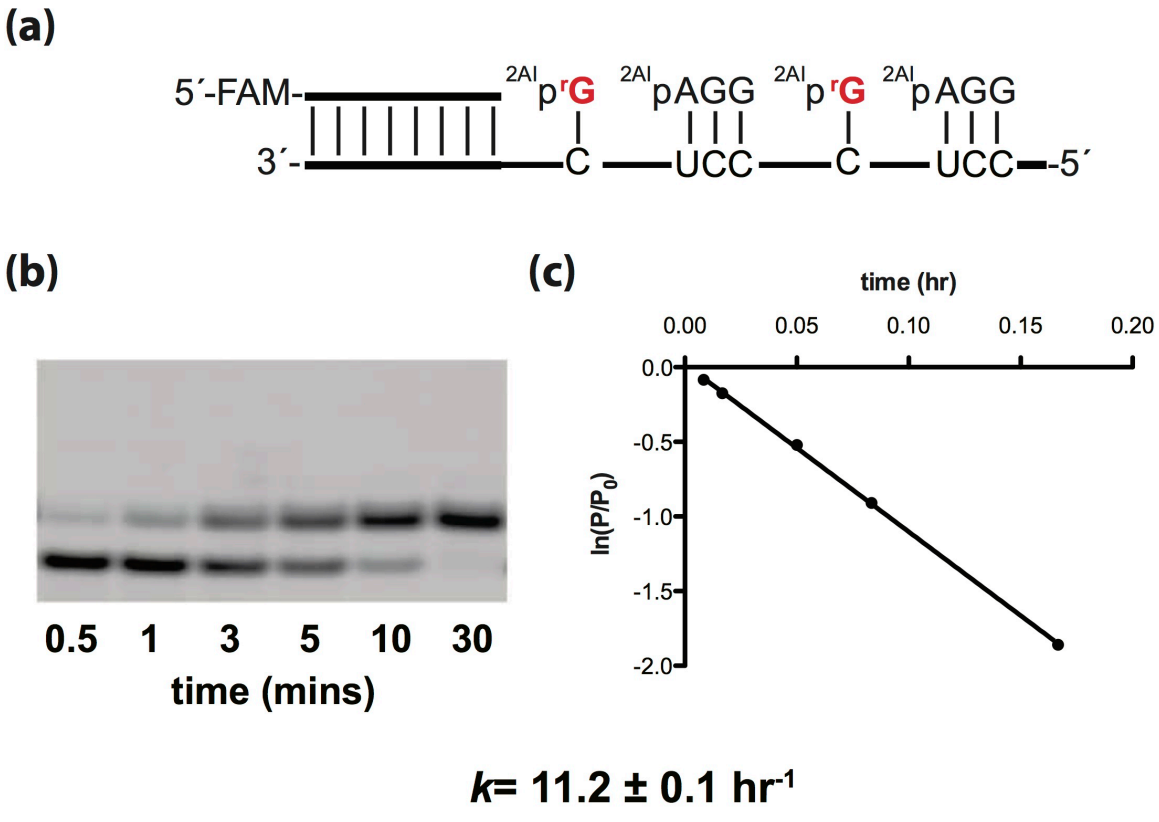


Figure S.3.13: Representative ^{31}P NMR spectrum of the hydrolysis of rA imidazolium bridged dimer under nonenzymatic primer extension conditions (5 mM imidazolium bridged dimer, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0) after 240 minutes of reaction time. Peaks are referenced to rA bridged dimer (δ = -12.72).



k= 15.6 ± 0.4 hr^{-1}

Figure S.3.14: Primer extension reactions were carried out in triplicate using 20 mM 2AIprA, 0.5 mM 2AIpGAC, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction. (b) Representative PAGE analysis of result. (c) Plot of $\ln(P/P_0)$ as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.



$k = 11.2 \pm 0.1\ hr^{-1}$

Figure S.3.15: Primer extension reactions were carried out in triplicate using 20 mM 2AIprG, 0.5 mM 2AIpGAC, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction. (b) Representative PAGE analysis of result. (c) Plot of $\ln(P/P_0)$ as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

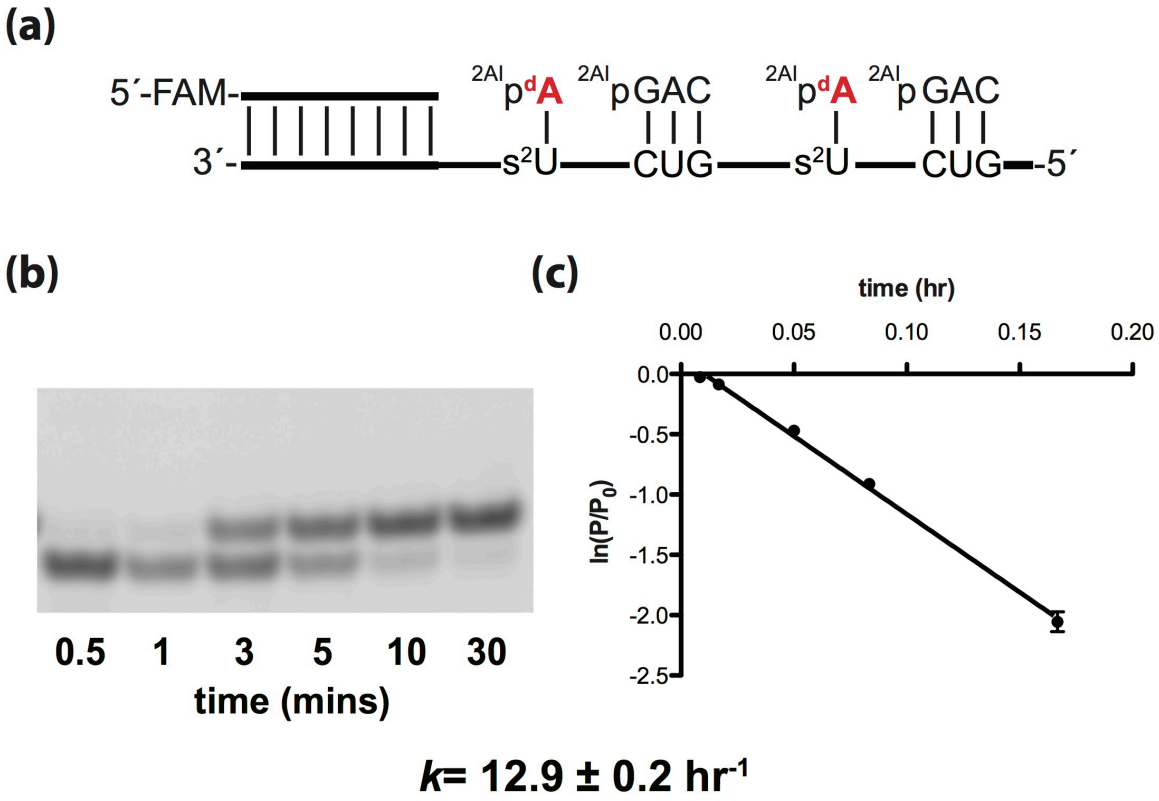
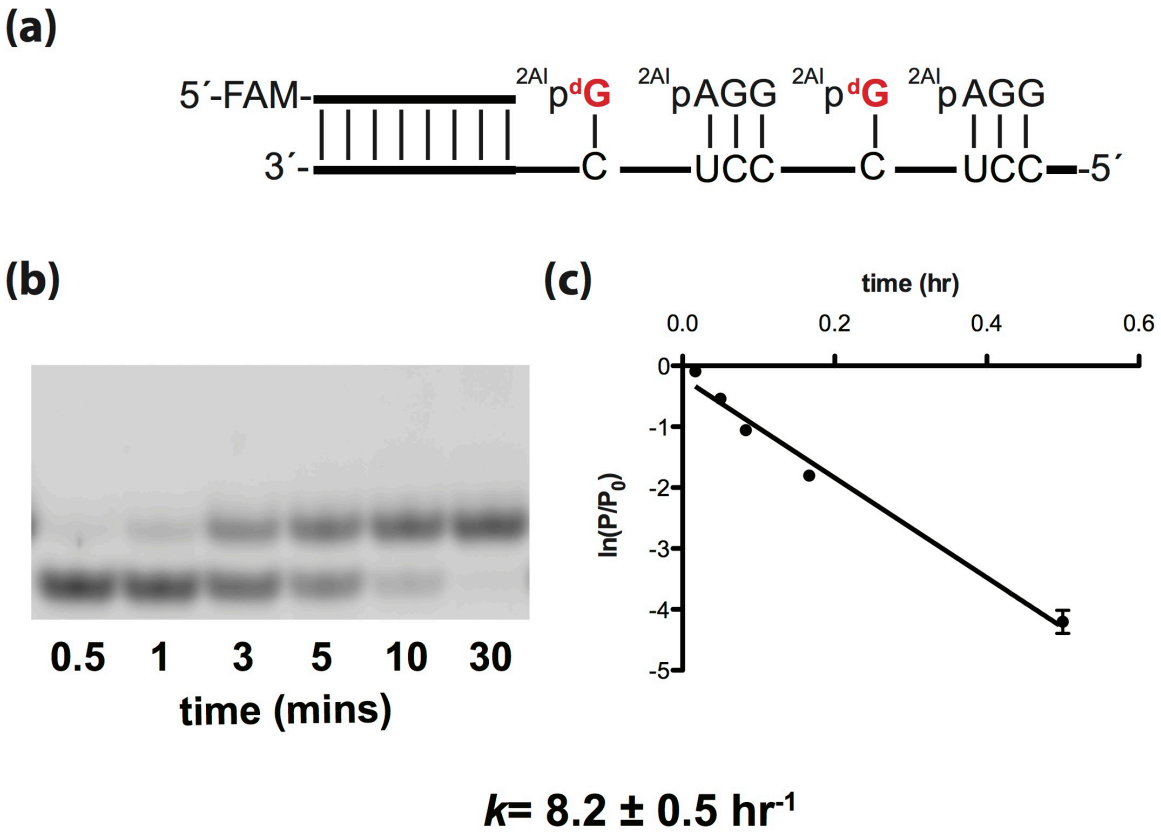
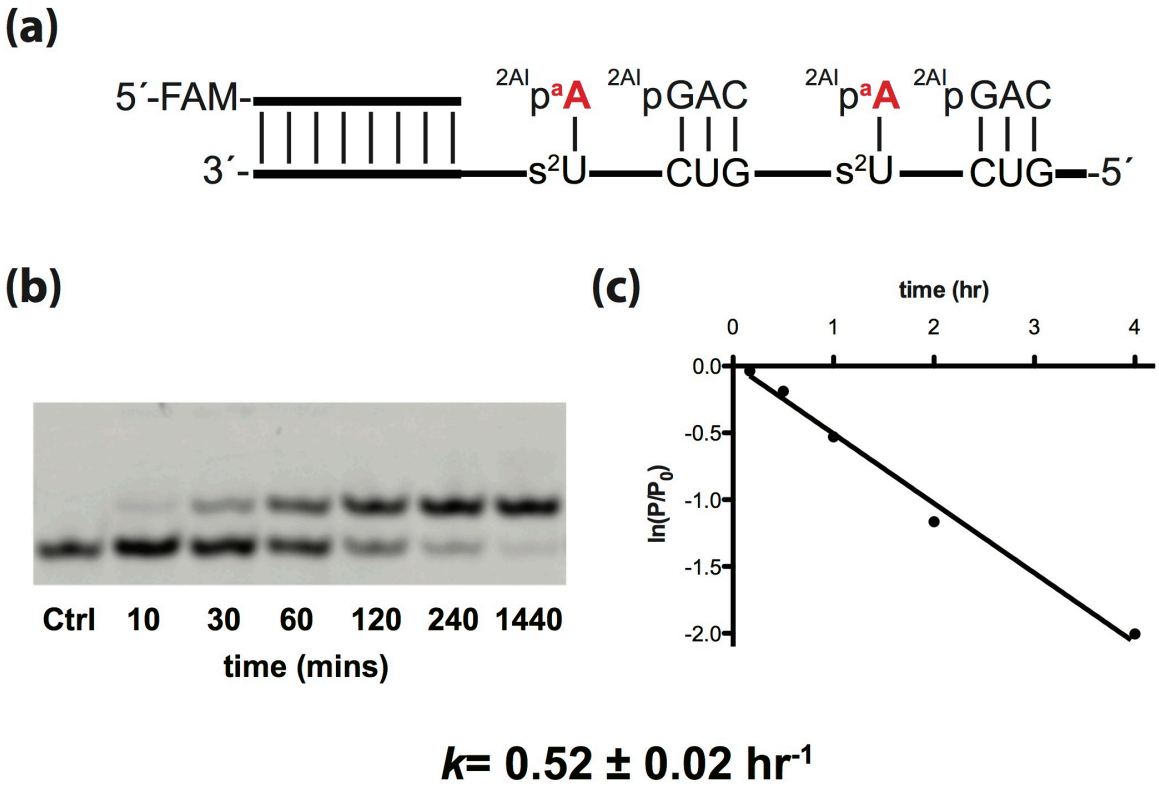


Figure S.3.16: Primer extension reactions were carried out in triplicate using 20 mM 2AIpdA, 0.5 mM 2AIpGAC, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction. (b) Representative PAGE analysis of result. (c) Plot of $\ln(P/P_0)$ as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.



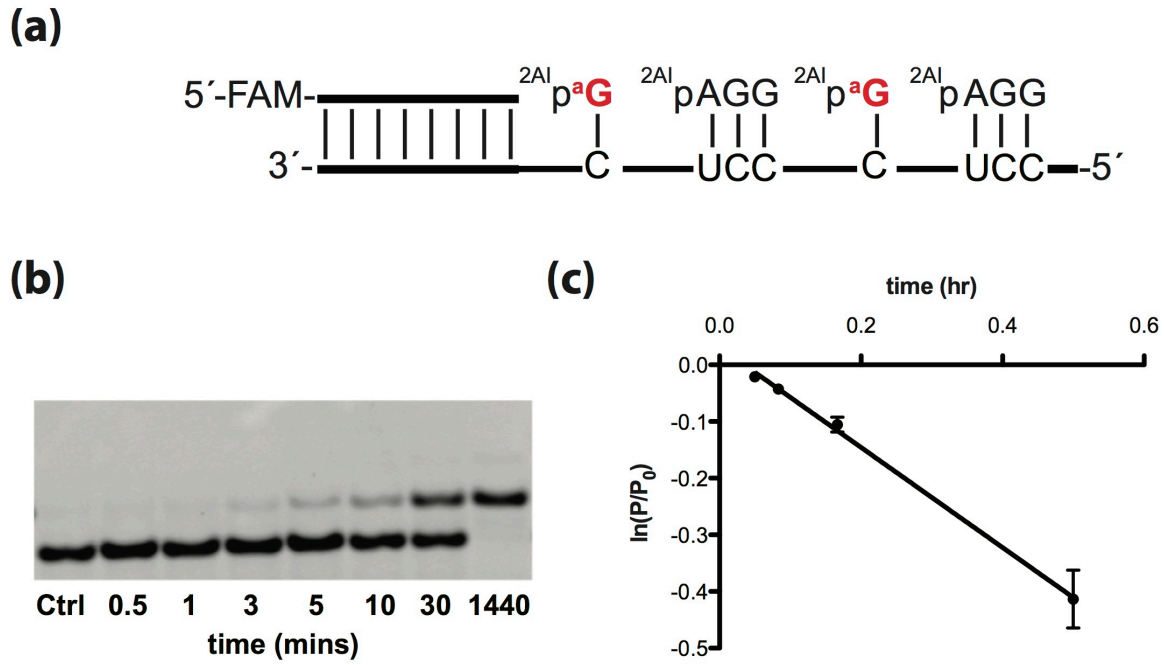
$k = 8.2 \pm 0.5\ hr^{-1}$

Figure S.3.17: Primer extension reactions were carried out in triplicate using 20 mM 2AIpdG, 0.5 mM 2AIpGAC, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction. (b) Representative PAGE analysis of result. (c) Plot of $\ln(P/P_0)$ as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.



$k = 0.52 \pm 0.02\ hr^{-1}$

Figure S.3.18: Primer extension reactions were carried out in triplicate using 20 mM 2AIparaA, 0.5 mM 2AIpGAC, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction. (b) Representative PAGE analysis of result. (c) Plot of $\ln(P/P_0)$ as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.



$k = 0.88 \pm 0.04\ hr^{-1}$

Figure S.3.19: Primer extension reactions were carried out in triplicate using 20 mM 2AIparaG, 0.5 mM 2AIpGAC, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction. (b) Representative PAGE analysis of result. (c) Plot of $\ln(P/P_0)$ as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

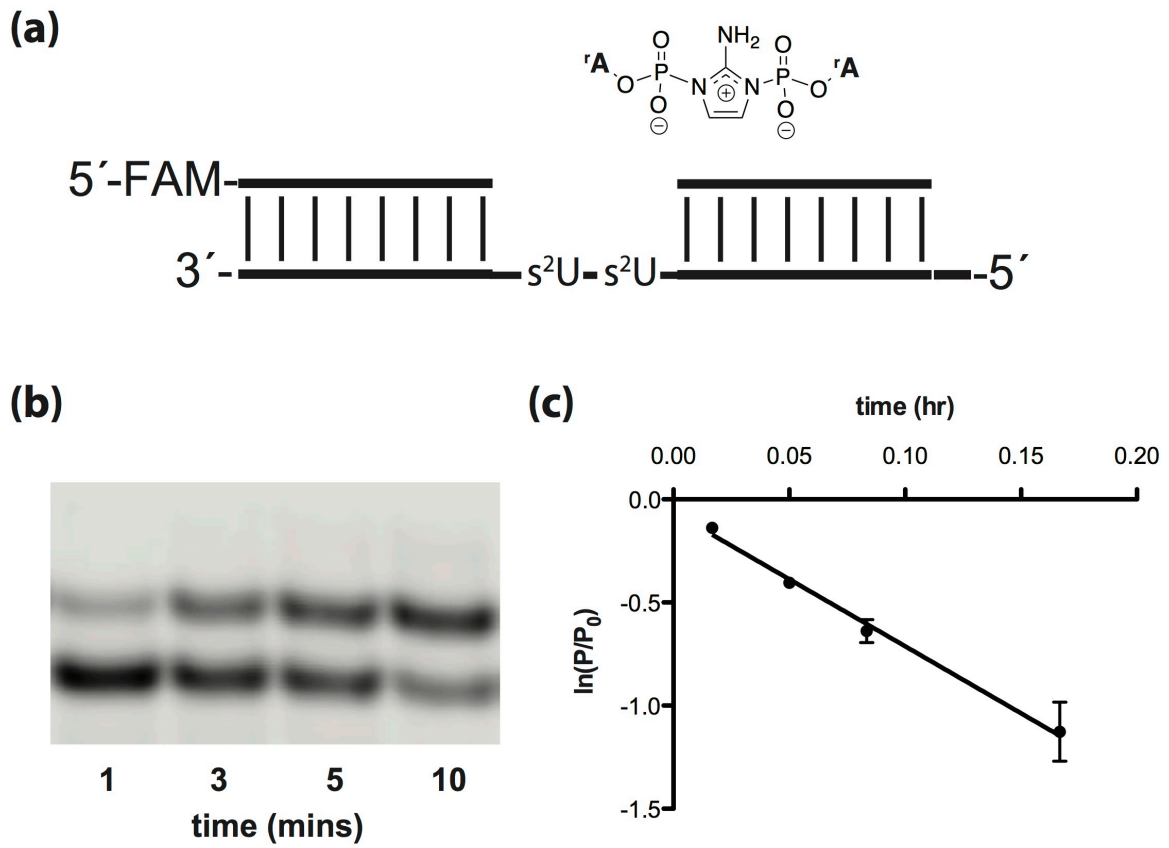


Figure S.3.20: Michaelis-Menten kinetics of the addition of 2AIprA imidazolium-bridged dimer. Primer extension reactions were carried out in triplicate using 0.05 mM of 2AIpA imidazolium-bridged dimer, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction. (b) Representative PAGE analysis of result. (c) Plot of $\ln(P/P_0)$as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

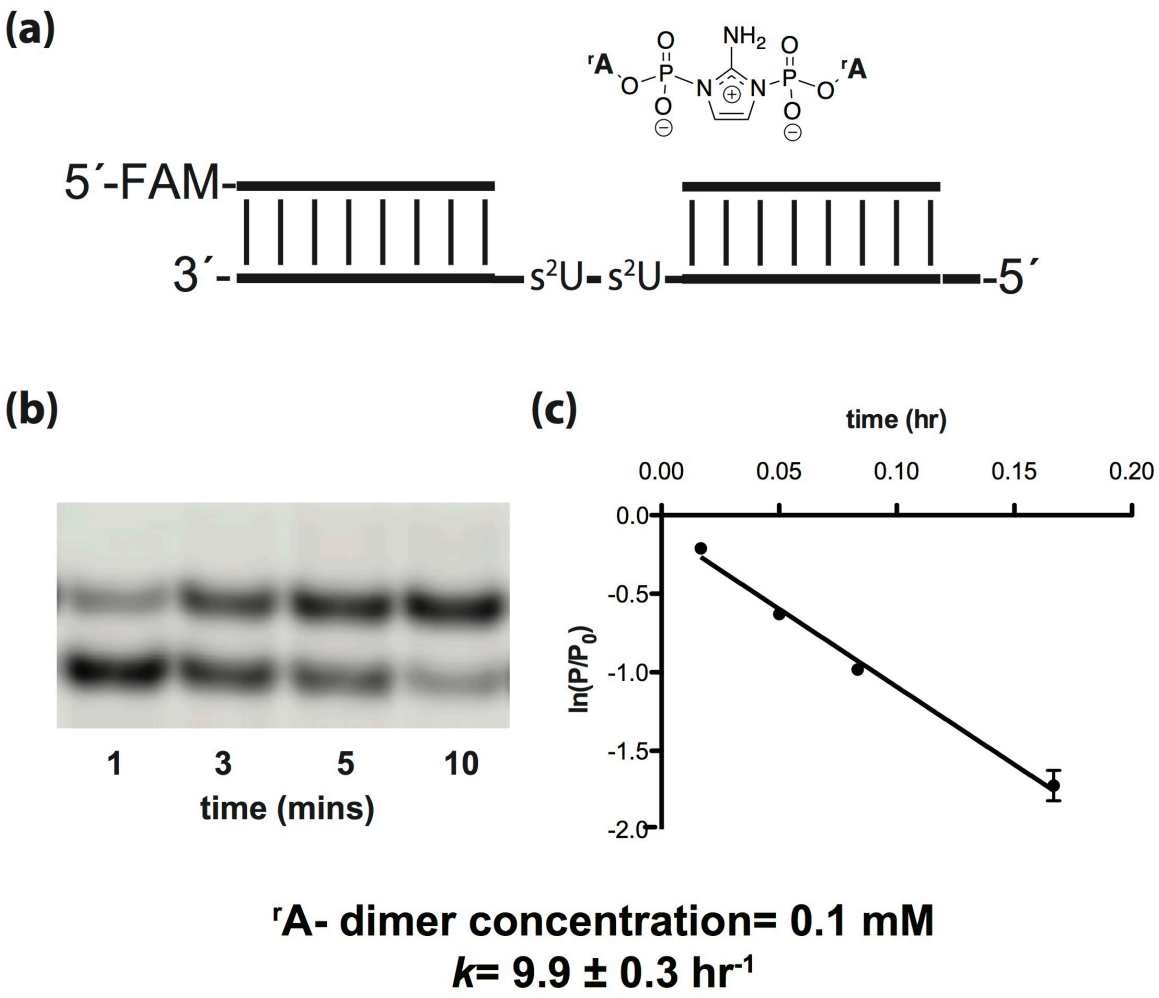


Figure S.3.21: Michaelis-Menten kinetics of the addition of 2AIprA imidazolium-bridged dimer. Primer extension reactions were carried out in triplicate using 0.1 mM of 2AIpA imidazolium-bridged dimer, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction. (b) Representative PAGE analysis of result. (c) Plot of $\ln(P/P_0)$as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

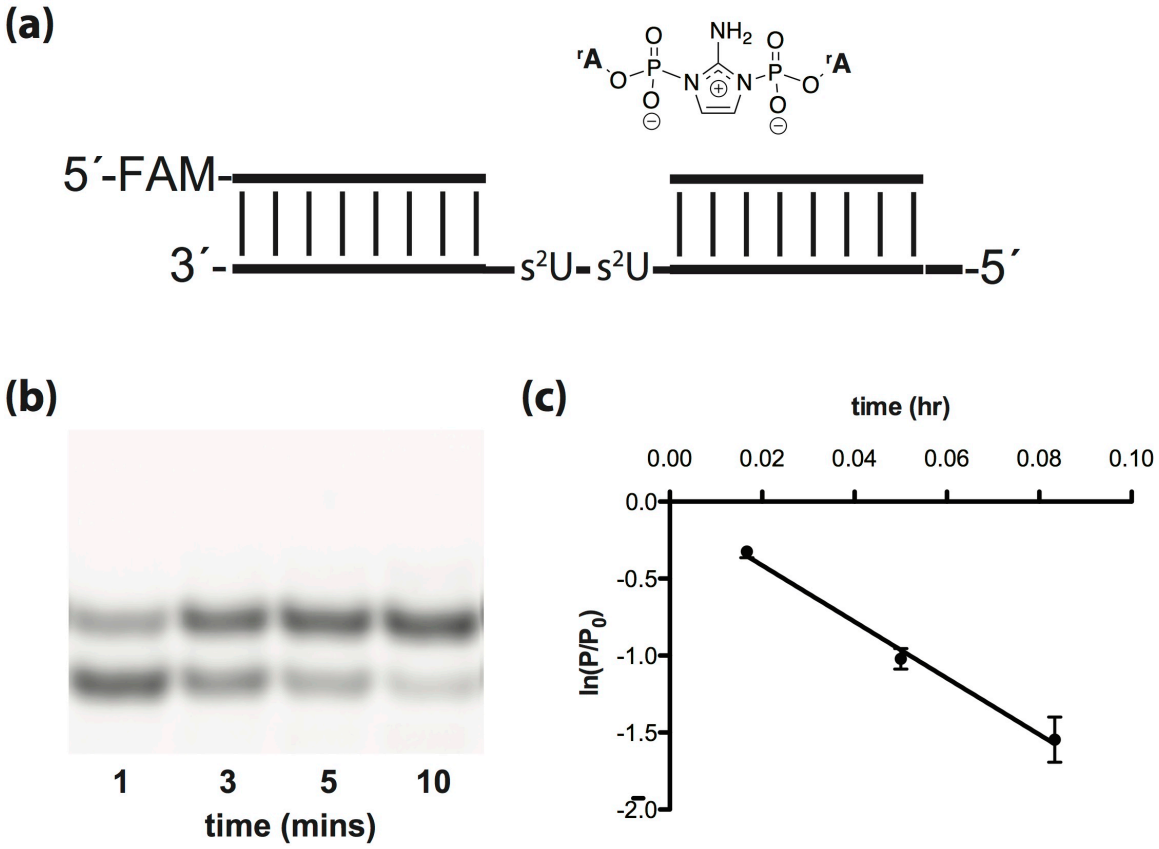


Figure S.3.22: Michaelis-Menten kinetics of the addition of 2AIprA imidazolium-bridged dimer. Primer extension reactions were carried out in triplicate using 0.5 mM of 2AIpA imidazolium-bridged dimer, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction. (b) Representative PAGE analysis of result. (c) Plot of $\ln(P/P_0)$as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

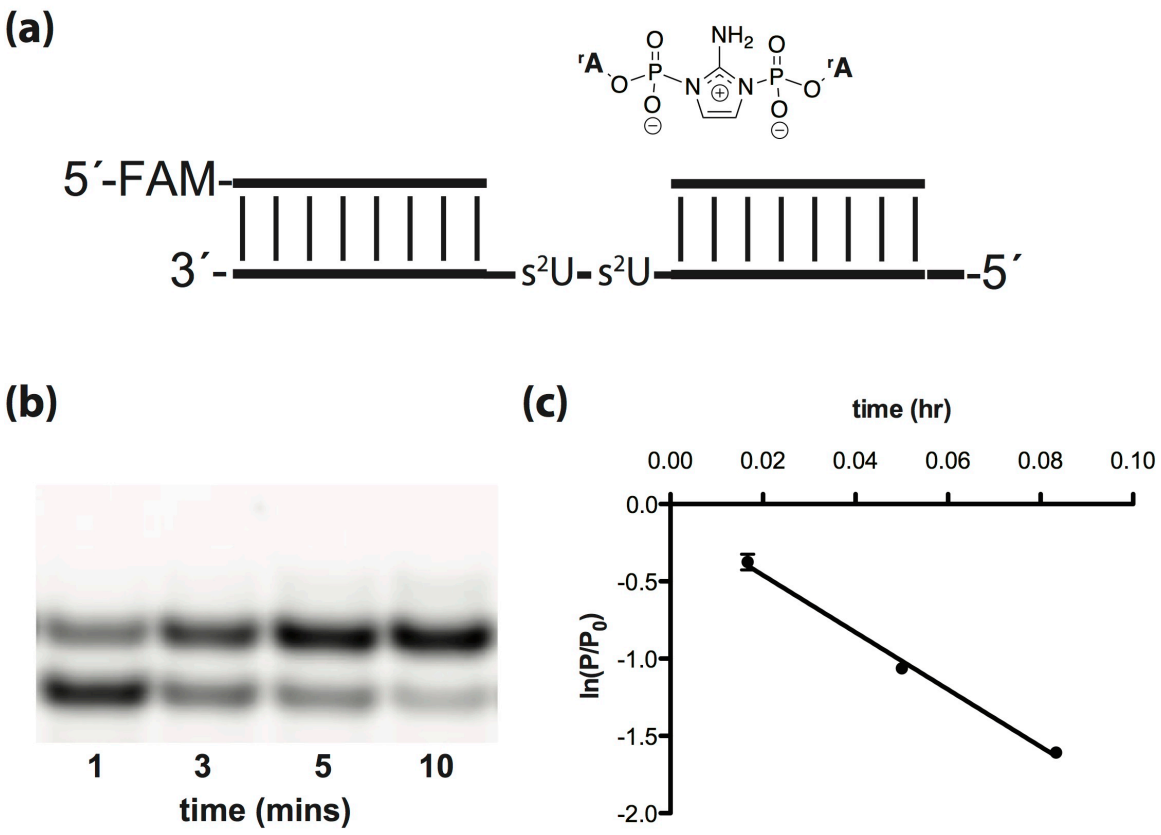
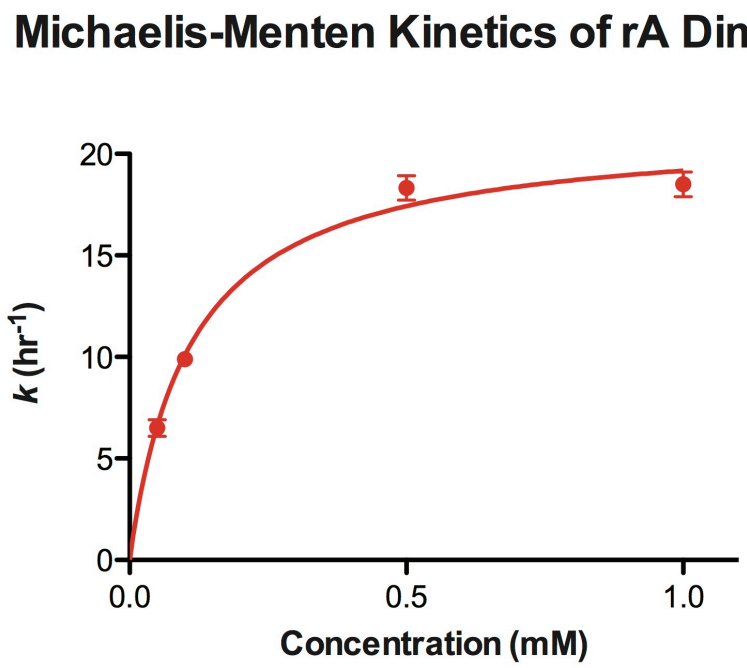


Figure S.3.23: Michaelis-Menten kinetics of the addition of 2AIprA imidazolium-bridged dimer. Primer extension reactions were carried out in triplicate using 1 mM of 2AIpA imidazolium-bridged dimer, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction. (b) Representative PAGE analysis of result. (c) Plot of $\ln(P/P_0)$as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.



$$K_m = 0.11 \pm 0.01 \text{ mM hr}^{-1}$$
$$V_{max} = 21.3 \pm 0.5 \text{ hr}^{-1}$$

Figure S.3.24: Michaelis-Menten kinetics of the addition of 2AIprA imidazolium-bridged dimer. Primer extension reactions were carried out in triplicate using various concentrations of 2AIpA imidazolium-bridged dimer, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0 and plot of k (hr^{-1}) as a function of the concentration (mM) of 2AIprA imidazolium-bridged dimer is shown as above. Michaelis-Menten parameters are K_M : 0.11(1) mM and V_{MAX} of 21.3(5) mM hr^{-1}.

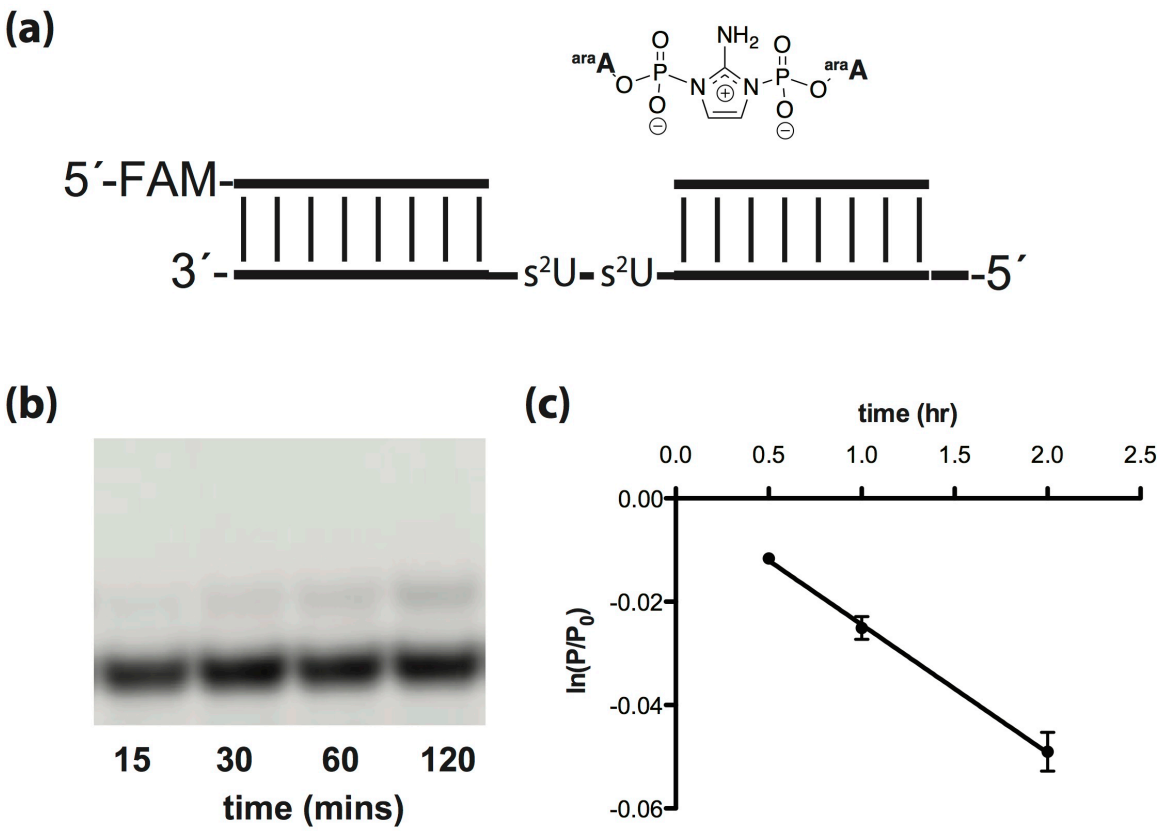


Figure S.3.25: Michaelis-Menten kinetics of the addition of 2AIparaA imidazolium-bridged dimer. Primer extension reactions were carried out in triplicate using 0.05 mM of 2AIparaA imidazolium-bridged dimer, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction. (b) Representative PAGE analysis of result. (c) Plot of $\ln(P/P_0)$as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

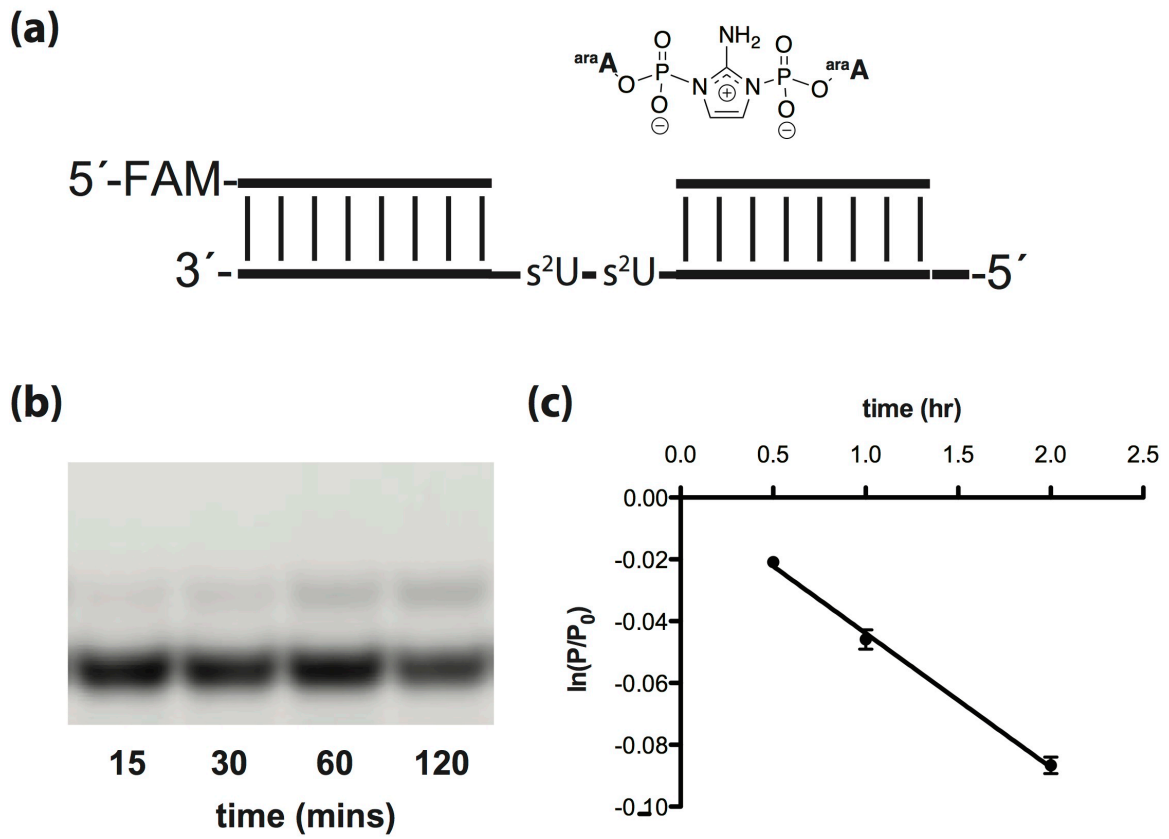


Figure S.3.26: Michaelis-Menten kinetics of the addition of 2AIparaA imidazolium-bridged dimer. Primer extension reactions were carried out in triplicate using 0.1 mM of 2AIparaA imidazolium-bridged dimer, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction. (b) Representative PAGE analysis of result. (c) Plot of $\ln(P/P_0)$as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

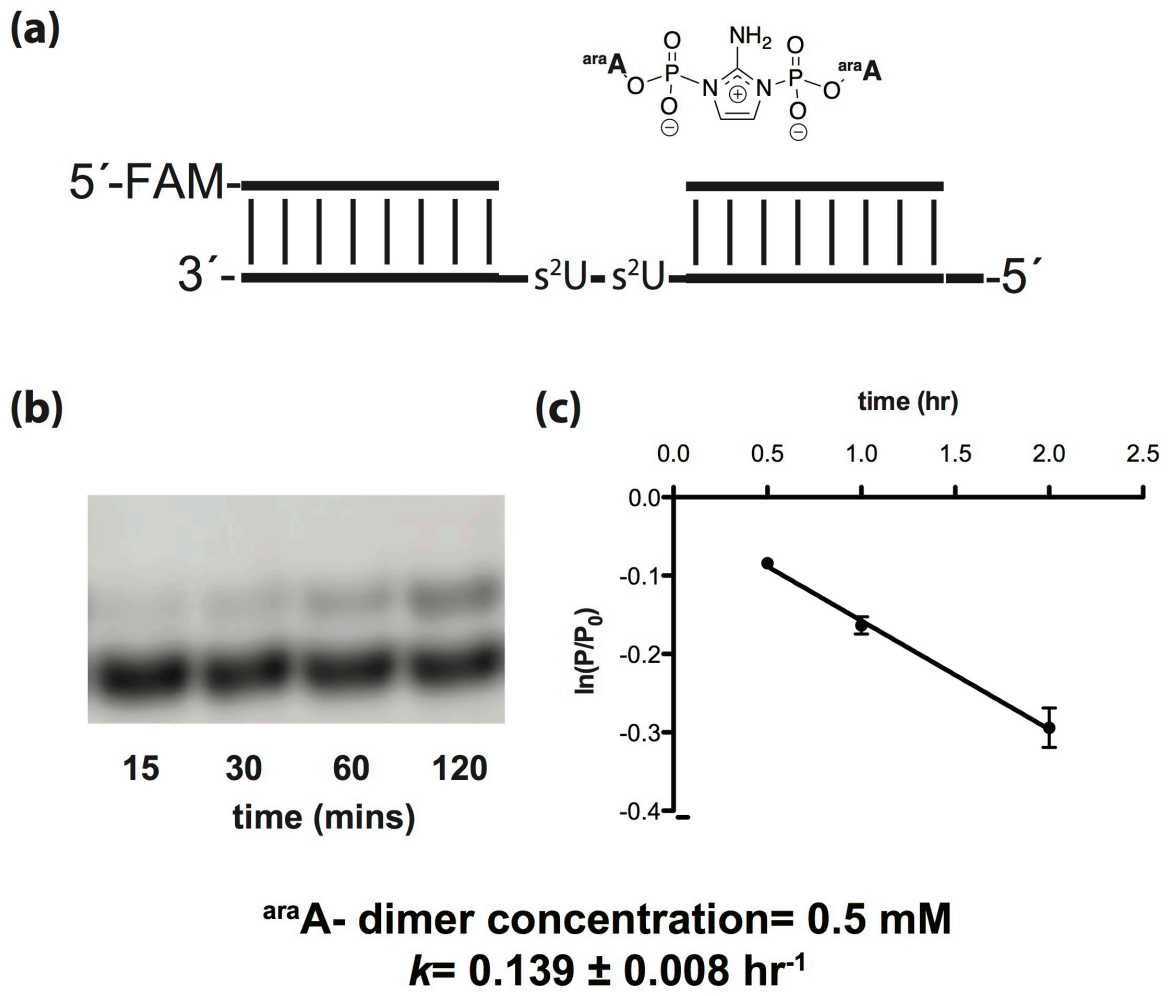


Figure S.3.27: Michaelis-Menten kinetics of the addition of 2AIparaA imidazolium-bridged dimer. Primer extension reactions were carried out in triplicate using 0.5 mM of 2AIparaA imidazolium-bridged dimer, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction. (b) Representative PAGE analysis of result. (c) Plot of $\ln(P/P_0)$as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

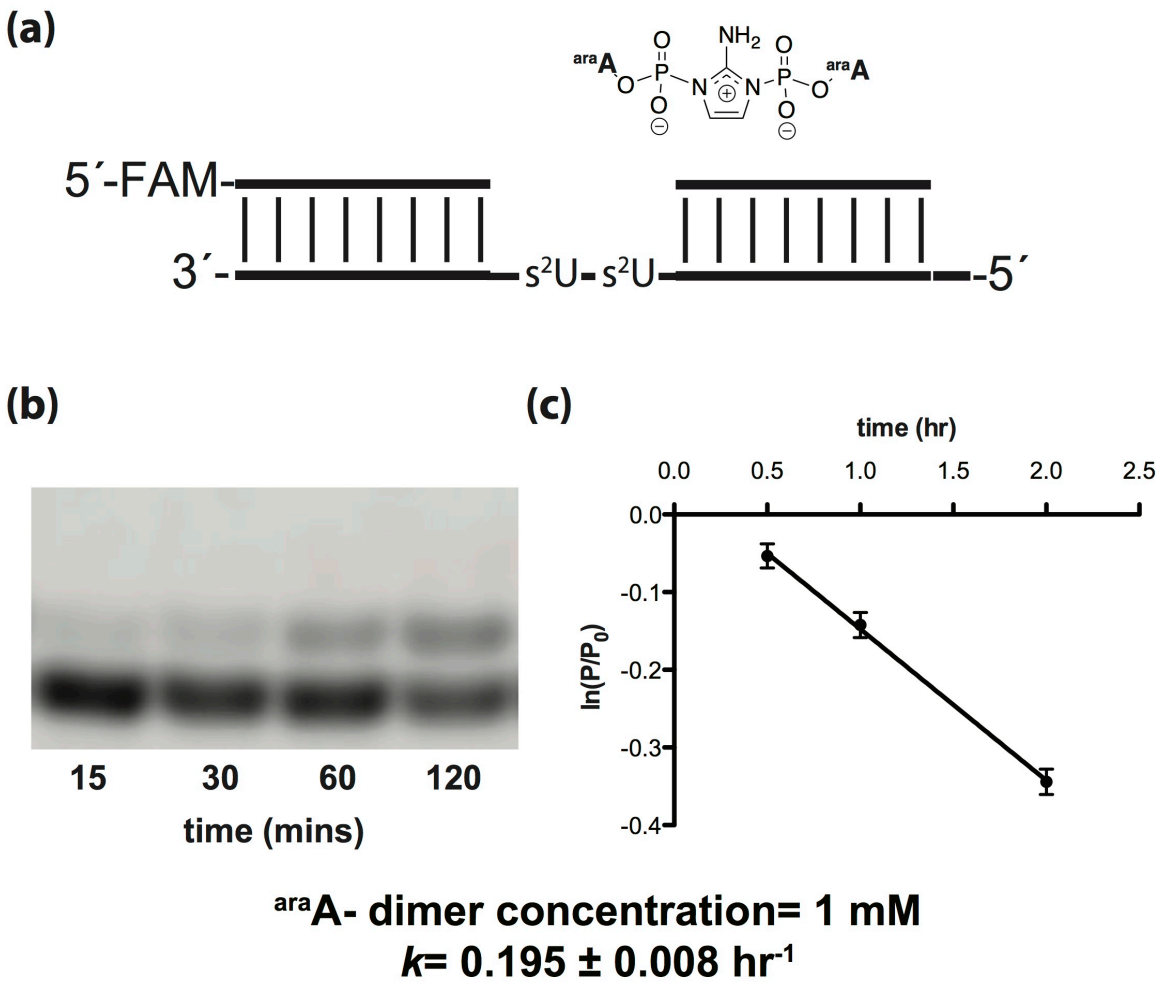


Figure S.3.28: Michaelis-Menten kinetics of the addition of 2AIparaA imidazolium-bridged dimer. Primer extension reactions were carried out in triplicate using 1 mM of 2AIparaA imidazolium-bridged dimer, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction. (b) Representative PAGE analysis of result. (c) Plot of $\ln(P/P_0)$as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

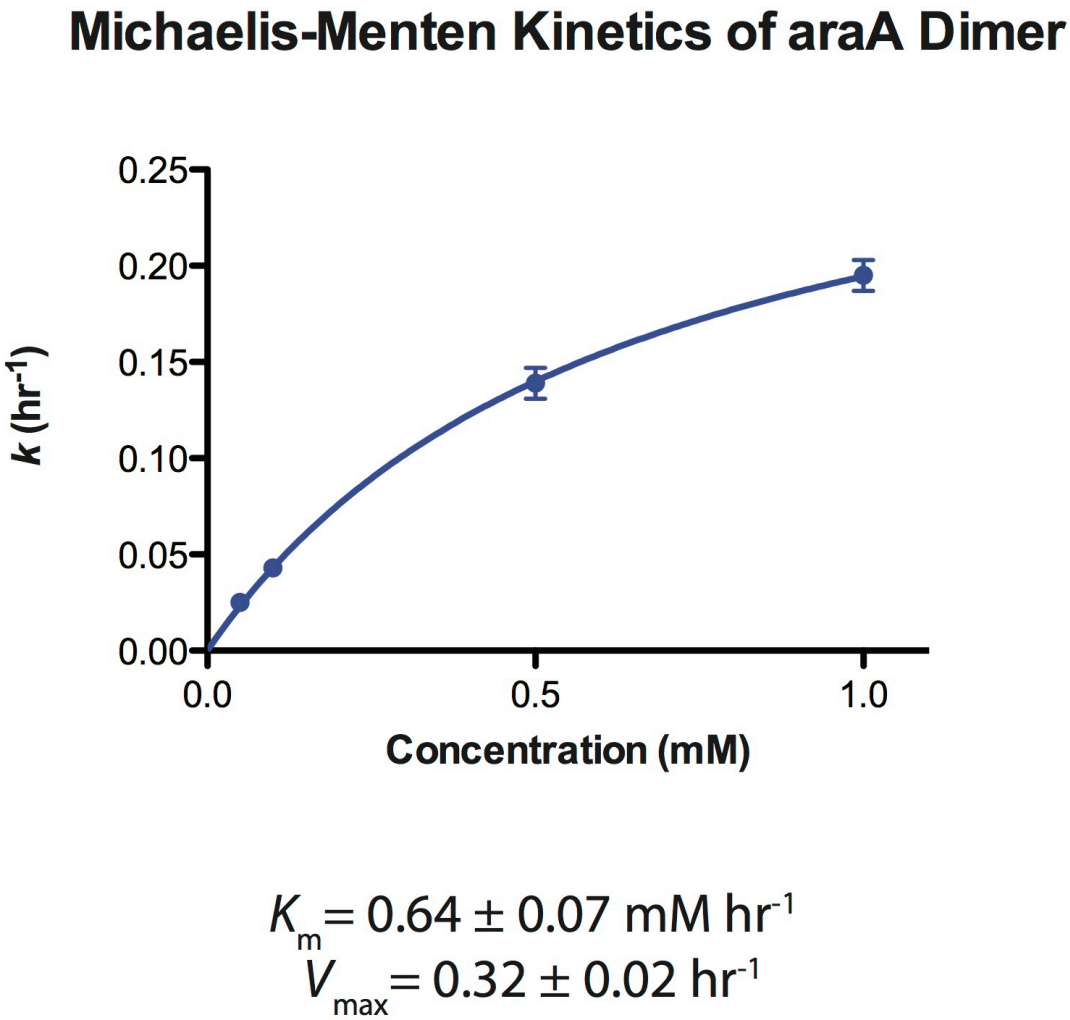
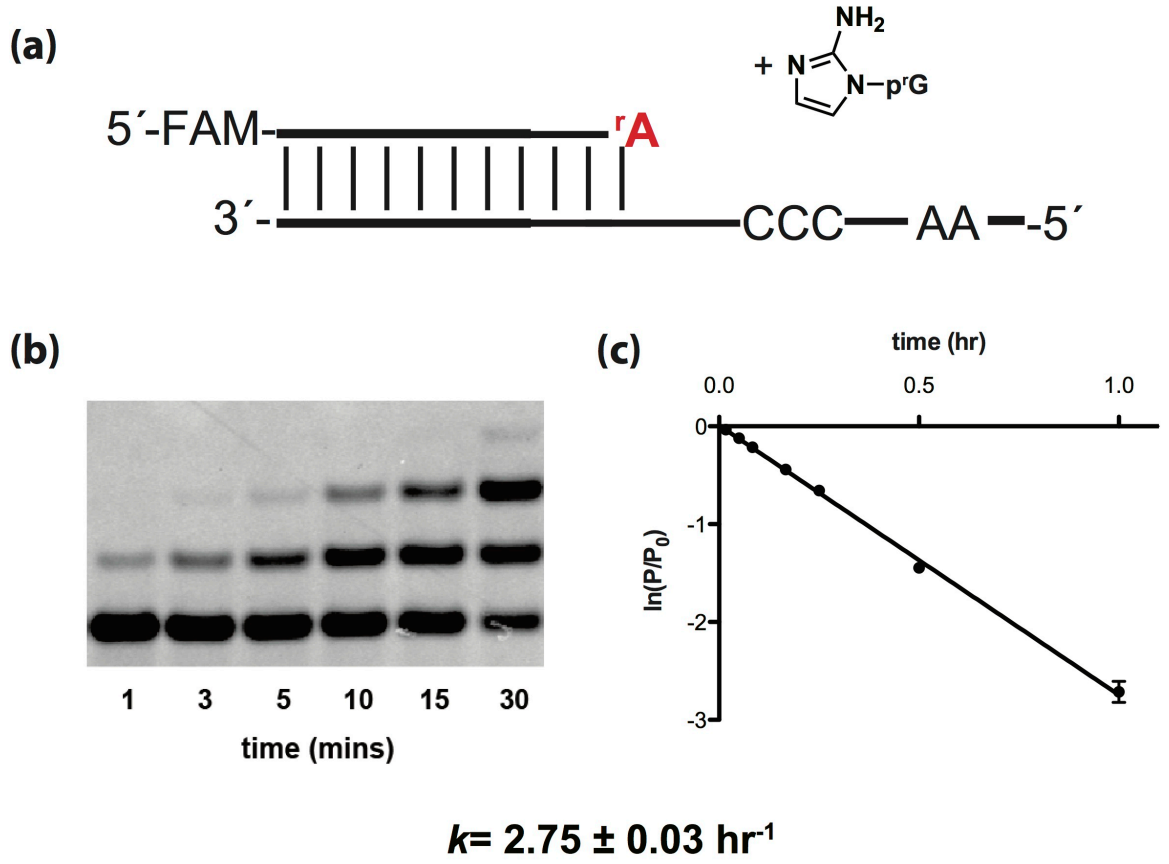
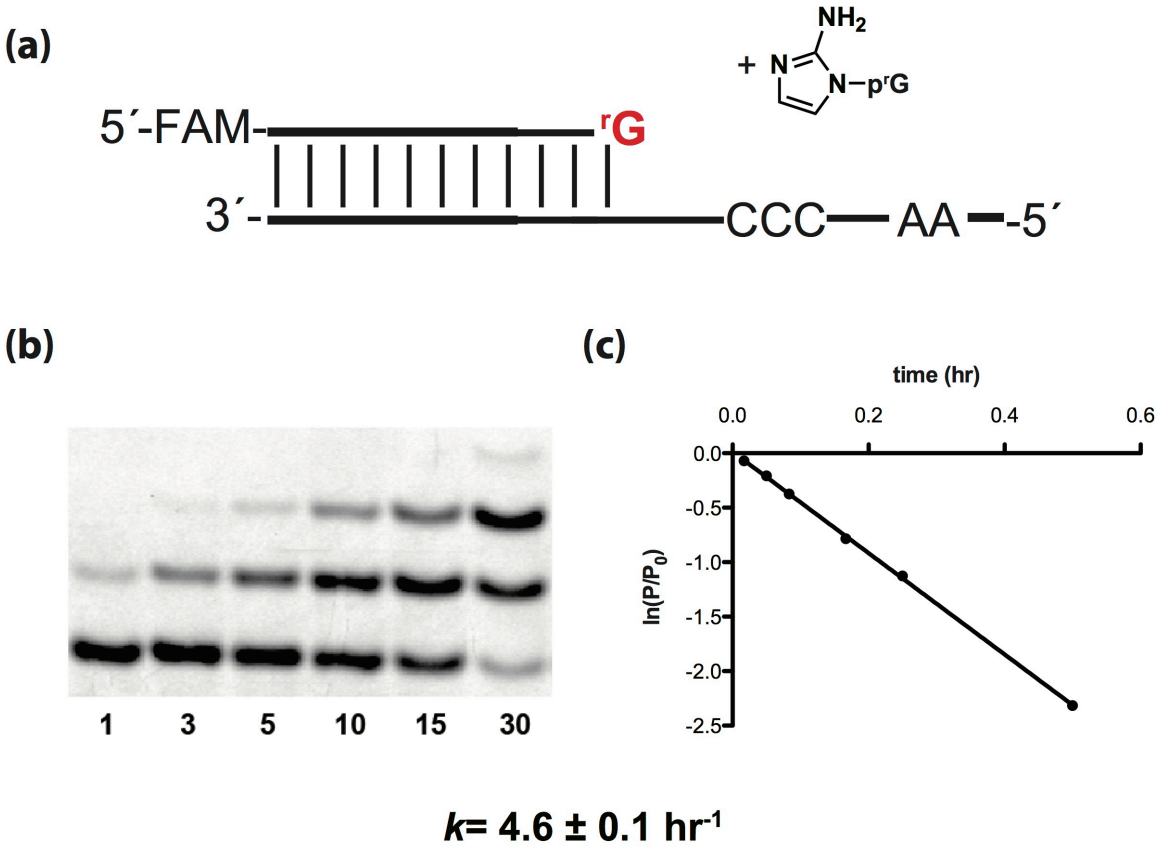


Figure S.3.29: Michaelis-Menten kinetics of the addition of 2AIparaA imidazolium-bridged dimer. Primer extension reactions were carried out in triplicate using various concentrations of 2AIparaA imidazolium-bridged dimer, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0 and plot of k (hr^{-1}) as a function of the concentration (mM) of 2AIparaA imidazolium-bridged dimer is shown as above. Michaelis-Menten parameters are K_M : 0.64(7) mM and V_{MAX} of 0.32(2) mM hr^{-1}.



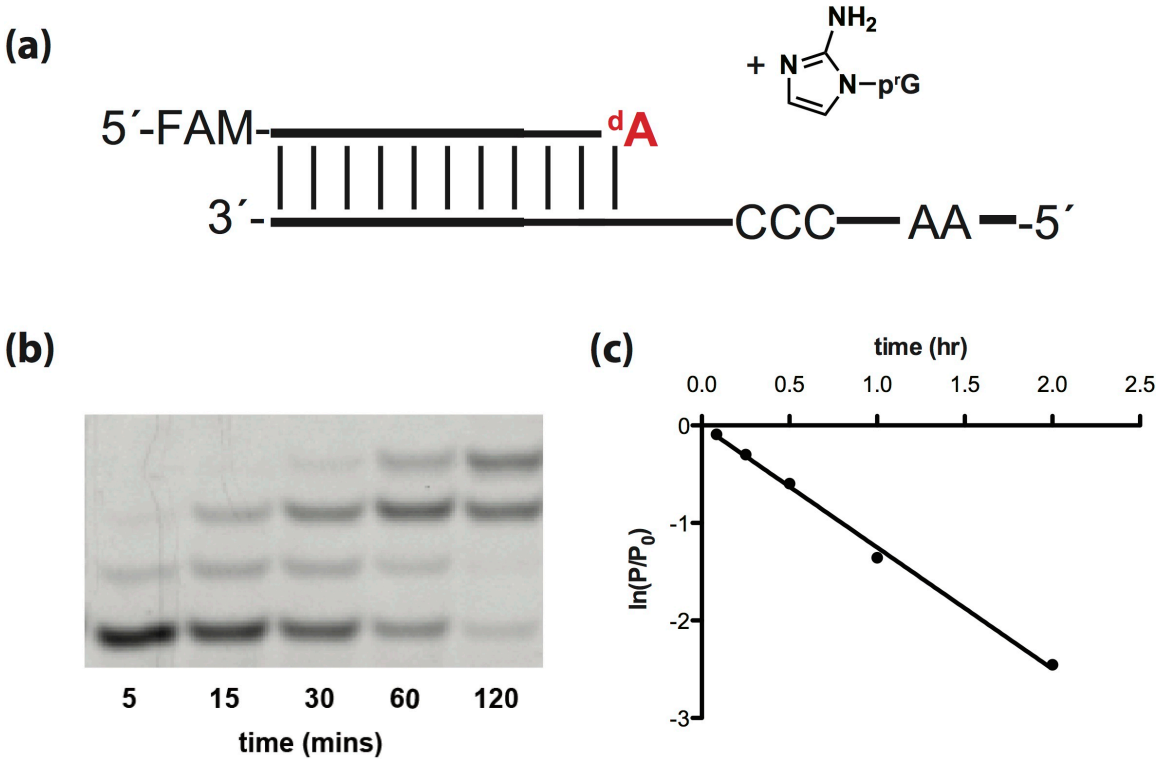
k= 2.75 ± 0.03 hr^{-1}

Figure S.3.30: Primer extension reactions were carried out in triplicate using 20 mM 2AIprG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction rA terminated primer. (b) Representative PAGE analysis of result. (c) Plot of $\ln(P/P_0)$ as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.



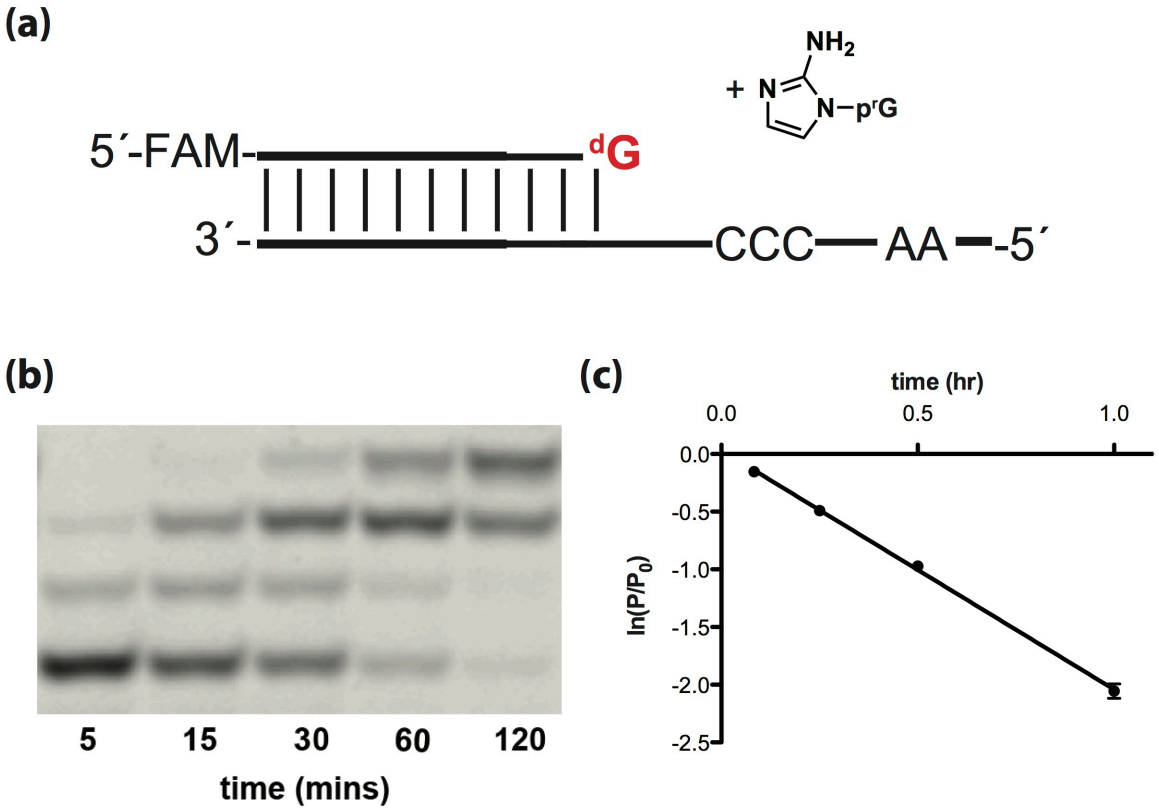
$k = 4.6 \pm 0.1\ hr^{-1}$

Figure S.3.31: Primer extension reactions were carried out in triplicate using 20 mM 2AIprG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction rG terminated primer. (b) Representative PAGE analysis of result. (c) Plot of $\ln(P/P_0)$ as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.



k= 1.25 ± 0.02 hr^{-1}

Figure S.3.32: Primer extension reactions were carried out in triplicate using 20 mM 2AIprG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction dA terminated primer. (b) Representative PAGE analysis of result. (c) Plot of $\ln(P/P_0)$ as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.



$k = 2.07 \pm 0.03\ hr^{-1}$

Figure S.3.33: Primer extension reactions were carried out in triplicate using 20 mM 2AIprG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction dG terminated primer. (b) Representative PAGE analysis of result. (c) Plot of $\ln(P/P_0)$ as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

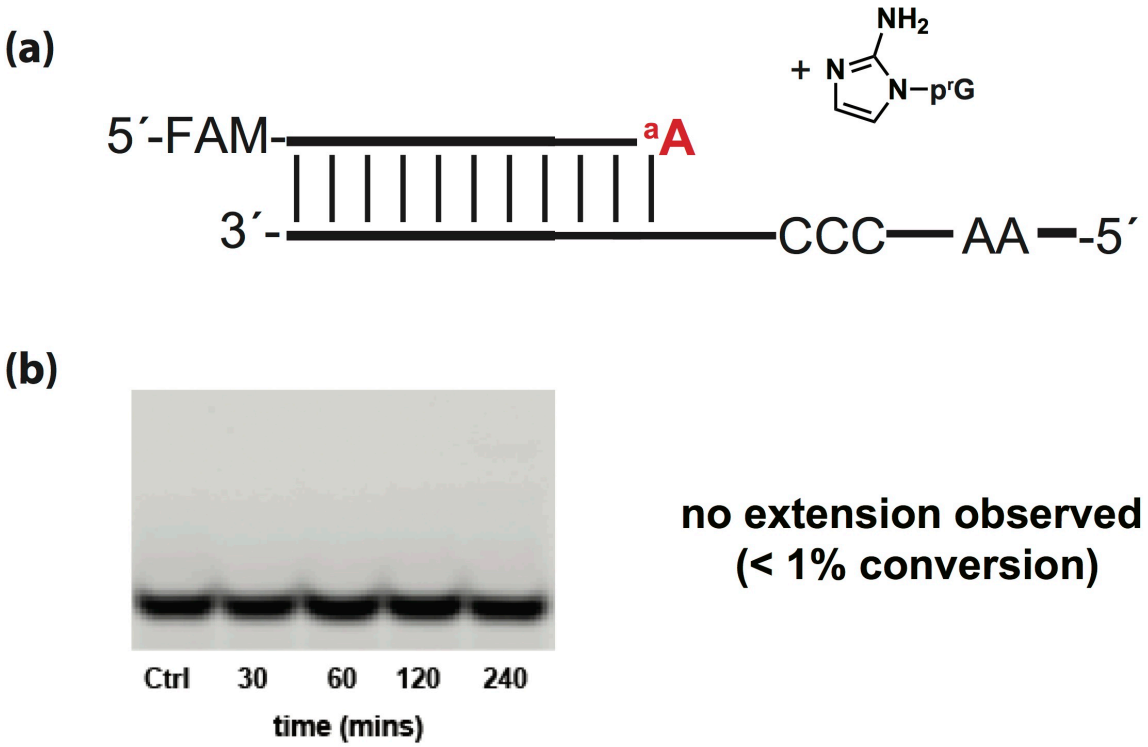


Figure S.3.34: Primer extension reactions were carried out in triplicate using 20 mM 2AIprG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction araA terminated primer. (b) Representative PAGE analysis of result.

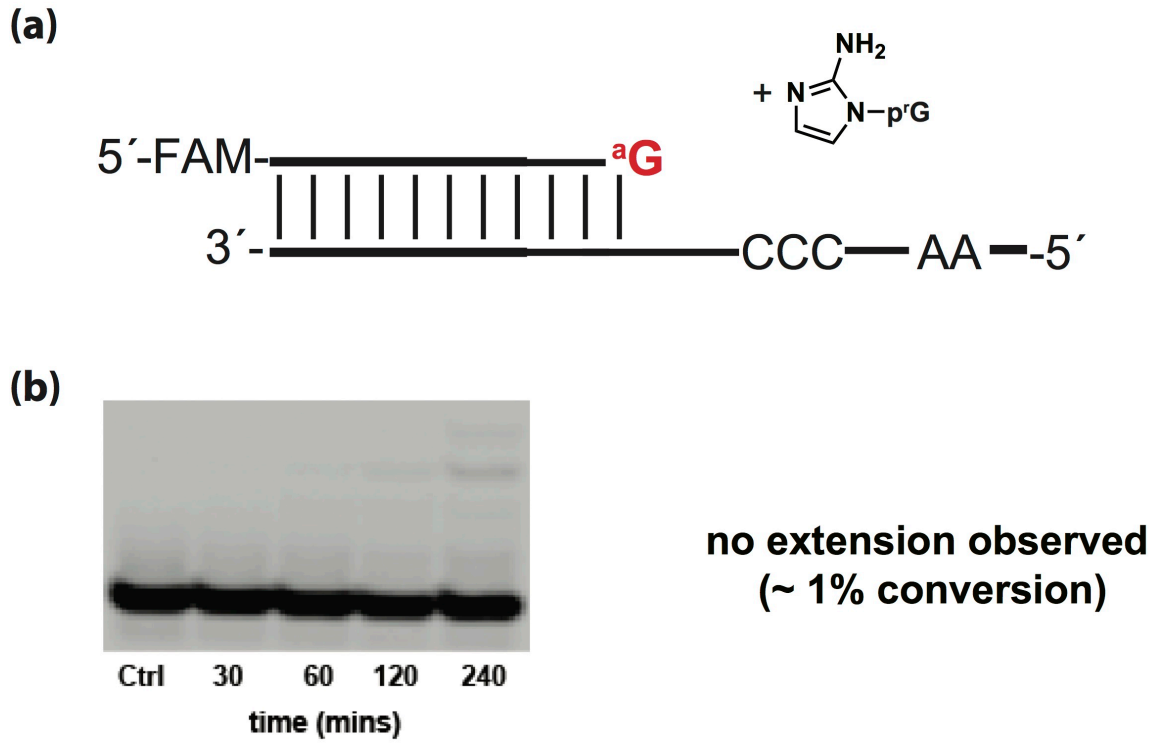


Figure S.3.35: Primer extension reactions were carried out in triplicate using 20 mM 2AIprG, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction araG terminated primer. (b) Representative PAGE analysis of result.

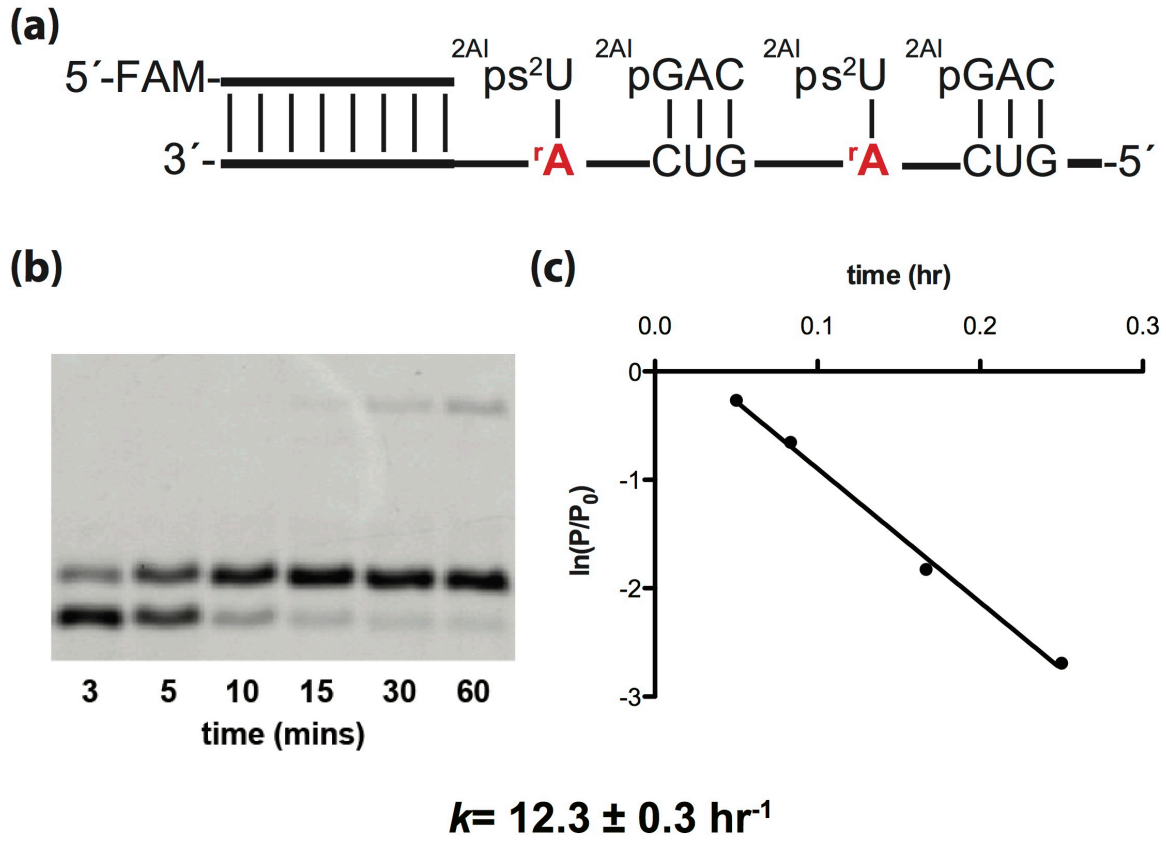


Figure S.3.36: Primer extension reactions were carried out in triplicate using 20 mM 2AIps^{2}U, 0.5 mM 2AIpGAC, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction using rA template. (b) Representative PAGE analysis of result. (c) Plot of $\ln(P/P_0)$ as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

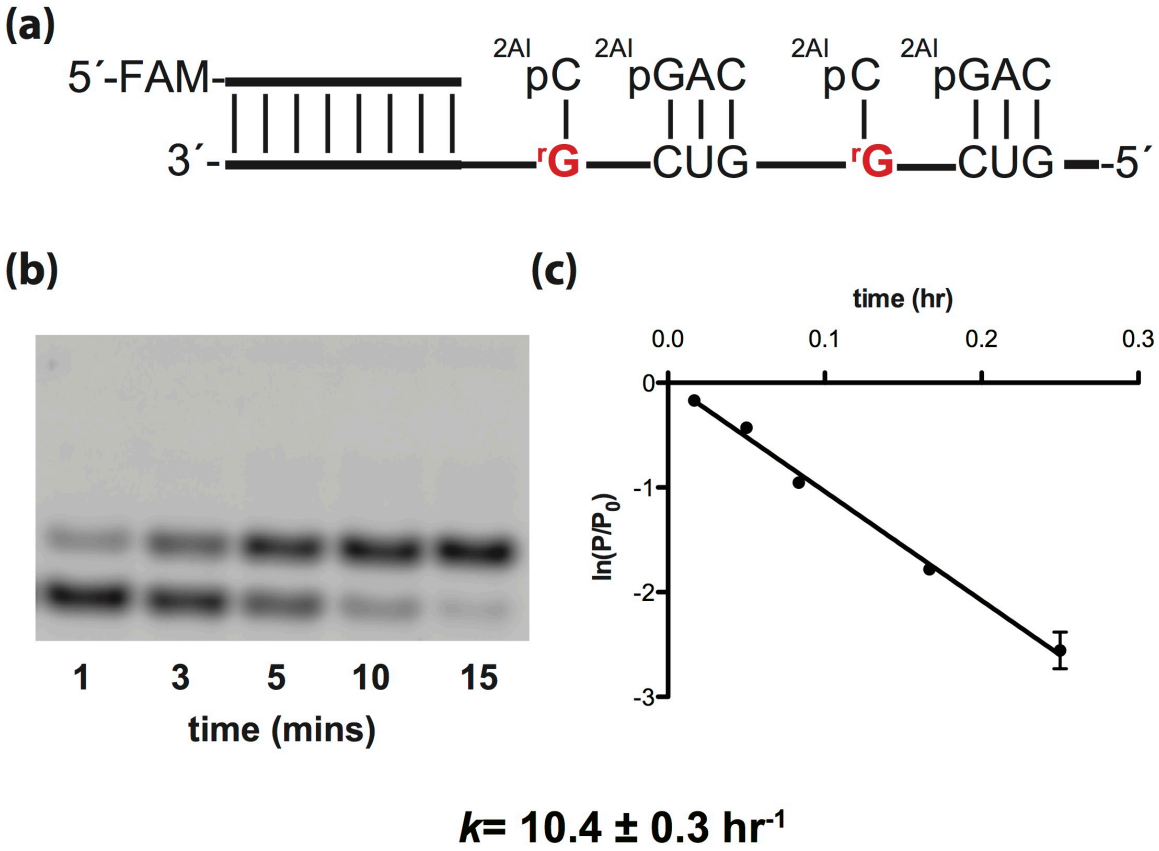
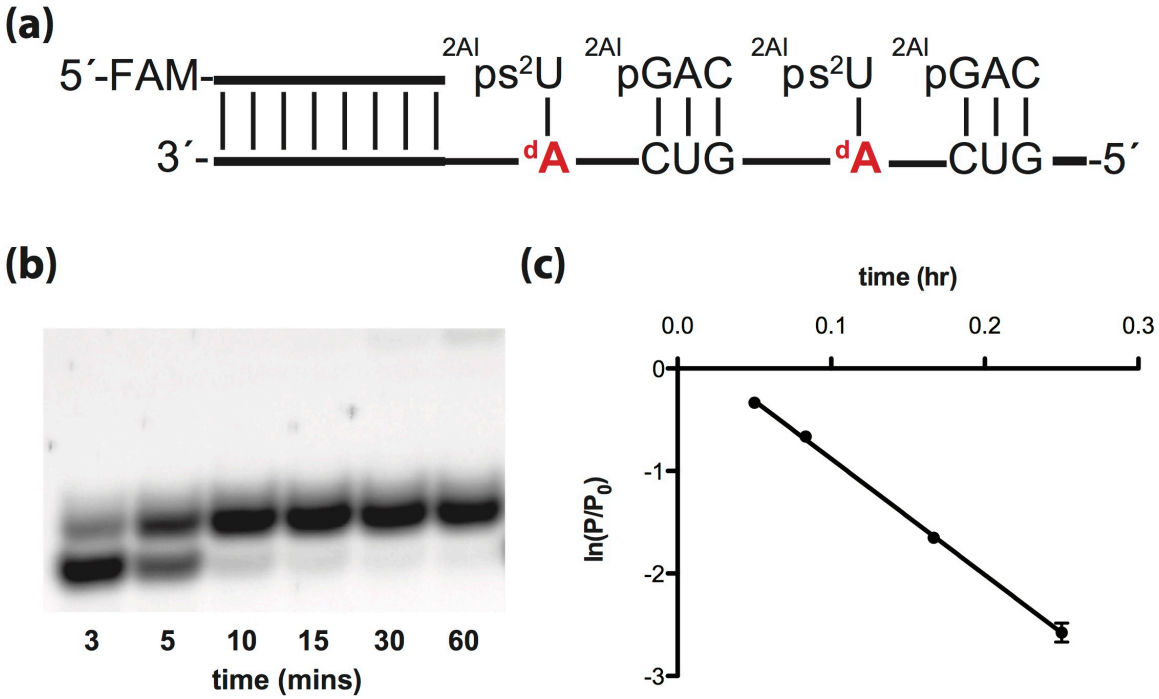
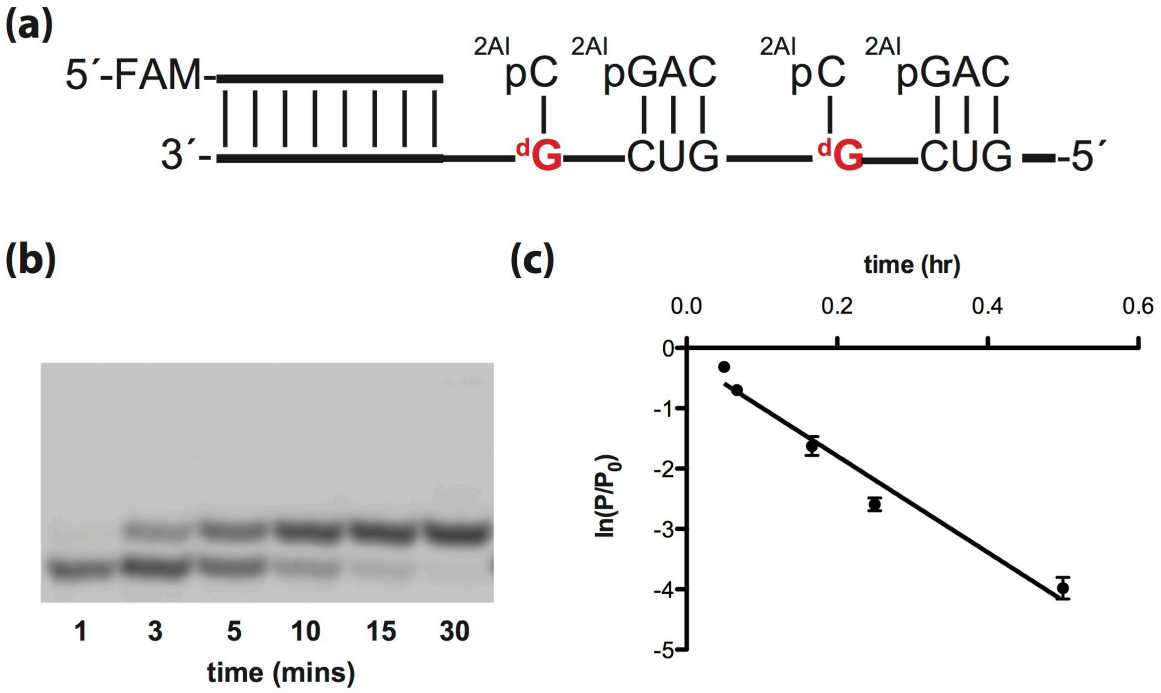


Figure S.3.37: Primer extension reactions were carried out in triplicate using 20 mM 2AIpC, 0.5 mM 2AIpGAC, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction using rG template. (b) Representative PAGE analysis of result. (c) Plot of $\ln(P/P_0)$ as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.



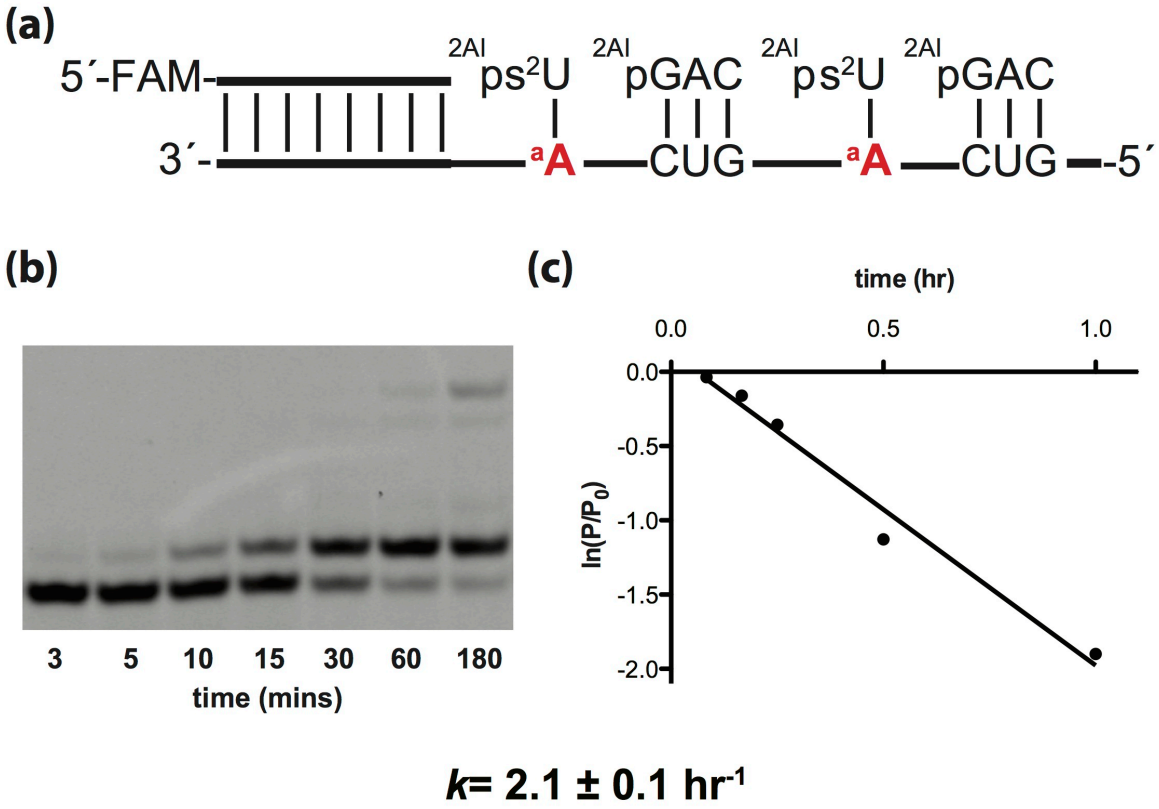
$$k = 11.3 \pm 0.2\ hr^{-1}$$

Figure S.3.38: Primer extension reactions were carried out in triplicate using 20 mM 2AIps^{2}U, 0.5 mM 2AIpGAC, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction using dA template. (b) Representative PAGE analysis of result. (c) Plot of ln(P/P_0) as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.



k= 8.0 ± 0.4 hr^{-1}

Figure S.3.39: Primer extension reactions were carried out in triplicate using 20 mM 2AIpC, 0.5 mM 2AIpGAC, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction using dG template. (b) Representative PAGE analysis of result. (c) Plot of $\ln(P/P_0)$ as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.



k= 2.1 ± 0.1 hr^{-1}

Figure S.3.40: Primer extension reactions were carried out in triplicate using 20 mM 2AIps^{2}U, 0.5 mM 2AIpGAC, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction using araA template. (b) Representative PAGE analysis of result. (c) Plot of ln(P/P_0) as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

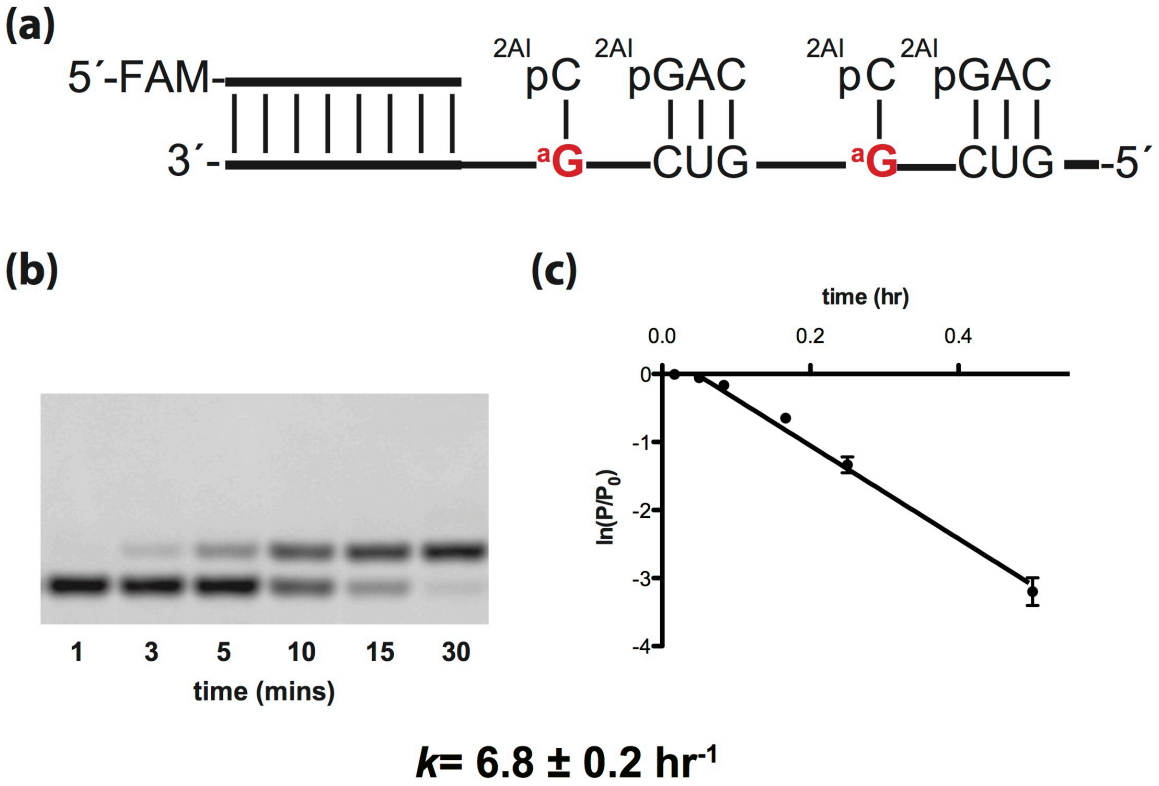


Figure S.3.41: Primer extension reactions were carried out in triplicate using 20 mM 2AIpC, 0.5 mM 2AIpGAC, 50 mM $MgCl_2$, 200 mM Na^+-HEPES pH 8.0. (a) Schematic representation of a primer extension reaction using araG template. (b) Representative PAGE analysis of result. (c) Plot of $\ln(P/P_0)$ as a function of time. The rate of extension was determined from linear least-squares fits of the data from three independent experiments.

3.7 Miscellaneous

Throughout first three years of my graduate studies, the enantioselective synthesis of α-allyl amino esters via hydrogen-bond-donor catalysis was completed[127]. Among many classes of unnatural amino acids, α-allyl amino acids have proven useful in numerous aspects, ranging from the derivatization of allyl functional group, preparation of glycopeptides, and generation of peptide staples and macrocycles[128]. The reported work has numerous advantages from previous works, such as high enantio- and diastereoselectivities, unprecedented catalyst loading of down to 1 mol%, and the practical accessibility of the starting material[127].

[127] Bendelsmith, A. J.; Kim, S. C.; Wasa, M.; Roche, S. P.; Jacobsen, E. N. Enantioselective Synthesis of α-Allyl Amino Esters via Hydrogen-Bond-Donor Catalysis. *J. Am. Chem. Soc.* **2019**, 141, 11414-11419.

[128] a) Lin, Y. A.; Chalker, J. M.; Davis, B. G. Olefin Metathesis for Site-Selective Protein Modification. *ChemBioChem*. **2009**. b) Lang, K.; Chin, J. W. Bioorthogonal Reactions for Labeling Proteins. *ACS Chem. Biol.* **2014**.