Towards Constructing Informational Replicating Systems in Nonenzymatic Environments

Towards Constructing Informational Replicating Systems in Nonenzymatic Environments

A dissertation presented

by

Weicheng Zhang

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

September 2018

Dissertation Advisor: Prof. Jack W. Szostak *Weicheng Zhang*

Towards Constructing Informational Replicating Systems in Nonenzymatic Environments

Abstract

Tremendous efforts have been made in building an artificial protocell that could undergo self-replication and Darwinian evolution. Self-replicating genetic polymers, which serve as major components of protocells, have been a major focus of the origins of life research for many decades. Numerous genetic polymers, especially RNA, have been studied in search of a model that allows effective self-replication in nonenzymatic environments. In this thesis, we aim to discover new alternative genetic polymers with enhanced copying chemistry and explore optimal reaction conditions for efficient copying of RNA.

Chapter 1 summarizes the key findings of the behavior of different types of nucleic acids in nonenzymatic, template-directed polymerization or replication.

Chapter 2 describes the development of a new model system using morpholino nucleic acid (MoNA) to enhance the reaction speed of the nonenzymatic replication of oligonucleotides. The template-directed polymerization of activated ribonucleotide monomers is generally slow because the 3′ primer hydroxyl group has relatively weak nucleophilicity. Previous studies have shown that replacing the hydroxyl group with more nucleophilic amine group generally results in faster primer extension. Herein, we have chosen to study MoNA as it contains a cyclic secondary amine which is expected to be

highly nucleophilic. In this study, we describe the synthesis of 2-methylimidazole activated MoNA monomers from their corresponding ribonucleotides 5′-monophosphates, as well as the synthesis of an RNA primer with a terminal MoNA nucleotide. We discover that activated G and C MoNA monomers enable more rapid and efficient extension of the morpholino-terminated primer on homopolymeric and mixed-sequenced RNA templates.

Chapter 3 describes the importance of the precise position of the hydroxyl at the 3′-end of the primer for the nonenzymatic template-directed primer extension. In this study, we aim to examine the effect of spatial displacement of the hydroxyl nucleophile. RNA primers terminating in either 2′-deoxy-guanosine or 3′-hydroxymethyl-2′,3′-dideoxyguanosine residues were synthesized and their rates in the primer extension reactions were compared. The decreased activity of the hydroxymethyl terminated primer suggests that the spatial preorganization of the deoxyribose and ribose sugars contributes significantly to the efficiency of nonenzymatic primer extension.

Chapter 4 presents a comprehensive analysis of the kinetic profile of nonenzymatic RNA primer extension, the thermodynamic favorability of duplex formation, and the rate of monomer hydrolysis in response to the presence of Mg^{2+}, Mn^{2+}, and $(Co(NH_3)_6^{3+}$. Multivalent metal ions are essential catalyst in the nonenzymatic replication of RNA. Metal ions play multiple key roles in nonenzymatic primer extension, including stabilizing the RNA duplex and activating the nucleophile and/or the electrophile at the reaction site. In this chapter, we demonstrate that the application of cooperative catalysis using metal ions with orthogonal properties results in a synergistic enhancement in reaction kinetics.

Table of Contents

Acknowledgements

I would like to first thank my Ph.D. advisor, Professor Jack Szostak for his ongoing support and exceptional guidance during my graduate studies. I have benefited tremendously from the outstanding research environment build and led by Jack. Coming to Harvard CCB program and working with Jack is one of the greatest decisions I have made in my research career. I was anxious about my scientific abilities, research techniques and my knowledge in prebiotic chemistry when I was doing my rotation in Jack's group. I was extremely surprised that Jack is almost always available to discuss about questions and problems I encounter in my research. Jack gave me the much-needed opportunity and supervision to complete my doctoral training. I am grateful to Jack not only for this valuable chance but also for his extraordinary scientific insights and inspirational guidance that come with his magnificent mentorship. He has been unbelievably kind and patient, giving me opportunities to find solutions, to explore new areas, to face setbacks, and ultimately to learn and discover. My graduate research experience with Jack will continue to shape who I am in the future.

I also want to thank my committee members, Prof. Dr. Gary Ruvkun and Prof. Dr. Radhika Subramanian, for their valuable feedback and tremendous support. Nothing in this thesis would be possible without their support.

I would like to acknowledge all the wonderful and supportive lab members in the Szostak lab. My first mentor in the lab, Ayan Pal, who has been a great friend over the years. I will never forget about all the organic reactions and primer extension experiments we set up together. Aaron Larsen and Tivoli Olsen are the most delightful bay mates. I would like to thank Anders Bjorkbom, Aaron Engelhart, Neha Kamat and Noam Prywes

for sharing their knowledge and teaching me new techniques. Shenglong Zhang, Li Li, Lin Jin, Anna Wang, Derek K. O'Flaherty, Kyle Strom, Constantin Giurgiu, Chris Kim, Travis Walton, Tivoli J. Olsen, Christian Hentrich, Tony Jia, Zoe Todd, Craig Blain, Daniel Duzdevich, and Enver Izgu, among many others, have been really helpful in scientific discussions. Fanny Ng, Pamela Svec, and Chris Kelleher are doing an excellent job in keeping the lab running.

Lastly, I wish to thank my family for their unconditional support and encouragement throughout my study.

Chapter 1

Towards constructing informational replicating systems

1.1 Introduction

How life originated on earth is still one of the biggest unsolved puzzles in science. It is hard to envision the transition from lifeless small molecules to sophisticated DNA-inheriting cells. The biological approach to this problem, also known as the top-down approach, focuses on the common ancestor of modern organisms.[1] Advancement in phylogenetic methods gives information about the genes that might be present in the last universal common ancestor (LUCA) and the physiology of LUCA.[2-3] However, currently recognized LUCA is already a complex life form that contains a rather significant amount of genetic material and relies on complex enzyme catalyzed functions, such as replication,[4] transcription,[5] translation, and energy metabolism.[3, 6] How LUCA arose from prebiotically plausible molecules without enzyme catalysis is still a question that needs to be addressed from a chemical perspective.[7] There are two major steps in the bottom-up approach: 1) the prebiotic synthesis of all necessary building blocks of a genetic system; 2) the assembly of those building blocks into a functional self-replicating system with the ability to evolve. In order to proceed with these two steps, one must identify this system's core building block, the information carrier. In order to be considered as a viable medium for information carrying, this material must be capable of promoting template-dependent nonenzymatic

polymerization. In this chapter, I will review different types of nucleic acids that have been considered as potential genetic materials. This review will focus on their behavior in nonenzymatic, template-directed polymerization or replication.

1.2 RNA

With the discovery of ribozymes in the early 1980s,[8-9] RNA-based systems, where RNA operates as both genetic material and catalyst, were proposed to emerge before the modern genetic system. This theory was then referred to as the "RNA World" hypothesis.[10] Although a ribozyme capable of catalyzing its own replication is yet to be found,[11] tremendous progress has been made towards the search for ribozymes that could efficiently catalyze the replication of RNA.[12-14] The origins of the first ribozyme is another piece of the puzzle that must be solved in order to validate the RNA World hypothesis. One of the plausible ways to obtain the first ribozyme is through the nonenzymatic templated-directed copying of RNA, which enables the replication of a large population of random-sequence oligonucleotides encapsulated in replicating protocells. Herein may allow some rare nucleic acid sequences that confer a selective advantage on their protocell to take over the population.

Experimental attempts to replicate RNA nonenzymatically began more than half a century ago.[11] The copying of oligonucleotides usually uses activated mononucleotides as substrates. However, nucleoside 5′-triphosphates (NTP), the extant substrates for enzyme-catalyzed oligonucleotide synthesis, are unreactive in a nonenzymatic environment.[15] This stimulated the discovery of more reactive alternatives, such as ribonucleotides activated with heterocycles. In 1968, nucleoside 5′-

phosphorimidazolides were shown to be viable substrates in the nonenzymatic template-directed polymerization of RNA.[16] In 1981, Inoue and Orgel tested the poly(C)-directed oligomerization of the derivatives of guanosine 5′-phosphorimidazolide and discovered that 2-methylimidazole (2-MeIm) is a far more efficient activating group.[17] In addition, products from the oligomerization of nucleoside 5′-phosphor-2-methylimidazolide were predominantly 3′,5′-linked oligomers.[17] Mononucleotides activated with pyridine,[18] and oxyazabenzotriazole (OAt)[19-20] also have sufficient reactivity for the copying of RNA in the absence of enzymes. Recently Li and his colleagues evaluated a larger pool of substituted phosphorimidazolides as substrates in the nonenzymatic copying of RNA and noted that the rate of reaction increases when phosphorimidazolide derivatives with a higher pK_a and a less bulky substituent are used. In this study, 2-aminoimidazole activated ribonucleotides were reported to allow RNA copying to proceed with superior rates and improved yields.[21]

The imbalanced rate of incorporation of different nucleobases is another major problem that stands in the way of efficient nonenzymatic RNA copying. Template copying with activated A and U monomers can be more than 100 fold slower than that with activated G and C monomers.[22-23] More concerning still, a significant portion of the activated monomers is hydrolyzed during the time required for 2-MeImpU to be added to the growing primer. Furthermore, in systems using mixed sequence templates, G:U and A:C wobble pairing can lead to mismatches. Mismatches can occur at rates as high as once for every 5–6 monomers incorporated.[24] Some mismatches have been demonstrated to strongly inhibit subsequent copying.[25] Modified nucleobase monomers offer a partial solution to these challenges. 2-thiouridine (s^2U) and 2-thioribothymidine

(s^2T) are nucleosides that can form more stable base pairs with A than U.[26] The G:U wobble pairing is disturbed by the single atom substitution of oxygen by sulfur.[27] The nonenzymatic template-directed primer extension using activated s^2U and s^2T monomers proceeds with improved rate and fidelity.[28] Other methods have been developed to address the inefficient copying of mixed sequence templates. Richert and his colleagues have demonstrated that downstream-binding RNA strands can accelerate the replication of RNA by assisting in the binding of activated monomer to the template.[29] Mixed-sequence RNA templates can be copied in high-yield by immobilizing the RNA duplex on magnetic iron oxide beads and periodically replacing the activating monomers and helper RNAs. Further improvement in this approach has been achieved by utilizing activated trinucleotide microhelpers.[30]

1.3 Alternative Genetic Materials

The challenges in creating an efficient model of nonenzymatic RNA replication suitable for a protocell environment have led to the search for alternative genetic molecules. Here I will review alternative genetic materials that have been examined in the context of nonenzymatic template-directed primer extension.

1.3.1 Phosphodiester Linkages

Hexitol nucleic acids (HNAs) are built up with a phosphorylated 1,5-anhydrohexitol backbone and standard nucleobases.[31] (Figure 1.1) Its six-membered ring is structurally similar to the furanose ring in its 2'-exo,3'-endo conformation.[32] HNA oligomers can hybridize with complementary DNA or RNA oligomers. The A-type

conformation of the resulting HNA/DNA and HNA/RNA duplexes has been confirmed by CD, NMR, and X-ray studies.[32-33] Altritol nucleic acid (ANA) is an analog of HNA with an additional hydroxyl group. (Figure 1.1) ANA can form more stable A-type double helical duplexes with DNA and RNA than the corresponding HNA. The additional hydroxyl group of ANA increases the hydration of the groove and therefore contributes to the stability of the duplex.[34]

The performance of HNA and ANA in nonenzymatic template-directed reactions was evaluated by the Orgel laboratory.[35] HNA and ANA are superior templates in the oligomerization of activated monoribonucleotides when compared to DNA and RNA. The superiority of HNA and ANA templates likely originates from their extensive pre-organization for the A structure. In contrast, activated HNA and ANA monomers (2-MeImpHG and 2-MeImpAG) failed to polymerize efficiently in the presence of all 4 different types of template. This is possibly caused by the unfavorable alignment between the hydroxyl group of the primer with activated phosphate of the monomer.[35]

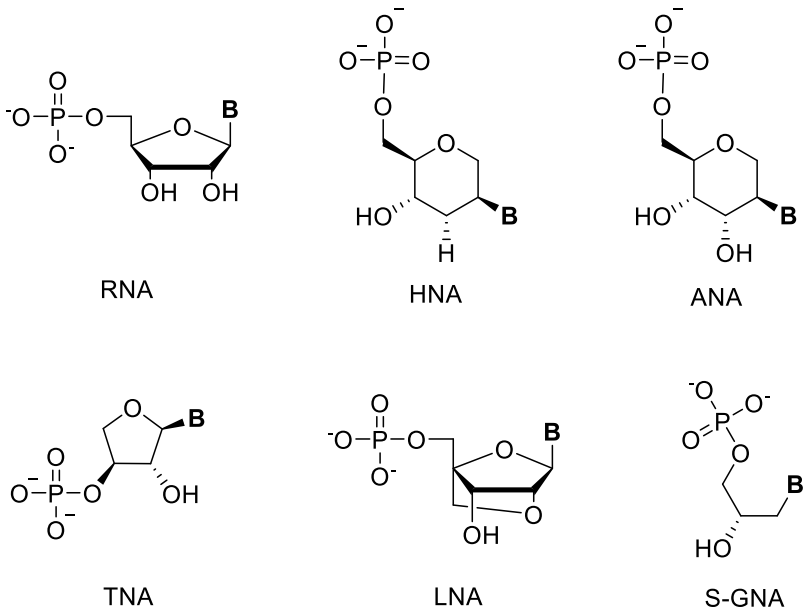


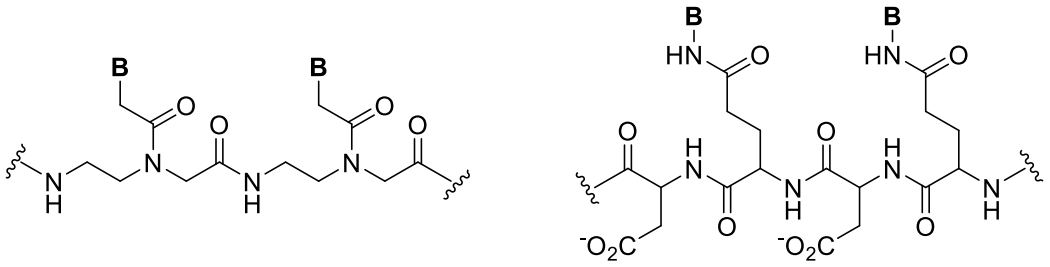
Figure 1.1: Alternative genetic materials with phosphodiester backbone linkages.

Threose nucleic acid (TNA) is one of the most well-known artificial genetic polymers introduced by Eschenmoser and his coworkers. Despite the fact that the backbone of TNA is shorter than that of RNA and DNA, TNA can cross pair with RNA and DNA to form heteroduplexes.[36] The use of a TNA template in the nonenzymatic oligomerization of RNA was reported by Heuberger and Switzer in 2006.[37] Although TNA did direct the polymerization of 2-MeImpG, TNA was found to be a less efficient template than both DNA and RNA. A series of other alternative nucleic acids that utilize phosphodiester linkages have since been discovered and synthesized. Many of them have been demonstrated to cross pair with RNA or DNA via Watson-Crick base pairs, such as locked nucleic acid (LNA), or to self-assemble into duplexes, such as glycol nucleic acid (GNA). However, these alternative backbones have not been carefully explored in terms of nonenzymatic polymerization. So far, no genetic molecule with phosphodiester backbone linkages has been found to be copied more effectively than RNA in enzyme-free reactions.

1.3.2 Amide Linkages

Peptide nucleic acids (PNA) were initially designed to target genes for antisense inhibition.[38] It can interact with many nucleic acids via Watson-crick base-pairs. The simple achiral structure of PNA makes it more accessible under prebiotic conditions.[39] Its amide backbone linkages are more resistant to hydrolysis than the phosphodiester linkages seen in DNA and RNA. PNA monomers are prone to cyclization, however, with the assistance of the condensing reagent EDC, PNA dimers can be used as the substrate in nonenzymatic DNA template-directed polymerization.[40] Furthermore, PNA

oligomers can act as templates to catalyze the polymerization of activated RNA monomers.[41] The uncharged backbone of the original PNA often leads to aggregation. To avoid this problem, Mittapalli and his coworkers introduced a dipeptide-based oligomer system (agPNA) which contains charged aspartate side chains.[42]



PNA agPNA

Figure 1.2: Structure of PNA and agPNA

1.3.3 Phosphoramidate Linkages

Replacing the hydroxyl nucleophile of RNA with the more nucleophilic amine group could enhance the rate of the nonenzymatic template copying. 3′-amino-3′-deoxynucleosides are synthesized in 1978 as a tool to enable the direct aminoacylation of tRNA.[43-44] Leslie Orgel and his colleagues compared the reactivity of 3′-amino-3′-deoxyuridine and uridine in the nucleophilic attack on ribonucleoside-5′-phosphorimidazolides and established that 3' amino group is far superior in reactivity compared to 3' hydroxyl group. They also showed that both RNA and DNA templates allow the phosphorimidazolide derivative of 3′-amino-3′-deoxynucleotide monomers to polymerize efficiently.[45-46]

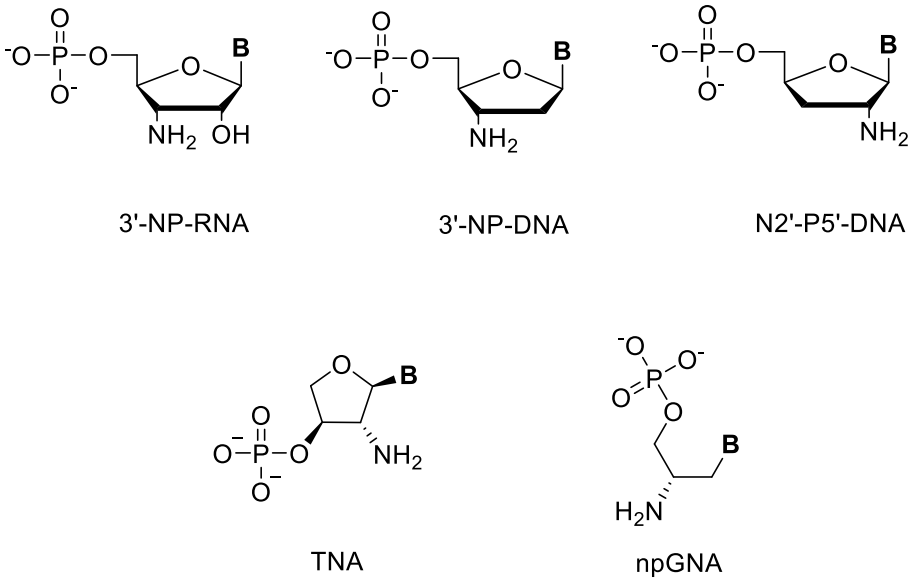


Figure 1.3: Alternative genetic materials with phosphoramidate backbone linkage.

In 2013, Zhang and his co-workers reported the synthesis of activated 3′-NP-DNA monomers with all four standard nucleobases. These monomers copy both RNA and DNA templates efficiently with the catalytic assistance of 2-hydroxyethyl-imidazole (HEI). During the copying of certain 4-nt template sequences under optimal conditions, quantitative conversion to the full-length product can be achieved within 5 minutes.[47] However, intramolecular monomer cyclization is promoted by the presence of HEI. With HEI, more than 90 percent of the activated 3′-NP-DNA monomers are converted to the unreactive 3′-5′ cyclized monomers within a few hours.[47-48] Similar to RNA systems, s^2U and s^2T derivatives of the activated 3′-NP-DNA monomers are significantly better substrates in the nonenzymatic copying of templates than the native activated U monomer.[49]

2′-NP-DNA is another system that has a large potential to function in a self-replicating model system. Imidazole activated 2′-amino-2′,3′-dideoxy-guanosine and -

cytidine (2′-NH_2-ImpddG and 2′-NH_2-ImpddC) can both be efficiently incorporated onto the end of primers under the direction of RNA, LNA, and 2'-5'-DNA templates.[50] The template-directed addition of 2′-NH_2-ImpddA and 2′-NH_2-ImpddU to primers are relatively slow when compared to G and C monomers. However, the application of C5-(1-propynyl)-U and diaminopurine nucleobase modifications can reduce the imbalance in rates in the primer extension reactions. Another advantage of the 2′-NP-DNA system is that the intramolecular cyclization of monomers is almost impossible due to ring strain.[50] A recent report shows that activated 2′-NH_2-ddG monomers can efficiently copy homopolymeric templates inside fatty acid vesicles in the presence of a low concentration of Mg^{2+}.[51]

The synthesis of 2′-amino modified TNA (2′-NH_2-TNA) nucleosides was reported recently. Interestingly, the rates of the nonenzymatic primer extension using activated 2′-NH_2-TNA monomer are much lower than the rates reported in the 2′-NP-DNA and 3′-NP-DNA systems. In fact, the rates of the template copying with activated 2′-NH_2-TNA monomer and the rates of the hydrolysis of activated 2′-NH_2-TNA monomers are not significantly different.[52] An analog of GNA with N2′-P3′ phosphoramidate linkages (npGNA) has also been evaluated. npGNA can form stable duplexes with its complementary npGNA. Activated npGNA monomers undergo rapid intramolecular cyclization, and therefore, are not good substrates for the primer extension. However, activated 2′-amino GNA dinucleotides are stable in solution and can ligate in a template-dependent manner to yield npGNA oligonucleotides.[53]

1.4 Prospective

Although there are some experimental demonstrations of the nonenzymatic template-directed copying of RNA, significant improvements in efficiency and accuracy are required to construct a robust model of RNA replication. Some of the alternative polymers reviewed in this chapter have significant advantages in replication reactions compared to RNA. However, deficiencies associated with those polymers, such as rapid cyclization and hydrolysis, must be addressed in order for them to function as effective genetic materials. The search for new alternative genetic polymers with superior chemical and physical properties and enhanced copying chemistry, as well as the search for reaction conditions leading to the enhanced copying of RNA, remain as active research fields, and that this has been the focus of the research reported in the following chapters.

1.5 References

1. Koonin, E. V., Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nature Reviews Microbiology* **2003,** *1*, 127.

2. Weiss, M. C.; Preiner, M.; Xavier, J. C.; Zimorski, V.; Martin, W. F., The last universal common ancestor between ancient Earth chemistry and the onset of genetics. *PLOS Genetics* **2018,** *14* (8), e1007518.

3. Weiss, M. C.; Sousa, F. L.; Mrnjavac, N.; Neukirchen, S.; Roettger, M.; Nelson-Sathi, S.; Martin, W. F., The physiology and habitat of the last universal common ancestor. *Nature Microbiology* **2016,** *1*, 16116.

4. Koonin, E. V.; Martin, W., On the origin of genomes and cells within inorganic compartments. *Trends in Genetics* **2005,** *21* (12), 647-654.

5. Burton, Z. F.; Opron, K.; Wei, G.; Geiger, J. H., A model for genesis of transcription systems. *Transcription* **2016,** *7* (1), 1-13.

6. Schlegel, K.; Leone, V.; Faraldo-Gómez, J. D.; Müller, V., Promiscuous archaeal ATP synthase concurrently coupled to Na+ and H+ translocation. *Proceedings of the National Academy of Sciences* **2012,** *109* (3), 947.

7. Sutherland, J. D., Opinion: Studies on the origin of life — the end of the beginning. *Nature Reviews Chemistry* **2017,** *1*, 0012.

8. 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,** *31* (1), 147-157.

9. 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,** *35* (3), 849-857.

10. Gilbert, W., Origin of life: The RNA world. *Nature* **1986,** *319*, 618.

11. Szostak, J. W., The Narrow Road to the Deep Past: In Search of the Chemistry of the Origin of Life. *Angew Chem Int Ed Engl* **2017,** *56* (37), 11037-11043.

12. Johnston, W. K.; Unrau, P. J.; Lawrence, M. S.; Glasner, M. E.; Bartel, D. P., RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension. *Science* **2001,** *292* (5520), 1319.

13. Wochner, A.; Attwater, J.; Coulson, A.; Holliger, P., Ribozyme-Catalyzed Transcription of an Active Ribozyme. *Science* **2011,** *332* (6026), 209.

14. Horning, D. P.; Joyce, G. F., Amplification of RNA by an RNA polymerase ribozyme. *Proceedings of the National Academy of Sciences* **2016,** *113* (35), 9786.

15. Orgel, L. E.; Lohrmann, R., Prebiotic chemistry and nucleic acid replication. *Accounts of Chemical Research* **1974,** *7* (11), 368-377.

16. Weimann, B. J.; Lohrmann, R.; Orgel, L. E.; Schneider-Bernloehr, H.; Sulston, J. E., Template-directed synthesis with adenosine-5'-phosphorimidazolide. *Science* **1968,** *161* (3839), 387.

17. Inoue, T.; Orgel, L. E., Oligomerization of (guanosine 5′-phosphor)-2-methylimidazolide on poly(C): An RNA polymerase model. *Journal of Molecular Biology* **1982,** *162* (1), 201-217.

18. Lohrmann, R.; Orgel, L. E., Preferential formation of (2'–5')-linked internucleotide bonds in non-enzymatic reactions. *Tetrahedron* **1978,** *34* (7), 853-855.

19. Hagenbuch, P.; Kervio, E.; Hochgesand, A.; Plutowski, U.; Richert, C., Chemical primer extension: efficiently determining single nucleotides in DNA. *Angew Chem Int Ed Engl* **2005,** *44* (40), 6588-92.

20. Kervio, E.; Sosson, M.; Richert, C., The effect of leaving groups on binding and reactivity in enzyme-free copying of DNA and RNA. *Nucleic Acids Research* **2016,** *44* (12), 5504-5514.

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

22. Joyce, G. F., Nonenzymatic template-directed synthesis of informational macromolecules. *Cold Spring Harb Symp Quant Biol* **1987,** *52*, 41-51.

23. Wu, T.; Orgel, L. E., Nonenzymatic template-directed synthesis on hairpin oligonucleotides. 3. Incorporation of adenosine and uridine residues. *Journal of the American Chemical Society* **1992,** *114* (21), 7963-7969.

24. Hagenbuch, P.; Kervio, E.; Hochgesand, A.; Plutowski, U.; Richert, C., Chemical Primer Extension: Efficiently Determining Single Nucleotides in DNA. *Angewandte Chemie International Edition* **2005,** *44* (40), 6588-6592.

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

26. 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 Research* **1997,** *25* (6), 1272-1280.

27. 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,** *38* (50), 16655-62.

28. 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,** *137* (7), 2769-75.

29. Vogel, S. R.; Deck, C.; Richert, C., Accelerating chemical replication steps of RNA involving activated ribonucleotides and downstream-binding elements. *Chem Commun (Camb)* **2005,** (39), 4922-4.

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

31. 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. *Journal of the American Chemical Society* **1999,** *121* (12), 2653-2656.

32. Winter, H. D.; Lescrinier, E.; Aerschot, A. V.; Herdewijn, P., Molecular Dynamics Simulation To Investigate Differences in Minor Groove Hydration of HNA/RNA Hybrids As Compared to HNA/DNA Complexes. *Journal of the American Chemical Society* **1998,** *120* (22), 5381-5394.

33. Hendrix, C.; Rosemeyer, H.; De Bouvere, B.; Van Aerschot, A.; Seela, F.; Herdewijn, P., 1′,5′-Anhydrohexitol Oligonucleotides: Hybridisation and Strand Displacement with Oligoribonucleotides, Interaction with RNase H and HIV Reverse Transcriptase. *Chemistry – A European Journal* **1997,** *3* (9), 1513-1520.

34. Allart, B.; Khan, K.; Rosemeyer, H.; Schepers, G.; Hendrix, C.; Rothenbacher, K.; Seela, F.; Van Aerschot, A.; Herdewijn, P., D-Altritol Nucleic Acids (ANA): Hybridisation Properties, Stability, and Initial Structural Analysis. *Chemistry – A European Journal* **1999,** *5* (8), 2424-2431.

35. Kozlov, I. A.; Zielinski, M.; Allart, B.; Kerremans, L.; Van Aerschot, A.; Busson, R.; Herdewijn, P.; Orgel, L. E., Nonenzymatic Template-Directed Reactions on Altritol Oligomers, Preorganized Analogues of Oligonucleotides. *Chemistry – A European Journal* **2000,** *6* (1), 151-155.

36. Brudno, Y.; Liu, D. R., Recent Progress Toward the Templated Synthesis and Directed Evolution of Sequence-Defined Synthetic Polymers. *Chemistry & biology* **2009,** *16* (3), 265-276.

37. Heuberger, B. D.; Switzer, C., Nonenzymatic Oligomerization of RNA by TNA Templates. *Organic Letters* **2006,** *8* (25), 5809-5811.

38. Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O., Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* **1991,** *254* (5037), 1497.

39. Nielsen, P. E., Peptide Nucleic Acids and the Origin of Life. *Chemistry & Biodiversity* **2007,** *4* (9), 1996-2002.

40. Schmidt, J. G.; Christensen, L.; Nielsen, P. E.; Orgel, L. E., Information transfer from DNA to peptide nucleic acids by template-directed syntheses. *Nucleic Acids Research* **1997,** *25* (23), 4792-4796.

41. Schmidt, J. G.; Nielsen, P. E.; Orgel, L. E., Information transfer from peptide nucleic acids to RNA by template-directed syntheses. *Nucleic Acids Research* **1997,** *25* (23), 4797-4802.

42. Mittapalli, G. K.; Osornio, Y. M.; Guerrero, M. A.; Reddy, K. R.; Krishnamurthy, R.; Eschenmoser, A., Mapping the Landscape of Potentially Primordial Informational Oligomers: Oligodipeptides Tagged with 2,4-Disubstituted 5-Aminopyrimidines as Recognition Elements. *Angewandte Chemie International Edition* **2007,** *46* (14), 2478-2484.

43. Azhayev, A. V.; Ozols, A. M.; Bushnev, A. S.; Dyatkina, N. B.; Kochetkova, S. V.; Victorova, L. S.; Kukhanova, M. K.; Krayevsky, A. A.; Gottikh, B. P., Aminonucleosides and their derivatives. IV I Synthesis of the 3′-amino-3′-deoxynucleoside 5′-phosphates. *Nucleic Acids Research* **1979,** *6* (2), 625-643.

44. Ozols, A. M.; Azhayev, A. V.; Dyatkina, N. B.; Krayevsky, A. A., Aminonucleosides and their Derivatives; VI1. A New Synthesis of 1,2, 5-Tri-O-acyl-3-azido-3-deoxy-β-D-ribofuranose. *Synthesis* **1980,** *1980* (07), 557-559.

45. Zielinski, W. S.; Orgel, L. E., Oligomerization of activated derivatives of 3'-amino-3'-deoxyguanosine on poly(C) and poly(dC) templates. *Nucleic Acids Res* **1985,** *13* (7), 2469-84.

46. Tohidi, M.; Zielinski, W. S.; Chen, C. H.; Orgel, L. E., Oligomerization of 3'-amino-3'deoxyguanosine-5'phosphorimidazolidate on a d(CpCpCpCpC) template. *J Mol Evol* **1987,** *25*, 97-9.

47. Zhang, S.; Zhang, N.; Blain, J. C.; Szostak, J. W., Synthesis of N3'-P5'-linked phosphoramidate DNA by nonenzymatic template-directed primer extension. *J Am Chem Soc* **2013,** *135* (2), 924-32.

48. Zhang, S.; Blain, J. C.; Zielinska, D.; Gryaznov, S. M.; Szostak, J. W., Fast and accurate nonenzymatic copying of an RNA-like synthetic genetic polymer. *Proc Natl Acad Sci U S A* **2013,** *110* (44), 17732-7.

49. Izgu, E. C.; Oh, S. S.; Szostak, J. W., Synthesis of activated 3′-amino-3′-deoxy-2-thio-thymidine, a superior substrate for the nonenzymatic copying of nucleic acid templates. *Chemical Communications* **2016,** *52* (18), 3684-3686.

50. 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,** *131* (40), 14560-70.

51. O'Flaherty, D. K.; Kamat, N. P.; Mirza, F. N.; Li, L.; Prywes, N.; Szostak, J. W., Copying of Mixed-Sequence RNA Templates inside Model Protocells. *Journal of the American Chemical Society* **2018,** *140* (15), 5171-5178.

52. Blain, J. C.; Ricardo, A.; Szostak, J. W., Synthesis and nonenzymatic template-directed polymerization of 2'-amino-2'-deoxythreose nucleotides. *J Am Chem Soc* **2014,** *136* (5), 2033-9.

53. Chen, J. J.; Cai, X.; Szostak, J. W., N2'-->p3' phosphoramidate glycerol nucleic acid as a potential alternative genetic system. *J Am Chem Soc* **2009,** *131* (6), 2119-21.

Chapter 2

Template-directed nonenzymatic primer extension using 2-methylimidazole-activated morpholino derivatives of guanosine and cytidine

[*W. Zhang, A. Pal and J. Szostak conceived of experiments, analyzed the data and wrote the text. W. Zhang and A. Pal performed experiments. A version of this chapter will be submitted for peer review and publication.*]

2.1 Abstract

Efforts to develop self-replicating nucleic acids have led to insights into the origin of life, and have also suggested potential pathways to the design of artificial life forms based on non-natural nucleic acids. The template-directed nonenzymatic polymerization of activated ribonucleotide monomers is generally slow because of the relatively weak nucleophilicity of the primer 3′-hydroxyl. To overcome this problem, several nucleic acids based on amino-sugar nucleotides have been studied, and as expected the more nucleophilic amine generally results in faster primer extension. Extending this logic, we have chosen to study morpholino nucleic acid (MoNA), because the cyclic secondary amine of the morpholino nucleotides is expected to be highly nucleophilic. We describe the synthesis of 2-methylimidazole activated MoNA monomers from their corresponding ribonucleotides 5′-monophosphates, and the synthesis of an RNA primer with a terminal MoNA nucleotide. We show that the activated G and C MoNA monomers enable rapid

and efficient extension of the morpholino-terminated primer on homopolymeric and mixed-sequenced RNA templates. Our results show that MoNA is an interesting non-natural informational polymer that is worthy of further study as a candidate self-replicating material.

2.2 Introduction

Constructing a model protocell with life-like properties that is capable of Darwinian evolution is one of the fundamental challenges in understanding the transition from chemical to biological evolution in the prebiotic environment of the early Earth.[1] With this goal in mind, significant efforts have been made to develop two mutually compatible and supportive self-replicating systems: membrane compartments to provide spatial localization, and a genetic polymer to serve as an informational and functional material. Both systems must replicate without the assistance of any complex protein enzymes or other evolved machinery. [2-5] Although membranes composed of fatty acids and related amphiphiles have been shown to be a promising basis for self-sustaining and self-replicating vesicles,[6] an efficient self-replicating genetic polymer is yet to be established. Many lines of evidence suggest that RNA is a reasonable candidate for the primordial genetic system owing to its ability to both encode genetic information and to carry out catalytic functions.[7-8] One possibility is that an initial phase of nonenzymatic chemical replication was sufficient to enable the evolution of the first functional ribozyme, which in turn led to strong selective pressure for increased RNA replication efficiency, and thus the gradual evolution of ribozyme catalyzed RNA replication.[9] Decades of research have established a working model of nonenzymatic template copying involving RNA primer

extension with activated nucleoside monophosphates (NMPs).[5, 10] However, despite recent progress, this approach remains limited to the copying of very short mixed sequence RNA templates, and improving the rate and extent of template copying is still a major challenge.[11] To address this problem, one strategy that has been explored is the use of nucleotide analogs that position a stronger amine nucleophile at the terminus of the growing primer chain.

Leslie Orgel and his colleagues showed that replacing the hydroxyl group with a more nucleophilic amine group at the 3′ position of uridine enhances the non-templated solution phase formation of dinucleotides from reaction with ribonucleoside-5′-phosphorimidazolides by one to two orders of magnitude. They also observed that 3′-amino-3′-deoxyguanosine-5′-phosphorimidazolides condense on both poly(C) and poly(dC) templates whereas the guanosine-5′-phosphor-imidazolide does not condense efficiently under similar conditions.[12-13] 3′-amino-2′,3′-dideoxyribonucleoside-5′-phosphor-2-methylimidazolides were found to polymerize efficiently on RNA templates with the catalytic assistance of 2-hydroxyethyl-imidazole (HEI).[14] However, in the presence of HEI the rate of intramolecular monomer cyclization to yield unreactive 3′-5′ cyclized monomers as a side-product also dramatically increased.[14-15] Imidazole activated 2′-amino-2′,3′-dideoxy-guanosine and -cytidine were both efficiently incorporated in template-copying reactions, while the intramolecular cyclization was significantly decreased due to ring strain.[16] Template-directed polymerization using an amine as the nucleophile at the growing end of the primer has also been observed for the npGNA[17] and 2′-NH_2-TNA[18] systems. Activated 2′-amino GNA dinucleotides can be used in nonenzymatic GNA

template-directed ligation to yield npGNA oligonucleotides, however, the addition of monomers is unsuccessful due to rapid intramolecular cyclization.[17]

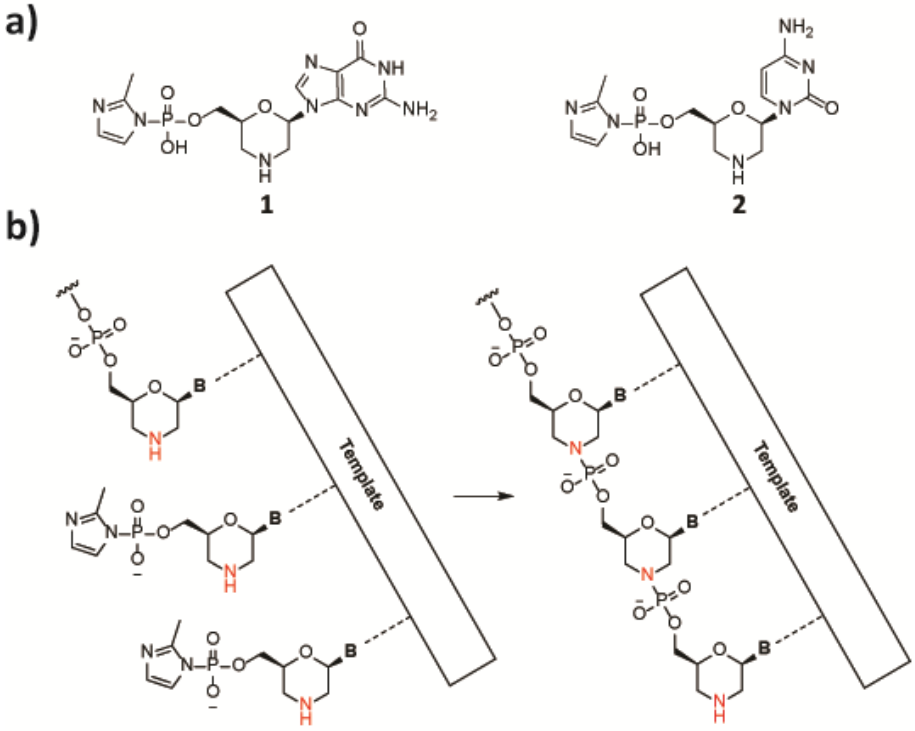


Figure 2.1: Morpholino nucleic acid (MoNA) system. (a) Structure of activated MoNA monomers: **1,** 5′-phosphor-2-methylimidazolide of morpholino guanosine (2-MeImpmG); **2**, 5′-phosphor-2-methylimidazolide of morpholino cytidine (2-MeImpmC). (b) Schematic representation of RNA template-directed primer extension of a morpholino-terminated primer in the presence of activated MoNA monomers. The attacking nucleophile is the morpholino amine marked in red.

Encouraged by previous results with nucleophilic amines, we have now investigated the use of morpholino nucleic acids (MoNA) (Fig.2.1a) in RNA template-directed nonenzymatic primer extension. MoNA is an interesting candidate because: a) the secondary morpholino amine should serve as a stronger nucleophile than a primary amine in water,[19] b) molecular modeling suggests that morpholino oligos in which the morpholine rings are in a chair conformation should form an A-form heteroduplex with an RNA oligonucleotide that is structurally similar to an RNA-RNA duplex.[20] , and c) the chair

conformation of the morpholine ring should prevent cyclization of the activated morpholino monomers. The chair conformation of the morpholine ring is supported by proton NMR analysis of a carbamate-linked morpholino oligonucleotide.[21] The A-form geometry of the morpholino oligo is important because it is the preferred geometry for template-directed nonenzymatic extension by activated monomers, since the A-form conformation is suitable for an in-line attack by the nucleophile of the growing primer on the phosphate of the incoming activated monomer.

Here we present the synthesis of 2-methylimidazole activated morpholino guanosine (2-MeImpmG) and cytidine (2-MeImpmC), and the synthesis of RNA primers terminating in a morpholino nucleotide. The monomers do not cyclize, but at neutral pH and high monomer concentration, non-templated oligomerization is significant. We show that the extension of an RNA primer terminated with a morpholino unit using low concentrations (5 mM) of 2-MeImpmG or 2-MeImpmC proceeds very rapidly on RNA templates.

2.3 Results

We first developed a two-step synthesis of 2-methylimidazole activated 5′-phosphoro-MoNA monomers. We began by converting NMPs to their corresponding 2-methylimidazole derivatives following a standard and well-established protocol.[22] Next, the ribose sugar was converted to a morpholine ring by oxidative opening of the 5-membered ribose ring with periodate. The resulting dialdehyde was treated with ammonium tetraborate to give a 6-membered morpholine ring, followed by reductive removal of the original 2′ and 3′ hydroxyls with cyanoborohydride (Figure 2.2a).

One of the limiting factors for the extent of template copying is the longevity of chemically activated mononucleotides in a closed reaction system. 2-methylimidazole activated monomers are known to be prone to hydrolysis and cyclization under template copying conditions. For example, in the presence of 100 mM Mg^{2+}, the rate of hydrolysis of 2-MeImpG at pH 7.57 is 0.022 h^{-1}, corresponding to a half-life of 32 h.[23] 3′-NH_2-2-

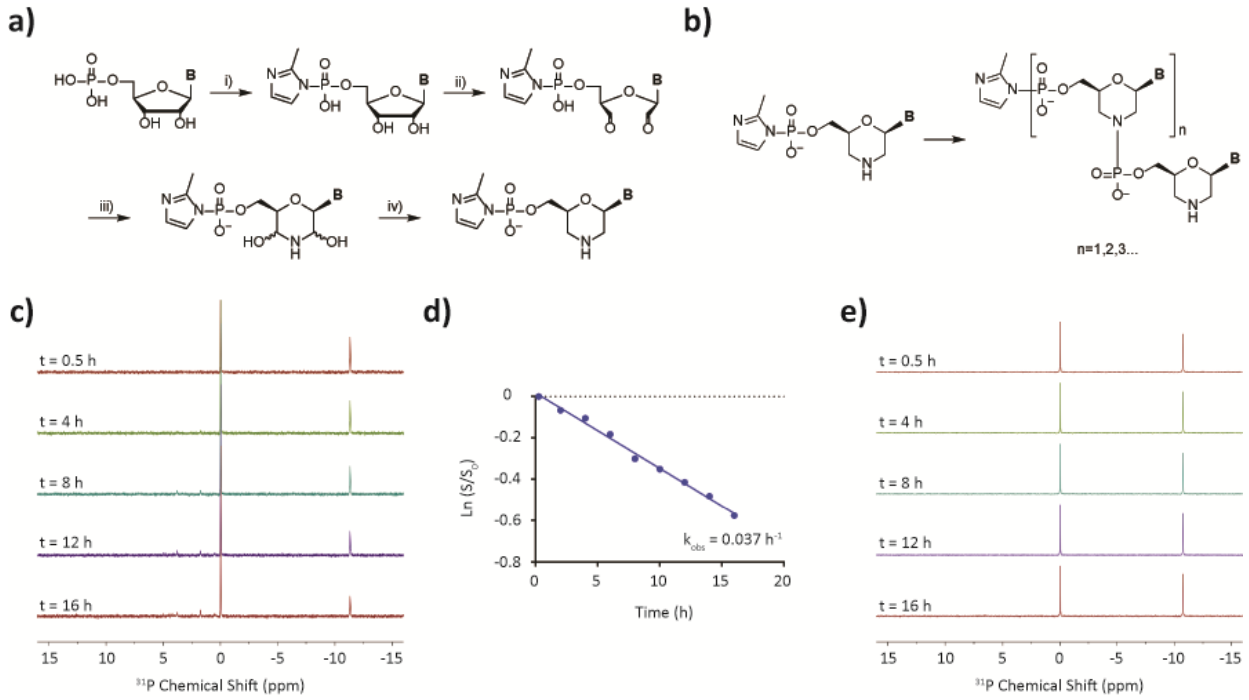


Figure 2.2: Preparation of activated MoNA monomers and NMR studies of the decay of 2-MeImpmG. (a) Synthesis of 2-methylimidazole-activated MoNA monomers. Reaction conditions: i) 2-methylimidazole, DPDS, TPP, TEA, DMSO, 20 ˚C, 2.5 h; ii & iii) $NaIO_4$, $(NH_4)_2B_4O_7$ MeOH, 20 ˚C, 0.5 h; iv) $NaBH_3CN$, 20 ˚C, 3 h. Reactions were performed in one pot for steps ii) to iv). (b) Schematic representation of the non-templated polymerization of activated MoNA monomers to yield activated oligomers. (c) Activated 2-MeImpmG monitored at δ = - 11.3 ppm by ^{31}P NMR spectroscopy at different time points. Phosphorus signals corresponding to internal phosphodiester linkages were observed between δ = 0 - 5 ppm. Phosphate buffer (δ = 0 ppm) was used as external reference. Reactions contained 10 mM 2-MeImpmG in 200 mM HEPES pH 7.5 and 15% D_2O. (d) Rate of decline of the ^{31}P phosphoramidate resonance from a plot of $ln(S/S_0)$ vs time. Solid line is linear fit with $R^2 > 0.99$ and $k_{obs} = 0.037\ h^{-1}$. (e) Stability of 2-MeImpmG at pH 12. Activated 2-MeImpmG monitored at δ = - 10.8 ppm by ^{31}P NMR spectroscopy at different time points. Phosphate buffer (δ = 0 ppm) was used as external reference. Solution contained 15mM 2-MeImpmG in H_2O with 15% D_2O.

MeImpddG, under its optimal primer extension condition in the presence of 100 mM HEI, undergoes cyclization at a rate of 0.53 h^{-1}, corresponding to a half-life of only 1.3 h.[14] Such undesired side reactions greatly impaired the efficiency and sustainability of template copying in these systems. We monitored the stability of the activated MoNA monomers at around neutral pH by ^{31}P NMR spectroscopy. Neither cyclization nor hydrolysis products were observed, even after one day. Instead, activated MoNA monomers were found to undergo non-templated polymerization to yield activated oligomers (Figure 2.2b). The accumulation of various activated oligomers and corresponding loss of activated monomer was evident from the decay of the ^{31}P signal related to the phosphoramidate linkage, and corresponding growth of the ^{31}P resonances due to internal phosphodiester linkages (Figure 2.2c). In the case of 2-MeImpmG at pH 7.5, the rate of decline of the ^{31}P phosphoramidate resonance was 0.037 h^{-1}, corresponding to a half-life of 19 h. These results were confirmed by liquid chromatography-mass spectrometry (LCMS) studies. Because the spontaneous oligomerization of the activated MoNA monomers to varying extents may complicate the interpretation of template-directed primer extension studies, we searched for conditions in which monomers could be stored without the formation of unwanted side products. Since all the side reactions of concern, i.e. cyclization, hydrolysis and oligomerization, required a protonated 2-methylimidazole leaving group, we envisioned that basic pH would be crucial. Following the reasoning, we found that the activated MoNA monomers were stable in the pH range from 10 to 12. For example, no change was observed in the ^{31}P NMR spectrum of a 10 mM solution of 2-MeImpmG at pH 12 after 16 hours (Figure 2.2e).

The morpholino-terminated primer was prepared following a similar strategy to that previously described by Richert and his coworkers for the synthesis of a 3′-amino terminated DNA primer.[24] A protected 5′-DMTr-morpholino residue was incubated with long-chain-alkylamine controlled pore glass (LCAA-cpg) previously treated with hexafluoroglutaric anhydride to prepare an appropriate solid support, which was subsequently used for standard solid-phase oligonucleotide synthesis to yield fluorescently labeled chimeric RNA/MoNA oligonucleotides (Figure 3a and Figure S1). Those oligonucleotides were used as primers for subsequent template-directed primer extension reactions with activated morpholino nucleotides and ribonucleotides.

We initially compared primer extension reactions primers ending in either the morpholino amine or the secondary alcohol of ribose, using the chimeric RNA/MoNA primer **P1** (5′-Cy5-GACUGACUGmG-3′) and the corresponding all RNA primer **P2** (5′-Cy5-GACUGACUGG-3′). Reactions were carried out on the C_4 RNA template **T1** in the presence of guanosine 5′-phosphoro-2-methylimidazolide (2-MeImpG). In the absence of divalent cations, primer **P1** yielded >95% of the N+1 product in 5 min (Figure A.2a) while when the same reaction was performed with **P2**, >98% of the primer remained unreacted even after 24 h. In the presence of 100 mM Mg^{2+}, *ca.* 81% of **P2** extended to various lengths after 1 h (Figure A.2b). The morpholino terminated primer is clearly much more reactive than the all RNA primer, even in the presence of Mg^{2+}, presumably due to the greater nucleophilicity of the morpholino amine than 3′-hydroxyl of ribose. We then compared the reactivity of a primer ending in a morpholino secondary amine with that of a primer ending in the primary amine of a 3′-amino-2′,3′-dideoxyribonucleoside (3′-NP-DNA). For this comparison, we used the chimeric RNA/MoNA primer **P3** (5′-Cy3-

AGUGAGUAACGmC-3′) and the 3′-amino-2′,3′-dideoxy C terminated primer **P4** (5′-Cy3-AGUGAGUAACGC$_{NH2}$-3′), on the C_4 RNA template **T2** in the presence of 2-MeImpG. Primer extension of primer **P3** yielded 93% of the N+1 product in 5 min, while primer **P4** yielded only 42% of the N+1 product in the same time. Kinetic analysis of the two reactions showed that primer extension from the morpholino-terminated primer **P3** was ca. 4 folds faster than from the 3′-amino terminated primer **P4** (Figure 2.3c and 2.3d).

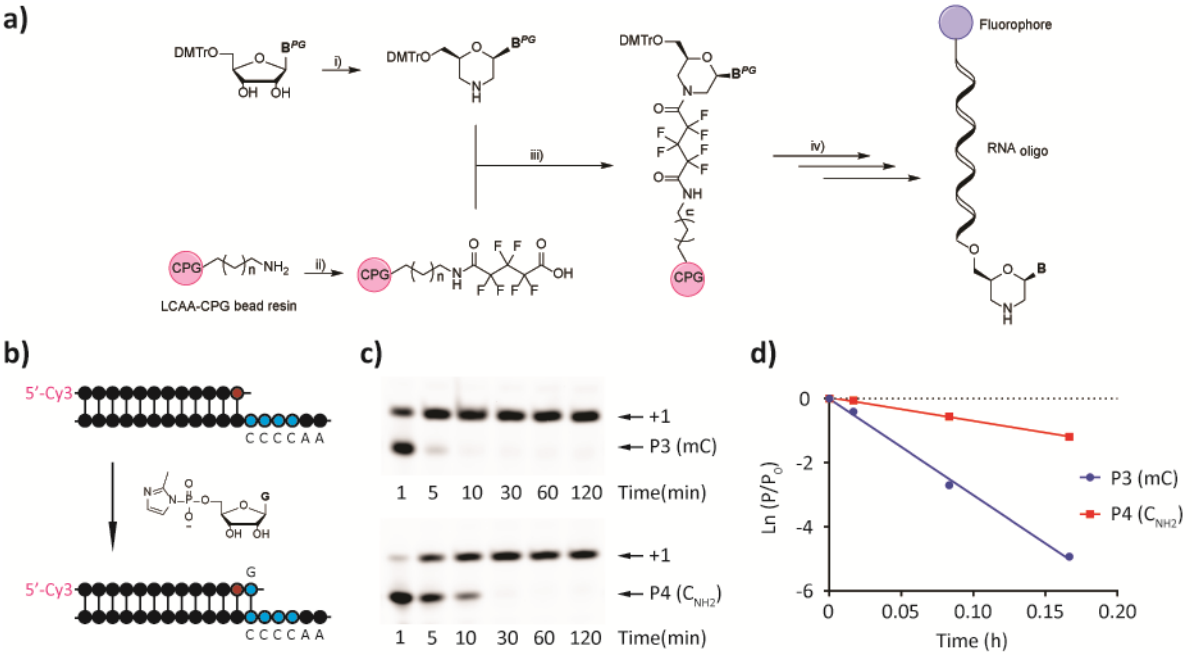


Figure 2.3: Preparation of fluorophore labeled 3′-morpholino-terminated RNA oligonucleotide and nonenzymatic primer-extension reactions with 3′-terminal modified primers. (a) Preparation of fluorophore labeled 3′-morpholino-terminated RNA oligonucleotide. Reaction conditions: i) A. $(NH_4)_2B_4O_7$, $NaIO_4$, MeOH, 20 ˚C, 0.5 h; B. $NaBH_3CN$, 20 ˚C, 2 h; ii) Hexafluoroglutaric anhydride, DIEA, DCM, 20 ˚C, 16 h; iii) PPh_3, DIEA, I_2, DMF, 20 ˚C, 16 h; iv) standard RNA synthesis followed by standard deprotection procedure. (b) Schematic representation of terminal modified primer **P3** or all RNA primer **P4** annealed to a C_4 RNA template (**T2**) in presence of 2-MeImpG. The monomers participate in primer extension reaction on the corresponding template region resulting in predominantly N+1 product. (c) High-resolution PAGE analysis of the primer-extension products at different reaction time. Reaction conditions: 1 μM primer **P3** (top) or **P4** (bottom), 5 μM C_4 template **T2**, 300 mM HEPES pH 7.5, 5 mM 2-MeImpG. (d) Determination of the rate of the primer extension from the morpholino-terminated primer **P3** and 3′-amino terminated primer **P4** via the plot of $\ln(P/P_0)$ vs time. Solid line is a linear fit with $R^2 > 0.99$ and k_{obs} of 30 h^{-1} with **P3** and 7.3 h^{-1} with **P4**.

Similar results were observed when both primers were tested on a G_4 RNA template in the presence of cytidine 5′-phosphoro-2-methylimidazolide (2-MeImpC) (Figure A.3). These results are consistent with the hypothesis that morpholino amine of MoNA nucleotides is a better nucleophile than both the secondary alcohol of RNA nucleotides and the primary amine of 3′-NP-DNA nucleotides, in the context of primer extension reactions.

Encouraged by the rapid and complete template-directed reaction of a primer ending in a morpholino nucleotide with an activated ribonucleotide, we proceeded to examine primer extension using activated morpholino nucleotides. We first considered the extension of primer **P1** on the homopolymeric C_4 and G_4 RNA templates (**T1** and **T3** respectively) (Figure 2.4). Polyacrylamide gel electrophoresis analysis of the reaction of the **P1/T1** complex in the presence of 5 mM 2-MeImpmG showed that >99% of the primer was consumed within 1 minute, resulting in mostly full-length N+4 product with some N+3 product (Figure 2.4b). A minor amount of overhanging N+5 product accumulated over an hour. Similar results were obtained when 5 mM 2-MeImpmC was added to the **P1/T3** complex(Figure 2.4e). After 1 min, the reaction was essentially complete with >98% of the primer converted to full length products (≥ +4 nts). The identities of the extended products were further confirmed by LC-MS (Figure 2.4c, 2.4f and A.4). When primer extension reactions were performed in the absence of template or in the presence of non-complementary templates, only traces of N+1 product formed after 1 h. Thus, non-templated primer extension is *ca.* 1000-fold slower than template-directed copying (Figure A.5). The rate of template copying did not vary noticeably between the primer terminated with mG (**P1**) and the primer terminated with mC (**P4**).

Remarkably, the latter reaction was at least one order of magnitude faster than the extension of the corresponding 3′-amino-2′,3′-dideoxy C terminated primer **P3** on the G_4 RNA template **T4** in the presence of 3′-NH_2-2-MeImpddC (Figure A.6).

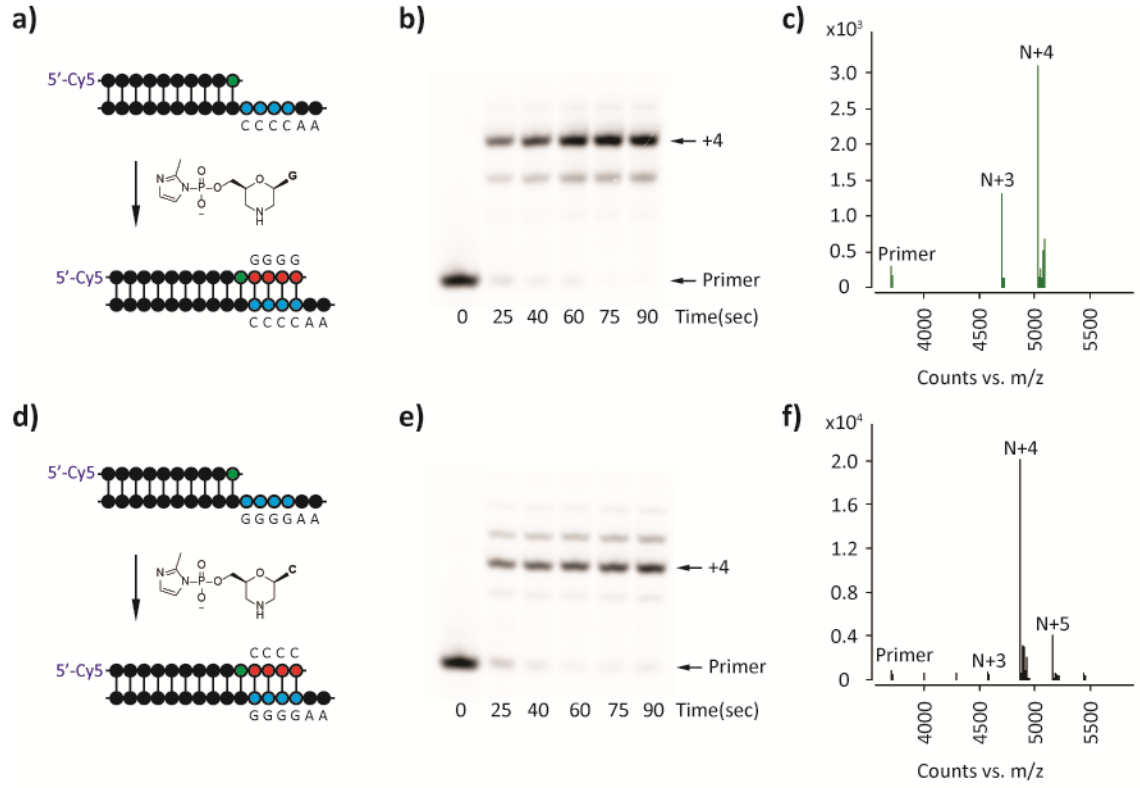


Figure 2.4: Nonenzymatic primer extension reactions on homopolymeric RNA templates using morpholino monomers 2-MeImpmG and 2-MeImpmC. (a) Schematic representation of primer **P1** annealed to a C_4 template (**T1**) in presence of 2-MeImpmG. Monomers participate in primer extension on the template resulting in the synthesis of chimeric RNA/MoNA oligomers. (b) High-resolution PAGE analysis of the primer extension products at increasing reaction times. Reaction conditions: 1 μM primer **P1**, 5 μM C_4 template **T1**, 300 mM HEPES pH 7.5, 5 mM 2-MeImpmG. (c) High-resolution MS analysis of the primer extension products formed after a 1-day incubation. Reaction conditions: 2.75 μM primer **P1**, 10 μM template **T1**, 200 mM HEPES pH 7.5, 5 mM 2-MeImpmG. (d) Schematic representation of primer **P1** annealed to a G_4 template (**T3**) in presence of 2-MeImpmC. (e) High-resolution PAGE analysis of the primer extension products at increasing reaction times. Reaction conditions: 1 μM primer **P1**, 5 μM G_4 template **T3**, 300 mM HEPES pH 7.5, 5 mM 2-MeImpmC. (c) High-resolution MS analysis of the primer extension products formed after a 1-day incubation. Reaction conditions: 2.75 μM primer **P1**, 10 μM template **T3**, 200 mM HEPES pH 7.5, 5 mM 2-MeImpmC.

We examined primer extension reactions with activated morpholino monomers under a range of pH and metal ion conditions, to assess the potential compatibility of this system with model protocell membranes. Primer extension reactions with primer **P1** on templates **T1** and **T3** in the presence of 2-MeImpmG and 2-MeImpmC respectively were indistinguishably rapid in the pH range from 6 to 8.5. At pH > 8.5, the reaction slowed down considerably (Figure A.7). As previously observed with amino-sugar nucleotides, the presence of divalent cations such as Mg^{2+} and Mn^{2+} had minimal effects on the kinetics of the primer extension reactions. This observation suggests that template copying in the morpholino system should be compatible with primitive fatty acid based vesicles, which are destabilized by high concentrations of divalent cations.[11]

The pattern of primer extension products obtained with the 3′-morpholino-terminated primer and morpholino-nucleotide monomers is strikingly different from that typically observed for an all RNA system. The template-directed addition of nucleotides to an all RNA primer occurs in a stepwise manner, with a distribution of +1 to +3 products observed at early time points, and longer addition products accumulating at later times. In contrast, the addition of morpholino-nucleotides to the 3′-morpholino-terminated primer appears to jump rapidly to full length products, with almost no +1 and +2 products observed even at early times when only a fraction of the primer has been converted to +4 full length product. One hypothesis that could explain this observation is that the addition of the first (and possibly second) morpholino nucleotides to a morpholino-terminated primer is much slower than subsequent additions. This could be caused by a transition in the geometry of duplex between that of the RNA:RNA homoduplex to that of the MoNA:RNA heteroduplex. To test this hypothesis, we purified

an RNA primer with three terminal morpholino nucleotides **P5** (5′-Cy5-GACUGACUGmGmGmG-3′) from a primer extension reaction of **P1** with 2-MeImpmG. Primer extension using **P5** as the primer on a C_4 template with morpholino-G monomers showed a more typical stepwise addition pattern (Figure A.8), consistent with the above hypothesis.

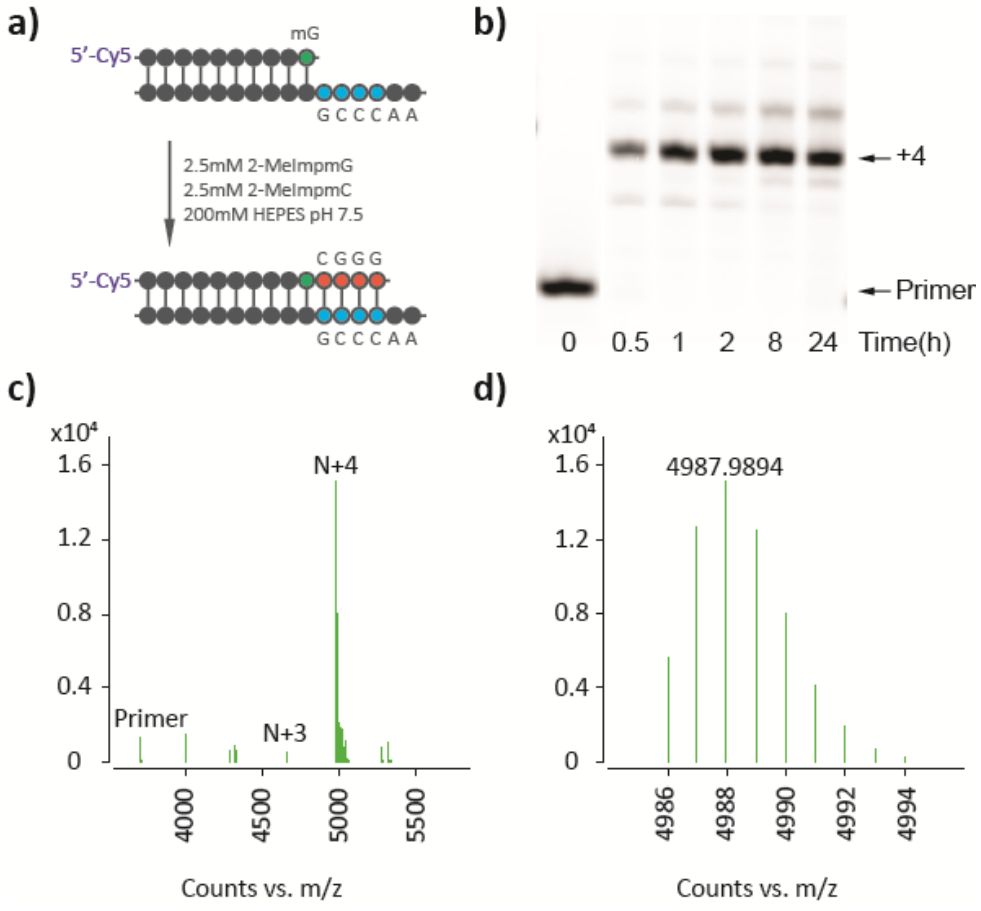


Figure 2.5: Nonenzymatic primer-extension reactions on a mixed sequence RNA template. (a) Schematic representation of primer **P1** annealed to template (**T5**) in presence of 2-MeImpmG and 2-MeImpmC, followed by primer extension. (b) High-resolution PAGE analysis of the primer-extension products at increasing reaction times. Reaction conditions: 1 μM primer **P1**, 5 μM template **T5**, 300 mM HEPES pH 7.5, 5 mM 2-MeImpmG, 5 mM 2-MeImpmC. (c) High-resolution MS analysis of the primer-extension products after 1 day. Reaction condition: 2.75μm primer **P1**, 10μm template **T5**, 200 mM HEPES pH 7.5, 3 mM 2-MeImpmG, 3 mM 2-MeImpmC. (d) Monoisotopic mass for the full-length primer +4 product: calculated mass, 4985.9979 Da; observed mass, 4985.9961 Da; error, 0.4 ppm.

Although homopolymeric templates have been traditionally used to investigate the efficiency of individual activated nucleotides, primer extension reactions on mixed sequence templates are more relevant to the ultimate goal of replicating functional sequences. We therefore prepared four mixed sequence C+G templates with template regions: 3′-GCCC, 3′-CGGG, 3′-GGGC, and 3′-CCCG-5′ (**T5**, **T6**, **T7**, and **T8** respectively) and studied the appearance of primer extension products in the presence of 5 mM each of 2-MeImpmG and 2-MeImpmC (Figure 2.5, Figure A.9-A.11). PAGE analysis showed that in all the cases the primers were consumed within 30 min, yielding predominantly N+3 and N+4 products. Primer extension on the templates that contained a different nucleotide at the last position of the templating region (**T7** and **T8**) took longer to complete the addition of the last nucleotide. The products of these reactions were further analyzed by LC-MS (Figure 2.5c). The extracted ion chromatograms confirmed that in all cases, the major products were the full-length correct sequences. The presence of the correct +1 to +3 products in the extracted ion chromatograms is consistent with the sequential addition of monomers to the growing primer. Only traces of sequences with a single incorrect base at the terminus were seen, suggesting not only that misincorporations were rare, but that continued primer extension following a mismatch was very slow, as previously observed for a DNA system with 3′-amino terminated DNA primers.[25] Surprisingly, the rates of primer extension on the mixed sequence templates **T5** and **T6** were slower than that on homopolymeric templates **T1** and **T3**. The reasons for the slow copying of these mixed sequence templates are unclear and are the subject of ongoing study.

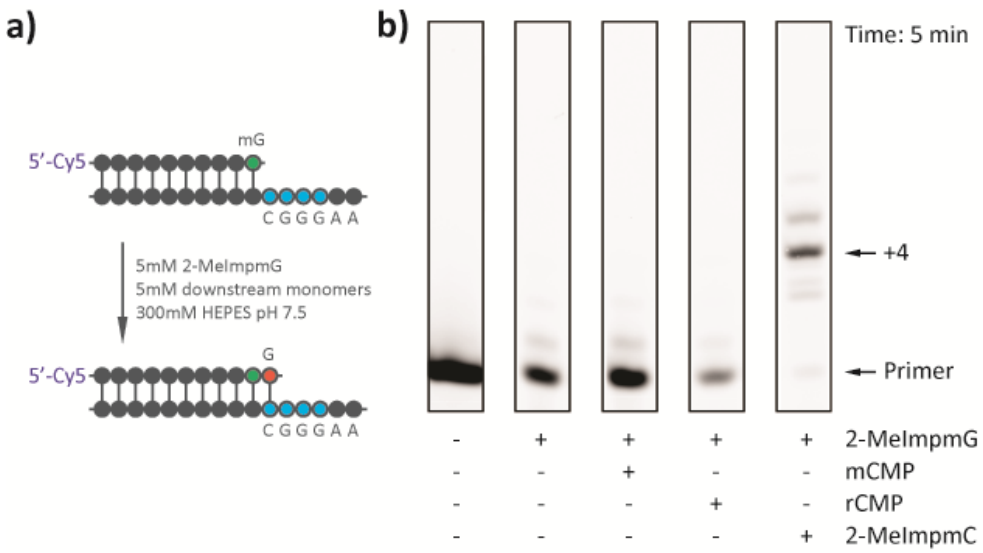


Figure 2.6: Fast primer extension with a morpholino-monomer requires an activated downstream monomer. (a) Schematic representation of primer **P1** annealed to template (**T5**) in the presence of 2-MeImpmG and various downstream monomers. (b) High-resolution PAGE analysis of the primer-extension products when different downstream monomers are used. Lane 1: primer only. Lane 2: activated mG only, no downstream monomer. Lanes 3,4: activated mG but unactivated mC or rC. Lane 5: activated mG and activated mC. The identity of nucleotides added to each reaction is indicated below each lane. Reaction conditions: 1 μM primer **P1**, 5 μM template **T5**, 300 mM HEPES pH 7.5, 5 mM 2-MeImpmG and 5 mM of downstream nucleotides. All reactions were quenched with urea at 5 minutes.

Template directed RNA polymerization with 2-methyl- or 2-amino-imidazole activated ribonucleotide monomers proceeds through the formation of a highly reactive imidazolium-bridged dinucleotide intermediate.[26] The intermediate binds to the template, and in the bound state its conformation is preorganized so as to favor reaction with the 3′-hydroxyl of the primer. In principle, two possible mechanisms could account for primer extension with a morpholino-terminated primer. The high nucleophilicity of the morpholino secondary amine could potentially allow for direct monomer addition, or alternatively, primer extension might still proceed largely or entirely though formation of an imidazolium-bridged dinucleotide intermediate, which then reacts rapidly with the

morpholino-terminated primer. To distinguish between these two mechanisms, we used the mixed sequence template with template region: 3′-CGGG-5′ (**T5**) and studied primer extension using primer **P1** in presence of 2-MeImpmG and different downstream monomers (Figure 2.6). PAGE analysis suggested that the rates of primer extension with mG were similar and very slow in the absence of any downstream monomer, or when unactivated morpholino-cytidine monophosphate (mCMP), or unactivated ribocytidine monophosphate (rCMP) were present. In all the three cases, only a small fraction of primer was converted to +1 products in 5 min, and > 20% of primer remained unreacted even after 24 h. In contrast, the addition of 2-MeImpmC resulted in a dramatic increase in the rate of primer extension, such that > 90% of the primer was converted to full length products (≥ +4 nts) in 5 min. This observation strongly suggests that template copying in the MoNA system proceeds via the formation of an imidazolium-bridged reactive intermediate, as previously observed for the all RNA system.

2.4 Discussion

Morpholino nucleotides are readily accessible through oxidation of the canonical ribonucleotides followed by reductive amination. Surprisingly, the corresponding morpholino oligonucleotides have not been characterized, although a neutral-backbone diamidate variant that is the basis of a recently approved therapeutic oligonucleotide has been studied in detail.[27-29] Our work provides a novel synthetic route to the charged backbone morpholino oligonucleotides, through the nonenzymatic copying of RNA templates with activated morpholino-nucleotide monomers. In the course of this work,

we have found that the 2-methylimidazole activated morpholino G and C nucleotides add very rapidly to a morpholino-terminated primer across from both homopolymeric and heteropolymeric RNA templates. Side by side comparisons reveal that the morpholino copying reactions are considerably faster than the corresponding reactions with 3′-amino, 2′,3′-dideoxy nucleotide terminated primer and monomers, which in turn are faster than an otherwise identical but all RNA system. These results are consistent with the expected greater nucleophilicity in water of the morpholino secondary amine, vs. the primary amine of 3′-amino, 2′,3′-dideoxy nucleotides, and the 3′-hydroxyl of ribonucleotides. However, there are clearly significant structural differences between the relatively rigid chair conformation of the morpholine ring, compared with the flexible 5-membered sugar ring of the 3′-amino or ribonucleotide sugars, and these structural differences are also likely to affect the observed rates of primer extension reactions. Nevertheless, an important implication of our observations is that a means of increasing the nucleophilicity of the ribose 3′-hydroxyl, for example by correctly positioning a nearby metal ion, might lead to enhanced rates of nonenzymatic primer extension in an all RNA system.

The rapid rates of template copying, together with the very slow rates of hydrolysis and cyclization of the activated monomers, suggests that the morpholino nucleic acids might serve as the basis for a nonenzymatically replicating genetic system. Furthermore, the fact that the template copying reaction proceeds without the assistance of divalent cations and over a wide range of pH values suggests that this genetic polymer should be compatible with replicating compartment systems, such as those formed by fatty acid vesicles. To develop the MoNA system as a general genetic

polymer, it will be necessary to demonstrate the copying of templates that contain all four nucleotides, and a means of copying the copies so that indefinite cycles of replication can be attained. If a replicating MoNA system can be developed, it may serve as the basis for an alternative protocell design that is not derived from either biology or prebiotic chemistry.

2.5 References

1. Szostak, J. W., The Narrow Road to the Deep Past: In Search of the Chemistry of the Origin of Life. *Angew Chem Int Ed Engl* **2017,** *56* (37), 11037-11043.

2. Szostak, J. W.; Bartel, D. P.; Luisi, P. L., Synthesizing life. *Nature* **2001,** *409* (6818), 387-90.

3. Chen, I. A.; Roberts, R. W.; Szostak, J. W., The emergence of competition between model protocells. *Science* **2004,** *305* (5689), 1474-6.

4. Ganti, T., Organization of chemical reactions into dividing and metabolizing units: the chemotons. *Biosystems* **1975,** *7* (1), 15-21.

5. Orgel, L. E., Molecular replication. *Nature* **1992,** *358* (6383), 203-9.

6. Gebicki, J. M.; Hicks, M., Ufasomes are stable particles surrounded by unsaturated fatty acid membranes. *Nature* **1973,** *243* (5404), 232-4.

7. Gilbert, W., Origin of life: The RNA world. *Nature* **1986,** *319*, 618.

8. Joyce, G. F., RNA evolution and the origins of life. *Nature* **1989,** *338* (6212), 217-24.

9. Ma, W.; Yu, C.; Zhang, W.; Hu, J., Nucleotide synthetase ribozymes may have emerged first in the RNA world. *RNA* **2007,** *13* (11), 2012-2019.

10. Weimann, B. J.; Lohrmann, R.; Orgel, L. E.; Schneider-Bernloehr, H.; Sulston, J. E., Template-directed synthesis with adenosine-5'-phosphorimidazolide. *Science* **1968,** *161* (3839), 387.

11. Szostak, J. W., The eightfold path to non-enzymatic RNA replication. *Journal of Systems Chemistry* **2012,** *3* (1), 2.

12. Zielinski, W. S.; Orgel, L. E., Oligomerization of activated derivatives of 3'-amino-3'-deoxyguanosine on poly(C) and poly(dC) templates. *Nucleic Acids Res* **1985,** *13* (7), 2469-84.

13. Tohidi, M.; Zielinski, W. S.; Chen, C. H.; Orgel, L. E., Oligomerization of 3'-amino-3'deoxyguanosine-5'phosphorimidazolidate on a d(CpCpCpCpC) template. *J Mol Evol* **1987,** *25*, 97-9.

14. Zhang, S.; Zhang, N.; Blain, J. C.; Szostak, J. W., Synthesis of N3'-P5'-linked phosphoramidate DNA by nonenzymatic template-directed primer extension. *J Am Chem Soc* **2013,** *135* (2), 924-32.

15. Zhang, S.; Blain, J. C.; Zielinska, D.; Gryaznov, S. M.; Szostak, J. W., Fast and accurate nonenzymatic copying of an RNA-like synthetic genetic polymer. *Proc Natl Acad Sci U S A* **2013,** *110* (44), 17732-7.

16. 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,** *131* (40), 14560-70.

17. Chen, J. J.; Cai, X.; Szostak, J. W., N2'-->p3' phosphoramidate glycerol nucleic acid as a potential alternative genetic system. *J Am Chem Soc* **2009,** *131* (6), 2119-21.

18. Blain, J. C.; Ricardo, A.; Szostak, J. W., Synthesis and nonenzymatic template-directed polymerization of 2'-amino-2'-deoxythreose nucleotides. *J Am Chem Soc* **2014,** *136* (5), 2033-9.

19. Brotzel, F.; Chu, Y. C.; Mayr, H., Nucleophilicities of primary and secondary amines in water. *J Org Chem* **2007,** *72* (10), 3679-88.

20. Moulton, J. D.; Yan, Y.-L., Using Morpholinos to Control Gene Expression. In *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc.: 2008.

21. Stirchak, E. P.; Summerton, J. E.; Weller, D. D., Uncharged stereoregular nucleic acid analogs: 2. Morpholino nucleoside oligomers with carbamate internucleoside linkages. *Nucleic Acids Res* **1989,** *17* (15), 6129-41.

22. Joyce, G. F.; Inoue, T.; Orgel, L. E., Non-enzymatic template-directed synthesis on RNA random copolymers. Poly(C, U) templates. *J Mol Biol* **1984,** *176* (2), 279-306.

23. Kanavarioti, A.; Bernasconi, C. F.; Doodokyan, D. L.; Alberas, D. J., Magnesium ion catalyzed P-N bond hydrolysis in imidazolide-activated nucleotides. Relevance to template-directed synthesis of polynucleotides. *J Am Chem Soc* **1989,** *111* (18), 7247-57.

24. Eisenhuth, R.; Richert, C., Convenient syntheses of 3'-amino-2',3'-dideoxynucleosides, their 5'-monophosphates, and 3'-aminoterminal oligodeoxynucleotide primers. *J Org Chem* **2009,** *74* (1), 26-37.

25. Rajamani, S.; Ichida, J. K.; Antal, T.; Treco, D. A.; Leu, K.; Nowak, M. A.; Szostak, J. W.; Chen, I. A., Effect of stalling after mismatches on the error catastrophe in nonenzymatic nucleic acid replication. *J Am Chem Soc* **2010,** *132* (16), 5880-5.

26. Walton, T.; Szostak, J. W., A Highly Reactive Imidazolium-Bridged Dinucleotide Intermediate in Nonenzymatic RNA Primer Extension. *J Am Chem Soc* **2016,** *138* (36), 11996-2002.

27. Nan, Y.; Zhang, Y.-J., Antisense Phosphorodiamidate Morpholino Oligomers as Novel Antiviral Compounds. *Frontiers in Microbiology* **2018,** *9* (750).

28. Summerton, J.; Weller, D., Morpholino antisense oligomers: design, preparation, and properties. *Antisense Nucleic Acid Drug Dev* **1997,** *7* (3), 187-95.

29. Moulton, H. M.; Moulton, J. D., *Morpholino oligomers : methods and protocols*. Humana Press: New York, 2017; p x, 284 pages.

Chapter 3

Effect of terminal 3′-hydroxymethyl modification of an RNA primer on nonenzymatic primer extension

[*A. Pal, W. Zhang and J. W. Szostak conceived of experiments, analyzed the data and wrote the text. R. Das synthesized the phosphoramidite; W. Zhang and A. Pal performed all other experiments. This chapter has previously been published with minor changes as the following publication:* Pal, A.*; Das, R. S.*; Zhang, W.*; McLaughlin, L. W.; Szostak, J. W. Effect of terminal 3′-hydroxymethyl modification of an RNA primer on nonenzymatic primer extension. *Chem. Commun.* **2016**, *52* (80), 11905–11907*. * denotes co-first author.*]

3.1 Abstract

The importance of the precise position of the hydroxyl at the 3′-end of the primer on the efficiency of nonenzymatic template-directed primer extension is not well understood. To test the effect of spatial displacement of the hydroxyl nucleophile, we synthesized RNA primers terminating in either 2′-deoxy-guanosine or 3′-hydroxymethyl-2′,3′-dideoxy-guanosine residues, and compared their activity in the primer extension reaction. The greatly decreased activity of the hydroxymethyl terminated primer suggests that the spatial preorganization of the deoxyribose and ribose sugars contributes significantly to the efficiency of nonenzymatic primer extension.

3.2 Introduction

Template-directed primer extension with chemically activated ribonucleotide monomers, such as nucleotide 5′-phosphorimidazolides, has been extensively studied as a model for nonenzymatic RNA replication during the origin of life.[1-4] In this model, activated nucleotide monomers first interact with the RNA primer-template complex by binding reversibly to complementary sites on the template. Extension of the primer occurs by nucleophilic attack of either the 2′- or 3′-hydroxyl of the 3′-terminal nucleotide of the primer on the activated 5′-phosphate of an adjacent template-bound monomer. A 2′-5′ or 3′-5′ phosphodiester bond is formed when the nucleophile displaces the leaving group on the phosphate of the activated monomer. This reaction is accelerated by the catalytic assistance of a divalent cation, such as Mg^{2+} and Mn^{2+}, although it is not known whether the metal ion interacts with the nucleophilic hydroxyl, the phosphate, or both. Various leaving groups on the phosphate of the incoming monomers, such as 2-methylimidazole,[5] imidazole,[5-6] and oxyazabenzotriazole,[7] have been studied in the primer extension reaction. Oligomerization on mineral surfaces has been carried out using 1-methyladenine[8] as leaving group to some success. Markedly improved polymerization, in terms of both rate and 3′-5′ regioselectivity, is achieved with 2-methylimidazole[9] as leaving group at a pH of 8 to 8.5, although hydroxyazabenzotriazole gives faster rates at pH 8.9 - 9.5.[10]

Previous studies have demonstrated that primer extension on A-form templates, e.g. RNA, exhibits better efficiency compared to B-form templates, e.g. DNA.[11-12] Moreover, our laboratory has shown that the sugar pucker of activated ribonucleotides switches from mostly 2′-endo in the free state to mostly 3′-endo upon binding to an RNA

template.[13] These observations suggest that an A-form primer-template complexed with activated ribonucleotides monomers in the 3′-endo conformation present the optimal conformation for the in-line attack of the nucleophile at the 3′ end of the primer on the 5′-phosphate of the adjacent activated monomer. A better nucleophile, such as the 3′-amine of N3'-P5'–linked phosphoramidate DNA (3′-NP-DNA), is known to improve the kinetics of the extension reaction.[14-15] However, the importance of the structural position and rigidity of the nucleophile at the 3′-position has not been systematically investigated. Based on the fact that both the 2′ and 3′ hydroxyls are able to attack the phosphate of the incoming nucleotide (to an extent controlled by leaving group and metal ion identity), it might seem that the precise position of the nucleophile is not critical. Furthermore, primers ending in a 2′-deoxynucleotide can still be extended at about 10% of the rate of primer

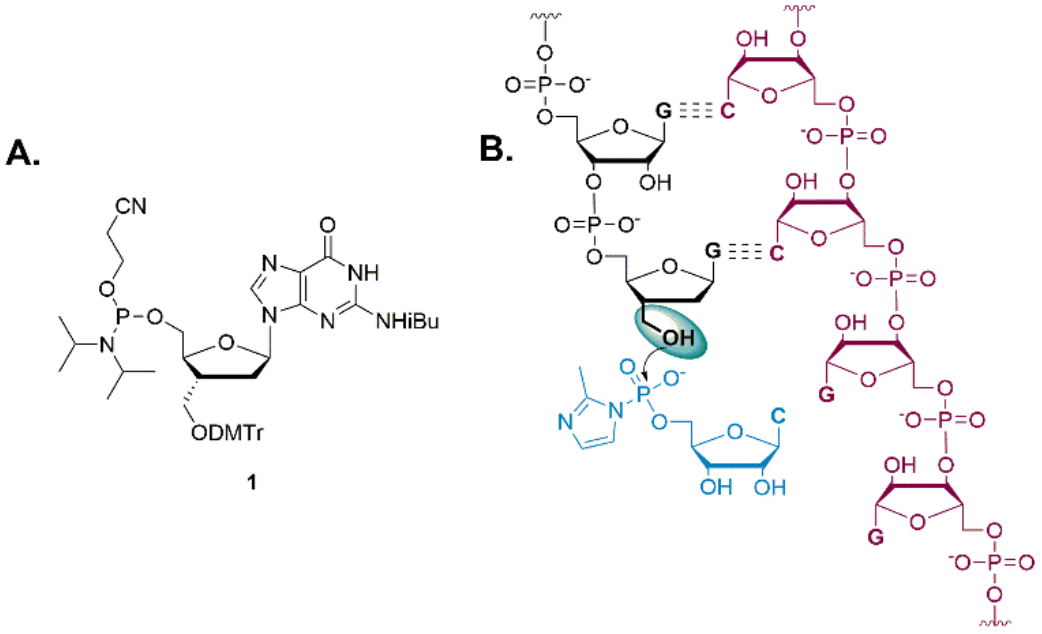


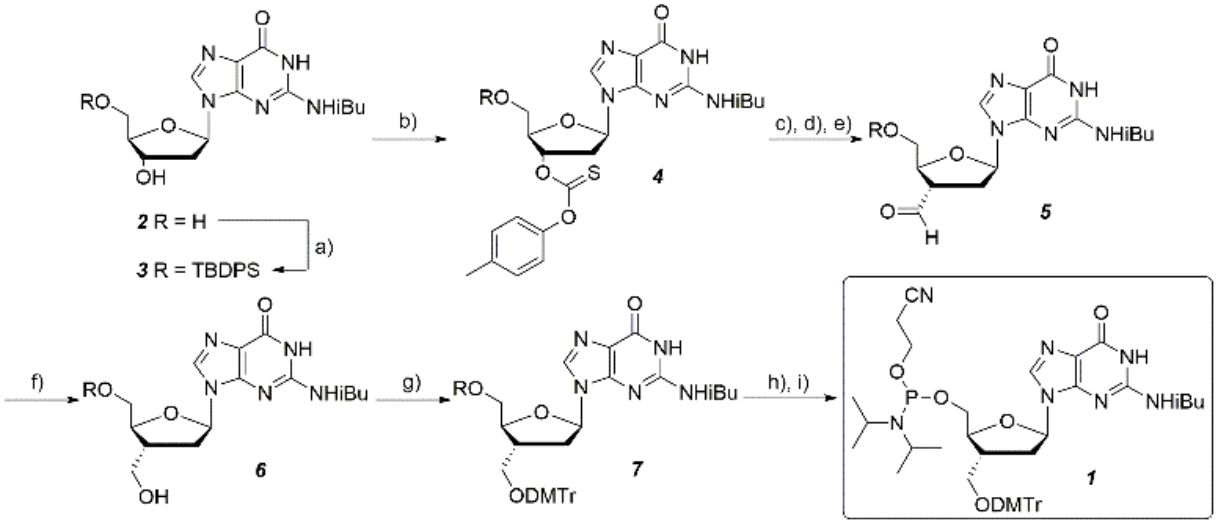
Figure 3.1: Ribonucleoside with 3’-hydroxymethyl modification. A. Structure of 5′-(β-cyanoethyl) phosphoramidite of 3′-hydroxymethyl-2′,3′-dideoxy-guanosine. B. Schematic representation of template-directed primer extension with 3′-hydroxymethyl-2′,3′-dideoxy-guanosine-terminated oligomer in the presence of 2-MeImpC.

ending in a ribonucleotides,[16] despite the higher pKa of the hydroxyl (~14 vs. 12.5).[16-17] The intrinsic sugar preference for the 2′-endo conformation could be countered by the cooperative induction of A-form helical geometry by the remainder of the RNA primer.[18] On the other hand, altritol and hexitol monomers are poor substrates for continued primer extension, although it is unclear whether this reflects poor positioning of the attacking nucleophile or the electrophilic phosphate.[19]

3.3 Results

To explore the structural requirements for placement of the nucleophile, we have replaced the terminal ribonucleotide G of a primer by 3′-hydroxymethyl-2′,3′-dideoxy-guanosine, represented as dG* in this manuscript. Different arguments suggest that this modified 3′-terminus could either enhance or interfere with primer extension. The hydroxymethyl group of dG*, being a primary alcohol, should be a better nucleophile than the 3′-hydroxyl of a 2′-deoxynucleotide, because of its lower pKa and decreased steric hindrance. On the other hand, the displacement of the hydroxyl of dG* away from the sugar, and its greater conformational flexibility, could decrease the rate of primer extension if spatial positioning is a critical factor in the reaction.

In order to investigate the effects of structural perturbation at the 3′-position on the primer extension reaction, we first had to prepare an RNA primer terminating in dG*. Here we describe the synthesis of 5′-(β-cyanoethyl) phosphoramidite of 3′-hydroxymethyl-2′,3′-dideoxy-guanosine and its use in the reverse solid-phase preparation of a primer terminating dG* (Figure 3.1). Several previous reports have described the synthesis of nucleotide analogs with an extended sugar phosphate



Scheme 3.1: Synthesis of phosphoramidite. Reaction conditions: a) TBDPSCl, DMAP, pyr, 20 ˚C, 36 h, 92%; b) O-(p-Tolyl)-chlorothionoformate, DMAP, ACN, 20 ˚C, 1 h, 94%, c) β-tributylstannylstyrene, AIBN, toluene, 20 to 80 ˚C, 48 h; d) OsO4, NMO, dioxane, 20 ˚C, 2 h; e) NaIO4, 20 ˚C, 3 h, 28% over 3 steps; f) NaBH4, MeOH, 0 to 20 ˚C, 30 min, 94%; g) DMTrCl, DMAP, pyr, 20 ˚C, 12 h, 93%; h) TBAF, THF, 20 ˚C, 8 h; i) 2-Cyanoethyl N,N,N,N-tetraisopropylphosphoramidite, tetrazole, DIEA, DCM, 20 ˚C, 2 h, 89% over 2 steps.

backbone, typically beginning with a modified abasic sugar or carbocycle[20-25] followed by glycosylation to incorporate the purine or pyrimidine nucleobase. This approach leads to the formation of unwanted regioisomers that are difficult to separate.[26] In addition, maintenance of the correct stereochemistry at the 3′-carbon has been difficult.[27-28] In our study, the 3′-hydroxymethyl-dG phosphoramidite was synthesized, purified, and characterized by a modification of the procedure originally reported by Mesmaeker et al.[29] that avoids these problems by 3′-homologation of 2′-deoxyguanosine.

To ensure the stereoselective introduction of the 3′-α-hydroxymethyl functionality, the β-face of the sugar ring was blocked by both the bulky tert-butyl diphenylsilyl group and the guanine nucleobase, so that the attack of an incoming nucleophile would occur

predominantly from the α-face. The 3′-hydroxyl was converted into its corresponding o-(p-tolyl)thiocarbonate ester 4. The ester 4 was reacted with β-tributylstannylstyrene in presence of AIBN to form a new C-C bond via β-elimination of tributylstannyl group. Since the β-face of the sugar ring is sterically blocked, styryl addition occurs preferentially through α-face. The resulting reaction mixture was filtered through a short silica pad to remove excess reagents and dried. The resulting crude product was used directly in the next step. The styryl group was cleaved by dihydroxylation using osmium tetroxide, followed by oxidation of the diol with sodium periodate to generate 3-α-C-formyl group in 5. The aldehyde of 5 was efficiently reduced to yield hydroxymethyl group of 6 by sodium borohydride. This 3-α-C-hydroxymethyl group was then protected with DMTr, and the 5′-alcohol of 7 retrieved by desilylation, followed by conversion to phosphoramidite 1 in excellent yield. This reverse amidite 1 was incorporated as the terminal residue of the oligomer by reverse solid-phase synthesis. The oligomer was purified by HPLC and characterized by LCMS. The dG*-terminated oligomer obtained was more than 97% pure with a trace level of N-1 oligomer.

To examine the effect of dG* on template-directed nonenzymatic primer extension, we used 5′-fluorophore tagged dG*-terminated RNA primer (P1, 5′-Cy3-GACUGACUGdG*-3′) complexed with a complementary RNA template strand (T, 3′-CUGACUGACCGGGGAA -5′) containing a homopolymeric G4 site to be copied by the activated monomer, cytidine-5′-phosphor-(2-methyl) imidazolide (2-MeImpC). As a control, a 5′-fluorophore tagged dG-terminated RNA primer (P2, 5′-Cy3-GACUGACUG dG-3′) was also prepared and time courses of primer extension with the two primers were compared (for up to 24 hours).

To determine the optimal conditions for the primer extension reaction, a series of experiments were performed with 1 μM primer, 5 μM template, and 100 mM of different divalent cations including Mg^{2+}, Mn^{2+} and Ca^{2+}, in the presence or absence of monovalent cation (Na^{+} or K^{+}) in the presence of 50 mM of activated monomers and 200 mM buffer (pH from 5.5 to 8.5). Surprisingly, we failed to observe any significant primer extension from the dG* primer P1 even after 1 day under most conditions. The best condition, with Mg^{2+} and pH 7.5, yielded less than 5% extended products after 24 hours, with +3 extended product being the predominant contributor (Fig. 3.2A). The kobs was approximately 0.0118 h^{-1}. We were unable to detect any +1 or +2 extended

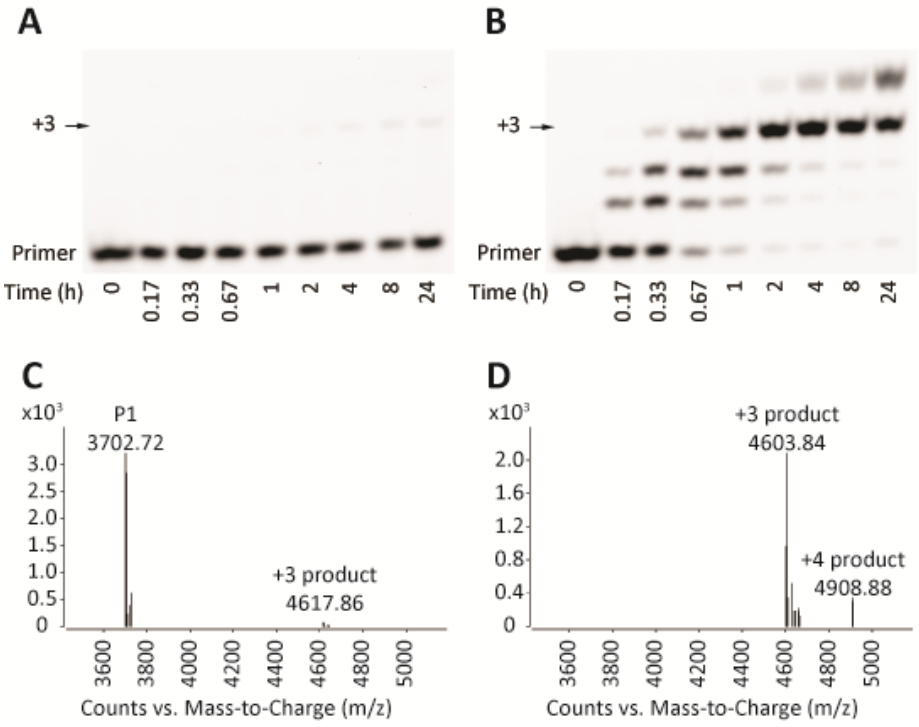


Figure 3.2: Gel electrophoresis and LCMS analysis of primer extension reaction. Reaction conditions:1 μM primer, 5 μM template T, 200 mM HEPES pH 7.5, 100 mM $MgCl_2$, 50 mM 2-MeImpC. A and B. Gel analysis of primer extension of primers P1 and P2 respectively; C and D. LCMS analysis of primer extension of primer P1 and P2 after 24 hours respectively.

products, most likely because after the very slow addition of first monomer, the addition of the following two residues is fast as the terminal residue of the +1 or +2 extended products is a ribonucleotide. On the other hand, P2 extended to varying extents depending on pH, metal ions and time (Figure 3.2B). At pH 7.5 with Mg^{2+}, full-length +3 products appeared after 2 hours with kobs of 2.79 h^{-1} (approx.). After 24 days, more than 95% of the P2 has extended, in striking contrast to the minimal extent of primer extension with P1.

To confirm that the small amount of P1 +3 product observed by PAGE analysis did indeed result from extension from the dG* residue (and not from trace levels of N-1 oligomer), we used LCMS to analyze the 24 hour time point of the primer extension reaction. While unreacted P1 remained as the predominant species (Figure 3.2C), approximately 3% of primer extended to +3 product (Fig. 3.2C; see also Supplementary Information Fig. B.3). These observations are consistent with the results of the gel electrophoresis analysis, and show that primer extension from a terminal dG* residue is very slow compared to extension from a terminal 2′-deoxynucleotide.

In conclusion, we have developed a synthetic route to the 5′-(β-cyanoethyl) phosphoramidite of 3′-hydroxymethyl-2′,3′-dideoxy-guanosine, allowing the dG* residue to be incorporated into an oligonucleotide by reverse solid-phase synthesis. Our results show that the template-directed nonenzymatic extension of a primer with a 3′-terminal dG* proceeds very slowly, at a rate roughly 0.4% that of a primer with a terminal dG residue. We suggest that two major factors may cause this very slow primer extension from dG*. First, the location of the hydroxyl in dG* is ~1.5 Å from the position of the normal 3′-hydroxyl. This altered location may make nucleophilic attack of the hydroxyl

on the phosphate of the incoming activated nucleotide very difficult, or may hinder coordination with the catalytic Mg^{2+} ion. Second, the conformational flexibility of the primary hydroxyl group may decrease the probability of the required spatial positioning of the nucleophile for an in-line attack to the 5′-phosphate of the activated monomer. In summary, our results suggest that the normal location of the nucleophile, on the 3′-carbon of the ribose or deoxyribose sugar, is required for efficient nonenzymatic template directed primer extension.

3.4 References

1. Orgel, L. E., Prebiotic Chemistry and the Origin of the RNA World. *Crit. Rev. Biochem. Mol. Biol.* **2004,** *39* (2), 99-123.

2. Wu, T.; Orgel, L. E., Nonenzymic template-directed synthesis on hairpin oligonucleotides. 2. Templates containing cytidine and guanosine residues. *J. Am. Chem. Soc.* **1992,** *114* (14), 5496-5501.

3. Wu, T.; Orgel, L. E., Nonenzymatic template-directed synthesis on hairpin oligonucleotides. 3. Incorporation of adenosine and uridine residues. *J. Am. Chem. Soc.* **1992,** *114* (21), 7963-7969.

4. Mansy, S. S.; Schrum, J. P.; Krishnamurthy, M.; Tobe, S.; Treco, D. A.; Szostak, J. W., Template-directed synthesis of a genetic polymer in a model protocell. *Nature* **2008,** *454* (7200), 122-125.

5. Lohrmann, R.; Orgel, L. E., Preferential formation of (2'–5')-linked internucleotide bonds in non-enzymatic reactions. *Tetrahedron* **1978,** *34* (7), 853-855.

6. Joyce, G. F.; Inoue, T.; Orgel, L. E., Non-enzymatic template-directed synthesis on RNA random copolymers. *J. Mol. Biol.* **1984,** *176* (2), 279-306.

7. Hagenbuch, P.; Kervio, E.; Hochgesand, A.; Plutowski, U.; Richert, C., Chemical Primer Extension: Efficiently Determining Single Nucleotides in DNA. *Angew. Chem.Int. Ed. Engl.* **2005,** *44* (40), 6588-6592.

8. 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.

9. Inoue, T.; Orgel, L. E., Oligomerization of (guanosine 5′-phosphor)-2-methylimidazolide on poly(C): An RNA polymerase model. *J. Mol. Biol.* **1982,** *162* (1), 201-217.

10. Vogel, S. R.; Deck, C.; Richert, C., Accelerating chemical replication steps of RNA involving activated ribonucleotides and downstream-binding elements. *Chem. Commun.* **2005,** (39), 4922-4924.

11. Kurz, M.; Göbel, K.; Hartel, C.; Göbel, M. W., Nonenzymatic Oligomerization of Ribonucleotides on Guanosine-Rich Templates: Suppression of the Self-Pairing of Guanosine. *Angew. Chem.Int. Ed. Engl.* **1997,** *36* (8), 842-845.

12. Kurz, M.; Göbel, K.; Hartel, C.; Göbel, M. W., Acridine-Labeled Primers as Tools for the Study of Nonenzymatic RNA Oligomerization. *Helvetica Chimica Acta* **1998,** *81* (5-8), 1156-1180.

13. Zhang, N.; Zhang, S.; Szostak, J. W., Activated Ribonucleotides Undergo a Sugar Pucker Switch upon Binding to a Single-Stranded RNA Template. *J. Am. Chem. Soc.* **2012,** *134* (8), 3691-3694.

14. Zhang, S.; Zhang, N.; Blain, J. C.; Szostak, J. W., Synthesis of N3′-P5′-linked Phosphoramidate DNA by Nonenzymatic Template-Directed Primer Extension. *J. Am. Chem. Soc.* **2013,** *135* (2), 924-932.

15. Zhang, S.; Blain, J. C.; Zielinska, D.; Gryaznov, S. M.; Szostak, J. W., Fast and accurate nonenzymatic copying of an RNA-like synthetic genetic polymer. *Proc. Nat. Acad. Sci.* **2013,** *110* (44), 17732-17737.

16. Wu, T.; Orgel, L. E., Nonenzymic template-directed synthesis on oligodeoxycytidylate sequences in hairpin oligonucleotides. *J. Am. Chem. Soc.* **1992,** *114* (1), 317-322.

17. Birnbaum, G. I.; Giziewicz, J.; Huber, C. P.; Shugar, D., Intramolecular hydrogen bonding and acidities of nucleoside sugar hydroxyls. Crystal structure and conformation of O2,2'-anhydro-1-.alpha.-D-xylofuranosyluracil. *J. Am. Chem. Soc.* **1976,** *98* (15), 4640-4644.

18. Wahl, M. C.; Sundaralingam, M., B-form to A-form conversion by a 3′-terminal ribose: crystal structure of the chimera d(CCACTAGTG)r(G). *Nucleic Acids Res.* **2000,** *28* (21), 4356-4363.

19. Kozlov, I. A.; Zielinski, M.; Allart, B.; Kerremans, L.; Van Aerschot, A.; Busson, R.; Herdewijn, P.; Orgel, L. E., Nonenzymatic Template-Directed Reactions on Altritol Oligomers, Preorganized Analogues of Oligonucleotides. *Chem. Eur. J.* **2000,** *6*, 151–155.

20. Boumchita, H.; Legraverend, M.; Bisagni, E., Synthesis of ring-enlarged cyclobut-A and cyclobut-G analogs as HIV inhibitors. Part 4. *Heterocycles* **1991,** *32* (9), 1785-1792.

21. Wachtmeister, J.; Classon, B.; Kvarnstrom, I.; Samuelsson, B., Synthesis of novel fluoro carbocyclic purine nucleoside analogs. *Nucleos. Nucleot.* **1997,** *16* (5 & 6), 809-814.

22. Wachtmeister, J.; Muhlman, A.; Classon, B.; Samuelsson, B., Synthesis of 4-substituted carbocyclic 2,3-dideoxy-3-C-hydroxymethyl nucleoside analogues as potential anti-viral agents. *Tetrahedron* **1999,** *55* (35), 10761-10770.

23. Moon, H. R.; Kim, H. O.; Chun, M. W.; Jeong, L. S.; Marquez, V. E., Synthesis of Cyclopropyl-Fused Carbocyclic Nucleosides via the Regioselective Opening of Cyclic Sulfites. *J. Org. Chem.* **1999,** *64* (13), 4733-4741.

24. Wachtmeister, J.; Björn, C.; Bertil, S.; Kvarnström, I., Synthesis of carbocyclic 2′3′-dideoxy-2′-fluoro-3′-C-hydroxymethyl nucleoside analogues as potential inhibitors of HIV and HSV. *Tetrahedron* **1997,** *53* (5), 1861-1872.

25. Janson, M.; Svansson, L.; Svensson, S. C. T.; Kvarnström, I.; Classon, B.; Samuelsson, B., Synthesis of Some Purine Carbocyclic Isosteres of 2′,3′-Dideoxy-3′-C-Hydroxymethyl Nucleosides as Potential Inhibitors of HIV. *Nucleos. Nucleot.* **1997,** *11* (10), 1739-1747.

26. Buenger, G. S.; Marquez, V. E., Carbocyclic ring-enlarged oxetanocin analogs. *Tetrahedron Lett.* **1992,** *33* (26), 3707-3710.

27. Kofoed, T.; Rasmussen, P. B.; ValentinHansen, P.; Pedersen, E. B., Oligodeoxynucleotides with extended 3'- and 5'-homologous internucleotide linkages. *Acta Chem. Scand.* **1997,** *51* (3), 318-324.

28. Vasseur, J. J.; Debart, F.; Sanghvi, Y. S.; Cook, P. D., Oligonucleosides: synthesis of a novel methylhydroxylamine-linked nucleoside dimer and its incorporation into antisense sequences. *J. Am. Chem. Soc.* **192,** *114* (10), 4006–4007.

29. Sanghvi, Y. S.; Bharadwaj, R.; Debart, F.; De Mesmaeker, A., Efficient and Stereoselective Synthesis of 3′-Deoxy 3′ -C-Branched-Chain Substituted Thymidine. *Synthesis* **1994,** (11), 1163-1166.

Chapter 4

The synergistic effects of metals on non-enzymatic nucleic acid copying

[*W. Zhang, A. Larsen and J. Szostak conceived of experiments, analyzed the data and wrote the text. W. Zhang and A. Larsen performed experiments. A version of this chapter will be submitted for peer review and publication.*]

4.1 Abstract

Multivalent metal ions have long been recognized as the essential catalyst in the nonenzymatic replication of RNA. During the catalytic process of the nonenzymatic primer extension, metal ions perform multiple functions, including stabilizing the RNA duplex and activating the nucleophile and/or the electrophile at the reaction site. Various metal ions perform those functions in different efficiency since different metal ion have unique characteristic properties. Here we carefully evaluate the kinetic profile of nonenzymatic RNA primer extension, the thermodynamic favorability of duplex formation and the rate of monomer hydrolysis in response to the presence of Mg^{2+}, Mn^{2+}, and $(Co(NH_3)_6^{3+}$. We demonstrate that the application of cooperative catalysis using metal ions with orthogonal properties results in a synergistic enhancement in reaction kinetics.

4.2 Introduction

Nonenzymatic replication of nucleic acids is believed to have been a critical step in the emergence of cellular life and offers a model for a self-replicating genetic system.[1] Similar chemistry may have enabled the evolution of fitness imparting nucleic acids – such as nucleic acid catalysts prior to the emergence of protein catalysts – and may be suitable for producing the genetic component of a simple artificial cell (i.e. a protocell).[2] Considerable progress has been made towards constructing a nonenzymatic RNA copying system with reasonable rate and fidelity. Copying rates have been improved by: i) the activation of the 5'-phosphate with reactive leaving groups such as 2-methylimidazole and 2-aminoimidazole;[3-4] ii) the use of modified nucleobases such as 2-thiouridine;[5] iii) the immobilization of the RNA template on a solid support,[6] and iv) the addition of microhelper RNAs that improve beneficial base stacking interactions between the growing primer and the incoming monomer.[7-8] Despite this progress, existing methods have proven insufficient for the realization of a self-replicating genetic system. Even systems using the most active leaving groups and optimized for template length, template sequence, and reagent concentration have apparently not enabled copying rates greatly exceeding the addition of a single nucleotide per hour.[9]

The presence of divalent metals has been shown to be vital for accelerating rates of nonenzymatic nucleic acid copying. In addition to facilitating base-pairing and duplex formation by shielding the repulsive negative charges of the nucleic acid phosphate backbone,[10-12] divalent metal ions also catalyze the bond forming reactions responsible for oligonucleotide polymerization. Without using any divalent metals as catalyst, only a

minimal level of reaction is observed.[13] Mg^{2+}, as the most abundant cellular divalent cation and an essential component of numerous enzymatic process,[14] is used most commonly as catalyst in the studies of nonenzymatic replication of nucleic acids. Other divalent metal ions, such as Zn^{2+} and Pb^{2+}, have been shown to catalyze the template-directed polymerization of activated mononucleotides.[15-17] Mn^{2+}, Ca^{2+}, Sr^{2+}, and Ba^{2+} are able to catalyze the nonenzymatic ligation of oligoribonucleotides.[18] In addition, Orgel et al. have reported that $(Co(NH_3)_6)^{3+}$ allows the oligomerization of 2-MeImpdG on poly(dC) template by potentially converting the structure of the DNA duplex into A form.[19-20]

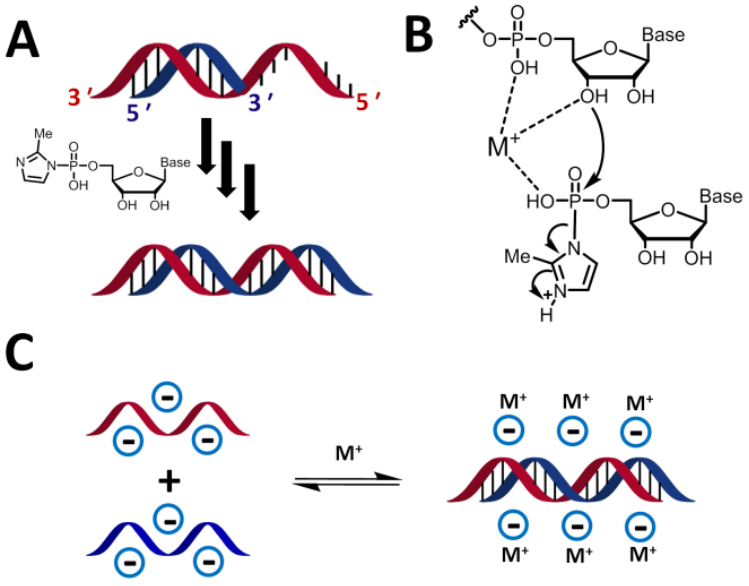


Figure 4.1: The role of metals in template directed, non-enzymatic nucleic acid replication. (A) Non-enzymatic nucleic acid replication scheme. The primer (blue) is annealed to a complementary template (red) and 2-methylimidazole-activated monomers are added to the 3′-OH end of the primer in a complementary fashion with regards to the template; (B) Metal catalysis accelerates non-enzymatic RNA copying, likely by enhancing the proximity and orientation of the nucleophile/electrophile, enhancing the nucleophilicity of the primer 3′-OH, and enhancing the electrophilicity of the 5′-phosphate at the reaction center. In this example, the leaving group is 2-methylimidazole; (C) The metal enhances the favorability of duplex formation by shielding the repulsive negative charges of the nucleic acid backbone.

Although the exact mechanism of the metal's catalytic effect on non-enzymatic RNA copying is not yet fully understood,[21] the mechanism is likely similar to that of enzymatic RNA copying where the metal enhances the proximity and orientation of the nucleophile/electrophile pair, enhances the nucleophilicity of the primer 3′-OH, and activates the 5′-phosphate at the reaction center (Figure 4.1B). For simplicity, we will refer to this mechanism simply as 'phosphate activation'. Unfortunately, metal-catalyzed phosphate activation also increases the hydrolysis of activated monomers or the formation of 2'-3' cyclic phosphates.[22-23] Both processes would greatly disrupt the process of RNA copying. Due to their different sizes, charge densities,[10] oxidation states, coordination spheres and pKa in aqueous complexes,[24] different metal species can be expected to affect duplex formation and phosphate activation differently. It is therefore conceivable that a proper combination of metal species may result in an optimized system that enables fast copying with minimal degradation of the activated monomer.

Here, we report the kinetic characterization of non-enzymatic RNA copying in the presence of Mg^{2+}, Mn^{2+}, and $(Co(NH_3)_6^{3+}$. We have also evaluated the role of these metal ions in the reaction by studying their effects on the thermodynamic favorability of duplex formation and the rate of monomer hydrolysis. We show that metal species that are well suited to improve the favorability of duplex formation are not necessarily well suited to improve the rates of primer extension and vice versa. Furthermore, by combining the actions of metals orthogonally suited to each process, we demonstrate rates of primer extension far in excess of those previously reported while avoiding increased rates of monomer hydrolysis.

4.3 Results

4.3.1 Nonenzymatic Primer Extension with Single Metal Species

To directly evaluate the effects of the different metals species on the rates of RNA copying, we performed RNA primer extension in the presence of various metal ions. We annealed a fluorophore-labeled, mixed sequence primer of 10 nucleotides in length (5′-Cy3-GACUGACUGG-3′) to a complementary, mixed sequence template of 16 nucleotides in length (5′-AACCCCCCAGUCAGUC-3′). This construct results in an overhang at the 5′-end of the template of 6 nucleotides with 4 internal cytosine residues, allowing for the covalent incorporation of up to 4 guanosines to the 3′-end of the primer. (Figure 4.2A) 100 mM Mg^{2+} can increase the rate of non-enzymatic primer extension by approx. 40 fold (6.7 h^{-1} in the presence of 100 mM Mg^{2+} vs 0.17 h^{-1} in the absence of any divalent metal). Most other divalent metal ions demonstrated inferior catalytic efficiency when compared to Mg^{2+}. The addition of 100 mM Mn^{2+} resulted in, by far, the greatest rate of primer extension (34.9 h^{-1}), more than 200 fold faster than the rate of primer extension without divalent metal ions. Hexaamminecobalt(III) chloride ($Co(NH_3)_6Cl_3$ referred to below as CoHex) was selected for investigation based upon its interesting properties and previous application in non-enzymatic DNA copying.[19] CoHex is an archetypal "Werner complex" and is often used as a substitute for Mg^{2+} to deconvolute the effects of inner shell and outer shell contacts in metal-catalyzed processes. While the size and shape of CoHex are similar to that of hexaaquamagnesium ($Mg(H_3O)_6^{2+}$),[25] the orbitals of the cobalt atom are inaccessible for bonding interactions due to the full coordination of the metal by six ammonia groups that are difficult to displace. Despite the lack of inner shell contact, 100 mM of CoHex

enhanced the rate of primer extension by ca. 7 fold (1.16 h^{-1} vs. 0.17 h^{-1}). (Figure 4.2D) Chloride counter ions were chosen in each case to minimize the possible effects of the counter-ion identity on our studies. It should be noted that most of the rates reported

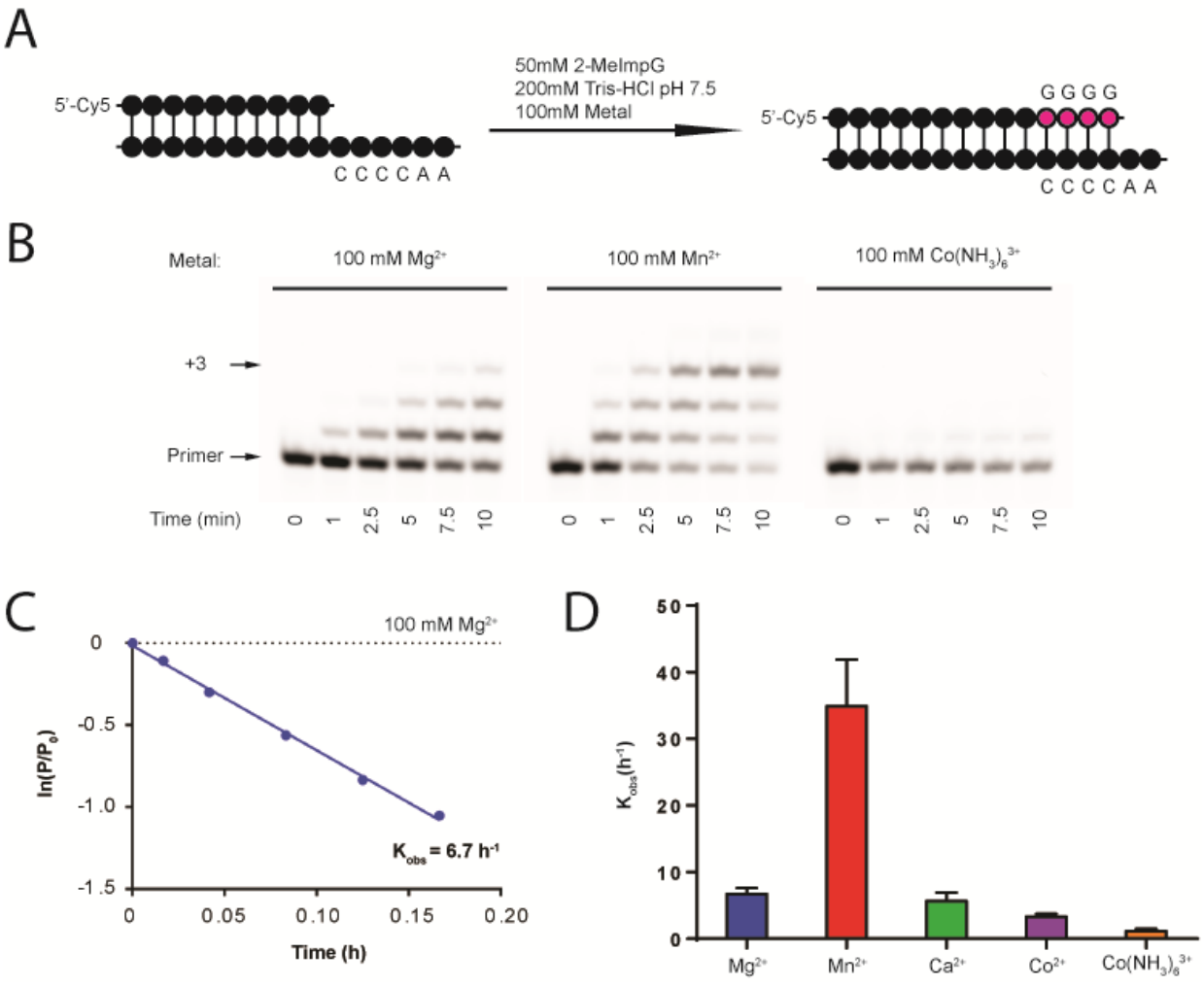


Figure 4.2: Nonenzymatic primer extension of 2-MeImpG on C_4 RNA template in presence of different metal species. (A) Schematic diagram of the primer extension reaction. (B) High-resolution PAGE analysis of the primer-extension products with the indicated metal ions as catalyst. The primer-extension reaction mixtures contained 1 μM Cy5-labeled RNA primer, 5μM template, 200 mM Tris-HCl (pH 7.5), 100 mM metal chloride and 50 mM 2-MeImpG monomer. (C) Determination of the rate of nonenzymatic primer extension in the presence of 100 mM Mg^{2+} via the plot of $\ln(P/P_0)$ versus time. Lines are linear fits, $R^2 \geq 0.99$. (D) Rates of primer extension with the indicated metal ions.

here are several fold greater than what has been previously reported for similar systems. This enhancement reflects improvements in the handling of the activated monomers as described in the supporting information.

To further evaluate the catalytic efficiency of each metal ion and their effect on bind of the activated monomer to the template, we measured the kinetic parameters (k_{max} and monomer K_m) of the primer extension reaction in the presence of metal ions of interest. Primer extensions were performed at various concentration of 2-MeImpG (0 - 100 mM). With 100 mM Mg^{2+}, the determined k_{max} was 22 h^{-1} and K_m was 36 mM. Using 100 mM Mn^{2+} as catalyst in the reaction resulted in a higher maximum rate of reaction, k_{max} = 87 h^{-1}. Mn^{2+} also resulted in a higher affinity of the activated monomer than Mg^{2+}, K_m = 12 mM. As expected, supplementing the reaction with 100 mM CoHex yielded a low maximum rate, k_{max} = 1.6 h^{-1}. However, the effect of CoHex on monomer affinity in the reaction was similar to Mg^{2+}. In the presence of 100 mM CoHex, the K_m of 2-MeImpG was determined to be 42 mM. (Figure C.1) We also investigated the effect of

Table 4.1: Rates of Nonenzymatic Primer Extension with Single Metal Species.

Metal	Rate (h^{-1}) 0 mM Metal	Rate (h^{-1}) 10 mM Metal	Rate (h^{-1}) 25 mM Metal	Rate (h^{-1}) 50 mM Metal	Rate (h^{-1}) 100 mM Metal
Mg^{2+}	0.17 ± 0.07	0.82 ± 0.11	1.86 ± 0.13	4.06 ± 0.77	6.71 ± 0.89
Mn^{2+}	0.17 ± 0.07	6.80 ± 1.00	12.8 ± 1.1	25.1 ± 1.8	34.9 ± 7.0
CoHex	0.17 ± 0.07	0.53 ± 0.07	0.77 ± 0.13	1.02 ± 0.27	1.16 ± 0.31

Primer extension reactions were performed as described in the supporting information. Rates were measured by primer disappearance visualized by separation by 20% (19:1) denaturing PAGE. Values are averages and errors are standard deviations calculated from triplicate reactions.

metal ion concentration on the rate of primer extension. In all cases, the rate of RNA copying increase with the concentration of metal ions. However, CoHex seems to reach its maximum catalytic capacity at a fairly low concentration; additional CoHex beyond 50 mM only gives a minor improvement in the rate of reaction. (Table 4.1 and Figure C.2)

4.3.2 Isothermal Titration Calorimetry

To better understand our results from the kinetic studies of primer extension reactions, we first examined the effect of the different metal species on the thermodynamics of nucleic acid duplex formation. We performed isothermal titration calorimetry (ITC) on RNA 7-mers at metal concentrations ranging from 0 - 100 mM (Figure 4.3). The sequences of the complementary RNA 7-mers (5′-ACUGUCA-3′ and 5′- UGACAGU-3′) were chosen for having a roughly equal mixture of purines/pyrimidines and to minimize the formation of structures other than the fully complementary heterodimer. Mg^{2+} had the most favorable effect on RNA duplex formation of the metal species tested. 100 mM Mg^{2+} improved the favorability of RNA duplex formation by –1.1 kcal mol^{-1} (Table 4.2). A concentration of 100 mM CoHex improved the favorability of RNA duplex formation to almost the same extent as 100 mM Mg^{2+} ($\Delta\Delta G$ = –0.9 and –1.1 kcal mol^{-1}, respectively). Mn^{2+} had a relatively insignificant effect ($\Delta\Delta G$ = –0.1 kcal mol^{-1}) which indicates that it did not significantly alter the favorability of RNA duplex formation. It is noteworthy that the effect of CoHex concentration on non-enzymatic RNA copying has a similar pattern as the effect of CoHex concentration on the favorability of RNA duplex formation. (Figure C.2C and

Figure 4.3B). This observation implies that the catalytic efficiency of CoHex is highly dependent on its ability to stabilize of RNA duplex.

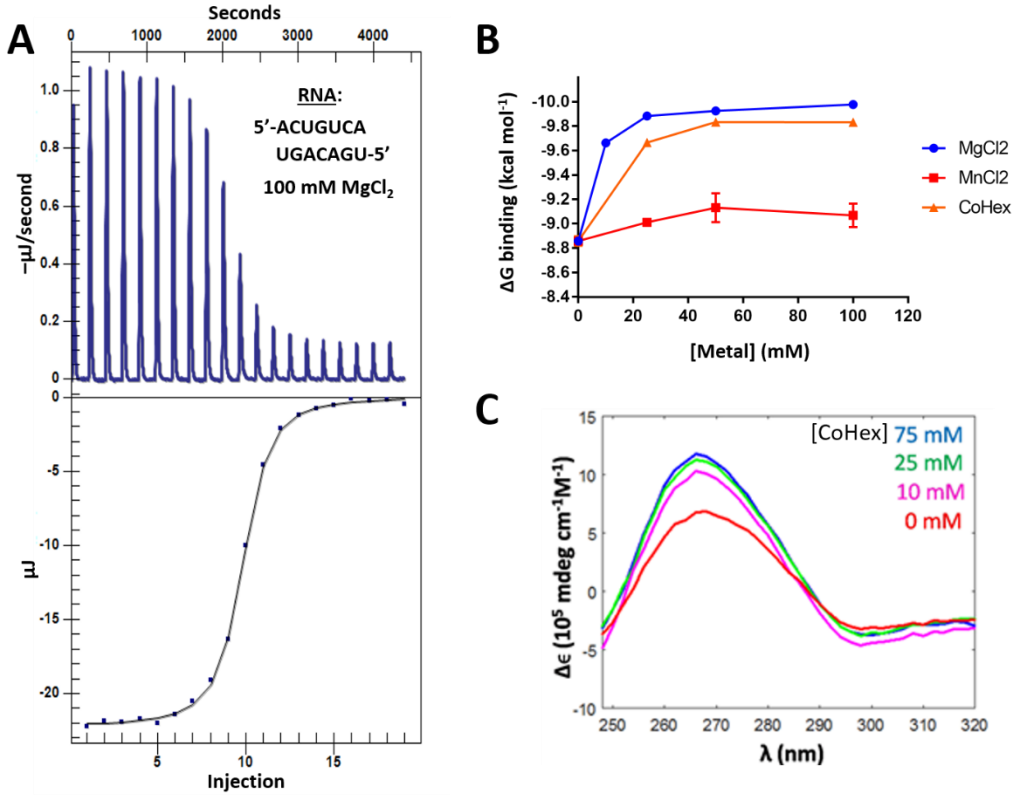


Figure 4.3: Metal-dependent thermodynamic favorability of RNA duplex formation. (A) Representative ITC data of RNA duplex formation. 5′-ACUGUCA (50 µM) in 200 mM Tris-HCl pH 7.5, 100 mM Mg^{2+} was titrated at 2.5 µL per injection into a solution containing its complement, 5′- UGACAGU (7.5 µM) in an identical buffer solution at 25 ˚C. Top: Raw curve of power versus time. Bottom: Fit of integrated peak areas (heat released per injection) yielding the thermodynamic parameters presented in Table 4.2; (B) ΔG of binding calculated from ITC for a complementary RNA 7-mer (bottom) at metal concentrations ranging from 0-100 mM. Error bars represent are double the difference between replicate measurements. (C) Circular dichroism Studies of dsRNA with increasing concentrations of CoHex. Overlapped spectra obtained by UV-CD spectroscopy at 50 ˚C with 100 µM total oligonucleotide in 200 mM HEPES at pH 7.5.

Table 4.2: **The thermodynamic parameters of duplex formation measured by isothermal titration calorimetry and thermal denaturation.**

Metal[a]	K_d^{ITC} [b] (nM)	ΔG^{ITC} [c] (kcal mol^{-1})	$\Delta\Delta G^{ITC}$ [d] (kcal mol^{-1})	T_M^{UV} [e] (˚C)	ΔG^{vH} [f] (kcal mol^{-1})
–	320(8)	-8.9(2)	-	315(2)	–8.5
Mg^{2+}	49(2)	–10.0(1)	–1.1	327(2)	–11.7
Mn^{2+}	225(19)	–9.0(1)	–0.1	320(2)	–10.8
CoHex	62(1)	–9.8(1)	–0.9	327(1)	-

[a] Metal concentrations are 100 mM except for blank entries which contained no added metal salts. [b] K_d^{ITC} values were evaluated directly from ITC titration data using an independent fit in NanoAnalyze software from TA instruments. ITC experiments were performed at 25 ˚C. Values of K_d^{ITC} are averages of duplicate experiments and errors correspond to double the difference between replicates; [c] ΔG^{ITC} was calculated using values of K_d^{ITC} from the ITC data according to $\Delta G^{ITC} = -RT \ln(K_d^{ITC})$ using the experimental temperature in K; [d] As compared to a solution without added metal salts; [e] T_M^{UV} was calculated from a double baseline, two-state fit of raw optimal melt data (see materials and methods for additional details). T_M^{UV} values are averages and the errors correspond to double the differences between the measured values; [f] ΔG^{vH} was calculated from thermal denaturation data collected in duplicate and according to $\Delta G^{vH} = \Delta H^{vH} - T\Delta S^{vH}$, where values of ΔH^{vH} and ΔS^{vH} were derived from linear fits of van 't Hoff plots of $1/T_M$ versus ln ($C_T/4$) where C_T is the total oligonucleotide concentration. The R^2 values of each plot exceed 0.99.

4.3.3 Thermal Denaturation of Nucleic Acid Duplexes

To further confirm the key results of the ITC experiments, thermal denaturation experiments were performed on the same RNA 7-mer duplexes. We measured the melting temperatures (TM) of the duplexes in 200 mM HEPES at pH 7.5 using variable temperature UV absorbance methods (Table 4.2).[26] The resulting melting curves were fit with a double-baseline, two-state model[27-28] (Figure 4.4) and the resulting values of

T_M^{-1} were plotted as functions of ln(C_T/4), where C_T is the total RNA concentration (Figure C.3). The slope and intercept of the resulting van 't Hoff plots afford estimations of ΔG^{vH} for duplex formation.[29-30] Duplicate melts were performed at oligonucleotide concentrations ranging from 6.75–200 μM with either no added metal salt or in the presence of 100 mM of Mg^{2+}, Mn^{2+}, or CoHex. Although the exact values of thermodynamic parameters obtained by van 't Hoff analysis of optical melting are not expected to perfectly agree with those obtained by ITC at constant temperature, agreement should be expected in the overall trends of ΔG (ΔG^{ITC} vs. ΔG^{UV}, Table 4.2).

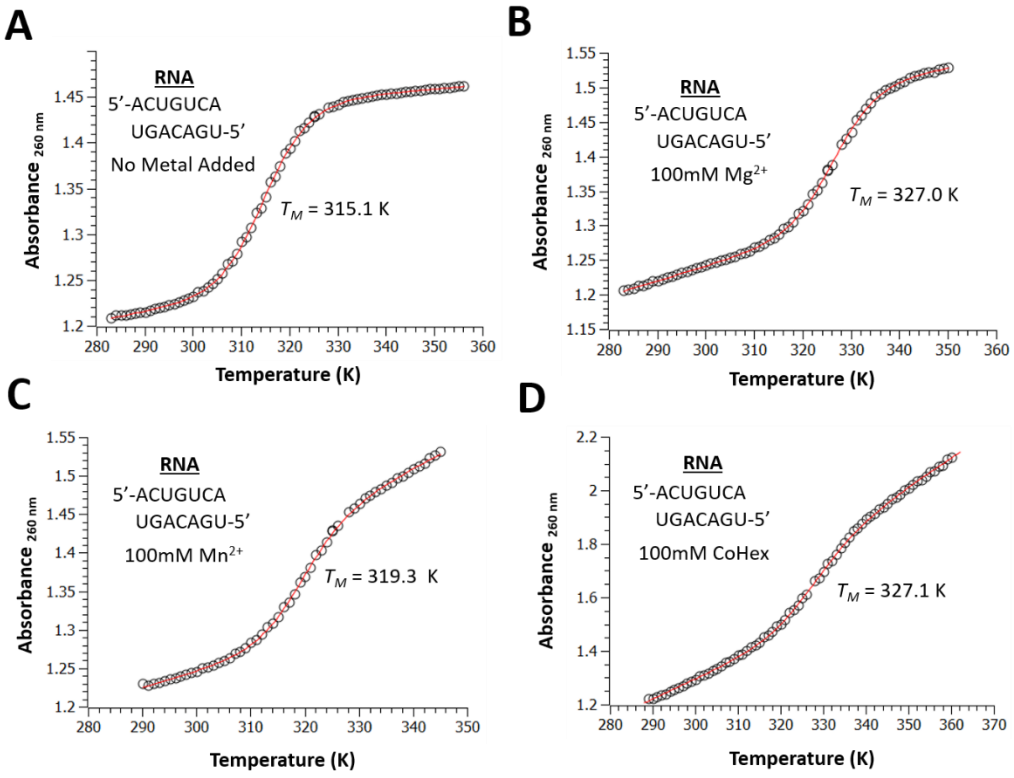


Figure 4.4: Thermal Denaturation of Nucleic Acid Duplexes in the presence of different metal ions. Representative melting curves (black circles) collected during the thermal denaturation of 200 μM RNA 7-mer in 200 mM HEPES, pH 7.5 with no additional metal ions or 100 mM of indicated metal ions. The double-baseline, two-state fitted model, shown in red line, was used to determine the T_m. (A) No metal ions added (B) 100 mM Mg^{2+} (C) 100 mM Mn^{2+} (D) 100 mM CoHex.

In the case of the RNA duplexes, the addition of 100 mM Mn^{2+} had a significant effect on the favorability of duplex formation ($\Delta\Delta G$ = –2.3 kcal mol^{-1}) but was still inferior to the effect of the addition of 100mM Mg^{2+} ($\Delta\Delta G$ = –3.2 kcal mol^{-1}). (Table 4.2) This comparison also agrees with the trends observed in the ITC studies. Because CoHex absorbs strongly at 260 nm, thermal denaturation studies could not be performed in the presence of 100 mM CoHex at oligonucleotide concentration below 200 μM. Thus, a concentration-dependent van 't Hoff analysis could not be performed using 100 mM CoHex. Despite the interference of CoHex, the signal at an oligonucleotide concentration of 200 μM is sufficient for the reliable measurement of the T_M. (Figure 4.4D) The addition of 100 mM CoHex had an identical impact on the T_M of the RNA 7-mer duplex as the addition of 100 mM Mg^{2+} (ΔT_M = 13 ˚C for both metal species). The above trends are all consistent with the results of the ITC experiments.

4.3.4 Circular Dichroism

Both thermodynamic studies described above indicate that CoHex strongly enhances the free energy of duplex formation for RNA. To better understand this phenomenon and to gain insight into the structural effects of CoHex on nucleic acid tertiary structure, we proceeded with circular dichroism (CD) studies using our RNA constructs in duplex form in the presence of various concentrations of CoHex (Figure 4.3C). As expected, the CD spectra of dsRNA tested featured a peak near 270 nm.[31] This peak at 270 nm in the CD spectrum of the dsRNA increased in intensity as the concentration of CoHex was increased from 0 to 75 mM. (Figure 4.3C) These results

further suggest that CoHex affects the detailed structure of dsRNA. The possible origin of this phenomenon may be that CoHex binds inside the narrow and deep major groove of A-form dsRNA. This possibility is supported by molecular dynamics studies.[32] CD studies also suggest that the structure of dsRNA is affected by increasing concentration of Mg^{2+} or Mn^{2+}. (Figure C.4)

4.3.5 Metal Catalyzed Hydrolysis of Activated-Substrates

The role of metal ions in catalyzing the bond formation responsible for oligonucleotide polymerization is similar to that in catalyzing the hydrolysis of activated monomers, where H_2O is used as nucleophile instead of the 2′ or 3′ hydroxyl on the terminus of the primer. To determine the extent to which the metal ions investigated here contribute to substrate degradation, we measured the metal-catalyzed rates of activated-monomer hydrolysis by HPLC (Table C.1). As expected, the metal species that afford the greatest rate enhancements for primer extension also resulted in the greatest rate enhancements for hydrolysis. 100 mM Mn^{2+} resulted in nearly triple the rate of activated monomer hydrolysis (34×10^{-3} h^{-1}) compared to that with 100 mM of Mg^{2+} (11×10^{-3} h^{-1}). The addition of 100 mM of CoHex has little or no effect to the rate of activated monomer hydrolysis. (8.4×10^{-3} h^{-1} vs 8.9×10^{-3} h^{-1} in the absence of multivalent metal).

4.3.6 Nonenzymatic Primer Extension with Combinations of Metal Species

The above primer extension results along with the results of the ITC and thermal denaturation studies suggested to us that CoHex may be well suited to shielding RNA backbone charges but poorly suited to phosphate activation. This insight is supported by the fact that CoHex is incapable of forming the inner shell contacts likely necessary for phosphate activation. Because the overall rate of RNA copying depends on both the formation of stable duplexes as well as phosphate activation, we became interested in the orthogonal use of these metal species. By depending on CoHex for charge shielding, we presumed that we may be able to either enhance or maintain high rates of RNA copying at lower concentrations of the metals responsible for phosphate activation (Mg^{2+} or Mn^{2+}) and potentially reducing the effects of these metals on monomer hydrolysis.

To investigate this possibility, we first performed primer extensions in buffers containing Mg^{2+} or Mn^{2+} in combination with different concentrations of CoHex (Figure 4.5 and Table 4.3). Primer extension reactions performed with the combination of 50 mM Mg^{2+} and 50 mM CoHex were found to be approximately 70% faster than the reaction with just 50 mM Mg^{2+} (6.9 h^{-1} compared to 4.1 h^{-1}). Meanwhile, the addition of just 10 mM CoHex to reactions containing 10 mM Mn^{2+} more than doubled the rate of primer extension observed for reactions with only 10 mM Mn^{2+} (13.9 h^{-1} compared to 6.8 h^{-1}). This acceleration far exceeds the sum of the independent effects of the metals at these concentrations (7.3 h^{-1}, Table 4.3). A similar multiplicative effect was present at all combinations of Mn^{2+} and CoHex tested (Table 4.3). The addition of 10 mM CoHex to

only 10 mM of Mn^{2+} resulted in a rate ca. 3 fold greater than that with 10 mM Mn^{2+} alone (19.0 h^{-1} vs. 6.80 h^{-1}, respectively) and ca. 16 fold greater than that with 100 mM CoHex alone (19.0 h^{-1} vs. 1.16 h^{-1}, respectively). By combining 50mM of Mn^{2+} with 100 mM CoHex, we achieved extension rates exceeding 40 h^{-1}, the highest rates observed in this study and greatly exceeding the highest rates previously reported in an all RNA

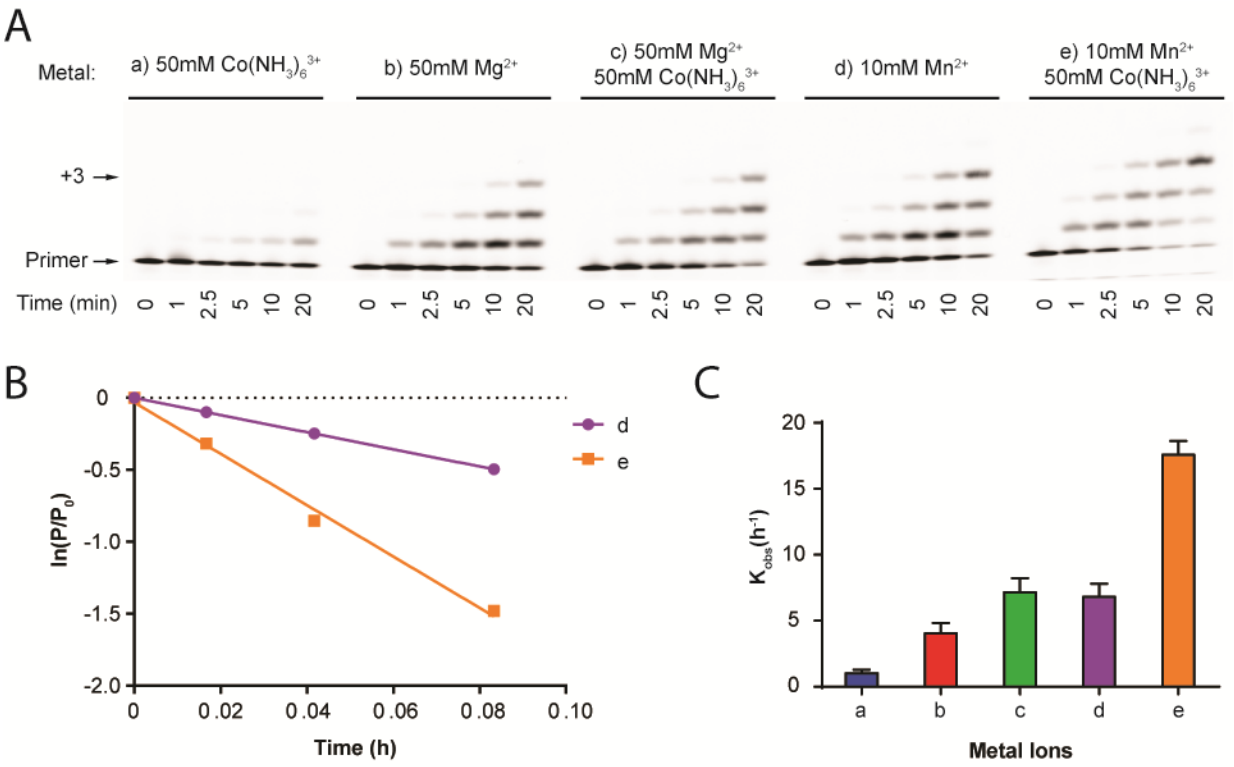


Figure 4.5: The synergistic effects of metal ions on non-enzymatic primer extension. (**A**) Select high-resolution PAGE analyses of primer-extension products under various conditions of metal concentration. The primer-extension reaction mixtures contained 1 μM Cy5-labeled RNA primer, 5μM template, 200 mM Tris-HCl (pH 7.5), 50 mM 2-MeImpG monomer and indicated amount of metal chloride. (**B**) The effect of 50mM CoHex on the rates of reactions containing 10 mM Mn^{2+}. The rate of primer extension was determined by the plot of the natural log of the fraction of unextended primer remaining compared to initial primer, $\ln(P/P_0)$, vs time. Lines are linear fits, $R^2 \geq 0.99$. (**C**) Rates of primer extension observed under various conditions of metal concentration.

system to the best of our knowledge. An ideal primer extension catalyst would greatly accelerate the rate of monomer incorporation without contributing to substrate degradation. Interestingly, when compared to 100 mM Mn^{2+}, the combination of 50 mM Mn^{2+} and 50 mM CoHex can provide an improved rate of primer extension (34.9 h^{-1} vs 39.9 h^{-1}) and a significantly reduced rate of activated monomer hydrolysis (34 x 10^{-3} h^{-1} vs 13 x 10^{-3} h^{-1}). (Table C.1)

Table 4.3: **Primer Extension Rates with Combinations of Metal Species**

[Mg^{2+}] (mM)	[CoHex] (mM)	Rate (h^{-1})	[Mn^{2+}] (mM)	[CoHex] (mM)	Rate (h^{-1})
10	50	2.3 ± 0.2	10	50	17.6 ± 1.1
25	50	3.8 ± 0.2	25	50	31.8 ± 3.1
50	50	6.9 ± 0.8	50	50	39.9 ± 1.9
100	50	9.2 ± 1.3	100	50	40.2 ± 4.5
50	10	5.4 ± 0.6	10	10	13.8 ± 1.6
50	25	6.4 ± 0.6	10	25	15.9 ± 1.8
50	50	6.9 ± 0.8	10	50	17.6 ± 1.1
50	100	8.3 ± 0.7	10	100	19.0 ± 0.2
			50	10	34.0 ± 2.5
			50	25	37.8 ± 4.3
			50	50	39.9 ± 1.9
			50	100	43.1 ± 1.9

Primer extension reactions were performed as described in the supporting information. Rates were measured by primer disappearance visualized by separation by 20% (19:1) denaturing PAGE. Values are averages and errors are standard deviations calculated from triplicate reactions.

4.4 Discussion

The concentration-dependent effects of metal ions on primer extension, duplex formation, and activated monomer hydrolysis were compared. The magnitudes of the effects roughly increased with metal concentration across these investigations. Increased concentrations of Mg^{2+} and CoHex effectively enhanced the favorability of RNA duplex formation. The fact that CoHex is inactive in inner-sphere contacts[33] suggests that outer shell contacts dominate the interactions between metal species and the phosphate backbone. However, increased concentrations of CoHex had an insignificant impact on the rate of primer extension compared to the hexaaqua Mg^{2+} and Mn^{2+}. A similar trend was observed for hydrolysis studies of activated monomers. This result suggests that inner-sphere binding is necessary for effective metal-dependent activation at the reaction site. The relatively small impact of CoHex on the rate of primer extension may result entirely from enhancing the structural stability of the primer-template-monomer reaction complex.

More difficult to justify are the differences between the effects of Mg^{2+} and Mn^{2+} ions. Both species are divalent and permit inner and outer shell contacts, yet, Mg^{2+} ions resulted in significantly more favorable duplex formation, but far inferior rates of primer extension and hydrolysis compared to Mn^{2+}. Both experimental and computational studies have shown that Mn^{2+} decreases the pKa of bound water more significantly than Mg^{2+}.[24] This suggests that a sugar hydroxyl coordinated to Mn^{2+} is more readily deprotonated and, therefore, a better nucleophile. Manganese ions' inferior ability to stabilize the RNA duplex when compared to Mg^{2+} is likely due to their differences in size and charge density.[34]

Combining the actions of CoHex and Mg^{2+} created a moderate enhancement in the rate of primer extension. Surprisingly, combining the actions of CoHex and Mn^{2+} resulted in rates several fold greater than what would be expected from adding their independent effects alone. This observed synergistic effect suggests that the actions of a metal ion well-suited to increasing the favorability of duplex formation may be orthogonal to the actions of a metal ion well-suited to phosphate activation. This hypothesis suggests that there exist multiple metal binding sites on the reaction complex and that different metals are differentially suited to interacting with each site. Mn^{2+} has an insignificant effect on duplex formation and a strong catalytic effect on phosphate activation, suggesting that it interacts more strongly with the sugar hydroxyl group on the primer and the phosphate of the incoming activated monomer. CoHex is an excellent complement to Mn^{2+} as it enhances duplex formation while having an insignificant effect on primer extension. Furthermore, if CoHex did bind with significant affinity near the reactive site phosphate, one would expect it to compete with the binding of the more effective catalyst, Mn^{2+}, inhibiting the action of Mn^{2+} and reducing the rate of the reaction. We observe no such effect. Our kinetic analysis of the primer extension reaction suggests that Mn^{2+} has a greater effect on binding of the activated monomer substrate to the template than either Mg^{2+} and CoHex. The combined effects of both ions apparently afford a strongly synergistic effect, likely because CoHex compensates for the shortcomings of Mn^{2+} and vice versa and neither ion significantly competes with the mechanism of action of the other. This model also explains why the combined effect of CoHex and Mg^{2+} is comparatively moderate. Both metals significantly enhance the favorability of duplex formation, and thus, likely compete for binding at the backbone

phosphates. Meanwhile, any CoHex bound near the reaction site is unlikely to significantly enhance phosphate activation due to the metal's inability to form inner shell contacts with phosphate oxygens, ribose hydroxyls, or surrounding waters.

The synergistic combination of CoHex and Mn^{2+} has one further major advantage. Elevated concentrations of Mn^{2+} result in the rapid hydrolysis of oligoribonucleotides and activated monomers while CoHex has a comparatively negligible effect. By augmenting low concentrations (i.e. 10 mM) of Mn^{2+} with greater amounts of CoHex, rates of primer extension can remain high while the rate of monomer hydrolysis remains low.

Our lab previously identified eight major obstacles preventing the realization of a non-enzymatic RNA replication system suitable for use in an artificial cell.[13] These included: the poor rate of template copying chemistry and high rate of metal-catalyzed monomer-substrate hydrolysis. This work directly addresses these limitations. By combining the actions of various metals with orthogonal and synergistic properties, we have demonstrated the ability to enhance the rate of RNA copying while avoiding the pitfalls normally associated with metal-dependent activation. The conditions that resulted in the first living cells likely featured complex mixtures of various metal species. It would be unsurprising if similar conditions were not crucial for the realization of a genetic system suitable for use in a simple artificial cell.

4.5 References

1. Wachowius, F.; Attwater, J.; Holliger, P., Nucleic acids: function and potential for abiogenesis. *Q Rev Biophys* **2017,** *50*, e4.

2. Sutherland, J. D., The Origin of Life—Out of the Blue. *Angewandte Chemie International Edition* **2016,** *55* (1), 104-121.

3. Inoue, T.; Orgel, L. E., Substituent control of the poly(C)-directed oligomerization of guanosine 5'-phosphoroimidazolide. *Journal of the American Chemical Society* **1981,** *103* (25), 7666-7667.

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

5. Heuberger, B. D.; Pal, A.; Del Frate, F.; Topkar, V. V.; Szostak, J. W., Replacing Uridine with 2-Thio-Uridine Enhances the Rate and Fidelity of Nonenzymatic RNA Primer Extension. *J. Am. Chem. Soc.* **2015,** *137* (7), 2769–2775.

6. Deck, C.; Jauker, M.; Richert, C., Efficient enzyme-free copying of all four nucleobases templated by immobilized RNA. *Nat Chem* **2011,** *3* (8), 603-608.

7. Vogel, S. R.; Deck, C.; Richert, C., Accelerating chemical replication steps of RNA involving activated ribonucleotides and downstream-binding elements. *Chem. Commun.* **2005,** (39), 4922-4924.

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

9. Prakash, T. P.; Roberts, C.; Switzer, C., Activity of 2′,5′-Linked RNA in the Template-Directed Oligomerization of Mononucleotides. *Angewandte Chemie International Edition in English* **1997,** *36* (13‐14), 1522-1523.

10. Koculi, E.; Hyeon, C.; Thirumalai, D.; Woodson, S. A., Charge density of divalent metal cations determines RNA stability. *Journal of the American Chemical Society* **2007,** *129* (9), 2676-2682.

11. Misra, V. K.; Draper, D. E., On the role of magnesium ions in RNA stability. *Biopolymers* **1998,** *48* (2-3), 113-135.

12. Misra, V. K.; Draper, D. E., The interpretation of Mg2+ binding isotherms for nucleic acids using Poisson-Boltzmann theory11Edited by B. Honig. *Journal of Molecular Biology* **1999,** *294* (5), 1135-1147.

13. Szostak, J. W., The eightfold path to non-enzymatic RNA replication. *Journal of Systems Chemistry* **2012,** *3* (1), 2.

14. Romani, A. M. P., CELLULAR MAGNESIUM HOMEOSTASIS. *Archives of biochemistry and biophysics* **2011,** *512* (1), 1-23.

15. Bridson, P. K.; Orgel, L. E., Catalysis of Accurate Poly(C)-directed Synthesis of 3'-5'-linked Oligoguanylates by Zn2+. *J Mol Biol* **1980,** *144*.

16. Lohrmann, R.; Orgel, L. E., Efficient catalysis of polycytidylic acid-directed oligoguanylate formation by Pb2+. *Journal of Molecular Biology* **1980,** *142* (4), 555-567.

17. Lohrmann, R.; Bridson, P. K.; Orgel, L. E., Efficient metal-ion catalyzed template-directed oligonucleotide synthesis. *Science* **1980,** *208* (4451), 1464.

18. Rohatgi, R.; Bartel, D. P.; Szostak, J. W., Kinetic and Mechanistic Analysis of Nonenzymatic, Template-Directed Oligoribonucleotide Ligation. *Journal of the American Chemical Society* **1996,** *118* (14), 3332-3339.

19. 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. *Journal of the American Chemical Society* **1999,** *121* (12), 2653-2656.

20. Robinson, H.; Wang, A. H. J., Neomycin, Spermine and Hexaamminecobalt(III) Share Common Structural Motifs in Converting B- to A-DNA. *Nucleic Acids Research* **1996,** *24* (4), 676-682.

21. Zhang, W.; Walton, T.; Li, L.; Szostak, J. W., Crystallographic observation of nonenzymatic RNA primer extension. *eLife* **2018,** *7*, e36422.

22. Kanavarioti, A.; Bernasconi, C. F.; Doodokyan, D. L.; Alberas, D. J., Magnesium ion catalyzed P-N bond hydrolysis in imidazolide-activated nucleotides. Relevance to template-directed synthesis of polynucleotides. *J Am Chem Soc* **1989,** *111* (18), 7247-57.

23. Butzow, J. J.; Eichhorn, G. L., Interaction of metal ions with nucleic acids and related compounds. XVII. On the mechanism of degradation of polyribonucleotides and oligoribonucleotides by zinc(II) ions. *Biochemistry* **1971,** *10*.

24. Jackson, V. E.; Felmy, A. R.; Dixon, D. A., Prediction of the pKa's of Aqueous Metal Ion +2 Complexes. *The Journal of Physical Chemistry A* **2015,** *119* (12), 2926-2939.

25. Grant, C. V.; Frydman, V.; Harwood, J. S.; Frydman, L., 59Co Solid-State NMR as a New Probe for Elucidating Metal Binding in Polynucleotides. *J. Am. Chem. Soc.* **2002,** *124* (16), 4458-4462.

26. Košutić, M.; Jud, L.; Da Veiga, C.; Frener, M.; Fauster, K.; Kreutz, C.; Ennifar, E.; Micura, R., Surprising Base Pairing and Structural Properties of 2′-Trifluoromethylthio-Modified Ribonucleic Acids. *J. Am. Chem. Soc.* **2014,** *136* (18), 6656-6663.

27. Schroeder, S. J.; Turner, D. H., Optical melting measurements of nucleic acid thermodynamics. *Methods Enzymol* **2009,** *468*, 371-87.

28. 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,** *38* (50), 16655-62.

29. Petersheim, M.; Turner, D. H., Base-stacking and base-pairing contributions to helix stability: thermodynamics of double-helix formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp. *Biochemistry* **1983,** *22* (2), 256-263.

30. Marky, L. A.; Breslauer, K. J., Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers* **1987,** *26* (9), 1601-1620.

31. Steely, H. T.; Gray, D. M.; Lang, D.; Maestre, M. F., Circular dichroism of double-stranded RNA in the presence of salt and ethanol. *Biopolymers* **1986,** *25* (1), 91-117.

32. Tolokh, I. S.; Pabit, S. A.; Katz, A. M.; Chen, Y.; Drozdetski, A.; Baker, N.; Pollack, L.; Onufriev, A. V., Why double-stranded RNA resists condensation. *Nucleic Acids Research* **2014,** *42* (16), 10823-10831.

33. Rowinska-Zyrek, M.; Skilandat, M.; Sigel, R. K. O., Hexaamminecobalt(III) – Probing Metal Ion Binding Sites in Nucleic Acids by NMR Spectroscopy. *Zeitschrift für anorganische und allgemeine Chemie* **2013,** *639* (8‐9), 1313-1320.

34. Marcus, Y., Ionic radii in aqueous solutions. *Chemical Reviews* **1988,** *88* (8), 1475-1498.

Appendix A

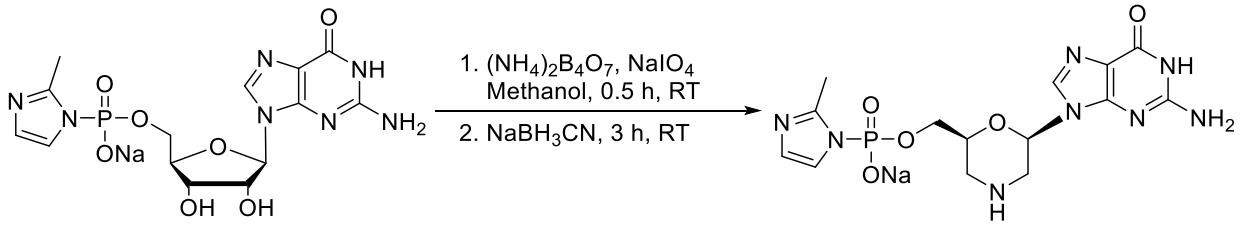
Supplemental Materials for Chapter 2

A.1 General information

All solvents and reagents were reagent grade, purchased commercially, and used without further purification unless specified. Oligonucleotides used as primers or templates were synthesized on an Expedite nucleic acid synthesizer (Applied BioSystems) or purchased from IDT (Coralville, IA) unless otherwise indicated. All chemicals were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise indicated. All nucleoside-5′-monophosphates were purchased as the free acids from Santa Cruz Biotechnology (Dallas, TX). Reverse phase flash chromatography was performed using a RediSep Rf Gold C18Aq 50 g column purchased from Teledyne Isco (Lincoln, NE). All Nuclear Magnetic Resonance (NMR) real-time studies and NMR spectra were recorded on a Varian Oxford AS-400 NMR spectrometer. ^{1}H and ^{31}P NMR spectra were acquired at 400 MHz and 161.8 MHz, respectively. Chemical shifts were reported in parts per million. Electrospray mass spectra were recorded on a Bruker Daltonics Esquire 6000 ESI-MS. LC-MS studies of oligonucleotides were carried out on Agilent Q-TOF LC/MS system.

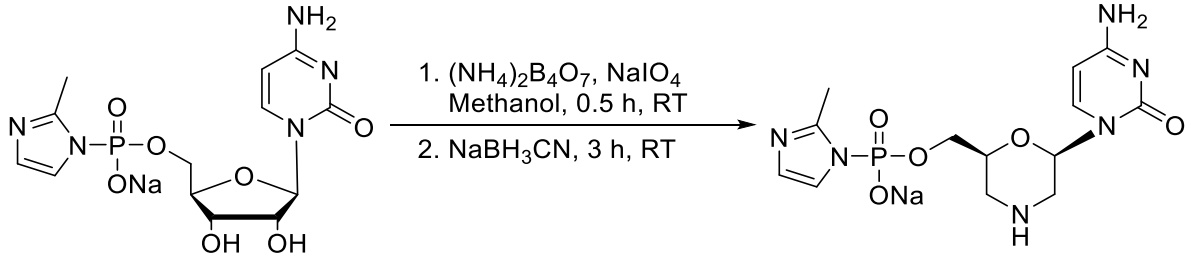
A.2 Synthesis of 5′-phosphor-2-methylimidazolide morpholino nucleosides.

A.2.1 Synthesis of 5′-phosphor-2-methylimidazolide morpholino guanosine (2-MeImpmG)



2-MeImpG (100 mg, 0.223 mmol), ammonium biborate (176 mg, 0.668 mmol, 3.0 equiv) and sodium periodate (143 mg, 0.668 mmol, 3.0 equiv) were stirred in anhydrous methanol (5mL) in an oven-dried round bottom flask at room temperature under an inert atmosphere. After 30 min, sodium cyanoborohydride (140 mg, 2.23 mmol, 10 equiv) was added to the reaction mixture and stirring was continued at room temperature. After 3 hours, solvent was removed in a rotary evaporator under reduced pressure. The resultant solid residue was dissolved in 3 mL buffer (20 mM TEAB buffer, pH 10) and purified by C18 column chromatography on a CombiFlash using gradient elution between (A) 20 mM TEAB buffer, pH 10 and (B) acetonitrile. The elution gradient was between 0% and 5% B over 4 CVs with a flow rate of 40mL/min. 0.1M NaOH was added drop wise to fractions containing product to adjust the pH to 10. The product was then lyophilized at -15 °C. HRMS (Q-TOF) m/z: $[M-H]^-$ calcd for $C_{14}H_{19}N_8O_5P$: 410.1216; Found: 410.1269

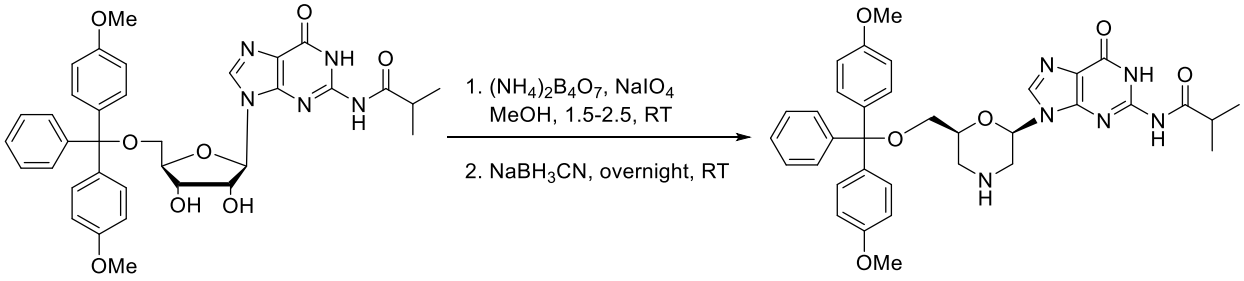
A.2.2 Synthesis of 5′-phosphor-2-methylimidazolide morpholino cytidine (2-MeImpmC)



2-MeImpC (100 mg, 0.244 mmol), ammonium biborate (193 mg, 0.733 mmol, 3.0 equiv) and sodium periodate (157 mg, 0.733 mmol, 3.0 equiv) were stirred in anhydrous methanol (5mL) in an oven-dried round bottom flask at room temperature under an inert atmosphere. After 30 min, sodium cyanoborohydride (154 mg, 2.44 mmol, 10 equiv) was added to the reaction mixture and stirring was continued at room temperature. After 3 hours, the solvent was removed in a rotary evaporator under reduced pressure. The resultant solid residue was dissolved in 3 mL buffer (20 mM TEAB buffer, pH 10) and purified by C18 column chromatography on a CombiFlash using gradient elution between (A) 20 mM TEAB buffer, pH 10 and (B) acetonitrile. The elution gradient was between 0% and 5% B over 4 CVs with a flow rate of 40mL/min. 0.1M NaOH was added drop wise to fractions containing product to adjust the pH to 10. The product was then lyophilized at -15 °C. HRMS (Q-TOF) m/z: $[M-H]^-$ calcd for $C_{13}H_{19}N_6O_5P$ 370.1155; Found: 370.1201.

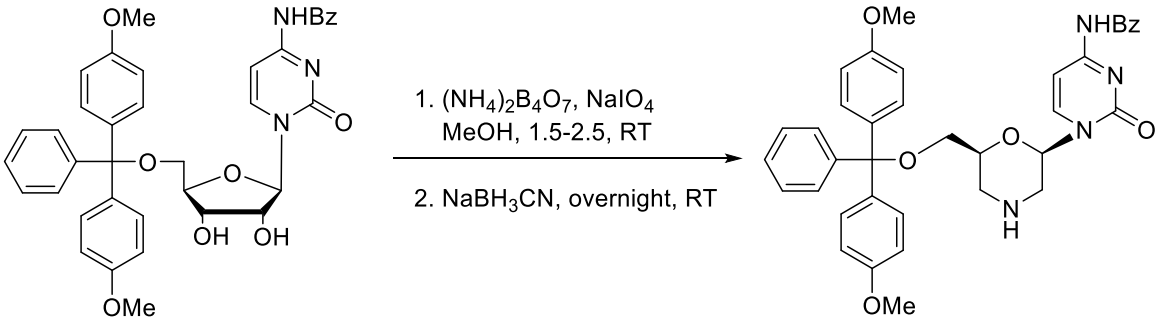
A.3 Preparation of solid supports loaded with morpholino nucleosides

A.3.1 Synthesis of 5′-O-DMTr-morpholino guanosine



5′-O-DMTr-G^{iBu} (5.00 g, 7.63mmol) was dissolved in 100 ml methanol. Ammonium biborate tetrahydrate (2.41 g, 9.15 mmol, 1.2 equiv) and sodium periodate (1.96 g, 9.15 mmol, 1.2 equiv) were added at room temperature with stirring. The reaction mixture was allowed to stir for 2 h. ESI-MS was used to monitor the reaction progress. After complete consumption of the starting material, the reaction mixture was centrifuged at 4000 rpm for 15 mins to remove excess reagents and byproducts. Sodium cyanoborohydride (958 mg, 15.3 mmol, 2.0 equiv) was added to the supernatant. After 30 min, another equivalent of sodium cyanoborohydride (479 mg, 7.63 mmol, 1.0 equiv) was added. The reaction mixture was allowed to stir for another 6 h, at which point the intermediate product was completely reduced. The solvent was then removed by rotary evaporation under reduced pressure. The resulting residue was dissolved in ethyl acetate (200 ml) and washed with brine three times. The solution was then dried over sodium sulfate and the solvent removed under reduced pressure. The crude product was purified by column chromatography (silica column, 0-10% MeOH in DCM) to afford the desired product as a white solid. ESI-MS m/z: $[M-H]^-$ calcd for $C_{35}H_{38}N_6O_6$ 637.28; Found: 637.3.

A.3.2 Synthesis of 5′-O-DMTr-morpholino cytidine



5′-O-DMTr-C^{Bz} (250 mg, 0.385 mmol) was dissolved in 10 ml methanol. Ammonium biborate tetrahydrate (122 mg, 0.462 mmol, 1.2 equiv) and sodium periodate (99 mg, 0.462 mmol, 1.2 equiv) were added at room temperature with stirring. The reaction mixture was allowed to stir for 2 h. ESI-MS was used to monitor the reaction progress. After complete consumption of the starting material, the reaction mixture was centrifuged at 4000 rpm for 15 mins to remove excess reagents and byproducts. Sodium cyanoborohydride (48 mg, 0.770 mmol, 2.0 equiv) was added to the supernatant. After 30 min, another equivalent of sodium cyanoborohydride (24 mg, 0.385 mmol, 1.0 equiv) was added. The reaction mixture was allowed to stir for another 6h, at which point the intermediate product was completely reduced. The solvent was then removed by rotary evaporation under reduced pressure. The resulting residue was dissolved in ethyl acetate (20 ml) and washed with brine three times. The solution was then dried over sodium sulfate and the solvent removed under reduced pressure. The crude product was purified by column chromatography (silica column, 0-10% MeOH in DCM) to afford the desired product as a white solid. ESI-MS m/z: $[M-H]^-$ calcd for $C_{37}H_{36}N_4O_6$ 631.26; Found: 631.4

A.3.3 Protocol for loading morpholino nucleosides on solid support

Native long chain alkylamine controlled pore glass (lacc-CPG) was coevaporated with toluene and dried under vacuum overnight. The dried CPG was mixed with DCM under an inert atmosphere. DIEA (12 equiv), and hexafluoroglutaric anhydride (10 equiv) were then added. The mixture was shaken at room temperature overnight. The CPG was then isolated by filtration and washed sequentially with 1% TEA in DMF, MeOH, and DCM. 5′-O-DMTr-morpholino guanosine (3 equiv) in toluene were then added. The mixture was dried by rotary evaporation under reduced pressure. DMF, PPh_3 (5 equiv), and DIEA (12 equiv) were added to the residue. Then a solution of iodine (5 equiv) in DMF was added dropwise with stirring. After 16 h, the derivatized CPG was isolated by filtration, washed sequentially with 1% TEA in DMF, MeOH and DCM, and dried under vacuum.

A.4 Primer extension reaction

A.4.1 General protocol for PAGE analysis of primer extension reactions

Representative reaction protocol: 6.0 μL of 1.0 M HEPES pH 7.5 buffer, 3.0 μL nuclease-free water, 3.0 μL 10 μM primer and 3.0 μL 50 μM template were combined in a thin-walled PCR micro-tube. 15 μL of 5 mM 2-MeImpmG was added to the lid of the micro-tube cap. The reaction was initiated when the activated nucleotides were spun down into the buffered primer/template solution. 2.0 μL aliquots were removed at desired time points, and were immediately quenched by addition to 38 μL quenching buffer (8.0 M urea in 1xTBE). The sample was incubated at 90 °C for 1 minutes before the primer extension products were separated by 20% (19:1) denaturing PAGE. The gel

was scanned with a Typhoon 9410 Variable Mode Imager. The bands were quantified using the ImageQuant TL software package.

A.4.2 General protocol for LC-MS analysis of primer extension reactions

A primer extension reaction mixture (30 µL) containing 7 µM primer, 12.5 µM template, 200 mM HEPES (pH 7.5), and 3 mM activated monomers was allowed to react for 1 day. Primer and extended products were separated from monomers by Sephadex G-25 size exclusion column chromatography (GE Healthcare, Buckinghamshire, UK). The solution was desalted by ion pairing reverse phase (IP-RP) purification on a C18 ZipTip pipette tip (Millipore, Billerica, MA). The tip was wetted with acetonitrile and equilibrated with 2 M TEAA prior to sample binding. Extensive washing with 10 mM TEAA was followed by elution in 50% acetonitrile. The eluate was dried under vacuum, followed by resuspension in LC-MS grade water. The eluted sample was analyzed on an Agilent 1200 HPLC coupled to an Agilent 6230 TOF equipped with a solvent degasser, column oven, auto sampler, and diode array detector. The sample was separated by IP-RP-HPLC on a 100 mm X 1 mm Xbridge C18 column with 3.5 µm particle size (Waters, Milford, MA) using gradient elution between (A) aqueous 200 mM 1,1,1,3,3,3-hexafluoro-2-propanol with 1.25 mM triethylamine, pH 7.0, and (B) methanol. The elution gradient was from 2.5% to 90% B over 40 min with a flow rate of 125 µL/min at 60 ˚C. Samples were analyzed in negative mode from 239 m/z to 3200 m/z with a scan rate of 1 spectrum/s and the following instrument settings: drying gas flow, 8 L/min; drying gas temperature, 325 ◦ C; nebulizer pressure, 30 psig; capillary

voltage, 3500 V; fragmentor, 200 V; and skimmer, 65 V. Extracted ion chromatograms were generated in Agilent's MassHunter Qualitative Analysis software.

A.5 Supporting Tables

Table A.1: Sequences of oligonucleotides used in this study

Name	Sequence (5′→3′)
P1	Cy5-GACUGACUGmG
P2	Cy5-GACUGACUGG
P3	Cy3-AGUGAGUAACGmC
P4	Cy3-AGUGAGUAACGC$_{NH2}$
P5	Cy5-GACUGACUGmGmGmG
T1	AACCCCCCAGUCAGUC
T2	AACCCCGCGUUACUCACU
T3	AAGGGGCCAGUCAGUC
T4	AAGGGGGCGUUACUCACU
T5	AACCCGCCAGUCAGUC
T6	AAGGGCCCAGUCAGUC
T7	AACGGGCCAGUCAGUC
T8	AAGCCCCCAGUCAGUC
T9	CCAAAACCAGUCAGUC
T10	AACCCCCCCCAGUCAGUC

Cy5 is cyanine 5; Cy3 is cyanine 3; mG is morpholino modified guanosine; mC is morpholino modified cytidine; C$_{NH2}$ is 3′-amino-2′,3′-dideoxycytidine.

A.6 Supporting Figures

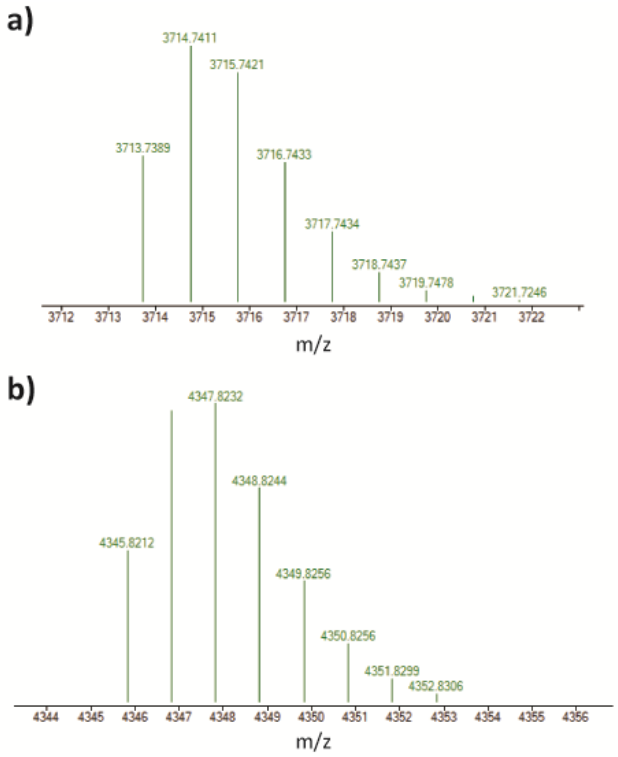


Figure A.1: LC-MS analysis of the fluorescent dye-labeled chimeric RNA/MoNA oligonucleotides. (a) 5′-Cy5 tagged 3′-morpholino G terminated RNA oligo (Cy5-GACUGACUGmG, **P1**). Chemical Formula: $C_{127}H_{157}N_{43}O_{70}P_{10}$. Exact mass: 3713.7424 Da; measured accurate mass: 3713.7389 Da. (b) 5′-Cy3 tagged 3′-morpholino C terminated RNA oligo (Cy3-AGUGAGUAACGmC, **P3**). Chemical Formula: $C_{145}H_{179}N_{53}O_{82}P_{12}$. Exact mass: 4345.8317 Da; measured accurate mass: 4345.8212 Da.

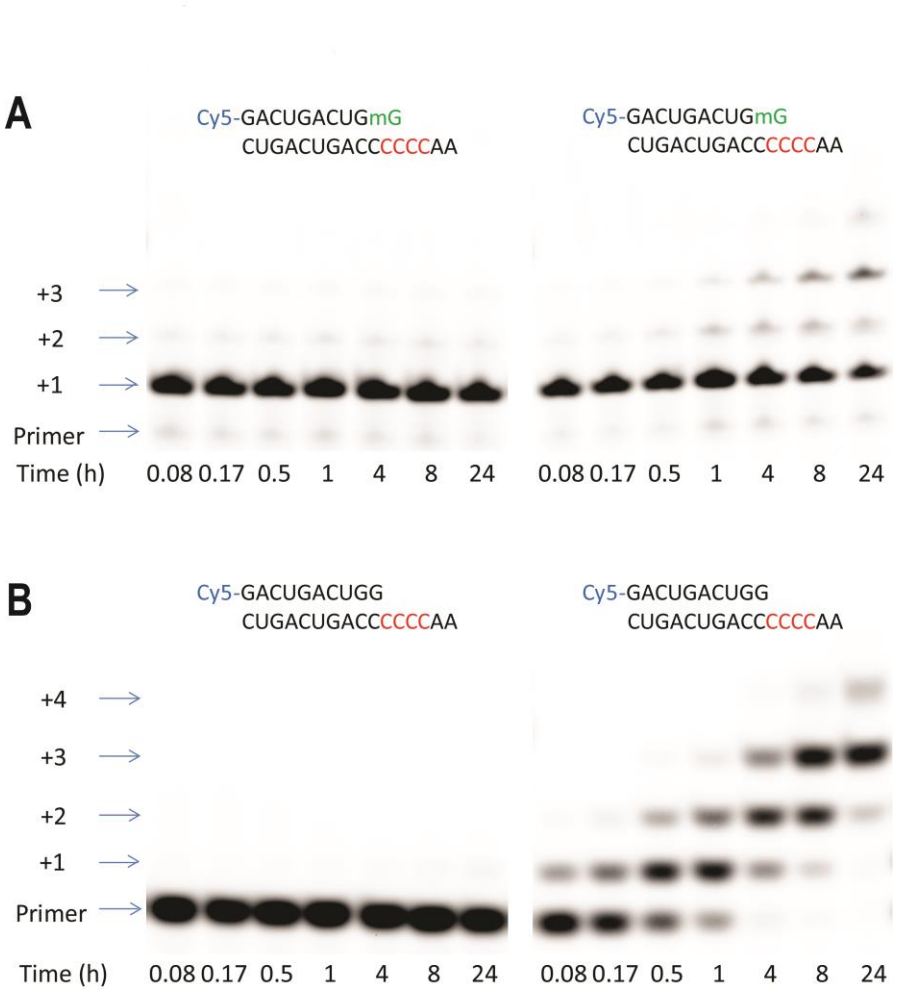


Figure A.2: Comparison between chimeric RNA/MoNA oligonucleotide and regular RNA oligonucleotide in nonenzymatic primer extension reaction in the presence of 2-MeImpG. (a) High resolution gel electrophoresis analysis of nonenzymatic primer extension reactions with 2-MeImpG on 3′-morpholino terminated oligonucleotide. Primer-extension reactions contained 1μM **P1**, 5μM **T1**, 200 mM HEPES pH 7.5, and 5 mM 2-MeImpG. 100 mM $MgCl_2$ was added to the reaction showed on the right. (b) High resolution gel electrophoresis analysis of nonenzymatic primer extension reactions with 2-MeImpG on RNA primer. Primer-extension reactions contained 1 μM **P2**, 5 μM **T1**, 200 mM HEPES pH 7.5, and 5 mM 2-MeImpG. 100 mM $MgCl_2$ was added to the reaction showed on the right.

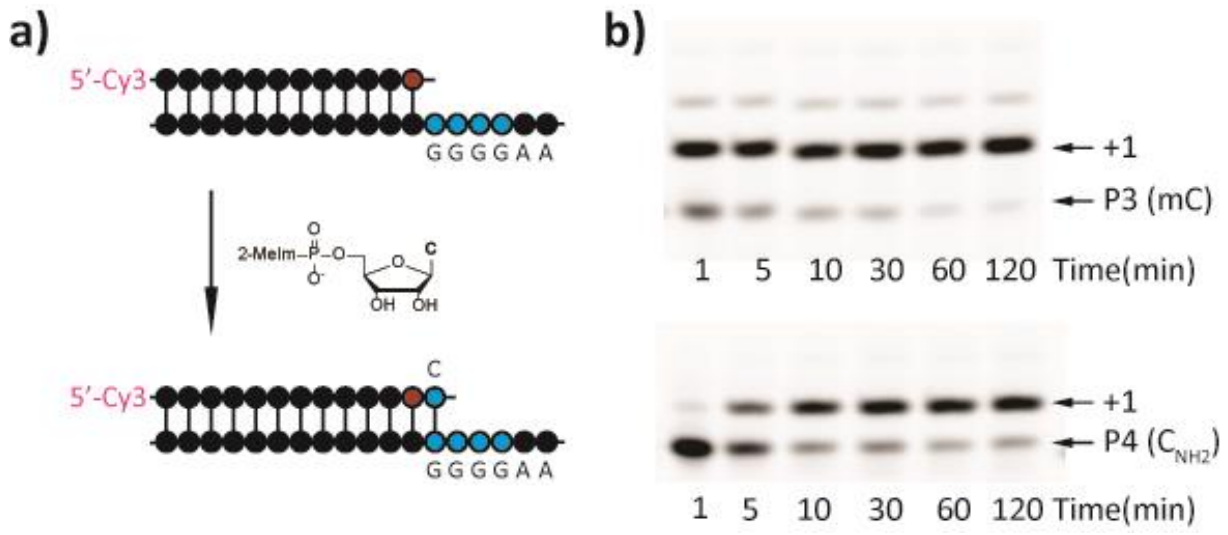


Figure A.3: Nonenzymatic primer-extension reactions with 3′-terminal amine and morpholino modified primers. (a) Schematic representation of terminal modified primer P3 or P4 annealed to a G_4 RNA template (T4) in presence of 2-MeImpC. The monomers participate in primer extension reaction on the corresponding template region resulting in predominantly N+1 product. (b) High-resolution PAGE analysis of the primer-extension products at different reaction time. Reaction conditions: 1 μM primer P3, 5 μM G_4 template T4, 200 mM HEPES pH 7.5, 5 mM 2-MeImpG. (c) PAGE analysis of the primer-extension products when P4 was used.

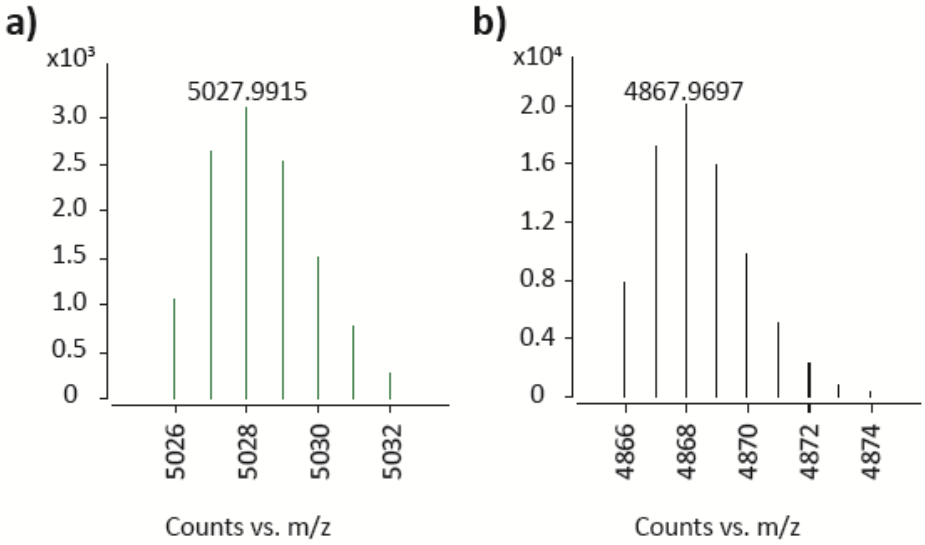


Figure A.4: LC-MS analysis of the major extended products in the template copying reaction with activated MoNA monomer. (a) Deconvoluted mass of +4 product form RNA C4 template **T1** copying reaction with 2-MeImpmG. Monoisotopic mass for the full-length primer +4 product: calculated mass, 5026.0164 Da; observed mass, 5025.9896 Da. (b) Deconvoluted mass of +4 product form RNA G4 template **T2** copying reactions with 2-MeImpmC. Monoisotopic mass for the full-length primer +4 product: calculated mass 4865.9918 Da; observed mass, 4865.9729 Da. Reaction conditions: 2.75 μM primer **P1**, 10 μM template **T3**, 200 mM HEPES pH 7.5, 5 mM 2-MeImpmC.

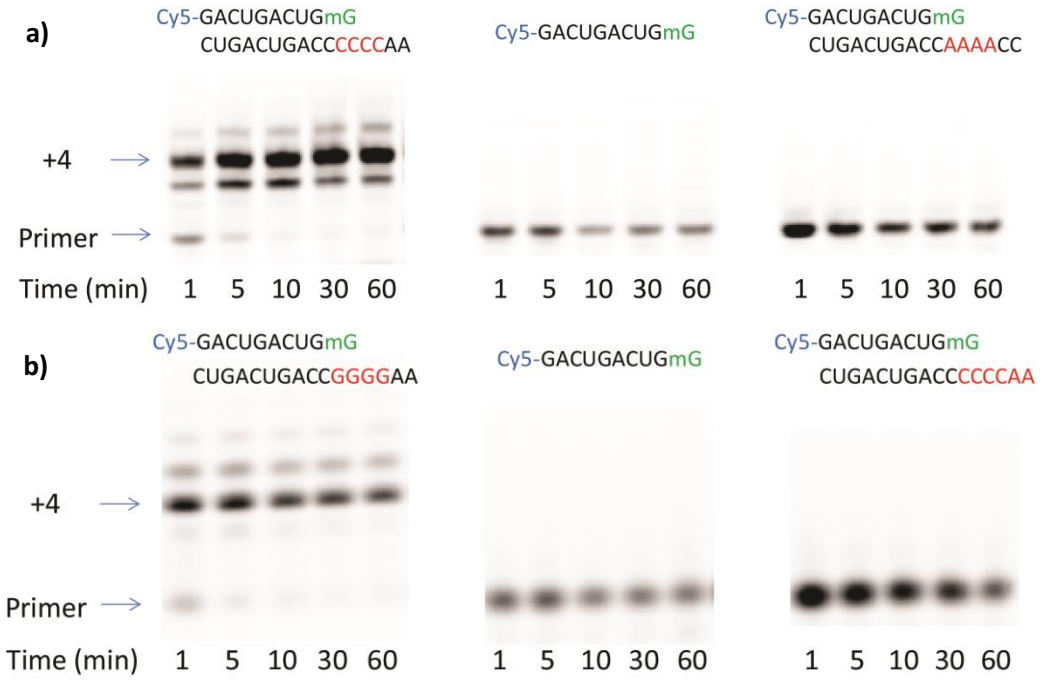


Figure A.5: Primer extension with activated morpholino monomer is template dependent. (a) High resolution gel electrophoresis analysis of primer extension reaction with 2-MeImpmG. Primer-extension reactions contained 1μM **P**, 5μM template (left to right, **T1, none, T9**), 200 mM HEPES pH 7.5, and 2.5 mM 2-MeImpmG. (b) High resolution gel electrophoresis analysis of primer extension reaction with 2-MeImpmC. Primer-extension reactions contained 1 μM **P1**, 5 μM template (left to right, **T3, none, T1**), 200 mM HEPES pH 7.5, and 2.5 mM 2-MeImpmC.

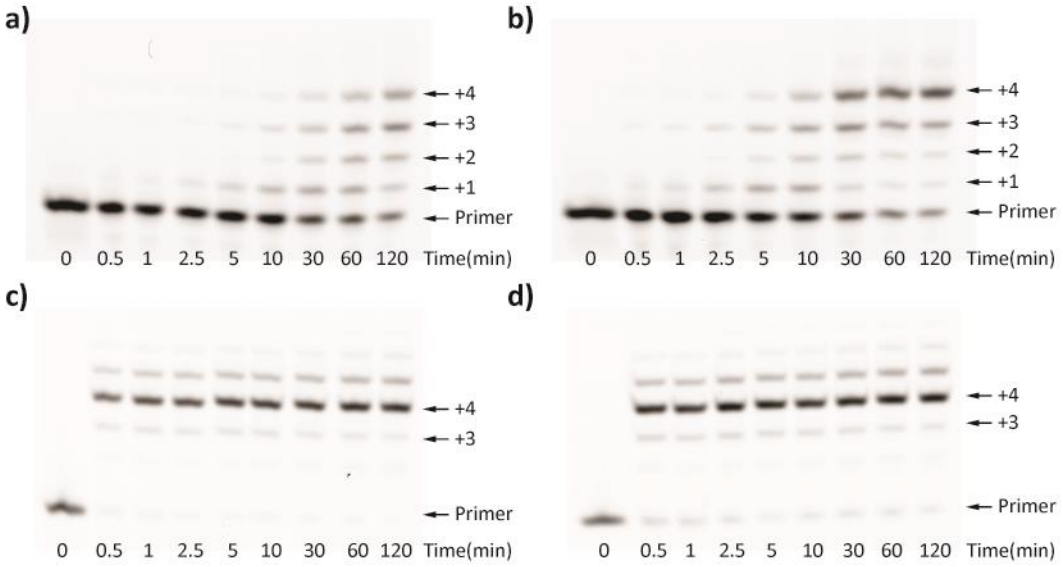


Figure A.6: Comparison of template copying reactions between 3′-NP-DNA system and MoNA system. (a) High resolution gel electrophoresis analysis of the extensions of 3′-amino modified primer **P3** (Cy3-AGUGAGUAACGC$_{NH2}$) on a G$_4$ RNA template **T4** in the presence of 3′-NH$_2$-2-MeImpddC. Primer-extension reactions contained 1 μM **P3**, 5 μM **T4**, 300 mM HEPES pH 7.5, and 5mM 3′-NH$_2$-2-MeImpddC. (b) Template copying with 3′-NH$_2$-2-MeImpddC in the presence of 100 mM HEI. (c) High resolution gel electrophoresis analysis of the extensions of 3′-morpholino modified primer **P4** (Cy3-AGUGAGUAACGmC) on a G$_4$ RNA template **T4** in the presence of 2-MeImpmC. Primer-extension reactions contained 1 μM **P4**, 5 μM **T4**, 300 mM HEPES pH 7.5, and 2-MeImpmC. (d) Template copying with 2-MeImpmC in the presence of 100 mM HEI.

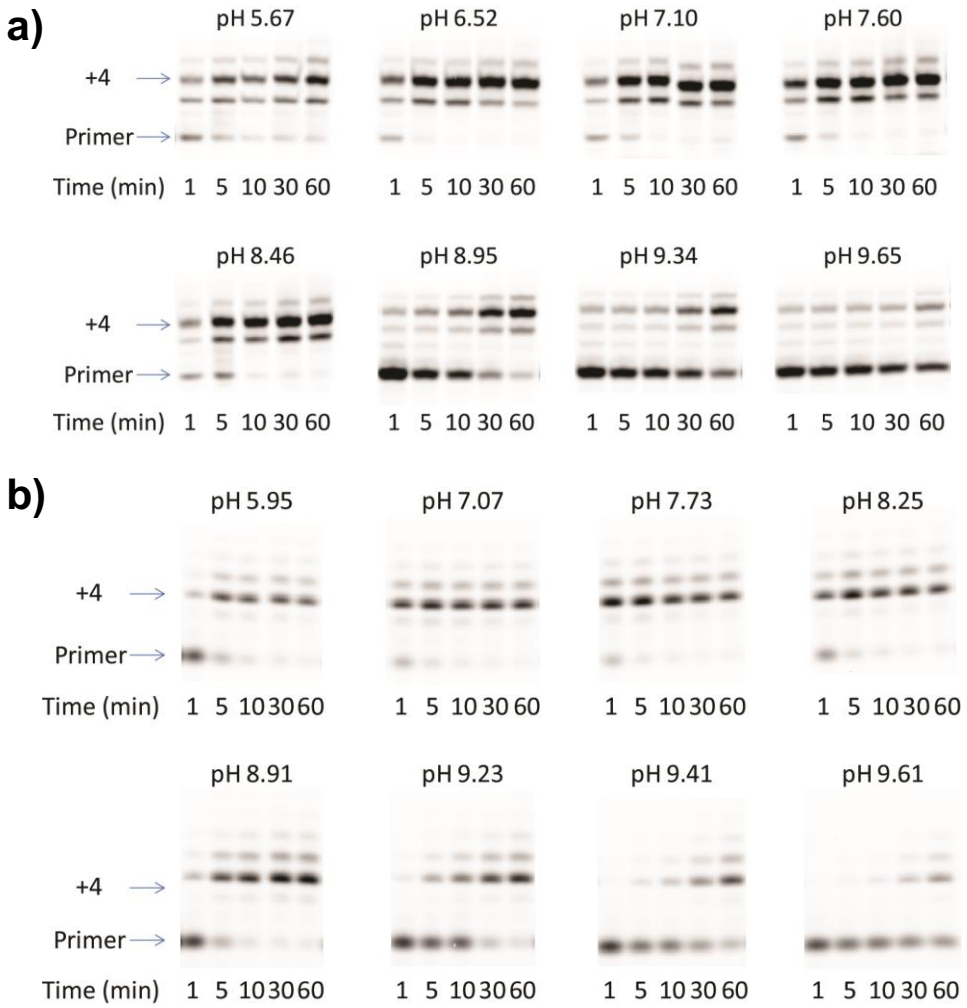


Figure A.7: pH dependence of template-dependent primer extension with morpholino monomers. (a) High resolution gel electrophoresis analysis of RNA C_4 template copying reactions with 2-MeImpmG at various pH values. Primer-extension reactions contained 1 μM **P1**, 5 μM **T1**, 200 mM buffer, and 2.5 mM 2-MeImpmG. (b) High resolution gel electrophoresis analysis of RNA G_4 template copying reactions with 2-MeImpmC at various pH values. Primer-extension reactions contained 1 μM **P1**, 5 μM **T3**, 200 mM buffer, and 2.5 mM 2-MeImpmC.

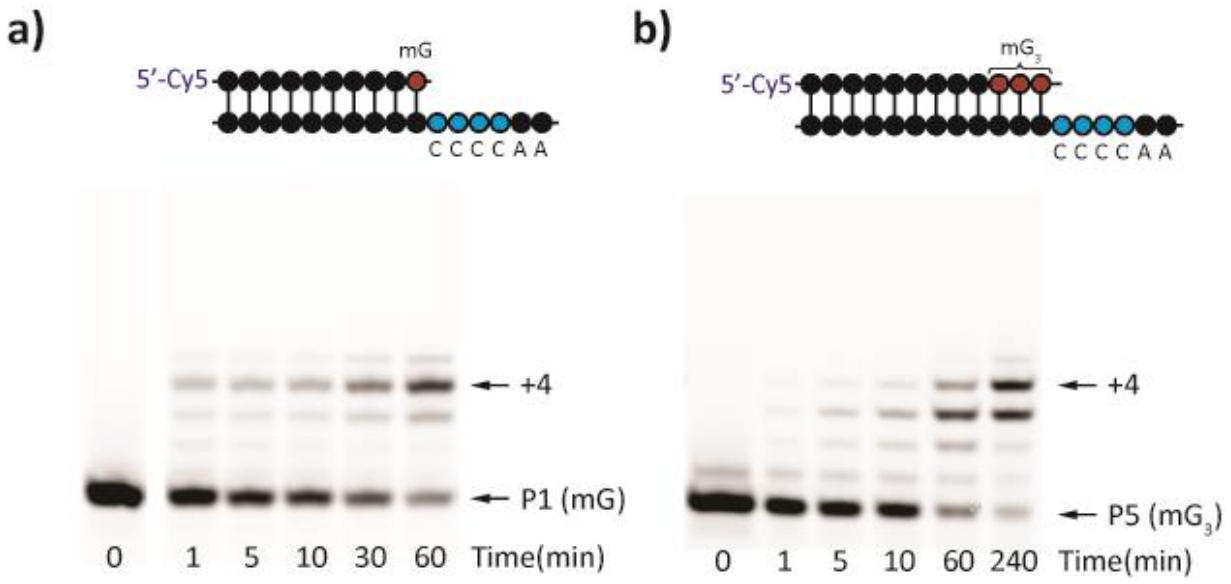


Figure A.8: Comparison between oligonucleotide with single morpholino modification and oligonucleotide with three terminal morpholino nucleotides in nonenzymatic primer extension reaction in the presence of 2-MeImpG. (a) Schematic representation of terminal modified primer **P1** (Cy5-GACUGACUGmG) annealed to a C_4 RNA template (**T1**, AACCCCCCAGUCAGUC). High resolution gel electrophoresis analysis of nonenzymatic primer extension reactions with 2-MeImpG on 3′-morpholino terminated oligonucleotide. Primer-extension reactions contained 1μM **P1**, 5μM **T1**, 150 mM CHES pH 9.5, and 5 mM 2-MeImpmG. (b) Schematic representation of oligonucleotide with three terminal morpholino nucleotides **P5** (Cy5-GACUGACUGmGmGmG) annealed to a C_4 RNA template (**T10**, AACCCCCCCCCAGUCAGUC). High resolution gel electrophoresis analysis of nonenzymatic primer extension reactions with 2-MeImpG on oligonucleotide with three terminal morpholino nucleotides **P5**. Primer-extension reactions contained 1μM **P5**, 5μM **T10**, 150 mM CHES pH 9.5, and 5 mM 2-MeImpmG.

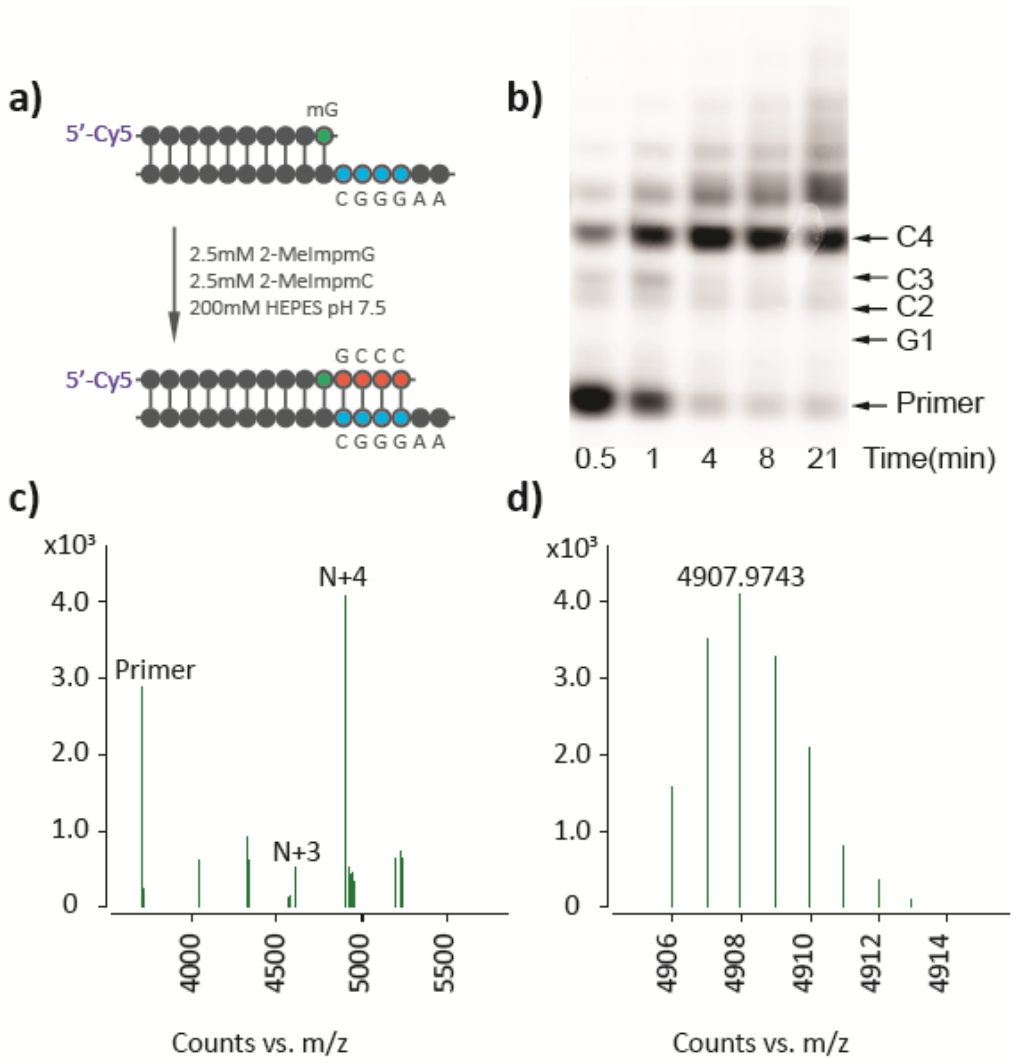


Figure A.9: Nonenzymatic primer-extension reactions on mix sequenced RNA template. (a) Schematic representation of an experimental set-up consisting of primer P1 annealed to a template (T6) in presence of activated MoNA monomers. 2-MeImpmG and 2-MeImpmC participate in the chemical extension reaction on the mixed sequenced templating region resulting in chimeric RNA/MoNA oligomers. (b) High-resolution PAGE analysis of the primer-extension products after different reaction time. Reaction conditions: 1μM primer P1, 5μM template T6, 200mM HEPES pH 7.5, 2.5mM 2-MeImpmG, 2.5mM 2-MeImpmC. (c) High-resolution MS anaylsis of the primer-extension products after 1 day. Reaction condition: 2.75μm primer P, 10μm template T6, 200mM HEPES pH 7.5, 3mM 2-MeImpmG, 3mM 2-MeImpmC. (d) Monoisotopic mass for the full-length primer +4 products: calculated mass, 4905.9979 Da; observed mass, 4905.9819 Da.

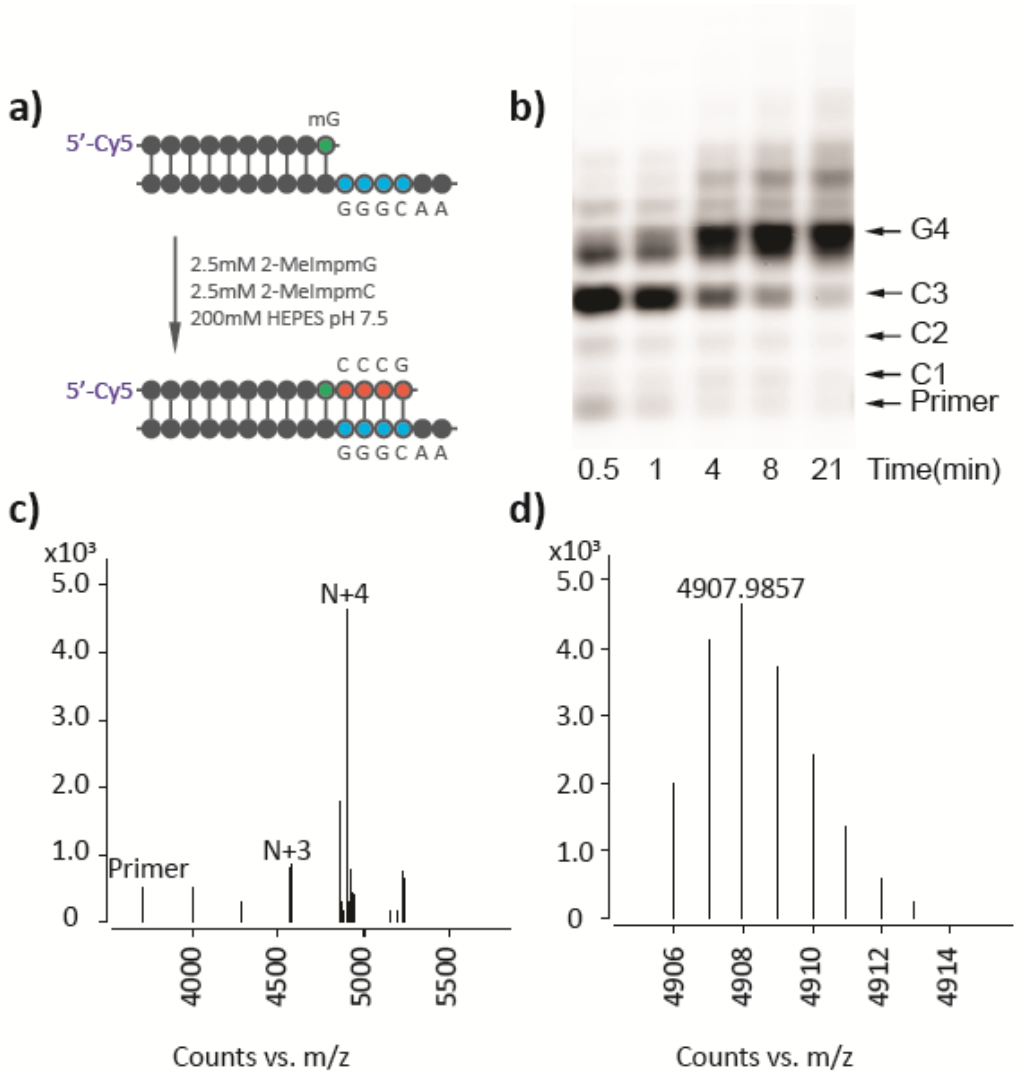


Figure A.10: Nonenzymatic primer-extension reactions on mix sequenced RNA template. (a) Schematic representation of an experimental set-up consisting of primer P1 annealed to a template (T7) in presence of activated MoNA monomers. 2-MeImpmG and 2-MeImpmC participate in the chemical extension reaction on the mixed sequenced templating region resulting in chimeric RNA/MoNA oligomers. (b) High-resolution PAGE analysis of the primer-extension products after different reaction time. Reaction conditions: 1μM primer P1, 5μM template T7, 200mM HEPES pH 7.5, 2.5mM 2-MeImpmG, 2.5mM 2-MeImpmC. (c) High-resolution MS anaylsis of the primer-extension products after 1 day. Reaction condition: 2.75μm primer P1, 10μm template T7, 200mM HEPES pH 7.5, 3mM 2-MeImpmG, 3mM 2-MeImpmC. (d) Monoisotopic mass for the full-length primer +4 products: calculated mass, 4905.9979 Da; observed mass, 4905.9858 Da.

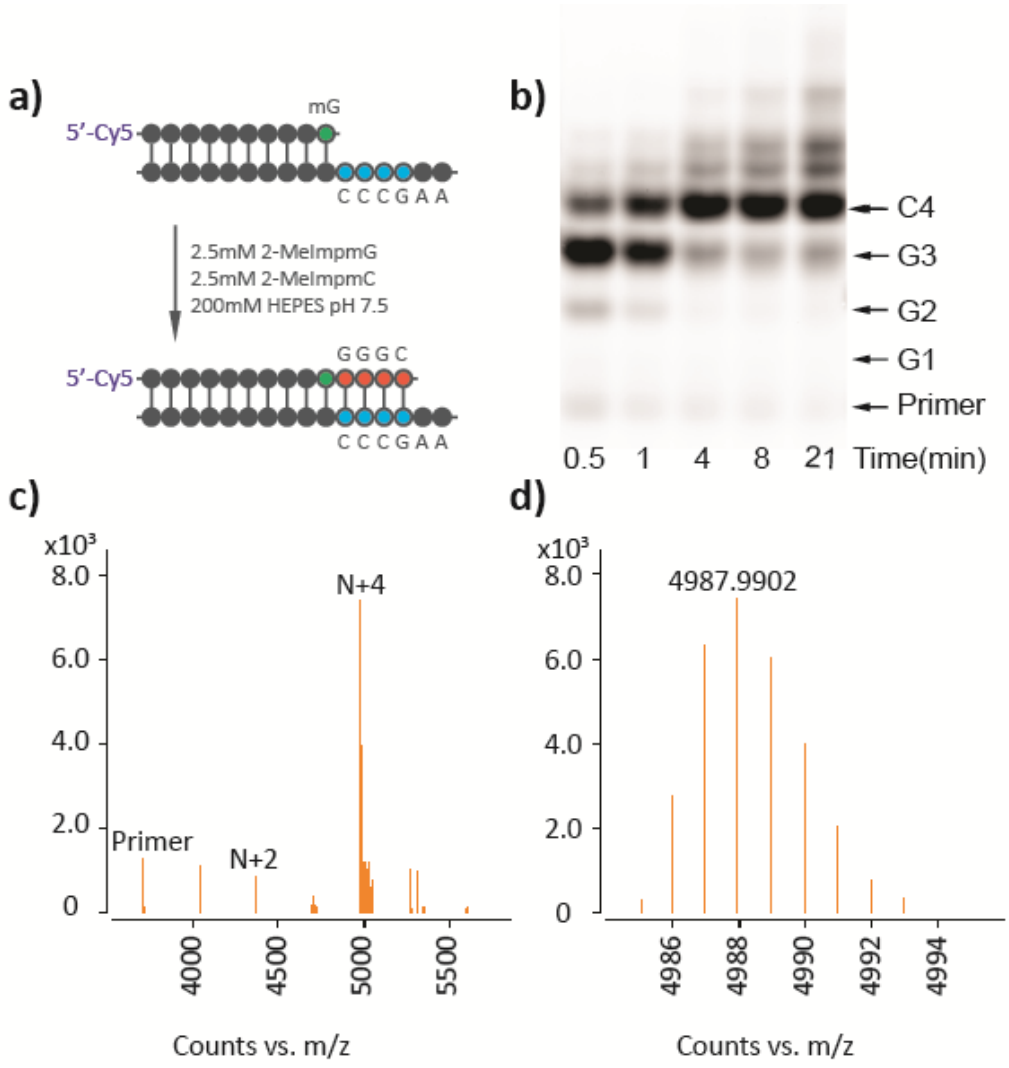


Figure A.11: Nonenzymatic primer-extension reactions on mix sequenced RNA template. (a) Schematic representation of an experimental set-up consisting of primer P1 annealed to a template (T8) in presence of activated MoNA monomers. 2-MeImpmG and 2-MeImpmC participate in the chemical extension reaction on the mixed sequenced templating region resulting in chimeric RNA/MoNA oligomers. (b) High-resolution PAGE analysis of the primer-extension products after different reaction time. Reaction conditions: 1μM primer P1, 5μM template T8, 200mM HEPES pH 7.5, 2.5mM 2-MeImpmG, 2.5mM 2-MeImpmC. (c) High-resolution MS anaylsis of the primer-extension products after 1 day. Reaction condition: 2.75μm primer P1, 10μm template T8, 200mM HEPES pH 7.5, 3mM 2-MeImpmG, 3mM 2-MeImpmC. (d) Monoisotopic mass for the full-length primer +4 products: calculated mass, 4985.9979 Da; observed mass, 4985.9974Da.

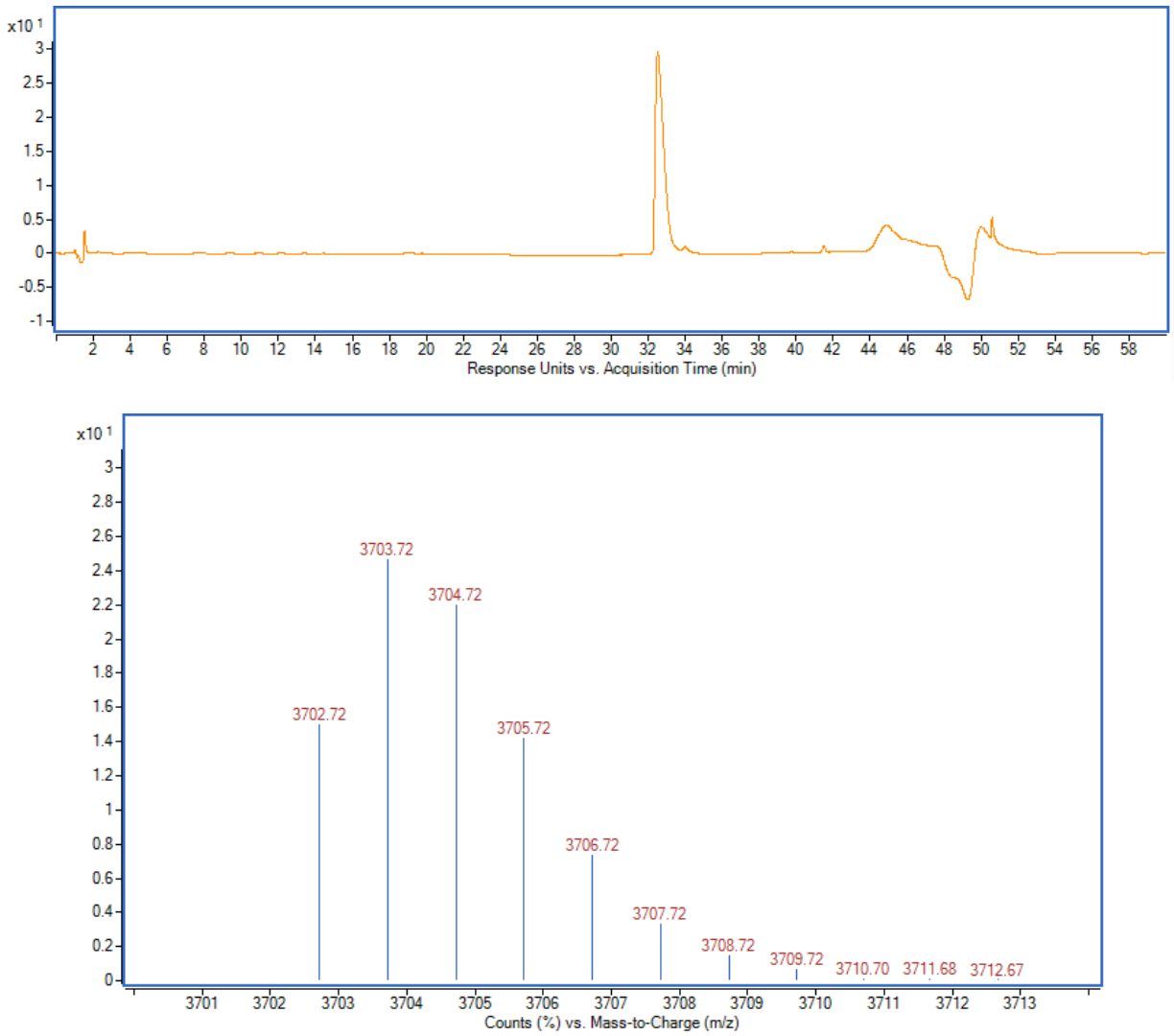
Appendix B

Supplemental Materials for Chapter 3

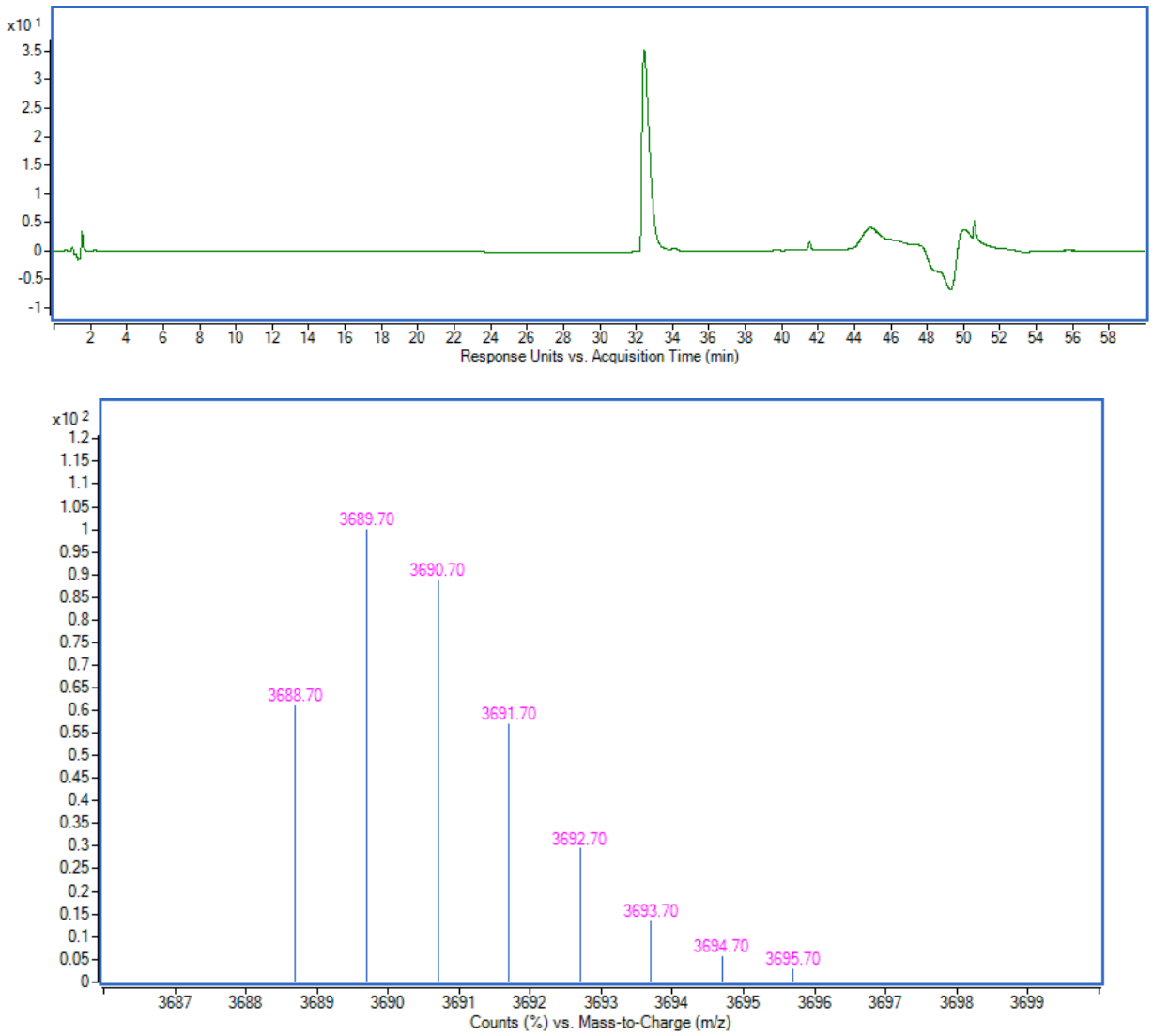
B.1 General information

Bulk solvents and chemicals were purchased either from Fisher Scientific (Clifton, NJ) or Sigma Aldrich (St. Louis, MO). Deuterated solvents were purchased from Cambridge Isotope Laboratories (Andover, MA). Water used for this study was obtained from a Barnstead Nanopure System (Thermo Scientific, Dubuque, IA). All reagents used were of highest purity available. The silica gel used for flash chromatography (40-63 μm) and gravitational column (63-200 μm) were purchased from Silicycle Inc (Québec City, Canada). NMR data were taken on a Varian 500 or 600 MHz spectrometer and analyzed using MestReNova software, referenced against residual solvent peak. Liquid chromatography followed by high-resolution mass spectroscopy (LC-HRMS) analysis in the ESI mode was carried out on an Agilent 6520 Q-TOF LC/MS-MS instrument. *β*-tributylstannylstyrene was prepared using a method published by Brooke et al.[1] N^2-Isobutyryl-2'-deoxyguanosine was synthesized according to a published procedure.[2]

B.2 LC-HRMS Analysis of the Oligomers



Fiugre B.1: LC and HRMS data for 5′-fluorophore tagged dG*-terminated RNA primer (P1, 5′-Cy3-GACUGACUG dG*-3′). Exact mass (Neg. mode): 3702.7264; measured accurate mass: 3702.7182.



Fiugre B.2: LC and HRMS data for 5′-fluorophore tagged dG-terminated RNA primer (P2, 5′-Cy3-GACUGACUG dG-3′). Exact mass (Neg. mode): 3688.7107; measured accurate mass: 3688.6978.

B.3 Gel Electrophoresis Results

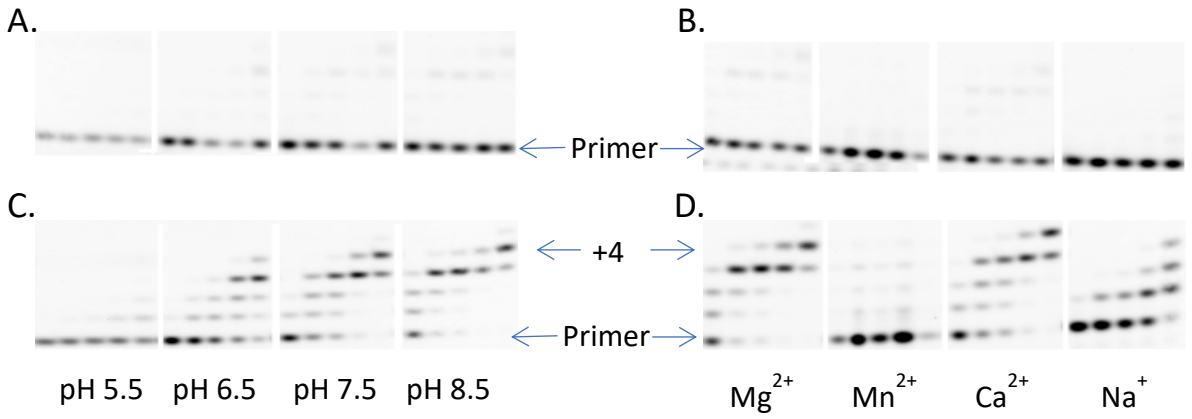
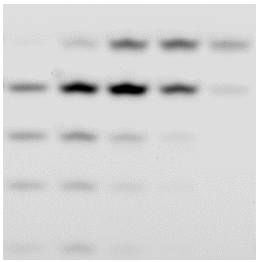


Figure B.3: Gel electrophoresis results of the primer extension reactions. Reaction conditions:1 μM primer (P1 in A and B; P2 in C and D), 5 μM template T, 200 mM Buffer, 100 mM metal salt, 50 mM 2-MeImpG. Time points: 5 min, 30 min, 1 hour, 4 hours and 1 day. A & C: pH screen with MgCl2 as metal salt; B & D: Metal salt screen with HEPES (pH 8.5) as buffer.

In figure B.3, the very faint extension bands observed for P1 in **A** and **B** can be due to: a) Insignificant extension of P1, and/or; b) Extension of an impurity, RNA 9-mer P3 (5′-Cy3-GACUGACUG-3′)

In the presence of Mg^{2+}, Mn^{2+}, or Ca^{2+}, P2 extends rapidly and starts to provide some +3 extension product in less than 5 minutes. P2 can also be extended without the catalytic effect of divalent cations. For example, some +3 product is observed after 4 hours in the presence of monovalent Na^{+}.



In figure B.3D, we did not observe any extension in the presence of Mn^{2+}. It is due to the precipitation observed at pH 8.5. We found Mn^{2+} act as an effective catalyst at pH 7.5 on the right. The time points are 5 min, 10 min, 30 min, 1 hour and 2 hour.

B.4 References

1. Brooke, D. G.; Morris, J. C., Total synthesis of hydroxystrobilurin A via Stille coupling. *Tetrahedron Letters* **2008,** *49* (15), 2414-2417.

2. Ti, G. S.; Gaffney, B. L.; Jones, R. A., Transient protection: efficient one-flask syntheses of protected deoxynucleosides. *J. Am. Chem. Soc.* **1982,** *104* (5), 1316-1319.

Appendix C

Supplemental Materials for Chapter 4

C.1 General Information

All solvents and reagents were reagent grade, purchased commercially, and used without further purification unless specified. Oligonucleotides used in the experiments were synthesized on an Expedite nucleic acid synthesizer from Applied BioSystems (Foster City, CA). The 2´-TBDMS-protected RNA phosphoramidites were obtained from ChemeGenes (Wilmington, MA). Reagents for oligo synthesis were obtained from Glen Research (Sterling, VA). Reverse-phase HPLC was performed on Agilent 1100 HPLC system from Agilent Technologies (Santa Clara, CA). All chemicals were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise indicated. Guanosine 5′-monophosphate were purchased as the free acids from Santa Cruz Biotechnology (Dallas, TX). Reverse phase flash chromatography was performed using a RediSep Rf Gold C18Aq 50 g column purchased from Teledyne Isco (Lincoln, NE). Isothermal titration calorimetry (ITC) experiments was performed using Nano ITC from TA Instruments (New Castle, DE). Thermal melting curves were collected using an Agilent Cary 60 UV-Vis spectrophotometer from Agilent Technologies (Santa Clara, CA).

C.2 Protocol for RNA Oligonucleotides Synthesis

All RNA oligonucleotides were chemically synthesized at 1.0-μmol scales by solid phase synthesis on an Expedite 8909 nucleic acid synthesis system. The 2´-TBDMS-protected RNA phosphoramidites were dissolved in acetonitrile to a concentration of 0.067 M. Synthesis was performed on the appropriate nucleoside immobilized via a succinate linker to controlled pore glass (CPG-1000). All oligonucleotides were prepared using a 5'-DMTr-off protocol. After the synthesis, RNAs were cleaved from the solid support and deprotected with 28-30% NH_4OH:40% methylamine (1:1 v/v) at 65 °C for 10 minutes. The solvent was completely removed using a Speed-Vac concentrator. The dried material was redissolved in 100 μL DMSO and treated with 125 μL of Et_3N•3HF at 65 °C for 2.5 h. The solution was cooled to room temperature and the RNA was precipitated by the addition of 25 μL 3M sodium acetate pH 5.5 and 1 mL 1-butanol. The solution was cooled to on dry ice for 30 minutes and the centrifuged for 30 minutes at 8000 g and 4 °C. The RNA was purified by reverse-phase HPLC on a ZORBAX Eclipse Plus C18 column 4.6 mm × 250 mm, 5 μ particle size (Agilent Technologies) equilibrated with 20 mM triethylammonium bicarbonate/2% acetonitrile, pH 7.5, and eluted with an acetonitrile gradient from 2% to 12% in 20 minutes. Purified nucleotides were lyophilized to dryness and stored at -80 °C.

C.3 Improved Procedures for 2-MeImpG Synthesis

Guanosine 5′-phosphor-2-methylimidazolide (2-MeImpG) was prepared by the following improved protocol. Guanosine 5′-monophosphate free acid (3 mmoles) was

fully dissolved in an aqueous solution of 2-methylimidazole (30 mmoles, 0.5 M) pH 5.5 and the mixture was lyophilized. The mixture of dried substrates was dissolved in 180 mL of dry DMSO, to which triphenylphosphine (30 mmoles), triethylamine (30 mmoles) and 2,2-dipyridyl disulfide (30 mmoles) was added. The reaction mixture was stirred for 3 hours at room temperature and monitored by ESI-mass spectroscopy. Upon completion, the reaction mixture was added dropwise to a precipitation solution containing 600 mL of acetone, 375 mL of anhydrous ether, 45 mL of triethylamine and 6 ml of saturated $NaClO_4$ in acetone. Continuous and vigorous stirring was maintained during the procedure and for an additional 15 min. The precipitate was collected by centrifugation. Crude product was purified by C18Aq Column with 20mM TEAB buffer and Acetonitrile. Desired fraction was immediately added to 20 fold volumn of a cold mixture of 25:25:1 acetone:diethyl ether:saturated $NaClO_4$ in acetone. The precipitate was collected by centrifugation and washed twice with 1:1 acetone:diethyl ether. Product was then dried on lyophilizer for 1 hour and stored at -80 °C.

C.4 Protocol for Isothermal Titration Calorimetry Experiments

All ITC experiments were performed using a TA Instruments Nano ITC. Samples were prepared by diluting stock solutions of HPLC-purified nucleotides into buffers containing 200 mM Tris-HCl, pH 7.5 at various metal ion concentrations. Nucleotide concentrations were verified by UV absorbance at 260 nm using extinction coefficients predicted by primary sequence. Nucleotide solutions of 50 μM were titrated into the sample cell containing 300 μLs of the complementary nucleotide at 7.5 μM. Samples

were stirred at 350 rpm and 25 °C with 220 seconds between injections of 1.5 μL. The values of K_d were measured by fitting the data to a binding isotherm using an independent fit model in the NanoAnalyze software from TA instruments.[1] ΔH, K_d, and stoichiometry (*n)* were fit simultaneously, and for each experiment, *n* was determined to be equal to 1 within the error of the fit. ΔG was calculated using the relationship $\Delta G = -RT \ln(K_d)$. Experiments were performed in triplicate and standard errors are reported as standard deviations. Control experiments with the inverted direction of titration revealed no significant differences.

C.5 Protocol for Optical Melting Experiments

Samples were prepared by diluting stock solutions of HPLC-purified oligonucleotides into buffers containing 200 mM Tris-HCl pH 7.5 and various metal ion concentrations. Oligonucleotide concentrations were verified by UV absorbance at 260 nm using extinction coefficients predicted by primary sequence. All samples were heated to 98 °C for 3 min and subsequently cooled to room temperature for 1 h before data was collected. Experimental RNA concentrations ranged from 200–6.75 μM and thermal melting curves were collected using an Agilent Cary 60 UV-Vis. Absorbance was recorded at 260 nm as the temperature was ramped between 4 and 89 °C at a rate of 1 °C min^{-1}. All data was collected in duplicate. Values of T_M were obtained by directly fitting the resulting melting curves to a double baseline, two state model.[2] These values were used to construct plots of T_M^{-1} versus $\ln(C_T/4)$, where C_T is the total RNA concentration. A linear least-squares fit was applied to the resulting van 't Hoff plots and

the Δ*H* and Δ*S* values were calculated from the slopes and intercepts according to the relationship:

$$\frac{1}{T_M} = \frac{R}{\Delta H} ln(C_T/4) + \frac{\Delta S}{\Delta H}$$

Values of Δ*G* were calculated according to the relationship: Δ*G* = Δ*H* - Δ*ST*, where *T* is temperature in K. Errors on the reported values correspond to half the difference between replicates.

C.6 General Protocol for PAGE Analysis of Nonenzymatic Primer Extension

Representative reaction protocol: 2.0 μL of 1.0 M Tris-HCl pH 7.5 buffer, 2.5 μL of nuclease-free water, 1 μL of 10 μM primer (5'-Cy5-GACUGACUGG), 1.0 μL of 50 μM template (5'-AACCCCCCAGUCAGUC), and 1 μL of 1 M $MgCl_2$ were combined in a thin-walled PCR microtube and mixed well. To initiate primer extension, 2.5 μL of 200 mM 2-MeImpG were added to the lid of the microtube cap. The reaction was initiated when the nucleotides were spun down into the buffered primer/template solution. Aliquots (1.5 μL each) were removed at 1, 2.5, 5, 7.5, and 10 min and were immediately quenched by addition to 43.5 μL of gel loading buffer [1X TBE (89 mM Tris, 89 mM boric acid, and 2 mM EDTA pH 8.0) in 8.0 M urea]. 5.0 μL of sample with primer extension products were separated by 20% (19:1) denaturing PAGE. The gel was scanned with a Typhoon 9410 Variable Mode Imager, and the bands were quantified using the accompanying ImageQuant TL software package.

C.7 Protocol for HPLC Analysis of Metal Catalyzed Hydrolysis of 2-MeImpG

2-MeImpG, purified by reverse phase liquid chromatography, was mixed with Tris-HCl buffer (200mM, pH 7.5) and various metal ions at different concentrations. 0.5 µL aliquots were removed at intervals of 120 minutes and were immediately separated by reverse-phase HPLC on a ZORBAX Eclipse Plus C18 column 4.6 mm × 250 mm, 5 µ particle size (Agilent Technologies) equilibrated with 20 mM triethylammonium bicarbonate/2% acetonitrile, pH 7.5, and eluted with an acetonitrile gradient. Peak area of remaining 2-MeImpG and hydrolyzed GMP were measured by UV-absorbance at 254 nm.

C.8 Supporting Figures

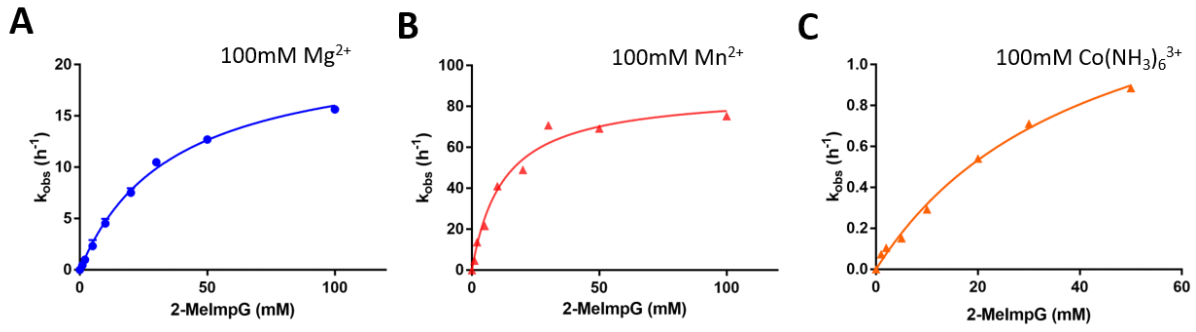


Figure C.1: Kinetic parameters of non-enzymatic primer extension with different metal ions. Plots of observed rate vs conc. of activated monomer (2-MeImpG). The primer-extension reaction mixtures contained 1 µM Cy5-labeled RNA primer, 5µM template, 200 mM HEPES (pH 7.5), 100 mM metal chloride and 0 – 100 mM 2-MeImpG monomer. (A) 100 mM Mg^{2+} (B) 100 mM Mn^{2+} (C) 100 mM CoHex.

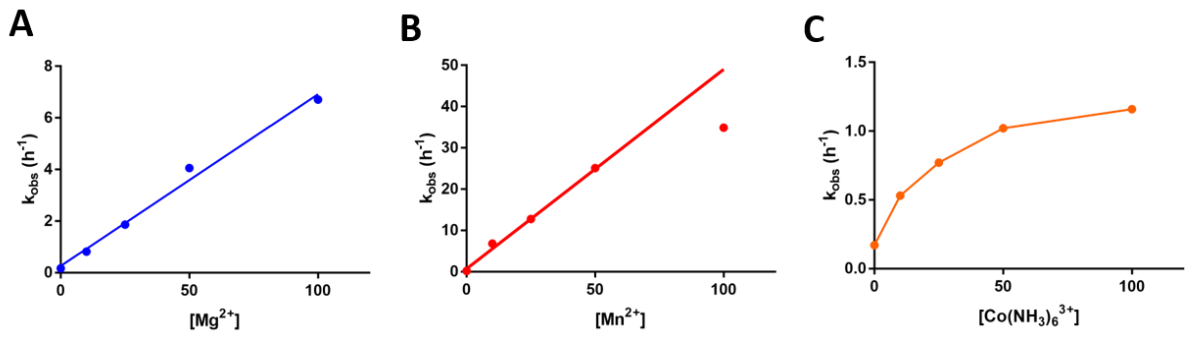


Figure C.2: Plots of observed rate vs conc. of metal ions. The primer-extension reaction mixtures contained 1 µM Cy5-labeled RNA primer, 5µM template, 200 mM Tris-HCl (pH 7.5), 0-100 mM metal chloride and 50 mM 2-MeImpG monomer. (A) 100 mM Mg^{2+} (B) 100 mM Mn^{2+} (C) 100 mM CoHex.

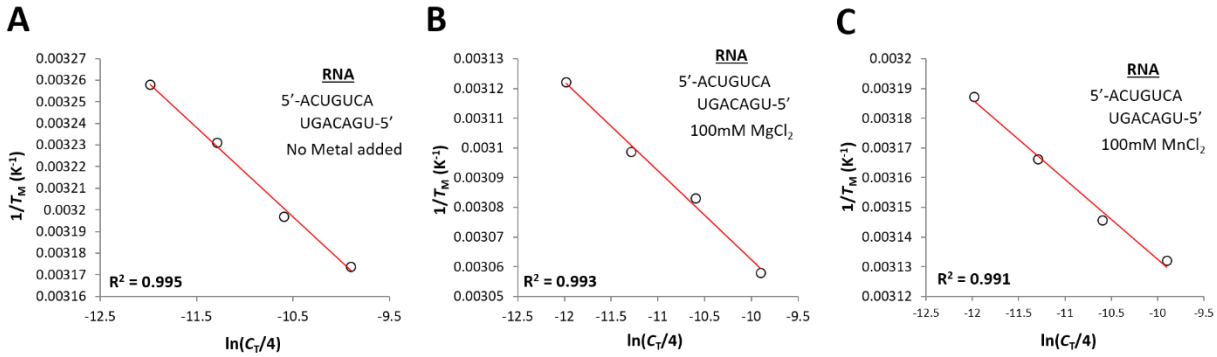


Figure C.3: van 't Hoff plots. Linear least-squared fits (red line) of a van 't Hoff plots of inverse melting temperatures (T_M^{-1}) collected from optical melts at different total oligonucleotide concentrations (C_T) of RNA 7-mers with 100 mM indicated metal ions and 200 mM HEPES buffer, pH 7.5. The R^2 values of each plot exceed 0.99.

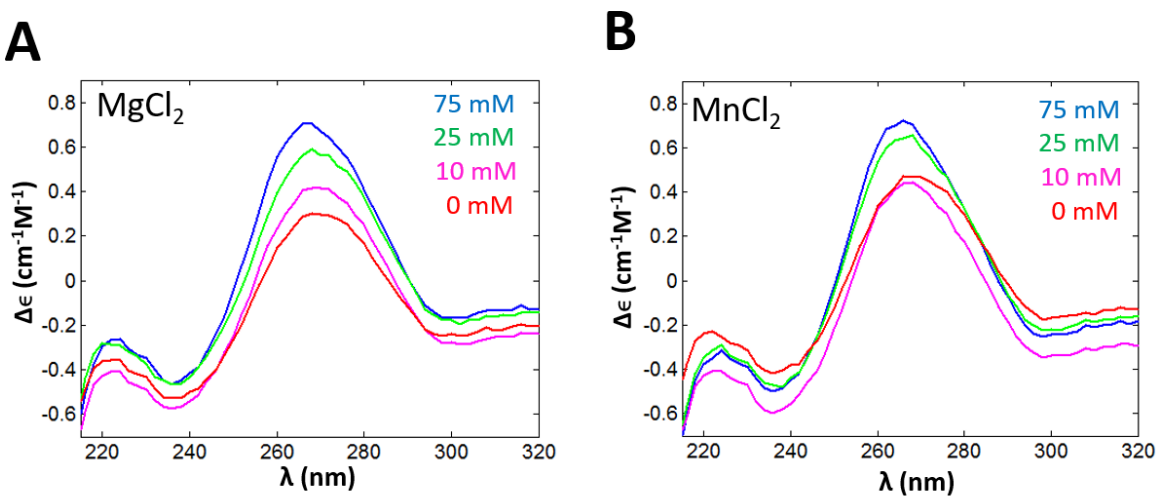


Figure C.4: Circular dichroism Studies of dsRNA with increasing concentrations of Mg^{2+} or Mn^{2+}. Overlapped spectra obtained by UV-CD spectroscopy at 50°C with 100 μM total oligonucleotide in 200 mM HEPES at pH 7.5.

C.9 Supporting Table

Table C.1: **Kinetics of the Hydrolysis of 2-MeImpG**

Entry	Metal Ions I	Metal Ions II	Rate (x 10^{-3} h^{-1})
1	-	-	8.9 ± 1.6
2	100 mM Mg^{2+}	-	11 ± 1
3	10 mM Mn^{2+}	-	10 ± 1
4	50 mM Mn^{2+}	-	14 ± 2
5	100 mM Mn^{2+}	-	34 ± 2
6	100 mM CoHex	-	8.4 ± 1.6
7	50 mM Mg^{2+}	50 mM CoHex	9.1 ± 1.5
8	10 mM Mn^{2+}	50 mM CoHex	10 ± 1
9	50 mM Mn^{2+}	50 mM CoHex	13 ± 2

Hydrolysis experiments were performed as described in the Materials and Methods. Rates were measured by reverse-phase HPLC. Values are averages and errors are standard deviations calculated from triplicate reactions.

C.10 References

1. Keeler, C.; Poon, G.; Kuo, I. Y.; Ehrlich, B. E.; Hodsdon, M. E., An explicit formulation approach for the analysis of calcium binding to EF-hand proteins using isothermal titration calorimetry. *Biophys J* **2013,** *105* (12), 2843-53.

2. Schroeder, S. J.; Turner, D. H., Optical melting measurements of nucleic acid thermodynamics. *Methods Enzymol* **2009,** *468*, 371-87.