



Investigating the Mechanism of Non-Enzymatic RNA Replication

Citation

Giurgiu, Constantin. 2021. Investigating the Mechanism of Non-Enzymatic RNA Replication. Doctoral dissertation, Harvard University Graduate School of Arts and Sciences.

Permanent link

<https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37368364>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

HARVARD UNIVERSITY
Graduate School of Arts and Sciences



DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the
Department of Chemistry & Chemical Biology
have examined a dissertation entitled:

Investigating the Mechanism of Non-Enzymatic RNA synthesis

presented by: Constantin Giurgiu

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

Signature

Typed name: Professor Jack W. Szostak

Signature

Typed name: Professor George Whitesides

Signature

Typed name: Professor Christina Woo

Date: 30 April 2021

Investigating the Mechanism of Non-Enzymatic RNA Replication

A dissertation presented by

Constantin Giurgiu

to

The Graduate School of Arts and Sciences

In partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Chemistry

Harvard University

Cambridge, Massachusetts

April 2021

© 2021 Constantin Giurgiu

All rights reserved.

Investigating the Mechanism of Non-Enzymatic RNA replication

Abstract

Before life became confoundingly complex it must have been simpler. In modern biology, polymerase enzymes leverage the molecular recognition of nucleic acids through Watson-Crick base pairs to efficiently synthesize genomic material. Templated RNA synthesis could have supported early life by enabling genomic replication in the absence of enzymatic catalysts. Despite encouraging early results, the model still has severe shortcomings: only some sequences are copied efficiently, multiple regioisomers are formed and RNA degradation is as fast as its synthesis. In my thesis work I show that 2-aminoimidazole, the leaving group used in non-enzymatic RNA copying, can be obtained prebiotically in similar conditions to those in which RNA is synthesized, suggesting a common origin for RNA synthesis and activation. 2-Aminoimidazole activated ribonucleotides enable faster RNA synthesis through the formation of an imidazolium bridged intermediate. I then show that the poor regioselectivity of RNA synthesis is overstated: fast reactions, in which the reactive intermediate can bind through Watson-Crick base-pairs to the template produce only the correct regioisomer. To further optimize the reaction, I set out to determine its mechanism. To this end, I developed a high-throughput assay for templated RNA synthesis and investigated the role of nucleic acid structure and of catalytic metal ions. I found that RNA is particularly adept at templated synthesis because of its preference for a specific conformation and that Mg ions catalyze RNA synthesis by forming a reactive anion. Lastly, RNA degradation is mitigated by templated synthesis, giving rise to a novel biopolymer: pyrophosphate linked RNA. My work adds additional support to the RNA world hypothesis while also revealing fundamental aspects of nucleic acid chemistry and developing the synthesis of alternative biopolymers. These findings point towards a mechanism of chemical selection that explains why RNA, and not its analogs, is present four billion years later.

Acknowledgements

I would first like to thank my family for being so incredibly supportive over the years and encouraging me to always strive for better. My parents, Luci and Costica, as well as my sister Monica, have been indispensable for me getting to this stage in my education.

Mama, Tata si Moni, va multumesc din suflet pentru cat m-ati ajutat, si sper ca sunteti mandri si multumiti de mine.

My advisor Jack has been wonderful in the six years I have spent in laboratory and I want to thank him for facilitating and supporting all the great science I have been part of. Our lab manager Fanny has been indispensable, and I do not think I would have been nearly half as productive as I have been without her consistent help. I want to thank all my collaborators, especially Li, Albert, Derek and Lijun for making the lab such a wonderful place to work in.

Lastly, I want to acknowledge all my friends for putting up with my incessant ramblings about the origin of life, and always being there when I needed anything. Special thanks to Amy, Pavel, Bernardo, Linus, Dio M., Jonathan, Brent, Emily and Dio D. for making my time in the US so pleasurable.

Contents

Title Page.....	I
Copyright.....	II
Abstract	III
Acknowledgements	IV
1. The Recent Past and Near Future of Non-Enzymatic RNA Replication	1
2. A common and potentially prebiotic origin for precursors of nucleotide synthesis and activation.....	26
3. A Mechanistic Explanation for the Regioselectivity of Non-Enzymatic RNA Primer Extension	35
4. A Fluorescent G-quadruplex Sensor for Chemical RNA Copying.....	49
5. Prebiotically Plausible ‘Patching’ of RNA Backbone Cleavage Through a 3’-5’ Pyrophosphate Linkage	58
6. Structure-activity relationships in nonenzymatic template-directed RNA synthesis.....	78
7. References	93

1. The Recent Past and Near Future of Non-Enzymatic RNA Replication

1.1 Introduction

The origin of life on Earth is one of the most important problems chemists are tasked with solving. The RNA world model assumes that modern biochemistry was predated by an earlier stage in which cells relied on RNA to store genetic information and carry out cellular processes. Prior to the evolution of ribozymes or enzymes, these first cells would have had to rely on non-enzymatic RNA replication to grow, replicate and evolve in a Darwinian sense. Cycles of non-enzymatic RNA replication would have eventually generated a ribozyme that provided a selective advantage to the protocells encoding it. At this point, the evolution of a replicase ribozyme would allow more efficient and accurate replication of ribozymes. This would generate a feedback loop where a replicase would sustain the replication and evolution of more ribozymes, which would in turn drive the evolution of more sophisticated replication machinery.

Despite 60 years of research, an experimental demonstration of continuous cycles of non-enzymatic RNA replication has yet to be achieved. One of us previously summarized the topic nine years ago and outlined eight major blockades for non-enzymatic RNA replication. Three of these hurdles stem from the inefficiency of the template copying chemistry; the template-directed synthesis of RNA proceeds with poor fidelity and regioselectivity and the rate of copying is not fast enough. Additional issues arise when considering the prerequisites for a second round of copying. The high melting temperature of the RNA duplexes generated by the copying of a template traps the two strands generating dead ends. Even provided that the two strands can be separated, they reanneal too quickly to allow the further rounds of copying. Additionally, multiple

rounds of copying require a constant supply of activated ribonucleotide species. Without a suitable activation chemistry that is compatible with the template-copying step, the activated species will hydrolyze and generate inhibitors. Finally, two major hurdles arise from the limitations of the protocellular model of RNA replication. Although divalent metal cations are required for the efficient copying of templates, they destroy the lipid envelopes of the protocells. Furthermore, these lipid envelopes are impermeable to oligonucleotides longer than trimers, which have been extensively used in experimental models of RNA copying. In this review we will survey recent progress in developing plausible prebiotic chemistry that brings us closer to non-enzymatic RNA replication, along with a perspective on future developments that could finally explain how RNA replication occurred during the origin of life.

1.2 How does RNA replication operate?

Experimental approaches to RNA replication have used the primer extension model. A complete cycle of replication involves a first copying event in which a double stranded RNA duplex is formed. This duplex then separates, and new primers anneal to the two strands, which will both serve as templates in subsequent copying steps. The copying step formally employs three different components: the primer, which contains the nucleophilic 3' moiety, the template, and the incoming electrophilic species (Figure 1.1). Pioneering research from

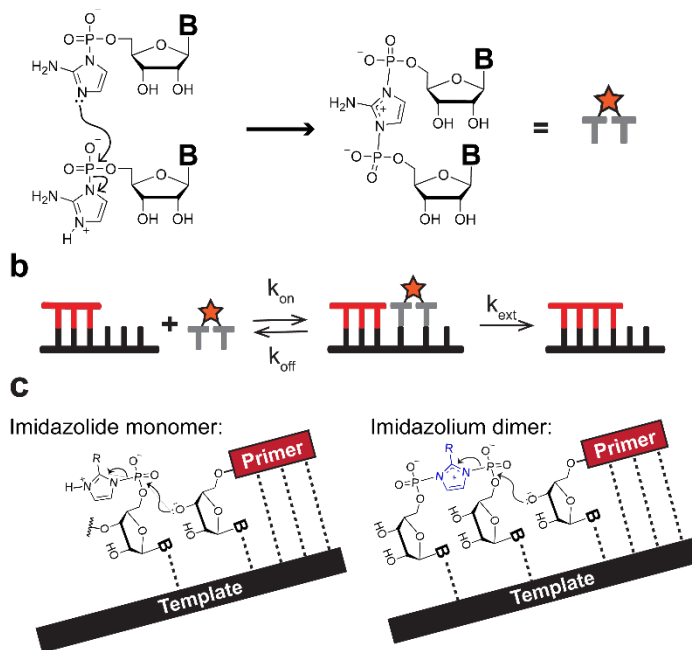


Figure 1.1. The primer-extension model for non-enzymatic RNA replication. (A) The formation of an imidazolium bridged intermediate from 5'-phosphoroimidazoles. (B) Two-step model for the copying of the template through an imidazolium bridged intermediate. (C) A side-by-side view of the classical monomer pathway and the imidazolium bridged dimer mediated primer extension.

Leslie Orgel's lab introduced 5'-phosphoroimidazolides as the reactive electrophilic species. These species were proposed to bind to the template through Watson-Crick base-pairing and react with the 3'-OH group of the terminal nucleotide on the primer via an in-line nucleophilic displacement reaction.

Early reports showed that the rate of ligation is considerably lower than the rate of polymerization. Considering that oligonucleotides bind stronger to the template than mononucleotides, it did not make sense that the rate of extension using a trimer or tetramer would be lower than the rate of polymerization of activated monomers. Additionally, Orgel's lab discovered that in the case of polymerization, the copying of the last nucleotide of the template was much slower than expected. These two observations suggested that our understanding of the reaction mechanism was incomplete. Kinetic studies from the Richert group suggested that the existence of a dimeric, imidazolium bridged reactive intermediate and invoked it as an explanation for the burst kinetics observed in the primer extension reaction¹. Our lab then characterized this intermediate and showed that it is formed from the reaction of two 5'-phosphoroimidazolides². We further showed that the contribution to the rate of the reaction by monomers is negligible compared to the dimeric intermediate². The template copying reaction can then be understood by using a two-step model which involves a reversible binding step of the reactive species to the pre-formed template-primer duplex, followed by the nucleophilic displacement of the leaving group.

The unusual structure and connectivity of the imidazolium bridged dinucleotide raised the question whether it can bind to the template in a productive fashion. Our lab first obtained a crystal structure of a primer-template duplex with a non-hydrolyzable, isosteric analog of the dinucleotide where the imidazolium bridged is replaced by a phosphate group³. This crystal structure demonstrated that a 5'-5' connected dinucleotide is capable of binding to the template via Watson-Crick base-pairs, and the primer 3'-OH is aligned to attack the phosphate atom closest to the primer. Although this structure suggests that the bridged dinucleotide analog is organized

upon binding to react with the primer-template duplex via a nucleophilic displacement reaction, it does not confirm the imidazolium bridged intermediate as an intermediate in the primer-extension reaction. In a follow-up study using time resolved X-ray crystallography, the same primer-template construct used previously was crystallized and soaked with activated mononucleotides⁴. Upon diffusing into the crystal, the activated mononucleotides bind to the template and react to form the imidazolium bridged dinucleotide. The dinucleotide then reacts with the 3'-terminus of the primer, and the primer is extended by one nucleotide, unambiguously proving the involvement of the imidazolium bridged dinucleotide in the primer extension reaction.

The intermediate needs to be stable enough to accumulate in appreciable quantities, but also reactive enough that when it is bound it extends quickly. Our lab discovered that if 2-aminoimidazole is used as the leaving group the imidazolium bridged intermediate is stabilized by the electron-donating exocyclic amine group. This extra stability increases the rate of the reaction considerably when compared to the previously used leaving groups⁵.

1.3 Main Body

1.3.1 Fidelity

Mutations introduced in the genetic material of organisms allow them to evolve novel features and adapt better to their environment. Evolution relies on a specific error threshold⁶. If the mutational rate is too high, newly acquired beneficial functions cannot be transmitted to descendants. Polymerase enzymes display an exquisite control of fidelity through multiple mechanisms. The active site architecture favors binding of the correct nucleotide in a Watson-Crick fashion, reducing the binding affinities for incorrect modes. If an incorrect nucleotide binds in the active site of the enzyme, the copying chemistry is considerably slowed down giving the nucleotide time to dissociate. The copying is either faster or slower than the association-

dissociation process. In contrast, for the uncatalyzed process polymerization is always slower than dissociation, regardless of orientation. The only catalyst available for the non-enzymatic version is the template. The template will favor the incorporation of the correct nucleotide to a certain extent. However, non-canonical base-pairing and wobble pairs will introduce mutations. Polymerases as catalysts have become very adept at amplifying the chemical bias towards the incorporation of correct nucleotides. Template directed non-enzymatic chemistry uses a different trick. When a wrong nucleotide is incorporated, the rate of the next addition is 10-100 times slower⁷. If a certain amount of time is allocated to the copying process, then the full-length products obtained will be much more homogenous than the fidelity of a single step would predict. The Chen lab has shown that such a mechanism of copying would explore the sequence space through large leaps rather than isolated point mutations. Again, the mechanism of RNA copying selects for the correct products. However, this improvement in fidelity comes at the cost of overall rate of the copying.

A complementary approach to improve the fidelity of the templated synthesis of RNA is to use prebiotically available non-canonical nucleotides with enhanced Watson-Crick base pairing. For example, replacing uridine with prebiotically available 2-thiouridine⁸ improves the rate of the reaction as well as its fidelity. The reaction is less error prone due to the destabilization of the G:U wobble pair by the presence of the sulfur atom. Additionally, the pre-organization of the 2-thiouridine nucleotide in the ³E (C3'-endo) conformation leads to an increase in rate. Conversely, the prebiotically plausible 8-oxo-purines⁹ have been found to be highly mutagenic when present in the template strands¹⁰. Both 8-oxo-guanosine and 8-oxo-adenosine are inferior templates, which discriminate poorly between the four canonical activated nucleotides. However, inosine supports fast and faithful template copying and could potentially replace guanosine in the early stages of evolution.

Our lab has recently developed a deep-sequencing assay for the primer extension reaction which allows for the simultaneous survey of a large number of template sequences¹¹. Using this assay, we were able to examine the influence of the four canonical nucleotides and the 12 imidazolium bridged dinucleotides that can be formed, on the error frequency¹². The fidelity of the primer extension reaction was found to be determined by the competition between the activated mononucleotides and bridged dinucleotides. Reaction conditions that favor activated mononucleotides generate more errors than conditions that favor the bridged dinucleotides, suggesting that primer extension through mononucleotides is the main source of errors. Considering that the bridged dinucleotides hydrolyze to a 5'-monophosphate and an activated mononucleotide, a continuously operating activation process that would convert the hydrolyzed mixture back to the bridged dimer would then also increase the fidelity of the reaction.

In modern biology, if the wrong nucleotide is incorporated proofreading enzymes can backtrack and hydrolyze the terminal phosphodiester bond using an exonuclease domain. Although such a feature is unavailable to non-enzymatic RNA replication, it could potentially be available. After a misincorporation there are three types of phosphodiester bonds in the primer-template complex. Most of the bonds will be incorporated in the helical duplex while others link the residues in the non-hybridized region of the template. The phosphodiester bond between the mis-incorporated residue is in a different chemical environment than the other two types. Because phosphodiester hydrolysis is known to be susceptible to changes in geometry around the bases¹³ and the sugars¹⁴ the three different bonds will hydrolyze at different rates through different transition states. Therefore, a small molecule catalyst such as a peptide or a metal ion complex could facilitate the hydrolysis of the bond connecting the misincorporated nucleotide without touching the other two types of phosphodiester bonds. The phosphodiester bond will be cleaved by intramolecular attack of the 2'-OH generating a cyclic phosphate which will hydrolyze further to a roughly equimolar

mixture of 2' and 3' phosphates. These ends could potentially be recycled and incorporated into a growing strand; however, a non-enzymatic proofreading process has not been discovered yet.

1.3.2 Regioselectivity

Almost all of RNA molecules in extant biology are 3'-5' linked. When RNA is polymerized biochemically, the polymerase active sites control the regioselectivity of the reaction by positioning the 3'-hydroxyl of the primer next to the 5'-monophosphate of the incoming nucleotide, generating a 3'-5' phosphodiester bond. The ribose 2'-OH is intrinsically more reactive than the 3'-hydroxyl, therefore in the absence of enzymes a mixture of regioisomers is expected¹⁵. Although a modest percentage of 2'-5' linkages are tolerated by catalytic RNA species, and the presence of the unnatural linkages in RNA duplexes leads to lower melting temperatures¹⁶, the 2'-5' linkage is a hydrolytic liability¹⁷. Thus, a copying process that generates large proportions of unnatural linkages would not be suitable for a primitive organism which relies on RNA to store its genomic information. Surprisingly, work from our laboratory has shown that a templated RNA copying process that proceeds through the bridged imidazolium intermediate generates almost exclusively 3'-5' linkages¹⁸. Our biochemical observations were explained by following the reaction taking place in a crystal. As soon as the imidazolium dimer is formed and bound to the template the 3'-OH of the primer is approximately 2.0 Å closer to the activated phosphorus atom of the imidazolium dimer⁴. The unnatural 2'-5' linkage is formed only in cases where Watson-Crick base-pairing of the bridged dimer to the template is not possible, as in the case of incorporation of a wrong nucleotide. The rate of copying in such cases is low and will therefore lead to stalled products. Additionally, primers with a terminal 2'-5' phosphodiester linkage react slower than their fully natural counterparts. When present in the templating region, a 2'-5' linkage is not heritable and directs the formation of 3'-5' products exclusively. However, the problem of regioselectivity still persists when copying long tracts of adenosine or uridine residues, for which the extension rate is still slow. A possible solution to this issue invokes the use of non-canonical

bases. Replacing uridine with 2-thiouridine in the template resulted in a dramatic increase in the regioselectivity of the reaction, suggesting yet another benefit of considering an expanded genetic alphabet.

If 2'-5' linkages cannot be avoided, can they be corrected? The Sutherland lab capitalized on the hydrolytic instability of the unnatural linkages as an entry into a cyclic mechanism that mutates 2'-5' linkages into 3'-5' linkages¹⁹. A double stranded RNA duplex with an internal 2'-5' linkage hydrolyses when incubated at slightly alkaline pH and 50 mM Mg²⁺. The hydrolysis proceeds through a cyclic phosphate intermediate which then is converted to an equimolar mixture of 3'- and 2-monophosphates. After acetylation of the monophosphates, the addition of an activating reagent selectively ligates the 3'-monophosphate to the 5'-hydroxyl of the other fractured end, transforming the half of 2'-5 linkages into 3'-5 linkages. Fortunately, the same conditions that hydrolyze the unnatural linkage also deacetylate the 2'-hydroxyl group and produce native RNA. The unacetylated 2'-phosphate regenerates the cyclic phosphate following activation. Mariani showed that after multiple cycles of hydrolysis, acetylation, and ligation, the natural 3'-5' linkages are enriched.

Despite their hydrolytic instability, 2'-5' linkages could prove useful for RNA replication. Because of the high melting point of RNA duplexes, product inhibition is a major obstacle before a complete cycle of replication can be achieved. Our lab has shown that introducing 2'-5' linkages in an RNA duplex considerably lowers its melting point. More importantly, the presence of 2'-5' linkages are surprisingly well tolerated by ribozymes and aptamers, which can function even with 25% of all their phosphodiester bonds being 2'-5'.

Fast and efficient template copying chemistry generates mostly 3'-5' linkages and backbone proofreading enriches them even further. Folded RNAs with a small proportion of 2'-5' linkages have decreased melting temperatures and can access multiple conformations easier than their

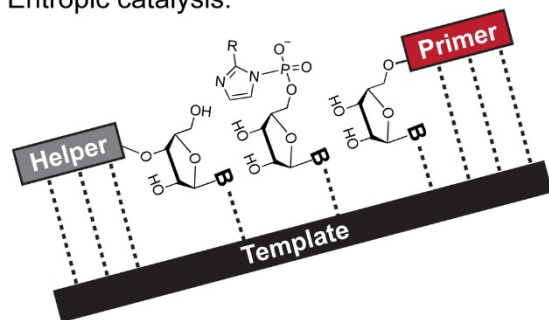
fully 3'-5' congeners²⁰. Considering that template directed oligonucleotide ligation produces 3'-5' linkages almost exclusively²¹ and that functional RNAs tolerate up to 25% 2'-5' linkages, the regioselectivity of non-enzymatic RNA copying is of more minor concern than initially anticipated.

1.3.3 The rate of the reaction

A major problem associated with RNA copying through polymerization is the poor copying of sequences containing adenosine and uridine residues. Unlike the copying of stretches of guanosine and cytosine residues, which is completed in minutes, the copying of A/U stretches is slower than the hydrolysis of the activated intermediates. For imidazole-based substrates, the 12 possible imidazolium bridged intermediates can bind the template through either 4, 5, or 6 hydrogen bonds. Additionally, stacking interactions^{22,23} also contribute to determine the strength of binding of the imidazolium bridged intermediate. Therefore, a dinucleotide intermediate with poor stacking and 4 hydrogen bonds like UU will achieve a maximal rate at much higher concentration than an intermediate that forms 6 hydrogen bonds and has good stacking (GG). At saturating concentrations, the A-U base-pairs should position the imidazolium bridged intermediates in a similar fashion as G-C base-pairs presumably leading to similar rates of primer extension. If there are no reactivation cycles operating, the initial amounts of dimers formed by mixing activated monomers will be sufficient to achieve maximal rates for only a subset of the possible 16 intermediates. This explains why tracts of guanosine or cytosine residues are copied quickly, while adenosine and uridine tracts lag. In the case of NP-DNA the rate of the primer extension reaction is much faster than for ribose, resulting in efficient copying even at sub-saturating concentrations. In the case of ligation reactions, poor binding is less of a problem, for example a tetramer reaches saturating concentrations at ~40 uM concentration. However, ligation reactions need an organocatalyst to reach comparable rates to primer extension through imidazolium bridged intermediates.

A strategy to enhance the rate of RNA copying for any sequences is to improve the affinity for the template of poorly binding monomers or intermediates. In an early example, the Richert group managed to copy through two adenosine residues²⁴ by conducting the reaction at low temperatures and using a downstream helper oligonucleotide²⁴. This helper nucleotide hybridizes downstream of the primer, leaving one binding site empty for the incoming nucleotide. The helper organizes the templating region into a helical geometry, facilitating the binding of the activated

Entropic catalysis:



Covalent catalysis:

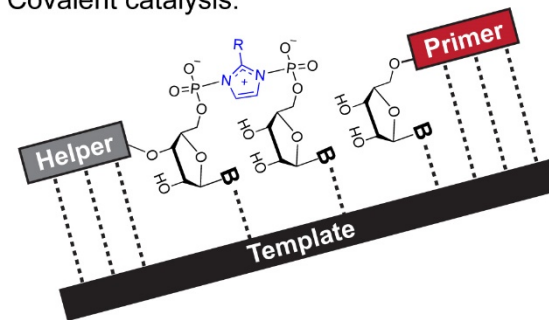


Figure 1.2. Downstream binders enhance the rate of the reaction.

ribonucleotide. In addition, the helical geometry is pre-organized to react, reducing the entropic costs of the primer extension reaction²⁵ (Figure 1.2). Our group took this methodology even further by using 5'-phosphorimidazolide trimers as a helper²⁶. Changing a 5'-phosphate or a 5'-hydroxyl to a 5'-phosphoroimidazolide dramatically increased the rate. Although this finding preceded the imidazolium bridged monomer, it became clear that the helper and the incoming nucleotide can react to form an imidazolium bridged tetramer, a superior electrophile. The imidazolium bridged oligonucleotides react even faster than the imidazolium bridged dimers, benefiting from the pre-organization conferred by the downstream binding region. Crystallographic studies using a non-hydrolyzable analog of the imidazolium bridge show that the bridged oligonucleotides bring the 3'-OH of the primer and the α -phosphorous atom of the bridge in closer proximity and at a better angle when compared a similar analog of the dimer. By combining activated trimers and mononucleotides, we have managed to copy mixed sequences in solution⁵, as well as inside model protocells²⁷.

A drawback of the method is that it requires 256 imidiazolium-bridged tetramers to copy an arbitrary sequence. However, in a system where longer RNA strands are available there will invariably be higher concentration of trimers. This is because the product distribution of all prebiotic methods of obtaining ssRNAs are heavily biased towards shorter sequences. The trimers can also permeate lipid vesicles, so it is reasonable to think they will be around in the environment. Interestingly, the Holliger group has shown that ribozyme mediated primer extension using trimers instead of monomers leads to an increase in the fidelity of the reaction, due to system-wide interactions between the activated trimers²⁸.

Even for efficient systems, like NP-DNA, a ladder of products is formed. What background process is responsible for capping these sequences? Dinucleotide pyrophosphates form abundantly when activated ribonucleotides react with monophosphates in the absence of templates^{29,30}. Additionally, they are the major by-products of the Sutherland isocyanide activation³¹. Our crystallographic studies have shown that pyrophosphate linked nucleotides mimic the dinucleotide intermediate, binding to the template but not reacting³. These species could be competitive inhibitors of templated RNA synthesis. If that were the case, exploiting the hydrolytic instability of the pyrophosphate bond should lead to better yields and higher rates in the primer extension reaction.

1.3.4 High melting temperatures of RNA duplexes

To have ongoing replication you first need to copy genetic material through some mechanism and then remove the product so that the turnover number of the template is larger than one. In any scenario of templated directed replication, the product of the copying step needs to fall off the template to allow the next round of replication. In the examples of nucleic acid replicators experimentally investigated, the rate-limiting step of the replication is the dissociation of the product from the template. If the product is too strongly bound to the template, then exponential

growth is not possible. Unlike exponential growth which supports Darwinian evolution and selection of the fittest, sub-exponential growth enables survival of everyone³². In the Darwinian view of evolution of self-replicators, the fastest replicator will out-compete the slower ones in the fight over resources, and it will be selected for^{33,34}. Sub-exponential growth does not allow for selection, it enables the coexistence of all templates, albeit at different concentrations. This behavior changes when catalytic RNA species are involved. Exponential replication of ligase ribozymes has been shown by Joyce and co-workers^{35,36}. The key to achieving exponential growth in non-enzymatic systems is to destabilize the product-template interface, while maintain the starting materials-template interface. Interestingly, small nucleic acids analogs containing unnatural linkages have been found to display this kind of behavior^{37,38}. Unfortunately, the opposite is true for RNA. A longer oligonucleotide (the product) binds the template stronger than the shorter starting materials.

Template directed RNA copying generates complementary nucleotides that are a barrier towards subsequent rounds of replication. The double stranded RNA molecules that are the final product of the copying chemistry have such high melting temperatures that make liberating functional RNA molecules difficult. If the simplest mechanism for strand separation, thermal denaturing, is not available, different strategies must be employed. Three groups have taken three different approaches to solve this problem, by tweaking three different variables: ionic strength, pH, and solvent viscosity. The Sutherland group observed that when the pH is reduced to 3 the melting point of the duplex is reduced by approximately 30 °C, regardless of the number of base pairs or GC content³⁹. Similarly, the Braun group showed that a thermal gradient in porous rocks induce a microscale liquid-gas cycle in the solution⁴⁰ (Figure 1.3). The condensation of the water

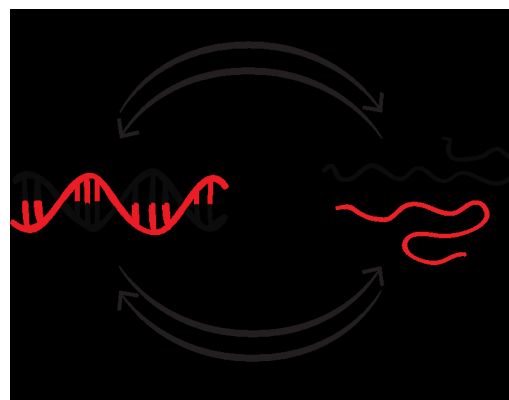


Figure 1.3. pH and salt driven strand separation.

droplets and subsequent reentry into solution reduces the local NaCl concentration at the liquid-gas interface⁴¹, which is where most of the nucleic acids localize in such a scenario. The nucleic acid duplexes dissociate under these low salinity conditions.

However, the various cycling methods (thermal, pH, salt, etc.) are all limited to short RNAs that can be denatured and none solves the fast-reannealing problem. Both the Sutherland and Braun methods destabilize Watson-crick base pairs, which limits ribozyme function and primer reannealing. As soon as the external stimuli are removed and Watson-Crick base pairing is allowed, the two strands will reanneal. Therefore, all these approaches only work if replication is via multiple ligation reactions with oligonucleotides that are long enough to bind and stay bound during the reannealing phase.

The cycling methods shown so far are thermodynamic switches for base-pairing, but a kinetic switch is needed to discriminate between the annealing of long and short oligonucleotides. Hud and coworkers have shown that by changing solvent viscosity a kinetic barrier is introduced for duplex formation⁴². In their system, long nucleic acid duplexes (>300 bp) are subjected to high temperature, and the single strands are liberated. As the solution is cooling, Watson-Crick base pairs reform. The increased viscosity of the solvent slows the diffusion rate of large single-stranded nucleic acids and favors intramolecular pairing. Because the sequences used are of a sufficient length, they will invariably have some degree of self-complementarity. This however means that this method will drastically reduce the sequence space of short oligonucleotides. Considering that the oligonucleotides that have intra-strand interactions are less likely to act as catalysts, perhaps this culling is beneficial in weeding out parasites. Bimolecular processes are affected more strongly by the viscosity since they involve diffusion of species through solution. Therefore, a hairpin folds at the same rate in water or in glycholine, while a 32 base-pair duplex forms considerably slower. Thus, the long strands fold instead of annealing. The intramolecular contacts kinetically trap the duplex, which prevents the double stranded RNA formation, and it

allows the annealing of smaller oligonucleotides on exposed regions. Over time, the smaller strands will invade the self-complementary regions of the templates, provided the correct combination of sequences is available in solution. A ligation step then delivers two new duplexes (Figure 1.4). The Hud lab has recently shown that genetic information from both strands of a gene length duplex (>300 base-pairs) can be copied using repeated heat cycles in glycholine⁴³. Additionally, the same conditions can be used to synthesize and liberate a ribozyme from a duplex and allow its folding and functioning.

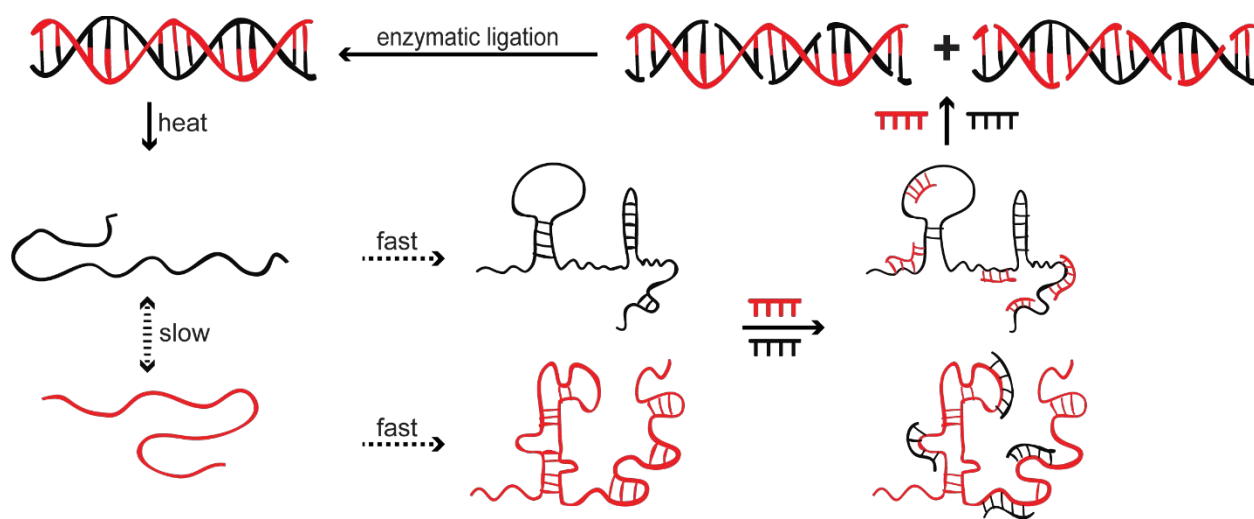


Figure 1.4. Viscosity mediated strand separation followed by short oligonucleotide annealing and enzymatic ligation. The Hud lab use deep eutectic solvents (DES) for these replication studies. A deep eutectic solvent is usually composed of two different solid components, which when combined have a lower melting point than the individual components. Deep eutectic solvents are a new class of ionic liquids, composed of large unsymmetrical ions with low lattice energy (quaternary ammonium salts) and a hydrogen bond donor (such as urea or glycerol). These eutectic solvents not only have increased viscosity but also support folding and catalysis of multiple classes of biopolymers. Although deep eutectic solvents are attractive alternative to aqueous solvents for the reasons outlined above, their prebiotic provenance is questionable. Wet-dry cycles promote evaporation and concentration and can generate concentrated solutions of formamide or urea, but the desired

properties of deep eutectic solvents come from their quaternary ammonium salts. The prebiotic synthesis of these salts is more problematic as of now. Additionally, only enzymatic copying of genetic information has been shown in eutectic solvents. To enable cycles of non-enzymatic RNA replication, a chemical ligation method should be used instead of enzymatic ligation. While it has not been tested whether aqueous methods of chemical ligation would perform as well in DES, intuitively it should be easier for the condensation reaction between an alcohol and a phosphate to occur in non-aqueous environments.

1.3.5 Strand displacement synthesis

Nature does not use any of these cycling methods. At no point in an organism's life cycle is Watson-Crick base pairing disabled globally. When any living organism is heated, put in low pH or low salt condition, it does not revert to a similar state after the stimulus has been removed. The organism dies. Nature uses an energy dissipative cycle to unwind DNA, using molecular machines called helicases. These enzymes are highly complex, multimeric units that use ATP hydrolysis to unzip base-pairs. As soon as the duplex is unzipped, the chemical copying proceeds at the dizzying speeds of polymerases. Thus, local destabilization of Watson-Crick base pairs coupled with fast primer extension enables strand displacement synthesis. The complex architecture of the multi-protein complexes that carries out this transformation hints at the difficulty of performing the task non-enzymatically. However, Szostak and Zhou demonstrated an impressive solution to this problem, enabling strand displacement synthesis in a non-enzymatic fashion⁴⁴. The local destabilization of base-pairs achieved by helicases is mimicked by non-enzymatic toehold migrations, using RNA oligonucleotides. The experimental setup is as follows: you have a primer annealed to a template. The positions downstream of the primer are occupied by a blocker oligonucleotide (Figure 1.5).

For RNA synthesis to continue, the blocker needs to come off so that the imidazolium bridged intermediate can bind and react. The blocker has an extraneous 5'-region overhang, which can be used to promote its dissociation. Oligos complementary

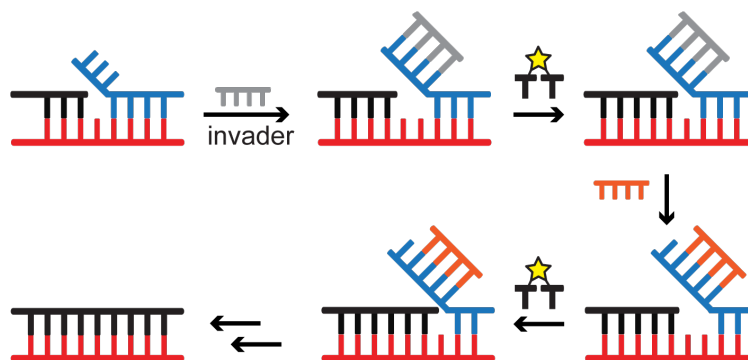


Figure 1.5. Strand-displacement synthesis mediated by invader oligonucleotides. Blue: blocker strand. Grey and Orange: Invader strands.

to this region and the template region, invaders are then added. The system can then exist in two states: an open state where the blocker is still bound to the template, or a closed state where the invader has displaced the blocker liberating the template. Because either state contains the same number of Watson-Crick base pairs, they are occupied to roughly the same extent. Only the later state is conducive to primer extension. The occupancy of the open state and the rate of copying determines the efficiency of non-enzymatic strand displacement synthesis. As soon as an extra nucleotide is added to the primer, the equilibrium between states is disturbed, as the blocker cannot bind to the template with the same efficiency. At this point, the initial invader is replaced by a downstream invader that again partitions the system between an open and closed state. This process can then repeat iteratively until the entire template is copied. Therefore, to copy a sequence of length L you will need $L-k+1$ invader, where k is the length of the invader. The less invaders are required, the more prebiotically plausible the process is. However, as the invaders get longer, the on/off rates for invader binding and dissociation decrease. This hinders the establishment of the open/closed equilibrium in turn hindering RNA copying. The authors demonstrated efficient strand displacement synthesis using hexamers. However, RNA performed poorly when asked to a longer, mixed sequence. This result was expected, as RNA requires downstream binders for efficient extensions, and these species cannot bind in the presence of the blocker. Considering that RNA copying with downstream binders is far more efficient, if invaders

uncover multiple positions on the template, then a monomer bridged to a di- or trinucleotide could bind to the template and react quickly enough with the primer.

1.3.6 Primer independent replication

Despite the usefulness of the primer extension model for experimental studies, its prebiotic plausibility is questionable. In a prebiotic environment where multiple oligonucleotides of various lengths coexist, many different pairings can occur. Therefore, our simplified experimental approach in which template copying starts at a defined position in the context of an RNA duplex is not an accurate representation. Most realistically, both activated and non-activated short nucleotides will anneal at different positions on the template, leaving gaps in between⁴⁵. These gaps offer binding positions for imidazolium bridged dinucleotides so that the copying of the template can be initiated at these nucleation spots. A combination of polymerization and ligation events will then lead to a double stranded product. Furthermore, for multiple cycles of replication to happen, primers need to be constantly supplied to the protocell. However, oligonucleotides longer than four cannot permeate the membranes of the protocells. If longer nucleotides were to participate in the replication cycles, they would need to be generated inside the vesicles. The prebiotic synthesis of nucleic acids generates a reservoir of nucleotides, which can then be activated. These activated nucleotides have been shown to generate single stranded nucleic acids under harsh conditions which involve either freezing⁴⁶ or evaporation⁴⁷ of the solution. If either of these protocols could be adapted such that lipid vesicles could tolerate the conditions, then the longer oligomers could be produced inside vesicles.

An additional issue with the primer-extension model is the difficult copying of the last nucleotide of the template. The imidazolium bridged intermediate cannot bind to the template with only one position available. The last nucleotide will then be copied via an activated mononucleotide, which will be slow, inaccurate and proceed with low regioselectivity. We have recently proposed an

alternative way in which the genome of primordial organisms could have been structured, as a virtual circular genome (Figure 1.6).

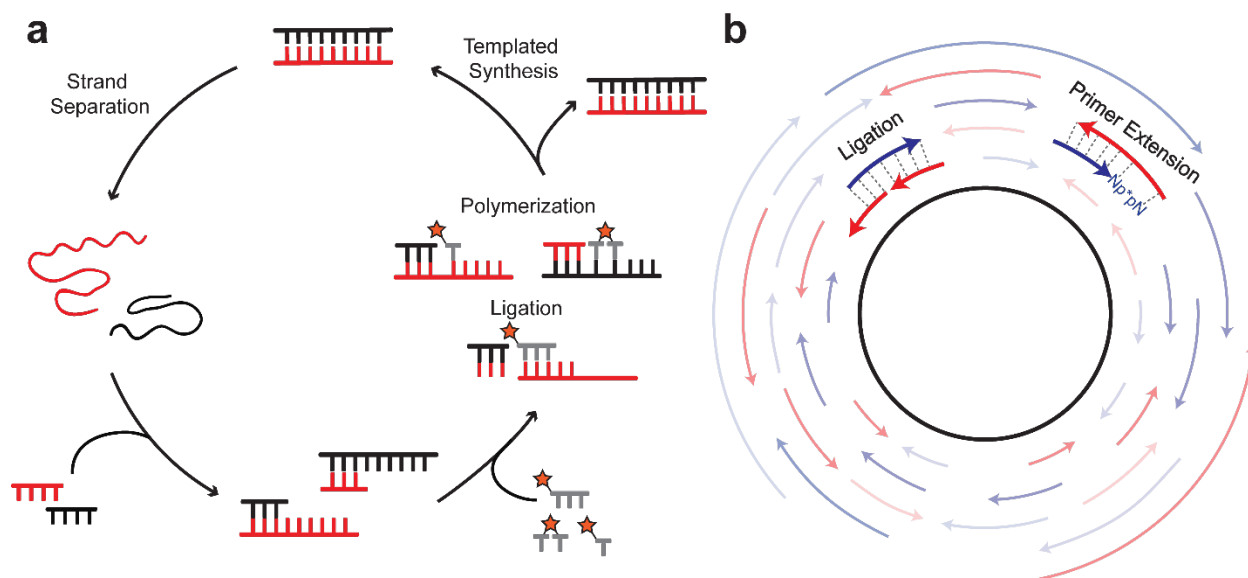


Figure 1.6 The virtual circular genome as an alternative to the primer extension model. (A) A complete cycle of replication in the primer extension model, showing the templated synthesis of double stranded RNA and strand separation of the final product. (B) In the virtual circular genome all possible combinations of oligonucleotides corresponding to the virtual circular genome shown in black are present in solution. These oligonucleotides can participate in ligation and primer extension reactions to drive the replication of the genome.

This virtual circular genome is composed of multiple short RNA oligonucleotides that begin and end at all possible positions on either strand of a virtual circular sequence. These oligonucleotides can anneal in numerous ways, forming a collection of dynamic, interconvertible configurations. We have shown that under these conditions the copying of the genome can occur by the extension of only one nucleotide, obviating the need for copying long sequences. Additionally, since all possible oligonucleotides corresponding to a single genome are present, all invaders and helpers needed for strand displacement synthesis and downstream binder promoted copying are always available. Therefore, the virtual circular genome is a promising alternative to the primer extension model, but experimental investigations are needed to confirm its usefulness.

1.3.7 Reactivation chemistry

The formation and replication of any kind of self-assembling complex requires kinetic barriers that prevent it from decaying into equilibrium⁴⁸. RNA oligomers are thermodynamically unstable with respect to their monomers; the free energy of hydrolysis of a internucleotidic phosphodiester bond is -5.3 kcal/mol⁴⁹. The condensation of ribonucleotide monomers into oligomers is therefore an unfavorable, endergonic process and requires a coupled form of chemical activation. In Nature, polymerase enzymes use the transesterification of phosphoanhydride bonds to form internucleotidic phosphodiester bonds. A fast-copying reaction requires activated ribonucleotides (monomers, dimers, oligomers etc.) that are high in energy. For non-enzymatic polymerizations, hydrolysis of the activated species is a major side reaction. Water is only a slightly weaker nucleophile than a secondary alcohol⁵⁰ and it is present in 56M concentration.

How does then an activated ribonucleotide react rapidly with a secondary alcohol but not with ubiquitous water? Nature has solved this problem by using nucleoside triphosphates. Although high in energy, these monomers are kinetically stable towards hydrolysis, and require the exact positioning enforced by an enzyme to react. The triphosphate moiety becomes activated only when bound to a primer-template-enzyme complex. In the absence of these partners, it is almost inert to nucleophiles. Prebiotic chemistry cannot access such specialized tools from the onset, so it must rely on kinetically labile monomers such as phosphoroimidazolides. Unfortunately, these monomers hydrolyze quickly enough to push the system towards equilibrium. Therefore, a continuous supply of chemical energy is required to convert the monophosphates back into the imidazolides. 2-Aminoimidazole is a privileged molecule: its 2'-amino group is particularly adept at stabilizing the imidazolium intermediate, it can be easily obtained prebiotically⁵¹ and it bears striking resemblance to two other azoles involved in RNA synthesis and maintenance^{52,53}.

An initial foray into the synthesis of ribonucleotide phosphorimidazolides was undertaken by the Fahrenbach lab, in which sodium hypochlorite and cyanide combine with imidazole to give diimidazole imine, which subsequently converts 5'-monophosphates to phosphoroimidazolides⁵⁴. The Sutherland lab recently developed a multi-component reaction, in which an isocyanide, aldehyde and an imidazole moiety convert a nucleotide monophosphate to the activated phosphoroimidazolides³¹ (Figure 1.7).

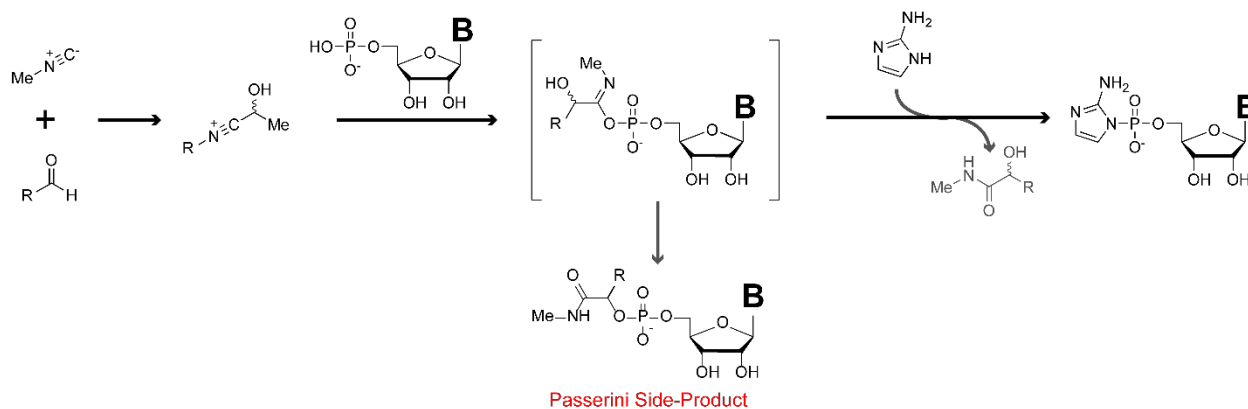


Figure 1.7. Prebiotically plausible multi-component pathway for the activation of 5'-mononucleotides.

All four canonical nucleotides react in good yields to form the desired 2-aminoimidazolides selectively with no modifications on the base being observed. Multiple cycles of reactivations are possible in the same vessel, a feature which is particularly useful for non-enzymatic polymerization.

Our lab has found that replacing acetaldehyde with 2-methylbutyraldehyde increases the yield of phosphoroimidazolides, minimizing the yield of the Passerini side-product⁵⁵. The reaction unfortunately requires high concentrations of imidazole to generate the activated species, but imidazole breaks down the bridged imidazolium intermediate, inhibiting the copying reaction. Replacing the imidazole with an activated phosphoroimidazolide in the activation reaction generates the imidazolium bridged dimer in good yield⁵⁵. Interestingly, the activation conditions improve the reaction of two 5'-phosphoroimidazolide nucleotides, suggesting an unknown mechanism operates in which methylisocyanide reacts directly with a 5'-phosphoroimidazolide.

Unfortunately, the current conditions employed for the generation of imidazolium bridged dimers do not support non-enzymatic RNA copying inside simple fatty acids vesicles. The carboxylate head groups are also activated by isocyanide and once they are modified the vesicles fall apart. The second-generation activation chemistry from the Sutherland lab circumvents this problem by using carboxylic acids instead of aldehydes as a partner for isocyanides. In these acidic activation conditions nucleotide monophosphates are activated in the presence of vesicles, with minimal damage to the lipid components⁵⁶. However, template directed RNA synthesis displays strong pH dependence, performing best at slightly alkaline pH 8. Additional efforts are required to enable continuous activation and RNA replication in fatty acid vesicles. However, it is perfectly reasonable to assume that even in a tight pH window and with the right kind of prebiotic additives, a phosphate can be activated in the presence of a carboxylate.

1.3.8 Divalent metal cations

RNA synthesis, folding and function is mediated by divalent metal ions. The rates of all these processes are sometimes drastically affected by the concentration of free metal ions in solution. Divalent cations are also critical to non-enzymatic templated RNA synthesis. Unfortunately, this requirement clashes with the prospect of RNA replication in fatty acid vesicles. Since fatty acid vesicles are formed by alternating carboxylates and carboxylic acids headgroups, which interact with Mg^{2+} , their stability depends on the abundance of metal ions in solution. Simple fatty acid vesicles cannot tolerate the concentrations of Mg^{2+} at which RNA copying is feasible. Our group circumvented this issue by using a chelator, citrate, to carry out non-

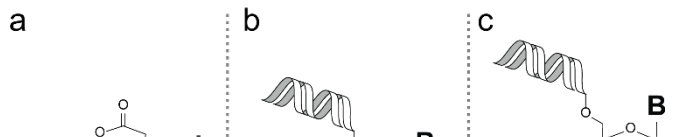


Figure 1.8. The involvement of magnesium in the primer extension reaction. (A) Putative structure for the Mg-citrate chelate. (B) Mg-alkoxide formed at the 3'-terminus of the primer. (C) A possible chelating agent on the 2'-OH of the terminal nucleotide of the primer.

enzymatic RNA copying inside fatty acids vesicles⁵⁷. Citrate binds Mg^{2+} , presumably in a tricoordinate fashion, leaving three empty sites of coordination on the octahedral Mg^{2+} ion (Figure 1.8). The empty sites can then participate in catalyzing the non-enzymatic copying reaction. This form of chelated Mg is also tolerated by the fatty acid vesicles, maintaining their permeability. Interestingly, similar catalytic assistance has since been observed with ribozymes and chelated-Mg. Yamagami and Bevilacqua reported that both amino-acid magnesium chelates⁵⁸ and nucleic acid magnesium chelates⁵⁸ improve the rates of ribozymes in physiological conditions. Similarly, the capacity of chelated magnesium to catalyze is linked to its available coordination sites. As noted by the authors, these interactions signal a catalytic role for these molecules in cellular conditions and possibly prebiotic conditions, where they could act as cofactors. Indeed, modern cofactors have been postulated to be remnants of the RNA world more than forty years ago⁵⁹.

In the absence of a prebiotic pipeline for citrate production and since the rate of RNA copying is reduced by chelation of the magnesium, a more plausible way of protecting vesicles is needed. A possible approach is to change the chemical composition of the lipids. Cyclophospholipid vesicles are more resistant to metal ions⁶⁰ and are prebiotically available through the cyclization of glycerol-phospholipids⁶¹. Other types of compartments have also been explored by Poudyal and collaborators who have also carried out non-enzymatic RNA synthesis inside coacervates⁶², small liquid droplets of two immiscible liquid phases that have been proposed as an alternative to lipid vesicles⁶³. Although most studies on non-enzymatic RNA copying use magnesium, other metals could have potentially catalyzed the process. Our lab has shown that ferrous iron, present in large quantities on the anoxic early earth, can replace magnesium as a catalyst⁶⁴. Iron (II) was shown to outperform magnesium in the ligation and polymerization reaction of nucleotides activated with 2-methylimidazole at neutral and acidic pH values. Divalent manganese, which has a similar radius to Fe^{2+} and Mg^{2+} but is a stronger Lewis acid could potentially improve the reaction even further. However, both Fe^{2+} and Mn^{2+} are likely to be incompatible with lipid vesicles.

A better understanding the role of divalent cations in the non-enzymatic copying reaction would provide avenues to lower cations concentration requirements below the limit tolerated by fatty acid vesicles. Recent experiments from our lab suggest that the metal interacts with the primer through its inner sphere of coordination. The working hypothesis is that the metal assists the deprotonation of the 3'-hydroxyl through an inner sphere contact (Figure 8b), but such interactions are yet to be observed in crystal structures of the copying reaction. Kinetics data shows that reaction rates at pH values can be rescued by increasing Mg^{2+} concentration, suggesting that Mg^{2+} is assisting with the deprotonation of the 3'-hydroxyl⁶⁵. Additional studies are required to clarify the role of the divalent cations. Fatty acid vesicles could be protected by increasing the local concentration of magnesium near the 3'-hydroxyl but lowering the overall concentration in solution. Polymerase enzymes use peptides with negatively charged residues to bring Mg close to the reaction center. A similar mechanism could potentially be operating in non-enzymatic RNA copying. If a magnesium chelating group were attached reversibly to the primer it could bring the cation close to the reaction center, while maintaining a low concentration in solution. An ester linkage at the 2'-hydroxyl (Figure 8c) is a particularly attractive idea since it strongly resembles the termini of charged tRNA and it suggests a link between coded protein synthesis and non-enzymatic RNA synthesis. However, this in turn raises the challenge of chemically synthesizing that linkage prebiotically. Alternatively, ribozymes are capable of aminoacylation, but they require magnesium concentrations above the ones tolerated by fatty acid vesicles. The Sutherland lab has extended their activation chemistry to enable a phosphate relay of amino acids to the 2'-hydroxyl of terminal nucleotides in RNA oligomers. Using this chemistry, this type of conjugates can be accessed and investigations into the birth of translation could be asked.

1.4 Conclusion and Outlook

The work reviewed above shows that over the last ten years the non-enzymatic replication of RNA has seen considerable improvements. The incorporation of an incorrect nucleotide leads to a stalling in copying, guaranteeing that if a longer copying product is generated it is unlikely to contain mismatched nucleotides, effectively raising the fidelity of the process. The regioselectivity of the reaction is excellent when the correct activated imidazolium bridged intermediate can bind to the template, consistent with the observation that the rate of the reaction and the regioselectivity are strongly correlated. Although the copying of adenosine and uracil residues is generally slow, when a downstream binder can assist the reaction proceeds smoothly, opening new sequence space for non-enzymatic copying. The strand inhibition problem can now be circumvented by environmental fluctuations in salt concentrations and acidity, and by using solvents that can form eutectic phases. Additionally, strand displacement synthesis can occur if complementary oligonucleotides are available in the environment. Most importantly, a new mild activation chemistry that improves the yield and rate of non-enzymatic RNA copying has been discovered and can potentially work alongside lipid vesicles. Although the role of Mg is not clear, the discovery of simple chelators has shown that the non-enzymatic synthesis of RNA is compatible with lipid vesicles.

Despite all these advances, the copying of long RNA strands is still a difficult process and cycles of RNA replication have yet to be demonstrated experimentally. The virtual circular genome is an attractive alternative to the classical primer extension model especially because the genome can be replicated without having to copy long stretches of a linear template. This is enabled by the constant reshuffling of oligonucleotide arrangements, triggered by fluctuations in reactions conditions (pH, salt, temperature). Environmental conditions and their plasticity are crucial for the success of non-enzymatic RNA replication. Although most of the issues discussed in this review have potential solutions, these solutions need to be mutually compatible for RNA replication to take place. Therefore, identifying an environment that supports copying, strand

separation and activation simultaneously inside a lipid compartment is the greatest challenge on the way to achieving non-enzymatic RNA replication. For example, is it necessary to constantly deliver activated nucleotides to the reaction medium or is it advantageous to start with a fixed reservoir of nucleotides and constantly deliver methylisocyanide to the reaction? A microfluidic platform would allow such question to be answered and would provide a more realistic experimental setup. Additionally, the virtual genome model also imposes some analytical challenges that will bring about new ways of examining this process.

2. A common and potentially prebiotic origin for precursors of nucleotide synthesis and activation

Reproduced with permission from Fahrenbach, A. C. *et al.* Common and Potentially Prebiotic Origin for Precursors of Nucleotide Synthesis and Activation. *J. Am. Chem. Soc.* **139**, 8780–8783 (2017). Copyright 2017 American Chemical Society.

Imidazoles are thought to have played⁶⁶ important roles in prebiotic chemistry prior to their current biological significance as components of the histidine residues in proteins, where they function as, for example, charge-relay agents in catalytic triads⁶⁷. One of the most important potential roles of imidazoles in prebiotic chemistry came to light through the efforts of Orgel and coworkers^{68–70}, who demonstrated that nucleoside 5'-phosphoro-imidazolides, and especially nucleotides activated with 2-methylimidazole, allow for nonenzymatic template-directed^{71,72} RNA synthesis yielding predominantly the canonical 3'-5' phosphodiester linkage. These phosphoro-imidazolides are more reactive than nucleoside triphosphates as a result of their labile P–N bonds⁷³ and their propensity to form a reactive imidazolium-bridged⁷⁴ dinucleotide. In addition to prebiotic activation chemistry, imidazoles have also been suggested to play catalytic roles⁷⁵, assisting in the oligomerization of amino acids⁷⁶, phosphates^{77,78} and nucleotides^{79,80}, and a number of plausible mechanisms for the prebiotic synthesis of imidazoles have been reported^{66,75,81–85} (further discussion in SI). Other than prebiotic chemistry, compounds containing imidazole motifs, including 2-aminoimidazole and 2-thioimidazole, are known for their medicinal properties⁸⁶.

We became interested in finding a prebiotic synthesis of 2-aminoimidazole (**2NH₂Im**), because we recently demonstrated⁸⁷ that ribonucleoside 5'-monophosphates activated with **2NH₂Im** offer a 10–100-fold enhancement in the rate of nonenzymatic template-directed RNA primer extension compared to those activated with 2-methylimidazole, i.e., in this context **2NH₂Im** is

the most effective leaving group known to date. Furthermore, the conjugate base of **2NH₂Im** is isosteric and isoelectronic with 2-aminooxazole⁸⁸ (**2NH₂Ox**) – a key intermediate in the prebiotic synthesis of cytidine and uridine 2',3'-cyclic phosphates, as reported by Sutherland and coworkers⁵³. The prebiotic synthesis⁵³ of **2NH₂Ox** is known, and begins with the addition of glycolaldehyde and cyanamide, followed by a series of steps that are facilitated by inorganic phosphate acting as both a pH buffer and as a general base catalyst. In this pathway, cyanamide first undergoes addition to the carbonyl of glycolaldehyde. Next, an intramolecular general-base-catalyzed attack of the glycolaldehyde-derived hydroxyl on the cyanamide-derived nitrile carbon leads to a five-membered ring. General-base catalysis by phosphate also likely accelerates C–H deprotonation in the following dehydration step leading to the aromatic **2NH₂Ox**.

Given their structural similarities, we wondered whether **2NH₂Im** and **2NH₂Ox** could share a common prebiotic synthetic pathway (Figure 2.1). Previously reported (non-prebiotic) syntheses of **2NH₂Im** make use of the dimethyl- or diethyl-acetal of α -aminoacetaldehyde^{89–91},

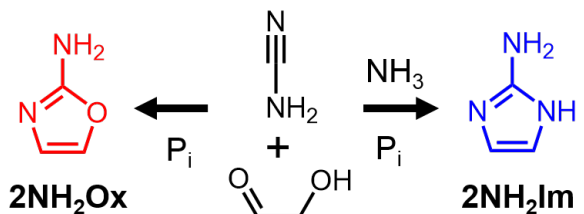


Figure 2.1. Common Prebiotic Synthetic Pathway for 2-Aminooxazole (**2NH₂Ox**) and 2-Aminoimidazole (**2NH₂Im**)

which presumably stabilizes the amino group by preventing enolization and subsequent exchange with water. We reasoned that in order to synthesize **2NH₂Im**, we would need to find a plausible route for the synthesis of α -aminoacetaldehyde. We hypothesized that simply adding a sufficiently high concentration of an ammonium salt to the mixture of glycolaldehyde, cyanamide and inorganic phosphate, would lead to the formation of at least some of this intermediate at equilibrium through an Amadori-type enol-mediated exchange mechanism (Fig. 2.2a, Figure S2.1). Alternatively, exchange of the glycolaldehyde-derived hydroxyl to an amine could take place after the addition of cyanamide to glycolaldehyde. To test this hypothesis, we carried out reactions with aqueous solutions of cyanamide and glycolaldehyde in the presence of 1 M

phosphate at varying concentrations of NH_4Cl at pH 7, 60 °C for three hours. In the absence of NH_4Cl , analysis of the reaction mixture by ^1H NMR spectroscopy revealed that **2NH₂Ox** is produced near exclusively (Fig. S2.2), as expected based on previous reports by Sutherland et al^{53,88}. The addition of 1 M NH_4Cl (Fig. 2.2b) resulted in the appearance of another resonance in the aromatic region of the proton NMR spectrum. Addition of an authentic standard of **2NH₂Im** confirmed that this resonance arises from the H4 and H5 protons of **2NH₂Im**. Increasing the

concentration of NH_4Cl to 5 M resulted in the almost exclusive formation of **2NH₂Im**. It seemed possible that **2NH₂Ox** was formed first and then converted to **2NH₂Im** in the presence of ammonium ions. However, when we treated **2NH₂Ox** with ammonium chloride under identical conditions, we did not observe the formation of **2NH₂Im** (Fig. S2.3). Quantification of the yield was carried out using a calibrated solution of 5'-cytidine monophosphate (CMP), which was added directly to the NMR tube; we used its H5 and H6 resonances as standards for integration. After one hour, the reaction was complete, and the yield was determined to be 15%. Although this yield is not as high as that previously reported⁵³ for **2NH₂Ox** (>80%), it is important to note that no other major products could be observed in the ^1H NMR spectrum.

Understanding how the ratio of **2NH₂Im** to **2NH₂Ox** varies under different pH regimes and ammonium ion concentrations is important for evaluating the reaction in the context of potential

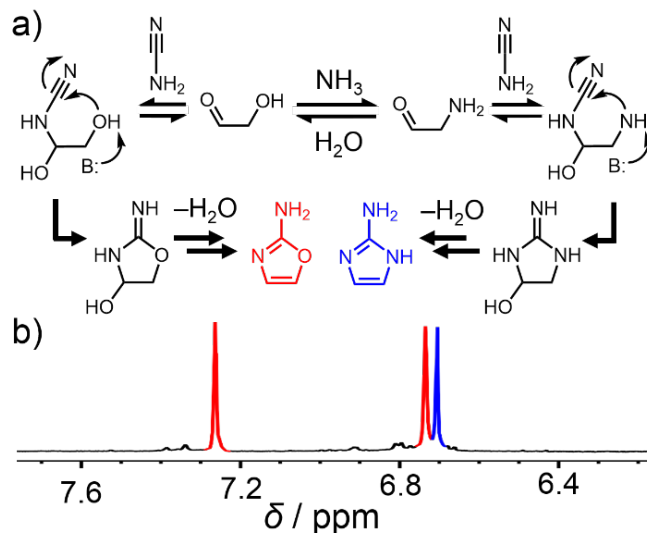


Figure 2.2. a) Possible mechanistic pathway for the synthesis of **2NH₂Ox** and **2NH₂Im**. The initial formation of α -aminoacetaldehyde from glycolaldehyde and ammonia is thought to be a potential intermediate on the path to **2NH₂Im**, although this exchange reaction may also occur after addition of cyanamide. Following intramolecular cyclization, dehydration leads to the final products. b) Partial ^1H NMR spectrum (400 MHz, D_2O) showing the H4/H5 aromatic resonances for **2NH₂Ox** (red) and **2NH₂Im** (blue) after reaction of glycolaldehyde, cyanamide, sodium phosphate and ammonium chloride, all at 1 M, for 3 hours at pH 7 and 60 °C.

geochemical scenarios. We systematically examined the reaction at pH 4, 5.5, 7 and 8.5 with NH_4Cl concentrations of 0, 0.5, 1, 2, 3, 4 and 5 M. All reactions were monitored by ^1H NMR after one, two and three hours, and the ratios of **2NH₂Im** to **2NH₂Ox** were measured by resonance integration (Table S2.1). The ratios after three hours are plotted in Figure 2.3. For all pH values the ratio of **2NH₂Im** to **2NH₂Ox** increases with increasing NH_4Cl , while mildly acidic pH also tends to increase this ratio. At least part of this effect of pH can be explained by the

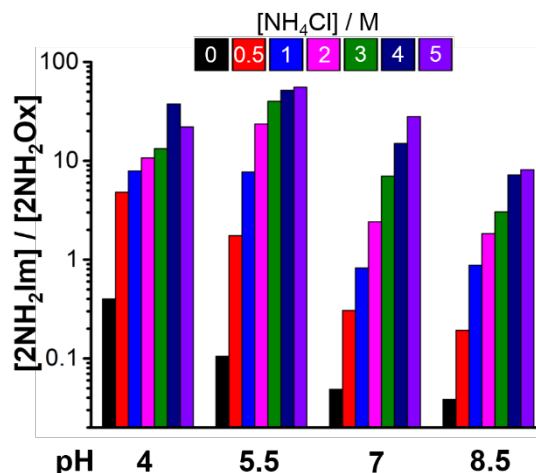


Figure 2.3. Bar graph displaying the ratios of **2NH₂Im** to **2NH₂Ox** at varying pH and NH_4Cl concentrations. All ratios were determined by ^1H NMR spectroscopy from reactions that were carried out at 60 °C with 1 M sodium phosphate and monitored over a period of 3 h. The ratio reaches a maximum at pH 5.5 and 5 M NH_4Cl .

fact that increasing acidity does not tend to favor dehydration of the 4-hydroxy intermediate of **2NH₂Ox**, a minor species which can be detected under the acidic conditions tested. A maximum ratio of $[\text{2NH}_2\text{Im}]/[\text{2NH}_2\text{Ox}] = 56$ was obtained at a pH value of 5.5 with 5 M NH_4Cl after 3 hours. For most cases, the ratios tended to increase over time. The reason for this increase is not the increase in concentration of **2NH₂Im** after one hour, but rather the decrease in the yield of **2NH₂Ox** (Table S2.1). This decrease in the yield observed over time for most conditions tested is likely a reflection of the greater reactivity of **2NH₂Ox**, possibly forming a distribution of side-products most of which are not easily identified by ^1H NMR. Similar yields for both **2NH₂Ox** and **2NH₂Im** were obtained when either 100 or 50 mM concentrations of the starting materials were used. We carried out a preparative scale reaction for the synthesis of **2NH₂Im** at a pH of ~5.3 employing 1 M $\text{NH}_4\text{H}_2\text{PO}_4$ and 5 M NH_4HCO_2 and obtained a 41% isolated yield (Figure S2.4). Given the general mechanistic features of the above synthetic reactions, we anticipated that the prebiotic formation of imidazoles may be rather general. We set out to make 2-thioimidazole (Fig. S2.5). Heating a 1 M solution of NH_4SCN and glycolaldehyde at pH 4 in 4 M NH_4Cl at 60 °C for

24 hours led to the formation of 2-thioimidazole with no other major products observable by ^1H NMR, although the yield was relatively low at 6.2%.

In order to gain more insight into the distribution of products formed during the formation of **2NH₂Im** and **2NH₂Ox**, we monitored the reactions by Q-TOF LCMS (positive mode) using the same conditions as shown in Figure 2. First, the yields and ratios of [**2NH₂Im**]/[**2NH₂Ox**] determined by LCMS are consistent with those determined by NMR (Table S2.2). Other than the m/z values arising from **2NH₂Im** and **2NH₂Ox**, we also observed a major peak at $m/z = 145.06$ (Fig. S2.6), and a minor peak at $m/z = 144.08$ (Fig. S2.7), consistent with the addition products of glycolaldehyde with **2NH₂Ox** and **2NH₂Im**, respectively. These potential addition products could be either the hemiaminal obtained from the attack of the exocyclic amines of **2NH₂Ox** or **2NH₂Im**, or glycols formed by the nucleophilic attack of the C5/C4 carbons on the carbonyl group of glycolaldehyde.

Knowing that **2NH₂Ox** and **2NH₂Im** can be made in the same reaction flask in approximately equimolar ratios at pH 7 and $\sim 1\text{ M NH}_4\text{Cl}$, we asked whether these two compounds would display sufficiently different reactivities such that **2NH₂Ox** could be channeled towards nucleotide synthesis, while **2NH₂Im** would be preserved for later nucleoside 5'-phosphate activation. The next step on the pathway from **2NH₂Ox** to the cytidine and uridine cyclic phosphates involves a cycloaddition reaction with glyceraldehyde, which generates a mixture of ribo- and arabino-furanosyl aminooxazolines⁵³ as the major products. For **2NH₂Im** to be preserved for subsequent nucleotide activation, it would have to be significantly less reactive with glyceraldehyde than **2NH₂Ox**. To address this question, we reacted a 1:1 mixture of **2NH₂Ox** and **2NH₂Im** with glyceraldehyde at 40 °C in the presence of inorganic phosphate for 24 hours while monitoring the reaction mixture by high-resolution LCMS (Fig. 2.4). The mixture of pentose aminooxazolines was detected in the reaction even after only ~ 10 minutes of reaction time, and their concentration reached a maximum after approximately three hours (Fig. S2.8). After 24 hours, nearly all the

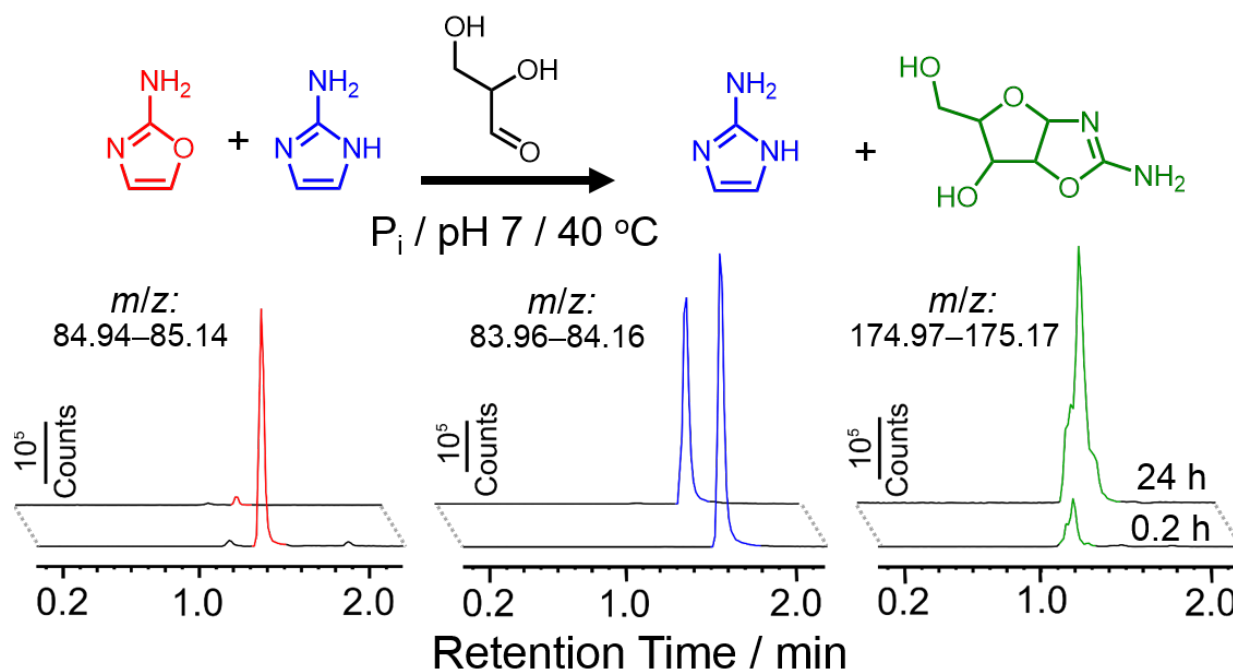


Figure 2.4. Selective cyclization of rac-glyceraldehyde with **2NH₂Ox** in the presence of **2NH₂Im**. The reaction was carried out at 1 M of each component and was monitored by high-resolution (Q-TOF) LCMS using a C₁₈ column. All traces are extracted ion chromatograms for *m/z* values that correspond to the [M + H]⁺ ions for **2NH₂Ox** (red), **2NH₂Im** (blue) and the mixture of aminooxazoline stereoisomers (green). All chromatograms were extracted with a tolerance ± 0.1 Da. The 24 h chromatograms have been displayed offset by ~10 s for clarity.

2NH₂Ox had been depleted from the reaction mixture, while ~80% of the **2NH₂Im** remained. Analysis by ¹H NMR spectroscopy confirmed this result (Fig. S2.9). We also detected lesser amounts of a product with an *m/z* of 174.05, a value which is consistent with the [M + H]⁺ mass of the analogous product but cyclized with **2NH₂Im**. Although it would appear that the **2NH₂Im** can also react with glyceraldehyde, based on the initial rates measured from control experiments reacting **2NH₂Im** and **2NH₂Ox** with glyceraldehyde individually in separate solutions, we estimate that the reaction with **2NH₂Ox** is about an order of magnitude faster. Indeed, the reaction of **2NH₂Im** with glyceraldehyde is sufficiently slow such that the reaction stops after the consumption of only ~20% of **2NH₂Im**. The likely reason for this effect is that the isomerization of glyceraldehyde to dihydroxyacetone is a competing process occurring at a similar rate, preventing the reaction from going to a completion. We suspect that the slower reaction kinetics of **2NH₂Im** compared to **2NH₂Ox** are at least in part a result of the greater aromatic stability of **2NH₂Im**, the greater nucleophilicity of **2NH₂Ox**, or both. We also examined the case of a “one-

pot” reaction, in which a solution of 1 M cyanamide, glycolaldehyde, glyceraldehyde, ammonium chloride and sodium phosphate at pH 7 was heated to 40 °C and monitored by LCMS over time. Analysis of the data (Fig. S2.10) reveals an initial rapid increase in the concentration of **2NH₂Ox**, which reaches a maximum around 1.5 hours, and thereafter slowly decreases. The **2NH₂Im**, on the other hand, shows a slower initial accumulation which levels out after about 5 hours – no subsequent decrease in concentration is observed. In addition, ribose and arabinose aminooxazolines are still obtained under these conditions. These results provide additional evidence that the greater stability of **2NH₂Im** allows it to persist and potentially accumulate for later activation chemistry.

Finally, we demonstrate that *N*-cyano-2-aminoimidazole (**2NH₂ImCN**) is capable of activating CMP to furnish (Figure 2.5) cytidine-5'-phosphoro-(2-aminoimidazole) (**2NH₂ImpC**). There have been several reports in the literature suggesting that *N*-cyanoimidazole can serve^{92,93} as an

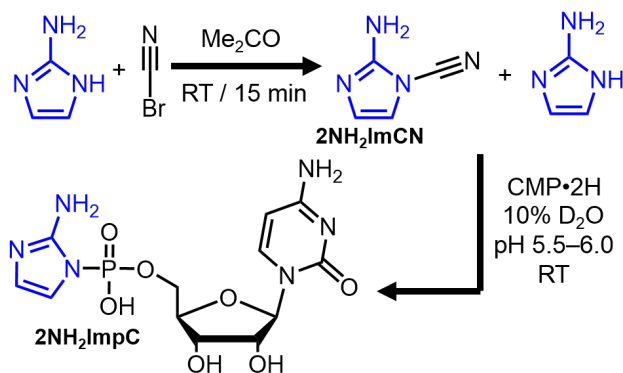


Figure 2.5. Synthesis of Cytidine-5'-phosphoro-(2-aminoimidazole) Making Use of *N*-cyano-2-aminoimidazole

activating agent for the formation of phosphodiester bonds. **2NH₂ImCN** appeared attractive as a potentially prebiotic activating agent, because it represents one of the simplest possible extensions of the chemistry reported here, i.e., oxidative coupling with hydrogen cyanide. We formed **2NH₂ImCN** by reacting⁹⁴ **2NH₂Im** with cyanogen bromide in acetone at room temperature for 15 min (Fig. S2.11). The reaction was concentrated to near dryness, at which point an aqueous solution of CMP with 10% D₂O was added and the pH adjusted between 5.5 and 6. After about ten minutes at room temperature, ³¹P NMR spectroscopic analysis revealed ~20% conversion to **2NH₂ImpC** (Fig. S2.12 and S2.13). A maximum of about 40% conversion was obtained after 1.6 hours. By addition of another freshly prepared batch of **2NH₂ImCN**, a final

conversion of ~75% was achieved. While the prebiotic synthesis of **2NH₂ImCN** has not yet been achieved, this synthesis of **2NH₂ImpC** serves as a proof of concept, highlighting the potential usefulness of *N*-cyanoimidazoles as prebiotic activating agents.

In summary, we have demonstrated a prebiotically plausible synthetic pathway for 2-aminoimidazole that shares a common origin with the synthesis of 2-aminooxazole. Recently, Powner et al. showed that 2-aminothiazole can be efficiently synthesized in a similar fashion as **2NH₂Ox** and **2NH₂Im**, starting from cyanamide and β -mercaptoacetaldehyde⁵². 2-Aminothiazole forms stable crystalline amins with aldehydes, but not with ketones, allowing for the concomitant accumulation and purification of reactive aldehydes, and the chemical selection of proteinogenic aminoacids. Remarkably, all three compounds in the series 2-amino-oxazole/imidazole/thiazole seem to have important potential prebiotic roles. In the shared pathway for **2NH₂Ox** and **2NH₂Im** production, the relative yield of each species depends on the pH and ammonium chloride concentration, with higher ammonium ion concentrations and mildly acidic pH favoring **2NH₂Im**. At neutral pH, this proposed pathway for the prebiotic synthesis of **2NH₂Im** requires high concentrations of aqueous ammonia, on the order of 1 M to generate comparable amounts of **2NH₂Im** and **2NH₂Ox**. There are several scenarios in which ammonium ions could have been generated in a concentrated form within an ancient aqueous reservoir. In one scenario proposed by Sutherland⁹⁵, cyanide produced in the atmosphere rains out and is captured by ferrous ions as ferrocyanide; salts of ferrocyanide then precipitate and accumulate over long periods of time. Subsequent thermal processing of deposits of magnesium ferrocyanide by magma or impacts would generate magnesium nitride, which upon moistening would hydrolyze, releasing ammonia. At the same time, calcium ferrocyanide generates calcium cyanamide upon heating, which in the presence of water releases cyanamide. Other pathways to ammonia sources have also been suggested (See SI). Finally, the greater stability/slower reactivity of **2NH₂Im** in comparison to **2NH₂Ox** suggests that the former could accumulate over

time even as the latter is continuously processed into intermediates on the path to nucleotide synthesis. Having a mechanism for the simultaneous production of **2NH₂Im** and **2NH₂Ox** suggests the tantalizing prospect of an ancient prebiotic reaction network that could have led to both nucleotide synthesis and to the subsequent chemical activation of those nucleotides in a manner suitable for efficient nonenzymatic template-directed replication. While we have shown that *N*-cyano-2-aminoimidazole can serve as an activating agent, one of the great challenges ahead is understanding how such an activating agent, or another mechanism altogether for the prebiotic activation of nucleoside 5'-monophosphates, could have arisen from such a network.

3. A Mechanistic Explanation for the Regioselectivity of Non-Enzymatic RNA Primer Extension

Reproduced with permission from Giurgiu, C., Li, L., O’Flaherty, D. K., Tam, C. P. & Szostak, J. W. A Mechanistic Explanation for the Regioselectivity of Nonenzymatic RNA Primer Extension. *J. Am. Chem. Soc.* **139**, 16741–16747 (2017). Copyright 2017 American Chemical Society.

Introduction

Elucidating a pathway from prebiotic chemistry to Darwinian evolution is a fundamental problem in chemistry. Darwinian evolution requires a replicator that can propagate genetic information⁹⁶, and RNA is an promising candidate for such a task. RNA stores information in its sequence of four bases, and folds into complex three-dimensional structures capable of catalysis and molecular recognition. In 1968 Crick, Orgel and Woese proposed that all life on Earth started from RNA, a scenario which later became known as the “RNA World”^{97–99}. In this scenario, early life forms had an RNA genome that encoded RNA enzymes¹⁰⁰. The original RNA world hypothesis suggested that genomic replication would be a RNA catalyzed process, executed by a polymerase ribozyme. Despite recent progress^{101,102}, major improvements in the accuracy and efficiency of such ribozymes are needed to permit self-replication. A plausible mechanism through which a RNA-based system could have evolved efficient polymerase ribozymes is through cycles of non-enzymatic RNA replication.¹⁰³ Over the last 50 years, considerable effort has gone into achieving non-enzymatic RNA replication in the laboratory, with little success. These attempts have highlighted several major issues that must be understood to enable genetic replication in an RNA world setting¹⁰³. One such problem is the poor regioselectivity of the template directed-synthesis of RNA: the phosphodiester bond connecting the ribonucleotides can form between the 5’ phosphate group of one nucleotide and the hydroxyl group on either the 2’ or 3’ carbon atom of

the next nucleotide. In biology, templated RNA synthesis is performed by polymerase enzymes, which exclusively form 3'-5' phosphodiester bonds. In the absence of strict steric constraints imposed by enzyme active sites, a mixture of products is obtained^{104,105}.

The formation of 2'-5' linkages was long considered a fatal flaw of template-directed RNA synthesis because it was thought that linkage heterogeneity would impede the evolution of functional RNA molecules¹⁰³. However, our group has shown that nucleic acids with non-heritable backbone heterogeneity can generate nucleotide binding aptamers through in vitro evolution¹⁰⁶. Furthermore, some RNA aptamers and ribozymes containing up to 25% 2'-5' linkages retain molecular recognition and catalytic properties¹⁶. These recent findings suggest that the poor regioselectivity of non-enzymatic RNA primer extension is not an insurmountable hurdle to chemical RNA replication. On the contrary, backbone heterogeneity assists with another difficult challenge, the strand separation problem. RNA duplexes of over 30 base pairs are unlikely to thermally denature in conditions compatible with the copying chemistry. Therefore, a primer-template complex that is efficiently extended inhibits further rounds of template-directed synthesis. Backbone heterogeneity lowers the melting point of RNA duplexes, allowing strands to separate at lower temperatures^{16,107}. In addition, 2'-5' linked RNA strands retain templating abilities. Switzer and colleagues have shown that primer extension proceeds in high yield, but with a lower rate, on a 2'-5' linked template¹⁰⁸. As a result, backbone heterogeneity potentially assisted the evolution of function in the early stages of the RNA world.

The regioselectivity of the reaction is affected by the metal catalyst¹⁰⁹, the leaving group on the activated monomers¹¹⁰ and by using activated oligomers instead of monomers²¹. Despite extensive efforts, it is not known precisely how these factors operate.

Our recent discovery of a superior leaving group for RNA primer extension⁸⁷ and of an alternative mechanism for the reaction⁷⁴ has prompted us to revisit the regioselectivity problem. In this mechanism, two monomers react with each other to form an imidazolium-bridged dimer, which

then presumably binds to the template at two adjacent sites. Here, we examined the influence of the leaving group, nucleobase and of the templating sequence on the regioselectivity of chemical RNA primer extension. Surprisingly, fast and efficient reactions preferentially formed 3'-5' linkages. In all instances where we observed high 3'-5' regioselectivity the following two criteria were invariably met: (1) Watson-Crick base pairing was observed at the extension site and at the adjacent upstream and downstream positions and (2) the downstream binding nucleotides had a leaving group on their 5' end. Unnatural 2'-5' linkages are formed in RNA primer extension during the copying of the last base of the template, through mismatched copying, and when multiple adenosine or uridine monomers are added sequentially. The relatively low frequency of such events potentially leads to a proportion of 2'-5' linkages that lowers the melting temperature of the RNA duplex, without preventing the evolution of functional RNA sequences. This leads us to consider that the backbone heterogeneity arising from chemical RNA primer extension was not an issue in the origin of life.

Results

We have adapted Orgel's well established method of quantifying the regioselectivity of templated non-enzymatic RNA primer extension¹¹¹. In this method, the RNA primer was extended by the templated addition of 5'-phosphorimidazolid monomers (Figure 3.1a). The reaction products were then treated with either RNase T1 (Figure 3.1b) or RNase A (Figure 3.1c), endoribonucleases that specifically cleave 3'-5' linkages to 3'-phosphates after guanosine residues, and pyrimidine residues, respectively. The sequences of the primers were designed such that the enzymes cleave at a single site, after the first bond that is formed during the primer extension reaction. Thus, if the first added residue is connected through a 3'-5' linkage, the primer would be cleaved. In contrast, if the first residue is connected through a 2'-5' linkage the extension products would be insensitive to cleavage. The RNase T1 digestion assay can detect as little as 6%, while the RNase A assay can detect as little as 2% 2'-5' linkages (Figures S3.1 and S3.2).

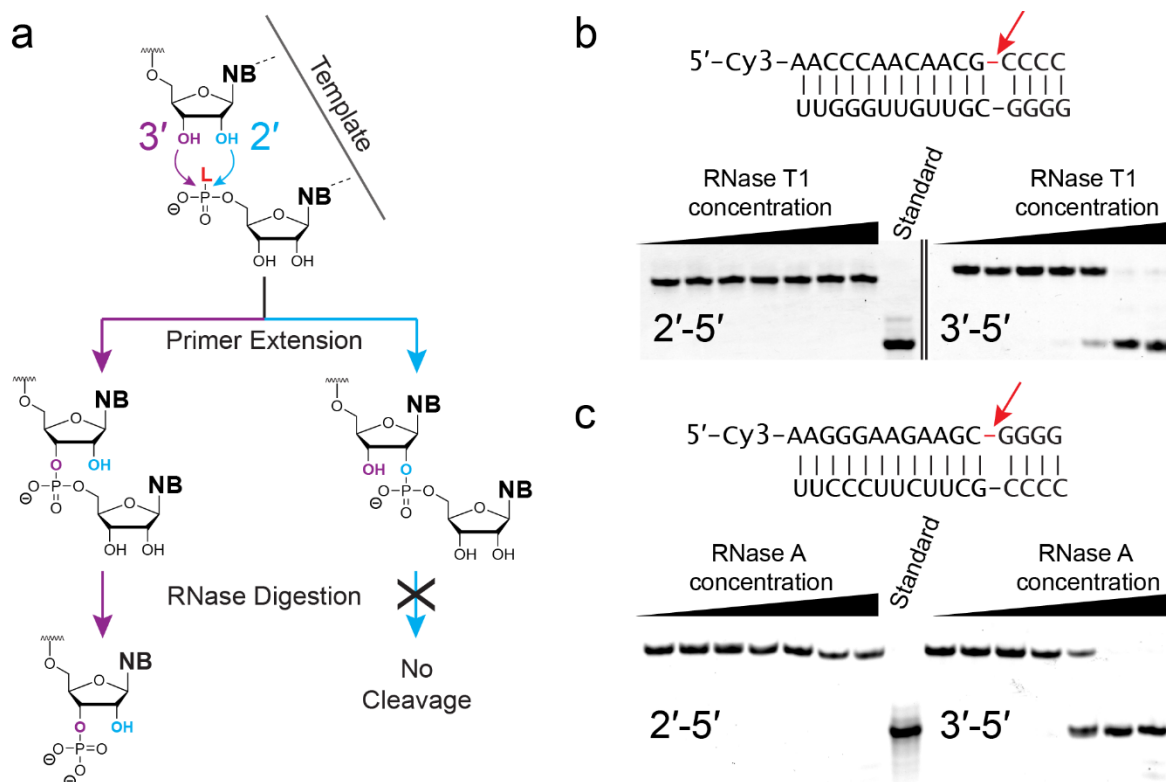


Figure 3.1. An assay for determining the percentage of 2'-5' linkages in an oligonucleotide. (a) Templated RNA primer extension forms either a 2'-5' linkage (cyan pathway) or a 3'-5' linkage (magenta pathway). Endoribonucleases selectively cleave 3'-5' linkages; NB stands for nucleobase. (b) Top: sequence of the primer-templated duplex used in the RNase T1 assay – if the red bond indicated by the arrow is a 3'-5' linkage it will be cleaved by the enzyme. Bottom: PAGE analysis of the enzymatic digest confirms that a 2'-5' extended primer is not cleaved by the enzyme at any of the concentrations tested, while the 3'-5' linkage is fully cleaved at high RNase concentration. Standard refers to the synthetically obtained 3'-phosphate primer. Assay dynamic range and detailed experimental conditions can be found in the supplemental information. (c) Top: sequence of the primer-templated duplex used in the RNase A assay, notations identical to panel (b); Bottom: PAGE analysis of the RNase A assay, analogous to panel (b) bottom.

We first examined how the nucleobase on the incoming monomer affects the ratio of 2'-5' to 3'-5' linkages (Figure 3.2). Our group has previously identified 2-aminoimidazole activated 5'-nucleotide monophosphates (2-AImpN) as the most efficient monomers for RNA polymerization to date⁸⁷. Primer extension with 2-AImpG **2a** and 2-AImpC **3a** monomers was fast and regioselective, and the proportion of 2'-5' linkages formed was below the detection limit of the assays. In addition, the regioselectivity of the reaction was identical over a wide range of concentrations of the activated monomer **2a** (Figure S3.3).

However, 2-AImpA **1a** and 2-AImpU **4a** additions were considerably slower and formed 17% and 46% 2'-5' linkages respectively. A and U pair through two hydrogen bonds while G and C pair

through three hydrogen bonds, suggesting that the strength of the Watson-Crick interaction plays an important role in the regioselectivity of the reaction. We next measured the regioselectivity of the primer extension using 2-thiouridine (s^2U) phosphorimidazolid monomer **5a**. If base

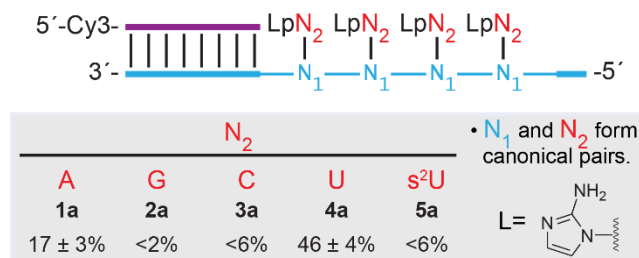


Figure 3.2. Nucleobase influence on regioselectivity. Enzymatic digestion was carried out with RNase A for the purine monomers and RNase T1 for the pyrimidine monomers. The percentages represent the proportion of 2'-5' linkages formed and were obtained from four independent experiments.

pair stability is important, the greater thermodynamic stability of the $s^2U:A$ base pair compared to the canonical $A:U$ pair should decrease the proportion of 2'-5' linkages. The s^2U activated monomer **5a** formed no detectable 2'-5' linkages, supporting this hypothesis.

The regioselectivity of chemical RNA primer extension depends on the imidazole leaving group used^{30,112}. For example, reactions using imidazole activated monomers are biased towards the formation of 2'-5' linkages¹¹³, whereas the 2-methylimidazole group is selective for 3'-5' linkages³⁰. We examined how the leaving group influences the regioselectivity of the reaction when using different guanosine monomers. The proportion of 2'-5' linkages obtained correlates

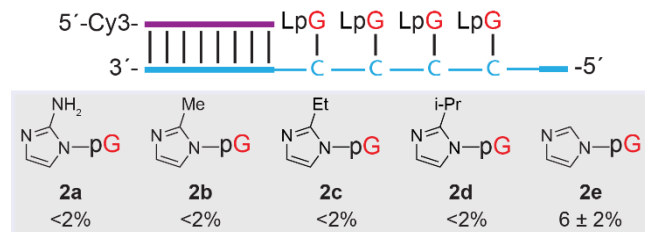


Figure 3.3. Leaving group influence on regioselectivity. Enzymatic digestion was carried out with RNase A. The percentages represent the proportion of 2'-5' linkages formed and were obtained from four independent experiments.

with the rate of extension; the more reactive 2-aminoimidazole and 2-methylimidazole leaving groups led to 3'-5' linkages exclusively, while the slowest of the series, imidazole, produced 6% 2'-5' linkages (Figure 3.3).

The nature of the leaving group does not have as large an influence on the regioselectivity as on the rate of the reaction. Although there is a 7-fold difference in reaction rates between **2a** and **2e**, both leaving groups form less than 10% 2'-5' linkages. The identity of the nucleobase seems to be the main determinant of regioselectivity. Thus, the recognition of the monomers by the template

is crucial for a regioselective reaction. However, the Orgel group has shown that the Pb^{2+} catalyzed polycytidilic acid templated oligomerization of guanosine monomer **2c** produces 2'-5' linked polyguanylic acid with remarkably fast rates.^{114,115} The difference in regioselectivity observed between our experiments and the Orgel experiments presumably arises from the use of a different metal catalyst and possibly the formation of RNA triplex structures in the latter.

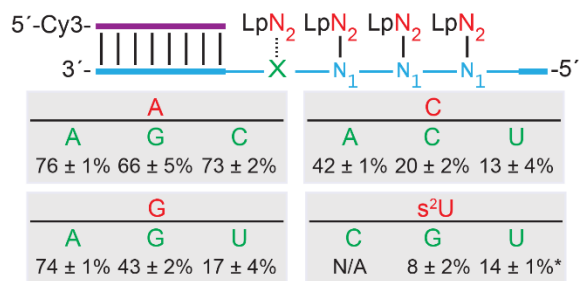


Figure 3.4. Mismatches lead to decreased regioselectivity. (a) RNase A was used for the enzymatic digestion of G and A monomers, and RNase T1 was used for C monomers. The nucleotides N₁ and N₂ form canonical base pairs; X and N₂ form mispairs. L stands for the 2-aminoimidazole leaving group. The percentages represent the proportion of 2'-5' linkages formed and were obtained from four independent experiments. No extension products were observed when the s²U monomer was added across a templating cytosine residue. *The value was obtained from three independent experiments.

Assuming the template and monomer interact in a canonical Watson-Crick fashion, we investigated the cases in which such a pairing was not possible. Wu and Orgel reported that primer extension with 2-methylimidazole activated 5'-guanosine phosphorimidazolid **2b** on a G template results in substantial 2'-5' linkage formation¹¹⁶. We explored the different possibilities of mismatch incorporation, using 2-aminoimidazole activated nucleotide monophosphates (Figure 3.4). The

regioselectivity of the reaction decreased when compared to the case of canonical base pair formation. Purine-purine mismatches form the highest percentages of 2'-5' linkages we have observed in any condition tested. Pyrimidine-pyrimidine mismatches form approximately 3-fold less 2'-5' linkages. A possible explanation for this phenomenon is that the template can form a one nucleotide loop when the first templating nucleotide is a pyrimidine, and the products obtained are the result of frameshift mutations. Interestingly, the 2-thiouridine monomer **5a** adds to the primer in poor yields in all three cases examined (no addition was observed when the templating nucleotide was cytosine), but with high 3'-5' regioselectivity.

The rate of non-enzymatic primer extension decreases in the absence of activated downstream binding nucleotides¹¹⁶. To investigate the effect of downstream binding on regioselectivity, we

designed a primer-template combination in which a single nucleotide is available to template the extension (Figure 3.5). In this case all three nucleotides (A, G, C) formed between 15 to 49% of 2'-5' linkages. Similar to the experiments presented in Figure 2 the regioselectivity did not differ greatly between the 2-methylimidazole and 2-aminoimidazole leaving groups, for any given nucleobase. However, for the imidazole derivative **2e**, twice the fraction of 2'-5' linkages formed when compared to the faster and more efficient **2a** and **2b** monomers.

Copying of the last nucleotide of a template also resulted in decreased regioselectivity, which implies that downstream binding nucleotides promote 3'-5' regioselective reactions.

The mechanism through which downstream binders influence the reaction could involve base stacking, leaving group – leaving group interactions, or a mixture of both.

Alternatively, two activated monomers can interact covalently to form an imidazole-bridged dimer⁷⁴, and the conformational constraint resulting from the formation of two base pairs could influence regioselectivity. To better

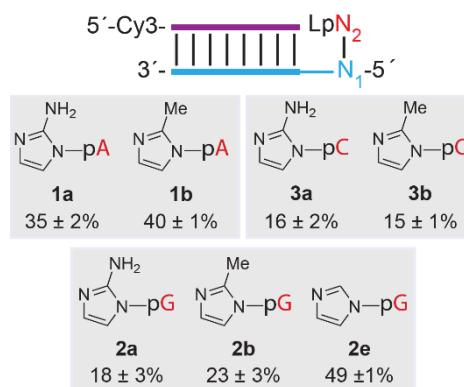


Figure 3.5. Copying of the last nucleotide generates backbone heterogeneity. For G and A monomers enzymatic digestion was carried out with RNase A; for the C monomers RNase T1 was used. The percentages represent the proportion of 2'-5' linkages formed and were obtained from four independent experiments.

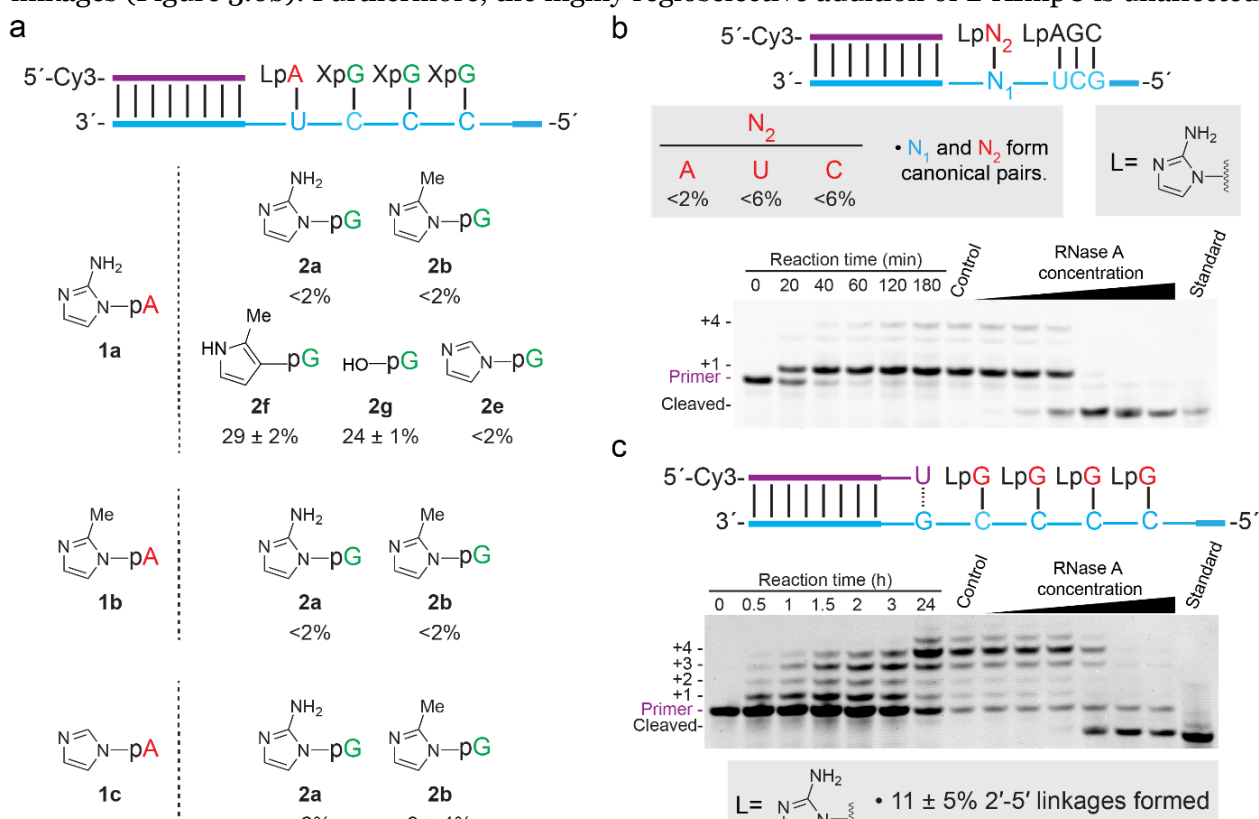
understand how this interaction influences the regioselectivity of the reactions, we varied the leaving groups on the adenosine monomers and on the downstream binding guanosine helpers when copying a UCCC template (Figure 3.6a). Unexpectedly, almost all of the combinations of imidazole activating groups we have tested formed 3'-5' linkages exclusively. When the adenosine monomer was omitted, the rate, yield and regioselectivity of the reaction decreased dramatically, suggesting that the correct monomer is added (Figure S3.4). Additionally, when the reaction was carried out with a mixture of all four canonical nucleotides, all of the products were 3'-5' linked. Considering that the mismatch addition opposite a templating uracil residue formed at least 13%

of 2'-5' linkages, and that the incorporation of a mismatched nucleotide causes a sharp decrease in rate⁷, we conclude that the fidelity of the reaction was high.

The regioselectivity suffered when no leaving group was present on the downstream binding nucleotides **2g**, or when a non-hydrolysable analogue of 2-methylimidazole **2f** was used. Leaving groups on neighboring nucleotides could interact via π -stacking, through the formation of hydrogen bonds between a protonated imidazole group and an adjacent unprotonated one, or through a cation- π interaction between two such groups (Figure S3.5). Although the 2-methylimidazole analogue **2f** can participate in these interactions, it forms 29% of 2'-5' linkages when used as a helper. In fact, there is little difference in the proportion of 2'-5' linkages formed between the analogue **2f** and guanosine monophosphate **2g**. Thus, hydrogen bonding and

stacking interactions between leaving groups do not fully determine the regioselectivity of the reaction.

Downstream binding 5'-activated oligonucleotides can enhance the rate and fidelity of primer extension²⁶, especially in troublesome A and U rich regions. Having determined that activated downstream monomers induce the formation of 3'-5' linkages we tested whether an activated oligonucleotide would similarly affect the regioselectivity of the reaction. In the presence of the 5'-phosphorimidazole trimers, both A and U monomers added without forming any 2'-5' linkages (Figure 3.6b). Furthermore, the highly regioselective addition of 2-AImpC is unaffected



by the presence of the trimer. Thus, downstream binders not only enhance the rate of the copying reaction for A and U rich regions, but also favor the formation of 3'-5' internucleotide linkages.

We next investigated the effect of a fraying base pair at the 3' end of the primer. This situation appears in RNA primer extension when a mismatched nucleotide is added to the primer. The rates of reaction post-mismatch are an order of magnitude slower⁷, which lead us to hypothesize that the regioselectivity of the reaction should decrease concomitantly. Introducing a C/A mismatch at the 3'-end of the primer significantly reduced the regioselectivity of 2AImpG **2a** addition; 2'-5' linkages were formed in 40% yield (Figure S3.6). Replacing the C/A mismatch with a G/U wobble pair improves the regioselectivity of the reaction. The reaction was slower than in the case of a G/C canonical pair, but the regioselectivity was similar (Figure 3.6c). Only 11% of 2'-5' linkages are obtained, 4-times less than in the case of the A/C mismatch.

The formation of 2'-5' linkages in the primer extension reaction raises the question of their heritability over multiple rounds of primer extension. Prakash and Switzer have shown that primer extension on a template

that has four consecutive 2'-5' linkages immediately after the primer annealing site proceeds 17 times slower than on the fully 3'-5' linked template. The reaction produces an equal mixture of 2'-5' and 3'-5' linkages¹⁰⁸. We examined the effect of a single 2'-5' linkage immediately after the primer annealing site (Figure 3.7). Although the reaction is 5.5 times slower than with the 3'-5' linked template, the regioselectivity is high: no detectable 2'-5' linkages were formed. Therefore, a single 2'-5' linkage in the template does not get passed on to the daughter strands,

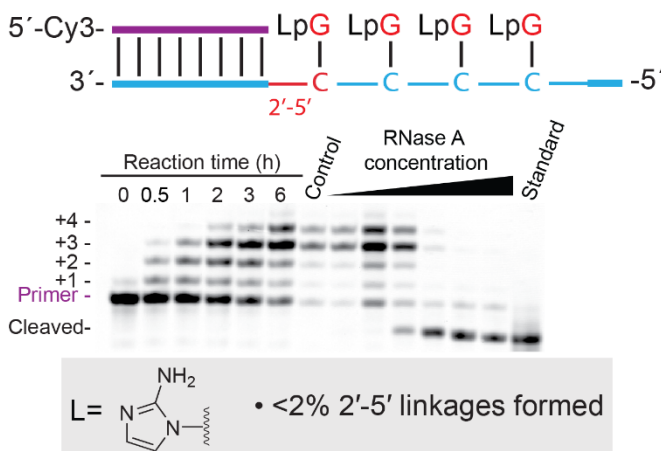


Figure 3.7. A 2'-5' linkage in the template is not heritable. The electrophoretogram shows the results of the primer extension RNase A enzymatic digestion. Control refers to omission of the enzyme; standard refers to the 3'-phosphate primer. The percentage value represents the proportion of 2'-5' linkages formed and was obtained from four independent replicates.

and is lost after a round of chemical RNA primer extension.

Discussion

The mechanism of template directed primer extension with 5'-phosphorimidazolid substrates was assumed to involve the binding of the activated monomer to the template, followed by a nucleophilic attack by either the 2' or 3' ribose hydroxyl on the phosphate group of the monomer. Our group recently suggested that an alternative mechanism is potentially operating⁷⁴. In this

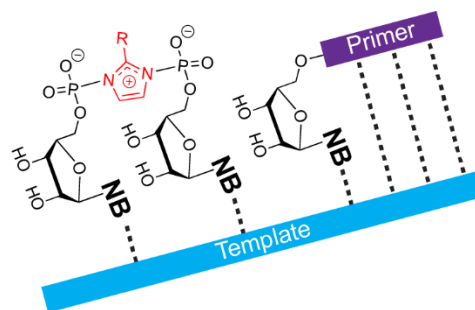


Figure 3.8. The imidazolium bridged dimer presumably binds to the template through Watson-Crick base pairs at two adjacent sites. The imidazolium core is highlighted in red.

mechanism, two monomers react with each other to form an imidazolium-bridged dimer, which then binds to the template, presumably at two adjacent sites (Figure 3.8). Our experiments show that Watson-Crick base pairing at both the reaction site and the site immediately downstream is required for 3'-5' regioselectivity, a result consistent with the binding of an imidazolium-bridged dinucleotide intermediate. Therefore, the poor regioselectivity observed with A:U base pairs was surprising, since a dinucleotide intermediate binding to the template by two Watson-Crick base pairs would be expected to give the same constrained geometry as a C-C or G-G intermediate, thus high regioselectivity. Possibly, the weak stacking interactions between two adjacent U nucleotides in either the imidazolium bridged intermediate or the template lead to a disordered geometry of the reaction center. Conversely, the high regioselectivity of the 2-thiouridine monomer (**5a**) can be explained by its enhanced pre-organization in the 3'-endo conformation¹¹⁷, which leads to a more rigid geometry of the reaction center. The poor regioselectivity and reactivity of the A and U monomers could stem from a common cause: poor conformational constraint of the reaction center. Indeed, we were able to observe a strong correlation between the yield of a reaction and its preference for the formation of 3'-5' linkages (see Figure S3.7 for a detailed discussion). This explanation is also consistent with the loss of reactivity and selectivity

observed with other perturbations such as mismatches. Although not sufficient for a regioselective reaction, the formation and binding of an imidazolium bridged dinucleotide intermediate pre-organizes the reaction center, and increases the proportion of 3'-5' linkages formed. For example, in cases where the formation of a heterodimer intermediate is required (Figure 6a), we see a marked increase in the proportion of 2'-5' linkages formed when one of the monomers cannot participate in the dimer formation reaction. Thus, when 2-AImpA is used together with guanosine monophosphate **2g**, or the non-hydrolyzable analogue **2f**, only the AA bridged dimer can be formed, but not the AG dimer that would correctly pair with the template. Consequently, the regioselectivity of the reaction decreases considerably compared to the cases in which the AG dimer is easily formed.

The nature of the internucleotide linkage formed in non-enzymatic primer extension is determined in part by the relative reactivity of the ribose 2' and 3' hydroxyl groups. Two independent studies looked at the distribution of 3'-5' and 2'-5' internucleotide linkages when 5'-phosphorimidazolid monomers reacted with phosphate-capped mononucleotides, in the absence of a template^{15,118}. In these studies, the 2'-hydroxyl group was more reactive than its 3'-counterpart across all nucleotide and imidazole leaving group combinations. We find that the reactivity of the hydroxyl groups is reversed when the nucleophile is contained in a RNA duplex. Presumably in the primer-template duplex the accessibility towards electrophiles of the 3'-OH group is increased at the expense of the 2'-OH group. Additionally, our current findings show that the leaving group of the activated monomers has little influence on the regioselectivity of the primer extension. This result agrees with a previous study from our group²¹ which shows that the RNA-templated ligation of activated oligonucleotides displays a strong preference for 3'-5' linkages, even when the leaving group was inorganic pyrophosphate. Presumably, the high regioselectivity of templated ligation and primer extension reactions is a consequence of the reactions occurring in a pre-organized extended Watson-Crick duplex.

Prakash and Switzer showed that primer extension on a fully 2'-5' linked template forms an equal mixture of 2'-5' and 3'-5' linkages¹⁰⁸. Here we show that fast reactions on a 3'-5' linked template proceed with high 3'-5' regioselectivity. Additionally, we show that a single 2'-5' linkage in the template does not affect the high 3'-5' regioselectivity of the reaction. These two results suggest that a 2'-5' linkage is not heritable, while for efficient reactions, 3'-5' linkages will carry over to the daughter strands. Furthermore, since the extension across a 2'-5' linkage is 5-fold slower than for 3'-5' linkage, the fully 3'-5' templates will outcompete the linkage heterogeneous ones. These observations are important to the origin of life, because all known processes through which RNA monomers polymerize in the absence of a template form a considerable proportion of 2'-5' linkages^{46,47,119}. Consequently, after multiple rounds of non-enzymatic RNA primer extension, the 3'-5' linkages will be enriched whereas the 2'-5' linkages will be depleted.

The Sutherland group recently demonstrated a mechanism for enriching 3'-5' linked RNA starting from a pool of RNAs containing mixed linkages¹⁹. The 2'-5' linkage is more hydrolytically labile than the 3'-5' linkage¹³. Acetylation of a mixture of 2' and 3' terminal RNA phosphates is selective for the 2' hydroxyl groups and subsequent templated ligation then forms 3'-5' linked RNA¹²⁰. If the hydrolysis and ligation processes are coupled in an energy dissipative cycling process¹²¹, they yield a plausible mechanism for the enrichment of 3'-5' linkages. We show here that simple copying chemistry inherently favors 3'-5' linkages. However, considering the difference in hydrolytic stability between the 3'-5' and 2'-5' phosphodiester linkages, a RNA duplex containing multiple 2'-5' linkages in each strand would be hydrolyzed during the repeated heating-cooling cycles presumably required to enable multiple rounds of primer extension. The recycling chemistry of Sutherland et al. could enable the repair of such hydrolytic damage, while simple copying chemistry would be unable to do so. There is also the possibility that hydrolysis is slow, for example at low Mg^{+2} levels. Additionally, if strand displacement synthesis is possible, the need for temperature cycling is alleviated. Under such conditions, our results suggest that

cycles of replication would lead to the depletion of 2'-5' linkages. Of course, the two approaches are not mutually exclusive and may both have operated to some extent.

Backbone heterogeneity will be generated in the event of a mismatch or when a terminal nucleotide is copied. The ligation of such oligonucleotides will form long RNA strands with interspersed 2'-5' linkages, which could evolve function and have lower duplex melting temperature than the canonically linked isomers. Thus, the balance of the two opposing selection forces i. e. fast replication and ease of strand displacement, will determine the percentage of 2'-5' linkages in a pool of RNA molecules.

Conclusion

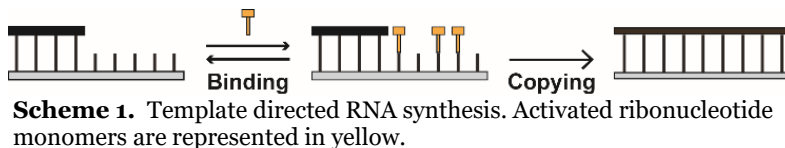
We observe that the rate and yield of chemical RNA primer extension are correlated with its regioselectivity. The reactions that rapidly proceed to completion contain mostly 3'-5' linkages. Both the high reactivity and the regioselectivity of the reaction have a common origin: a properly pre-organized geometry in the reaction center. In all instances of fast and regioselective reactions the RNA monomers can form Watson-Crick base pairs with the template at the extension site and the adjacent downstream position. In addition, the downstream binding nucleotide must have a leaving group on its 5' phosphate. Our results are consistent with a reaction mechanism involving a 5'-5' imidazolium bridged intermediate formed between the incoming monomer and a downstream monomer or an oligonucleotide. Reaction conditions that would perturb the binding of the intermediate to the template lead to a simultaneous decrease in reaction rate and regioselectivity.

4. A Fluorescent G-quadruplex Sensor for Chemical RNA Copying

Reproduced with permission from Giurgiu, C., Wright, T. H., O’Flaherty, D. K. & Szostak, J. W. A Fluorescent G-Quadruplex Sensor for Chemical RNA Copying. *Angew. Chemie Int. Ed.* **57**, 9844–9848 (2018). Copyright 2018 Wiley.

The RNA world hypothesis is a central concept in origin of life research. First proposed by Orgel, Crick and Woese^{97–99} and then elegantly enunciated by Gilbert¹⁰⁰, the hypothesis states that early life forms relied on RNA for both catalysis and the transmission of genetic information. In this context, the primitive RNA genome was hypothesized to be replicated by an RNA enzyme, a replicase. However, the evolution of an effective RNA replicase is thought to have required an initial phase of chemical RNA replication. Chemical RNA replication starts with the binding of activated ribonucleotides to an RNA template through Watson-Crick base pairing. The ribonucleotide monomers react with each other to form a double-stranded RNA duplex; subsequent amplification may then require separation of the strands, so that they can then act as templates for further copying rounds. Multiple rounds of copying would lead to an exponential increase in the population of RNAs, and selective pressures for more efficient replication would eventually give rise to catalytic RNAs. Although significant progress in reconstructing the initial steps of such a process has been made over the last 50 years⁶⁹, several problems persist, including the poor rate and fidelity of non-enzymatic copying, the difficulty of achieving multiple rounds of replication, and incompatibility with prebiotically plausible cellular membranes¹⁰³.

A useful laboratory model for the study of templated RNA synthesis is non-enzymatic



primer extension (Figure 4.1). The reaction is performed by adding activated ribonucleotide

monomers to a pre-formed primer-template duplex, after which the product distribution is analysed by various methods. Conventional methods used to measure the rate of and to characterize the products of templated RNA synthesis include high performance liquid chromatography (HPLC), polyacryl-amide gel electrophoresis (PAGE), and mass spectrometry, all of which are cumbersome and low throughput. The current methods of analysis are often denaturing and do not provide information about the folding ability of the RNA strand that is generated in the copying process. The folding ability of the RNA strands is crucial if RNA replication is hypothesized to have given rise to RNAs with structure and function. As such, a quick and facile assay for RNA copying that also informs on the structural integrity of the newly synthesized RNA strand is needed. Such an assay would accelerate the discovery of catalysts and conditions for the template directed chemical synthesis of RNA. Here we present a high-throughput fluorescence assay which enables the screening of multiple reaction conditions in a fraction of the time required by conventional methods. The assay relies on the formation of a correctly folded RNA G-quadruplex, which is, in turn, conditional on the success of the primer extension reaction.

We set out to adapt a previously reported assay for enzymatic primer extension developed by Kankia and colleagues.^{122,123} The assay is based on a DNA primer that contains the fluorescent adenine analog 2-aminopurine.¹²⁴ 2-aminopurine fluoresces when exposed to the solvent, but is quenched when incorporated in a double-stranded duplex. Starting with a primer-template duplex, the addition of two guanosines at the 3' end of the primer by a polymerase enzyme in the presence of K^+ removes the primer from the template to form a parallel G-quadruplex structure. In this structure, the 2-aminopurine nucleobase is exposed to the solvent, resulting in a 90-fold increase in fluorescence.¹²⁵ We hypothesized that changing the template and primer from DNA to RNA would not affect the formation of the secondary structure since RNA is known to form stable parallel G-quadruplexes.¹²⁶ Furthermore, instead of using a polymerase and nucleoside

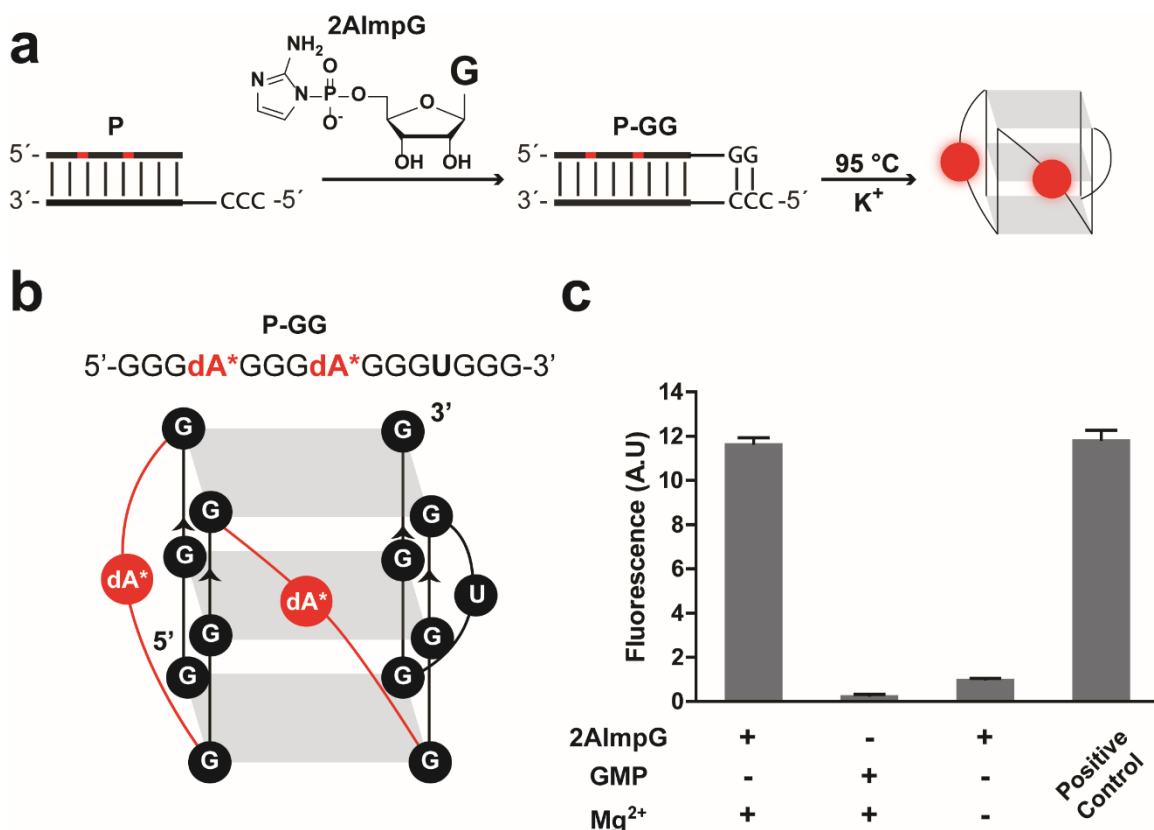


Figure 4.2. (a) Non-enzymatic RNA primer extension using 2-aminoimidazole activated guanosine phosphoroimidazole (2AImpG) followed by a heating step liberates a fluorescent G-quadruplex. The red squares represent the 2-aminopurine residues (b) The P-GG oligonucleotide forms a G-quadruplex that contains two solvent exposed 2'-deoxyribose 2-aminopurine residues (highlighted in red and labelled dA*). (c) Incubation of P with 2AImpG for 3 hours results in similar levels of fluorescence to identically treated P-GG (positive control). In the second column 2AImpG was replaced by an equal amount of guanosine monophosphate (GMP). In the third column 20 mM EDTA was added to ensure that trace Mg²⁺ is chelated. Data are reported as the mean \pm the standard error of the mean from triplicate experiments.

triphosphates to extend the primers, we used 5'-phosphorimidazolide guanosine monomers (Figure 4.2a). Our group has previously demonstrated that non-enzymatic primer extension can be correlated with an increase in fluorescence of the malachite green aptamer¹²⁷. Here we opted to use mononucleotides activated with 2-aminoimidazole (2-AI), since the reaction proceeds quickly and in high yield⁸⁷, and a potentially prebiotic synthetic path to 2AI has been reported recently⁵¹.

We began by synthesizing the RNA sequence corresponding to the fully extended primer with two 2-aminopurine residues (P-GG), which is expected to form a parallel G-quadruplex (Figure 4.2b). We reasoned that having two 2-aminopurine residues would improve the signal intensity and thus

the sensitivity of the assay. In the presence of K^+ , the oligonucleotide forms a G-quadruplex, as shown by its characteristic circular dichroism spectrum¹²⁵, as well as UV-melting data (Figure S4.1). In addition, we observed a significant shift in the melting temperature of the duplex formed by **P-GG** and its complementary strand in the presence of KCl, which is not observed in the case of **P** (Figure S4.1), as previously reported for the DNA analogues¹²⁵. The different behaviour of the two duplexes suggested that only **P-GG** might form a fluorescent G-quadruplex following denaturation, and that conversion of **P** to **P-GG** could therefore be monitored by a fluorescence assay. To test this idea, pre-annealed **P** duplex was incubated with guanosine 2-aminophosphorimidazolid (2AImpG), in the presence of 50 mM Mg^{2+} , in a model primer extension reaction. The reaction was then quenched by adding EDTA, and the mixture was heated in the presence of KCl. The measured fluorescence was identical to that of the identically treated **P-GG** duplex (Figure 4.2c). In contrast, when guanosine monophosphate (GMP) was used instead of 2AImpG, we observed a 40-fold lower fluorescence intensity relative to **P-GG**. Most previous work on non-enzymatic primer extension has been carried out in the presence of > 0.1 M Mg^{2+} concentration¹⁰³. At lower concentrations, the reaction rate decreases considerably. When no divalent cations were present during the reaction, the fluorescence was similar to the GMP reaction, as expected. Taken together, these results show that the increase in fluorescence is consistent with primer extension.

To validate the assay, we synthesized a modified **P** primer that contains an orthogonal fluorescent label (6-FAM) at its 5' end. The modified primer enabled us to simultaneously monitor a copying reaction by 2-aminopurine fluorescence and PAGE (Figure 4.3). We observed an increase in 2AP fluorescence over the time course of the reaction which was strongly correlated with the extension of the primer by at least 2 nucleotides (Figure S4.2a). To show that the +1-extended primer does not contribute to the fluorescence, we synthesized the **P-G** primer and showed that it fluoresces similarly to the unextended primer **P** (Figure S4.2b). We confirmed that the assay is suitable for

high-throughput screening given its Z' -factor of 0.90¹²⁸. The Z' -factor is a statistical parameter developed to measure the reproducibility of an assay for high-throughput screening by determining the separation between positive and negative controls. A Z' of 1 is ideal, while a value below 0.5 is considered inadequate for high-throughput screening.

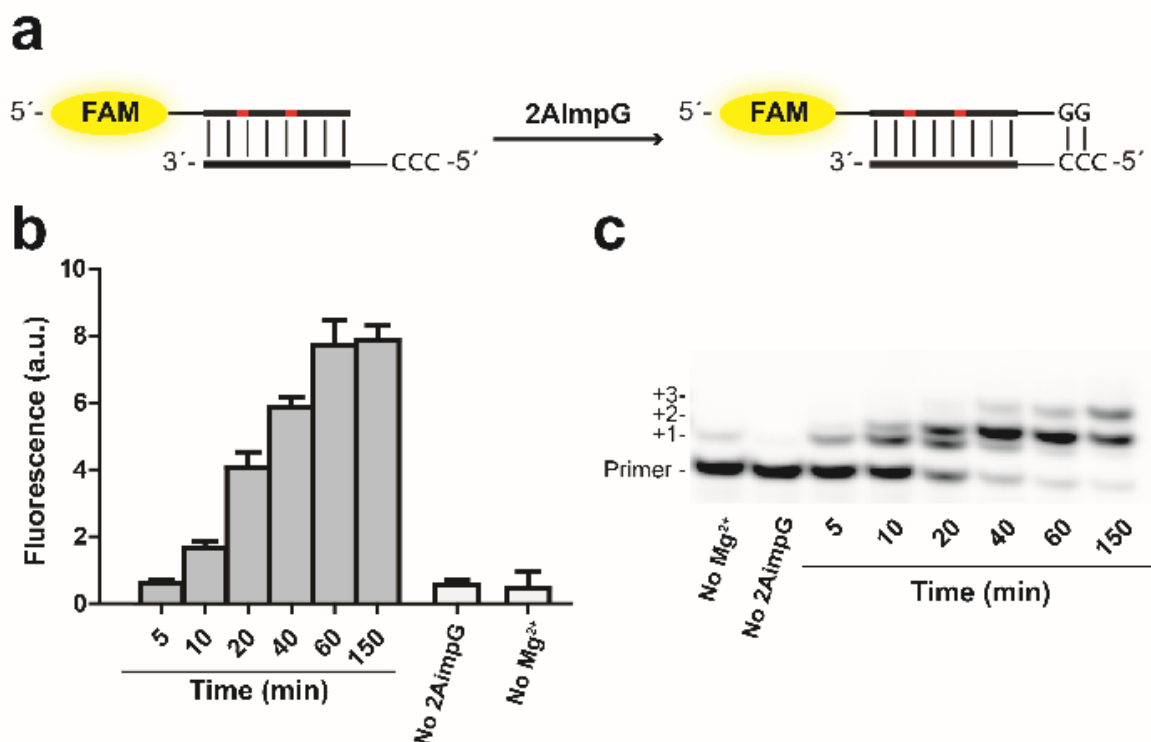


Figure 4.3. Results from the fluorescence assay for primer extension are consistent with the standard PAGE assay for primer extension (a) A primer extension reaction was carried out with a dually labelled primer, with two internal 2-aminopurine residues and an additional 5'-fluorescein modification (FAM). (b) 2-aminopurine fluorescence was measured at different times for the primer extension reaction. Data points are reported as the mean \pm s.e.m. from triplicate experiments. (c) Electrophoretogram showing the primer extension of the FAM labelled primer. The primer extension reaction was carried out at pH 8.0 in the presence of 50 mM Mg²⁺, using 10 mM of 2AImpG.

Non-enzymatic RNA synthesis of homopolymeric G and C stretches has been shown to be relatively quick and high yielding¹²⁹. In contrast, copying mixed sequences or homopolymeric A and U stretches is slow and often results in a mixture of truncated products and unreacted primer. The poor efficiency of mixed sequence copying precludes sequence general RNA replication and the generation of catalytic RNAs. To examine the copying of mixed sequences we removed the 3'-UG sequence of the primer **P**. When the resulting oligonucleotide is used as a primer in a copying reaction, extension by three nucleotides is required to generate a fluorescence signal. We then

examined primer extension on templates designed to give products containing both purine and pyrimidine nucleotides (CGGG and UGGG). In comparison to the homopolymeric G extension, which is complete in 1 hour, the synthesis of the CGG moiety plateaus after 3 hours, reaching 80% of its maximum fluorescence level (Figure 4.4a). In contrast, the UGGG extension, although as efficient as the CGGG extension, requires 24 hours to plateau (Figure 4.4b). When the activated pyrimidine nucleotides are replaced by the respective monophosphates, the reaction rate and efficiency is drastically reduced. However, a significant increase in fluorescence is

observed when 2AImpG and UMP are incubated with the primer-template complex for extended periods of time, presumably due to the formation of a mismatched product through G-A mispairing or through a strand slipping process¹³⁰, in which the template forms a mononucleotide loop, excluding the adenosine residue from the templating region.

Divalent metal cations are known to catalyze the chemical copying of RNA, but the mechanism of catalysis remains incompletely understood. A potential mechanism by which divalent cations increase the rate of primer extension is by presenting a coordinated hydroxide to

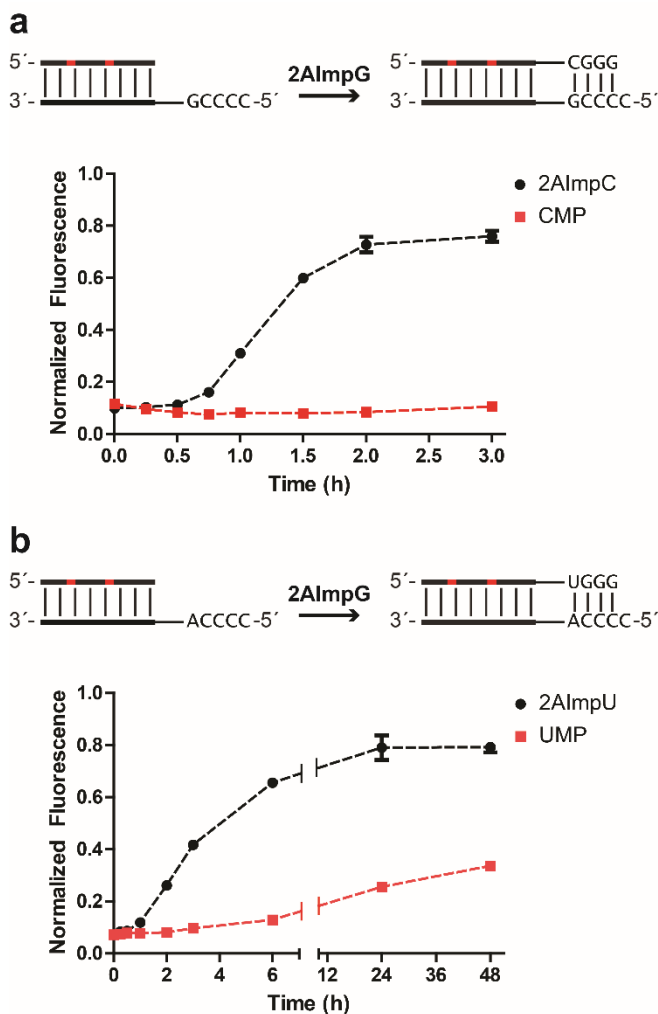


Figure 4.4. Monitoring copying of mixed sequences. (a) Fluorescence time course of the copying reaction on a GCCCC template. Reactions were initiated by adding 2AImpG and 2AImpC (black) or CMP (red) (b) Fluorescence time course of the copying reaction on a ACCCC template. The reactions were initiated by adding 2AImpG and 2AImpU (black) or UMP (red). All data points are reported as the mean \pm s.e.m. from triplicate experiments. Fluorescence was normalized with respect to the identically treated P-GG duplex.

the diol, or coordinating it, or either of its hydroxyl groups directly, to facilitate diol deprotonation. The resulting alkoxide, a strong nucleophile, would attack the activated phosphate and form a phosphodiester bond. If that were the case, then the reaction rate would depend on the ability of the metal to act as a Lewis acid, provided that nucleophilic attack is the rate-determining step. We therefore looked at how different concentrations of two different cations (Mg^{2+} and Ca^{2+}) and 2AImpG affect the rate of primer extension. Using our assay, we surveyed 84 different reaction conditions in a single experiment. At all metal concentrations,

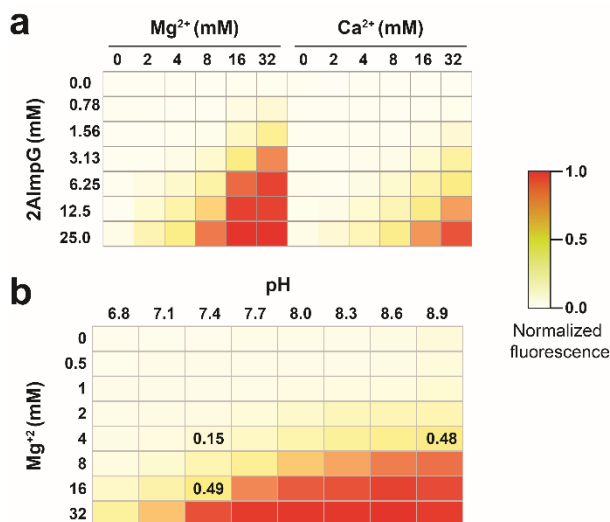


Figure 4.5. (a) Mg^{2+} outperforms Ca^{2+} as a catalyst of non-enzymatic primer extension. Reactions were incubated for 135 minutes at room temperature. Fluorescence values were normalized with respect to a 25 mM 2AImpG, **P-GG** replicate for each of the different metal concentrations. (b) Non-enzymatic RNA primer extension works best at higher pH values and higher Mg^{2+} concentrations. Values were normalized with respect to the control incubated at each of the Mg^{2+} concentrations. The numerical values of the normalized fluorescence represent the conditions referred to in the main text

Mg^{2+} is a better catalyst than Ca^{2+} for the primer extension reaction, and more **P-GG** is formed (Figure 4.5a). The difference in pK_a values of the hexaaquaions, $[\text{M}(\text{OH}_2)_6]^{2+}$, of the two metals¹³¹ ($\text{pK}_a = 11.2$ for Mg^{2+} and 12.7 for Ca^{2+}) show that Mg^{2+} is more efficient in promoting the deprotonation of a coordinated water molecule relative to Ca^{2+} . If a similar mechanism is operating in the primer extension reaction, higher pH values should relieve the metal dependence of the reaction. To test this hypothesis, we conducted a two-dimensional screen of pH values and different Mg^{2+} concentrations. The reaction fails at low pH values (≤ 7.1), or at low Mg^{2+} concentrations (≤ 4 mM). However, at intermediate metal concentrations the yield of the reaction is rescued by increasing the value of the pH. For example, increasing the pH from 7.4 to 8.9 at a fixed concentration of 4 mM Mg^{2+} generates similar amount of **P-GG** product as increasing the concentration of Mg^{2+} from 4 mM to 16 mM at a fixed pH of 7.4 (Figure 4.5b). This result supports

the hypothesis that the divalent cations help in the deprotonation of the ribose secondary hydroxyl groups. Incubating the primer-template duplex at the highest pH value and Mg^{2+} concentration, in the absence of 2AImpG, does not lead to an increase in fluorescence (Figure S4.3). This result confirms that values measured in the two-dimensional screen correspond to successful primer extension and not to any other processes that might increase the fluorescence, such as hydrolysis of the primer-template duplex. Further investigations into the role of divalent metals in the primer extension reactions are currently being carried out in our laboratory.

The templated chemical copying of RNA is one of many hypotheses that aims to explain the transition from inanimate, chemical matter to self-replicating entities capable of Darwinian evolution. Although encouraging early results have sustained research in the area for over 50 years, considerable advances are still required to achieve continued cycles of nonenzymatic RNA replication. To address the remaining hurdles, we developed a fluorescence-based assay for non-enzymatic copying of RNA. Under conditions in which RNA copying proceeds quickly and in good yield, a large gain in signal is observed that is strongly correlated with the formation of a fluorescent RNA G-quadruplex. We have shown that the method can be used to measure the kinetics of the copying reaction. In addition, the assay can be used in high-throughput screens of the parameters affecting the copying process.

An ideal fluorescence assay for chemical RNA copying would be able to measure the extent of copying for any given template sequence in real-time. Our assay relies on the formation of a G-quadruplex and is therefore limited to generating guanosine-rich RNA strands. Additionally, after the copying is concluded a heating step is required to liberate the G-quadruplex from its complementary strand. However, a simple heating step is more amenable to highly parallel screening than traditional PAGE assays or mass spectrometry, which typically require lengthy sample preparations. However, mass spectrometry can identify mismatched copying, as well as the identity of the mismatched nucleotide, while PAGE, despite being laborious, is sensitive and

versatile, and remains the gold standard for quantification. Our newly developed fluorescence assay complements mass spectrometry and PAGE with its much higher throughput. The parallel nature of the assay enabled us to observe the combinatorial effects of multiple variables such as the concentration of activated monomer, the pH of the reaction, as well as the concentration and the nature of the divalent metal cations on the efficiency of the copying reaction. In addition, the assay is not limited to RNA, and could potentially work for any nucleic acid that can form G-quadruplexes.

Prebiotic synthetic pathways to biomolecules have benefited greatly in recent years from adopting a systems chemistry approach¹³². The chemical routes to the different classes of biomolecules (nucleic acids, lipids, amino acids) have been evaluated by taking into consideration the possible interactions between the starting materials, the types of chemistries and the nature of the reagents used. As a result, more robust pathways, common to all three classes of biomolecules, have been uncovered⁹⁵. The same approach has not yet been used to explore superior routes to chemical RNA replication. The non-enzymatic assay developed herein should facilitate a systems approach to the long-standing problem of chemical RNA replication.

5. Prebiotically Plausible ‘Patching’ of RNA Backbone Cleavage Through a 3′-5′ Pyrophosphate Linkage

Reproduced with permission from Wright, T. H. *et al.* Prebiotically Plausible “Patching” of RNA Backbone Cleavage through a 3′–5′ Pyrophosphate Linkage. *J. Am. Chem. Soc.* **141**, (2019). Copyright 2019 American Chemical Society.

Introduction

The RNA world hypothesis proposes an early stage in the evolution of life in which RNA was responsible for both catalysis and genetic inheritance^{98,100,133,134}. In existing biology, ribonucleotide monomers are joined to form polynucleotides in a template-directed process by the enzyme-catalyzed reaction of 3′-hydroxyl groups with nucleoside 5′-triphosphates, forming a 3′-5′ phosphodiester-linked backbone. Before the evolution of protein enzymes, a replicase made of RNA may have catalyzed RNA synthesis, a possibility explored by the *in vitro* selection and evolution of ribozymes with ligase¹³⁵ and polymerase^{28,101,136} activity. A more challenging problem is how these ribozymes could have emerged from a non-enzymatic process of chemical RNA replication. Chemical RNA replication starts with the binding of energy-rich activated ribonucleotides to an RNA template through Watson-Crick base pairing. The ribonucleotide monomers react with each other to form phosphodiester bonds, leading to a double-stranded RNA duplex; subsequent amplification requires separation of the strands, so that they can act as templates for further cycles of copying. Mechanistic studies which revealed the critical role of 5′-5′ imidazolium-bridged dinucleotides as the active species in primer extension⁷⁴, and the discovery of 2-aminoimidazole as a superior 5′-phosphate activating group^{51,87}, have recently enabled the copying of short mixed sequence RNA templates within vesicle models of early cells¹³⁷. Despite this progress, the chemical copying of RNA is not currently able to produce RNA products of the length and complexity necessary to sustain ribozyme evolution^{65,103}.

Any model proposing RNA or RNA-like polymers as early carriers of genetic information must also account for cleavage of the phosphodiester backbone¹³⁸, which competes with copying chemistry. The cleavage of RNA strands results in a mixture of 2' and 3' phosphate terminated strands via an initial transesterification reaction followed by hydrolysis of a cyclic phosphate intermediate (Figure 5.1a). Cleavage of RNA strands is greatly accelerated by divalent metal cations, such as magnesium, that are also required for the copying chemistry to proceed at a reasonable rate. This degradation pathway constrains plausible rates of non-enzymatic RNA synthesis and

would result in a net loss of material, and thus information, from a primordial genetic system if there were no mechanism for recycling of the phosphate-terminated strands¹³⁸. 2',3'-cyclic phosphate terminated strands can participate in templated ligation reactions^{139,140}, albeit with slow rates and poor regiospecificity, largely yielding 2'-5' linked products. The Sutherland lab¹⁹ has demonstrated a prebiotically plausible cycle of backbone repair that converts 2'-5' to 3'-5' phosphodiester linkages via iterative cycles of strand cleavage and repair, in which a 3' phosphate plays a central role as a substrate for ligation chemistry. The identification of complementary pathways for RNA salvage or repair following strand cleavage^{19,120}, or the discovery of additional

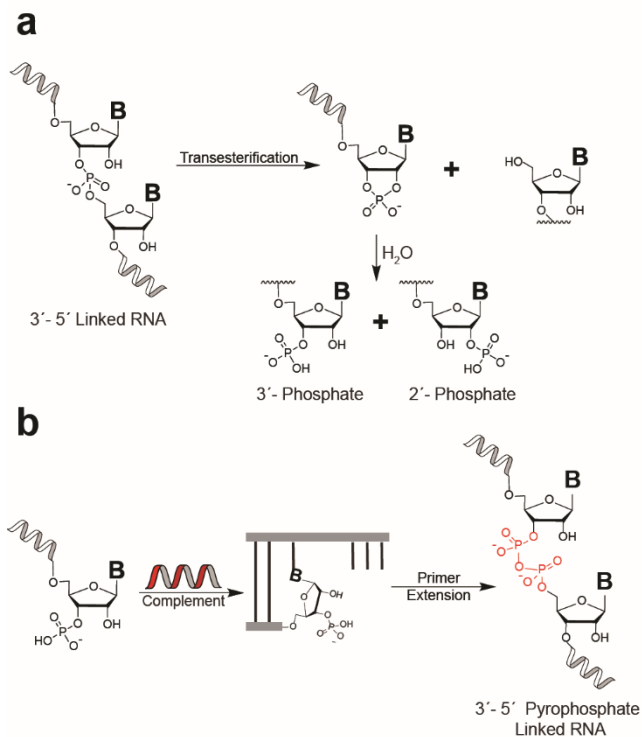


Figure 5.1 RNA strand cleavage and possible recovery of genetic information via pyrophosphate bond formation. (a) The cleavage of 3'-5' phosphodiester-linked RNA leads to a mixture of 2' and 3' monophosphate terminated strands, via a cyclic phosphate intermediate. (b) Non-enzymatic primer extension starting from a 3'-monophosphate leads to 3'-5' pyrophosphate-linked RNA, a possible mechanism for salvage of truncated RNA strands.

functions for phosphate-terminated strands¹⁴¹, could further strengthen the case for RNA as the earliest carrier of genetic information.

Consideration of potentially prebiotic nucleoside phosphorylation reactions provides further motivation for studying terminally phosphorylated RNA. Phosphorylation of nucleosides typically gives varying mixtures of 5'-, 3'- and 2'-phosphorylated products as well as the 2',3'-cyclic phosphate. Although regioselective 5'-phosphorylation has been demonstrated, specific conditions such as the addition of borate minerals¹⁴² or the use of gas-phase reactions are required¹⁴³. The diversity of pathways to nucleoside phosphorylation suggests that a mixture of phosphorylation states may have been likely at the monomer level. Heterogeneity of phosphorylation state at the 2' and 3' position of nucleotide monomers could therefore be incorporated into RNA strands via templated copying or non-templated reactions, in addition to the hydrolytic pathway.

Considering this implied presence of terminal phosphates in any 'RNA world' scenario and bearing in mind the greater nucleophilicity of phosphate relative to hydroxyl groups, we became interested in exploring whether phosphate-terminated RNA primers could participate in RNA copying chemistry. We hypothesized that the reaction of terminally phosphorylated primers with incoming nucleotides could lead to a mechanism for preservation of genetic information via formation of a pyrophosphate linkage (Figure 5.1b). Early work from Schwartz and Orgel demonstrated that 3'-phosphorylated, 2'-deoxy, imidazole-activated nucleotides efficiently polymerize on DNA templates to form 3'-5' pyrophosphate-linked oligomers¹⁴⁴. In turn, the 3'-5' pyrophosphate-linked oligomers could be used as templates for further synthesis, again employing activated 3'-phosphate, 2'-deoxy monomers, indicating that a pyrophosphate-linked backbone may not preclude cycles of non-enzymatic replication¹⁴⁵. Unfortunately, these early results were never extended to RNA-like systems.

Here, we report the results of our initial investigations into the behavior of phosphate-terminated RNA primers in non-enzymatic primer extension. We confirm that terminally phosphorylated RNA primers participate in primer extension reactions, leading to RNA polymers containing a pyrophosphate linkage. We have undertaken a thorough study of the stability of a single pyrophosphate linkage embedded within an RNA strand, which revealed pronounced lability towards cleavage reactions in the presence of magnesium ions. However, similarly to ‘native’ RNA, the pyrophosphate linkage is protected from strand cleavage in the context of a duplex and in the presence of magnesium chelators. These observations enabled us to survey the kinetic parameters of primer extension from phosphate-terminated primers and following incorporation of a pyrophosphate bond. We further demonstrate that pyrophosphate-linked RNA can function as a template to direct the polymerization of canonical ribonucleotides, pointing towards a possible non-enzymatic salvage pathway for primordial RNA.

Results

We first synthesized RNA primers terminated with either a 2' or 3' monophosphate, using a solid phase approach (SI Figure 1a). Using strong-anion exchange chromatography¹⁹, we confirmed the regioisomeric purity of the individual primers; no contamination with the alternative terminal

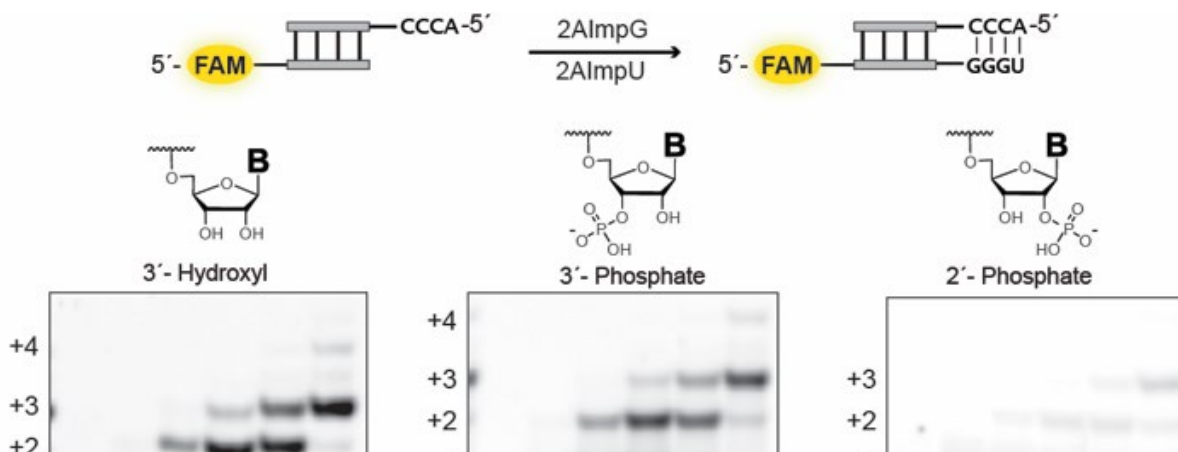


Figure 5.2. Efficient non-enzymatic copying of RNA commences from a primer with a terminal 3'-monophosphate. The time course of primer extension with guanosine 2-aminophosphorimidazolidine (2AImpG) and uridine-2-aminophosphorimidazolidine (2AImp-U) monomers was monitored using polyacrylamide gel electrophoresis (PAGE). All reactions were performed at pH 8.0, 200 mM HEPES, 50 mM Mg^{2+} and 200 mM citrate, with 10 mM each of guanosine 2-aminophosphorimidazolidine and uridine-2-aminophosphorimidazolidine.

phosphate was observed in either case (Figure S5.1b). We therefore tested the 2' and 3' phosphorylated primers in a primer-extension assay (Figure 5.2), employing a templating region with the sequence 3'-CCCA-5', which can be extended efficiently by addition of activated guanosine and uridine ribonucleotide monomers. Citrate-chelated magnesium was used in these assays as it enables RNA copying chemistry to proceed within vesicles composed of fatty acids⁵⁷, which are the most likely candidates for protocell membranes but are destabilized by Mg^{2+} at low millimolar concentrations¹⁴⁶. The two phosphorylated primers, and a non-phosphorylated control, were annealed separately with the template, incubated with 50 mM Mg^{2+} , 200 mM citrate and 10 mM of both guanosine 2-aminophosphorimidazolid (2AImpG) and uridine-2-aminophosphorimidazolid (2AImpU) and monitored over the course of 24 hours. All three primer-template duplexes were extended, although extension in the case of the 2'-phosphate terminated primer was very poor. Surprisingly, the 3' phosphate-terminated primer gave comparable extension to that obtained in the 3'-hydroxyl system, with 81% of +3 products after 24 hours (cf. 87% of +3 product for 3'-hydroxyl).

We hypothesized that extension of the phosphate-terminated primers could be due to formation of a pyrophosphate linkage by reaction of the 2' or 3' terminal phosphate with the activated 5'-phosphate of the incoming monomer. However, we aimed to exclude an alternate possibility in which the reaction proceeds through the non-phosphorylated hydroxyl, as chemical primer extension can proceed via either the 2' or 3' hydroxyl group¹⁸. To confirm the presence of

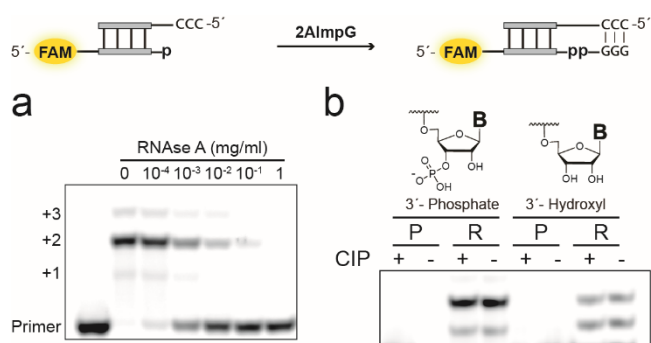


Figure 5.3. Confirmation of 3'-5' pyrophosphate linkage by enzymatic digestion assays. (a) PAGE analysis of RNase A treatment. RNase A treatment of a primer extension reaction commenced from a primer bearing a terminal 3'-monophosphate leads to regeneration of the primer. (b) PAGE analysis of calf intestinal phosphatase (CIP) treatment of primer extension reaction mixtures. 'P' refers to primer and 'R' refers to reaction mixture derived from primer extension. Lanes 1 and 2: 3'-phosphorylated primer standard. Lanes 3 and 4: Reaction mixture from extension of the 3'-phosphorylated primer with (Lane 3) and without (Lane 4) CIP added. Lanes 5 and 6: Non-phosphorylated RNA primer. Lanes 7 and 8: Reaction mixture from extension of the same RNA primer.

a pyrophosphate linkage in the reaction products derived from the 3'-phosphorylated primer, we made use of the fact that RNase A is known to cleave pyrophosphate bonds¹⁴⁷ while leaving 2'-5' phosphodiester linkages intact¹⁸ (Figure 5.3a). The primer utilized was designed so that it is cleaved by RNase A at a single site, after the first bond-forming step of the primer extension reaction. Thus, for the 3'-terminal phosphate reaction, we expected cleavage if the addition of the first monomer occurs through the 3' phosphate and no cleavage if the reaction instead proceeded through the 2' hydroxyl. Treating a primer extension reaction mixture with increasing concentrations of RNase A resulted in cleavage of the +1 to +3 bands and the appearance of a band with the same gel mobility as the original primer. As RNase A requires a free 2'-OH for cleavage and should lead to regeneration of the 3'-monophosphate, this result rules out formation of a 2'-5'-phosphodiester bond during primer extension and strongly suggests the presence of a 3'-5'-pyrophosphate linkage.

If primer extension proceeded from the free terminal hydroxyl and not from the terminal phosphate group, a phosphate monoester internal to the RNA chain would remain in the +1 and higher extension products. To test this hypothesis, we treated the reaction mixtures with calf intestinal phosphatase (CIP), which has been shown to cleave internal phosphate monoesters¹⁴⁸ (Figure 5.3b). For reaction mixtures derived from the 3'-phosphorylated primer, gel bands corresponding to extended products were not affected by CIP treatment. Taken together, these results obtained from enzymatic digestion support formation of a pyrophosphate linkage during primer extension initiated from a terminal phosphate and rule out the alternative possibility of a 2'-5' phosphodiester-linked structure.

To further investigate the properties of pyrophosphate-linked RNA, we required a method to generate pure, single-stranded RNA containing site-specific pyrophosphate linkages, for use both as primers and as templates in downstream studies. Notably, there exist no synthetic or enzymatic routes to produce RNA containing defined 3'-5' pyrophosphate linkages. We thus devised a

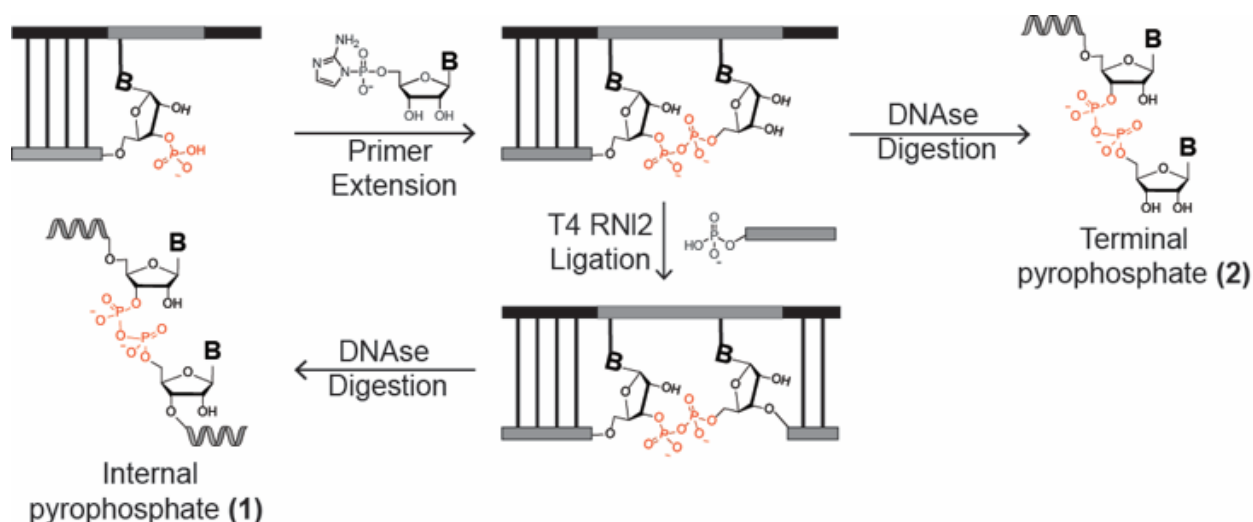


Figure 5.4. Synthetic strategy to access RNAs containing a single, defined pyrophosphate linkage. An RNA primer containing a terminal 3'-phosphate monoester is annealed to a chimeric DNA/RNA template. Primer extension with activated guanosine monomers generates a 3'-5' pyrophosphate linkage. Digestion of the DNA residues of the template with Turbo DNase generates a ssRNA containing a terminal pyrophosphate linkage. Alternatively, T4 RNA ligase 2 can be used to ligate an RNA oligonucleotide downstream in excellent yields (>95%). The same DNase treatment can then be used to generate ssRNA containing an internal pyrophosphate linkage. Grey: RNA, Black: DNA. See the Supporting Information for detailed experimental procedures.

strategy that enabled us to generate pure (>95% by PAGE analysis) RNA strands containing a single pyrophosphate linkage (Figure 5.4). Our approach begins with a primer extension reaction using a 3'-phosphate terminated primer and a mixed sequence template which allows us to control the number and identity of added nucleotides simply by adjusting the reaction times or adding/removing nucleotide phosphorimidazolides from the reaction mixture. Critically, we use a DNA-RNA hybrid template in which the region to be copied is RNA but the remainder of the template is DNA, to allow for DNase digestion of the template which facilitates recovery of the modified primer. Incubation of the primer template duplexes with 2-aminoimidazole activated monomers leads to robust conversion of the primer to extended products in which the +1 nucleotide is connected to the RNA primer by a pyrophosphate linkage. Following desalting to remove unreacted monomer, a 5'-phosphorylated ligator RNA oligonucleotide is annealed and ligated to the pyrophosphate-containing primer by T4 RNI2 ligase. The ligation step combined with design of the primer and template sequence allows for the generation of essentially any RNA sequence in the final product. After ligation, treatment of the reaction mixture with DNase

followed by gel purification affords reasonable yields (typically 20-30%, based on known input of primer) of high purity (>95% by PAGE analysis), single stranded ‘pyrophosphate-containing RNA’. To produce ssRNA species containing a terminal pyrophosphate (Figure 4), a template containing only a single binding site for the activated G-G dinucleotide intermediate is employed using a 5′-CC-3′ templating region and the ligation step is omitted. The +1 primer extension product, bearing a terminal pyrophosphate linkage, can be isolated directly following DNase digestion. Although a recent report detailed the synthesis of pyrophosphate-linked DNA via a solid-phase synthesis approach¹⁴⁹, our method represents the first route to RNA containing defined pyrophosphate bonds.

We were interested in exploring the relative stability of the pyrophosphate linkage within RNA, to determine whether it could support the chemical replication of genetic information, and whether it could in principle enable ribozyme function. We prepared two single-stranded RNAs containing either an internal (AAGGGAAGAAGC-pp-GGGCUAGCAUGAC **1**) or terminal pyrophosphate linkage (AAGGGAAGAAGC-pp-G **2**), using the strategies outlined above. The pyrophosphate-RNAs were then incubated at 22 °C in a pH 8.0 solution (conditions typical for primer extension reactions) in the presence or absence of magnesium, which was included as either the free cation or in citrate-chelated form (Table 5.1).

Table 5.1. Hydrolytic half-life of a single 3′-5′ pyrophosphate linkage within RNA*

	ssRNA	dsRNA	Stabilization
<i>Internal pyrophosphate</i>			
Mg ²⁺	0.12 ± 0.01	690 ± 20	5500 ± 200
Mg-citrate	3.7 ± 0.1		
Citrate	30 ± 1		
EDTA	390 ± 20		
<i>Terminal pyrophosphate</i>			
Mg ²⁺	0.20 ± 0.01	26 ± 1	130 ± 10
Mg-citrate	3.7 ± 0.2		

*The half-life values are presented with the standard error of the mean and reported in hours.

We included citrate as it protects both RNA and fatty acid vesicles from magnesium-induced degradation, while still allowing RNA copying reactions to proceed, albeit at a reduced rate⁵⁷. In the presence of 1 mM EDTA, which should chelate trace divalent cations, the pyrophosphate linkage was relatively stable, with a half-life of 16 days. In contrast, incubation with free magnesium (50 mM) led to rapid cleavage of the pyrophosphate bond, with a half-life measured to be on the order of minutes for both internal and terminal pyrophosphate linkages. Even at low millimolar concentrations of magnesium ions, the half-life of the pyrophosphate bond was only extended slightly (Table S5.1). The product of strand cleavage was determined by LC-MS (SI Figure 5.2a) and CIP digestion (Figure S5.2b) to be the 2'-3'-linked cyclic phosphate, consistent with a mechanism in which the α -phosphorus atom is subject to nucleophilic attack by the adjacent 2'-hydroxyl group. Incubating the pyrophosphate-RNA with magnesium citrate afforded modest protection, extending the half-life of the linkage from minutes to hours. Notably, the reduction in the rate of cleavage afforded by citrate chelation is much larger than the effect of chelation on the rate of the copying chemistry⁵⁷. Duplex formation has been shown to protect 3'-5' linked phosphodiester bonds in RNA from strand cleavage, while promoting cleavage in the case of 2'-5' bonds^{13,21}. This observation has been used to explain the ultimate 'selection' of the canonical 3'-5' linkage in extant RNA. We therefore examined the effect of duplex formation on the rate of cleavage of the pyrophosphate linkage, to determine whether pyrophosphate bonds, once formed, could be retained in RNA duplex structures and ultimately serve as components of templates for further copying cycles. For the RNA 25mer **1** containing an internal pyrophosphate, addition of the complementary strand 'buries' the pyrophosphate bond in an extensive duplex region. Incubating this duplex with free Mg^{2+} revealed a significant stabilization effect, with the half-life of the pyrophosphate bond increased from minutes to 28 days; a stabilization factor of 5500. Protection in the case of the terminal pyrophosphate was far more modest, consistent with reduced helicity at the primer terminus.

Together, our degradation studies constrain the conditions under which the pyrophosphate linkage could support the storage and propagation of early genetic information. Pyrophosphate retention is possible within a duplex but is strongly disfavored in single-stranded RNA in the presence of Mg^{2+} . The fate of a pyrophosphate bond at the terminal end of a duplex will depend on the relative rates of downstream extension or ligation reactions and strand cleavage and is thus expected to also depend on the RNA sequence (see Discussion below).

A critical factor determining the proportion of pyrophosphate-containing RNAs within a prebiotic population of polynucleotides would be the relative rates of reaction of phosphate and hydroxyl terminated primers with activated nucleotides. We were therefore interested in comparing the kinetics of primer extension for the initial reaction step. We measured the rates of primer extension on a template designed to provide a single binding site for C*C dimer (5'-5' aminoimidazolium-bridged cytidine dinucleotide), the reactive species in primer extension using 2-aminoimidazole activated cytidine ribonucleotides (Figure 5.5). Using varying concentrations of C*C, we obtained Michaelis-Menten parameters for reactions using primers terminated in either a hydroxyl or monophosphate group at the 3' position. We performed these experiments with magnesium-citrate, which protects against pyrophosphate degradation on the time-scales

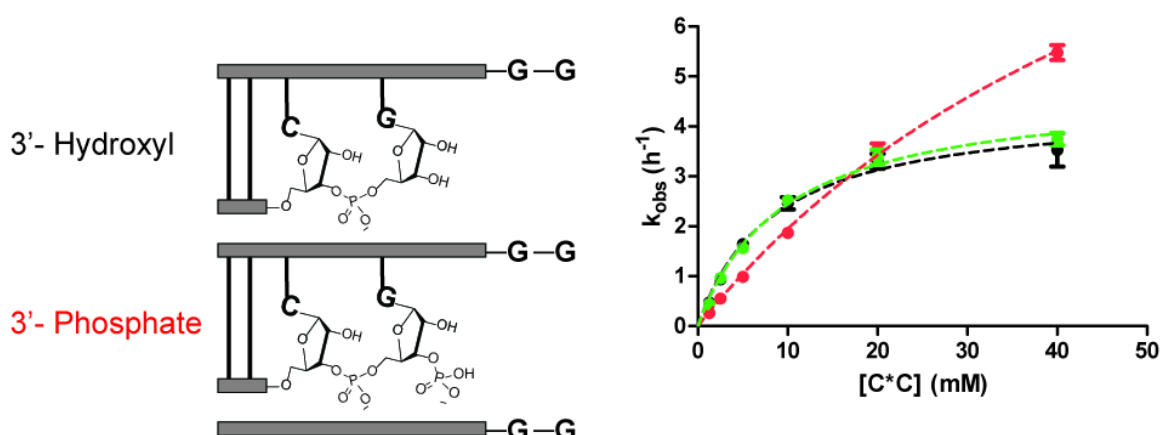


Figure 5.5. Michaelis-Menten analysis of non-enzymatic copying from 3'-hydroxyl and 3'-monophosphate terminated primers, and for a 3'-hydroxylated terminus following a pyrophosphate bond. Left: Schematic representation of the primer-template duplexes analyzed. Right: Michaelis-Menten curves and obtained kinetic parameters. Plotted is k_{obs} (h⁻¹) against the concentration of C*C dimer. All reactions were performed at pH 8.0, 200 mM HEPES, 50 mM Mg^{2+} and 200 mM citrate.

examined. Notably, we observed both higher v_{\max} (14.1 h⁻¹ vs. 4.4 h⁻¹ obtained for hydroxyl) and K_M values (62.6 mM vs. 8.4 mM for hydroxyl) for the phosphate-terminated primer (Figure 5.5). The observed binding defect could be due to either charge repulsion between the negatively charged phosphate and the dimer species or to steric hindrance. The rate enhancement may be rationalized by consideration of the active nucleophile in each case. The actual nucleophilic species in primer extension with ‘native’ RNA is likely a Mg-bound alkoxide, which is poorly populated at pH 8⁶⁵. The nucleophile for the 3'-phosphorylated primer is most likely a phosphate monoester dianion, which is more highly populated at pH 8¹⁵⁰. The difference in the population of these two nucleophiles under primer extension conditions may therefore explain the observed rate difference.

Previous reports have demonstrated a stalling effect on primer extension following the incorporation of mismatched bases¹⁵¹ and certain non-canonical nucleotides¹⁰. Our examination of the stability of the +1 extended product indicates that downstream extension is required to afford protection against cleavage of the newly formed pyrophosphate bond. We therefore sought to quantify the relative rates of reaction for primers in which the terminal 3' nucleotide is joined by either a pyrophosphate or phosphodiester bond (Figure 5.5). The kinetic parameters for incorporation of the next nucleotide were almost identical for the pyrophosphate and phosphodiester linked systems (v_{\max} of 4.8 h⁻¹ following a pyrophosphate linkage vs. 4.4 h⁻¹ obtained for phosphodiester-linked). Importantly, the defect observed for binding of the imidazolium bridged dimer intermediate C*C to the 3' phosphate terminated primer was almost completely ameliorated (K_M 9.6 mM vs. 8.4 mM for phosphodiester-linked). This result implies that after the initial pyrophosphate linkage formation, the downstream steps in primer extension proceed essentially as if the primer contains only native RNA.

To better understand how RNA can accommodate the 3'-5' pyrophosphate linkage during primer extension, we turned to crystallography. A 14mer self-complementary RNA strand containing a

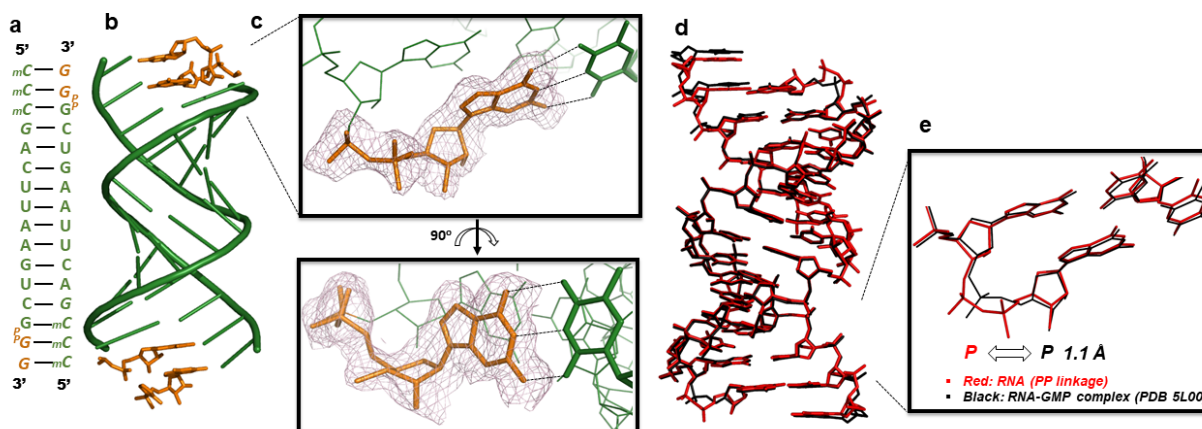


Figure 5.6 (a-e). Crystal structure of RNA containing a pyrophosphate linkage. (a) Schematic of the RNA-monomer complex used for crystallization. Green: RNA/LNA duplex. Italics represent LNA. Orange: pyrophosphate linkage connecting the +1 G nucleobase, and G monomer at +2 position. (b) Overall structure of the RNA-monomer complex. (c) The local structure of the pyrophosphate linkage and the extended guanosine, which forms a Watson-Crick pair with the template methylcytidine. The $2F_o - F_c$ omit map (counter level of 1.0σ) indicates the pyrophosphate linkage and guanosine are ordered. (d) Superimposed structures of the pyrophosphate-containing RNA and the native RNA-GMP complex (PDB 5L00). Red: pyrophosphate-RNA. Black: native RNA. (e) Local structural comparison of the pyrophosphate linkage and the phosphodiester linkage.

terminal 3'-monophosphate (5'-*mCmC*CGACUUAAGUCG_p-3', italic: locked methylcytidine nucleotides, Figure 5.6a) was incubated with 2-AImpG under primer extension conditions (10 mM MgCl₂, 10 mM Tris pH 8.0, 8 h). The extension product was crystallized, and the complex structure determined to 1.6 Å resolution. The duplexes are slip-stacked with one another in an end-to-end fashion, as observed in our previous structure of RNA-GMP complexes¹⁵² (Figure 5.6b). The crystal structure clearly indicates extension of the 3'-phosphorylated RNA by one guanosine nucleotide via pyrophosphate bond formation at both ends of the duplex (Figure 5.6c). The newly incorporated guanosine and the pyrophosphate linkage are well ordered. Canonical Watson-Crick base pairing is observed between the pyrophosphate-linked terminal 3'-guanosine and the templating methylcytidine and the guanosine sugar pucker is in the 3'-endo conformation. Both canonical base pairing and the 3'-endo conformation are favored for non-enzymatic primer extension, suggesting a structural rationale for the correspondence of reaction rates we observed for extension of phosphodiester-linked and pyrophosphate-linked RNA (Figure 5.5). To afford a more detailed comparison with phosphodiester-linked RNA, we superimposed

the pyrophosphate-RNA structure over the native RNA-GMP structure (PDB 5L00) we obtained previously (Figure 5.6d). The two structures possess strikingly similar geometry, again suggesting that the pyrophosphate group does not cause significant structural perturbation to the reaction center. Due to the flexibility of the phosphodiester backbone and the accommodating capacity of the RNA major groove, the pyrophosphate moiety is accommodated with only a slight structural shift (Figure 5.6e).

If pyrophosphate-linked RNA can act as a template for primer extension with canonical, 3'-hydroxyl containing ribonucleotide monomers, a mechanism for re-enrichment of phosphodiester-linked RNA could operate over cycles of replication, as the altered backbone structure will not be passed on to daughter strands. We were therefore curious whether a single 3'-5' pyrophosphate linkage internal to an RNA template influences the ability of the template to direct monomer incorporation. For these experiments, we prepared 20mer ssRNA **3** (AUCGAAGGGppGGCAACACGAC), which contains a single pyrophosphate linkage, as the template strand. The template was designed to contain a stretch of guanosine residues 5'-GppGGG-3' such that we could directly compare the kinetic parameters of template-copying across different systems using the same C*C dimer employed in our previous kinetic experiments. We examined three cases that differ only in the position of the primer strand relative to the pyrophosphate bond within the template (Figure 5.7). In the first case, the pyrophosphate bond joins the two nucleotides of the dimer binding site (Figure 5.7a). In the second case, the pyrophosphate bond is located after the primer annealing site, such that the expected distance between the primer 3'-hydroxyl and the 5'-phosphate of the incoming dimer is greater than in the case of phosphodiester-linked RNA (Figure 5.7b). Finally, in the third case examined the primer is 'clamped' over the pyrophosphate linkage, rendering it internal to the primer-template duplex structure (Figure 5.7c). Using varying concentrations of C*C, we evaluated initial rates of primer extension and determined Michaelis-Menten parameters for the three cases using hydroxyl-

terminated primers and either the pyrophosphate-containing template or a fully RNA template of the same sequence. For all three situations, extension could be observed, and rates determined, for concentrations of imidazolium bridged intermediate as low as 2 mM. This result indicates that pyrophosphate-linked RNA can indeed act as template for the incorporation of canonical ribonucleotide monomers (Figure 5.7). Binding of the imidazolium-bridged C*C dimer to the template was only strongly affected when the pyrophosphate linkage directly connects the two bases of the binding site (~7-fold increase in K_M observed, Figure 6a). However, in all three cases, the maximum rate (v_{max}) values obtained were lower for the pyrophosphate-linked templates. When comparing the three cases, the observed rates of copying across the pyrophosphate template were higher, and the difference in maximum rate between RNA and pyrophosphate templates was less pronounced, when the primer ‘clamps’ over the pyrophosphate linkage (Figure 6c). This may imply that conformational distortion in the ternary complex of primer: template: bound dimer due to the pyrophosphate bond becomes less significant as the linkage is buried within an RNA duplex.

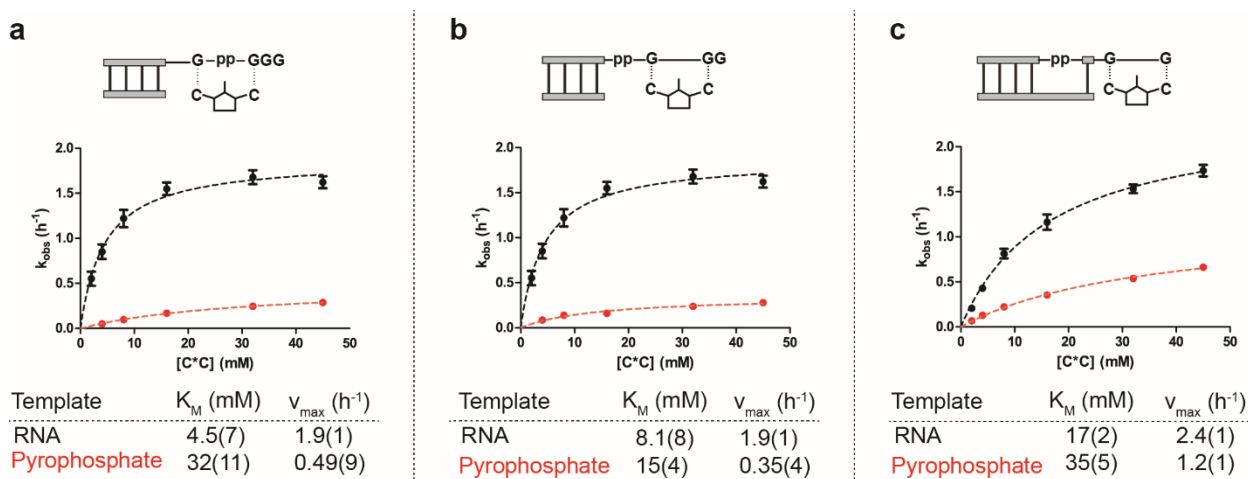


Figure 5.7. (a-c) Michaelis-Menten analysis of non-enzymatic copying across 3'-5' pyrophosphate-linked template 3, compared with phosphodiester-linked control. Top: Schematic representation of the primer-template duplexes analyzed, showing the binding site for the C*C dimer in each case. Bottom: Michaelis-Menten curves and obtained kinetic parameters. Plotted is k_{obs} (h^{-1}) against the concentration of C*C dimer. All reactions were performed at pH 8.0, 200 mM HEPES, 50 mM Mg^{2+} and 200 mM citrate. Further experimental details and data analysis procedures can be found in the Supporting Information.

Discussion

We have found that RNA strand cleavage and hydrolysis is not a dead-end for primitive, RNA-based genetic systems. Instead, the 3'-phosphorylated RNAs that result can participate in non-enzymatic primer extension when supplied with activated nucleotides. Upon reaction of phosphate-terminated primers with incoming nucleotides, a 3'-5' pyrophosphate linkage is formed. By surveying the kinetic parameters for both the synthesis and degradation of this linkage, we have been able to place reasonable constraints on the likelihood of pyrophosphate-linked RNA playing a role in early systems of genetic inheritance.

The kinetic stability of the phosphodiester linkage is a critical feature enabling the storage and use of genetic information¹⁵³. In an RNA world dependent on ribozymes for catalysis, the rate of cleavage of the phosphodiester bond should also determine the lifetime, and thus biosynthesis requirements, of the functional machinery of the protocell¹³⁸. In the transition from prebiotic chemistry to the first genetic polymers, the phosphodiester linkage may have been 'selected' from a range of alternative backbones that were feasible from the standpoint of chemical reactivity but were not stable enough or for other structural or kinetic reasons were unable to support the long-term storage and transmission of genetic information or the functioning of ribozymes. Using only canonical ribonucleotide monomers as precursors to primitive RNA oligonucleotides, four principal backbone linkages can be considered which are derived from either templated copying reactions or the cleavage pathways discussed in this manuscript: 3'-5' and 2'-5' phosphodiester linkages and the 3'-5' and the 2'-5' pyrophosphate linkages explored here. Other linkages, such as 5'-5' pyrophosphate linkages^{29,154}, have been observed in non-enzymatic polymerization of activated nucleotides. These linkages can disrupt information transfer by interfering with the directionality of copying chemistry. Our initial primer extension results suggest that 2'-5' pyrophosphate linkages were unlikely to play a significant role in early forms of RNA-based genetics, as initiation of copying chemistry from a 2'-phosphate is strongly disfavored. In contrast, copying from 3'-phosphate terminated primers is robust and comparable to extension

from 3'-hydroxyl groups. When considering degradation reactions, the rate of cleavage of a 3'-5' pyrophosphate linkage within single-stranded RNA in the presence of magnesium ions is orders of magnitude greater than that of either a 2'-5' or 3'-5' phosphodiester linkage^{13,21}. However, duplex formation protects 3'-5' phosphodiester-linked bonds from cleavage, by constraining RNA in a helical conformation that disfavors attack on the phosphodiester unit by the adjacent 2'-hydroxyl. We observed a similar effect for 3'-5' pyrophosphate linkages, suggesting a similar mechanism of protection. In the case of 2'-5' phosphodiester linkages, duplex formation has the opposite effect, promoting attack by the 3'-hydroxyl and leading to more rapid cleavage. We expect a similar effect to operate for 2'-5'-linked pyrophosphate bonds, although the inefficiency of primer extension from 2'-phosphates has prevented us from directly testing this hypothesis. The observation of enhanced stability for the 3'-5' phosphodiester linkage in duplex structures has previously been used to support the idea of 'selection' over time for the more stable 3'-5' linkage in the context of genetic inheritance¹³. Our results suggest that susceptibility to strand cleavage in the presence of magnesium ions, which are required for RNA folding and catalysis, could explain why biology does not currently employ 3'-5' pyrophosphate linkages in the RNA backbone. However, before the emergence of enzymatic mechanisms for synthesis and proof-reading of genetic polymers, pyrophosphate linkages could have been tolerated as part of duplex structures and thus have contributed to the transmission of primitive genetic information.

We have shown that pyrophosphate linkages in template strands do not block template-directed primer extension. This result suggests pyrophosphate linkages may have been tolerated in primordial RNA-based genetics and raises the prospect of a non-enzymatic 'salvage' and proof-reading cycle for cleaved RNA (Figure 5.8). Following a cleavage(which could take place in a single strand or within a duplex structure), a cyclic phosphate terminus results and is further hydrolyzed to a mixture of 2' and 3' terminal phosphates, with the latter being slightly favored¹⁵⁵. If a 3'-phosphorylated RNA is bound by a template strand, reaction with activated monomers could

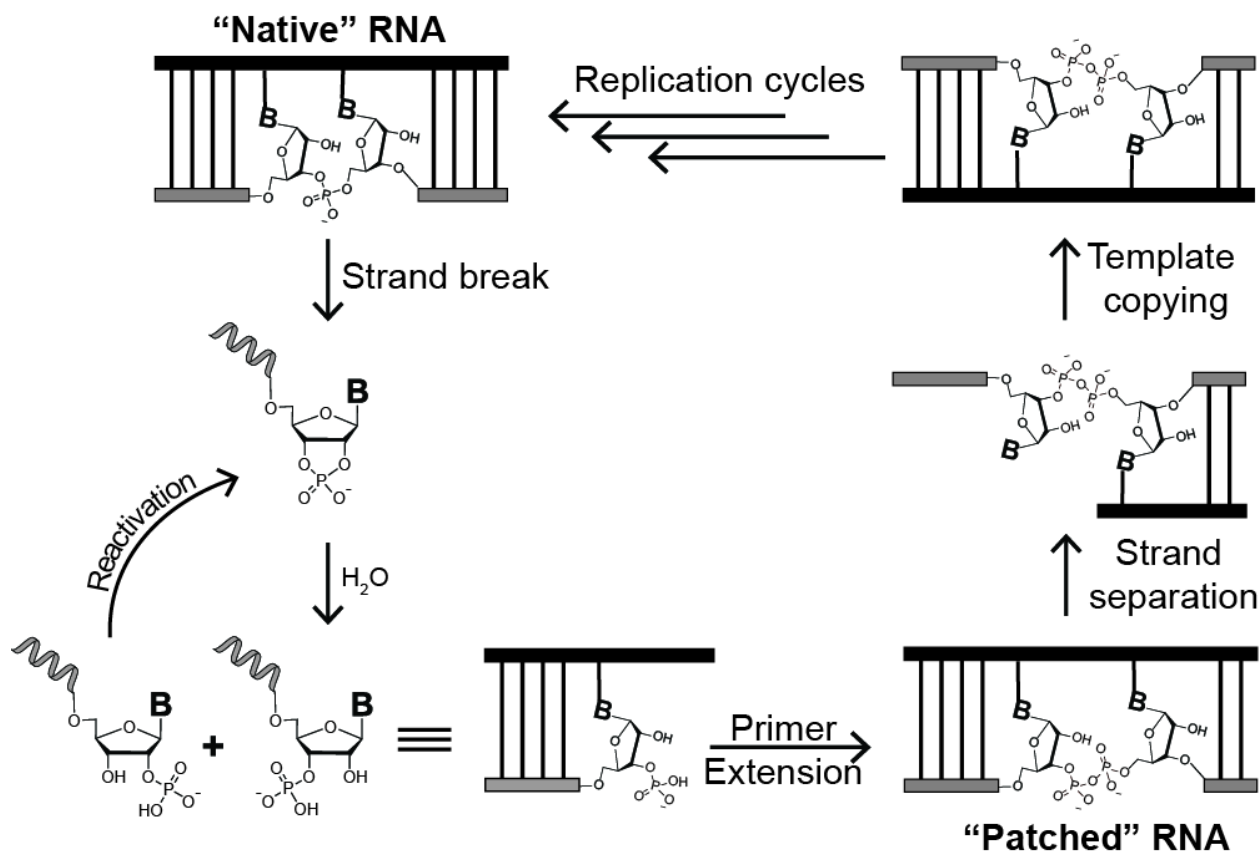


Figure 5.8. A non-enzymatic pathway for the recovery of genetic information via a pyrophosphate 'patch'. (From top left) Following a cleavage event, a cyclic phosphate terminus is formed which hydrolyzes to a mixture of 3' and 2' monophosphate-terminated strands. Primer extension from a 3'-monophosphate terminus leads to duplex RNA containing a single 3'-5' pyrophosphate linkage. Strand displacement and templated copying with canonical ribonucleotides leads to re-enrichment of 'native', 3'-5' phosphodiester-linked RNA over multiple cycles of replication. The unreactive 2' monophosphate-terminated strands can be reactivated to the cyclic phosphate and re-enter the cycle.

'patch' the cleaved strand via formation of a pyrophosphate linkage. This would result in a duplex structure bearing a single pyrophosphate linkage, the lifetime of which depends on the relative rate of pyrophosphate cleavage compared to primer extension or ligation reactions. Cleaved pyrophosphate bonds can simply re-enter the cycle while those sequences able to extend efficiently (or undergo rapid ligation reactions) will 'bury' the labile pyrophosphate bond within a duplex structure, affording some protection from cleavage. 2'-phosphorylated RNA will not form pyrophosphate linkages in appreciable quantities but can re-enter the cycle as a 2',3' cyclic phosphate upon activation and cyclization¹⁹.

For regeneration of the original phosphodiester-linked RNA strand, strand separation and copying of the pyrophosphate-linked RNA (now acting as template) is required. If only canonical ribonucleotide monomers are available, we have demonstrated that copying over pyrophosphate linkages is feasible, albeit slower than for purely phosphodiester-linked RNA. Non-enzymatic separation of RNA strands and multiple cycles of copying (replication) have yet to be demonstrated experimentally^{39,103}. Assuming such a cycle, phosphodiester bond formation over a pyrophosphate template would re-enrich canonical RNA in a crude, non-enzymatic form of backbone proof-reading. The extreme susceptibility of pyrophosphate linkages to metal-ion catalyzed cleavage in single-strands renders this pathway unlikely under a model of non-enzymatic replication that requires temperature cycling, as this promotes a fully single-stranded state under conditions likely to promote cleavage.

However, if a strand displacement synthesis is possible, temperature cycling is not required, and the salvage pathway may be feasible. Mariani et al.¹⁹ have demonstrated a different cycle of backbone repair that converts 2'-5' to 3'-5' phosphodiester linkages via iterative cycles of strand cleavage, 2'-acetylation and non-enzymatic RNA ligation¹²⁰. Such a cycle provides a pathway for the repair of genetic information that also relies on the formation of a 3'-phosphate as the critical intermediate. In conditions that promote the chemical activation of phosphate groups, ligation with a 5'-hydroxyl is thus a complementary pathway to that described in this manuscript. Under such activating conditions, if a 5'-phosphate is present on a bound ligator, pyrophosphate formation should also occur. Hydrolysis of a 3'-5' pyrophosphate linkage produces such a 5'-phosphate on the downstream cleaved fragment, which could feed back into the system as an input for repeated ligation with a 3'-phosphate or ligation to an oligo bearing a 3'-hydroxyl. A network of interconnecting salvage pathways for RNA may therefore operate under non-enzymatic conditions, all of which depend on the unique reactivity of 3'-phosphates generated upon hydrolytic cleavage.

Acetylation of the vicinal 2'-hydroxyl group of a terminal 3'-phosphate before primer extension^{19,120} should render the resultant pyrophosphate linkage stable to strand cleavage as the 2'-hydroxyl would be incapable of transesterification. Such prebiotically plausible 'protection' could greatly enhance the retention of pyrophosphate linkages within a population of RNAs and bolster the chances of repair by enhancing the kinetic stability of pyrophosphate-containing templates. As a proof of principle that protecting the 2'-hydroxyl would not inhibit primer extension from a terminal 3'-monophosphate, while reducing the problem of strand cleavage, we synthesized a primer with a 2'-OMe, 3'-phosphorylated terminus and compared its activity with our original 2'-OH, 3'-phosphorylated primer in a primer extension assay with free magnesium cations (Figure S5.3). Over 18 hours, both primers were extended up to +6 products, with a slightly greater yield of +6 products observed for the 2'-O-methylated primer. Pleasingly, minimal remaining primer (5%) was observed for the 2'-O-methylated system during the course of the assay. In contrast, the reaction with a free 2'-hydroxyl yielded 22% of a product that co-migrates with the original primer band, consistent with cleavage leading to a cyclic phosphate terminated strand (evidence for the co-migration of our 3'-phosphorylated primer and its derived cyclic phosphate is presented in Figure S5.2b). The fraction of remaining primer observed for the 2'-O-methylated system correlates with that observed in identical reactions conducted with non-phosphorylated RNA, in which native phosphodiester bonds are formed that do not appreciably cleave on this time scale. This result suggests that prebiotically-plausible 2'-OH protection could indeed facilitate the increased retention of pyrophosphate bonds in early RNA polymers and perhaps expand potential roles for such linkages in the RNA world. Work towards this end is currently underway in our laboratory.

This study examined primer extension using only canonical ribonucleotides reacting with phosphate groups introduced to the 3' ends of primers by solid-phase synthesis. A remaining question is the efficiency of primer extension using 2' or 3'-phosphorylated monomers. Schwartz

and Orgel have shown that 3'-monophosphate, 2'-deoxy modified monomers can polymerize on DNA templates to form 3'-5' pyrophosphate-linked oligomers¹⁴⁴. Isolated pyrophosphate-linked oligonucleotides could also be used as templates for further polymerization, employing the same activated monomers¹⁴⁵. These results imply that a full cycle of non-enzymatic replication may be possible in a system composed of phosphorylated monomers and pyrophosphate-linked genetic polymers. Here, we examined the consequences of a single pyrophosphate linkage in the context of RNA copying. Primer extension using 3'-phosphorylated ribonucleotide monomers, or in a system containing monomers with a mixture of phosphorylation states, should introduce multiple pyrophosphate linkages. The implications of multiple pyrophosphate bonds for the function of RNA, as both genetic material and catalyst, remain to be explored.

6. Structure-activity relationships in nonenzymatic template-directed RNA synthesis

The catalytic synthesis of sequence defined polymers is a hallmark of biology. Organisms devote considerable resources to the template-directed synthesis of DNA, RNA, and proteins, which coordinate to carry out all cellular processes. These interrelated synthetic programs may have originated some four billion years ago when life first appeared on Earth. According to the RNA World hypothesis¹⁰⁰, life started with self-replicating protocells that stored their genetic information in RNA and carried out cellular processes with the help of RNA enzymes. A possible prebiotic pathway to primordial self-replicating RNA is through the template-directed nonenzymatic synthesis of RNA. In this process, a strand of RNA directs the synthesis of its complement by bringing the necessary components together through Watson-Crick base-pairs. Considerable effort from multiple research groups has gone into understanding and optimizing this process in the past decades, but substantial problems remain unsolved^{103,156}. For example, the rate and fidelity of chemical RNA replication are not high enough to sustain the reproduction and evolution of a primitive cellular organism, and the relevant reaction mechanisms are still poorly understood. A deeper understanding of the mechanism of RNA copying and the factors that contribute to fast and high-yielding reactions is necessary to improve nonenzymatic RNA replication.

We use nonenzymatic primer extension as a laboratory model for prebiotic RNA replication (Figure 6.1). Two RNA strands, a primer and a template, anneal to form a duplex which contains a 5'-overhang. Chemically activated nucleotides bind to this overhang and the primer is extended one nucleotide at a time. We have recently shown that chemically activated phosphoroimidazolid mononucleotides do not participate directly in the reaction, but instead form an imidazolium-

bridged dinucleotide, which is the reactive electrophilic species⁷⁴. The imidazolium-bridged dinucleotide, once bound to the template through Watson-Crick base pairing, reacts quickly with the 3'-end of the primer to generate extension products. Understanding the structure and reactivity of the electrophile⁷⁴ has helped rationalize previously unexplained results and contributed to developing an improved RNA copying process⁸⁷. Unfortunately, the participation of the 3'-OH nucleophile and the catalytic metal ion are less well understood. Previous reports hint that the active nucleophilic species is a 3'-alkoxide,

generated with the aid of a Mg^{2+} cation⁶⁵. It has also been known for decades that the reaction works best when the sugar ring of the terminal primer nucleotide is in the ³E (C3'-endo) conformation^{157,158}, but the reasons for this preference are not understood. Besides the conformational effects that might be at play, the acidity of the reacting hydroxyl could affect the rate. A more acidic 3'-OH group is easier to deprotonate and thereby generate the reactive alkoxide species. However, a better understanding of the transition state of the reaction is required before the differences in rate can be adequately interpreted. For example, if the O-H bond is broken in the rate determining step of the reaction, then the rate will be strongly influenced by the acidity of the 3'-OH. Conversely, if the rate limiting step is the attack of the previously formed alkoxide on the adjacent phosphate of the imidazolium-bridged dinucleotide,

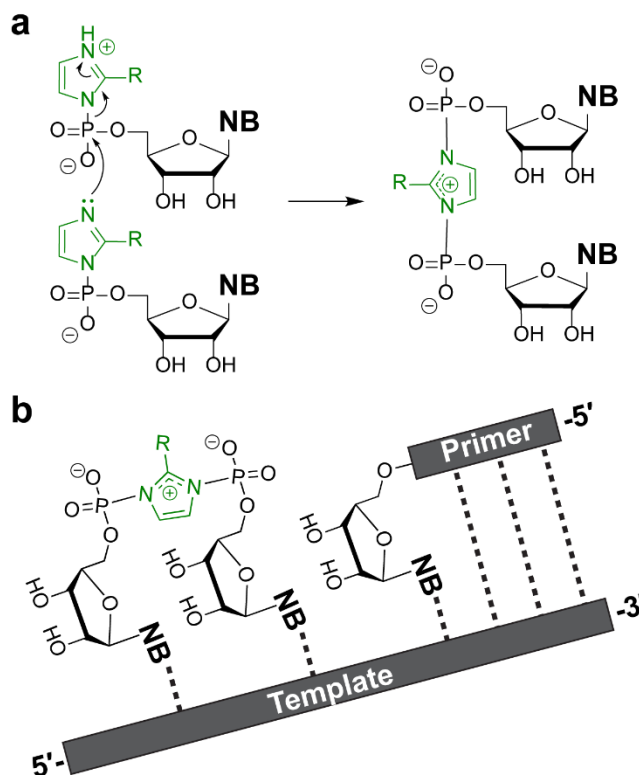


Figure 6.1. Primer extension occurs via two sequential reactions. (A) Formation of the imidazolium-bridged dinucleotide intermediate, the reactive species in nonenzymatic RNA primer extension. (B) The imidazolium-bridged intermediate anneals to the primer-template duplex and reacts with the 3'-end of the primer.

then steric congestion surrounding the hydroxyl group will strongly affect the energy of the transition state and the rate of the reaction.

In this report we combine kinetic and structural studies to determine the prerequisites for efficient templated nonenzymatic synthesis of nucleic acids. We show that the rate limiting step of the reaction involves the nucleophilic attack of a Mg-alkoxide and therefore the reaction is insensitive to differences in hydroxyl acidity. However, because the rate limiting step involves an increase in steric bulk around the nucleophilic hydroxyl group, the steric environment around the hydroxyl moiety strongly influences the rate.

Results

We set out to determine the structure-activity relationships for nonenzymatic primer extension using a set of primers with modified guanosine nucleotides at their 3'-termini. The set of nucleotides includes several potentially prebiotic alternatives to ribonucleotides (threose, arabinose and deoxyribose nucleotides), as well as several synthetic modified nucleotides that serve as probes for the mechanism of the reaction. We used X-ray crystallography to determine the conformation of each of these sugar-modified G nucleotides when present at the end of an oligoribonucleotide, base-paired to a template strand, and adjacent to a template bound GpppG (Figure 6.2a), an isosteric,

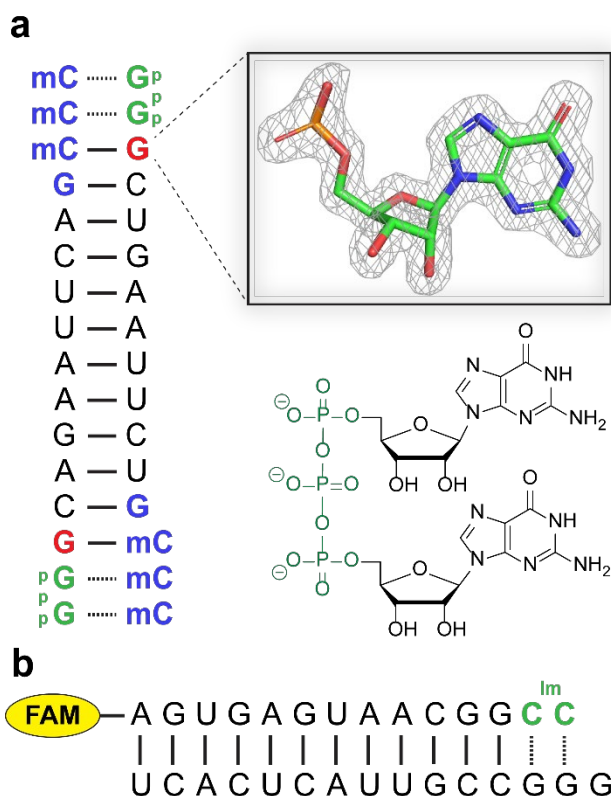


Figure 6.2. Constructs used to correlate structure and reactivity. (A) Self-annealing oligonucleotide construct used for X-ray crystallography studies. Blue: locked nucleic acid (LNA) residues; mC is 5-methyl-cytosine. The crystallography studies are focused on the 3'-terminal nucleotide of the primer, in red. The inset shows a $2F_o - F_c$ omit map contoured at 1.5σ of the structure of this terminal nucleotide for the all-RNA primer³. The GpppG analog is green on the sequence diagram and its structure is shown below the inset. (B) Fluorescently labelled primer-template duplex used for kinetic measurements. The imidazolium-bridged dinucleotide is in green. FAM is 5(6)-carboxyfluorescein.

unreactive analog of the imidazolium-bridged dinucleotide³. A similar construct was used to determine the primer extension reaction rates for each modified G nucleotide. In each case a primer-template duplex reacts with a saturating concentration² (20 mM) of the imidazolium bridged dinucleotide (Figure 6.2b).

The modified primers we studied display a wide range of reactivities in a standard primer extension reaction, with differences of at least two orders of magnitude between the most reactive and least reactive modifications of the 3'-terminal primer nucleotide (Figure 6.3). In an effort to explain these rate variations, we looked for correlations between rate and sugar conformation. Five membered rings avoid an unstable planar conformation by puckering in two different ways: one atom is out of the plane of the other four, resulting in an envelope conformation (E), or two atoms can be above and below the plane of the other three, resulting in a twist conformation (T). In theory, furanoses can adopt multiple twist and envelope conformations that are energetically similar²⁰; in practice, decades of structure determination by crystallography and NMR have shown that 2'-deoxyribonucleotides prefer the ²E (2'-*endo*) conformation in which the C2' carbon atom is above the plane of the other four atoms, whereas ribonucleotides prefer the ³E (3'-*endo*) conformation in which the C3' carbon atom is above the plane of the other four atoms. Most modifications we tested crystallize in the ³E conformation, except for the arabino-modifications FANA and ANA (Figure 6.3) which crystallized in the ⁰T₄ and ²E conformations. Interestingly, the arabino-modifications are the slowest reacting species, especially arabino-G, for which we could not measure the rate of extension. Since the crystal construct is a symmetric self-annealing duplex, each crystal structure contains two of the 3' modifications. The crystal structure of the arabinose modified primer is the only one that crystallizes in two different conformations. On one end, the imidazolium bridged dinucleotide analog is disordered and forms a non-canonical base-pair at the position downstream of the primer. The same structure has previously been observed in a construct with guanosine monophosphates instead of the imidazolium bridged

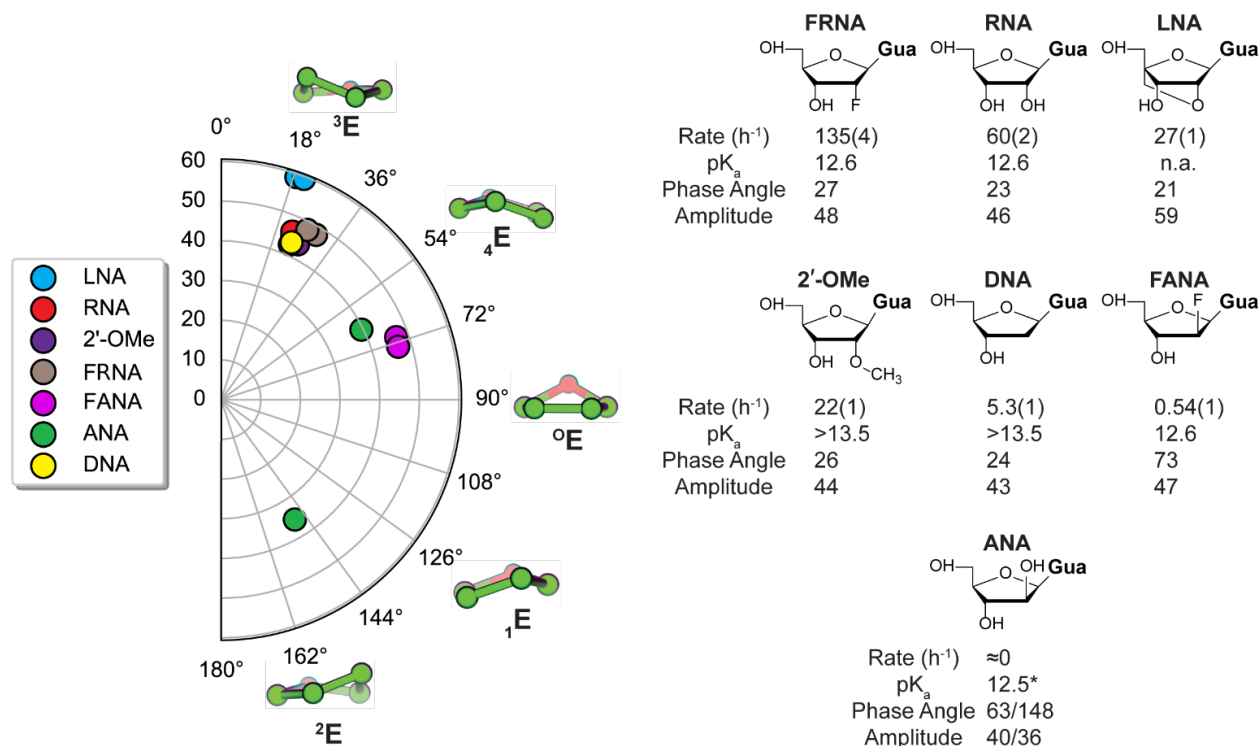


Figure 6.3. Conformational landscape of the primer 3'-terminal nucleotide modified sugars. Left side: conformation of each of the 3'-terminal sugars from crystal structures of modified primer/template/GpppG complexes. Since the crystal construct is symmetric, each modification is represented twice. Each point on the polar plot corresponds to a value of the phase angle (0° - 180°) and the amplitude of the pucker (0-60). Five representative envelope conformations are shown, each corresponding to a 36° shift in phase angle. Right: chemical structure of each modification, its extension rate relative to that of the RNA primer, and the experimentally determined pK_a value of the nucleoside. The phase angle and puckering amplitudes are listed for each nucleotide. For the modifications that crystallized symmetrically only one pair of values were shown. The complete list of structural parameters is presented in Supplementary Table S5. The reaction rates were determined in triplicate and the standard error of the mean is reported *The pK_a value of the arabinose nucleoside (ANA) is from reference 13.

dinucleotide¹⁵⁹. The arabinose sugar is in the ^2E conformation and there is a putative hydrogen bond between its 2'-OH and the N7 atom of the guanosine moiety of the imidazolium bridged dinucleotide analog (Figure S6.1). In this conformation no reaction is possible, since the O3'-P distance is almost 12 Å. At the other end of the duplex, the imidazolium bridged dinucleotide analog is well-ordered, forms canonical Watson-Crick base-pairs with the template, and the arabinose sugar is in the $^0\text{T}_4$ conformation. Here the O3'-P distance is 5.14 Å; while this distance is longer than for all other analogs, the difference is small and may not fully account for the extremely slow reaction rate.

Although the sugars that crystallize in the 3E conformation react considerably faster than the ANA and FANA analogs, the reasons for this difference cannot be deduced solely from the structural parameters (Table S6.1). For example, the two fluoro-modifications, FRNA (3E) and FANA (OT_4), have their 3'-hydroxyls at a similar distance to the adjacent phosphate (4.5Å and 4.9Å respectively), as well as similar angles between 3'-OH and the bridging P-O bond of GpppG (133° and 138°), but FRNA reacts 250 times faster. Although the 0.5Å difference could potentially explain the difference in reactivity, the 3'-OH of the LNA modification is 0.9Å closer to the adjacent phosphate when compared to FRNA, but LNA reacts 5-time slower than FRNA. Additionally, FRNA, RNA, DNA and 2'-OMe have very similar distances and angles of approach but the fastest (FRNA) reacts 25-times faster than the slowest (DNA).

Since structural parameters such as nucleophile-electrophile distance and the angle of approach of the 3'-OH group towards the P-N bond of the electrophile do not by themselves account for the measured differences in rate we asked the whether the acidity of the 3'-OH group of the primer correlates with the rate of the primer extension reaction. There are two possible ways in which the acidity of the hydroxyl group could have an impact on the rate of the reaction. First, when comparing rates at constant pH values, if the modifications lead to various extents of 3'-OH deprotonation, the differences in rate will reflect the degrees of deprotonation rather than the inherent reactivities of the alkoxides. Second, the inherent reactivities of the alkoxides should correlate with the acidity of the alcohols. An alkoxide that is more difficult to form is also more reactive. Provided all other steric and electronic factors are identical, the rate of phosphate transfer should increase as the pKa of the attacking nucleophile increases when the reaction is carried out at pH values where the alcohol is fully deprotonated. This has indeed been observed by Piccirilli for RNA analogs in trans-esterification reactions¹⁶⁰.

We therefore determined the 3'-OH pKa values of the modified mononucleosides by 1H NMR using the method of Chattopadhyaya¹⁶¹. Because the acidity of the 3'-OH in most nucleotides is

low, only a handful of pKa values could be determined in aqueous solution. For example, only a lower bound of 13.5 could be estimated for the 3'-OH pKa of 2'-deoxy and 2'-OMe derivatives. Interestingly, the acidities of the 3'-OH in RNA, FRNA and FANA are identical within experimental error. Although FRNA and FANA lack the 2'-OH which can stabilize the anion through a hydrogen bond, the inductive effect of the fluorine atom appears to be equally well suited to stabilizing the negative charge of the O3' alkoxide. Because the inductive effect of the fluorine atom is transmitted through bonds, the spatial orientation of the 2'-substituent is irrelevant, thus FANA and FRNA have identical acidities.

The fact that the nucleoside 3'-OH pKa values do not reflect the trends in reactivity suggests that under reaction conditions the deprotonation of the primer 3'-OH is quite different. Although the 3'-hydroxyl group in RNA is very weakly acidic, it is expected to be a much stronger acid when coordinated with Mg^{2+} . Because of the inherent technical difficulties associated with determining the pKa of the 3' -OH of the primer in a duplex and in the presence of Mg^{2+} by the NMR method used for the nucleosides, we used a kinetic method of determining the pKa. We determined a pH-rate profile for most of the modifications, by measuring the reaction rate at pH values ranging from 7-10. Because the reaction rates at high pH were too fast to accurately measure with the initial experimental setup, we changed the template so as to create a binding site for a di-adenosine

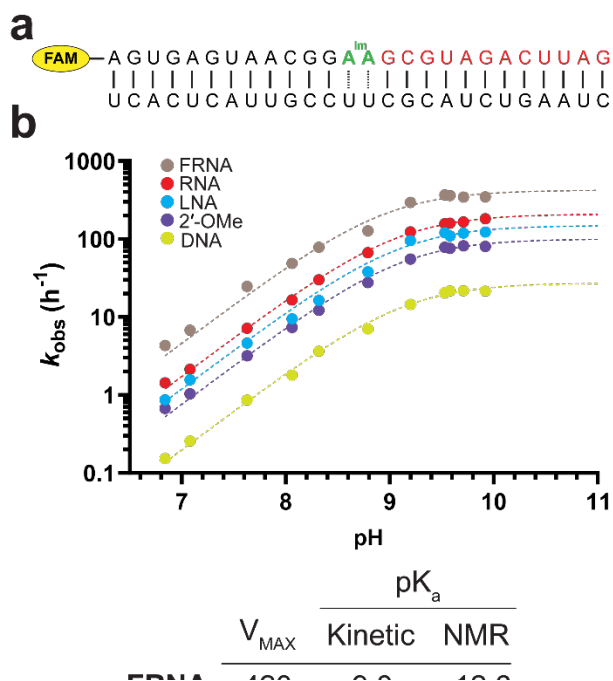


Figure 6.4. Determination of the 3'-OH pKa for five modified nucleotides. (A) The construct used for the primer extension experiments. The oligonucleotide in red is a downstream binder which pre-organizes the duplex¹⁶⁹ and increases the affinity of the template for the imidazolium bridged intermediate. (B) pH-rate profiles for primer extension with primers containing five distinct 3'-terminal residues. Each reaction was carried out in triplicate, and only the mean values were used to fit the data. The table contains the pKa and V_{MAX} , determined from the pH-rate profiles. The right most column shows pKa values determined by 1H NMR in the absence of Mg^{2+} .

imidazolium bridged intermediate, which reacts more slowly than its cytosine analog (Figure 6.4a). Despite the absolute differences in rate between the two constructs, the relative rates of the modifications follow the same trend. The rates increase linearly with pH until a plateau region is reached. The pKa values are then determined by assuming that only the Mg-alkoxide reacts and then fitting the data to the Henderson-Hasselbalch equation. Considering that the nucleosides have different acidities depending on the C2' substituent, it was somewhat surprising that for all the modifications the kinetically determined pKa values were very similar, converging around 9.1 (Figure 6.4b). Although the kinetically determined pKa values for the primer modifications appear to be identical, they need to be interpreted carefully since kinetic measurements carried out at pH values over 9 may be affected by deprotonation of the imino protons of the uracil and guanine moieties of U and G, which have a pKa of ~10. Once the nucleobases are deprotonated, they cannot engage in Watson-Crick base-pairing. Therefore, it is difficult to distinguish whether the reaction rate plateaus because of complete deprotonation of the 3'-OH or whether the primer/template/imidazolium-bridged dinucleotide complex begins to dissociate because of nucleobase deprotonation. Additionally, for the slower reacting modifications such as FANA, the imidazolium-bridged dinucleotide hydrolyzes on a comparable time scale, making it impossible to measure the rate. However, the pH dependence of the reaction changes drastically when replacing Mg^{+2} with Ca^{+2} , suggesting that the kinetically determined pKa values do indeed correspond to deprotonation of the 3'-OH⁶⁵.

Since the kinetically determined acidity constants fail to explain the differences in reactivity between the different modifications, other factors must be at play. However, interpretations of steric and electronic effects in the primer extension reaction hinge on the identity of the rate determining step.

We therefore asked whether deprotonation of the 3'-hydroxyl was rate limiting by measuring solvent kinetic isotope effects (SKIE), which can be used to determine whether or not the breaking of an O-H bond is involved in the rate-limiting step of a reaction. In the case of primer extension, measuring the SKIE will determine whether the reaction mechanism involves the formation of magnesium alkoxide prior to the rate-limiting step or whether the deprotonation of the 3'-OH is the rate determining step of the reaction. If the former is correct, then no SKIE is expected since the O-H bond is broken prior to the rate limiting step. If the deprotonation of the 3'-OH is the rate determining step, then a large SKIE is expected, since the zero-point energy of an O-H bond is different from that of an O-D bond. Alternatively, the deprotonation and nucleophilic attack could be concerted, in which case an intermediate SKIE is expected.

To determine where deprotonation of the 3'-OH lies on the primer extension reaction pathway, we determined the pL-rate profiles (where pL is either pH or pD) in both H₂O and D₂O (Figure 6.5). Between pH values of 7 and 8.5 the slope of the pH rate profile is close to unity, after which it begins to plateau. The kinetically determined value of the 3'-OH pK_a is 9.1, suggesting that Mg²⁺ coordination increases the acidity of the alcohol by four orders of magnitude. A similar pD-rate profile is observed in D₂O, with a determined pK_a value of 9.8 for the 3'-OD. A shift in acidity between the two solvents of this magnitude is expected because of the lower zero-point energy

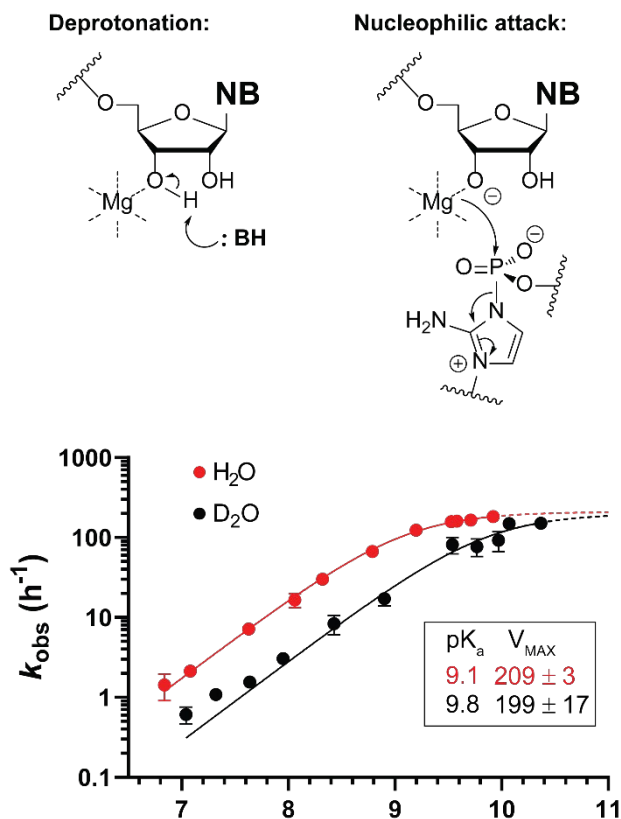


Figure 6.5. The kinetic solvent isotope effect for non-enzymatic template directed RNA primer extension. (A) Two mechanisms representing limiting cases for the rate determining step of the reaction. (B) Graph showing pL-rate profiles for primer extension in H₂O and D₂O. Rates were determined in triplicate. The mean value and its 95% confidence interval are shown on the graph.

of the O-D bond compared to the O-H bond, and it mirrors results reported for the hammerhead ribozyme¹⁶². In the linear section of the pL-rate profile the rates in H₂O are roughly three times higher than the rates in D₂O and converge at higher pL values as the deprotonation of the 3'-OH becomes complete. Considering that the 3'-OD is a weaker acid in heavy water than the 3'-OH in light water, the difference in rates at any given pL value can be explained by the different degrees of deprotonation of the hydroxyl group.

After correcting for pK_a differences, the rates in H₂O and D₂O are virtually identical. The absence of a SKIE suggests that there is no O-H bond breaking in the rate limiting step. This implies that the reactive species, the Mg-alkoxide, is formed before the rate limiting step and therefore, the rate limiting step of the reaction is the nucleophilic attack of the Mg-alkoxide on the adjacent phosphate of the imidazolium-bridged dinucleotide. Any stereoelectronic effects that facilitate this step will therefore lead to an acceleration in rate.

Considering that the FANA and FRNA modifications have identical pK_a values, similar distances between the 3'-OH and the adjacent phosphate group of the substrate, as well as similar O-P-N angles of attack, what is the reason for their 250-fold difference in reactivity? The reasons for this difference can be deduced by examining the crystal structures of FANA and FRNA, specifically the differing environments around the 3' -OH groups. In the ³E conformation adopted by FRNA, the reacting 3'-OH group is pseudo-equatorial, with the molecule adopting staggered conformations along the C3'-C4' bond and the C3'-C2' bond. On the other hand, the ⁰T₄ conformation of FANA forces an eclipsed arrangement of the substituents along the C3'-C4' and C3'-C2' bonds. This results in the reacting OH, which is now pseudo-axial, being eclipsed by the H2' and H4' atoms, as well as a pseudo-axial 1-3 interaction with the H1' (Figure 6.6). A reaction occurring on the 3'-OH is retarded by steric bulk when the rate determining step involves increasing bulk. Because in the transition state of the primer reaction the alkoxide approaches the scissile P-N bond there is an increase in bulk around the nucleophile. When the two reactants are

approaching each other, the 1,2 and 1,3 eclipsing interactions in the ${}^{\circ}\text{T}_4$ conformation will destabilize the reacting complex. Therefore, a pseudo-axial alcohol reacts more slowly than its pseudo-equatorial analogue because its transition state is more strongly destabilized. Similar behavior was first observed by Barton for sterols, where an equatorial alcohol is easier to esterify than its axial isomer¹⁶³.

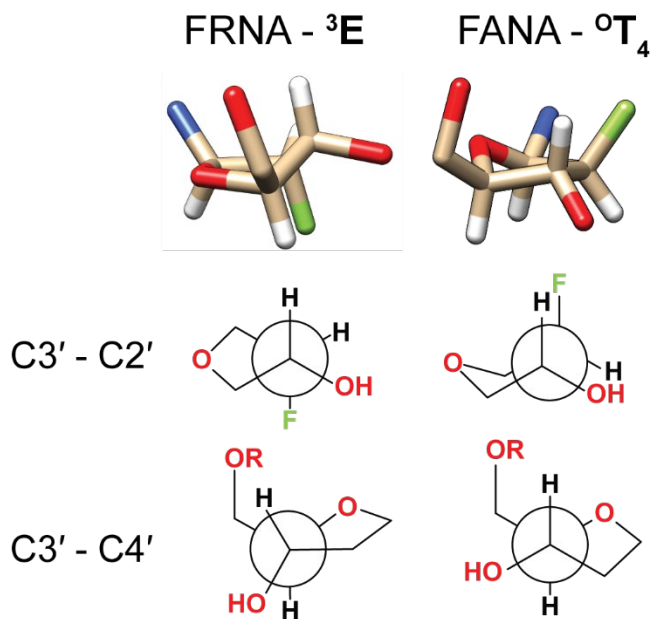


Figure 6.6. Crystal structures of the sugar moiety of 2'-fluoro guanosine and 2'-fluoroarabino guanosine. Hydrogen atoms are represented in white, oxygen in red, fluorine in green and nitrogen in blue. The bottom of the figure shows the Newman projections across two different bonds for the two modifications.

Even for the modified nucleotides that adopt a ${}^3\text{E}$ sugar conformation there is a 25-fold difference in rates, revealing that conformation is not the sole determinant of reactivity. Considering that the kinetically determined pK_a values for these five modifications are virtually identical, what then is the reason for the different reactivities? Strikingly, there is a strong correlation between the rates of the reactions and the electronegativity of the 2' substituent amongst the five modifications (Figure 6.7). This correlation suggests that the reaction is affected by either a stereoelectronic or an inductive effect that stems from the 2' substituent. For a stereoelectronic effect to directly stabilize the transition state, suitable overlap is required between the orbitals of the forming P-O bond and the C-X bond, where X is the 2' substituent. However, in the ${}^3\text{E}$ conformation the dihedral angle between these two bonds is closer to a perpendicular orientation of the orbitals than to the preferred parallel

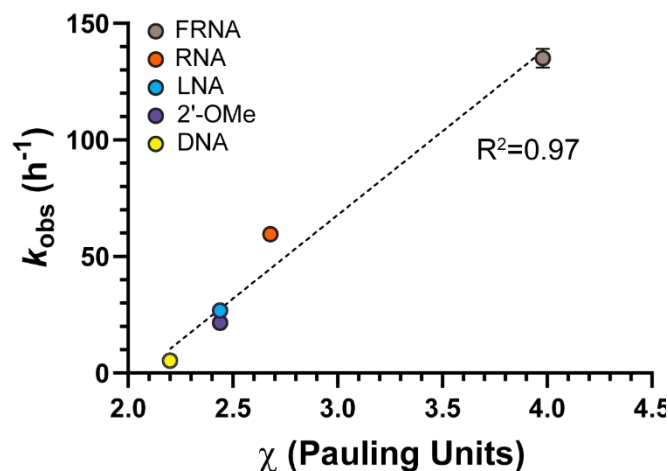


Figure 6.7. The reaction rate of the 3'-terminally modified primers correlates with the electronegativity of the 2'-substituent. We assumed that the electronegativity of the -OMe group and the OCH₂R group in LNA are the same, and we used the value reported by Bratsch¹⁷⁰.

arrangement. Considering the poor orbital overlap as well as the relatively large distance between the C-X bond and the P-O bond, it is unlikely that a stereoelectronic effect can stabilize the transition state. A more reasonable explanation for the influence of the 2' substituent on the rate of the reaction invokes a field effect. The reactive species in the primer extension reaction is the negatively charged 3'-

alkoxide. A more electronegative 2' substituent will have a larger nearby partial negative charge, which more strongly destabilizes the alkoxide, raising the energy of the ground state of the reaction. In the transition state the 3'-OH group pivots closer to the adjacent phosphate of the imidazolium bridged intermediate, possibly through an increase in the amplitude of the sugar pucker, and the electrostatic repulsion between the 2' and 3' groups is presumably reduced. Thus, an electronegative substituent alters the energy of the ground state more strongly than the energy of the transition state, resulting in an increase in rate. For DNA, where the 2' substituent is the hydrogen atom, this effect does not operate and therefore it reacts the slowest.

An alternative potential explanation for the influence of the electronegativity of the 2' substituent on rate in this series of modifications arises from the relationship between reaction rate and the puckering amplitude of the furanose rings. The rate of the reaction and the puckering amplitude are correlated with the electronegativity of the 2' substituent for four of the five ³E modifications. Only the conformationally constrained LNA diverges from this correlation (Figure S6.2).

Considering that conformational preferences are dictated by the strength of the gauche effects in furanose rings¹⁶⁴, and that a more electronegative substituent induces a stronger gauche effect¹⁶⁵,

FRNA is as expected the most strongly puckered of the unconstrained nucleotides. A similar effect is observed in the conformational preferences of mononucleotides in solution, where an electronegative substituent in the 2' position biases the conformational equilibrium towards the ³E conformation¹⁶⁶. If the conformational dynamics from the ground state to the transition state involve enhanced sugar puckering in order to allow the 3'-OH to approach the phosphorous atom, then a modification that is pre-organized by a stronger puckering will require less energy to reach the transition state and therefore will react faster. Therefore, the stereoelectronic effect generated by the 2' substituent pre-organizes the ground state of the reaction. The conformationally restricted LNA is limited in the ways the ring can rearrange and reacts more slowly than expected based on its puckering amplitude. Ultimately, a more detailed knowledge of the structure of the transition state of the primer extension reaction is needed to decide on the merit of these two alternative explanations.

Having unearthed the factors that govern the reactivity of the 3'-OH in the furanose series we expanded our investigation to include sugar derivatives that cannot form the canonical 3'-5' phosphodiester bond. Threose nucleic acids (TNA) have been proposed as prebiotically plausible alternatives to RNA by Eschenmoser and colleagues¹⁶⁷.

We have synthesized a primer terminating in a TNA residue and measured the extension rate (Figure 6.8). Surprisingly, the TNA primer reacts only 6-times slower than RNA, suggesting that the 2'-OH of the threose moiety is a competent nucleophile in the primer extension reaction. To rationalize this result, we have determined the crystal structure of the TNA terminated primer together with the analog of the imidazolium-bridged dinucleotide. The threose sugar crystallizes in a ⁴E (C4'-exo) conformation, in agreement with our previously determined structure¹⁶⁸. The threose nucleotide is linked to the primer through a 3'-3' phosphodiester bond which results in the rotation of the sugar so that the TNA C2'-OH is relocated to the same position where the RNA

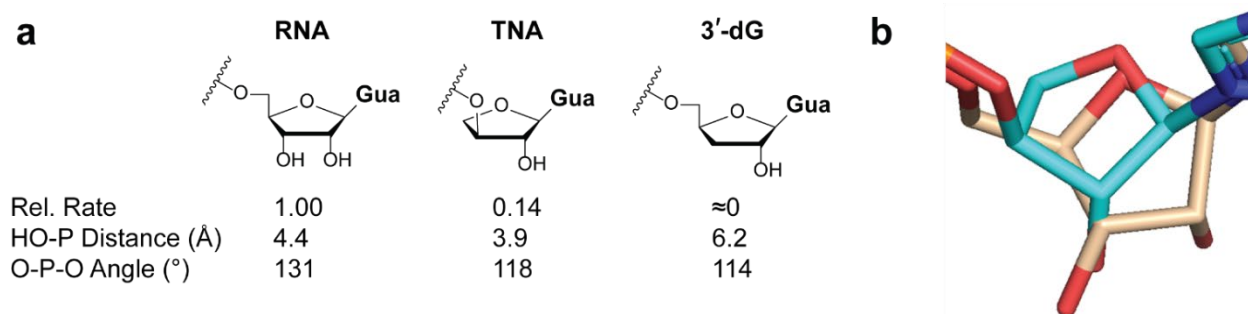


Figure 6.8. Primer extension with three different terminal nucleotides. (A) The relative rates and structural parameters for the TNA and 3'-deoxyguanosine modifications as determined from the crystal structures. Each rate has been determined in triplicate. (B) Top-view of the superimposed RNA (gold) and TNA (cyan) structure showing the rotation of the TNA ring.

C3'-OH resides. This rotation brings the TNA 2'-OH close to the adjacent phosphate group of the imidazolium-bridged dinucleotide analog. However, because the TNA sugar is in a 4E conformation, the TNA 2'-OH is eclipsed by the TNA H1'. This steric congestion is likely to be responsible for the decrease in rate between TNA and RNA. These results agree with our recent report on the reactivity of threose nucleotide systems¹⁶⁸. On the other hand, a primer terminating in a 3'-deoxyguanosine residue is completely unreactive in the primer extension reaction. The crystal structure of 3'-deoxyguanosine containing complex is identical to that of the all-RNA complex, apart from the absence of the 3'-OH. Therefore, the 3'-deoxyguanosine 2'-OH is far from the phosphate group and cannot react to form a 2'-5' phosphodiester bond. Furthermore, the 3'-deoxyribose is in the 3E conformation, in which the 2'-OH fully eclipsed by its neighboring H-atoms.

Conclusion

Modifying the 3'-terminus of the primer has a large impact on the rate of the primer extension reaction and the fastest reacting modifications all crystallize in the 3E conformation. We first showed that the rate of the reaction is surprisingly independent of the acidity of the 3'-OH, since in the presence of Mg^{2+} all modifications have very similar acidities. We then determined that the rate-limiting step of the reaction is the attack of the Mg-alkoxide on the adjacent phosphate group of the imidazolium bridged intermediate by measuring the solvent kinetic isotope of the reaction.

Since the transition state of this state increases in steric bulk compared to the ground state, the steric environment of the nucleophile strongly affects the rate of the reaction. When the terminal nucleotide on the primer is in the ³E conformation the steric congestion around the 3'-OH group is relieved and thus the reaction is faster. In addition, a higher puckering amplitude of the ring in the ³E conformation results in a shorter 3'-OH - phosphate distance, as can be seen from the crystal structure of LNA (Figure 6.3 and Table S6.1). Finally, having an electronegative substituent at the 2' position speeds up the reaction, presumably by destabilizing the ground state through a field-effect. This effect, alongside its preference for the ³E conformation, explains why RNA outcompetes its alternatives (including potentially prebiotic alternatives such as ANA and TNA) in the template-directed copying of nucleic acids.

7. References

1. 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 Res.* **44**, 5504–5514 (2016).
2. Walton, T. & Szostak, J. W. A Kinetic Model of Nonenzymatic RNA Polymerization by Cytidine-5'-phosphoro-2-aminoimidazolid. *Biochemistry* **56**, 5739–5747 (2017).
3. Zhang, W. *et al.* Insight into the mechanism of nonenzymatic RNA primer extension from the structure of an RNA-GpppG complex. *Proc. Natl. Acad. Sci.* **114**, 7659 LP – 7664 (2017).
4. Zhang, W., Walton, T., Li, L. & Szostak, J. W. Crystallographic observation of nonenzymatic RNA primer extension. *Elife* **7**, e36422 (2018).
5. Li, L. *et al.* Enhanced Nonenzymatic RNA Copying with 2-Aminoimidazole Activated Nucleotides. *J. Am. Chem. Soc.* jacs.6b13148 (2017) doi:10.1021/jacs.6b13148.
6. Eigen, M. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**, 465–523 (1971).
7. Leu, K. *et al.* Cascade of Reduced Speed and Accuracy after Errors in Enzyme-Free Copying of Nucleic Acid Sequences. *J. Am. Chem. Soc.* **135**, 354–366 (2013).
8. Xu, J. *et al.* A prebiotically plausible synthesis of pyrimidine β -ribonucleosides and their phosphate derivatives involving photoanomerization. *Nat. Chem.* **9**, 303–309 (2017).
9. Stairs, S. *et al.* Divergent prebiotic synthesis of pyrimidine and 8-oxo-purine ribonucleotides. *Nat. Commun.* **8**, (2017).

10. Kim, S. C., O’Flaherty, D. K., Zhou, L., Lelyveld, V. S. & Szostak, J. W. Inosine, but none of the 8-oxo-purines, is a plausible component of a primordial version of RNA. *Proc. Natl. Acad. Sci.* **115**, 13318 LP – 13323 (2018).
11. Duzdevich, D., Carr, C. E. & Szostak, J. W. Deep sequencing of non-enzymatic RNA primer extension. *Nucleic Acids Res.* **48**, e70–e70 (2020).
12. Duzdevich, D. *et al.* Competition between bridged dinucleotides and activated mononucleotides determines the error frequency of nonenzymatic RNA primer extension. *bioRxiv Mol. Biol.* 1–11 (2021) doi:10.1093/nar/gkab173.
13. Usher, D. A. & McHale, A. H. Hydrolytic stability of helical RNA: a selective advantage for the natural 3’,5’-bond. *Proc. Natl. Acad. Sci. U. S. A.* **73**, 1149–1153 (1976).
14. Guo, F. *et al.* Effect of Ribose Conformation on RNA Cleavage via Internal Transesterification. *J. Am. Chem. Soc.* **140**, 11893–11897 (2018).
15. Kanavarioti, A., Lee, L. F. & Gangopadhyay, S. Relative reactivity of ribosyl 2’-OH vs. 3’-OH in concentrated aqueous solutions of phosphoimidazolid activated nucleotides. *Orig. Life Evol. Biosph.* **29**, 473–487 (1999).
16. Engelhart, A. E., Powner, M. W. & Szostak, J. W. Functional RNAs exhibit tolerance for non-heritable 2’-5’ versus 3’-5’ backbone heterogeneity. *Nat. Chem.* **5**, 390–4 (2013).
17. Usher, D. A. & McHale, A. H. Hydrolytic Stability of Helical RNA: A Selective Advantage for the Natural 3’,5’-Bond. *Proc. Natl. Acad. Sci. U. S. A.* **73**, 1149 (1976).
18. Giurgiu, C., Li, L., O’Flaherty, D. K., Tam, C. P. & Szostak, J. W. A Mechanistic Explanation for the Regioselectivity of Nonenzymatic RNA Primer Extension. *J. Am. Chem. Soc.* **139**, 16741–16747 (2017).
19. Mariani, A. & Sutherland, J. D. Non-Enzymatic RNA Backbone Proofreading through

- Energy-Dissipative Recycling. *Angew. Chemie Int. Ed.* **56**, 6563–6566 (2017).
20. Li, L. & Szostak, J. W. The free energy landscape of pseudorotation in 3'-5' and 2'-5' linked nucleic acids. *J. Am. Chem. Soc.* **136**, 2858–2865 (2014).
 21. Rohatgi, R., Bartel, D. P. & Szostak, J. W. Nonenzymatic , Template-Directed Ligation of Oligoribonucleotides Is Highly Regioselective for the Formation of 3' - 5' Phosphodiester Bonds. *J. Am. Chem. Soc.* **118**, 3340–3344 (1996).
 22. Norberg, J. & Nilsson, L. Stacking Free Energy Profiles for All 16 Natural Ribodinucleoside Monophosphates in Aqueous Solution. *J. Am. Chem. Soc.* **117**, 10832–10840 (1995).
 23. Jafilan, S., Klein, L., Hyun, C. & Florián, J. Intramolecular Base Stacking of Dinucleoside Monophosphate Anions in Aqueous Solution. *J. Phys. Chem. B* **116**, 3613–3618 (2012).
 24. Vogel, S. R., Deck, C. & Richert, C. Accelerating chemical replication steps of RNA involving activated ribonucleotides and downstream-binding elements. *Chem. Commun.* 4922–4924 (2005) doi:10.1039/b510775j.
 25. Zhang, W. *et al.* Structural Rationale for the Enhanced Catalysis of Nonenzymatic RNA Primer Extension by a Downstream Oligonucleotide. *J. Am. Chem. Soc.* **140**, 2829–2840 (2018).
 26. 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* **5**, e17756 (2016).
 27. O'Flaherty, D. K. *et al.* Copying of Mixed-Sequence RNA Templates inside Model Protocells. *J. Am. Chem. Soc.* **140**, 5171–5178 (2018).
 28. Attwater, J., Raguram, A., Morgunov, A. S., Gianni, E. & Holliger, P. Ribozyme-catalysed

- RNA synthesis using triplet building blocks. *Elife* **7**, e35255 (2018).
29. Sulston, J., Lohrmann, R., Orgel, L. E. & Miles, H. T. Nonenzymatic synthesis of oligoadenylates on a polyuridylic acid template. *Proc. Natl. Acad. Sci. U. S. A.* **59**, 726–733 (1968).
 30. Inoue, T. & Orgel, L. E. Oligomerization of (guanosine 5'-phosphor)-2-methylimidazolidine on poly(C). An RNA polymerase model. *J. Mol. Biol.* **162**, 201–217 (1982).
 31. Mariani, A., Russell, D. A., Javelle, T. & Sutherland, J. D. A Light-Releasable Potentially Prebiotic Nucleotide Activating Agent. *J. Am. Chem. Soc.* **140**, 8657–8661 (2018).
 32. von Kiedrowski, G. Minimal Replicator Theory I: Parabolic Versus Exponential Growth. **3**, 113–146 (1993).
 33. Szathmáry, E. Simple growth laws and selection consequences. *Trends Ecol. Evol.* **6**, 366–370 (1991).
 34. Szathmáry, E. & Gladkih, I. Sub-exponential growth and coexistence of non-enzymatically replicating templates. *J. Theor. Biol.* **138**, 55–58 (1989).
 35. Lincoln, T. A. & Joyce, G. F. Self-Sustained Replication of an RNA Enzyme. *Science* (80-). **323**, 1229–1232 (2009).
 36. Paul, N. & Joyce, G. F. A self-replicating ligase ribozyme. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12733–12740 (2002).
 37. Luo, P., Leitzel, J. C., Zhan, Z.-Y. J. & Lynn, D. G. Analysis of the Structure and Stability of a Backbone-Modified Oligonucleotide: Implications for Avoiding Product Inhibition in Catalytic Template-Directed Synthesis. *J. Am. Chem. Soc.* **120**, 3019–3031 (1998).
 38. Abe, H. & Kool, E. T. Destabilizing Universal Linkers for Signal Amplification in Self-

- Ligating Probes for RNA. *J. Am. Chem. Soc.* **126**, 13980–13986 (2004).
39. Mariani, A., Bonfio, C., Johnson, C. M. & Sutherland, J. D. pH-Driven RNA Strand Separation under Prebiotically Plausible Conditions. *Biochemistry* **57**, 6382–6386 (2018).
40. Ianeselli, A., Mast, C. B. & Braun, D. Periodic Melting of Oligonucleotides by Oscillating Salt Concentrations Triggered by Microscale Water Cycles Inside Heated Rock Pores. *Angew. Chemie - Int. Ed.* **58**, 13155–13160 (2019).
41. Morasch, M. *et al.* Heated gas bubbles enrich, crystallize, dry, phosphorylate and encapsulate prebiotic molecules. *Nat. Chem.* **11**, 779–788 (2019).
42. He, C., Gállego, I., Laughlin, B., Grover, M. A. & Hud, N. V. A viscous solvent enables information transfer from gene-length nucleic acids in a model prebiotic replication cycle. *Nat. Chem.* **9**, 318–324 (2017).
43. He, C., Lozoya-Colinas, A., Gállego, I., Grover, M. A. & Hud, N. V. Solvent viscosity facilitates replication and ribozyme catalysis from an RNA duplex in a model prebiotic process. *Nucleic Acids Res.* **47**, 6569–6577 (2019).
44. Zhou, L. *et al.* Non-enzymatic primer extension with strand displacement. *Elife* **8**, e51888 (2019).
45. Szostak, J. W. An optimal degree of physical and chemical heterogeneity for the origin of life? *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **366**, 2894–901 (2011).
46. Monnard, P.-A., Kanavarioti, A. & Deamer, D. W. Eutectic Phase Polymerization of Activated Ribonucleotide Mixtures Yields Quasi-Equimolar Incorporation of Purine and Pyrimidine Nucleobases. *J. Am. Chem. Soc.* **125**, 13734–13740 (2003).
47. Verlander, M. S., Lohrmann, R. & Orgel, L. E. Catalysts for the self-polymerization of

- adenosine cyclic 2',3'-phosphate. *J. Mol. Evol.* **2**, 303–316 (1973).
48. Pascal, R., Pross, A. & Sutherland, J. D. Towards an evolutionary theory of the origin of life based on kinetics and thermodynamics. *Open Biol.* **3**, 1–9 (2013).
 49. Dickson, K. S., Burns, C. M. & Richardson, J. P. Determination of the free-energy change for repair of a DNA phosphodiester bond. *J. Biol. Chem.* **275**, 15828–15831 (2000).
 50. Phan, T. B. & Mayr, H. Comparison of the nucleophilicities of alcohols and alkoxides. *Can. J. Chem.* **83**, 1554–1560 (2005).
 51. Fahrenbach, A. C. *et al.* Common and Potentially Prebiotic Origin for Precursors of Nucleotide Synthesis and Activation. *J. Am. Chem. Soc.* **139**, 8780–8783 (2017).
 52. Islam, S., Bučar, D.-K. & Powner, M. W. Prebiotic selection and assembly of proteinogenic amino acids and natural nucleotides from complex mixtures. *Nat Chem* **advance on**, (2017).
 53. Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239–242 (2009).
 54. Yi, R., Hongo, Y. & Fahrenbach, A. C. Synthesis of imidazole-activated ribonucleotides using cyanogen chloride. *Chem. Commun.* **54**, 511–514 (2018).
 55. Zhang, S. J., Duzdevich, D. & Szostak, J. W. Potentially Prebiotic Activation Chemistry Compatible with Nonenzymatic RNA Copying. *J. Am. Chem. Soc.* **142**, 14810–14813 (2020).
 56. Liu, Z. *et al.* Harnessing chemical energy for the activation and joining of prebiotic building blocks. *Nat. Chem.* **12**, 1023–1028 (2020).
 57. Adamala, K. & Szostak, J. W. Nonenzymatic Template-Directed RNA Synthesis Inside

- Model Protocells. *Science (80-.)*. **342**, 1098–1100 (2013).
58. Yamagami, R., Bingaman, J. L., Frankel, E. A. & Bevilacqua, P. C. Cellular conditions of weakly chelated magnesium ions strongly promote RNA stability and catalysis. *Nat. Commun.* **9**, 1–12 (2018).
 59. White, H. B. Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol.* **7**, 101–104 (1976).
 60. Toparlak, D., Karki, M., Egas Ortuno, V., Krishnamurthy, R. & Mansy, S. S. Cyclophospholipids Increase Protocellular Stability to Metal Ions. *Small* **1903381**, 1–8 (2019).
 61. Gibard, C., Bhowmik, S., Karki, M., Kim, E.-K. & Krishnamurthy, R. Phosphorylation, oligomerization and self-assembly in water under potential prebiotic conditions. *Nat. Chem.* **10**, 212–217 (2018).
 62. Poudyal, R. R. *et al.* Template-directed RNA polymerization and enhanced ribozyme catalysis inside membraneless compartments formed by coacervates. *Nat. Commun.* **10**, 1–13 (2019).
 63. Oparin, A. I. The Origin of Life and the Origin of Enzymes. *Advances in Enzymology and Related Areas of Molecular Biology* 347–380 (1965)
doi:doi:10.1002/9780470122723.ch7.
 64. Jin, L., Engelhart, A. E., Zhang, W., Adamala, K. & Szostak, J. W. Catalysis of Template-Directed Nonenzymatic RNA Copying by Iron(II). *J. Am. Chem. Soc.* **140**, 15016–15021 (2018).
 65. Giurgiu, C., Wright, T. H., O’Flaherty, D. K. & Szostak, J. W. A Fluorescent G-Quadruplex Sensor for Chemical RNA Copying. *Angew. Chemie Int. Ed.* **57**, 9844–9848 (2018).

66. Vázquez-Salazar, A. *et al.* Can an Imidazole Be Formed from an Alanyl-Seryl-Glycine Tripeptide under Possible Prebiotic Conditions? *Orig. Life Evol. Biosph.* **47**, 345–354 (2017).
67. Dodson, G. & Wlodawer, A. Catalytic triads and their relatives. *Trends Biochem. Sci.* **23**, 347–352 (1998).
68. Inoue, T. *et al.* Template-directed synthesis on the pentanucleotide CpCpGpCpC. *J. Mol. Biol.* **178**, 669–676 (1984).
69. Blain, J. C. & Szostak, J. W. Progress toward synthetic cells. *Annu. Rev. Biochem.* **83**, 615–40 (2014).
70. Wu, T. & Orgel, L. E. Nonenzymatic template-directed synthesis on hairpin oligonucleotides. 3. Incorporation of adenosine and uridine residues. *J. Am. Chem. Soc.* **114**, 7963–7969 (1992).
71. Fahrenbach, A. C. Template-directed nonenzymatic oligonucleotide synthesis: lessons from synthetic chemistry. *Pure Appl. Chem.* **87**, 205–218 (2015).
72. Izgu, E. C. *et al.* Uncovering the Thermodynamics of Monomer Binding for RNA Replication. *J. Am. Chem. Soc.* **137**, 6373–6382 (2015).
73. Li, L., Lelyveld, V. S., Prywes, N. & Szostak, J. W. Experimental and Computational Evidence for a Loose Transition State in Phosphoroimidazolid Hydrolysis. *J. Am. Chem. Soc.* (2016) doi:10.1021/jacs.6b00784.
74. Walton, T. & Szostak, J. W. A Highly Reactive Imidazolium-Bridged Dinucleotide Intermediate in Nonenzymatic RNA Primer Extension. *J. Am. Chem. Soc.* **138**, 11996–12002 (2016).
75. Oró, J., Basile, B., Cortes, S., Shen, C. & Yamrom, T. The prebiotic synthesis and catalytic

- role of imidazoles and other condensing agents. *Orig. Life* **14**, 237–242 (1984).
76. LOHRMANN, R. & ORGEL, L. E. Prebiotic Activation Processes. *Nature* **244**, 418–420 (1973).
 77. Weber, A. L. Formation of pyrophosphate, tripolyphosphate, and phosphorylimidazole with the thioester, N, S-diacetylcysteamine, as the condensing agent. *J. Mol. Evol.* **18**, 24–29 (1981).
 78. Keefe, A. D. & Miller, S. L. Potentially prebiotic syntheses of condensed phosphates. *Orig. life Evol. Biosph. J. Int. Soc. Study Orig. Life* **26**, 15–25 (1996).
 79. Oró, J. & Stephen-Sherwood, E. The prebiotic synthesis of oligonucleotides. *Orig. Life* **5**, 159–172 (1974).
 80. Pongs, O. & Ts'o, P. O. P. Polymerization of unprotected 2'-deoxyribonucleoside 5'-phosphates at elevated temperature. *J. Am. Chem. Soc.* **93**, 5241–5250 (1971).
 81. Shen, C., Yang, L., Miller, S. L. & Oró, J. Prebiotic synthesis of imidazole-4-acetaldehyde and histidine. *Orig. life Evol. Biosph. J. Int. Soc. Study Orig. Life* **17**, 295–305 (1987).
 82. Grimmett, M. R. & Richards, E. L. The reaction of DL-glyceraldehyde with ammonia. *Aust. J. Chem.* **17**, 1379–1384 (1964).
 83. Ferris, J. P., Narang, R. S., Newton, T. A. & Rao, V. R. Mechanistic studies on the photochemical conversion of enamionitriles to imidazoles. *J. Org. Chem.* **44**, 1273–1278 (1979).
 84. Bigot, B. & Roux, D. Photochemical conversion of 1-amino-2-cyanoethylene to imidazole: an ab initio SCF-CI study. *J. Org. Chem.* **46**, 2872–2879 (1981).
 85. Ferris, J. P. & Orgel, L. E. An Unusual Photochemical Rearrangement in the Synthesis of

- Adenine from Hydrogen Cyanide¹. *J. Am. Chem. Soc.* **88**, 1074 (1966).
86. Zhang, L., Peng, X.-M., Damu, G. L. V, Geng, R.-X. & Zhou, C.-H. Comprehensive Review in Current Developments of Imidazole-Based Medicinal Chemistry. *Med. Res. Rev.* **34**, 340–437 (2014).
 87. Li, L. *et al.* Enhanced Nonenzymatic RNA Copying with 2-Aminoimidazole Activated Nucleotides. *J. Am. Chem. Soc.* **139**, 1810–1813 (2017).
 88. Anastasi, C., Crowe, M. A., Powner, M. W. & Sutherland, J. D. Direct assembly of nucleoside precursors from two- and three-carbon units. *Angew. Chemie - Int. Ed.* **45**, 6176–6179 (2006).
 89. Lawson, A. 63. The reaction of cyanamide with α -amino-acetals and α -amino-aldehydes. *J. Chem. Soc.* 307–310 (1956) doi:10.1039/JR9560000307.
 90. Weinmann, H., Harre, M., Koenig, K., Merten, E. & Tilstam, U. Efficient and environmentally friendly synthesis of 2-amino-imidazole. *Tetrahedron Lett.* **43**, 593–595 (2002).
 91. Lancini, G. C. & Lazzari, E. A new synthesis of alkyl and aryl 2-aminoimidazoles. *J. Heterocycl. Chem.* **3**, 152–154 (1966).
 92. Ferris, J. P., Huang, C. H. & Hagan, W. J. N-cyanoimidazole and diimidazole imine: water-soluble condensing agents for the formation of the phosphodiester bond. *Nucleosides Nucleotides* **8**, 407–14 (1989).
 93. Chen, H. *et al.* Template-directed Chemical Ligation to Obtain 3'-3' and 5'-5' Phosphodiester DNA Linkages. *Sci. Rep.* **4**, 4595 (2014).
 94. McCallum, P. B. W., Grimmett, M. R., Blackman, A. G. & Weavers, R. T. Reaction of Imidazoles with Cyanogen Bromide: Cyanation at N 1 or Bromination at C 2? *Aust. J.*

- Chem.* **52**, 159–166 (1999).
95. Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D. & Sutherland, J. D. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat. Chem.* **7**, 301–307 (2015).
 96. Szostak, J., Bartel, D. & Luisi, P. Synthesizing life. *Nature* 387–390 (2001).
 97. Crick, F. H. C. The origin of the genetic code. *J. Mol. Biol.* **38**, 367–379 (1968).
 98. Orgel, L. E. Evolution of the genetic apparatus. *J. Mol. Biol.* **38**, 381–393 (1968).
 99. Woese, C. *The Genetic Code*. (Harper and Row, 1967).
 100. Gilbert, W. Origin of life: The RNA world. *Nature* **319**, 618 (1986).
 101. Horning, D. P. & Joyce, G. F. Amplification of RNA by an RNA polymerase ribozyme. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 9786–9791 (2016).
 102. Wochner, A., Attwater, J., Coulson, A. & Holliger, P. Ribozyme-catalyzed transcription of an active ribozyme. *Science* **332**, 209–12 (2011).
 103. Szostak, J. W. The eightfold path to non-enzymatic RNA replication. *J. Syst. Chem.* **3**, 2 (2012).
 104. Kozlov, I. A. & Orgel, L. E. Nonenzymatic Template-directed Synthesis of RNA from Monomers. *Mol. Biol.* **34**, 781–789 (2000).
 105. Joyce, G. F. Nonenzymatic template-directed synthesis of informational macromolecules. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 41–51 (1987).
 106. Trevino, S. G., Zhang, N., Elenko, M. P., Lupták, A. & Szostak, J. W. Evolution of functional nucleic acids in the presence of nonheritable backbone heterogeneity. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 13492–13497 (2011).

107. Giannaris, P. A. & Damha, M. J. Oligoribonucleotides containing 2',5'-phosphodiester linkages exhibit binding selectivity for 3',5'-RNA over 3',5'-ssDNA. *Nucleic Acids Res.* **21**, 4742–4749 (1993).
108. Prakash, T. P., Roberts, C. & Switzer, C. Activity of 2',5'-Linked RNA in the Template-Directed Oligomerization of Mononucleotides. *Angew. Chemie Int. Ed.* **36**, 1522–1523 (1997).
109. Bridson, P. K. & Orgel, L. E. Catalysis of accurate poly(C)-directed synthesis of 3'-5'-linked oligoguanylates by Zn²⁺. *J. Mol. Biol.* **144**, 567–577 (1980).
110. Joyce, G. F., Inoue, T. & Orgel, L. E. Non-enzymatic template-directed synthesis on RNA random copolymers. Poly(C, U) templates. *J. Mol. Biol.* **176**, 279–306 (1984).
111. Wu, T. & Orgel, L. E. Nonenzymic template-directed synthesis on oligodeoxycytidylate sequences in hairpin oligonucleotides. *J. Am. Chem. Soc.* **114**, 317–322 (1992).
112. Inoue, T. & Orgel, L. E. Substituent control of the poly(C)-directed oligomerization of guanosine 5'-phosphorimidazolid. *J. Am. Chem. Soc.* **103**, 7666–7667 (1981).
113. Weimann, B. J., Lohrmann, R., Orgel, L. E., Schneider-Bernloehr, H. & Sulston, J. E. Template-directed synthesis with adenosine-5'-phosphorimidazolid. *Science* **161**, 387 (1968).
114. Lohrmann, R. & Orgel, L. E. Efficient catalysis of polycytidylic acid-directed oligoguanylate formation by Pb²⁺. *J. Mol. Biol.* **142**, 555–567 (1980).
115. Lohrmann, R., Bridson, P. K. & Orgel, L. E. Efficient metal-ion catalyzed template-directed oligonucleotide synthesis. *Science* **208**, 1464–1465 (1980).
116. Wu, T. & Orgel, L. E. Nonenzymatic template-directed synthesis on hairpin oligonucleotides. 2. Templates containing cytidine and guanosine residues. *J. Am. Chem.*

- Soc.* **114**, 5496–5501 (1992).
117. Agris, P. F. *et al.* Thiolation of uridine carbon-2 restricts the motional dynamics of the transfer RNA wobble position nucleoside. *J. Am. Chem. Soc.* **114**, 2652–2656 (1992).
 118. Lohrmann, R. & Orgel, L. E. Preferential formation of (2'–5')-linked internucleotide bonds in non-enzymatic reactions. *Tetrahedron* **34**, 853–855 (1978).
 119. Ferris, J. P., Hill, A. R., Liu, R. & Orgel, L. E. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* **381**, 59–61 (1996).
 120. Bowler, F. R. *et al.* Prebiotically plausible oligoribonucleotide ligation facilitated by chemoselective acetylation. *Nat. Chem.* **5**, 383–389 (2013).
 121. Sutherland, J. D. Opinion: Studies on the origin of life — the end of the beginning. *Nat. Rev. Chem.* **1**, 0012 (2017).
 122. Taylor, A. *et al.* Isothermal quadruplex priming amplification for DNA-based diagnostics. *Biophys. Chem.* **171**, 1–8 (2013).
 123. Kankia, B. I. Self-dissociative primers for nucleic acid amplification and detection based on DNA quadruplexes with intrinsic fluorescence. *Anal. Biochem.* **409**, 59–65 (2011).
 124. Ward, D. C., Reich, E. & Stryer, L. Fluorescence Studies of Nucleotides and Polynucleotides: I. FORMYCIN, 2-AMINOPURINE RIBOSIDE, 2,6-DIAMINOPURINE RIBOSIDE, AND THEIR DERIVATIVES . *J. Biol. Chem.* **244**, 1228–1237 (1969).
 125. Johnson, J., Okyere, R., Joseph, A., Musier-Forsyth, K. & Kankia, B. Quadruplex formation as a molecular switch to turn on intrinsically fluorescent nucleotide analogs. *Nucleic Acids Res.* **41**, 220–228 (2013).
 126. Joachimi, A., Benz, A. & Hartig, J. S. A comparison of DNA and RNA quadruplex

- structures and stabilities. *Bioorg. Med. Chem.* **17**, 6811–6815 (2009).
127. Adamala, K., Engelhart, A. E. & Szostak, J. W. Generation of functional RNAs from inactive oligonucleotide complexes by non-enzymatic primer extension. *J. Am. Chem. Soc.* **137**, 483–489 (2015).
 128. Zhang, J.-H. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J. Biomol. Screen.* **4**, 67–73 (1999).
 129. Kozlov, I. A. & Orgel, L. E. Nonenzymatic Template-directed Synthesis of RNA from Monomers. *Mol. Biol.* **34**, 921–930 (2000).
 130. Garcia-Diaz, M. & Kunkel, T. A. Mechanism of a genetic glissando*: structural biology of indel mutations. *Trends Biochem. Sci.* **31**, 206–214 (2006).
 131. Hawkes, S. J. All Positive Ions Give Acid Solutions in Water. *J. Chem. Educ.* **73**, 516 (1996).
 132. Powner, M. W. & Sutherland, J. D. Prebiotic chemistry: a new modus operandi. *Philos. Trans. R. Soc. London B Biol. Sci.* **366**, 2870–2877 (2011).
 133. Crick, F. H. C. The Origin of the Genetic Code. *J. Mol. Biol.* **38**, 367–379 (1968).
 134. Joyce, G. F. & Szostak, J. W. Protocells and RNA Self-Replication. *Cold Spring Harb. Perspect. Biol.* **10**, (2018).
 135. Bartel, D. P. & Szostak, J. W. Isolation of new ribozymes from a large pool of random sequences [see comment]. *Science* **261**, 1411–1418 (1993).
 136. 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 (80-.).* **292**, 1319–1325 (2001).

137. O’Flaherty, D. K. *et al.* Copying of Mixed-Sequence RNA Templates inside Model protocells. *J. Am. Chem. Soc.* **140**, 5171–5178 (2018).
138. Lazcano, A. & Miller, S. L. The origin and early evolution of life: prebiotic chemistry, the pre-RNA world, and time. *Cell* **85**, 793–798 (1996).
139. Usher, D. A. & McHale, A. H. Nonenzymic Joining of Oligoadenylates on a Polyuridylic Acid Template. *Science (80-.)*. **192**, 53 (1976).
140. Lutay, A. V., Zenkova, M. A. & Vlassov, V. V. Nonenzymatic recombination of RNA: Possible mechanism for the formation of novel sequences. *Chem. Biodivers.* **4**, 762–767 (2007).
141. Biron, J.-P., Parkes, A. L., Pascal, R. & Sutherland, J. D. Expeditious, potentially primordial, aminoacylation of nucleotides. *Angew. Chem. Int. Ed. Engl.* **44**, 6731–6734 (2005).
142. Kim, H.-J. *et al.* Evaporite Borate-Containing Mineral Ensembles Make Phosphate Available and Regiospecifically Phosphorylate Ribonucleosides: Borate as a Multifaceted Problem Solver in Prebiotic Chemistry. *Angew. Chemie Int. Ed.* n/a-n/a (2016) doi:10.1002/anie.201608001.
143. Hu, J., Lei, W., Wang, J., Chen, H.-Y. & Xu, J.-J. Regioselective 5’-position phosphorylation of ribose and ribonucleosides: phosphate transfer in the activated pyrophosphate complex in the gas phase. *Chem. Commun.* **55**, 310–313 (2019).
144. Schwartz, A. W. & Orgel, L. E. Template-Directed Synthesis of Novel, Nucleic Acid-Like Structures. *Science (80-.)*. **228**, 585–587 (1985).
145. Visscher, J., Bakker, C. G., van der Woerd, R. & Schwartz, A. W. Template-directed oligomerization catalyzed by a polynucleotide analog. *Science (80-.)*. **244**, 329–331

- (1989).
146. Chen, I. A., Salehi-Ashtiani, K. & Szostak, J. W. RNA catalysis in model protocell vesicles. *J. Am. Chem. Soc.* **127**, 13213–13219 (2005).
 147. Jardine, A. M. *et al.* Cleavage of 3',5'-pyrophosphate-linked dinucleotides by ribonuclease A and angiogenin. *Biochemistry* **40**, 10262–10272 (2001).
 148. Konarska, M., Filipowicz, W., Domdey, H. & Gross, H. J. Formation of a 2'-phosphomonoester, 3',5'-phosphodiester linkage by a novel RNA ligase in wheat germ. *Nature* **293**, 112–116 (1981).
 149. Anderson, B. A. & Krishnamurthy, R. Heterogeneous Pyrophosphate-Linked DNA–Oligonucleotides: Aversion to DNA but Affinity for RNA. *Chem. – A Eur. J.* **24**, 6837–6842 (2018).
 150. Phillips, R., Eisenberg, P., George, P. & Rutman, R. J. Thermodynamic data for the secondary phosphate ionizations of adenosine, guanosine, inosine, cytidine, and uridine nucleotides and triphosphate. *J. Biol. Chem.* **240**, 4393–4397 (1965).
 151. Rajamani, S. *et al.* Effect of Stalling after Mismatches on the Error Catastrophe in Nonenzymatic Nucleic Acid Replication. *J. Am. Chem. Soc.* **132**, 5880–5885 (2010).
 152. Zhang, W., Tam, C. P., Wang, J. & Szostak, J. W. Unusual base-pairing interactions in monomer-template complexes. *ACS Cent. Sci.* **2**, 916–926 (2016).
 153. Westheimer, F. H. Why nature chose phosphates. *Science (80-.)*. **235**, 1173 LP – 1178 (1987).
 154. Rodriguez, L. & Orgel, L. E. Pyrophosphate formation as the most efficient condensation reaction of activated nucleotides. *J. Mol. Evol.* **32**, 101–104 (1991).

155. Oivanen, M., Kuusela, S. & Lonnberg, H. Kinetics and Mechanisms for the Cleavage and Isomerization of the Phosphodiester Bonds of RNA by Bronsted Acids and Bases. *Chem. Rev.* **98**, 961–990 (1998).
156. Sosson, M. & Richert, C. Enzyme-free genetic copying of DNA and RNA sequences. *Beilstein J. Org. Chem.* **14**, 603–617 (2018).
157. Wu, T. & Orgel, L. E. Nonenzymatic template-directed synthesis on oligodeoxycytidylate sequences in hairpin oligonucleotides. *J. Am. Chem. Soc.* **114**, 317–322 (1992).
158. Kozlov, I. A. *et al.* Nonenzymatic synthesis of RNA and DNA oligomers on hexitol nucleic acid templates: the importance of the A structure. *J. Am. Chem. Soc.* **121**, 2653–2656 (1999).
159. Kim, S. C. *et al.* A Model for the Emergence of RNA from a Prebiotically Plausible Mixture of Ribonucleotides, Arabinonucleotides, and 2'-Deoxynucleotides. *J. Am. Chem. Soc.* **142**, 2317–2326 (2020).
160. Ye, J.-D., Li, N.-S., Dai, Q. & Piccirilli, J. A. The Mechanism of RNA Strand Scission: An Experimental Measure of the Brønsted Coefficient, β_{nuc} . *Angew. Chemie Int. Ed.* **46**, 3714–3717 (2007).
161. Velikyan, I., Acharya, S., Trifonova, A., Földesi, A. & Chattopadhyaya, J. The pKa's of 2'-Hydroxyl Group in Nucleosides and Nucleotides. *J. Am. Chem. Soc.* **123**, 2893–2894 (2001).
162. Takagi, Y. & Taira, K. Analyses of kinetic solvent isotope effects of a hammerhead ribozyme reaction in NH_4^+ and Li^+ ions. *Nucleic Acids Res. Suppl.* 273–274 (2002) doi:10.1093/nass/2.1.273.
163. Barton, D. H. R. The conformation of the steroid nucleus. *Experientia* **6**, 316–320 (1950).

164. Plavec, J., Tong, W. & Chattopadhyaya, J. How do the gauche and anomeric effects drive the pseudorotational equilibrium of the pentofuranose moiety of nucleosides? *J. Am. Chem. Soc.* **115**, 9734–9746 (1993).
165. Thibaudeau, C., Plavec, J., Garg, N., Papchikhin, A. & Chattopadhyaya, J. How Does the Electronegativity of the Substituent Dictate the Strength of the Gauche Effect? *J. Am. Chem. Soc.* **116**, 4038–4043 (1994).
166. Guschlbauer, W. & Jankowski, K. Nucleoside conformation is determined by the electronegativity of the sugar substituent. *Nucleic Acids Res.* **8**, 1421–1433 (1980).
167. Schöning, K.-U. *et al.* Chemical Etiology of Nucleic Acid Structure: The α -Threofuranosyl-(3'→2') Oligonucleotide System. *Science (80-.)*. **290**, 1347 LP – 1351 (2000).
168. Zhang, W. *et al.* Structural interpretation of the effects of threo-nucleotides on nonenzymatic template-directed polymerization. *Nucleic Acids Res.* **49**, 646–656 (2021).
169. Zhang, W. *et al.* Structural Rationale for the Enhanced Catalysis of Nonenzymatic RNA Primer Extension by a Downstream Oligonucleotide. *J. Am. Chem. Soc.* **140**, 2829–2840 (2018).
170. Bratsch, S. G. A group electronegativity method with pauling units. *J. Chem. Educ.* **62**, 101–103 (1985).

