Towards Self-Replicating Informational Polymers

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th>Citation</th>
<th>Prywes, Noam. 2016. Towards Self-Replicating Informational Polymers. Doctoral dissertation, Harvard University, Graduate School of Arts &amp; Sciences.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessed</td>
<td>August 7, 2017 12:14:12 AM EDT</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:33493609">http://nrs.harvard.edu/urn-3:HUL.InstRepos:33493609</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>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 <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>

(Article begins on next page)
Towards self-replicating informational polymers

A dissertation presented

by

Noam Prywes

to

The Department of Chemistry and Chemical Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

In the subject of

Chemistry

Harvard University

Cambridge, Massachusetts

May, 2016
Towards self-replicating informational polymers

Abstract

The capability to transmit information from generation to generation is an essential feature of life. In all terrestrial life, DNA and RNA contain information in the form of a sequence of monomers and are copied in every generation. The RNA world hypothesis posits that there was a time in the history of life when all cellular functions were accomplished by RNA catalysts. However, the initial emergence of those RNA catalysts is not yet fully understood.

Nonenzymatic RNA polymerization has been proposed as a potential stepping-stone from prebiotic chemistry to the RNA world. In searching for alternatives to the chemically trapped triphosphate nucleotides found in modern biology, chemical modifications of RNA have been discovered that allow for the copying of RNA. However, the copying of sequences rich in adenine and uracil residues remains a significant challenge. In chapter 2 we use chemically activated oligonucleotides as catalysts to copy all four monomers sequentially, potentially creating a route for the copying of any sequence without a polymerase and contributing to a model for the emergence of evolution.

Replacing uracil with 2-thiouracil and 2-thiothymine, modified forms of uracil found in modern life, has proven to improve the reactivity and fidelity of nonenzymatic
RNA polymerization. In chapter 3, we tested these alternatives to uracil as substrates and components of an RNA polymerase ribozyme. We discovered that they were superior in the context of ribozyme mediated RNA polymerization both in terms of faster rate and higher fidelity. We then synthesized ribozymes in which every instance of uracil was replaced by either 2-thiouracil or 2-thiothymine and found them to retain some activity. We hypothesize that these alternative nucleobases could have conferred significant benefits to early life forms.

To explore alternative genetic systems to DNA and RNA, in chapter 4 of this thesis we synthesized an organic-soluble copolymer with two different monomers capable of reversible covalent bond formation with one another. We show that information copying is possible with this polymer by synthesizing a polymer with a sequence complementary to a template while it is still covalently bound to that template, demonstrating that nucleic acids are not the only molecules capable of information storage and replication.

Together, these results assist in the construction of artificial life and expand the possibilities for the emergence of life.
# Table of contents

Abstract iii

Table of contents v

Acknowledgements vii

## Chapter 1: Introduction
1. Introduction 1
2. Polymer backbone 8
3. Nucleobase 23
4. Leaving group 28
5. Nucleophile 36
6. Contextual modifications 40
7. Prospective 47
8. References 49

## Chapter 2: Catalysis of nonenzymatic copying of RNA templates by activated helper oligonucleotides 59
1. Abstract 59
2. Introduction 60
3. Results and discussion 61
Chapter 3: Thiolated uridine substrates and templates improve the rate and fidelity of ribozyme catalyzed RNA copying

3.1: Abstract
3.2: Introduction
3.3: Results and discussion
3.4: Materials and methods
3.5: References

Chapter 4: A novel, self-replicating, informational polymer

4.1: Abstract
4.2: Introduction
4.3: Results and discussion
4.4: Materials and methods
4.5: References

Chapter 5: Conclusion

5.1: Summary of findings
5.2: Future directions
5.3: Prospective
5.4: References
Acknowledgements

First, I would like to thank my advisor, Jack Szostak. Jack has provided more support than I rightly could have expected. His shocking experimental insights over the years would have been enough without the patience and kindness with which they were offered. Jack made himself extremely available to talk about science; I frequently abused his open door policy to discuss various experimental peculiarities or the latest scientific literature. Most of all, Jack demonstrates how joyous a life of scientific curiosity can be.

I am deeply indebted to many of the other scientists in the laboratory. To Ben Heuberger, for getting me started and believing in my ideas from the beginning; to Aaron Engelhart and Matt Powner, for showing me how to collaborate and how to achieve big things quickly; to Seung Soo Oh and Ayan Pal, for essential help with work in the third chapter of this thesis; to Christian Heinrich, for his literary and musical taste, as well as helpful German translation; to Li Li for his encouragement, assistance and conversations; to Neha Kamat and Tivoli Olsen, for being constant friends and mentors; to Pam Svec, for being a cheerful presence and making the work of every lab member possible; to Fanny Ng for continuing that essential work; to Craig Blain, for pioneering much of the difficult mechanistic work in our lab and for bringing a level of scientific rigor to bear that served as an anchor in my work; to Itay Budin, for showing
me that for some people, everything they touch turns to gold; and, finally, to Tony Jia, for helping me think through problems, editing and Japanese translation.

I have had the great honor of mentoring a series of talented rotation students and undergraduates. The rotation students, Fred Rubino, Alix Chan, Daniel Strassfeld, Holly Rees and Jeff Bessen, all substantively contributed to this dissertation and, in many cases, changed the direction of our experiments within the all too short month I had to work with each of them. I was fortunate to have more time with each of the undergraduates with whom I worked. Dipti is a natural born scientist who always makes sure to let her curiosity lead her. Alison was capable of completing any task effectively with cheer and always ready for the next challenge. Francesca’s dedication was inspiring and her determination in the face of experimental difficulty unmatched. Finally, Yale did not need a mentor, and I did not serve as one, we were, and remain, colleagues. Our ongoing discussions on scientific literature always leave me freshly excited about the enterprise of science.

I would like to thank my friend and roommate David Westwood for making me feel supported and at home, for the haircuts and for introducing me to aspects of Go I didn’t know existed. I frequently leaned on Carolyn Brotherton for her unparalleled and sober synthetic acumen. I am also grateful for her example that a degree in chemistry can lead to unexpectedly exciting careers. I would like to thank Emily Ricq for her excitement to talk about science and her boundless talent and creativity.
I count myself uniquely fortunate to have lived a year at Nashton and to have known its fine residents. My conversations with Miriam Huntley sitting high above New Hampshire are ones I will cherish my whole life. Her hard work is an inspiration, and her keen intellect an object of envy. Coleman Connelly enriched and broadened my life by bringing to our conversations a perspective so deep and so broad that all I could do was lose myself in his wisdom. Anders Hansen showed me how dangerous a sharp mind is when paired with a strong work ethic; especially since he seems to be having so much fun producing such beautiful science. Joshua Klobas and Dougal Maclaurin, each in very different ways, followed their curiosity along intellectual switchbacks and made amazing advances in the process. Alexandria Symonds and Hillary Busis hosted me in their singular apartment in the Lower East Side, which became a respite and my home away from home.

Kyle Strom believed in the idea that became the fourth chapter of this thesis and without him it would have been nothing more than that. Thank you for making my dreams come true. I have never had the opportunity to work with a colleague as a true peer and the experience has been overwhelming. I am so proud of what we have done.

Most of all, I am thankful to everyone who came to the hundreds of meetings of the ASAASA. The community we built continues to be the most meaningful intellectual space in my life.
Finally, I am immeasurably thankful for the love and support of my family. I am always moved by my father’s example of a life in science and his unwavering commitment to experimentalism. His advice has also been essential throughout my studies. I am thankful to my mother for finding joy in her academic pursuits as well as for her constant love and encouragement. I am thankful to my brother for his wit, and for knowing me so well. Your examples of lives lived in the pursuit and love of knowledge and truth are almost more than I can bear.
CHAPTER 1

The replication of informational polymers before enzymes

1.1 Abstract

Transmission of genetic information is coordinated and catalyzed by highly evolved protein enzymes in extant organisms on Earth. How this complex process occurred before the advent of enzymes is a puzzle in origins of life research. Whatever the first genetic material was, it must have initially replicated without the assistance of genetically encoded enzymes – that is to say, in a manner very different from modern, protein based replication. Limited nonenzymatic copying of genetic polymers has been demonstrated in a number of contexts, but many challenges remain with respect to robustness, generality, yield, and fidelity. A major research focus in this field has been on developing monomers capable of assembling in a sequence specific fashion on a template. In this review, we highlight the most notable advances in efforts to construct a self-replicating informational polymer. In particular, we compare the plausibility of scenarios in which RNA arose directly from prebiotic chemistry with scenarios in which an alternative informational polymer arose first and was later
replaced by RNA. We conclude that, although many challenges remain to be overcome in prebiotic RNA synthesis and replication, the lack of a superior alternative to RNA leaves it as the most promising candidate for the first genetic polymer.

1.2 Introduction

The overarching goal of origin of life research is to understand the order of chemical events on the early Earth that led to primitive biology. One important piece of the puzzle is the chemistry that gave rise to the replication of the hereditary materials RNA and DNA. In searching for an answer, researchers noted that RNA acts both as an information carrier (Volkin 1956) and as a catalyst (Kruger 1982, Guerrier-Takada 1983). It has been inferred that RNA likely preceded DNA and proteins in those roles in early life. This theory is known as the RNA world hypothesis (Gilbert 1986). Accordingly, the biochemical processes associated with RNA have become central figures in theories of the origin of life.

The discovery of functional RNAs helps to explain evolution before the advent of proteins but fails to explain how RNA replication began. No polymerase ribozyme has yet been found or developed that is capable of self-replication (Szostak 2012), but even if one existed it would still pose a paradox: in order for a replicase ribozyme to exist in the RNA world, this ribozyme itself
must be the product of a polymerization reaction (Robertson 2012). This raises the question, how was the first replicase produced? One possibility is that the first replicase was the product of untemplated, random chemical polymerization of nucleotides and it survived by copying a second similarly produced ribozyme. Since this scenario relies on two exceedingly unlikely events occurring in close proximity to one another, it has generally been regarded as unlikely.

1.2.1 Nonenzymatic replication of RNA

The alternative to the “replicase first” model is nonenzymatic RNA template copying. Since the 1960s (Pollmann 1961, Weimann 1968) single-stranded poly(C) RNA has been known to be capable of catalyzing the assembly of short complementary oligonucleotide strands from activated guanosine mononucleotides in the absence of any enzymes. This provides a model for early RNA replication where the only catalyst is the template being copied. Paired with renewed interest in the prebiotic synthesis of RNA monomers (Powner 2009), a pathway from small abiotic molecules to self-replicating RNA seems increasingly plausible.

In practice, nonenzymatic RNA replication faces serious challenges. The rate of nonenzymatic template-directed polymerization is slow due to the
limited reactivity of the activated RNA nucleoside monophosphates (Joyce 1987). Furthermore, there is a troubling imbalance in the rates of polymerization for each nucleotide; for example, while a G monomer can polymerize on a template of poly(C) within hours, a U monomer cannot polymerize at all on poly(A) before the template begins to succumb to hydrolysis (Wu 1992a,b,c). These problems, among others (e.g. fidelity, annealing, regiospecificity) (Szostak 2012), must be addressed in order to validate the plausibility of nonenzymatic polymerization as a mechanism for RNA copying prior to the emergence of replicase.

1.2.2 Criteria for the evaluation of alternatives to RNA

In order to circumvent the various problems posed by nonenzymatic RNA replication, the idea that RNA arose directly from prebiotic chemistry, “RNA-first” (Robertson 2012) has been challenged by the theory that an alternative polymer arose first and later disappeared as it was replaced by RNA as the genetic polymer in life – “RNA-later” (Hud 2013).

On the grounds of simplicity, the “RNA-first” view is at an inherent advantage to the “RNA-later” view because the “RNA-first” view does not require any novel chemical species. The lack of evidence of a precursor to RNA indicates either a complete displacement by RNA or else that no such precursor ever existed. Any
alternative to RNA must be shown to be fundamentally more prebiotically plausible than RNA. Naturally, the point at which the parsimony of the “RNA-first” theory breaks down in the face of a plausible alternative polymer is a matter of debate.

In assessing the plausibility of progenitors to RNA, four criteria must be met (Fig. 1.1) (Carothers 2006). First, the chemical synthesis of the alternative polymer must be compatible with the prebiotic chemistry of the early Earth, at least to the same extent as, and ideally more so, than RNA (Fig. 1.1a). This includes compatibility with the compartmentalization system of early cells (Carothers 2006). Necessarily, the synthesis of alternatives to RNA will be less studied, but some synthetic advantage with respect to RNA would make a strong case for such a polymer.

Second, the alternative polymer must be capable of promoting template copying by sequence-dependent, nonenzymatic polymerization (Fig. 1.1b). Despite the technical difficulties associated with achieving multiple rounds of nonenzymatic polymerization of arbitrary sequences of RNA (Robertson 2012, Szostak 2012), the fact that RNA is capable of nonenzymatic polymerization at all (Sulston 1968, Wu 1992a,b,c) is an example of a chemical predisposition that is highly suggestive. Thus, any proposed progenitor to RNA must be capable of nonenzymatic template copying at least to the same extent seen with RNA. Ideally any proposed alternative would not suffer from the challenges RNA
faces in nonenzymatic polymerization – any advantage in this respect is
evidence in favor of the “RNA-later” view.

Figure 1.1: Criteria for alternatives to RNA. (a) Prebiotic synthesis of genetic polymer
components. (b) Nonenzymatic, templated polymerization of a polymer. (c) Folding of an
information containing polymer into an active catalyst in a sequence-defined manner. (d)
Transfer of information from one polymer backbone structure (black) to another (pink).
Third, any alternative to RNA must be able to fold into a conformation that has an evolutionarily selectable function, either catalytic or structural (Fig. 1.1c). If a polymer only contains information and cannot catalyze chemical reactions then the selective advantage it would provide to early life forms would not depend on the information it contains. This would preclude evolution since there would be no connection between genotype and phenotype. Since RNA is capable of facilitating a wide array of chemical transformations (e.g., formation or cleavage of phosphodiester bonds (Kruger 1982, Johnston 2001), formation of peptide bonds (Zhang 1997), as well as binding to small ligands (Ellington 1990), to name a few) alternative polymers that are incapable of folding into catalysts would make for poor evolutionarily progenitors.

Finally, the alternative polymer backbone should ideally be able to form a Watson-Crick base-paired duplex with RNA. That way, genetic information can be transferred through base pairing from the proposed progenitor to RNA (Eschenmoser 1999). Base pairing between the alternative polymer and RNA could allow for the nonenzymatic polymerization of RNA in a template-directed fashion (Fig. 1.1d). If base pairing with RNA were not possible, RNA would have had to develop without the benefit of a template on which to polymerize and there would be no way to transfer information from the progenitor to RNA.
In this review we highlight major advances towards nonenzymatic, template-directed polymerization of high-energy monomers with the objective of understanding the chemistry required for information transfer on the early Earth. To this end, we have categorized nucleotide modifications based on: (1.3) the polymer backbone, (1.4) the nucleobase (NB), (1.5) the leaving group (LG), and (1.6) the nucleophilic group. The role of contextual modifications to experimental conditions will be discussed in section (1.7).

1.3 Polymer backbone

It has now been demonstrated that the nucleic acid polymers known to be capable of undergoing Darwinian evolution are not limited to RNA and DNA. Recent breakthroughs from the Chaput (Yu 2012) and Holliger (Pinheiro 2012) laboratories showed that alternative nucleic acids can also be effective for the storage and transfer of genetic information as well as for the evolution of novel functions including ligand binding and catalysis. A wide array of alternative backbones have been studied because, while even subtle alterations in nucleobase structure significantly affect base pairing and stacking, backbone heterogeneity is, in many cases, tolerated by DNA and RNA as heteroduplex partners.

In this section we first review the chemical attributes relevant to the role of RNA in early life and then present a series of notable alternatives which we compare
to RNA by the criteria laid out above.

1.3.1 RNA

RNA has been studied for its potential as the first genetic polymer since the 1960s when it was first proposed as a progenitor to DNA as the informational repository of life (Orgel 1968, Crick 1968, Woese 1968). The capacity of RNA to store information and to act as a genetic polymer was recognized long before its catalytic capabilities which were discovered decades later. By the 1980s, after the discovery of RNaseP (Guerrier-Takada 1983) and self-splicing introns (Kruger 1982), it was widely accepted that there existed an “RNA world” where RNA both transmitted information and served as a catalyst (Gilbert 1986). Even though RNA displays remarkable functional versatility, the pathway from prebiotic chemistry to the RNA world remains unclear. Specifically, although ribozymes of any required function are likely evolvable, evolution requires replication, and replication without a ribozyme catalyst remains problematic.

Two steps between the abiotic early earth and the RNA world have been the focus of much of the research in the origin of life field over the last 50 years: how RNA monomers could have formed from prebiotic chemistry and how RNA monomers could copy RNA oligomers without the assistance of enzymes. The “RNA-first” view of the origin of life must supply satisfying demonstrations of each of these steps in
order to be accepted; until that happens, “RNA-later” theories will be able to point to large holes in the “RNA-first” theory as a motivating factor.

To address the issue of monomer synthesis, initially, the retrosynthetic strategy most widely explored was one in which ribose is synthesized separately from the nucleobases followed by their condensation into complete nucleosides. Small amounts of ribose (Benner 2012, Shapiro 1988) and the canonical nucleobases (Oró 1960, Basile 1984, Robertson 1995, Shapiro 1999) can be readily synthesized from formaldehyde and cyanide (albeit as minor components in complex mixtures), but no efficient condensation has yet to be discovered. Among the nucleobases, only adenine and guanine could be condensed with ribose and, even then, at poor yields (<10%) (Fuller 1972). To this day, the prebiotic synthesis of ribose and the nucleobases remains an active field of study (Meinert 2016), partly because of the fact that nucleobases are found in prebiotic meteorites (Callahan 2011). The presence of lead (Zubay 1998) and borate (Ricardo 2004) have both been found to change the equilibrium distribution of the products of the formose reaction to favor ribose. However, neither ribose (Benner 2012) nor the natural nucleobases (Fuller 1972, Oró 1960) have ever been produced with high yield and purity; they are always found as minor products in highly heterogeneous mixtures.

Recent improvements to an alternative retrosynthetic analysis have allowed for the synthesis of the C and U 2′-3′ cyclic phosphate ribonucleotides, without ever
producing the free nucleobases or ribose, in five steps in much higher yields (Powner 2009). The involvement of multiple steps has left some unconvinced (Benner 2012), and the absence of a similar synthesis for A and G is a continuing challenge. Nevertheless, the remarkable success of the new approach has breathed new life into the field.

Even if a robust synthesis of all of the individual RNA monomers were to be discovered, nonenzymatic copying of RNA sequences would still have to be demonstrated. Orgel and co-workers (Sulston 1968) first successfully produced poly(G) sequences on poly(C) templates by employing heterocyclic leaving groups. The limitations of nonenzymatic RNA polymerization were explored systematically in 1992 when it was shown that templates that contained two or more adjacent A or U bases were exceedingly difficult to copy (Wu 1992a,b,c). This, in addition to the discovery that the enantiomeric form of the G monomer acts as a covalent chain terminator for polymerization, caused many researchers to search for alternatives to RNA (Joyce 1987).

Widespread acceptance of the “RNA-first” theory would seem to require at a minimum simpler pathways to both the synthesis of the nucleotide monomers and nonenzymatic template directed polymerization. Syntheses of the purine nucleotides must be demonstrated (and the pyrimidine syntheses simplified if possible) and an explanation of the transition from the 2′-3′ cyclic phosphate of the apparent early
monomers to the 5’ phosphorylated mononucleotides used by modern life must be found. Nonenzymatic synthesis of arbitrary RNA templates must be demonstrated over multiple rounds of replication, ideally within a vesicle. If these goals are achieved, any “RNA-later” theory will be left with fewer arguments based on the intractability of RNA as the first genetic polymer.

1.3.2 DNA

Although the use of DNA in life is widely considered to be a later evolutionary adaptation, due to its relative stability to hydrolysis as compared to RNA (Robertson 2012), several recent results complicate this belief. While it is true that DNA monomers are synthesized from RNA monomer precursors in modern life (Benner 1989), it is possible that DNA and RNA coexisted on the early earth. A recent study (Powner 2012) demonstrated a partial synthesis of a DNA nucleotide through some of the same steps that were reported for RNA (Powner 2009). Nonenzymatic polymerization of DNA monomers on a DNA template has been demonstrated (Kozlov 1999b) and enzymes with DNA backbones (DNAzymes) are commonly evolved (Breaker 1994, Joyce 2009, Ponce-Salvatierra 2016).

Though DNA fails to measure up to RNA by each of the above criteria, it is still notable for not strictly failing any one of them. The lack of DNAzymes in modern biology and the biosynthetic pathway of DNA monomers strongly indicate that the
wholesale storage of genetic information in DNA happened in a post-RNA world. However, DNA and RNA are highly compatible: certain aptamers composed of the same sequence of DNA or RNA can both bind specifically to the same ligands, albeit with different $K_d$ values. Those same aptamers, composed of a mixture of DNA and RNA monomers can also bind their target ligands specifically (Trevino 2011). Thus, there may be some room in the RNA world for DNA: it is possible that a progenitor to RNA existed that was a heterogeneous mix or DNA and RNA nucleotides. Such heterogeneous mixtures may also have included other sugars (e.g. ANA, see below) as well as RNA with both 2′-5′ and 3′-5′ linkages (Engelhart 2012). Deciding between RNA and DNA will perhaps depend on which is easier to synthesize prebiotically. If an equivalent synthesis were found for each, it would be appropriate to conclude that they may have coexisted.

1.3.3 Six membered sugars

Motivated by a desire to systematically explore alternatives to RNA, the Eschenmoser laboratory synthesized a series of alternative sugar-phosphate backbone polymers and assessed them based on their ability to base pair with one another and with RNA (Eschenmoser 1999, 2011). Remarkably, many nucleic acids based on alternative sugars base-paired both to themselves and to RNA and DNA. Because nucleotides with alternative sugars are all structurally very similar to ribonucleotides,
there is little reason to believe that any one of them is easier to synthesize prebiotically. However, depending on the prebiotic synthetic pathway, four and six carbon sugars can be formed by the condensation of two identical precursor molecules containing two or three carbons respectively, while any five carbon sugar must be formed from the condensation of at least two different precursors. If the synthetic pathway is one of individual carbons adding, as it is in the formose reaction, then this advantage disappears.

Many of the synthetic nucleic acids displayed unexpected properties, for example, oligonucleotides with a pyranosyl RNA (p-RNA, an isomer of RNA) backbone (Fig. 1.2) can form Watson-Crick double helices that are more thermally stable than RNA (Pitsch 1995, 2003). Additionally, tetranucleotide oligomers of p-RNA with terminal 2', 3' cyclophosphates were shown to ligate in a template directed fashion to produce longer sequences. However, p-RNA oligomers fail to form stable duplexes with complementary RNA. Therefore, any information stored in p-RNA could not be directly transferred to RNA (Eschenmoser 1999). Furthermore, the high level of structural rigidity of p-RNA does not bode well for its utility as a catalyst (Bolli 1997). Due to a lack of the enzymes required for molecular evolution of this backbone, work exploring it as an alternative to RNA has largely ended in favor of more tractable backbones that can pair with RNA.
Hexitol and altritol closely related hexose sugars that can be used to generate nucleic acids with backbones similar to p-RNA (Fig. 1.2). They were both studied by the Orgel laboratory which found that, unlike p-RNA, both HNA and ANA form stable heteroduplexes with DNA and RNA (Van Aerschot 1995, Kozlov 2000). Because HNA/RNA and ANA/RNA heteroduplexes are more rigidly constrained to an A-form helix conformation than RNA/RNA homoduplexes, HNA and ANA act as better templates than RNA for the nonenzymatic polymerization of RNA monomers (Kozlov 1999a,b,c). Based on these observations, it would seem that HNA or ANA might be
excellent alternatives to RNA because they are superior nonenzymatic polymerization substrates while being of a similar structure to RNA and thus equally prebiotically plausible. However, crucially, imidazole-activated monomers of ANA and HNA did not oligomerize efficiently on any backbone (Kozlov 1999c) for unknown reasons, perhaps due to subtle conformational effects. Thus, even though enzymes composed of HNA have been evolved (Taylor 2014, 2015) and likely could be with ANA, these backbones are much less likely than RNA to have been the primordial genetic polymer because they appear to be unable to copy themselves by nonenzymatic polymerization.

1.3.5 Threofuranose Nucleic Acid (TNA)

Perhaps the most surprising alternative sugar-phosphate backbone nucleic acid discovered by Eschenmoser and coworkers is l-threofuranose nucleic acid (TNA, Fig. 1.2), a polymer derived from a tetrose sugar (Schöning 2000), which has a repeat unit that has one fewer atom in it than DNA or RNA. It came as a surprise that, despite having a shorter backbone, TNA is capable of forming stable Watson-Crick duplexes with complementary TNA, DNA and RNA sequences. Importantly, thermal melting experiments showed that TNA-RNA duplexes are significantly more stable than TNA-DNA duplexes, presumably because TNA strictly adopts an A-form-like helical geometry (Ebert 2008, 2010). In 2006, Heuberger and Switzer reported nonenzymatic
polymerization of guanosine 5′-phosphoro(2-methyl) imidazolide (2-MeImpG) assisted by a polycytosine-TNA hairpin template (Heuberger 2006). Blain et al. demonstrated the polymerization of 2′-amino-guanosine 5′-phosphoro-(2-methyl) imidazolide (2′-NH₂-2-MeImpG) monomers to form N2′-P3′-linked phosphoramidate TNA on RNA, DNA, and TNA templates. Unexpectedly, TNA oligonucleotide templates were found to be a less effective than both DNA and RNA with regard to the polymerization rate. The poor performance of the TNA template was attributed, ironically, to the constrained backbone, the source of its tight binding to RNA oligomers: since TNA oligomers are shorter than RNA, they must be in a maximally extended (and perhaps less favorable) conformation in order to base pair with RNA monomers.

TNA has also been used as an alternative system in enzymatic information transfer. Notably, chimeric DNA-TNA templates have been shown to allow DNA polymerase-mediated TNA synthesis (Chaput 2003, Yu 2012, 2013, Pinheiro 2012, Dunn 2014), albeit with limited efficiency (Ichida 2005). The efficiency was improved when guanine was replaced by 7-deazaguanine to prevent spurious G:G mispairing (Dunn 2015). Chaput and co-workers have succeeded in the in vitro evolution a TNA aptamer composed of a three-letter genetic alphabet (Yu 2012), which can bind to human thrombin with an activity similar to analogous RNA aptamers (Bock 1992, Kubik 1994).
TNA is therefore capable, like RNA, of forming folded catalysts and somewhat less advantageous than RNA in nonenzymatic polymerization. Since threose is a sugar, like ribose, and is perhaps no more or less likely than ribose to be the product of prebiotic chemistry, we might therefore conclude that TNA is not likely to have preceded RNA as the genetic material of life on Earth. The fact that threose has one fewer carbon than ribose means that it could be strictly “simpler” from a synthetic standpoint, and thus at a distinct advantage to RNA as a product of prebiotic chemistry (Orgel 2000). On the other hand, the ease of synthesis of a sugar in a prebiotic setting may not simply be a function of the complexity of the sugar (Anastasi 2007). Attempts to synthesize TNA by the same prebiotic route as RNA have not been successful (Islam 2013) and the glycosylation reaction between the canonical nucleobases and threose may be impossible. As prebiotic syntheses of sugars and nucleotides are developed it may become apparent if the structures of ribose or threose are more predisposed to be synthesized on the early Earth.

1.3.6 Xeno- Nucleic Acid (XNA)

The observation that DNA and RNA can form active enzymes raises the question of what other backbones have the ability to evolve into folded catalysts. Recently, the Holliger laboratory has developed polymerases capable of polymerizing
a series of alternative sugar backbones collectively referred to as xeno-nucleic acids (XNAs) on DNA templates (Pinheiro 2012). These versatile polymerases allowed them to select for endonuclease ribozymes from libraries of HNA, CeNA, LNA (locked) ArNA (arabinose) and FANA (Fig. 1.2) (Taylor 2014, 2015). Similarly, the Chaput laboratory developed the tools for (Yu 2013) and performed (Yu 2012, Dunn 2014) in vitro evolution with a TNA backbone (see above). This work demonstrated that once novel nucleic acids capable of base pairing with DNA are identified, tools for their evolution in the laboratory can be developed and the capacity of the new nucleic acid for function can be quickly demonstrated. Nonenzymatic polymerization and the potential for prebiotic synthesis have not been explored with these alternative backbones (with the exceptions of TNA and HNA, see above); this would be the next step in testing the feasibility of CeNA or ArNA as the first genetic polymers (FANA out of contention because the fluorine atom renders the prebiotic synthesis untenable and LNA because of its overly complex sugar structure). Perhaps the greatest contribution of this work is the fact that all of the studied backbones showed function, which suggests that functionality is not a rare trait among nucleic acids in alternative backbones and that demonstrations of prebiotic synthesis and nonenzymatic polymerization are more important, and difficult, tests for candidate pre-RNA polymers.
1.3.7 Glycol Nucleic Acid (GNA)

Motivated by the same logic that suggests that TNA may be easier to synthesize prebiotically than RNA because it is a simpler sugar, the three-carbon sugar, glycerol, has been studied as an alternative backbone. In 2005, Meggers and co-workers reported the synthesis of glycol nucleic acid (GNA, Figure 1.2), an acyclic propylene glycol obtained from glycidol (Zhang 2005). Similar three-carbon sugars have been observed as intermediates in prebiotic chemistry reactions (Patel 2015). It is therefore surprising that GNA-GNA duplexes have higher melting temperatures than RNA-RNA duplexes of the same sequence (Zhang 2005). The greater entropic cost of binding the more flexible, acyclic GNA backbone is offset by an even greater enthalpic benefit. Additionally, one enantiomer of this backbone, (S)-GNA can cross-pair with RNA, opening the door to information transfer.

There is, however, a drawback to the simplicity of GNA in nonenzymatic polymerization: cyclic sugars help to maintain a steric separation between the nucleophile and electrophile of a monomer, and without this separation rapid monomer cyclization may preclude polymerization. However, this deficiency could potentially be overcome by an alternative activation scheme (for example, if the monomer could be activated in a cyclic form that opened upon polymerization monomer cyclization would not be a problem). To date, no GNA enzymes have been
developed; if one could be produced it would be interesting to study the effect of the higher backbone flexibility on the sorts of enzyme structures that would be discovered.

1.3.8 Polyamides

Amino acids are known to form under prebiotic conditions and amide linkages are more stable to hydrolysis than the sugar-phosphodiester linkages of RNA. As such, peptide backbones (Fig. 1.2) have been studied for their potential in origin of life as pre-RNA informational polymers (Nelson 2000). Standard peptide backbones with side chains modified with nucleobases (uracil or thymine groups) proved limited in their ability to form stable Watson-Crick hybridization with DNA or RNA (Buttrey 1975). An alternative backbone, in which nucleobases are attached by an installed exocyclic amine that is condensed with a glutamic acid residue on a polypeptide chain at a every other position (with aspartic acid residues spaced between them) (Fig. 1.2 agPNA), was synthesized and shown to base pair to DNA and RNA (Mittapalli 2007). However, the necessity of dimeric units (Glu-Asp) and the prebiotically implausible modifications to the nucleobases make this system unattractive as a progenitor to RNA.
Another alternative peptide nucleic acid (PNA), composed of \(N\)-2-(aminoethyl)glycine (AEG) units, was designed to mimic the spacing of DNA and was found to base pair with DNA in a sequence specific fashion (Nielsen 1991, 1994, Egholm 1993). Orgel and co-workers showed that it was possible to polymerize such peptide nucleic acid (PNA) molecules, in particular, a PNA dimer with the sequence GG on poly(C) DNA or PNA templates by \textit{in situ} carbodiimide mediated 2-methyl imidazolidine activation of the carboxylate of a PNA-GG dimer (Bohler 1995). The reciprocal was also shown to be possible; 2-methyl imidazole activated RNA monomers could polymerize on a PNA template (Bohler 1995, Schmidt 1997). As with RNA, templates other than polycytidine were less efficient, making the copying of mixed sequence templates difficult. However, the demonstration that nonenzymatic information transfer was possible from PNA to PNA, from PNA to RNA, and from RNA to PNA shows that the PNA backbone is positioned to act as a direct informational transfer vehicle to RNA in an enzyme-free reaction setting. Unfortunately, AEG-derived PNAs are achiral and very hydrophobic, and they aggregate in the absence of a complementary base-pairing partner. This deficiency could potentially be alleviated by adding a charged amino-acid side chain at the gamma position (a location where an enantiogenic moiety has already been installed successfully (Dragulescu-Andrasi 2006)). No prebiotic synthesis exists for this
backbone, but if one were, for AEG-PNA or a similar structure, the stability of the backbone would make it an attractive alternative to RNA.

1.4 Nucleobase

There are two classes of alternatives to the Watson-Crick base pairs found in DNA and RNA (Fig. 1.3a,b): slight modifications that enhance affinity or specificity (Fig. 1.3a,b), and complete reconstructions with novel hydrogen bond pairing patterns or even completely different means of base pairing (Fig. 1.3c,d).

1.4.1 Slight modifications

Any nucleobase modification, like the backbone modifications described above, must be at least as good if not a significant improvement over RNA by one of the criteria laid out in section 1.2.2 above to be considered as a likely progenitor of RNA. Slightly modified forms of the natural nucleobases have been studied primarily with the aim of improving the rate and fidelity of nonenzymatic polymerization; under the assumption that they are similarly prebiotically plausible in terms of synthesis and would not impair the catalytic efficacy of the nucleic acid. For example, Orgel and co-workers showed that swapping 2,6-diaminopurine (D) (Fig. 1.3a) for adenine in RNA
templates improves the efficiency of nonenzymatic template-directed polymerizations, in some cases even to the extent of the G-C base pair, due to the extra hydrogen bond. Base-pairing can further be enhanced by introducing moieties in the nucleobases for additional stacking. When a 5-propynyl group was attached to the C5 position of uracil, it was found to pair with D with even higher efficiency (Chaput 2002). D has two known drawbacks however: it has increased propensity for adopting other tautomeric forms causes a decrease in fidelity (Schrum 2009) and its fluorescent properties leave it more vulnerable to light-mediated decay than adenine.

One of the major challenges in nonenzymatic primer extension reactions is the fidelity of polymerization: how accurately can the information coded on a template be transferred to the complementary strand. It has been shown that replacing the oxygen atom at the 2-position on uracil or thymine (O2) with sulfur increases both the fidelity and the rate of template copying (Heuberger 2015). This observation is attributed to the preferential stabilization of the 3′-endo conformation of the monomer sugar, which enhances binding and reactivity (Zhang 2013a, Larsen 2015). These modifications are relatively minor while being compatible with the same prebiotic chemistry scenarios as the canonical nucleobases. Additionally, they have been shown to be compatible with ribozyme activity (Prywes 2016), further validating them as potential progenitors to uridine.
Figure 1.3: The structures of (a) the base pairs of adenine (A) to uracil (U) or thymine (T) and diaminopurine (D) to U or T; (c) the base pair of guanine (G) to cytosine (C) wobble pairs of G with U or T, G with 2-thiouracil (s^2U) or 2-thiothymine (s^2T) and the base pair of A and s^2U or s^2T; (d) the rosette formed by melamine (M) and barbituric acid (BA); (e) the rosette fromed by A and cyanuric acid (CA).
1.4.2 Complete reconstructions

Due to difficulties associated with the prebiotic synthesis of the canonical nucleobases (Oró 1960, Ferris 1968), more divergent alternative nucleobases have been investigated for their potential as precursors to the natural nucleobases (G, C, A and U) with the hope that they could transfer information to RNA later (Hud 2013). One possibility is that a system of purine pairs formed an alphabet (Heuberger 2008). Alternatively, pyrimidine-only systems have been explored (Mittapalli 2007) as well as systems involving triplexes (Zhou 2015, Piao 2015) or even six membered “rosettes” (Fig. 1.3c) (Cafferty 2013). In a recent study, Krishnamurthy and Hud demonstrated that two prebiotically plausible heterocycles, melamine (M) and barbituric acid (BA) (Fig. 1.3c) spontaneously and selectively react with ribose-5-phosphate in water to form nucleotides favoring the β-anomers over the α-anomers (Cafferty 2016). Notably, these nucleotides, like their constituent nucleobases, formed base pairs and self-assembled into long, stable, non-covalently stacked hexameric rosettes (Fig. 1.3c). The resemblance of the base pairing pattern of these heterocycles to that of adenine and uridine has been taken to suggest a possible ancestral connection, in fact cyanuric acid (CA), which has a very similar structure to BA, was recently shown to form stable hydrogen bonded complexes with three polyA strands of DNA (Fig. 1.3d) (Avakyan 2016) demonstrating some level of compatibility between BA or CA and A.
Additional alternative nucleobases have been studied for their role as a possible additional base pair in nucleic acids that can act orthogonally to the natural pairs in the context of a living cell (reviewed in Malyshev 2015). These alternative nucleobases have been successfully used in ribozyme selections (Sefah 2014, Kimoto 2013) and as a third base pair in a living organism (Malyshev 2014). However, these bases are not more easily synthesized under prebiotic constraints than the natural nucleobases and, while important for the constructions of modified life forms, provide little insight from an origin of life perspective.

The most important question facing any proposed alternative nucleobase system is how to transition to RNA. The transition from D to A or 2-thioU to U can be explained in a rather trivial manner. However, the transition from M, BA and CA to the canonical RNA nucleobases requires more elaborate steps that must each be efficient and compatible with the complex prebiotic chemistry. First, the alternative bases must be shown to be capable of storing and passing on useful information. Specifically, covalent strands must be formed in a prebiotic manner, nonenzymatic polymerization must be demonstrated and enzymes must be evolved. After that, it must be shown that RNA can base pair with these nucleic acids based on alternative nucleobases, and that the alternative nucleic acid can template the polymerization of RNA monomers.
Transitions between very different nucleobase alphabets raise other important questions, for example: Would there be transitional polymers containing nucleobases from both sets? Would both alphabets form under the same prebiotic conditions and if so in what ratios? These difficulties would seem to erect very high barriers for potential alternative nucleobases to overcome. Until answers to these questions are provided, the simplest solution, that the natural RNA nucleobases arose first (especially in light of there being at least some trace of them in prebiotic chemical systems like meteorites) would seem to be the default hypothesis.

1.5 Leaving Group

In modern biology the leaving group in enzymatic polymerization of a nucleotide triphosphate is the pyrophosphate moiety. Nucleoside 5′-triphosphates (NTPs) are stable in solution and make excellent substrates for enzymes that are capable of screening the charge of the NTPs to allow for the nucleophilic displacement (Westheimer 1987). Unfortunately, for nonenzymatic polymerization, pyrophosphate makes for a poor leaving group as it is kinetically trapped. In aqueous solutions it cannot be easily displaced by the 3′ end of a growing nucleotide chain because the strong negative charge of the triphosphate blocks the approach of the hydroxyl
nucleophile. As such, research into nonenzymatic polymerization has employed alternative leaving groups.

1.5.1 5’ or 3’?

In principle, RNA polymerization in the RNA world could have proceeded in either the 5’ to 3’ direction, or the reverse, while in modern biology synthesis is exclusively in the 5’ to 3’ direction. This is a consequence of the pyrophosphate leaving group on the canonical NTPs is found exclusively at the 5’ end. As such, much of the work in nonenzymatic polymerization has focused on model systems with 5’-phosphate activation. On the other hand, the work of Powner et al. resulted in nucleotide monomers that are weakly activated as 2’-3’ cyclic phosphates (Powner 2009). Regarding 5’ activation of RNA monomers, the modern 5’-NTPs may have been replaced earlier, more reactive nucleotides such as the 2-methylimidazolides. In fact, if pyrophosphate was available on the early earth, it could have displaced other leaving groups, such as 2-methylimidazole, to form nucleotide triphosphates directly; if NTPs were already available, the evolution of polymerases that use NTP substrates would have been much easier. The question of the chemical origin of 5’ phosphorylation and activation, however, remains to be answered. For 3’ activation of RNA (Fig. 1.4b), nonenzymatic polymerization of monomers has not been successful
(Verlander 1973). No process by which 3’ activation is converted to 5’ activation has been demonstrated. Nonenzymatic polymerization in the “reverse” (3’ to 5’) direction has only been demonstrated when an amine nucleophile is used at the 5’ end of a primer (Kaiser 2012). Polymerization in either direction involves the same chemical transformation, phosphoryl transfer, but, for standard hydroxyl RNA monomers has only been observed in the 5’ to 3’ direction. This leaves open the question of whether steric constraints fundamentally favor 5’ to 3’ polymerization. If so, some mechanism for the transformation to 5’ phosphates will be required for any proposed prebiotic nucleotide synthesis that results in 3’ phosphates.

![Diagram](image.png)

**Figure 1.4:** (a) Examples of leaving groups from enzymatic and nonenzymatic polymerization reactions. (b) Protecting group strategy for the polymerization of RNA with a leaving group at the 3’ hydroxyl. (c) Alternative nucleophiles for nonenzymatic polymerization of ribose sugar containing polymers.
**Heterocycles:**

The identity of the chemical activating groups used in laboratory RNA copying experiments has historically been guided primarily by chemical constraints rather than prebiotic plausibility. The first use of chemically activated nucleotides came in 1968 when Orgel and co-workers (Weimann 1968) showed that nucleoside monophosphates pre-activated with imidazole (Fig. 1.1a,b) could be polymerized on a template without the aid of an enzyme. Generation of imidazole has been proposed to be prebiotically plausible (Oró 1984, Sleeper 1978), and work with imidazole leaving groups continues to this day.

Alternatives to imidazole have been found but they are few in number. 2-methylimidazole was recognized early on for its improved regiospecificity and reaction rate in nonenzymatic polymerization reactions (Inoue 1981). Since then, only one other family of heterocyclic leaving groups has been found that is of comparable reactivity: the benzotriazoles (Fig 1.4a) (Vogel 2005). Neither of these alternatives to imidazole is more amenable to prebiotic synthesis than imidazole itself. More prebiotically plausible alternatives like 1-methyladenine (Fig 1.4a) have been used (Prabahar 1997) but have generally been found to be inferior to 2-methylimidazole. 2-methylimidazole was discovered in a screen of very small scope (Inoue 1981) and the benzotriazoles were chosen based on their utility in peptide coupling chemistry. A
more comprehensive screen of heterocycles might therefore uncover improved leaving groups that are also more prebiotically plausible.

The prebiotic presence of heterocycles is necessary for the production of activated monomers but not sufficient; the activation of nucleoside monophosphates through condensation with heterocycles is achieved in the laboratory using synthetic reagents like 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) or dipyridyldisulfide (DPDS) which are not likely to be found on the early earth. How then were activated monomers generated? There has been some work done with N-cyanoimidazole (NCI) (Bowler 2013) demonstrating that chemical ligation is possible with this single, simple molecule but no prebiotic synthesis of heterocycle-activated nucleotides has yet been demonstrated.

Early life scenarios where activated monomers exist in high concentrations are difficult to justify, therefore, their use has sometimes been viewed as a model. However, RNA polymerization requires activation with good leaving groups, so alternative scenarios must be considered. It is possible that low concentrations of activated monomers were sufficient to foster slow early polymerization; low temperature reactions (Jauker 2015) or even polymerization in ice (Attwater 2013) might facilitate binding when monomers are dilute. If high concentrations are required, alternative leaving groups and nucleotide activation chemistries will have to
be discovered. Finally, *in situ* activation provides a scenario where purified activated nucleotides are unnecessary, and a steady-state supply of activated nucleotides might exist at high enough concentration for polymerization to proceed.

1.5.2 *In situ* activation:

The presence of a constant supply of monomers with highly reactive leaving groups constitutes an unlikely scenario for the geochemistry relevant to the early Earth. As such, alternative polymerization reaction conditions that allow *in situ* activation of monomers are desirable. For example, activation of a nucleoside phosphate by the addition of EDC into the reaction medium leads to 3’ to 5’ polymerization on a complementary RNA template (Sulston 1968, Jauker 2015), though not very efficiently. A screen of leaving groups for *in situ* activation was tested in a 3’-amino system (see “nucleophilic activation” section below) and 1-methylimidazole was found to promote fast polymerization with the fewest side reactions (Röthlingshöfer 2010).

While the prospect of one pot synthesis is attractive, especially in a prebiotic setting where multiple steps are less plausible, there is one major drawback. An activating agent has the potential to derivatize the existing nucleobases, templates or growing primers. If, for example, the nucleophile at the 3’ end of a growing primer
reacts with an activating agent, the growing chain will be prematurely terminated and no further polymerization will be possible. Röthlingshöfer et al. found that, as they increased the pH of their reactions from 7 to 9 the proportion of such terminated products increased with respect to polymerization products (Röthlingshöfer 2010). An ideal activator, therefore, must be highly chemoselective for phosphates. This tradeoff between reactivity and specificity is one of the major challenges for in situ activation of monomers.

In a recent study, Jauker et al. (Jauker 2015) improved on the earlier strategy (Sulston 1968) of using a carbodiimide activating reagent by including 1-ethylimidazole. Unlike the standard imidazole-activated nucleoside monophosphates, N1-alkylated imidazole analogs are more reactive due to the positive charge (Fig. 1.4a), which makes them better leaving groups and further activates the phosphorous center. This allowed nonenzymatic copying of any individual monomer and some pairs of monomers at high yields with the assistance of helper oligomers (see section 1.7). Even with in situ formed cationic leaving groups, these reactions were in many cases rather slow (as long as 21 days for a single addition). However, they did not appear to produce alkylated side products, speaking to their specificity.

Finally, in situ activation was also carried out to enhance the reactivity of preformed activated nucleotides, mainly nucleoside phosphoroimidazolides, using 1-
(2-hydroxyethyl)-imidazole (HEI) (Fig 1.4a) (Mansy 2008, Schrum 2009, Zhang 2013a,b, Blain 2014, Izgu 2016). HEI likely improves the rate of primer extension by exchanging with imidazole group to form a more reactive, zwitterionic intermediate.

*In situ* activation is an attractive prospect because it allows multiple reactions to proceed in one-pot. Like all one-pot reactions in which multiple, sequential chemical transformations are occurring together, the possibility of cross-reactivity between the different reactions, activation and polymerization in this case, is a central challenge. If sufficiently specific and prebiotic conditions could be found to overcome this challenge, they would be very good candidates for pre-RNA world conditions.

1.5.3  **Cyclic phosphates:**

2′-3′ cyclic phosphates of RNA monomers are the result of the prebiotic synthesis pathway for of cytidine and uridine proposed by Powner et al. 2009 (but maybe there are as yet undiscovered pathways leading directly to 5′-phosphates) and have also been explored as activated monomers. The 5-membered ring containing the phosphate is only weakly reactive towards nucleophilic ring opening. Therefore, the rate of nonenzymatic template-directed monomer addition is slower than that for nucleotides with traditional leaving groups at their 5′ ends. As in the case in 5′-activated monomers, hydrolysis limits the lifetime of 2′-3′ cyclic phosphates and thus,
the yield of polymerization reactions. Moreover, the regioselectivity of the reaction strongly favors the formation of a 2′-5′ linkage over the canonical 3′-5′ linkage (Verlander 1973), in fact, the 2′-5′ linkage is the exclusive product under certain conditions (Renz 1971). Recently, Sutherland and co-workers reported a prebiotically plausible polymerization of 3′,5′-RNA from a ribonucleoside-2′,3′-cyclic phosphate via selective 2′-O-acetylated, 3′-P activated intermediates (Fig. 1.4b) (Bowler 2013). Acetylation of mixed oligomers possessing either 2′- or 3′-terminal phosphates is selective for the 2′-hydroxy group of the latter. Whether or not this strategy can be expanded to monomers remains to be seen.

1.6 Nucleophile

The scope of nucleophiles used in nonenzymatic polymerization is significantly more restricted than that of leaving groups. This is primarily because very few nucleophiles can replace the 3′-hydroxyl of a ribonucleotide monomer without significantly perturbing the helical geometry of the growing duplex or the electronic requirement of the nucleophile.
1.6.1 RNA vs. DNA:

Pioneering studies from the Orgel laboratory demonstrated that the two vicinal hydroxyls of RNA react faster than the one 3’ hydroxyl of DNA, by about one order of magnitude, due to increased nucleophilicity (Kozlov 1999b). The drawback of two hydroxyls is the possibility of forming 2′-5′ linkages as well as the desired 3′-5′ linkages. Yet, the polymerization can still be regiospecific depending on the leaving group, e.g. 2-MeIm (Inoue 1981), or the divalent metal catalyst (Bridson 1980), both of which can by unknown mechanisms promote the formation of 3′-5′ linkages.

1.6.2 Amino nucleic acids:

The only alternative to a hydroxyl nucleophile at the 3′ (and/or the 2′) position of a nucleotide that has been explored extensively is the amine (Fig 1.4c). Orgel and co-workers examined the polymerization efficiency of 3′-amino-3′-deoxyguanosine-5′-phosphoro(2-methyl) imidazolide on both poly(C) DNA and RNA templates (Zielinski 1985, Tohidi 1987). Their results showed that both the yield and regiospecificity of the full-length oligomers were superior to those obtained with the standard ribo-G monomer, suggesting that replacing the hydroxyl group on the sugar with the more nucleophilic amino group can enhance the rate of the nonenzymatic polymerization of nucleic acids.
Since the polymerization of amino-nucleophile monomers does not require the presence of Mg$^{2+}$, such monomers are promising as components of a protocell. Deoxyribonucleotides featuring an amino group on the sugar have been employed to copy RNA templates with high efficiency both in solution and inside lipid vesicles (Fig. 1.4c) (Mansy 2008). Later, Schrum and co-workers reported (Schrum 2009) the nonenzymatic copying of both homopolymeric and heteropolymeric (mixed-sequence) templates using the entire 2′-NP-DNA monomer set (G, C, and analogs of A and U). Although the activated 2′-aminonucleotides were stable under the primer-extension conditions and able to rapidly polymerize on short homopolymeric RNA and LNA templates (4-mers), the rate of copying a mixed-sequence 15-mer RNA template was significantly lower, affording truncated 2′-NP-DNA products. This observation was attributed to the incorporation of mismatched bases, which caused stalling of the primer extension reaction.

In an effort to identify a system with a better inherent fidelity of polymerization, another phosphoramidate system, 3′-NP-DNA, was developed. Zhang and co-workers showed (Zhang 2013a) that activated 3′-amino-2′,3′-dideoxynucleoside monophosphates (3′-NH$_2$-2-MeImpddNs) (Fig. 1.1a) rapidly polymerized on short (4-nt) homopolymeric DNA, RNA, and locked nucleic acid (LNA) templates. Zhang et al. later reported (Zhang 2013b) that the fidelity for copying the (G)$_2$(A)$_2$ 3′-NP-DNA template increased upon use of the 2-thio ($s^2$) analog.
of the T monomer, 3′-NH₂-2-MeImpdds₂T (Fig. 1.3). More recently, Izgu et al. quantitatively investigated the effects of these 2-thio and 3′-amino modifications on RNA template-copying rates (Izgu 2016) finding that thiolated thymidine monomers polymerize 5 times faster than their standard thymidine counterparts.

While the amino modified systems react significantly faster than the standard ribose analogs, rapid 3′-5′ cyclization of the monomers is a problem unique to the amino nucleophile; a problem that can be overcome by making cyclization sterically impossible, as is the case for the activated 2′-aminonucleotides (Schrum 2009). The half-lives of 3′-NH₂-2-MeImpddNs are similar to those of activated hydroxyl-nucleophile monomers, 1-3 days, though the presence of the catalyst HEI reduces the halflife sharply to approximately 1 hour. Nevertheless, it remains possible that, due to the fast and accurate polymerization observed in this system, sufficiently long and information-rich 3′-NP-DNA products can be produced nonenzymatically despite monomer cyclization. Whether or not amino-monomers could have been synthesized prebiotically, they present a promising avenue towards the construction of a protocell because of their compatibility with fatty-acid vesicles and their improved polymerization rates. Future work will determine whether or not the N-P bond-containing backbone has the ability to fold into tertiary structures with evolutionarily useful functional properties.
1.6.3 Other nucleophiles:

Nonenzymatic polymerization with sulfur or selenium as the nucleophile has not been attempted because sulfur is known to be a very poor nucleophile for phosphorus electrophiles (Dantzman 1996). Other alternatives like amino-oxy or hydrazide nucleophiles might be effective but would add additional atoms to the backbone, which would have an unknown effect. The Huisgen cycloaddition (click) reaction has been used in place of a nucleophile in the backbone (Isobe 2008) though it is clearly not prebiotically plausible, nor has nonenzymatic polymerization been demonstrated.

1.7 Contextual modifications

In addition to modifying nucleotide structure, a number of other strategies have been employed to increase the yield and fidelity of the nonenzymatic polymerization reactions. These strategies primarily consist of using additives and strand functionalizations that assist the polymerization reaction. As in the previous sections, these strategies can be judged on two scales, how useful they are for the reaction and how plausible they are in a prebiotic setting.
1.7.1 Helper RNA

In recent studies, the Richert group (Deck 2011, Jauker 2015) has used “microhelper” RNAs to improve the binding of monomers to the primer-template duplex (Fig. 1.5a). Microhelper RNAs are short, RNA oligomers; in the case of these studies they were trimers or tetramers. These microhelpers have sequences that allow them to bind reversibly just downstream of the monomer binding site such that there is only one exposed (single stranded) base on the template where the monomer can bind. The additional base stacking face provided by the microhelper is thought to strengthen the binding of the monomer. However, base stacking may not be the only relevant non-covalent interaction capable of assisting in polymerization. Orgel and coworkers observed (Wu 1992b), similarly, that downstream-binding mononucleotides increased the extent of polymerization. They also noticed that the identity of the leaving group attached to the 5’ phosphate had a dramatic effect on polymerization efficiency: if imidazole was used as the leaving group, rather than the more efficient 2-methylimidazole, in the helper position (one position downstream of the adding nucleotid) polymerization proceeded at a much lower yield than when 2-methylimidazole was used. The mechanism of this effect is unknown.

Since short oligomers may be produced as abortive products of polymerization, they are highly plausible in the context of any nonenzymatic polymerization reaction.
In addition to assisting in monomer polymerization they may also serve as primers or as oligonucleotides to be ligated.

**Figure 1.5:** (a) Helper oligonucleotides. (b) Bead immobilization. (c) Reversible termination. (d) Mineral surface catalysis. (e) Metal ion catalysis. (f) Template termination assistance.

### 1.7.2 Immobilization

In addition to using helper RNAs, the Richert group took inspiration from the standard solid-phase oligomer synthesis protocol and used a solid support (Fig. 1.5b,c) (Deck 2011). The primer strand in these reactions was bound to a magnetic iron-oxide
bead through a biotin-streptavidin interaction so that the monomer (and microhelper RNA) could be refreshed continuously (Fig. 1.5b). By immobilizing the primer and replacing the monomer, the background reaction of monomer hydrolysis was no longer problematic because hydrolyzed monomer simply flowed out of the reaction. Without this constant input of fresh reagents and output of reacted ones, the hydrolyzed monomers would have acted as competitive inhibitors to the polymerization reaction stalling it. While this approach is very effective in improving the yield of the reaction, it is not obvious what prebiotic conditions would mimic or lead to the immobilization of a growing RNA strand – it is possible that a vesicle could be semipermeable in such an advantageous way, but hydrolyzed monomers are actually less capable of permeating membranes than activated monomers.

Building on this strategy, Richert et al. (Kaiser 2012) developed a stepwise, bidirectional template-copying assay, in which an immobilized primer strand was extended in either the 3’-to-5’ or the 5’-to-3’ direction by addition of a benzotriazole-activated nucleotide that was protected either at the 3’ or 5’ end, respectively (panel B). Once extended, the primer terminus was deprotected and subjected to the next round of monomer addition-deprotection. In addition, a similar reaction setting, but with a native primer strand that was annealed to an immobilized template, was employed to extend the primer in both 3’ and 5’ directions simultaneously in the presence of two
different activated monomers. This strategy, while synthetically elegant, is completely dependent on the presence of a chemist and thus unsuitable for the ancient Earth.

1.7.3 Effect of Mineral Surfaces

Montmorillonite, one of the most common members of the smectite family of clay minerals is an aluminum silicate derivative, and is formed by the weathering of volcanic ash. Catalytically active montmorillonites are obtained when the alkali and alkaline earth metal ions (with the exception of Mg$^{2+}$) are exchanged for Na$^+$. The mechanism of the catalytic function relies on the intercalation of activated monomers into the layers of montmorillonite platelets in the presence of divalent metal cations; the high concentration of monomers facilitates the polymerization of RNA oligomers. The observation of the formation of 2′,3′-cyclic phosphates from 3′-nucleoside monophosphates in the presence of montmorillonite (Ferris 1986, 1988) suggested that it could act as a prebiotic catalyst for the formation of the phosphodiester bonds of RNA (Fig. 1.5d). Oligomers up to 10 bases long were observed in the montmorillonite-catalyzed polymerization reaction (Ferris 1992) with the 3′,5′-phosphodiester bond as the major product using nucleoside phosphorimidazolide monomers. Further, using activated nucleotides, montmorillonite catalyzed the extension of a decameric primer reaction to yield oligomers of length 30-50 (Ferris 1996, 2002). Ferris and co-workers
showed that the clay-catalyzed oligomerization reaction can influence the sequence and chiral selectivity of the polynucleotide products, as reactions containing equal amounts of the four nucleobases lead to products primarily containing purines at their 5'-ends; this is due to the fact that pyrimidine nucleotides elongate at a significantly slower rate than their purine counterparts. Finally, when polymerization reactions contain mixtures of right- and left-handed RNA monomers, monmorillonite catalysis produces and excess of homochiral RNA oligonucleotide products; this remarkable feat of molecular assembly is especially suggestive of a prebiotic role for montmorillonite, or a similar mineral catalyst, in light of the fact that the mineral itself is achiral.

1.7.4 Metal ions

Mg$^{2+}$ is the divalent cation of choice for the stabilization of ribozymes in modern biology (Bowman 2012). It is also the metal ion most commonly used in nonenzymatic polymerization reactions, primarily because the fastest rates of nonenzymatic polymerization of RNA have been reported in the presence of Mg$^{2+}$ (though a more comprehensive screen of metal ions might demonstrate otherwise). For reactions involving a hydroxide nucleophile, reactions where divalent metal cations are not included exhibit extremely slow polymerization (Fig. 1.5e) (Inoue 1982). It should be
noted that this is not in the case for reactions involving amine nucleophiles. Several alternatives to Mg\(^{2+}\) have been explored, however, and some have significant advantages: The Orgel group also experimented with Pb\(^{2+}\) and Zn\(^{2+}\), finding that Pb\(^{2+}\) favored 2′-5′ linkages while Zn\(^{2+}\) favored 3′-5′ linkages in the formation of poly(G) oligomers, possibly due to coordination of the N\(^7\) atom of the guanine nucleobase (Lohrmann 1980). The difference in reactivity between divalent cations is likely rooted in the precise geometry adopted by the RNA molecules upon metal binding (Lohrmann 1980), however, the pKa of the metal may also be a significant factor.

### 1.7.5 Template-end assistance

The rate of a primer extension reaction directed by the terminal nucleotide of a template is slow compared to primer extension on internal template positions (Fig. 1.5f) (Pisch 2003, Wu 1992a,b,c). This presents a serious problem for nonenzymatic polymerization because incomplete copying would cause sequence shortening after every copying step and, in the absence of prebiotic telomerase, irreversible loss of information. Richert and co-workers showed that acyl substitutions at the 5′-carbon of the terminal nucleotide of a DNA template can elevate both the rate and the fidelity of the extension reaction of a 3′-amino-terminal DNA primer using 2-methylimidazolide-DNA monomers on DNA templates (Rojas Stutz 2001, 2006). Use of 3′-terminal
carboxamide substituents on 2′-deoxyribose systems such as stilbene and tetrahydroquinoline enhanced the rate of the incorporation of all four nucleobases across from the terminal template position. Substitutions on the primers were also investigated. For example, Richert and co-workers showed that a 5-propargylamine side chain on the 3′-terminal 3′-amino-2′,3′-dideoxyuridine residue of a primer enhanced the rate of primer extension (Baumhof 2006). Unfortunately, no plausible prebiotic synthesis of such terminal modifications is known. It is possible that homopolymeric sequences at the ends of templates could, by slipping or another mechanism, cause regeneration of terminal sequences; such a scheme remains to be experimentally demonstrated.

1.8 Prospective

Nonenzymatic polymerization of information-containing materials thus far has only been demonstrated with activated nucleotides not found in life. Even using unnatural components, the overall success of nonenzymatic polymerization as a method for replication on the early Earth is debatable due to low reaction rates and fidelities; an early genetic system that is composed only of RNA would have faced many other obstacles as well, as outlined above. Nevertheless, due to the existence of very short catalytic RNA sequences (Chumachenko 2009, Turk 2011), nonenzymatic
polymerization may not need to be particularly efficient in order to successfully produce the first low-activity ribozymes. In other words, ribofuranosyl nucleotides could form short (5–40 nucleotides) (Turk 2011, Hendry 1995) functional RNA fragments capable of facilitating chemical transformations, including RNA polymerization. Short RNA fragments are also capable of assembling through base pairing (non-covalently) into larger functional complexes (Doudna 1991, Green 1992, Whoriskey 1995, Vaidya 2012), meaning that nonenzymatic polymerization of short fragments could still allow for the construction of large, intricate ribozymes. If such a system could operate inside of a compartment, RNA catalysis of reactions between small molecules (ligation of nucleotides, formation of peptide bonds, etc.) within that compartment could potentially drive selective growth and division and initiate Darwinian evolution.

All alternative polymers to RNA mentioned in this review have significant advantages compared to RNA in one or more areas, but are accompanied by disqualifying deficiencies. As laid out above, these alternative systems must be prebiotically synthesizable, replicate efficiently, fold into functional catalysts and be capable of transferring information to RNA.

In this review we have discussed alternative polymers that could plausibly transfer information to RNA. Recent work ignoring this constraint has sought to
discover alternative genetic polymers that, while not relevant from the perspective of the emergence of life on earth, help to delineate the criteria that are necessary for the formation of hereditary systems more generally. The supramolecular assembly of non-biological informational polymers has recently been demonstrated (Sadownik 2016, Zhu 2010). Such systems may be radically incompatible with life on earth, requiring non-aqueous solvents or participating in informational transfer that is very different from the sequence of letters from a limited alphabet that is observed in RNA. If such a system could be developed in which arbitrary information could be stored and copied, it would expand the chemical space relevant towards the formation of life and could, thereby, influence the search for extraterrestrial life.

1.9 Chapter 1 References


Baumhof, P., Griesang, N., Bächle, M. & Richert, C. Synthesis of Oligonucleotides with 3’-Terminal 5-(3-Acylamidopropargyl)-3’-amino-2’,3’-dideoxyuridine Residues and


Chaput, J. C., Sinha, S. & Switzer, C. 5-Propynyluracil-diaminopurine: an efficient


Heuberger, B. D., Pal, A., Del Frate, F., Topkar, V. V. & Szostak, J. W. Replacing uridine with 2-thiouridine enhances the rate and fidelity of nonenzymatic RNA primer


Robertson, M. P. & Miller, S. L. An efficient prebiotic synthesis of cytosine and uracil.


Taylor, A. I. & Holliger, P. Directed evolution of artificial enzymes (XNAzymes) from


Yu, H., Zhang, S. & Chaput, J. C. Darwinian evolution of an alternative genetic system
CHAPTER 2

Catalysis of nonenzymatic copying of RNA templates by activated helper oligonucleotides

2.1 Abstract

The nonenzymatic replication of RNA is a potential transitional stage between the prebiotic chemistry of nucleotide synthesis and the canonical RNA world in which RNA enzymes (ribozymes) catalyze replication of the RNA genomes of primordial cells (Patel 2015, Powner 2009). However, the plausibility of nonenzymatic RNA replication is undercut by the lack of a protocell-compatible chemical system capable of copying RNA templates containing all four nucleotides. Over 30 years ago, Orgel et al. showed that 2-methylimidazole activated G and C monomers could copy short templates by primer extension (Joyce 1986), but the copying of templates containing all four nucleotides remains problematic (Wu 1992a, Deck 2011). We show that short 5′-activated oligonucleotides act as catalysts that accelerate primer extension, and allow for the one-pot copying of mixed sequence RNA templates. The fidelity of the primer extension products resulting from the sequential addition of activated monomers,
when catalyzed by activated oligomers, is sufficient to sustain a genome long enough to encode active ribozymes.

### 2.2 Introduction

In order to serve as a bridge from prebiotic chemistry to primordial life, the nonenzymatic copying of RNA templates by suitably activated nucleotides (Fig. 2.1a) must occur quickly enough to replicate functional RNA sequences faster than they degrade. However, the nonenzymatic copying of sequences containing all four nucleotides has not yet been possible for a number of reasons (Szostak 2012), most notably the slow rate of primer extension with adenosine and uridine monomers (Wu 1992a, Deck 2011, Joyce 1987). Indeed, two adjacent A or U residues in the template will stop polymerization entirely (Patel 2015, Powner 2009, Joyce 1987) under normal conditions; even partial copying requires extreme conditions, such as the eutectic phase of frozen samples (Joyce 1986, Stribling 1991), which are not compatible with replication within protocells. Extensive optimization of reaction conditions with the aim of improving the rate of primer extension with all four monomers, including varying the choice of divalent cation and leaving group, improved the rate and regiospecificity of polymerization for G and C but not A and U (Wu 1992a, Deck 2011, Inoue 1981, Hagenbuch 2005, Lohrmann 1980). Recent biophysical studies suggest that low A and U monomer affinity for the template is not the only reason for this
problem (Szostak 2012, Izgu 2015) since A binds only 3-fold more weakly than G, but primer extension with A is at least 100-fold slower (Wu 1992a, Deck 2011, Joyce 1987, Heuberger 2015) than with G. Moreover, Richert et al. have shown that downstream helper oligonucleotides (Deck 2011) that provide incoming monomers with an additional binding surface improve the binding of A and U monomers, but the underlying difference in rates remains. The use of helper oligonucleotides allowed primer extension to proceed on a template containing all four bases over the course of several weeks, where previously such templates could not be copied at all. However, this advance required the use of an immobilized template, such that each monomer-helper pair could be added, allowed to react for several days to a week, and then washed out prior to the addition of the next monomer-helper pair. This approach is clearly incompatible with RNA replication within protocells.

2.3 Results and Discussion

Given the historical and continuing problems in demonstrating efficient nonenzymatic template directed primer extension, one might wonder whether this biologically inspired model is appropriate for replication within primordial cells. An alternative scenario that might seem more reasonable in the context of prebiotic chemistry involves template copying by the initial formation of short oligonucleotides, followed by ligation events that generate longer oligonucleotide intermediates and
eventually a full length copy (Szostak 2011). To test the viability of this hierarchical assembly model, we compared the rate of primer extension with activated monomers to the rate of ligation of two template bound oligonucleotides (Fig. 2.1b,c reactions 1 and 3). In both cases, the chemical reaction is the same, i.e. attack of the 3′-hydroxyl of an oligonucleotide on the 5′-phosphate of a downstream nucleotide, with displacement of the 2-methylimidazole leaving group (Fig. 2.1b). Moreover, the reacting nucleotides are the same in both cases: a G at the 3′-end of the primer and a G monomer or a G at the 5′-end of an oligonucleotide. We expected the ligation reaction to be faster, because the oligonucleotides to be ligated are both stably bound to the template, and because the fully base-paired nicked duplex should be largely in the optimal A-type conformation (Zhang 2012), whereas the primer/template complex with template bound monomers would be significantly more disordered. To our surprise, the primer extension reaction proceeded almost 100 times more rapidly than the ligation reaction (Fig. 2.1d). In our search for an explanation for this unexpected observation, we reasoned that one of the major differences between the two scenarios was the presence of multiple adjacent 5′-activated nucleotides downstream of the primer in the case of monomer addition, but not ligation, where there is only a single activated nucleotide. Indeed, the case for a role of a second downstream activated nucleotide is consistent with the well known difficulty of extending a primer to the last nucleotide of a template (Wu 1992b). Furthermore, Orgel et al. showed over 20
years ago that efficient primer extension requires the presence of two activated monomers adjacent to the primer. Remarkably, the reaction is fastest when both monomers are activated with 2-methylimidazole instead of either or both being activated with imidazole (Wu 1992c), suggesting a possible catalytic role for a physical interaction of the leaving groups of adjacent monomers.

In order to directly test the possibility that a 2-methylimidazole leaving group on the 5′-phosphate of a downstream monomer or oligonucleotide (Fig. 2.1b) would increase the rate of reaction between a primer and an adjacent monomer, we designed a series of templates for nonenzymatic RNA polymerization and ligation (Fig. 2.1c). We then measured the rate of addition of a monomer to a primer in the presence or absence of downstream nucleotides, with or without 5′-activation of the downstream nucleotides (Fig. 2.1c-f). Rates were calculated assuming a pseudo-first order rate equation (Fig. 2.2), because the concentrations of primer and template were far below that of the monomer, which would not change significantly during the reaction. Primer extension by addition of a single monomer, in the absence of any additional downstream mono- or oligo-nucleotides, was very slow and comparable in rate to the ligation of the primer to a 5′-activated AGC trinucleotide (Fig. 2.1c-f, reactions 2 and 3). The binding of the unactivated AGC trimer, with either a 5′ hydroxyl or a 5′ phosphate, downstream of an activated G monomer conferred only a modest increase in the rate of addition of the G monomer to the primer (Fig. 2.1c-f, reactions 4 and 5),
Figure 2.1: Catalysis of nonenzymatic primer extension by activated downstream nucleotides. (a) Structure of 2-methylimidazole-activated guanosine-5′-monophosphate, and its schematic representation. (b) Schematic of the RNA primer extension reaction. N1 represents an individual ribonucleotide monomer in position to react with the primer, N2 represents either a monomer or an oligomer downstream, with a leaving group capable of interacting with N1. Dashed lines: Potential interactions between leaving groups or between the downstream leaving group and the upstream nucleotide N1. (c) Schematics of the RNA primers, templates, monomers and oligomers used in d - f. Templates are complementary to the displayed monomers and oligomers. Template 2 has a U following the C to which the G monomer is bound to prevent downstream binding of G. (d) Primer extension by polymerization or ligation on templates 1 or 3, respectively. Fits describe ln(fraction primer remaining) vs time, giving an apparent first-order rate constant. (e) Primer extension assay for the experiments described in c, showing reaction progress after 10 minutes. In lane 6, a primer + 4 band can be observed, representing the slow addition of the activated trimer after the monomer. (f) Pseudo-first order rates of the reactions described in c. Error bars indicate S.E.M; all experiments were performed in triplicate or greater. Reaction conditions: 10 µM primer, 11 µM template, 200 mM CHES pH 9.0, 200 mM MgCl2, 50 mM monomer, 1 mM trimer.
consistent with previous reports of unactivated downstream ‘helper’ oligonucleotides (Deck 2011, Jauker 2015, Kervio 2010). Such helper oligonucleotides are thought to act largely by increasing monomer affinity to the primer/template complex by providing an additional base-stacking surface. In contrast, when the same AGC trimer was activated as a 5′-phosphoro-2-methylimidazolide and used as a downstream helper oligonucleotide, the rate of primer extension (by addition of a G monomer) was increased by over two orders of magnitude (Fig. 2.1c-f, reaction 6) compared to reactions with unactivated trimers. The rate of primer extension with a single G in the presence of an activated downstream trinucleotide was about twice as fast as the rate when the primer was followed by up to four sequential activated G monomers (Fig. 2.1c-f, reaction 1 and 6).

To systematically study the rate of nonenzymatic primer extension as a function of the length of the downstream activated ‘helper’ oligonucleotide, we synthesized a series of activated oligonucleotides of different lengths (Fig. 2.3a). At saturating concentrations (Fig. 2.4) of each of the oligonucleotides, the rate of primer extension improved as the length of the helper increased from mononucleotide to dinucleotide to trinucleotide, and then stayed constant with the tetranucleotide (Fig. 2.3a). At subsaturating concentrations of helper oligonucleotide, the rate of the primer extension reaction increased with increasing concentrations of di- or tri-nucleotide helper (2.4). Because the maximum rate of the trimer-assisted reaction was
Figure 2.2: Determination of the rates of primer extension reactions. (a) Schematic of a primer extension reaction. (b) Gel showing primer (starting material) and extended primer. (c) Plot showing the decay of primer as a percentage of total fluorescence intensity on the gel. (d) Following the assumption of a pseudo-first order rate curve, the ln(fraction of primer remaining) plotted over time gives a plot where the slope is the rate for the reaction.

significantly faster than that of the dimer-assisted reaction, we proceeded with trimers as the downstream helper oligonucleotides for the remainder of this study.

We then tested the ability of a downstream 2-methylimidazole-activated AGC trimer to catalyze template-directed RNA primer extension with all four individual 2-methylimidazole activated monomers – A, G, C and U. We also tested the 2-thiouridine monomer, which we have previously shown to be superior to uridine in
Figure 2.3: Nonenzymatic primer extension using all four monomers. (a) RNA primer extension assay using alternative “helper” oligomers and corresponding, complementary templates. The respective rates are shown at bottom. (b) Primer extension assay for each of the four monomers, as well as 2-thiouridine, in the absence and presence of activated trimer. Shown at bottom of both a and b is a bar graph with the respective rates. Reaction conditions: 10 µM primer, 11 µM template, 200 mM Tris pH 8.0, 100 mM MgCl₂, 50 mM monomer, 10 mM 2-MelmpAG, 1 mM 2-MelmpAGC, 1 mM 2-MelmpAGGC. The gels in a and b show reaction progress after 10 minutes and one hour respectively. Error bars indicate S.E.M; all experiments were performed in triplicate or greater.
nonenzymatic primer extension in terms of both increased rate and improved fidelity (Heuberger 2015). In each case, the rate of primer extension increased by at least two orders of magnitude compared to the corresponding reactions without any helper oligonucleotides (Fig. 2.3b). Remarkably, whereas the rate of primer extension with A and U monomers was previously too slow to measure (Wu 1992a, Heuberger 2015) – conservatively estimated to be at least three orders of magnitude slower than primer extension with G or C monomers – the difference in polymerization rates was reduced to roughly one order of magnitude when assisted by downstream activated trimers (Fig. 2.3b).

In order to assess the fidelity of nonenzymatic trimer-assisted primer extension, we measured the rates of monomer addition for all monomer-template combinations, including both matches and mismatches (Fig. 2.5a). We calculated the fidelity for each template as the rate of matched monomer addition divided by the sum of matched and mismatched monomer addition rates. By averaging the fidelities of the four templates we found the overall fidelity of trimer-assisted polymerization to be 98%. Assuming an error threshold of one mutation per genome (Eigen 1971), this value of fidelity allows for effective genome sizes of 50 nucleotides, long enough to produce functional ribozymes (Ferré-D’Amaré 2010). The fastest mismatch reaction was primer extension by G during copying of a template U, which was approximately 5% as fast as primer extension with A on the same template (Fig. 2.5a, 2.6). This is likely due to
the formation of a G:U wobble base pair (Heuberger 2015), and is consistent with previous studies of nonenzymatic primer extension reactions (Rajamani 2010). All other mismatches were at least two orders of magnitude slower than their matched counterparts. We also tested the effect of replacing the canonical U monomer with 2-thiouridine and 2-thioribothymine on the polymerization rate and found, in line with
previous reports (Heuberger 2015), that this sulfur substitution improved the rates by approximately one order of magnitude (Fig. 2.5b), bringing the rate of primer extension closer to the rates of G and C addition. More rigorous tests of fidelity, in which monomers can compete for binding on a template in all possible sequence contexts, are ongoing.

![Figure 2.5](image)

**Figure 2.5: Fidelity of trimer-assisted primer extension.** (a) Schematic of the fidelity assay. Using the same RNA primer and trimer, four different templates, one with each base, were paired with each of the four 2-methylimidazole activated monomers to test the relative rates of each matched and mismatched pair. (b) Heat map showing the relative rates of each two-letter pairing. Black indicates a rate of 0.001 h\(^{-1}\), white indicates a rate of 100 h\(^{-1}\). (c) Relative rates of the modified U monomers, 2-thiouridine and 2-thioribothymidine compared to that of a canonical, activated U monomer on a template containing an A. Reaction conditions as in figure 2. Error bars indicate S.E.M. All experiments were performed in triplicate or greater.
Figure 2.6: Rates of trimer-assisted primer extension for all monomer/template combinations. (a) Schematic of the primer extension reaction. (b) Rates of monomer additions on different templates for all monomer-template pairs.
Encouraged by the relatively fast and accurate addition of all four monomers to a primer, we then attempted to copy short RNA templates containing all four nucleotides. In order to iterate the process of trimer assisted monomer addition, after a downstream activated trimer catalyzes the addition of the first monomer to the primer, the helper trimer must dissociate to allow for the binding of the next monomer-trimer pair. To test the feasibility of this mode of primer extension, we designed a template that contained binding sites for all four monomers and synthesized the appropriate activated trimers (Fig. 2.7a). We used 2-thiouridine monomer in place of uridine (see above), and all four monomers (A, G, C and 2-thiouridine) were present in every reaction. In the absence of activated trimers, we observed almost no primer extension (Fig. 2.7b lane 1); even though all four monomers were present, their combined rate of addition was not measurable. Adding just the first helper trimer resulted in rapid addition of the first monomer (C) to the primer, followed by slow addition of the first trimer (AGC) to the growing primer. Similarly, adding two, three or four trimers together resulted in the generation of primers extended by two, three or four monomers respectively (Fig. 2.7b). These reactions converted primer into a product with four sequentially added monomers with an ~80% yield after 16 hours; single monomer addition reactions reached >95% yield after only 10 minutes (Fig. 2.3b). The difference between the fast rates observed in the trimer-assisted addition of a single base (Fig. 2.3b) and the slower rate of trimer-
assisted polymerization of all four bases simultaneously (Fig. 2.7b) may be due to the competition of trimers for overlapping binding sites on the template. This hypothesis is supported by the decreased rate of trimer assisted primer extension with a single monomer in the presence of additional overlapping downstream trimers (Fig. 2.8).

The copying of templates containing two consecutive A or U monomers is also made possible by the use of activated helper oligonucleotides (Fig. 2.9).

The experiments described above demonstrate for the first time a simple and robust means of nonenzymatically copying mixed sequence RNA templates in a one-pot reaction. Short, activated oligonucleotides – themselves plausibly generated by either templated or untemplated monomer polymerization – are efficient catalysts of high fidelity primer extension with all four RNA monomers. Thus short activated oligonucleotides can play dual roles, acting either as catalysts of chain growth by monomer addition, or directly as replication intermediates, since they can also be incorporated into a growing chain by ligation. Template copying in a complex but realistic milieu containing both activated monomers and oligomers could therefore occur via a hybrid process combining primer extension with monomers and oligonucleotide ligation. Further advances are required to enable the copying of longer and potentially functional RNA sequences. We suggest that the identification of a plausible source of chemical energy that could drive the re-activation of monomers and oligomers following hydrolytic loss of the leaving group is a key
Figure 2.7: Primer extension with all four monomers in a one-pot reaction. (a) Schematic of a primer extension reaction incorporating all four RNA monomers, with four respective downstream helper trimers. (b) All four monomers and the activated trimers indicated below the gel were added to each reaction. For example, in lane 3 all four monomers and the activated AGC and GCG trimers were mixed with primer, template, buffer and Mg$^{2+}$. Higher bands represent either the ligation of activated trimers or the polymerization of G monomers without trimer assistance. The gel shows reaction progress after 16 hours. (c) Timecourse of the reaction with all four trimers. A 48 hour timepoint of the reaction without any trimers (left lane of b) is contrasted with a timecourse (timepoints at 5 and 30 minutes, 2, 24 and 48 hours). The percentage of polymerization products that have been extended by at least four nucleotides is plotted over time. Reaction conditions: 2.5 µM primer, 2.5 µM template, 100 mM HEPES pH 8.0, 100 mM MgCl$_2$, 20 mM monomer, 100 µM trimer.
Figure 2.8: Inhibition of primer extension by overlapping helper-trimers. (a) Schematic of the primer extension reaction. (b) Rate of primer extension as a function of increasing GCG trimer concentration.

Figure 2.9: Extension of primers by multiple consecutive A or U nucleotides. (a) Schematic of a primer extension reaction wherein two consecutive A nucleotides can polymerize. The gel shows extension without trimers and with one or two trimers after 16 hours. (b) Same as in (a) but with two U additions. In the right half 2-MeImpU is replaced with 2-MeImp-2-thioU.
missing component in efforts to reconstitute nonenzymatic RNA replication. A chemical environment that could maintain a fully activated pool of substrates, and avoid accumulation of the inhibitory by-products resulting from hydrolysis, might be sufficient to drive the complete copying of longer RNA templates, thus setting the stage for the emergence of Darwinian evolution.

2.4 Materials and Methods

Chemicals

Guanosine 5’-monophosphate was purchased as the free acid from Santa Cruz Biotechnology (Dallas, TX). 2-thiouridine-5’-phosphoro-2methylimidazolide and 2’,3’-diacetyl-nucleosides were purchased from ChemGenes (Wilmington, MA). Phosphoramidite nucleotides and all other oligonucleotide synthesis reagents were purchased from either ChemGenes or Bioautomation (Plano, TX). Tris(hydroxymethyl)aminomethane (Tris)-HCl, ThermoScript reverse transcriptase as well as reaction buffer and NTPs were purchased from Thermo Fisher Scientific (Waltham, MA). DNA and RNA primers and were purchased from Integrated DNA Technologies (Coralville, IA). All oligonucleotide sequences are listed below. All other chemicals were purchased from Sigma Aldrich Corporation (St. Louis, MO).
Gel Electrophoresis

Gels were prepared using the SequaGel – UreaGel system from National Diagnostics (Atlanta, GA). Gels were prepared to 20% acrylamide and scanned using a Typhoon Scanner 9410 (GE Healthcare, Little Chalfont, Buckinghamshire, UK).

Activated monomer and oligomer synthesis

RNA oligonucleotides were prepared by standard phosphoramidite oligonucleotide synthesis using a MerMade 6 DNA/RNA synthesizer (Bioautomation, Plano, TX) or starting with 2’,3’-diacetyl nucleosides using standard manual coupling procedures (Beaucage 1981). 5’-phosphates were installed using bis-cyanoethyl-N,N-diisopropyl phosphoramidite from ChemGenes.

The ancient teachers of this science promised impossibilities, and performed nothing. The modern masters promise very little; they know that metals cannot be transmuted, and that the elixir of life is a chimera. But these philosophers, whose hands seem only made to dabble in dirt, and their eyes to pore over the microscope or crucible, have indeed performed miracles. They penetrate into the recesses of nature, and show how she works in her hiding places. They ascend into the heavens: they have discovered how the blood circulates, and the nature of the air we breathe. They have acquired new and almost unlimited powers; they can command the thunders of
Mononucleotide monophosphates and oligonucleotide monophosphates were activated using a modified published protocol\textsuperscript{23}. As an example, 2-MeImpAGC (the 2-methylimidazolide of the trimer 5′-phosphoro-AGC) was synthesized by first dissolving 5 mg (5 µmoles) of 5′-phosphoro-AGC in 1 mL of dimethyl sulfoxide (DMSO). To that mixture 160 mg (2 mmoles) of 2-methylimidazole, 56 mg triphenylphosphine (260 µmoles), 64 mg 2,2′-dipyridyldisulfide (290 µmoles) and 40 µL triethylamine (550 µmoles) were added. After stirring overnight at room temperature, the mixture was precipitated in 10 mL of a 400:250:30:1 mixture of acetone:diethylether:triethylamine:saturated solution of NaClO\textsubscript{4} in acetone. The precipitate was pelleted by centrifugation (3000 rpm, 5 minutes) and washed twice with a 1:1 mixture of acetone:diethylether and once with pure diethylether. After decanting the solvent, the pellet was dried under vacuum, resuspended in deionized water and purified by high-performance liquid chromatography (HPLC) over a C18 column (Alltima C18 5µm, Thermo Fisher, 250 mm x 10 mm) with 25 mM triethylammonium bicarbonate (TEAB) pH adjusted to 7.5 in 2% (v/v) acetonitrile for mobile phase A and acetonitrile for mobile phase B over a gradient beginning at 100% A and falling to 80% A over 20 minutes with a flow rate of 3 mL/minute. The fraction containing the product was verified by electrospray ionization mass spectrometry.
(ESI/MS) in negative mode (m/z = 996), frozen in liquid nitrogen and lyophilized overnight to yield a white powder. The powder was dissolved in water and the concentration of the activated trimer was determined using a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, Waltham, MA) and calculated assuming $\varepsilon_{260} = 34,170 \text{ M}^{-1} \text{ cm}^{-1}$ (Cavaluzzi 2004).

Monomer stocks were adjusted to pH 7 before adding into the reaction mixture.

**Primer extension reactions**

Nonenzymatic RNA primer extension reactions exploring the effect of downstream nucleotides and oligonucleotides on 2-MeImpG addition were conducted under the following conditions: 200 mM MgCl$_2$, 200 mM sodium N-cyclohexyl-2-aminoethanesulfonic acid (CHES) pH 9, 50 mM monomer, 1 mM trimer, 10 $\mu$M primer and 11 $\mu$M template. The primer sequence was the same for all reactions (6-carboxyfluorescein (FAM)-5′-GAC UGG-3′ for kinetics, cyanine 3 (Cy3)-5′-GCG UAG ACU GAC UGG-3′ for gels). The templates used in Fig. 2.1 had the following sequences:

1) 5′-AAC CCC CCA GUC -3′

2, 4-6) 5′-GUC CCA GUC -3′

3) 5′-GCU GCU GCU CCA GUC AGU CUA CGC-3′
Reaction progress was assessed by gel electrophoresis (20% acrylamide denaturing urea gel) after 10 minutes. Reaction rates were calculated by quantifying primer conversion to products using a Typhoon Scanner 9410 (GE Healthcare, Little Chalfont, Buckinghamshire, UK). Band intensities were quantified using ImageQuant TL software (GE Healthcare, Little Chalfont, Buckinghamshire, UK). The negative log of the fraction of unreacted primer was plotted against time, in hours. A linear regression was performed and the slope of the fit as plotted was reported as the pseudo-first order rate $k_{\text{obs}}$.

Primer extension reactions exploring the effect of different monomers and different oligomers on the rate of polymerization were conducted under the following conditions: 100 mM MgCl$_2$, 200 mM Tris-HCl pH 8, 25 mM monomer, 1 mM trimer, 10 µM primer and 11 µM template. The downstream helper oligomers were added in the following concentrations: 25 mM 2-MeImpA and 1 mM activated oligonucleotide (AG, AGC or AGGC). The reaction progress was assessed by gel electrophoresis (20% acrylamide denaturing urea gel) after 10 minutes. Reaction rates were calculated as above.

Primer extension reactions exploring the addition of multiple monomers and trimers in one reaction had the following conditions: 100 mM MgCl$_2$, 200 mM Tris-HCl pH 8, 25 mM monomers, 500 µM 2-methylimidazole-activated AGC, CGG and GGG, 100 µM activated GCG, 10 µM primer and 11 µM template. The reaction
progress was assessed by gel electrophoresis (20% acrylamide denaturing urea gel) after 18 hours.

2.5 Chapter 2 References


Joyce, G. F., Inoue, T. & Orgel, L. E. Non-enzymatic template-directed synthesis on


CHAPTER 3

Thiolated uridine substrates and templates improve the rate and fidelity of ribozyme-catalyzed RNA copying

3.1 Abstract

Ribozyme-catalyzed RNA polymerization is inefficient and error prone. Here we demonstrate that two alternative bases, 2-thio-uridine ($s^2$U) and 2-thio-ribothymidine ($s^2$T), improve the rate and fidelity of ribozyme catalyzed nucleotide addition as NTP substrates and as template bases. We also demonstrate the functionality of $s^2$U and $s^2$T-containing ribozymes.

3.2 Introduction

The RNA world hypothesis posits that early life forms utilized RNA molecules, both as genetic polymers and as chemical catalysts (Gilbert 1986). Central to this theory is the emergence of an RNA enzyme (ribozyme) capable of catalyzing RNA polymerization (Robertson 2012). The plausibility of such a ribozyme polymerase is supported by the existence of numerous and critical catalytic RNAs in contemporary organisms such as self-splicing introns
(Herschlag 1990), the peptidyl transferase core of the ribosome (Noller 1992) and a multitude of self-cleaving RNAs (Serganov 2007).

Inspired by biological ribozymes, researchers have successfully evolved ribozymes in the laboratory capable of RNA-catalyzed RNA polymerization starting from pools of random oligonucleotides (Bartel 1993, Ekland 1996). In vitro evolution and subsequent engineering have yielded, from such random pools, a class of ribozymes capable of polymerizing RNA in a template directed manner (Johnston 2001), culminating recently with the development of the tC19Z ribozyme which is able to polymerize a strand of RNA longer than itself (Wochner 2011). Despite this progress, even state of the art ribozyme polymerases achieve very low yields of full-length products on templates longer than 10 nucleotides and operate with poor fidelity, with UTP addition across from a G template being the most frequent error due to their propensity to form a stable G-U wobble base-pair (Wochner 2011). Consequently, a ribozyme with the capacity for complete and efficient self-replication has yet to be isolated.

One compelling method for potentially improving the fidelity and efficiency of ribozyme polymerases and ultimately achieving ribozymes capable of self-replication, is to explore the use of non-canonical genetic polymers. Nucleic acids with alternative backbone linkages, sugar structures or nucleobases have been shown to fold into functional enzymes and aptamers
with various advantages compared with their Watson-Crick counterparts (Attwater 2013, Kimoto 2013, Pinheiro 2012, Sefah 2013). Ribozymes with backbones modified to include a mixture of 2′-5′ and 3′-5′ phosphate linkages have also been shown to be active (Cozens 2015, Engelhart 2013). 2′-5′ linkages are a natural byproduct of non-enzymatic polymerization under prebiotic conditions and are more amenable to duplex dissociation following replication, a necessity for subsequent copying of daughter strands. Aptamers evolved to contain one alternative nucleobase, 7-(2-thienyl)imidazo[4,5-b]pyridine, displayed tighter ligand binding than their wildtype counterparts (Kimoto 2013). In a separate study, when two alternative nucleobases, 2-amino-8-(1′-β-D-2-deoxyribofuranosyl)-imidazo[1,2-a]-1,3,5-triazin-4(8H)one and 6-amino-5-nitro-3-(1′-β-D-2′-deoxyribofuranosyl)-2(1H)-pyridone, were used, a tight binding (low nM affinity) aptamer emerged from an earlier round of in vitro evolution than is typical (Sefah 2013).

As the high error rates observed in ribozyme-catalyzed RNA replication are largely a consequence of G-U wobble base pairing (Ekland 1996, Wochner 2011), we chose to explore 2-thio-uridine (s²U) and 2-thio-ribo-thymidine (s²T) as potential alternatives to uridine in the context of increasing ribozyme efficiency and fidelity because these uridine analogues destabilize wobble pairing with guanine and form more stable pairs with adenine (Kumar 1997,
Sheng 2014, Siegfried 2010, Sintim 2006, Testa 1999). We have previously demonstrated the utility of $s^2U$ and $s^2T$ for improving fidelity and efficiency of non-enzymatic RNA polymerization (Heuberger 2015). Furthermore, thiolated uracil-analogues are abundant in biology; notably, $s^2U$ in the tRNA anticodon loop plays a context-dependant role in prohibiting or tolerating wobble pairing (Rodriguez-Hernandez 2013, Vormbrock 1974, Yoshida 1971). $s^2T$ has been found to increase the thermostability of tRNA in archaea (Horie 1985) and thermophilic bacteria (Watanabe 1974). More recently, Attwater et al. showed that $s^2UTP$ may serve as a more facile alternative to UTP as a substrate for ribozyme polymerization; however in the context of the highly evolved tC19Z ribozyme, incorporation of $s^2U$ leads to a block in transcription (Attwater 2013).

3.3 Results and Discussion

We addressed the question of ribozyme efficiency and fidelity in the simplified context of single-nucleotide addition. The b1-233t ribozyme catalyzes the addition of one to three ribonucleotides to an external RNA primer in a template directed fashion (Ekland 1996). Prior characterisation of b1-233t revealed that addition of UTP across from a G (erroneous, wobble-pairing template) proceeds more efficiently than UTP addition across from an A (correct template). We sought to improve the efficiency and fidelity of this ribozyme by
using $s^2$U and $s^2$T, two U analogues with increased affinity for A and reduced affinity for G (Heuberger 2015, Siegfried 2010, Testa 1999). Following the procedure of Ekland and Bartel (Ekland 1996), we combined the b1-233t ribozyme with a fluorescently labeled primer paired to a template with a single, variable nucleotide—the first unpaired template base 3’ of the primer-template duplex (Figure 3.1). UTP, $s^2$UTP or $s^2$TTP were added to a reaction mixture containing the b1-233t ribozyme (2.5 µM), primer (1 µM), template (2 µM) and buffer (30 mM Tris pH = 8, 60 mM MgCl₂, 200 mM KCl and 600 µM EDTA). We then quantified primer extension rates across a range of NTP concentrations for templates containing either an A or a G base at the coding site and determined the Michaelis-Menten parameters $K_M$ and $k_{cat}$.

Consistent with previous reports, UTP is added at comparable rates across from an A-containing template ($k_{cat} = 4.81 \times 10^{-2} \text{hr}^{-1}$, $K_m = 7.34 \text{mM}$, $k_{cat}/K_m = 6.56 \times 10^{-3} \text{mM}^{-1} \text{hr}^{-1}$) or a G-containing template ($k_{cat} = 5.68 \times 10^{-2} \text{hr}^{-1}$, $K_m = 9.31 \text{mM}$, $k_{cat}/K_m = 6.10 \times 10^{-3} \text{mM}^{-1} \text{hr}^{-1}$) (Figure 3.2g,j, 3.3c,d, 3.4a, Table 3.1).

Interestingly, the thiolated UTP analogue $s^2$UTP exhibits more stable binding across from an A-containing template than UTP ($K_M$ UTP = 7.34 mM, $K_M$ $s^2$UTP = 2.84 mM, Figure 3.2g,h, 3.3c,d, Table 3.1). Furthermore $s^2$UTP is polymerized with diminished efficiency across from a G-containing template compared to UTP ($k_{cat}/K_m$ UTP = $6.1 \times 10^{-3} \text{mM}^{-1} \text{hr}^{-1}$, $k_{cat}/K_m$ $s^2$UTP = $1.6 \times 10^{-3} \text{mM}^{-1} \text{hr}^{-1}$),
confirming the hypothesis that substituting \( s^2 \)UTP for UTP in single nucleotide primer extension reactions inhibits the formation of G-templated wobble basepairs (Figure 3.4a).

**Figure 3.1: b1-233t ribozyme schematic.** (a) Schematic of the b1-233t ribozyme. The primer (red) is extended by a single nucleotide (blue) directed by a template (grey). The template is base paired to both the primer and the ribozyme (black). (b) Sequence of the 3′ portion of the primer (red) and the full sequence of the template (grey). N indicates the position in the template that is varied in this study. \( N' \) indicates the NTP. Structures of \( s^2 \)U and \( s^2 \)T are shown.

<table>
<thead>
<tr>
<th>Template</th>
<th>A</th>
<th>G</th>
<th>U</th>
<th>( s^2 )U</th>
<th>( s^2 )T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monomer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( K_M ) (mM)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>-</td>
<td>-</td>
<td>5.4±1.9</td>
<td>3.8±0.7</td>
<td>2.8±0.2</td>
</tr>
<tr>
<td>G</td>
<td>-</td>
<td>-</td>
<td>3.1±0.4</td>
<td>1.9±0.2</td>
<td>4.2±0.7</td>
</tr>
<tr>
<td>U</td>
<td>7.3±0.6</td>
<td>9.3±1.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( s^2 )U</td>
<td>2.8±0.5</td>
<td>4.4±0.4</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( s^2 )T</td>
<td>3.9±0.3</td>
<td>10.2±4.1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( k_{cat} ) (hr⁻¹)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>-</td>
<td>-</td>
<td>0.05±0.009</td>
<td>0.26±0.02</td>
<td>0.16±0.004</td>
</tr>
<tr>
<td>G</td>
<td>-</td>
<td>-</td>
<td>0.016±0.0009</td>
<td>0.16±0.008</td>
<td>0.028±0.002</td>
</tr>
<tr>
<td>U</td>
<td>0.048±0.002</td>
<td>0.057±0.005</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( s^2 )U</td>
<td>0.030±0.002</td>
<td>0.0071±0.0003</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( s^2 )T</td>
<td>0.058±0.002</td>
<td>0.020±0.005</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**Table 3.1: Kinetic parameters.** \( k_{cat} \) and \( K_M \) values and standard errors calculated from the curves in figure 3.2 using Prism 6™.
Figure 3.2: Saturation curves for NTP substrates. For each concentration of NTP a polymerization rate was calculated. These curves were fit to Michaelis-Menten parameters and $k_{cat}$ and $K_M$ values were extracted (Table 3.1).
**Figure 3.3: Kinetic parameters for ribozyme catalyzed primer extension.** Values from Table 3.1 plotted as bar graphs. 

- **a** and **b** display $k_{\text{cat}}$ and $K_M$ values when uridine and uridine analogs are in the template. 
- **c** and **d** display $k_{\text{cat}}$ and $K_M$ values for uridine and uridine analogs as NTP substrates. 

The convention used is N on N', where N is the NTP and N' is in the template.
Figure 3.4: Ribozyme kinetics and specificity. (a) Gel image showing the extension of the primer (bottom band) by the indicated nucleotide triphosphate (blue) when paired with the indicated template uridine analog in the template (grey) after 3 days. Reaction conditions were as specified by Bartel and colleagues (Ekland 1996). In the bar graph below, kcat/KM for each of the pairings is plotted. (b) Same as in (a) but with uridine analogs as NTPs and with A and G in the template. (c) The specificity of the reaction was calculated as the kcat/KM value for the U:A pair divided by that of the U:G pair.

We also observed the same trend for s²TTP: addition of s²TTP across from an A-containing template increased kcat and lowered KM compared to UTP while addition of s²TTP across from a G-containing template was impaired both in addition rate and binding strength compared to UTP (Table 3.1). We found that both s²UTP and s²TTP exhibited at least a six-fold better selectivity for Watson-Crick template primer extension than UTP (Figure 3.4c). As NTP substrates, the thiolated uracil analogues that we tested each substantially improved both the
efficiency and selectivity of ribozyme-catalyzed primer extension. Despite the utility of $s^2$UTP and $s^2$TTP as substrates for single-nucleotide addition, thiolated uracil analogues have been shown, in the context of the tC19Z polymerase ribozyme, to impair the copying of long templates (Attwater 2013). We reason that these nucleobases may interfere with tertiary structural interactions between the growing primer strand and the ribozyme, which was evolved with canonical nucleobases. Further evolution with thiolated nucleobase substrates might overcome this hurdle.

In addition to assessing $s^2$UTP and $s^2$TTP as NTP substrates, we also evaluated $s^2$U and $s^2$T as template coding bases for ribozyme-catalyzed primer extension. Using identical reaction conditions as those described above, we determined kinetic parameters for b1-233t ribozyme-catalyzed ATP and GTP addition to a primer across from a template containing U, $s^2$U or $s^2$T in the coding position (Figure 3.4b, Table 3.1). When U was replaced by $s^2$U as the template coding base for ribozyme-catalyzed primer extension and ATP is used as a substrate, the $K_M$ decreased by 30% and $k_{cat}$ increased five-fold (Figure 3.3a,b, Table 3.1). Though replacing U with $s^2$U in the template coding position increases the overall efficiency of ATP primer extension, it also serves as a more efficient template for GTP, diminishing the selectivity of addition by 55% as compared to U (Figure 3.4b–c). This is in accord with previous studies that have
shown that s²U can, under some circumstances, pair more tightly to G than U does (Johansson 2008, Rodriguez-Hernandez 2013). Replacing U with s²T in the template coding position, however, improves both the efficiency and selectivity of ribozyme-catalyzed primer extension. As a template for ATP addition, s²T reduces the $K_M$ two-fold and increases the $k_{cat}$ threefold as compared to a U-containing template (Figure 3.3a,b, Table 3.1). As a template for GTP addition, a s²T-containing template does not significantly affect the efficiency of primer extension compared to a U-containing template ($p = 0.33$, unpaired t test). Consequently, a s²T-containing template improves both the efficiency and the selectivity of accurate primer extension by six-fold and five-fold, respectively, when compared to a canonical U-containing template (Figure 3.4b,c). Taken together, these data suggest that thiolated uracil analogues generally improve multiple parameters of ribozyme-catalyzed monomer addition both as templates and as substrates.

Given the efficiency of thiolated uracil analogues as templates and substrates of nucleotide addition, we asked whether ribozymes fully substituted with s²U or s²T were capable of catalyzing RNA primer extension (Figure 3.5, 3.6a). Consequently, we sought to synthesize modified versions of the b1-233t ribozyme containing s²U or s²T in place of U. We first transcribed a 26-mer DNA oligonucleotide template containing 8 As with T7 RNA Polymerase (T7-RNAP)
using a reaction mixture comprised of GTP, CTP, ATP and s²UTP (75mM each) and assayed the resulting transcript by liquid chromatography mass spectrometry (LCMS) (Figure 3.5). The major product (m/z = 8989.8828) agreed with the predicted molecular weight for the s²U-containing transcript (m/z =
8989.8827), confirming the previously established (Nakayama 1984) capacity for T7-RNAP to polymerize RNA oligonucleotides using thiolated nucleotide triphosphates as substrates.

Next, we transcribed a DNA template for the b1-233t ribozyme in vitro using T7-RNAP, substituting s²UTP or s²TTP for UTP in the reaction mix. T7-RNAP was able to synthesize full-length transcripts with both s²UTP and s²TTP (Figure 3.6a). We term these new, s²U or s²T-containing versions of the ribozyme b1-233t-s²U and b1-233t-s²T, respectively. We assessed the capacity of these modified ribozymes to catalyze single-nucleotide RNA polymerization. Remarkably, despite analogue substitution at all uracil positions, both b1-233t-s²U and b1-233t-s²T were able to extend an RNA primer, albeit at much slower rates than the wildtype b1-233t ribozyme, with b1-233t-s²T requiring three days to achieve appreciable primer extension at the detection limit of our assay (Figure 3.6b). We reasoned that the class 1 ligase ribozyme, the evolutionary precursor to b1-233t might also be amenable to substitution of s²U and s²T substitution. We generated two variants of the class 1 ligase by replacing UTP with s²UTP or s²TTP in the in vitro transcription reaction and assessed the capacity of each ribozyme to ligate itself to an external primer (Bartel 1993). Again, both thiolated ligase variants were functionally active albeit at rates markedly lower than the wildtype class 1 ligase (Figure 3.6c). Though
functional ribozymes have been evolved to contain non-canonical nucleotides (Hollenstein 2009, Tarasow 1997), to our knowledge, there is only one previous example (Sun 2013) wherein a previously isolated ribozyme can be completely retrofitted with a new nucleobase and maintain functionality. In a diverse prebiotic chemical space, tolerance for chemical substitutions may have conferred catalytic RNAs a fitness advantage.

*In vitro* evolution would likely yield more efficient thiolated-ribozyme polymerases. The ability to reverse-transcribe RNA containing s²U and s²T would be critical for such experiments (Ellington 1990). We tested whether SuperScript™ reverse transcriptase could reverse transcribe b1-233t-s²U and b1-233t-s²T (Figure 3.6d, Table 3.2), and confirmed that RNA containing s²U and s²T can indeed be reverse-transcribed by conventional methods; thus, all of the tools required to evolve a thiolated ribozyme are readily accessible.

Here we have comprehensively assessed s²U and s²T as template bases and as NTP substrates for the b1-233t ribozyme (Ekland 1996). We show that, in the context of this particular ribozyme, s²U and s²T have several advantages over U in terms of both the rate and the fidelity of primer extension. We have also demonstrated that both of these thiolated uracil analogues are compatible substrates for T7 RNA polymerase, thus allowing us to synthesize several ribozymes containing s²U or s²T in place of U. Remarkably, s²U or s²T
substitutions are tolerated in the b1-233t ribozyme at all U residues simultaneously. The capacity for primer extension is maintained in both variants, albeit at substantially reduced efficiency. Not only do thiolated uracil analogues improve several dimensions of RNA-catalyzed nucleotide addition, ribozymes containing s²U and s²T are functionally active and can be generated.
by protein polymerases. These thiolated ribozyme variants can be reverse transcribed paving the way for future \textit{in vitro} evolution of thio-nucleotide substituted ribozymes.

3.4 Materials and Methods

In-vitro transcription

The b1-233t ribozyme (Table 3.2) was \textit{in vitro} transcribed using T7-MEGAscript™ (Invitrogen, Carlsbad CA). Template DNA (Integrated DNA Technologies, Coralville IA) was diluted in water to a concentration of 10µM. 2µl of template DNA was added to an 18µl reaction mixture containing 2µl of each rNTP (75mM) as well 2µl 10x T7-MEGAscript™ reaction buffer, 2µl T7 enzyme mix and 6µl water as per the manufacturer’s instructions. Transcription reactions were carried out overnight at 37°C.

RNA Sequences and Sources

Table 3.2 lists all of the sequences and sources of all oligonucleotides used in this chapter.
Transcript purification

The transcripts generated by *in vitro* transcription (IVT) were purified by denaturing polyacrylamide gel electrophoresis (PAGE). An equal volume of 8M urea (Sigma-Aldrich, St. Louis MO) was added to the crude transcription product. The crude product with urea was heated at 95°C for 2 minutes and run at 55W on a 20% PAGE gel. Desired gel bands were excised from the gel and crushed. RNA was extracted from the gel sections using 500mM NH₄OAc (Ambion®- Thermo Fisher Scientific, Waltham MA). After rocking at 4° overnight, samples were centrifugally purified using a 200 micron filtration column (Merck, Kenilworth NJ) at 13000 rpm for 90 seconds. The flow-through was diluted in 2 volumes of EtOH and precipitated on dry ice (-78°C) for 30 minutes. RNA was separated at 13000 rpm for 30 minutes at 4°C. The supernatant was then aspirated and RNA pellets were re-suspended in 10µl of

<table>
<thead>
<tr>
<th>RNA name</th>
<th>Description</th>
<th>Sequence</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>b1-233t</td>
<td>Ribozyme polymerase</td>
<td>5’-GGAAAAAGACAAAUUCUGGCCUCAGAGUGAAGAACUC UUCUGUGCAGGGAUGCAGAGGCGGCUCGUCGGUUG CUUUAACGGCAACGGUUCUCACAAAGCCA-3’</td>
<td><em>in vitro</em> transcription</td>
</tr>
<tr>
<td>b1-233t_G</td>
<td>Ribozyme template (unpaired 5’ G)</td>
<td>5’-GGCUAUAGGACUGGAACCA-3’</td>
<td>IDT</td>
</tr>
<tr>
<td>b1-233t_U</td>
<td>Ribozyme template (unpaired 5’ U)</td>
<td>5’-GGCUAUAGGACUGGAACCA-3’</td>
<td>IDT</td>
</tr>
<tr>
<td>Class 1 ligase</td>
<td>Ligase ribozyme</td>
<td>5’-GGACACUAUAAGGGGUAAAGAGAGA AAAUUCUGACUCAGCUUGAGAACACCUJUCG GAUGCAGGGAAGCAGGCCCGGUGUCUUAACGCCAACGGUUCACAAAGUGA-3’</td>
<td><em>in vitro</em> transcription</td>
</tr>
<tr>
<td>b1-233t_P</td>
<td>Fluorescien labelled b1-233t primer</td>
<td>5’-Fluor-dAdAdAdACCUGUC-3’</td>
<td>IDT</td>
</tr>
</tbody>
</table>

**Table 3.2: RNA Sequences and Sources**
purified water. RNA concentration was determined by spectrophotometry using a NanoDrop™ 2000 instrument. Purified RNA was diluted in distilled water to a working concentration of 10µM.

**Ribozyme Reaction Conditions**

The b1-233t ribozyme polymerase and thiolated-uracil variants, were generated by IVT and used to assay the rate of primer extension. Primer extension reactions were carried out at a final volume of 20µl in 200µl PCR tubes. Reactions were conducted in a final concentration of 300µM Tris (pH = 8), 600µM MgCl₂ and 2M KCl. Ribozyme RNA was heated at 95°C for 2 minutes and cooled to room temperature gradually to allow secondary structure formation. 2 µl of ribozyme RNA was then added to the reaction for a final concentration of 1 µM. The desired template as well as the desired monomer (either ATP, GTP, UTP, s²UTP or s²rTTP) were both added to the reaction mix (1 µM final) as well as a 5’ fluorescein labeled primer (2 µM final) (See Table 3.2 for sequences). Reactions proceeded at room temperature. At the designated time points, 5µl of the reaction was removed and stored at -80°C. The class 1 ligase ribozyme and thiolated-uracil variants were generated by IVT. Ligation reactions were carried out at room temperature by combining each ligase variant (1 µM final) with the 5’ fluorescein labeled primer (0.5 µM) in ligation reaction buffer (30mM TRIS pH = 8, 60mM MgCl₂, 200mM KCl, 600µM EDTA) unless otherwise specified. (See Table 3.2 for sequences).
Quantifying Primer Extension

Ribozyme reactions were quenched in 100mM EDTA and 8M urea. Samples were subjected to denaturing PAGE at 32W. Gels were scanned with a Typhoon 9410 scanner (GE Healthcare, Little Chalfont UK) using a fluorescein filter ($\lambda_{nm} = 526$-$532$) at a resolution of 100 microns. Primer extension was quantified using ImageQuant 5.2 software (GE Healthcare). Each gel was background corrected and the extension ratio was calculated as the fraction of the plus one product divided by the un-extended primer.

Determining Kinetic Parameters

Individual timepoints were plotted in Excel™ (Microsoft, Redmond WA). The natural log was determined for each timecourse and the slope of a linear fit was taken as the rate. The rates for each concentration series were then plotted in Prism 6™ (GraphPad, San Diego CA) and standard Michaelis-Menten parameters $k_{cat}$ and $K_m$, along with standard errors, were calculated.

Liquid Chromatography-Mass Spectrometry

IVT was performed in accordance with the above method, replacing U with an equal concentration of $s^2$U. Yield was compared by 20% denaturing PAGE. Gels were scanned and quantified in accordance with the previously described procedure. $s^2$U incorporation was verified by Liquid Chromatography-Mass Spectrometry (LCMS).
LCMS was performed with a 6520 Q-TOF mass analyzer (Agilent Technologies, Santa Clara CA) and 1200 series High Performance Liquid Chromatograph (HPLC) with an XBridge C18 column (3.5 µm, 1x100 mm, Waters, Milford MA). Mobile phase A was aqueous 200 mM hexafluorisopropanol and 3 mM Tris/Borate/EDTA at pH = 7, and mobile phase B was methanol. HPLC was performed for 10 µL of a 200 µM solution at a linear increase of 5% to 20% B over 30 min at 0.1 mL/min at 60 °C. A$_{260\text{nm}}$ was used to monitor sample elution and the eluate was passed directly to an electrospray ionization source with 325 °C drying nitrogen gas flowing at 8.0 L/min, a nebulizer pressure of 30 psig and a capillary voltage of 3500 V. Agilent MassHunter Qualitative Analysis software was used for Q-TOF derived MS data and the spectra were deconvoluted using the maximum entropy method.

3.5 Chapter 3 References


CHAPTER 4

A novel, self-replicating, information-bearing polymer

4.1 Abstract

Nucleic acid polymers are the only polymers known that can contain information and be copied. In order to demonstrate that these two properties are present elsewhere in chemical space, we have synthesized a synthetic polymer with a specified monomer at each position, which can also template its own replication. In place of hydrogen-bond-mediated base pairing found in nucleic acids, we employed reversible imine bond formation. The repeating aryl-yne backbone of the polymer is readily synthesized using standard Sonogashira chemistry. Here we demonstrate the copying of this polymer in a model dimer system. This polymer, or others like it, provide alternatives to DNA for arbitrary data storage. Additionally, this polymer exhibits all of the informational properties that allow DNA to serve as the basis for heredity in life on earth.
4.2 Introduction

The storage of information in polymers is a ubiquitous feature of life on earth. In addition to DNA and RNA – proteins, oligosaccharides, polyketides and polyterpenes are just some classes of sequence-defined biological polymers. Outside of biology, creating sequence-defined polymers has proven to be much more challenging. Synthetic polymers capable of storing information have been developed (Young 1994) and recent techniques have allowed for the synthesis of a wide diversity of sequence-controlled polymers (Lutz 2015, Ouahabi 2015, Roy 2015, Colquhoun 2014, Lutz 2013, Zhu 2010, Elliot 2006, Lewandowski 2013, Lutz 2014). Typically, these polymers are synthesized in the same manner as oligonucleotides or peptides, i.e. by making use of protecting groups to allow for the addition of monomers one at a time; most of these polymers also have only two letters (binary) for the sake of simplicity.

Unlike all other information-bearing polymers, DNA and RNA can be copied with high fidelity. This ability is essential in their capacity as the informational repositories of life. DNA and RNA, through the geometry of base pairing, serve as templates for their own duplication. Unlike protein synthesis, which requires an adapter molecule, tRNA, DNA and RNA are synthesized directly on a template strand. Polymerases are required for this process, but primarily in their capacity to lower the activation barrier for the phosphoryl-
transfer chemistry required in oligonucleotide polymerization; the recognition of correct base pairing is largely achieved by the template strand. In life today proteins comprise most of the catalytic machinery of cells, but there is no mechanism by which proteins can serve as templates for their own synthesis; for this reason the information required to produce proteins, and ultimately all other sequence-defined biopolymers, is stored in DNA.

Other sequence-defined polymers capable of templated copying have been discovered. However, they have all been modified nucleic acids, which are as structurally different from DNA as RNA is (Pinheiro 2012, Eschenmoser 2011, Ura 2009, Lincoln 2009, Hayden 2008, Isobe 2008). The replication of these xeno-nucleic acid (XNA) polymers is achieved with mutated transcriptases and reverse-transcriptases whose normal function in biology is to convert between DNA and RNA; the nucleobases of XNA polymers are identical to those found in DNA and RNA. Aside from nucleic acid polymers, no example of template-directed sequence-copying of a polymer has been discovered or constructed.

In an attempt to synthesize such a polymer capable of template-directed sequence-copying outside of the nucleic acid chemical space, we first must choose alternative chemical reactions to those found in the base-pairing and polymerization reactions of nucleic acids (Fig. 4.1). Importantly, the reaction, or set of reactions, we choose for base pairing must be reversible. In nucleic acids,
base pairing is achieved by highly reversible hydrogen bonds, which equilibrate easily. However, base pairing in nucleic acids is weak and in order for polymerization to proceed, nucleotide monomers are bound to pockets within polymerases that contribute additional non-covalent binding interactions that assist in reducing the dissociation rate sufficiently to allow for polymerization to occur (Westover 2004). For that reason we decided that, since we would not have the benefit of a polymerase, a stronger base-pairing interaction would be necessary. We chose reversible covalent bond formation for this purpose.

Figure 4.1: Informational polymer copying scheme. A ladder-like duplex structure held together by base pairs (left; wedges: base pairs, black circles: polymer links) can be split into two strands complemented by free monomers (middle) upon the addition of an excess of free monomers (blue) that can exchange with the base pairs of the full duplex. Finally, these complemented complexes can polymerize back into duplexes as linkers (blue circles) connect adjacent monomers to one another.

Any polymer consisting of at least two different subunits can be used to encode arbitrary information. In the name of simplicity we decided that a single base pair, consisting of two different constituent monomers, would be suitable
for the construction of a synthetic genetic polymer (Fig. 4.1). For this purpose we chose imine formation (and dissociation) as our base-pairing interaction; imines form and dissociate readily and are robust to diverse chemical conditions. Our two-letter alphabet therefore contains just the functional groups of aldehyde and amine; from an information perspective these can be thought of as ones and zeros.

In choosing backbone polymerization chemistry we were guided by two principles: efficiency of chemical reactivity and simplicity. Since the backbone links need not be reversible we looked towards so-called “click” reactions. Among those, the Sonogashira reaction (Sonogashira 2002) involved components (aryl iodides and terminal alkynes) that were particularly easy to either find commercially or to generate synthetically. Thus, we chose to build a stable aryl-yn backbone using standard Sonogashira chemistry.

Once those two reactions were chosen we designed a system wherein the requisite functional groups are connected. For every monomer there must be one base-pairing functional group and two backbone functional groups. We designed monomers such that they could not react with one another until we added a separate connector molecule (Scheme 4.1). R groups in the linker can be chosen to modulate solubility, just as the phosphate group in DNA and RNA maintain high levels of solubility in water; additional chemical functionality
(e.g. fluorescence, electrical charge, reactive functional groups, etc.) may be added as desired. We rely on proximity to ensure that the sequence is copied correctly as only adjacent monomers will react with one another due to their high effective molarity. Unlike DNA and RNA, which are directional, containing 3’ and 5’ ends, each link in this system is symmetric. This symmetry has several consequences; for example, sequences that are the reverse of one another are indistinguishable (e.g. 10010 and 01001 are the same molecule). This reduces the informational capacity of our polymer by half. However, since every additional polymer unit doubles the informational capacity of this polymer, in order to encode the same amount of information as a directional polymer only a single additional unit would have to be added.

Molecules with a similar structural framework as in Scheme 4.1 have been studied by Moore and colleagues for their self-assembly properties (Zhao 2002a,b,c, Hartley 2007, Elliot 2011). Only homopolymers (polymers containing exclusively aldehydes or anilines) have been studied, but the conclusions reached about base pairing strengths are likely generalizable. A recent study from the Scott laboratory explored polymers with the same base pairing chemistry but a different backbone structure (Wei 2015).

The structures in this study were chosen with ease of synthesis in mind, we would like to emphasize that we anticipate that many other structures would also
Scheme 4.1: Proposed monomers, connector and backbone. In order to create a system that matches the outline described in Figure 4.1, we designed the scheme above. Benzaldehyde and aniline monomers (in orange and green respectively) are joined by conjugated dialkynyl-aryl connectors. Information copying begins with the splitting of a fully duplexed pair of polymers, in which all constituent monomers are bound through imine bonds to their complements in the opposite strand (top). Dialkynyl monomers are added which split the duplex into two complexes wherein each constituent monomer is now bound to a free monomer (bottom). Finally, the original duplex is reproduced upon polymerization with a diiodo connector molecule.

satisfy the criteria we have laid out (Scheme 4.2). Other base pairing chemistries we believe may work include boronate ester condensation, reversible Diels-Alder
reactions and thioester formation. For the backbone there are even more possible chemical structures including azide-alkyne click chemistry, olefin methathesis or even simple S$_2$N$_2$ chemistry. In this study we develop a synthetic genetic polymer composed of imine base pairs and a Sonogashira backbone to serve as a test case against which other chemistries can be compared.

**Scheme 4.2: Potential alternative backbone and base pair structures.** Presented in comparison with RNA are several additional alternative backbone and base pair structures. This is a small set of possible examples and does not explore the many possible structures that could connect the backbone to the base pair. The disulphide chemistry is distinct from the remaining chemistries in that it is symmetric. In order for it to carry information it would have to be paired with an additional, orthogonal base pair.
4.3 Results and Discussion

In order to study the base pairing and polymerization reactions in a model duplex, we set out to generate a model system that consisted of only two base pairs. This would allow us to optimize base pairing and polymerization chemistries in an analytically tractable way (Scheme 4.3).

Scheme 4.3: Homodimeric macrocycle. B1 and B1’ were generated by general procedure A (see section 4.4.1.1, 4.4.2.2, 4.4.2.3) and combined to form macrocycle A1. B1 and B1’ were combined at a ratio of 1:1 in chloroform and heated to 120 °C under a gentle flow of argon. The resulting brown mixture was resuspended in chloroform and crystallized by the addition of benzene to produce pure A1.

Because of the symmetric nature of our polymer, there are only 3 possible single-stranded dimer molecules, B1, B1’, and I2 (Schemes 4.3 and 4.4). By the same token, there are only two possible duplexes, A1 and A2 (Schemes 4.3 and 4.4). B1 and B1’ are complements of one another and, through replication, act as templates for the synthesis of one another. Scheme 4.3 shows how we generated macrocycle (duplex) A1 for the purpose of initiating a replication cycle. In order to study the structure of A1 in greater detail, we obtained a crystal structure of A1 (Fig. 4.2). Notably, the nitrogen atoms (blue spheres) found in the two imine bonds (base pairs)
are unexpectedly both found pointing outwards. For steric reasons (Scheme 4.3) we expected to find one pointing into the center of the macrocycle A1.

Scheme 4.4: Heterodimeric macrocycle. P2 was synthesized by general procedure A (see sections 4.4.1.1 and 4.4.2.1). A mixture of I2 and A2 was generated by general procedure A (see section 4.4.1.1). This mixture was dissolved in chloroform and heated to 120 °C under a gentle flow of argon. The resulting brown mixture was resuspended in chloroform and crystallized by the addition of benzene to produce pure A2.

We initiated replication cycles at the duplex stage in order to demonstrate a complete replication cycle (Fig. 4.3). Transimination of the imine bonds of the parent duplex A1 separated the two strands while simultaneously providing the amine complement necessary for replication. The equilibrium between imines formed from the amines of B1′ and the aldehydes of B1 could be driven nearly entirely (>95% as access by 'H NMR) to complement dimer C1 and diamine B1′ by the addition of 30 equiv. of 3-ethynylaniline and a catalytic amount of TFA (Fig. 4.3). This equilibrium required a large excess of 3-ethynylaniline to maintain. As such, attempts to isolate C1 by removing the excess of 3-ethynylaniline in presence B1′ resulted in the reverse transimination reaction and the reformation of duplex A1 along with various polymers. The mixture of C1 and B1′ was therefore characterized in presence of 3-
ethynyl aniline (Section 4.4.1.2 below). Figure 4.3 shows the $^1$H NMR and MALDI MS spectra of the resulting mixture. We compared these spectra with those of isolated C1 to confirm the assignment.

Complementation of B1, as well as the excess 3-ethynylaniline, was accomplished by the addition of 60-fold excess of 3-ethynylbenzaldehyde. 4Å molecular sieves were found sufficient to prevent hydrolysis of C1 and drive the imine condensation reactions to completion, resulting a mixture of 1:1:30:28, C1:C1′:3-ethynylbenzaldehyde:dimonomerimine (general procedure B, section 4.4.1.2 below). The excess monomers and base paired monomers were removed by heating the mixture to 120 °C under a vacuum pressure of $2 \times 10^{-2}$ mbar. The solution, now devoid of amine nucleophiles, is not in equilibrium; the excess of aldehydes ensures that during the heating under vacuum no free amines are produced. If free amines were present during heating, the system would equilibrate, reforming duplex A1. Further NMR and MS spectra for this mixture were obtained (Fig. 4.3) before macrocyclization.

In order to distinguish the product of a templated macrocyclization reaction from the transamination of C1 and C1′, which would reform A1, we subjected the mixture of C1 and C1′ to general procedure C (4.4.1.3) in the presence of methyl-3,5-diiodobenzoate rather than 2-(2-methoxyethoxy)ethyl 3,5-diiodobenzoate. The methylester served as a mass tag: the resulting macrocycles, D1 and D1′ are distinguishable from A1 by both $^1$H NMR and MALDI-MS allowing us to verify that
we are observing a macrocyclization reaction and not simply a Sonogashira coupling of free monomers. In order to achieve replication we could have instead used 2-(2-methoxyethoxy)ethyl 3,5-diiodobenzoate as the linker.

We dissolved the purified mixture of C1 and C1’ in DMF to allow for a high reaction temperature (60 °C) at a low concentration (6 μM) to favor macrocyclization. The solution was rigorously degassed to prevent Glaser coupling of the terminal alkynes. After the addition of 5 eq. methyl-3,5-diiodobenzoate, 10 eq. DABCO, 1 eq. Pd(PPh₃)₄ and 0.5 eq. CuI the reaction was heated to 60 °C for 16 hours.

Figure 4.2: Crystal structure of A1.
Figure 4.3: Templated copying of the homodimeric system. (a) Scheme (b) Steps a and b were conducted in an NMR tube containing 0.5 mL 0.02% TFA/CDCl₃ and 10 mg 4 Å mol. sieves, starting with 0.01 M A₁. a) 30 equiv 3-ethynylaniline. b) 60 equiv 3-ethynylbenzaldehyde. c) 2x10⁻² mbar, 120 °C, neat. d) 5 equiv methyl 3,5 diiodobenzoate, 1 equiv Pd(PPh₃)₄, 0.5 equiv CuI, 6 uM degassed DMF, 60° C.

The resulting mixture was recrystallized in benzene and ¹H NMR analysis yielded a spectrum of macrocycles that is distinct from the one in the ¹H NMR spectrum of A₁. The ¹H NMR peaks of the similar, but not identical, macrocycles D₁ and D₁’ are not distinguishable. MALDI-MS analysis revealed a single peak with m/z = 811.438, in accord with the calculated molecular weight of 811.28 for both D₁ and D₁’ (they are isomers and therefore share a mass). Notably absent were peaks at m/z =
899.33 or 723.23 indicating \textbf{A1} or a macrocycle with methyl-ester links on both halves respectively. This demonstrates that no transimination occurred during the reaction; if Sonogashira coupling were occurring off of our templates and the dimers were in equilibrium, a mixture of all three masses would be expected at a ratio of 1:2:1.

Satisfied that we had successfully copied the \textbf{A1} system, we turned our focus to macrocycle \textbf{A2}, a heterodimeric duplex (Scheme 4.4). \textbf{A2} is an isomer of \textbf{A1} where a single imine bond is reversed. To synthesize \textbf{A2} we first synthesized precursor \textbf{P2} by general procedure A (4.4.2.1). \textbf{P2} was reacted with 3-ethynylbenzaldehyde to produce a mixture of the heterodimer \textbf{I2}, the macrocycle \textbf{A2} and some unidentified polymers. Heating and recrystallization furnished pure \textbf{A2}.

\textbf{A2} was subjected to precisely the same procedure as \textbf{A1}. Unlike \textbf{A1}, however, the symmetry of \textbf{A2} caused there to be only one set of molecules along the replication pathway (Fig. 4.4). Thus, $^1$H NMR and MALDI-MS spectra show only one set of peaks throughout. After addition of excess 3-ethynylaniline (general procedure B, 4.4.1.2) \textbf{B2} was produced. Subsequent addition of 3-ethynylbenzaldehyde followed by heat-purification delivered pure \textbf{C2} (4.4.1.2). Finally, a Sonogashira coupling with methyl-3,5-diodobenzoate followed by recrystallization (4.4.1.3) generated \textbf{D2}. $^1$H NMR and MALDI-MS confirmed each of these transformations (Fig. 4.4).
Having discovered reaction conditions capable of copying a dimer, we next intend to turn our attention to longer polymers. As the polymers increase in length we will particularly monitor the effect of length on solubility. As the polymers grow longer, the possibility for macrocyclizations other than the ones leading to the desired duplexes (Scheme 4.1) will become conformationally feasible. If difficulties of this sort are encountered two strategies may be employed to combat them: first, it is possible
to continue to adjust the reaction conditions. By modulating the temperature, concentration, metal ligands and the identity of the base, the Sonogashira reaction kinetics can be biased towards the production of the desired product. The second strategy is to modify the backbone or base pair structures. Some additional options can be found in scheme 4.2. Of particular interest are backbones that have fewer rotatable bonds, in particular, 1,4-aryl linkers would result in polymers with extremely long persistence lengths, a feature that might be of particular benefit for information storage integrity.

By demonstrating the templated copying of a sequence specific polymer, we have shown, in this study, that structures that share no structural relationship to DNA, may serve as informational polymers. One implication of this work for origin of life scenarios on earth and elsewhere is that the structures of DNA and RNA are not uniquely suited to the transmission of information. DNA and RNA may be more likely than other polymers to form in various prebiotic environments (on earth or extra-terrestrially); however, since there is only one form of life from which to draw this conclusion the exploration of alternatives like this one may help inform a more comprehensive catalog of the possible space of informational polymers.

We hope to demonstrate in the future that the reaction conditions we have outlined in this study are compatible with longer polymers of the aryl-yne backbone from this study. Longer homopolymers with this structure have been synthesized
(Elliott 2011) but no heteropolymer syntheses have yet been reported. If we can demonstrate the efficient copying of longer polymers we will next attempt to demonstrate some method by which polymer sequences can be determined (see chapter 5). Since backbone linkers can be varied between rounds of replication (e.g. methyl-3,5-diiodobenzoate vs. 2-(2-methoxyethoxy)ethyl 3,5-diiodobenzoate section 4.4.1.3 below) we are hoping to explore the transfer of information from one backbone structure to another. Finally, if polymers like these can fold into functional catalysts in a sequence dependent fashion and if they can be encapsulated, then they would be suitable substrates for the genetic material of a protocell.

4.4 Materials and Methods

4.4.1 General procedures

4.4.1.1 General procedure A – Sonogashira chemistry

\[
\text{Ar}^1 \overset{\text{Pd(PPh}_3\text{)}_4, \text{CuI, TEA}}{\longrightarrow} \text{Ar}^2 - I + \text{Ar}^1 - \overset{\text{CuI}}{\longrightarrow} \text{Ar}^1 - \text{Ar}^2
\]

The aryl alkyne and aryl iodide, at the given ratio of equivalents, were dissolved in tetrahydrofuran (0.1 M) in a screw-capped vial containing a magnetic stir bar. 5 equiv. of TEA, with respect to the aryl iodide, was added followed by 10 mol % Pd(PPh\textsubscript{3})\textsubscript{4} and 5 mol % CuI. The vial was sealed with a screw cap and left stirring between 1 and 24 hours at room temperature. The reaction was monitored by silica TLC (10% EtOAc in hexanes) until it reached full conversion whereupon it was diluted
in ethyl acetate and mixed with silica. The solvent was removed in vacuo and the crude mixture was purified by silica gel flash chromatography using a CombiFlash (Teledyne Isco Lincoln, NB).

4.4.1.2 General procedure B – Duplex complementation

Step 1 – Amine complementation (a)

Approximately 1 mg of dissolved in 0.5 mL 0.02% TFA/CDCl$_3$ over 10 mg 4 Å molecular sieves in an NMR tube. 30 equivalents of 3-ethynylaniline were added to the tube and transimination was monitored by $^1$H NMR (500MHz) until the reactants reached equilibrium and the amount of the duplex was reduced to <3% (approximately 2 hours). A 1 µL aliquot of the solution was subjected to MALDI-TOF spectroscopy.
To the NMR tube an additional 20 mg of 4 Å molecular sieves and 60 equivalents of 3-ethynylbenzaldehyde were added. Imine formation was monitored by $^1$H NMR until the free amines of both the polymer and the excess 3-ethynylaniline were consumed (about 2 hours). The molecular sieves were pelleted via centrifugation and the CDCl$_3$ supernatant decanted into a 5 mL glass vial. The solvent was removed under a stream of N$_2$ air and the resulting white solid sublimated under reduced pressure at 120 °C for 16 hours using a BÜCHI GKR-50 glass tube oven (BÜCHI Labortechnik Flawil, Switzerland). The sublimate was found to contain a 1:1 mixture of 3-ethynylbenzaldehyde and (E)-N,1-bis(3-ethynylphenyl)methanimine with >99% mass recovery. The unsublimated residue was found to contain nearly pure polymer complement as determined by $^1$H NMR and MALDI-TOF spectroscopy.
To the polymer complement formed via General Procedure B, 5 equivalents of methyl-3,5-diiodobenzoate and 10 equivalents of DABCO were added and the relative stoichiometry was verified by $^1$H NMR. The mixture was diluted in rigorously degassed anhydrous DMF to a concentration of 6 µM. 0.5 mg/mL 4 Å molecular sieves, 1 equivalent of Pd(PPh$_3$)$_4$, and 0.5 equivalents of CuI were added, and the reaction was allowed to stir under N$_2$ at 60 °C for 16 hours. The DMF was removed in vacuo, the residue was taken up in CHCl$_3$ (10 mL), the 4 Å molecular sieves were pelleted via centrifugation, the CHCl$_3$ was decanted and then removed in vacuo. The resulting light brown oil was recrystallized from a 3:1 mixture of benzene:CHCl$_3$ (1 mL). The resulting colorless crystals were characterized by $^1$H NMR and MALDI-TOF spectroscopy.

4.4.2 Chemical syntheses and characterizations
4.4.2.1 2-(2-methoxyethoxy)ethyl 3-((3-aminophenyl)ethynyl)-5-iodobenzoate

(P2)

Prepared using general procedure A: 3-ethynylaniline (50 µL, 0.44 mmol) was reacted with 2-(2-methoxyethoxy)ethyl 3,5-diiodobenzoate (1.2 g, 5.7 eq., 2.52 mmol), in the presence of Pd(PPh₃)₄ (60 mg, 0.06 mmol), CuI (5 mg, 0.026 mmol), TEA (2 mL) in tetrahydrofuran (4 mL) overnight at room temperature. CombiFlash silica column purification gave (P2) (gradient of 0-100% EtOAc in Hexanes, 120 mg, 58%) as a light brown oil. The identity of P2 was confirmed by ¹H-NMR (400 MHz, CDCl₃) δ 8.32 (dd, J = 1.6, 1.6 Hz, 1H), 8.14 (dd, J = 1.6, 1.6 Hz, 1H), 8.03 (dd, J = 1.6, 1.6 Hz, 1H), 7.14 (dd, J = 8.1, 7.5 Hz, 1H), 6.93 (ddd, J = 7.7, 1.6, 1.0 Hz, 1H), 6.84 (dd, J = 2.4, 1.6 Hz, 1H), 6.69 (ddd, J = 8.1, 2.4, 1.0 Hz, 1H), 4.53 – 4.47 (m, 2H), 3.87 – 3.81 (m, 2H), 3.72 (br s, NH2), 3.72 – 3.67 (m, 2H), 3.60 – 3.55 (m, 2H), 3.40 (s, 3H); ¹³C NMR (100 MHz, CDCl₃) δ 164.8, 146.6, 144.3, 138.1, 132.2, 132.1, 129.6, 125.9, 123.3, 122.4, 118.0, 116.1, 93.4, 92.2, 86.3, 72.1, 70.8, 69.3, 64.8, 59.4.
4.4.2.2  2-(2-methoxyethoxy)ethyl 3,5-bis((3-aminophenyl)ethynyl)benzoate
(B1')

Prepared using general procedure A:  3-ethynylaniline (236 µL, 2.1 mmol) was
reacted with 2-(2-methoxyethoxy)ethyl 3,5-diiodobenzoate (200 mg, 0.2 eq., 0.42
mmol), in the presence of Pd(PPh₃)₄ (10 mg, 0.01 mmol), CuI (2 mg, 0.01 mmol), TEA (1
mL) in tetrahydrofuran (2 mL) overnight at room temperature.  CombiFlash silica
column purification gave (B1') (gradient of 0-100% EtOAc in Hexanes) as a light
brown oil.  The identity of B1' was confirmed by ¹H NMR (500 MHz, Chloroform-d) d
8.13 (d, J = 1.6 Hz, 2H), 7.82 (t, J = 1.6 Hz, 1H), 7.15 (dd, J = 8.1, 7.6 Hz, 2H), 6.95 (ddd, J
= 7.6, 1.6, 1.0 Hz, 2H), 6.86 (dd, J = 2.5, 1.6, 2H), 6.69 (ddd, J = 8.1, 2.5, 1.0 Hz, 2H), 4.55 –
4.49 (m, 2H), 3.89 – 3.83 (m, 2H), 3.75 (br s, 4NH) 3.74 – 3.69 (m, 2H), 3.61 – 3.56 (m,
2H), 3.40 (s, 3H); ¹³C NMR (125 MHz, cdcl3) d 165.6, 146.5 (2C), 138.5, 132.3 (2C), 130.9,
129.6 (2C), 124.4 (2C), 123.5 (2C), 122.4 (2C), 118.0 (2C), 116.0 (2C), 91.3 (2C), 87.2 (2C),
72.1, 70.8, 69.4, 64.7, 59.3.
Prepared using general procedure A with a 5 eq. excess of 3-ethynylbenzaldehyde over 2-(2-methoxyethoxy)ethyl 3,5-diiodobenzoate. The identity of B1 was confirmed by $^1$H NMR (500 MHz, Chloroform-d) $\delta$ 10.03 (s, 2H), 8.20 (d, $J = 1.6$ Hz, 2H), 8.05 (dd, $J = 1.9$, 1.4 Hz, 2H), 7.90 – 7.86 (m, 3H), 7.79 (ddd, $J = 7.7$, 1.4, 1.4 Hz, 2H), 7.56 (dd, $J = 7.7$, 7.7 Hz, 2H), 4.56 – 4.51 (m, 2H), 3.90 – 3.85 (m, 2H), 3.73 – 3.70 (m, 2H), 3.61 – 3.57 (m, 2H), 3.40 (s, 3H); $^{13}$C NMR (125 MHz, cdcl3) $\delta$ 191.6 (2C), 165.3, 138.6, 137.4 (2C), 136.7 (2C), 133.2 (2C), 132.9 (2C), 131.2, 129.6 (2C), 129.5 (2C), 124.0 (2C), 123.9 (2C), 89.6 (2C), 89.1 (2C), 72.1, 70.8, 69.4, 64.8, 59.3.
2-(2-methoxyethoxy)ethyl 3,5-bis((3-((E)-3-ethynylbenzylidene)amino)phenyl)ethynyl)benzoate

(C1)

C1’ was prepared by the addition of an excess of 3-ethynylbenzaldehyde to B1’ in chloroform. The mixture was heated to 120 °C to form the desired imine bonds as well as to remove by evaporation both the chloroform and the excess 3-ethynylbenzaldehyde. The identity of C1’ was confirmed by $^1$H NMR (500 MHz, Chloroform-d) δ 8.45 (s, 2H), 8.18 (d, J = 1.6 Hz, 2H), 8.04 (dd, J = 1.5, 1.5 Hz, 2H), 7.91 (ddd, J = 7.8, 1.5, 1.5 Hz, 2H), 7.88 (t, J = 1.6 Hz, 1H), 7.62 (ddd, J = 7.7, 1.5, 1.5 Hz, 2H), 7.48 – 7.37 (overlap, 8H), 7.23 (ddd, J = 7.4, 2.2, 1.6 Hz, 2H), 4.55 – 4.51 (m, 2H), 3.89 – 3.85 (m, 2H), 3.73 – 3.70 (m, 2H), 3.61 – 3.57 (m, 2H), 3.40 (s, 3H), 3.14 (s, 2H).
C1 was prepared by the addition of an excess of 3-ethynylaniline to B1′= in chloroform. The mixture was heated to 120 °C to form the desired imine bonds as well as to remove by evaporation both the chloroform and the excess 3-ethynylaniline. The identity of C1′ was confirmed by $^1$H NMR (500 MHz, Chloroform-d) d 8.45 (s, 2H), 8.20 (d, J = 1.6 Hz, 2H), 8.10 (dd, J = 2.0, 1.4 Hz, 2H), 7.90 (overlap, 3H), 7.67 (ddd, J = 7.7, 1.4, 1.4 Hz, 2H), 7.50 (dd, J = 7.7, 7.7 Hz, 2H), 7.41 – 7.33 (overlap, 6H), 7.23 (dt, J = 7.1, 2.0 Hz, 2H), 4.56 – 4.51 (m, 2H), 3.90 – 3.86 (m, 2H), 3.74 – 3.70 (m, 2H), 3.61 – 3.58 (m, 2H), 3.41 (s, 3H), 3.11 (s, 2H).
A1 was formed by the addition of an equimolar mixture of B1 and B1’ in chloroform. The mixture was heated to 120 °C to form the desired imine bonds as well as to evaporate the chloroform. Finally, A1 was dissolved in chloroform and recrystallized in benzene. The identity of A1 was confirmed by $^1$H NMR (500 MHz, Chloroform-d) $d$ 8.58 (s, 2H), 8.18 (d, $J = 1.6$ Hz, 2H), 8.17 (d, $J = 1.6$ Hz, 2H), 8.13 (dd, $J = 1.8$, 1.6 Hz, 2H), 7.938 (ddd, $J = 7.7$, 1.7, 1.4, 2H), 7.933 (t, $J = 1.6$ Hz, 2H), 7.91 (t, $J = 1.6$ Hz, 1H), 7.66 (ddd, $J = 7.7$, 1.6, 1.2 Hz, 2H), 7.52 (dd, $J = 2.1$, 1.6 Hz, 2H), 7.50 (dd, $J = 7.7$, 7.7 Hz, 2H), 7.45 (ddd, $J = 7.5$, 1.6, 1.5 Hz, 2H), 7.42 (dd, $J = 7.5$, 7.5 Hz, 2H), 7.34
(ddd, J = 7.5, 2.1, 1.5 Hz, 2H), 4.58 – 4.51 (overlap, 4H), 3.92 – 3.86 (overlap, 4H), 3.77 – 3.70 (overlap, 4H), 3.63 – 3.59 (overlap, 4H), 3.42 (s, 6H); $^{13}$C NMR (125 MHz, cdcl3) d 165.5, 165.4, 159.3 (2C), 151.0 (2C), 139.4, 139.2, 136.6 (2C), 134.3 (2C), 133.0 (2C), 132.24 (2C), 132.22 (2C), 131.15, 131.09, 129.8 (2C), 129.6 (2C), 129.3 (2C), 128.8 (2C), 124.9 (2C), 124.2 (2C), 124.1 (2C), 123.8 (2C), 123.6 (2C), 122.2 (2C), 90.8 (2C), 90.5 (2C), 88.6 (2C), 88.2 (2C), 72.2 (2C), 70.8 (2C), 69.4 (2C), 64.75, 64.73, 59.4.

The structure of A1 was further confirmed by X-Ray Crystallography: Crystals were formed by mixing equal molar amounts of a benzene solution of B1 and B1′ and leaving them overnight at room temperature after which small white needles formed. A crystal mounted on a diffractometer and collected data at 100 K. The intensities of the reflections were collected by means of a Bruker APEX II DUO CCD diffractometer (CuK$_\alpha$ radiation, $\lambda=1.54178$ Å), and equipped with an Oxford Cryosystems nitrogen flow apparatus. The collection method involved 1.0° scans in $\omega$ at -30°, -55°, -80°, 30°, 55°, 80° and 115° in 2$\theta$. Data integration down to 0.84 Å resolution was carried out using SAINT V8.34 C (Bruker diffractometer, 2014) with reflection spot size optimisation. Absorption corrections were made with the program SADABS (Bruker diffractometer, 2014). The structure was solved by Intrinsic Phasing methods and refined by least-squares methods again $F^2$ using SHELXT-2014 (Sheldrick, 2015) and
SHELXL-2014 (Sheldrick, 2015) with OLEX 2 interface (Dolomanov, et al., 2009). Non-hydrogen atoms were refined anisotropically, and hydrogen atoms were allowed to ride on the respective atoms. Crystal data as well as details of data collection and refinement are summarized in Table 1. The Ortep plots were produced with SHELXL-2014 program.

<table>
<thead>
<tr>
<th></th>
<th><strong>A1 macrocycle</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crystal data</strong></td>
<td></td>
</tr>
<tr>
<td>Chemical formula</td>
<td>C₃₈H₃₂NO₄</td>
</tr>
<tr>
<td>M_r</td>
<td>566.64</td>
</tr>
<tr>
<td>Crystal system, space group</td>
<td>Monoclinic, P2₁/n</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>100</td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>24.5782 (7), 4.6272 (1), 25.8074 (7)</td>
</tr>
<tr>
<td>b (°)</td>
<td>93.140 (2)</td>
</tr>
<tr>
<td>V (Å³)</td>
<td>2930.62 (13)</td>
</tr>
<tr>
<td>Z</td>
<td>4</td>
</tr>
<tr>
<td>Radiation type</td>
<td>Cu Ka</td>
</tr>
<tr>
<td>m (mm⁻³)</td>
<td>0.66</td>
</tr>
</tbody>
</table>

**Table 4.1: Experimental details**
<table>
<thead>
<tr>
<th>Crystal size (mm)</th>
<th>0.24 × 0.02 × 0.02</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection</td>
<td></td>
</tr>
<tr>
<td>Diffractometer</td>
<td>Bruker D8 goniometer with CCD area detector diffractometer</td>
</tr>
<tr>
<td>Absorption correction</td>
<td>Multi-scan</td>
</tr>
<tr>
<td></td>
<td>SADABS</td>
</tr>
<tr>
<td>$T_{\text{min}}, T_{\text{max}}$</td>
<td>0.818, 0.863</td>
</tr>
<tr>
<td>No. of measured, independent and observed $[I &gt; 2s(I)]$ reflections</td>
<td>48046, 5104, 3670</td>
</tr>
<tr>
<td>$R_{\text{int}}$</td>
<td>0.076</td>
</tr>
<tr>
<td>$(\sin q/l)_{\text{max}}$ (Å$^{-1}$)</td>
<td>0.596</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
</tr>
<tr>
<td>$R[F^2 &gt; 2s(F^2)], wR(F^2)$, $S$</td>
<td>0.058, 0.165, 1.08</td>
</tr>
<tr>
<td>No. of reflections</td>
<td>5104</td>
</tr>
<tr>
<td>No. of parameters</td>
<td>619</td>
</tr>
<tr>
<td>No. of restraints</td>
<td>522</td>
</tr>
<tr>
<td>H-atom treatment</td>
<td>H-atom parameters constrained</td>
</tr>
<tr>
<td>$D_p_{\text{max}}, D_p_{\text{min}}$ (e Å$^{-3}$)</td>
<td>0.28, -0.33</td>
</tr>
</tbody>
</table>

**Table 4.1 (continued)**

**Proof:**

Show that the set of positive integers cannot be partitioned into a finite number of arithmetic progressions with distinct steps. An arithmetic progression with step size \( d \) is the set \( \{a, a + d, a + 2d, \ldots\} \).

Proof: Suppose not, consider the sum

\[
\sum_{k=0}^{\infty} z^k
\]

where \( z \) is a complex number. Since we can partition the positive integers into arithmetic progressions \( \{a_i + kd_i\} \) we can write the above sum like this:

\[
\sum_{k=0}^{\infty} z^{a_1+kd_1} + \ldots + \sum_{k=0}^{\infty} z^{a_n+kd_n}.
\]

Each of these sums are a geometric series and equals

\[
\frac{z^a}{1 - z^d}
\]

when \( |z| < 1 \). However the original series is also a geometric series and equals

\[
\frac{1}{1 - z}
\]

when \( |z| < 1 \). So we get that

\[
\frac{1}{1 - z} = \frac{z^{a_1}}{1 - z^{d_1}} + \cdots + \frac{z^{a_n}}{1 - z^{d_n}}.
\]

This is a contradiction for a number of reasons but one that will work is that if \( z \neq 1 \) but \( z^{d_i} = 1 \) the RHS blows up and the LHS stays bounded so they can’t be equal.

If you forgot:

\[
(1 - z) \sum_{k=0}^{\infty} z^k = \sum_{k=0}^{\infty} z^k - \sum_{k=1}^{\infty} z^k = 1
\]

So

\[
\frac{1}{1 - z} = \sum_{k=0}^{\infty} z^k.
\]

I use this to compute both sides of the equation above.
Figure 4.5. Perspective views showing 50% probability displacement.
P2 was reacted with a 2 eq. excess of 3-ethylbenzaldehyde by general procedure A. CombiFlash silica column purification gave a mixture of A2 and I2. The mixture was heated to 120 °C to form the desired imine bonds as well as to evaporate the chloroform. Finally, A2 was dissolved in chloroform and recrystallized in benzene. The identity of A2 was confirmed by $^1$H NMR (500 MHz, Chloroform-d) δ 8.60 (s, 2H), 8.20 (t, $J = 1.6$ Hz, 2H), 8.19 (t, $J = 1.6$ Hz, 2H), 8.13 (t, $J = 1.7$ Hz, 2H), 7.97 (ddd, $J = 7.8$, 1.7, 1.2 Hz, 2H), 7.94 (t, $J = 1.6$ Hz, 2H), 7.68 (ddd, $J = 7.6$, 1.6, 1.2 Hz, 2H), 7.53 (dd, $J = 2.1$, 1.6 Hz, 3H), 7.51 (dd, $J = 7.8$, 7.6 Hz, 2H), 7.47 (ddd, $J = 7.6$, 1.6, 1.4 Hz, 2H), 7.43 (dd, $J = 7.6$, 7.6 Hz, 2H), 7.36 (ddd, $J = 7.6$, 2.1, 1.4 Hz, 2H), 4.57 – 4.53 (m, 4H), 3.91 –
3.87 (m, 4H), 3.76 – 3.71 (m, 4H), 3.63 – 3.59 (m, 4H), 3.42 (s, 6H); $^{13}$C NMR (125 MHz, 
CDCl$_3$) d 165.5 (2C), 159.4 (2C), 151.0 (2C), 139.3 (2C), 136.6 (2C), 134.4 (2C), 133.1 (2C), 132.3 (2C), 132.2 (2C), 131.1 (2C), 129.8 (2C), 129.6 (2C), 129.3 (2C), 128.7 (2C), 124.8 (2C), 124.3 (2C), 124.1 (2C), 123.8 (2C), 123.6 (2C), 122.4 (2C), 90.9 (2C), 90.4 (2C), 88.7 (2C), 88.1 (2C), 72.2 (2C), 70.8 (2C), 69.4 (2C), 64.7 (2C), 59.4 (2C).

4.4.3 Chemicals

All chemicals were purchased from either Sigma (St. Louis MO), Matrix Scientific (Elgin, SC) or Spectra Group Ltd. (Millbury, OH).

4.4.4 MALDI-TOF

Matrix Assisted Laser Desorption/Ionization (MALDI) Time-of-Flight (TOF) Mass Spectrometry – Five portions of a 0.5 uL solution of CHCl$_3$ saturated with 2,5-dihydroxybenzoic acid matrix (Sigma) were deposited on a MTP 384 polished steel BC target plate (Bruker), allowing to solvent to evaporate between each addition. A single 0.5 uL aliquot of analyte in CDCl$_3$ was then added on top of 2,5-dihydroxybenzoic acid film and the sample analyzed with an autoflex speed LRF MALDI-TOF mass spectrometer (Bruker).
4.5 Chapter 4 References


Chapter 5

Conclusion

5.1 Summary of findings

This dissertation has been a series of explorations in the field of polymer-based information storage and replication. The first chapter summarized the state of the field, exploring the various polymers that have been proposed as the first carriers of information in life. We concluded that there are significant challenges in demonstrating the viability of any theory in which RNA arose first due both to the continuing failure of the field to discover a satisfactorily prebiotic synthesis of RNA monomers as well as the lack of a simple, protocell-compatible procedure by which RNA can be copied non-enzymatically (Hud 2013). However, as we detail in the first chapter, these same deficiencies, along with many others, can be found in all other theories based on alternative genetic polymers as well. It is tempting, given the lack of an obvious alternative to RNA, to conclude that, despite its obstacles, RNA must have been first. Importantly though, RNA has been studied more extensively than any alternative polymer by far. We highlight the most promising alternatives to RNA and recommend that research continue in the search for novel alternatives to RNA, and in the synthesis and polymerization of RNA monomers under prebiotic conditions.
In chapter one we explored the limitations of the state of the art of nonenzymatic RNA copying and identified the slow rate of A and U copying as a key challenge (Wu 1992). In the second chapter of this thesis we achieved faster polymerization rates of A and U monomers by employing helper oligonucleotides (Deck 2011) activated at their 5’-phosphates with the same 2-methylimidazole leaving groups found appended to the polymerizing monomers. This project ended with a demonstration that an RNA template containing all four RNA monomers can be copied in a single reaction in one day using our method. Previously, long incubation times (~17 days) and multiple reaction conditions were necessary to achieve this goal, casting doubt on the plausibility of nonenzymatic RNA copying of arbitrary RNA sequences. The results from chapter two of this thesis open the door to further demonstrations that nonenzymatic polymerization may have bridged the gap between prebiotic chemistry and the RNA world.

The transition between nonenzymatic polymerization and an RNA world where ribozymes catalyze RNA polymerization raises questions about what bearing the constraints of each setting have on one another. The third chapter takes as its premise that thiolated uridine analogs have been found to react more quickly and with a higher fidelity that uridine in nonenzymatic primer extension reactions (Heuberger 2015). In this chapter we asked the further question, what consequences does the presence of these analogs have in the context of ribozyme-catalyzed RNA copying?
We first tested two thiolated uridine analogs alongside uridine in a kinetic analysis and found that they both outperformed uridine for both efficiency and fidelity. We next synthesized two related phosphoryl-transfer ribozymes in which uridine is replaced at each position by a thiolated analog and found them to retain function, albeit much diminished. We consider this function to be a suitable starting point for a molecular evolution experiment so we demonstrated that reverse transcription of these ribozymes is possible using a standard T7 reverse transcriptase enzyme. These findings demonstrate both that the transition from nonenzymatic RNA copying in the presence of thiolated uridine analogs to an RNA world could have been smooth, and that ribozymes are generally robust to certain nucleobase changes.

In the fourth chapter we develop a polymer framework that mimics DNA in schematic. In order to synthesize a molecule that is capable of storing information and being copied we took the ladder-like structure of DNA as a guide. We replaced the sugar-phosphate backbone of DNA with an alternating aryl-alkyne backbone. For the rungs of the ladder we replaced the base pairs of nucleobases with imine bonds. We chose reversible covalent bonds in place of the hydrogen bonds of DNA in order to facilitate nonenzymatic, templated polymerization. In this chapter we synthesize both possible duplex structures of length two and demonstrate a full round of replication. This is an example of an alternative genetic polymer to DNA. In theory, there is no function of DNA that cannot be achieved with this polymer. This work is a
demonstration that information storage is achievable by polymers that are very different from DNA. Since other polymers, besides the aryl-alkyne polymer described in chapter four or DNA, surely have the same capability, the search for extraterrestrial life should accordingly make no assumption about the ubiquity of DNA in living systems.

5.2 Future Directions

5.2.1 Future directions for nonenzymatic polymerization

There are two primary goals in the field of nonenzymatic polymerization. The first is to discover a system of reactions that could have copied information prior to the emergence of protocellular life. The second is to construct a system that can copy information in a growing and dividing compartment.

The technique explored in the second chapter of this thesis, the use of activated oligonucleotides as catalysts, was successful in extending a primer by each of the four letters of RNA in sequence, making progress towards a pre-cellular replicating polymer. However, it is clear that longer sequences are desirable. Unfortunately, the interactions between the various trimers used and their competition for template binding prevented the polymerization of much longer sequences from being produced. Further work in optimizing reaction
concentrations, choosing sequences that have more compatible binding
interactions and even changing the activation chemistry may permit the
copying of longer, ideally arbitrarily long, templates.

Because nonenzymatic polymerization is meant to act as a stepping-stone
between prebiotic chemistry and the RNA world, one way to demonstrate that
it is possible to achieve that transformation is to synthesize an active ribozyme
nonenzymatically. Under this premise, non-enzymatic synthesis of the
hammerhead ribozyme, which can be split into two parts (Engelhart 2013, Sun
2013) and tested for activity, is a critical future direction. Besides longer primer
extension, three other technical challenges must be overcome before the
construction of protocells is possible: template copying without a primer,
multiple rounds of copying and copying within a vesicle. Preliminary work
from our laboratory has shown that trimers are long enough to act as primers,
meaning that the helper oligonucleotides described in chapter two could also
act as primers. In order to copy a template multiple times, strand separation
must be promoted and strand reannealing must be prevented. 2′-5′ linkages in a
template are known to lower the melting temperature of an RNA duplex while
still permitting ribozyme function (Engelhart 2013). That modification,
combined with the addition of cationic peptides which have been shown to slow
reannealing while still allowing nonenzymatic polymerization (Jia et al.
forthcoming), might allow for multiple rounds of copying. Finally, if all of these modifications can be combined inside of a vesicle, which can grow and divide, it may become possible to observe the beginnings of life-like behaviours, in particular, evolution.

5.2.2 Future directions for thiolated uridine analogs

In chapter three of this thesis we specifically describe a logical extension of our study, namely a molecular evolution experiment involving a modified RNA alphabet. A ribozyme could be evolved, either from random sequence or starting from an already functional sequence, with $s^2$T in place of U. We predict that such an evolution experiment may yield ribozymes with superior characteristics compared to current ribozyme polymerases (Attwater 2013).

5.2.3 Future directions for novel, replicable, informational polymers

Though the work described in chapter four is preliminary, we have demonstrated that the copying of dimers is possible using our novel polymer. In the future, an exploration of longer polymers is of great interest to us. Before we can continue to explore the chemical space of genetic polymers we must first be satisfied that the polymer we have constructed can be copied at longer lengths.

In exploring chemical space, we may be interested in devising alternative base
pairs and backbones (as described in chapter four). In addition, more than one set of base pairs could be used at once. This would be a significant technical challenge because the base pairs would have to be strictly orthogonal. However, the bigger the monomer alphabet, the more information can be stored in a polymer of a given length.

In order to demonstrate the utility of our new polymer we will have to show that it can carry readable information. To this end, a method of sequencing said polymer is necessary. Mass spectrometry is one method by which we could sequence short polymers, but longer polymers will require more creative means. In this endeavor, the analogy to DNA sequencing is particularly illustrative. Replication in the presence of chain terminating linkers, like in Sanger sequencing, could produce the fragments necessary for sequencing. Alternatively, if a new strand can be formed one base at a time rather than all at once, fluorescent monomers could be used to observe the identity of monomers in sequential additions. Our polymer could also be derivatized to allow some form of nanopore sequencing. Without knowing the chemical structure of the polymer it is impossible to speculate as to which approach will be the most effective.

Perhaps the most exciting avenue of further research is the question of whether or not it will be possible to construct polymers that fold into a specific conformation that are also catalytic. For the reasons laid out in the first chapter of this thesis, the construction of such a polymer would imply the possibility of a connection between
sequence and function that underpins the beginning of life on earth. Such a polymer may be capable of evolution, and, if embedded in a compartment, could begin to function in much the same way we imagine RNA protocells might. Additional base pairs would be particularly useful for this goal, although functional RNAs have been discovered with only two letters (Reader 2002).

Moreover, were there not a minimum, the smallest bodies would have infinites. Since then a half-of-half could still be halved, with limitless division less and less. Then what the difference ’twixt the sum and least? None: for however infinite the sum, yet even the smallest would consist the same of infinite parts. But since true reason here protests, denying that the mind can think it, convinced thou must confess such things there are as have no parts, the minimums of nature.

Continuing the analogy with biological life, we would like to explore the possibility of translating from our polymer to another one, with a different backbone and monomer alphabet. This goal has been achieved with DNA (Meng 2016) and other, non-replicable synthetic polymers (Lewandowski 2013). For the polymer from chapter four, synthetic adapter “tRNAs” could be constructed that have three aniline or aldehyde functional groups, which would bind to a template “mRNA” strand of our polymer. At the other end of the adapter molecules would be different functional groups and some other polymerization chemistry that would permit the polymerization of another sequence defined polymer that had some desired function.
In this case, as with proteins, a capacity for self-replication would not be required.

Finally, this project raises the question of whether or not a linear polymer is the only, or even the simplest, way in which information can be stored by molecules. Many theories in the origin of life field involve assemblies of molecules held together by non-covalent interactions that can grow and divide and, in that manner, transmit information (Markovitch 2014). However, the construction of such a system has not yet been demonstrated. Besides linear polymers and statistical assemblages, are there any additional molecular information storage strategies?

5.3 Prospective

The three central chapters of this thesis may seem to be only loosely connected. Nonenzymatic RNA polymerization is very different from ribozyme mediated RNA polymerization and even further removed from the nonenzymatic polymerization of a synthetic polymer. However, the deeper theme that ties them together is that all three projects are an exploration of the chemical space of genetic polymers. All three attempt to answer the question, “what are the necessary features of a polymer that allow it to be the basis for life?” at different levels of abstraction. Non-enzymatic polymerization of RNA is an unrealized goal whose fulfillment may be a crucial confirmation of the whole RNA world hypothesis. The exploration of thiolated analogs of uridine is an attempt to explore how beneficial slight nucleobase
modifications can be for RNA copying in an RNA world and what conclusions we can draw about their necessity in early life. The exploration of synthetic alternatives to RNA is an attempt to demonstrate that life is not restricted to the canonical chemistry of contemporary biology. Consequently, life may be much more abundant and various in its forms than we naively imagine. And if a broader milieu of life is theoretically possible it makes the search for life beyond our earth all the more challenging and tantalizing. The requirements for life are abstract; DNA-based life is the only kind life we have ever encountered but there is no reason to believe that it is the only possible form (Livneh 1986).

5.4 Chapter 5 References


