VISCOUS SOLVENTS AS AN ENVIRONMENT FOR NUCLEIC ACID REPLICATION

A Dissertation
Presented to
The Academic Faculty

By

Christine He

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in Chemical & Biomolecular Engineering

Georgia Institute of Technology
May 2017

VISCOUS SOLVENTS AS AN ENVIRONMENT FOR NUCLEIC ACID REPLICATION

Approved by:

Dr. Martha Grover, Advisor
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

Dr. Nicholas Hud, Co-advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Loren Williams
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Mark Styczynski
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

Dr. Clifford Henderson
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

Date approved: 08 February, 2017

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor Martha Grover for her research advice, thoughtful guidance, and warm encouragement during my time in graduate school. She has given me room to grow as an independent researcher and pursue what interests me, a privilege for which I am incredibly grateful. I am also thankful for the excellent example she has provided as a successful woman engineer and professor.

I would like to express my deepest gratitude to Nicholas Hud for welcoming me into his lab. My development as a researcher has been influenced enormously by his scientific insightfulness, enthusiasm, integrity, and devotion to the field of prebiotic chemistry. Thank you to Nick for your mentorship, and for all that you have taught me.

I thank my thesis committee members, Dr. Loren Williams, Dr. Mark Styczynski, and Dr. Clifford Henderson, for their suggestions and feedback on my research.

This work would not have been possible without funding from the NSF, NASA, and the James S. McDonnell Foundation. In particular, I have enjoyed and benefited greatly from being part of the collaborative Center for Chemical Evolution.

Thank you to all my friends at Georgia Tech for making my time in Atlanta enjoyable, both inside and outside the lab. In particular, I would like to thank Isaac Gállego for teaching me how to work in a biochemistry lab, contributing to much of the research presented here, and being a constant source of optimism and support. Thank you to Adriana Lozoya-Colinas and Gary Newnam for productive discussions and assistance with research, as well as all the members of the Grover and Hud groups—past and present—who have helped me over the years.

I would like to thank my parents and brother for their love and support. Finally, I would like to thank my husband, Jeff Camp, who has been my study partner and best friend since we were both nineteen years old. This thesis is dedicated to Jeff.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

NOMENCLATURE

bp	Base pairs
DNA	Deoxyribonucleic acid
ds	Double stranded
EtBr	Ethidium bromide
FAM	Carboxyfluorescein
kb	Kilobase
nt	Nucleotide(s)
PAGE	Polyacrylamide gel electrophoresis
RNA	Ribonucleic acid
ss	Single stranded
TAE	Tris/acetate/EDTA
TBE	Tris/borate/EDTA
Tris	Tris(hydroxymethyl)aminomethane

SUMMARY

Many hypotheses concerning the nature of early life assume that genetic information was once transferred through the template-directed synthesis of RNA, prior to the evolution of genetically encoded protein synthesis. However, despite more than half a century of research into the chemical origins of nucleic acids, a robust route to the abiotic synthesis of nucleic acid polymers is unclear. In particular, identifying the earliest mechanism for the enzyme-free replication of nucleic acids remains an elusive goal. A fundamental biophysical problem known as strand inhibition limits copying of a nucleic acid duplex: transferring information from a template sequence in the presence of its complementary strand is inhibited by the stability of the template duplex. Strand inhibition is a major bottleneck in understanding how an RNA system capable of sustained replication without the aid of enzymes evolved on the early Earth.

In this thesis, I describe a robust, prebiotically plausible route to enzyme-free replication of nucleic acids, driven by hot/cool cycles in viscous environments. Viscous solvents enable kinetic trapping of a nucleic acid duplex as single strands, providing a time window for the assembly and ligation of oligonucleotide substrates on the single stranded templates. I have shown that viscous solvents can be utilized to overcome the problem of strand inhibition, enabling copying of a gene-length template duplex (>300 nt), a process which is highly unfavorable in aqueous conditions. Additionally, viscosity enables copying of an RNA duplex template containing a hammerhead ribozyme motif, suggesting a potential route for the selection and amplification of catalytically active RNA (ribozymes) on the prebiotic Earth.

CHAPTER 1 : INTRODUCTION

1.1. Overview

Nucleic acids are the basis for storing and transferring genetic information in all living organisms. Many models of early evolution assume that the first informational polymer was a nucleic acid, based on the ability of nucleic acids to robustly transfer sequence information through Watson-Crick base pairing. It was discovered in the 1980's that RNA, in addition to its role as an informational polymer, possesses catalytic properties and carries out a range of key biochemical reactions[1], including the linking of amino acids within the ribosome[2]. These discoveries helped generate interest in an RNA World hypothesis, which proposes that RNA was the key biopolymer in early evolution, serving as both genetic information carrier and catalyst prior to the adoption of these roles by DNA and proteins, respectively[3]. Motivated by an RNA World scenario, researchers have utilized *in vitro* evolution to generate RNA sequences that function as RNA-templated RNA polymerases[4, 5], providing support for the idea that RNA might have once catalyzed its own replication prior to the evolution of genetically encoded protein synthesis. These advances—along with ongoing research which indicates that RNA's cellular role is more expansive than previously thought[6, 7]—suggest that the template-directed synthesis of RNA was a critical process in the transition from prebiotic chemistry to life.

However, despite more than half a century of research into the chemical origins of nucleic acids, a robust pathway from simple chemical precursors to self-replicating nucleic acid polymers has not been identified[8-10]. Open questions remain about every

stage in the abiotic formation of nucleic acids, including synthesis of mononucleotides[11], non-templated polymerization of mononucleotides into oligonucleotides[12], and information transfer from the resulting nucleic acid templates[13]. In particular, identifying a robust, prebiotically plausible mechanism for template-directed nucleic acid synthesis remains an elusive goal. In contemporary biology, DNA replication is a complex process, orchestrated by specialized enzymes (such as polymerases and helicases) working in a concerted, step-wise fashion[14, 15]. In contrast, replication of an early informational polymer must have occurred by a simpler, enzyme-free mechanism that may have borne little resemblance to the process of DNA replication in life today.

Much of the research focused on identifying a prebiotic route for nucleic acid replication has employed a mechanism resembling that utilized by biological polymerases, where a short oligonucleotide sequence—a primer—is annealed to a template strand and then extended at its 3' end by sequential addition of mononucleotides. To drive phosphate ester bond formation without the aid of enzymes, researchers "activate" the 5' phosphate of the mononucleotides with nitrogenous heterocycles—such as 2-methylimidazolide or oxyazabenzotriazole—that serve as leaving groups during phosphodiester bond formation. However, several problems limit this model of template-directed copying, including weak association of mononucleotides with the primer-template complex and template binding by inactivated mononucleotides (resulting from hydrolysis of the activated phosphate group)[16]. In addition, the primer extension reaction itself proceeds at slow rates, in low yields, and in a strongly sequence-dependent manner. In order to extend a primer with all four nucleobases, downstream "helper"

oligonucleotides are utilized to provide an additional stacking surface that stabilizes the association of the incoming mononucleotide with the primer-template complex[17].

Because of the limitations associated with the use of mononucleotides for template-directed copying, I have chosen to utilize oligonucleotides as substrates, which exhibit stronger template binding than mononucleotides. While the non-templated polymerization of mononucleotides into oligonucleotides remains an area of active research, if we assume that nucleic acid polymers long enough to potentially form a viable genetic system (>20-30 nt) were accessible as templates[12], then shorter oligonucleotides would have also been present in the prebiotic mixture at this stage of chemical evolution. An advantage of an oligonucleotide-based copying process is that it would not be limited to proceeding sequentially across the template in a 5' to 3' manner, as primer extension by mononucleotide addition proceeds during copying by a processive polymerase. Instead, copying with oligonucleotide substrates may proceed by simultaneous assembly of oligonucleotides at multiple locations on a template strand, followed by ligation of the assembled oligonucleotides and/or “gap-filling” by mononucleotides to complete synthesis of a complementary strand[13].

Another problem arises after the synthesis of a complementary sequence from a template strand has proceeded. The result of copying a nucleic acid template strand (or, rather, producing the complementary strand) is a duplex that must be separated into single strands before another round of template-directed synthesis can occur. In the absence of enzymes, elevated temperatures can drive the separation of a duplex into single strands. However, for copying to proceed, the temperature must then be lowered to the point at which mono- or oligonucleotide substrates can stably bind to their

complementary sites on the template strands. Reformation of the template duplex is kinetically and thermodynamically favored over template-substrate binding, so that once the temperature is lowered, the template duplex reforms and substrate binding is inhibited. This longstanding problem in enzyme-free, template-directed nucleic acid synthesis is known as *strand inhibition*[18, 19]. Few strategies have been developed to explicitly address the strand inhibition problem and existing approaches are far from robust or self-sustaining, requiring manual washing steps to separate copy strands from template strands[17, 20] or introduction of synthetically difficult chemical modifications to specific sites on the template or substrates[21-23]. These limitations motivate the continued search for more general, non-enzymatic processes that could have facilitated copying from long template strands of arbitrary sequence before the emergence of polymerase enzymes (or ribozymes)[24-26].

In identifying a mechanism for enzyme-free nucleic acid replication, it is important to consider the physical environment and solvent conditions in which such a process might have transpired. This thesis focuses on viscosity-dependent mobility as a mechanism for facilitating the template-directed synthesis of nucleic acids. Viscous environments could have been generated on the early Earth during periodic cycles of hot/cold, dry/wet conditions, driven by day/night or seasonal transitions. During periods of high temperature, evaporation of water from pools on dry land would concentrate organic compounds or salts into a viscous solution. Rehydration—through water uptake from a humid environment, rain, or tides—would bring the pool back to a hydrated, low viscosity state. Simulated wet/cool and hot/dry cycles have been utilized to promote the synthesis of polypeptides from amino acid and hydroxy acid building blocks (through a

depsipeptide intermediate), overcoming the difficulty of forming an amide bond—a dehydration-condensation reaction—in water[27]. We propose that similar fluctuations water activity, temperature, and viscosity could have facilitated copying of a nucleic acid duplex in prebiotic conditions.

In a viscous environment, diffusion of nucleic acid polymers should be slowed in a size-dependent manner. A highly viscous solvent can significantly retard the movement of long template strands, slowing the reformation of the template duplex. Additionally, intramolecular secondary structure (e.g., hairpins) on the template strands form quickly, serving to further kinetically trap the template as single strands. Meanwhile, mononucleotides and short oligonucleotides remain relatively mobile. We hypothesized that the difference in mobility between template strands and oligonucleotide substrates in a viscous environment would create a time window for assembly of oligonucleotide substrates on their complementary template strands, overcoming the problem of strand inhibition and enabling template-directed nucleic acid synthesis.

In addition to providing a mechanism to circumvent the strand inhibition problem, viscosity provides a potential mechanism for the selection and amplification of catalytically active RNA sequences (ribozymes) from a prebiotic pool of diverse sequences—a key process in the emergence of a putative RNA World. In contrast to aqueous environments, viscous environments may promote the replication of longer, highly structured nucleic acid sequences—which are less mobile and more readily kinetically trapped—over shorter, unstructured sequences. Concerns over how long template sequences can “outcompete” short sequences have existed since the 1960s, when Spiegelman demonstrated that short sequences are replicated more quickly than

longer sequences in an extracellular replication system, resulting in shortening of template sequences and loss of genetic information over time[25, 28]. Viscosity has the potential to reverse this trend, promoting the kinetic trapping and replication of long template sequences. Viscous environments also present a potential solution to the "replicator-catalyst" paradox of the RNA world, which arises from the fact that sequences with well folded intramolecular structures—a requirement for catalytic activity—are less accessible as templates for replication[29, 30]. In a viscous environment, the formation of a stable intramolecular structure on a template sequence is not necessarily detrimental, but might be highly beneficial for replication, preventing the re-annealing of complementary template strands into a duplex. Therefore, viscous environments have the potential to provide a novel route to enzyme-free nucleic acid replication that may explain the emergence of desirable template features that are not favored in aqueous environments.

In this thesis, I describe the implementation of a model prebiotic nucleic acid replication cycle in viscous environments that can be utilized to copy gene-length, mixed sequence templates. In Chapter 2, I present a proof-of-concept demonstration that viscous solvents can be used to overcome the problem of strand inhibition. Using a model DNA system, I demonstrate information transfer from a gene-length (>300 nt) region of a larger (3 kb) template duplex in glycholine, a viscous eutectic solvent composed of glycerol and choline chloride. In Chapter 3, I extend these results by demonstrating information transfer from an RNA template duplex whose length (>500 bp) is on the same order of length as the minimal size of an RNA-directed RNA polymerase[31]. Additionally, I probe the limits of viscosity-mediated information transfer by exploring

the effects of template length, substrate length, substrate purity, and solvent on the efficiency of the oligonucleotide assembly and ligation process. Finally, in Chapter 4, I describe the use of viscous solvents to promote a true replication cycle—i.e. simultaneous information transfer from both strands—of an RNA duplex containing a ribozyme sequence, and demonstrate the activity of the ribozyme with the newly synthesized strands.

1.2. References

1. Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR. Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* 1982, **31**(1)**:** 147-157.

2. Cech TR. The Ribosome Is a Ribozyme. *Science* 2000, **289**(5481)**:** 878-879.

3. Gilbert W. ORIGIN OF LIFE - THE RNA WORLD. *Nature* 1986, **319**(6055)**:** 618-618.

4. Ekland EH, Bartel DP. RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* 1996, **382**(6589)**:** 373-376.

5. Sczepanski JT, Joyce GF. A cross-chiral RNA polymerase ribozyme. *Nature* 2014, **515**(7527)**:** 440-442.

6. Storz G. An Expanding Universe of Noncoding RNAs. *Science* 2002, **296**(5571)**:** 1260-1263.

7. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. *Nat Rev Genet* 2009, **10**(2)**:** 94-108.

8. Oro J. Mechanism of Synthesis of Adenine from Hydrogen Cyanide under Possible Primitive Earth Conditions. *Nature* 1961, **191**(4794)**:** 1193-1194.

9. Joyce GF. The antiquity of RNA-based evolution. *Nature* 2002, **418**(6894)**:** 214-221.

10. Cafferty BJ, Hud NV. Abiotic synthesis of RNA in water: a common goal of prebiotic chemistry and bottom-up synthetic biology. *Current Opinion in Chemical Biology* 2014, **22:** 146-157.

11. Hud Nicholas V, Cafferty Brian J, Krishnamurthy R, Williams Loren D. The Origin of RNA and “My Grandfather’s Axe”. *Chemistry & Biology*, **20**(4)**:** 466-474.

12. Ferris JP, Jr. ARH, Liu R, Orgel LE. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* 1996, **381:** 59-61.

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

14. Bell SP, Dutta A. DNA replication in eukaryotic cells. *Annu Rev Biochem* 2002, **71:** 333-374.

15. Kunkel TA, Bebenek R. DNA replication fidelity. *Annu Rev Biochem* 2000, **69:** 497-529.

16. Kervio E, Sosson M, Richert C. The effect of leaving groups on binding and reactivity in enzyme-free copying of DNA and RNA. *Nucleic Acids Research* 2016.

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

18. Grossmann TN, Strohbach A, Seitz O. Achieving Turnover in DNA-Templated Reactions. *ChemBioChem* 2008, **9**(14)**:** 2185-2192.

19. Fernando C, Kiedrowski Gv, Szathmary E. A Stochastic Model of Nonenzymatic Nucleic Acid Replication: ‘‘Elongators’’ Sequester Replicators. *Journal of Molecular Evolution* 2007, **64:** 572-585.

20. Luther A, Brandsch R, Kiedrowski Gv. Surface-promoted replication and exponential amplification of DNA analogues. *Nature* 1998, **396:** 245-248.

21. Dose C, Ficht S, Seitz O. Reducing Product Inhibition in DNA-Template-Controlled Ligation Reactions. *Angewandte Chemie International Edition* 2006, **45:** 5369-5373.

22. Zhan Z-YJ, Lynn DG. Chemical Amplification through Template-Directed Synthesis. *Journal of the American Chemical Society* 1997, **119:** 12420-12421.

23. Kausar A, McKay RD, Lam J, Bhogal RS, Tang AY, Gibbs-Davis JM. Tuning DNA Stability To Achieve Turnover in Template for an Enzymatic Ligation Reaction. *Angewandte Chemie International Edition* 2011, **50:** 8922-8926.

24. Mutschler H, Wochner A, Holliger P. Freeze–thaw cycles as drivers of complex ribozyme assembly. *Nat Chem* 2015, **7**(6)**:** 502-508.

25. Kreysing M, Keil L, Lanzmich S, Braun D. Heat flux across an open pore enables the continuous replication and selection of oligonucleotides towards increasing length. *Nat Chem* 2015, **7**(3)**:** 203-208.

26. Walker SI, Grover MA, Hud NV. Universal Sequence Replication, Reversible Polymerization and Early Functional Biopolymers: A Model for the Initiation of Prebiotic Sequence Evolution. *PLoS ONE* 2012, **7**(4)**:** e34166.

27. Forsythe JG, Yu S-S, Mamajanov I, Grover MA, Krishnamurthy R, Fernández FM, *et al.* Ester-Mediated Amide Bond Formation Driven by Wet–Dry Cycles: A Possible Path to Polypeptides on the Prebiotic Earth. *Angewandte Chemie International Edition* 2015, **54**(34)**:** 9871-9875.

28. Mills DR, Peterson RL, Spiegelman S. An Extracellular Darwinian Experiment with a Self-duplicating Nucleic Acid Molecule. *Proceedings of the National Academy of Sciences* 1967, **58**(1)**:** 217-224.

29. Ivica NA, Obermayer B, Campbell GW, Rajamani S, Gerland U, Chen IA. The Paradox of Dual Roles in the RNA World: Resolving the Conflict Between Stable Folding and Templating Ability. *Journal of Molecular Evolution* 2013, **77**(3)**:** 55-63.

30. Attwater J, Wochner A, Holliger P. In-ice evolution of RNA polymerase ribozyme activity. *Nat Chem* 2013, **5**(12)**:** 1011-1018.

31. Wochner A, Attwater J, Coulson A, Holliger P. Ribozyme-Catalyzed Transcription of an Active Ribozyme. *Science* 2011, **332**(6026)**:** 209-212.

CHAPTER 2 : BUILDING A MODEL VISCOSITY-MEDIATED REPLICATION CYCLE WITH DNA[a]

2.1. Introduction

DNA replication is the molecular basis for genetic information transfer in all living organisms. However, despite the importance of DNA replication in contemporary life, its chemical origins are poorly understood. A "chicken and egg" problem arises from the fact that DNA replication is carried out by protein enzymes, but enzymes themselves are the end product of reading the information encoded in DNA. The RNA World hypothesis and other models of chemical evolution assume that early on in evolutionary history—prior to the emergence of coded proteins—nucleic acid polymers catalyzed the synthesis of their complementary sequences without the aid of protein enzymes. Despite continued enthusiasm for the RNA world hypothesis, a robust, prebiotically plausible mechanism for the template-directed synthesis of nucleic acids has yet to be identified.

A major obstacle to demonstrating multiple rounds of non-enzymatic nucleic acid replication is a long-standing problem known as strand inhibition (or product inhibition)[1, 2]. In extant life, enzymes perform the critical function of separating a nucleic acid duplex into single strands, enabling the base pairing of complementary mono- or oligonucleotides onto the exposed templates. It is generally accepted that, in prebiotic conditions, elevated temperatures could have provided a mechanism for denaturing a duplex into single strands. However, during sample cooling the duplex annealing rate is so rapid in water—relative to the rate of binding by complementary mono- or

[a] This chapter has been adapted from: He C*, Gállego I*, Laughlin B, Grover M, Hud NV. "A viscous solvent enables information transfer from gene-length nucleic acids in a model prebiotic replication cycle." *Nature Chemistry* 2016, In Press. It is reproduced with permission. © Nature Publishing Group.

oligonucleotides—that the duplex reforms before template-directed synthesis can proceed[3]. Therefore, transfer of sequence information from a nucleic acid template strand is inhibited by the presence of its complementary template strand. The thermodynamic and kinetic favorability of duplex formation increases with the number of base pairs[4], making strand inhibition a major bottleneck for the replication of a genetic polymer of any appreciable length.

Studies exploring prebiotic routes to nucleic acid replication have long been limited by strand inhibition. Though researchers have utilized *in vitro* evolution techniques to generate RNA-directed RNA polymerases[5, 6] capable of synthesizing copy strands from single-stranded RNA templates[7], the product of these polymerase reactions is a "dead end" template-product duplex that cannot undergo further rounds of template-directed synthesis due to strand inhibition. Attempts to circumvent the strand inhibition problem have included the immobilization of template strands on a solid surface to allow manual washing of the copy strands[8, 9] and the introduction of chemical modifications at specific template or substrate sites[10-12]—strategies which have not been applied to longer nucleic acid polymers (>30 nucleotides) of arbitrary sequence. These limitations motivate the continued search for non-enzymatic processes that could have originally facilitated nucleic acid replication[13, 14]—especially ones that could replicate templates of arbitrary sequence and length[15].

Here, we present a new approach to template-directed nucleic acid synthesis that can be applied to gene-length, double stranded nucleic acid templates. This approach utilizes solvent viscosity to slow duplex formation in a length-dependent manner[16], allowing oligonucleotides to assemble on a longer template before the template duplex

reforms. For a proof-of-concept system, we selected DNA as a model informational polymer, due to its greater ease of preparation and greater chemical stability compared to RNA. For a model viscous environment, the eutectic solvent glycholine (composed of glycerol and choline chloride in a 4:1 molar ratio) was chosen due to its ability to preserve nucleic acid secondary structure[17] and promote the folding of DNA origami, a complex process where hundreds of oligonucleotides assemble on a single stranded template[18]. We apply our solvent-mediated replication process to generate a complementary sequence from a 352 nucleotide (nt) region of a 3 kilobase pair (kb) DNA template. The 3 kb DNA template is of bacterial origin, and was chosen as a model of a mixed, arbitrary sequence (i.e. one that was not engineered to facilitate replication).

2.2. A Model Replication Cycle

The proposed process for viscosity-enabled, polymerase-free information transfer is divided into Steps 1-5 of a replication cycle, with Step 6 representing the return of the system to its original state (Figure 2.1). Our process was built step by step, and I detail the rationale behind and key results for each step.

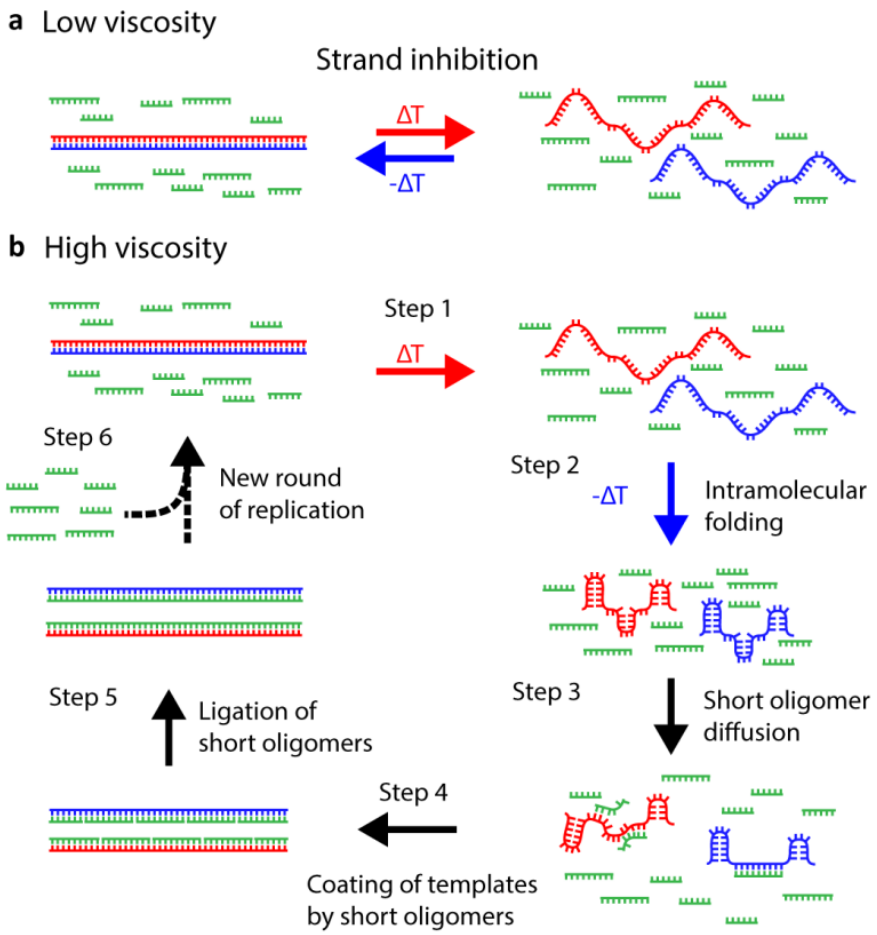

Figure 2.1. Schematic representation of the strand inhibition problem and a possible solution using a viscous solvent. **a**, Illustration of the strand inhibition problem for long complementary template strands in water. **b**, Proposed process for duplex replication in a viscous solvent. *Step 1*) Heating causes thermal denaturation of the template duplex and reduced viscosity of the solvent, allowing the template single strands to diffuse apart. *Step 2*) Upon cooling, solvent viscosity increases, kinetically trapping internal hairpins on the template single strands. *Step 3*) Oligonucleotide substrates, which are more mobile than the much longer templates, diffuse and bind to complementary sites on the template strands. *Step 4*) Eventually, the oligonucleotides completely coat and unfold the trapped template strands. *Step 5*) Ligation (chemical or enzymatic) of the bound substrate oligonucleotides completes the process of duplex replication. *Step 6*) Another round of viscosity-mediated replication can begin.

2.3. Step 1: Choosing glycholine as a solvent for viscosity-mediated information transfer

Testing the proposed process required a viscous solvent in which nucleic acid duplexes are stable at room temperature, but denature at low enough temperatures that the nucleic acids are not irreversibly altered (e.g., depurination) when thermally denatured (Figure 2.1, Step 1). We chose to focus on eutectic solvents, a class of solvents closely related to ionic liquids[19, 20]. These solvents are mixtures of two or more components whose melting point is significantly lower than that of either individual component. Previous studies performed by the Hud lab have shown that DNA and RNA form the full range of potential secondary structures—duplexes, triplexes, and quadruplexes—in anhydrous and hydrated eutectic solvents, though the thermodynamic stability and exact secondary structure of nucleic acid sequences are often altered compared to aqueous buffer[17, 21]. Eutectic solvents are miscible with water, hygroscopic, stable upon heating, and often composed of small molecules, making them attractive as potential solvents for day/night, hydration/evaporation cycles on the prebiotic Earth. Simulated day/night cycles in eutectic solvents have been shown to promote the formation of organophosphates [22, 23], overcoming the prebiotic "phosphorylation problem" present in aqueous conditions.

Based on its ability to promote folding of DNA origami[18], we chose to use glycholine, a solvent composed of a 4:1 molar ratio of glycerol and choline chloride that has a room temperature viscosity of 437 cP. The melting temperatures (T_m's) of three distinct DNA structures in anhydrous glycholine were determined to be 45 °C for a 17 nt DNA hairpin (7 bp stem), 49 °C for a 32 bp duplex, and 51 °C for a linearized 3 kb (2961 nt single strands) bacterial plasmid (sequences provided in Section 2.9.2. DNA

sequences). These T_m values are significantly lower and closer to each other than those measured in aqueous buffer (

Table 2.1). Because our aim is to develop a general approach to replicate arbitrary sequences, the 3 kb bacterial plasmid was chosen as a model for a mixed, heterogeneous sequence, i.e. a sequence that has not been specifically designed for the purpose of overcoming strand inhibition. The 32 bp duplex was designed specifically to minimize the formation of intramolecular hairpins on its single strands[17], allowing us to evaluate rates of intermolecular vs. intramolecular duplex formation by comparison with the 17 nt hairpin.

Table 2.1. Melting temperatures of DNA duplex species in glycholine and in aqueous buffer (20 mM Tris pH 7.5, 0.1 M NaCl).

Species	Base Pairs	%GC	T_m in Aqueous Buffer (°C)	T_m in Glycholine (°C)
17 nt hairpin	7 bp stem, 3 nt loop	41	79.5	44.7
32 bp duplex	32	47	72.2	49.0
3 kb duplex	2957	50	88.0	50.7

2.4. Step 2: Viscosity enables kinetic trapping of single stranded templates

The proposed cycle includes the intramolecular folding of single stranded templates (Figure 2.1, Step 2), which should be increasingly favored over duplex reformation as solvent viscosity increases. Figure 2.2 shows the kinetics of secondary structure reformation after heat cycling (thermal denaturation and cooling to 20 °C) for the 17 nt hairpin, 32 bp duplex, and 3 kb duplex DNA in aqueous buffer and in glycholine.

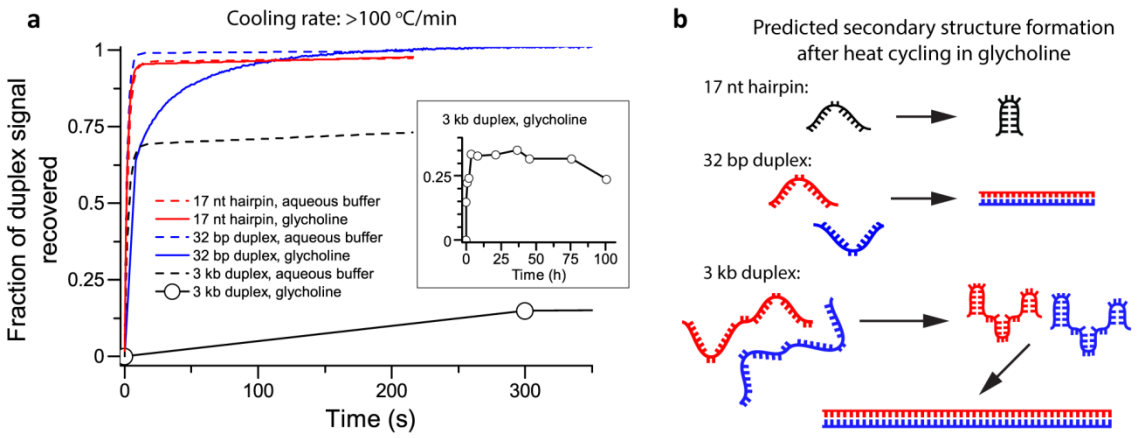

Figure 2.2. DNA duplex formation after heat cycling for a 17 nt hairpin, 32 bp duplex, and 3 kb duplex. **a,** Kinetics of duplex formation after thermal denaturation and cooling, as measured by UV-Vis absorbance and agarose gel electrophoresis. **b,** Predicted secondary structures of the 17 nt hairpin, 32 bp duplex, and 3 kb duplex.

The rate of hairpin reformation is not detectably slowed in viscous glycholine compared to reformation in aqueous buffer, reaching completion in less than 9 seconds in both solvents (Figure 2.2a). In contrast, the time required for reformation of the 32 bp duplex is increased by at least a factor of 15 in glycholine relative to aqueous buffer. Viscosity slows the kinetics of 32 bp duplex reformation because this process is bimolecular, involving the diffusion of single strands through the solvent[24]. In contrast, hairpin formation is unimolecular. The complementary sequences are located in relatively close proximity on the same strand, and duplex formation requires local rearrangements[25], which are potentially less sensitive to changes in viscosity.

The effect of glycholine on the reformation rate of the 3 kb duplex is striking, with only ~25% of the full length duplex reformed after 100 hours, compared to ~70% duplex reformation after 20 seconds in aqueous buffer (Figure 2.2a). The slow reformation of the 3 kb duplex in glycholine may be explained in part by the rapid

formation of intramolecular structures (e.g., hairpins) compared to intermolecular duplexes in glycholine. It is well known that genome-length, single-stranded DNA (ssDNA) generated via rapid cooling (e.g., >48 °C/min) from a thermally denatured state contains substantial intramolecular structure[26]. Thus, reformation of the 3 kb DNA duplex in glycholine is likely delayed both by slowed diffusion of the ssDNA through the viscous solvent and by the persistence of intramolecular structures, the latter presenting an additional kinetic barrier to duplex formation[26-28]. Figure 2.2b shows a schematic of predicted secondary structure formation after heat cycling for the 17 nt hairpin, 32 nt duplex, and 3 kb duplex DNA in glycholine.

The presence of intramolecular structure can be more directly observed with agarose gel electrophoresis, which physically separates different populations of nucleic acids with different hybridization states or secondary structures. After heating 3 kb DNA to denaturing temperatures and cooling to 20 °C, analysis by agarose gel electrophoresis shows that DNA separates into double stranded (ds) and single stranded (ss) states (Figure 2.3). The 3 kb single stranded state is referred to as 2961 nt ssDNA. Samples heat cycled in aqueous buffer and in glycholine exhibit comparable electrophoretic mobility of the 2961 nt ssDNA, suggesting the generation of ssDNA with a similar degree of intramolecular structure in these two solvents. As further evidence that the kinetically trapped ssDNA is forming a significant degree of intramolecular structure, the ssDNA band in Figure 2.3a is visible after staining by ethidium bromide (EtBr), a small molecule that intercalates between base pairs in double stranded DNA.

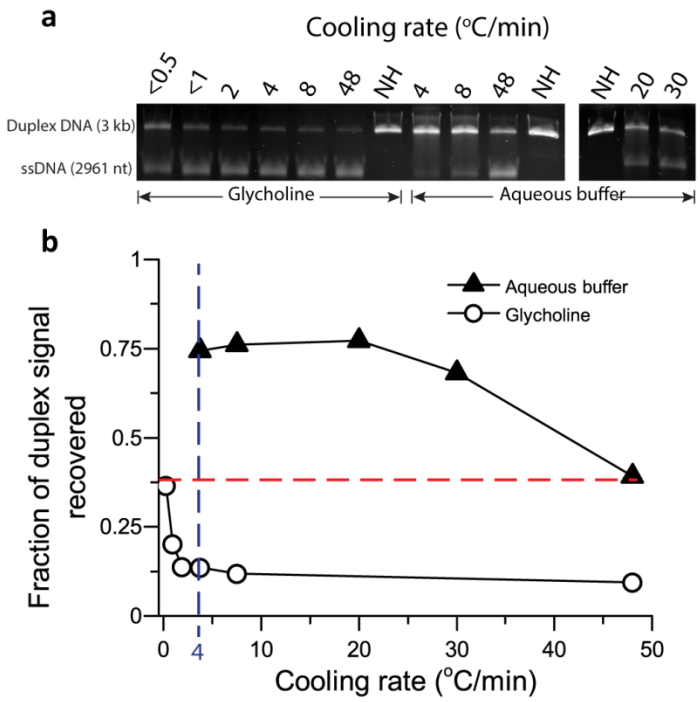

Figure 2.3. Return of 3 kb DNA to a duplex state after heat cycling. **a**, Agarose gel (2%) stained with ethidium bromide showing the separation of 3 kb DNA—after heat cycling—into duplex (less mobile band) and single stranded (more mobile band) states. **b**, The recovery of the duplex state for the 3 kb DNA as a function of different cooling rates, as measured by densitometry analysis of the duplex band intensity on the agarose gel.

Consistent with the contribution of intramolecular secondary structures to slowing intermolecular duplex formation, the rate of duplex reformation can be modulated by cooling rate. For example, in order to kinetically trap 63% of the 3 kb DNA as single strands in aqueous buffer, we must utilize a rapid cooling rate (48 °C/min) that is two orders of magnitude faster than that required in glycholine (<0.5 °C/min) (Figure 2.3, dashed red line). For all subsequent experiments, a cooling rate of 4 °C/min was employed, as it is the most gradual cooling rate that maximizes the difference in kinetic trapping of the single stranded (ss) 3 kb DNA between glycholine and aqueous buffer (Figure 2.3, dashed blue line).

2.5. Step 3: Oligonucleotide binding to kinetically trapped templates

As a next step, I show that an oligonucleotide substrate can bind to its complementary sequence on kinetically trapped single stranded templates (Figure 2.1, Step 3). We list this step as separate from complete coating of the target sequence because it is likely that binding will first occur by oligonucleotides to single stranded regions of the template strand, which are not blocked by hairpins or other intramolecular structures. Using the 3 kb DNA duplex as a template, we designed a 32 nt oligonucleotide, named *F0,* which binds to a site on the antisense template strand. The sequence of *F0* was chosen so that its template binding site is expected to be relatively accessible, i.e. less likely to be involved in an intramolecular structure (

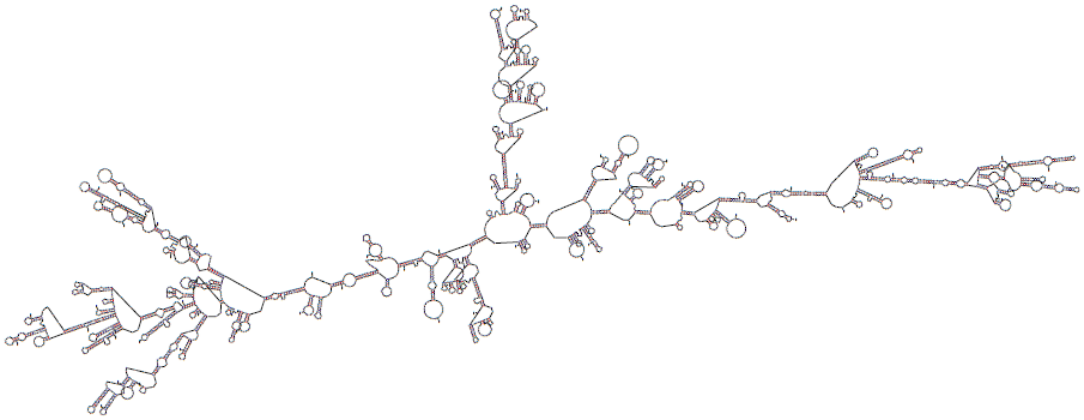

Figure **2.4**)

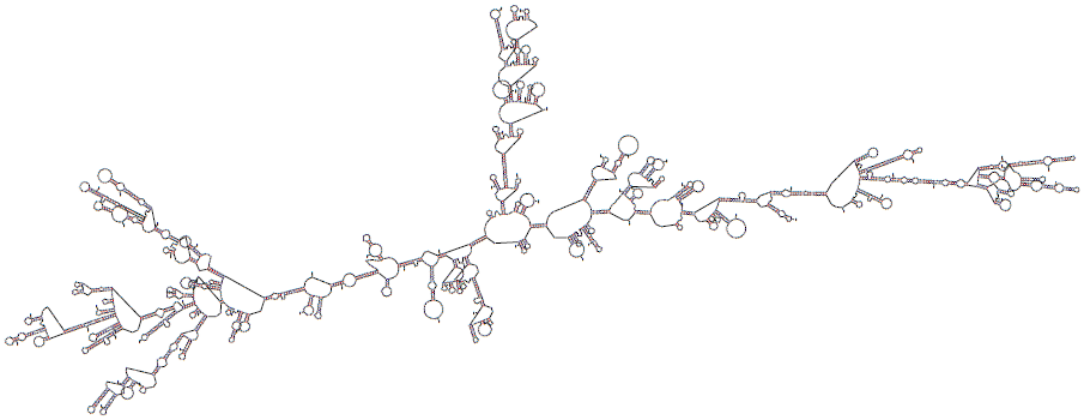

Figure 2.4. The predicted secondary structure of the antisense strand of the 3 kb duplex (pBlueScript SK II (-)) in aqueous buffer. Prediction was carried out using the Mfold web server[29].

F0 was tagged with fluorescein (FAM) at its 5' end to enable tracking via fluorescence. Glycholine samples containing the 3 kb DNA duplex and *F0* in a 1:20 molar ratio were thermal cycled (denatured at 95 °C and cooled to 20 °C at a rate of 4 °C/min), and the binding kinetics of *F0* to its target sequence at 20 °C were then monitored by agarose gel electrophoretic mobility. Refosrmation of the 3 kb duplex over time was followed by the increase in the intensity of the less mobile duplex band after

EtBr staining, and corresponding decrease in the intensity of the more mobile ssDNA band (Figure 2.5, EtBr channel). Thermal cycling of the 3 kb DNA in glycholine kinetically traps a significant fraction of the DNA in its single-stranded state for over 10 days. When the same gel is imaged through the FAM channel to track the position of *F0*, the single visible band (Figure 2.5b, FAM channel) has the same shape and mobility as the 2961 nt ssDNA band in the EtBr channel, confirming that *F0* is binding to its target site on the ssDNA in glycholine.

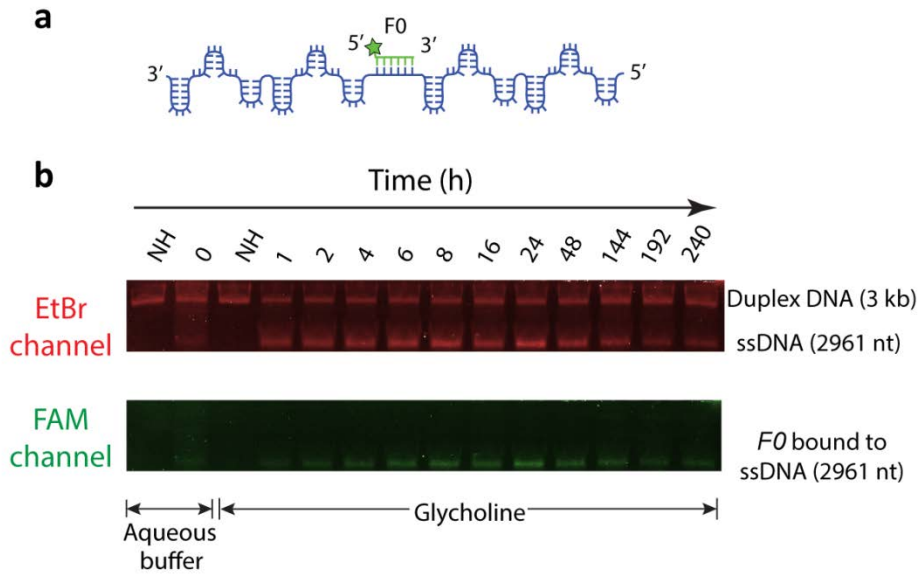

Figure 2.5. Binding of a single oligonucleotide (32 nt) to the 3 kb duplex template. **a**, Schematic illustrating the binding of *F0* (fluorescently tagged with FAM, i.e. carboxyfluorescein) to a specific site on one of the 2961 nt template strands. **b**, Agarose gel (2%) showing the result of heat cycling *F0* and the 3 kb duplex in glycholine, after staining with ethidium bromide (EtBr) and imaging through both an EtBr-specific channel (red) and a FAM-specific channel (green).

Fluorescence analysis shows that *F0* binding reaches a maximum around 8 hours after cooling, corresponding to approximately 60% loading of the *F0* target sites (Figure

2.6c). Over time, *F0* is gradually displaced by reformation of the 3 kb duplex. In contrast, immediately after thermal cycling in aqueous buffer, densitometry analysis indicates that less than 4% of the *F0* target sites are occupied (Figure 2.5b, second lane from the left).

2.6. Step 4: Oligonucleotides cooperatively bind to a gene-length template

In our proposed cycle, viscosity should enable full coating by substrate oligonucleotides of a continuous section on a single stranded template (Figure 2.1, Step 4). To test this possibility, I designed ten 32 nt oligonucleotides, five of which bind directly upstream of *F0* on the template strand (named *L1* through *L5*) and five which bind directly downstream (named *R1* through *R5*). When assembled on the 3 kb antisense strand and ligated together, this series of eleven oligonucleotides—with *F0* in the central position—forms a continuous 352 nt strand (Figure 2.6a).

When all eleven 32-mer oligonucleotides (*L5* through *R5*) are thermally cycled with the 3 kb DNA duplex, we observe that *F0* binds to its ssDNA target and remains bound for over 10 days (Figure 2.6b). In the EtBr image, two bands are observed with similar electrophoretic mobilities indicative of 2961 nt ssDNA. The FAM channel image reveals that F0 is bound to the less mobile of these two bands (asterisk), indicating that these upper and lower bands correspond, respectively, to 2961 nt ssDNA with and without the bound oligonucleotides.

When all eleven 32-mer oligonucleotides are present, the extent of *F0* binding increases from approximately 60% (when only *F0* is present) to approximately 100% loading (within measurement error) of the *F0* target sites when *L5* through *R5* are added (Figure 2.6c). The degree of *F0* template binding was quantified by the ratio of total *F0* fluorescence in the template-bound *F0* band to the total fluorescence in the template-

bound and unbound *F0* bands. However, this method of quantifying *F0* binding may overestimate the percentage of template-bound *F0* because the high concentration of *F0* in the unbound *F0* band may result in quenching of *F0* fluorescence, resulting in a measured *F0* binding peak that is higher than the theoretical maximum based on stoichiometry (Figure 2.6c).

The increase in *F0* binding with the addition of additional oligonucleotides suggests that *L5* through *R5* cooperatively unfold intramolecular structures along the single stranded template via toehold-mediated strand displacement[30-32], increasing the accessibility of target sites in the same region—including the *F0* binding site. This result illustrates how the binding of oligonucleotides to a double stranded template can be greatly enhanced by thermal cycling in a viscous solvent, effectively overcoming strand inhibition.

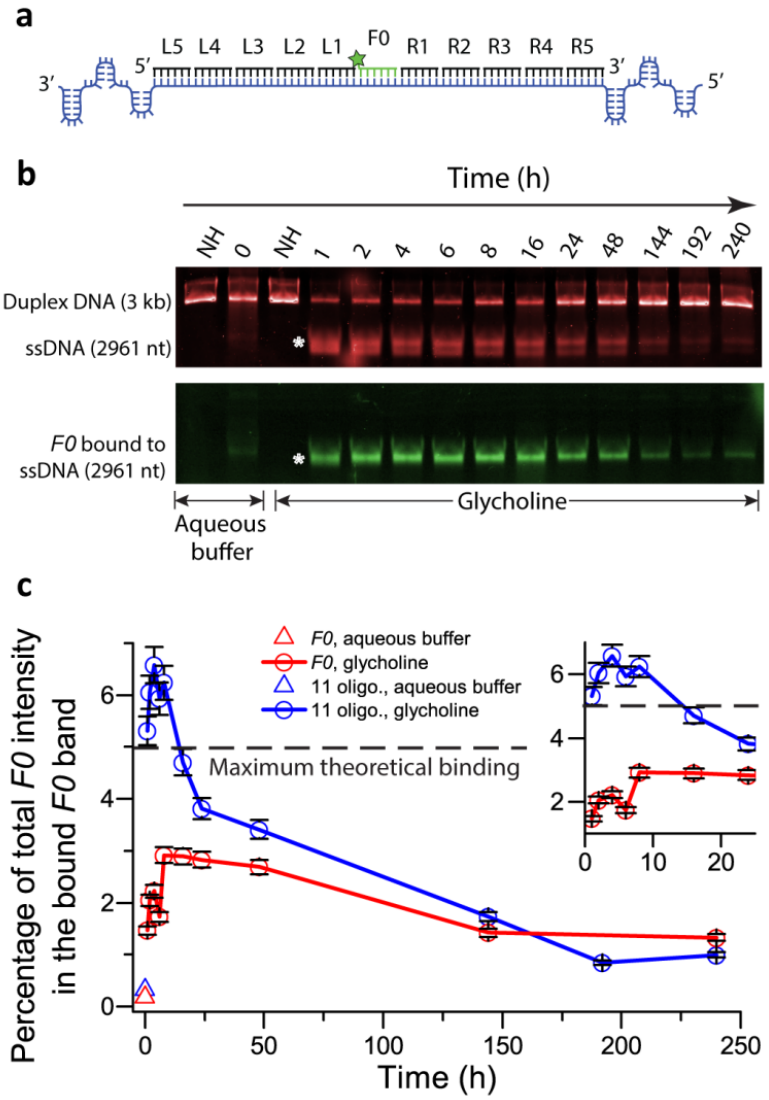

Figure 2.6. Kinetics of oligonucleotide binding to a denatured 3 kb DNA duplex. **a**, A schematic illustrating the binding of eleven oligonucleotides (32 nt each) to the 2961 nt template strand. *F0*, the central oligonucleotide, is labelled with fluorescein (FAM), represented by a green star. b, Agarose gel (2%) showing the results of thermal cycling *L5* through *R5* with 3 kb duplex DNA in glycholine. The gel was stained with ethidium bromide (EtBr), then imaged through EtBr-specific (red) and FAM-specific (green) channels. EtBr image shows the kinetics of 3 kb DNA annealing, while FAM image shows that F0 is binding to the ssDNA. **c**, The percentage of F0 bound to the 3 kb DNA over time. Error bars account for all known sources of error. The maximum theoretical limit of 5% bound F0 is based on a 20:1 molar ratio of F0 to 3 kb template.

2.7. Step 5: Viscosity-enabled information transfer from a 3 kb DNA duplex

As the final step in demonstrating information transfer from a duplex template (Figure 2.1, Step 5), we covalently linked the eleven 32-mers bound to the 2961 nt ssDNA template using T4 DNA ligase, a non-prebiotic but robust ligation method. Importantly, we note that ligase was added *after* samples were thermal cycled and maintained at 20 °C for 4 hours (see Section 2.9.7. DNA ligation procedure), at which point the oligonucleotides were already assembled on the template strand and strand inhibition had already been circumvented (as shown in Section 2.6. Step 4: Oligonucleotides cooperatively bind to a gene-length template). Therefore, the assembly of oligonucleotides on its kinetically trapped template strand was enabled by the viscous solvent, without the aid of a protein enzyme.

To ligate DNA samples after thermal cycling in 100% glycholine, samples were diluted to 15% (w/w) glycholine so that the glycerol content did not inhibit the enzymatic activity of T4 DNA ligase. After ligation, denaturing polyacrylamide gel electrophoresis (PAGE) reveals that when all eleven 32-mer substrates are present with 3 kb DNA, eleven product bands are observed (Figure 2.7b), which correspond to all potential polymers produced by ligation of the 32-mer substrates (Figure 2.7a). The maximum length product possible (352 nt), formed by ligation of all eleven 32-mers, is produced in a higher yield (by total mass) than any other ligation product (Figure 2.7c). Consistent with a template-directed process, the removal of a single oligonucleotide from the set of eleven oligonucleotides in the complete reaction mixture (Figure 2.7a) limits the length of the ligation products in a predictable manner (Figure 2.7b, lanes 6-8). For example, removing either L2 or R2 from the reaction mixture generates the same banding pattern

and product distribution (Figure 2.7c, blue and green curves) because both situations truncate the maximum possible product length in the same way.

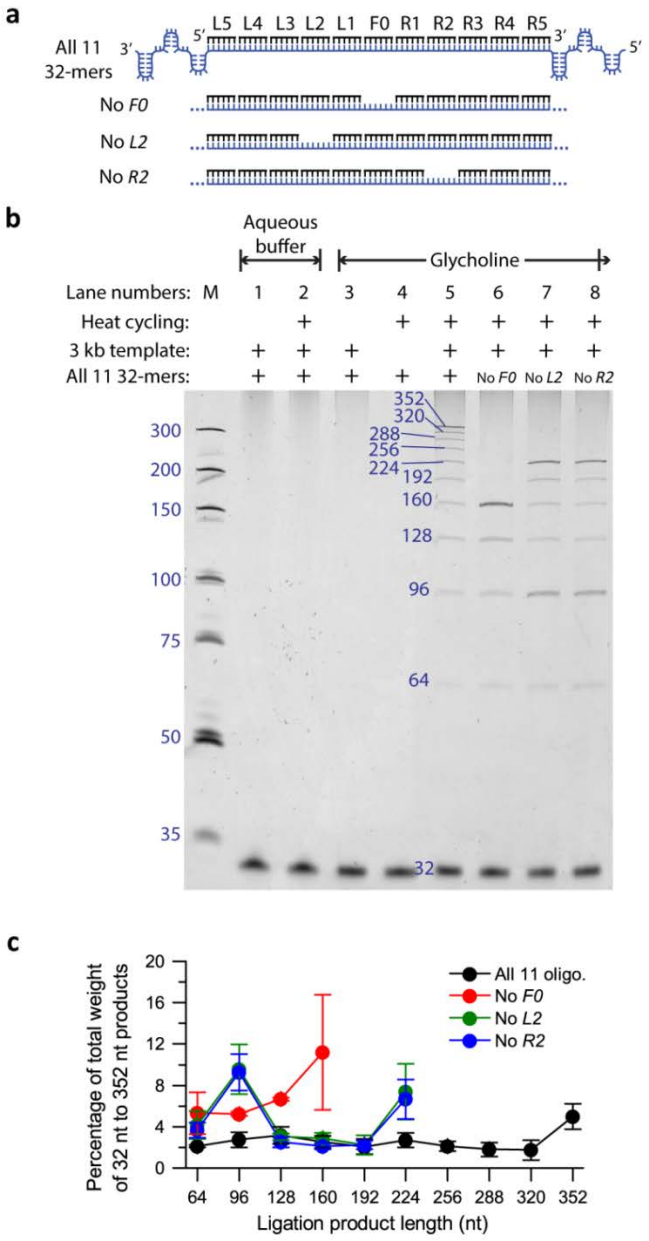

Figure 2.7. Viscosity-enabled gene-length information transfer from a 3 kb duplex. **a**, Schematic of the predicted products of systems composed of the 3 kb duplex and different combinations of oligonucleotide substrates that bind to an internal 352 nt region. **b**, Denaturing polyacrylamide gel (10%) stained with SYBR Gold showing products of thermal cycling the 3 kb duplex with complementary oligonucleotides. Reactions run in lanes 6, 7, and 8 were lacking *F0*, *L2* and *R2*, respectively. When the 3 kb DNA and the 32-mers are heat cycled in aqueous buffer no ligation products are observed (lane 2). **c**, Densitometry analysis showing the percentage of the total mass in each ligation product band. Data is averaged over three experimental repeats, with error bars indicating standard deviations. The ladder is ds DNA (GeneRuler Ultra Low Range Ladder, Thermo Fisher).

This experiment was carried out at a 4:1 molar ratio of oligonucleotides to template, but a significant yield of the maximum length product is still observed when this ratio is reduced to 1:1 (Figure 2.8).

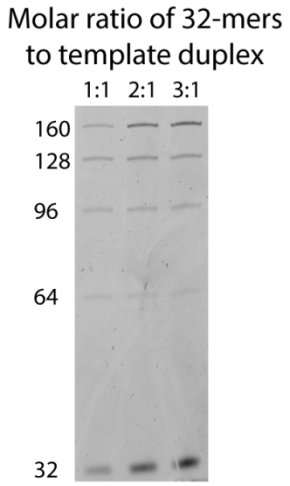

Figure 2.8. Denaturing polyacrylamide gel (10%) showing ligation results obtained from heating and cooling the 3kb DNA in glycholine with a 1- to 3-fold molar ratio of 32-mer oligonucleotides (*L1-L5* and *R1-R5)* to template. *F0* was not added to reaction mixtures, resulting in 5 product bands.

2.8. Conclusion

In this chapter, I have demonstrated that solvent viscosity can be utilized to overcome strand inhibition, enabling information transfer from a gene-length segment (352 nt) within a longer DNA template duplex (3 kb). Our approach utilizes viscosity to alter nucleic acid mobility and annealing kinetics in a length-dependent manner. Specifically, solvent viscosity is able to dramatically slow the reformation of long duplex templates, whereas the hybridization rates of much shorter oligonucleotide substrates are far less

retarded by solvent viscosity. Such size-dependent kinetics provides a time window for information transfer via the template-directed assembly of oligonucleotides.

The temperature and viscosity changes in our experiments reflect conditions that could have occurred on the early Earth in an aqueous pool containing small organic molecules, salts, and oligonucleotides. During the daytime, water would evaporate, concentrating solutes and increasing the viscosity of the solvent; elevated temperatures would have denatured template duplexes. During the nighttime, as the remaining viscous solvent cooled, monomers and short oligomers could diffuse and bind to complementary sequences. Ligation of the assembled oligonucleotides into copy strands would return the system to its initial duplex state. Hygroscopic organic solvents that are of low volatility, miscible with water, and permit the formation of nucleic acid secondary structures[17, 18] are particularly attractive as possible prebiotic milieus for viscosity-mediated replication, as such solvents transition easily between hot/anhydrous and cool/aqueous states.

In our DNA system, an enzyme was used to covalently link oligonucleotide substrates after assembly on a template, but this enzyme could be replaced in future experiments with non-enzymatic ligation chemistry[33]. This final ligation step may be promoted in a eutectic solvent, as non-aqueous solvents have been found to promote the formation of covalent bonds that are not thermodynamically favored in water[22, 34, 35]. Finally, our study, like others addressing problems associated with prebiotic biopolymer formation and replication, highlights the importance and potential value of considering the physical environment as an integrated, dynamic system when exploring how the first system capable of self-replication and chemical evolution might have emerged[13, 36-41].

2.9. Materials and methods

2.9.1. Materials

Glycholine was prepared from glycerol (Sigma, 99%) and choline chloride (Acros Organics, 99%) that was subjected to two rounds of recrystallization from ethanol. Glycerol and choline chloride were placed in a lyophilizer (Virtis) with a liquid nitrogen cold trap for 12 hours before being mixed together in a 4:1 molar ratio. Mixtures were placed in an 80 °C oven until fully liquefied, then dried in a vacuum centrifuge (Labconco) for at least 12 hours before use in experiments. The water content of glycholine prepared in this manner was measured to be 0.07% (wt/wt) using Karl-Fischer titration[18].

The 3 kb DNA template refers to the plasmid pBluescript II SK(-) (Agilent), which was transformed into DH5α *E. coli* cells, purified using a Qiagen plasmid preparation kit, and linearized with HindIII-HF (New England Biolabs).

DNA oligonucleotides were purchased from Integrated DNA Technologies and resuspended in 18.2 MΩ/cm water (Barnstead Nanopure™). For accurate sample manipulation and preparation, the amount of glycholine was measured by mass rather than by volume, and hence concentrations are reported as molal (*m*). To prepare DNA samples in glycholine, aqueous DNA stocks were added to a weighed amount of glycholine and mixed until homogeneous. Water was then removed by vacuum centrifugation for at least 12 hours. All DNA samples in aqueous buffer were prepared with 0.1 M NaCl, 20 mM Tris (pH 8).

2.9.2. DNA sequences

17 nt hairpin DNA:

5'-GCAAAACGAAGTTTTGC-3'

32 bp DNA duplex:

```
5'-GGTGTCAGTAAGCCATTCGAGATCCTCATAGT-3'
3'-CCACAGTCATTCGGTAAGCTCTAGGAGTATCA-5'
```

3 kb DNA duplex: pBluescript II SK(-) vector DNA, phagemid excised from lambda ZAPII, GenBank accession: X52330.1

Oligonucleotides L5-R5:

L5:	nt 2568-2599	*5′*-AGTTGCTCTTGCCCGGCGTCAATACGGGATAA-*3′*
L4:	nt 2600-2631	5′-TACCGCGCCACATAGCAGAACTTTAAAAGTGC-3′
L3:	nt 2632-2663	*5′*-TCATCATTGGAAAACGTTCTTCGGGGCGAAAA-*3′*
L2:	nt 2664-2695	*5′*-CTCTCAAGGATCTTACCGCTGTTGAGATCCAG-*3′*
L1:	nt 2696-2727	*5′*-TTCGATGTAACCCACTCGTGCACCCAACTGAT-*3′*
F0:	nt 2728-2759	5′-CTTCAGCATCTTTTACTTTCACCAGCGTTTCT-3′
R1:	nt 2760-2791	5′-GGGTGAGCAAAAACAGGAAGGCAAAATGCCGC-3′
R2:	nt 2792-2823	5′-AAAAAAGGGAATAAGGGCGACACGGAAATGTT-3′
R3:	nt 2824-2855	5′-GAATACTCATACTCTTCCTTTTTCAATATTAT-3′
R4:	nt 2856-2887	5′-TGAAGCATTTATCAGGGTTATTGTCTCATGAG-3′
R5:	nt 2888-2919	5′-CGGATACATATTTGAATGTATTTAGAAAAATA-3′

For sequences *L5-R5* listed above, the second column refers to the nucleotide position on the pBlueScript II SK(-) sense strand.

2.9.3. DNA thermal denaturation studies

UV absorbance was used to monitor the thermal denaturation of the 17 nt hairpin, while circular dichroism (CD) spectroscopy was used for the 3 kb DNA. The thermal denaturation of the 32 bp duplex was reported in a previous study[18]. Because of the difference in size/extinction coefficient between the 17 nt hairpin and 3 kb DNA, different concentrations were required to achieve comparable UV or CD signal from the two sequences: 35.6 μ*m* (hairpin in buffer); 0.188 μ*m* (3 kb DNA in buffer); 29.7 μ*m* (hairpin in glycholine). All UV and CD measurements were performed using 1 mm quartz cuvettes in a temperature-controlled UV-Vis spectrophotometer (Agilent 8453) or a temperature-controlled CD spectrophotometer (Jasco J-810) with nitrogen flow through the sample chamber at low temperatures. To determine T_m values, heating and cooling

traces were generated for each sample by recording spectra from 220-400 nm, at temperatures between 0 to 100 °C in intervals of 1 °C. Melting curves were generated using the signal at a single wavelength (Figure 2.9). For UV-Vis studies, the absorbance change at 260 nm was used; for CD studies, the ellipticity change at 248 nm was used. T_m values were determined using the method described by Mergny & Lacroix[42].

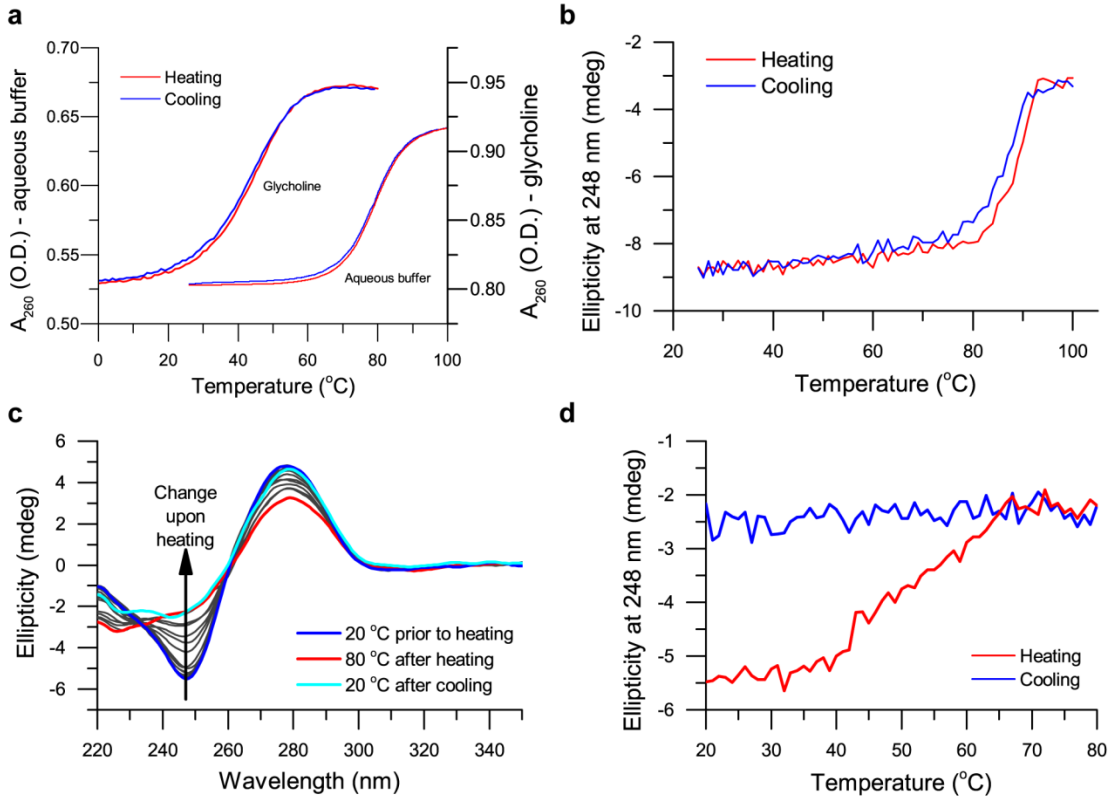

Figure 2.9. Thermal denaturation studies of the 17 nt hairpin and 3 kb DNA sequence. **a,** Thermal denaturation of the DNA hairpin in aqueous buffer and in glycholine. **b,** Thermal denaturation of the 3 kb DNA in aqueous buffer. **c,** CD spectra of the 3 kb DNA in glycholine at 20 °C (blue), at selected temperatures over the course of heating from 20 to 80 °C (grey), at 80 °C (red), and after cooling back to 20 °C (cyan). **d,** CD signal monitored at 248 nm for the 3 kb DNA in glycholine during heating (red) and cooling (blue).

2.9.4. Kinetics of double helix formation after thermal denaturation

UV-Vis absorbance was used to monitor the duplex reformation kinetics of the 17 nt hairpin in glycholine and aqueous buffer, the 32 bp duplex in glycholine and aqueous buffer, and the 3 kb duplex in aqueous buffer. Samples were prepared in 1 mm quartz cuvettes and heated to completely denature all secondary structure (glycholine samples to 80 °C for 5 minutes; samples in aqueous buffer to 95 °C for 5 minutes). After heating, sample temperature was quickly cooled to 20 °C by plunging cuvettes into an ice-water bath (time zero). Cuvettes were quickly dried and placed in the spectrophotometer, a process that required about 9 seconds. Sample spectra were then collected every 1 second.

Agarose gel electrophoresis was used to monitor the duplex reformation kinetics of the 3 kb duplex in glycholine because of the long time scale of the process (days). All samples were heated to the required denaturation temperature (80 °C for glycholine and 95 °C for aqueous buffer) for 5 minutes before being cooled to 20 °C at the appropriate rate on a thermal cycler (BioRad) and placed in a desiccator at 20 °C for the desired amount of time before running on an agarose gel.

2.9.5. PAGE purification of DNA oligonucleotides

Oligonucleotides were purified by PAGE (polyacrylamide gel electrophoresis) using a 40 cm long denaturing polyacrylamide gel (6 M urea, 15%) run in 1x TBE buffer. Approximately 100-200 µg of DNA was loaded into each lane with RNA loading dye (New England Biolabs). Prior to sample loading, the gel was pre-run at 75 W for ~2 hours (until the gel reached a temperature of 55 °C); after sample loading, the gel was run at 75 W for 2-2.5 hours. After electrophoresis, the gel was placed over a thin layer

chromatography plate containing a fluorescent indicator, and the DNA visualized by exposure to 254 nm light from a handheld UV lamp. The 32 nt oligonucleotide band was cut out from the gel with a razor blade, crushed through a syringe, and suspended in 20 mL of 0.2 M NaCl solution. The samples were frozen and placed into a hot water bath (95 ^{o}C) for 10 minutes to further crush the solid acrylamide pieces. After being allowed to shake on a rotary shaker overnight, samples were lyophilized, resuspended in 18.2 MΩ/cm water (Barnstead Nanopure™), purified with ethanol precipitation, and desalted on a Sephadex G-25 column (GE Healthcare Life Sciences Illustra NAP-5 column).

2.9.6. Conditions for agarose gel electrophoresis of glycholine samples

Prior to gel loading, all samples were brought to 50% (wt/wt) glycholine (water was added to the glycholine samples and glycholine was added to the aqueous buffer samples), which was found to be the optimal conditions for running on a gel[18]. Agarose gels (UltraPure agarose, Life Technologies) were run in 1x TAE buffer for 10 minutes at 35 W and then for 1.5 hours at 70 W. After running, gels were stained with ethidium bromide and imaged using a Typhoon Trio+ laser scanner (GE Healthcare). The “EtBr channel” refers to 532 nm excitation with a 610 nm emission filter. The “FAM channel” refers to 488 nm excitation with a 526 nm emission filter.

To show that adding water/glycholine to glycholine/water samples, respectively, did not alter the DNA binding state, control experiments were performed with sequences *F0* (tagged with 5'-fluorescein) and *L5* through *R5* in a 20:1 molar ratio with the 3 kb template duplex (Figure 2.10). In one experiment, the samples were heated and cooled in 100% glycholine before being diluted to 50% glycholine (w/w) immediately before

loading in the gel. In the other experiment, the samples were thermally cycled in 100% glycholine and kept at 20 °C for 4 hours—after which *F0* binding reaches a peak—and then diluted to 50% glycholine. Comparison of the *F0* binding kinetics for these two experiments show that the decay of the bound-*F0* band intensity after dilution to 50% glycholine is on par with that in 100% glycholine up to 12 hours after cooling; after 12 hours, *F0* does appear to be displaced more quickly in 50% glycholine than in 100% due to faster reformation of the 3 kb DNA duplex in a more water-like environment. This result indicates that dilution to 50% glycholine should not affect short oligonucleotide binding on the time scale of running an agarose gel, and that any binding seen on the gel occurs in 100% glycholine rather than in 50% glycholine after the dilution step.

2.9.7. DNA ligation procedure

Samples were prepared as 3 mg aliquots of glycholine with 54.7 n*m* of template duplex and each of the 11 oligonucleotides (or a specific subset of the 11) in the appropriate molar ratio. After the DNA samples were heated and cooled on a thermal cycler, they were diluted with water in order to permit ligation by T4 DNA ligase, whose activity—like that of many enzymes—is inhibited by high concentrations of glycerol. Because the thermal stability of a given duplex is higher in aqueous conditions than in glycholine, any assembled oligonucleotides should remain bound to the template upon dilution with water (see experiments in Section 2.9.6. Conditions for agarose gel electrophoresis of glycholine samples). Each 3 mg glycholine samples was diluted with 10x T4 DNA ligase buffer, and 50 U of T4 DNA ligase (New England Biolabs) to a final solution concentration of 15% glycholine (wt/wt). A concentration of 15% glycholine (wt/wt) was determined to be the limit above which T4 DNA ligase exhibits significantly reduced

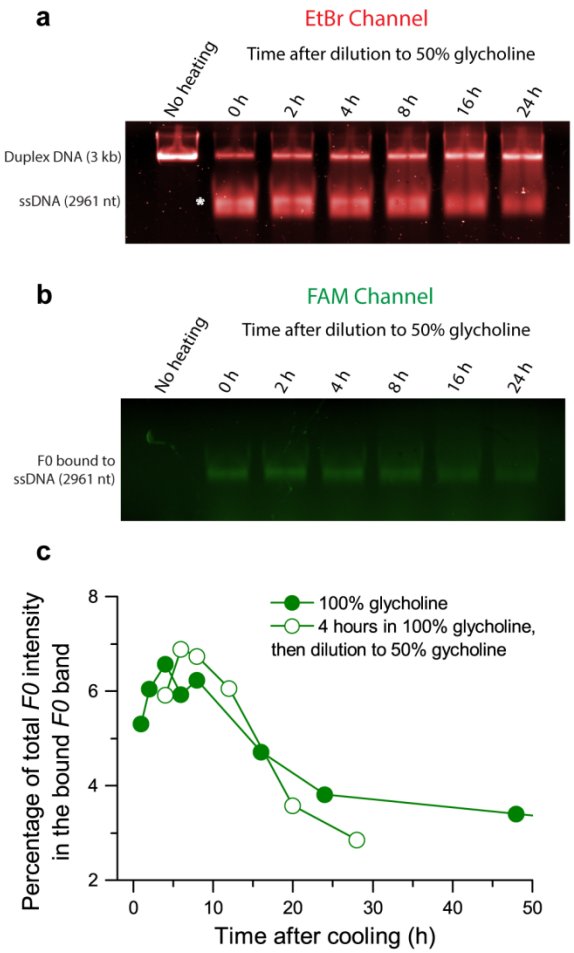

Figure 2.10. Samples containing *L5-R5* (including FAM-labeled *F0*) and 3 kb template duplex were heated and cooling in 100% glycholine for 4 hours, then diluted to 50% (wt/wt) glycholine.**a,** Agarose gel (2%) imaged with the ethidium bromide (EtBr) filter showing the time course of duplex reformation by 3 kb DNA after dilution. Asterisk indicates the position of the bound *F0* band. **b,** The same gel imaged with the FAM filter showing the binding kinetics of *F0* after dilution. **c,** Densitometry analysis comparing the kinetics of *F0* binding after dilution to 50% glycholine versus in 100% glycholine.

activity compared with standard buffer conditions (Figure 2.11). The samples were incubated at 37 °C for 1 hour, heated to 65 °C for 10 min to deactivate the T4 DNA ligase, and mixed with RNA loading dye (New England Biolabs) before loading into a

denaturing polyacrylamide gel (8 M urea, 1x TBE buffer). After running under denaturing conditions (14 W), gels were stained with SYBR Gold and imaged on a Typhoon FLA 9500 laser scanner (GE Healthcare) with a 473 nm laser and a long pass filter at 510+ nm.

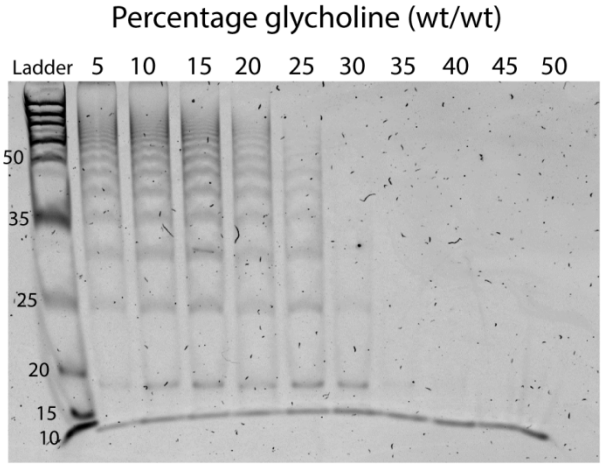

Figure 2.11. Denaturing polyacrylamide gel (10%) showing the efficiency of T4 DNA ligase in a range of glycholine concentrations. A DNA system consisting of two tiling half-complement 12-mers was used. The 12-mers are designed to tile on each other; one sequence contains a 5'-phosphate, allowing it to be ligated, and producing strands of 12, 24, 36…etc bp in length. Sequences of the 12-mers are: (1) 5'-phos-GAATGGGTAGAC-3'; (2) 5'-CATCTGCTTACC-3'. The ladder is ds DNA (GeneRuler Ultra Low Range Ladder, Thermo Fisher).

2.10. References

1. Grossmann TN, Strohbach A, Seitz O. Achieving Turnover in DNA-Templated Reactions. *ChemBioChem* 2008, **9**(14)**:** 2185-2192.

2. Fernando C, Kiedrowski Gv, Szathmary E. A Stochastic Model of Nonenzymatic Nucleic Acid Replication: ‘‘Elongators’’ Sequester Replicators. *Journal of Molecular Evolution* 2007, **64:** 572-585.

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

4. SantaLucia J, Hicks D. The thermodynamics of DNA structural motifs. *Annu Rev Biophys Biomolec Struct* 2004, **33:** 415-440.

5. Ekland EH, Bartel DP. RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* 1996, **382**(6589)**:** 373-376.

6. Sczepanski JT, Joyce GF. A cross-chiral RNA polymerase ribozyme. *Nature* 2014, **515**(7527)**:** 440-442.

7. Wochner A, Attwater J, Coulson A, Holliger P. Ribozyme-Catalyzed Transcription of an Active Ribozyme. *Science* 2011, **332**(6026)**:** 209-212.

8. Luther A, Brandsch R, Kiedrowski Gv. Surface-promoted replication and exponential amplification of DNA analogues. *Nature* 1998, **396:** 245-248.

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

10. Dose C, Ficht S, Seitz O. Reducing Product Inhibition in DNA-Template-Controlled Ligation Reactions. *Angewandte Chemie International Edition* 2006, **45:** 5369-5373.

11. Zhan Z-YJ, Lynn DG. Chemical Amplification through Template-Directed Synthesis. *Journal of the American Chemical Society* 1997, **119:** 12420-12421.

12. Kausar A, McKay RD, Lam J, Bhogal RS, Tang AY, Gibbs-Davis JM. Tuning DNA Stability To Achieve Turnover in Template for an Enzymatic Ligation Reaction. *Angewandte Chemie International Edition* 2011, **50:** 8922-8926.

13. Mutschler H, Wochner A, Holliger P. Freeze–thaw cycles as drivers of complex ribozyme assembly. *Nat Chem* 2015, **7**(6)**:** 502-508.

14. Kreysing M, Keil L, Lanzmich S, Braun D. Heat flux across an open pore enables the continuous replication and selection of oligonucleotides towards increasing length. *Nat Chem* 2015, **7**(3)**:** 203-208.

15. Walker SI, Grover MA, Hud NV. Universal Sequence Replication, Reversible Polymerization and Early Functional Biopolymers: A Model for the Initiation of Prebiotic Sequence Evolution. *PLoS ONE* 2012, **7**(4)**:** e34166.

16. Kramers HA. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* 1940, **7**(4)**:** 284-304.

17. Mamajanov I, Engelhart AE, Bean HD, Hud NV. DNA and RNA in Anhydrous Media: Duplex, Triplex, and G-Quadruplex Secondary Structures in a Deep

Eutectic Solvent. *Angewandte Chemie International Edition* 2010, **49**(36)**:** 6310-6314.

18. Gállego I, Grover MA, Hud NV. Folding and Imaging of DNA Nanostructures in Anhydrous and Hydrated Deep-Eutectic Solvents. *Angewandte Chemie International Edition* 2015, **54**(23)**:** 6765-6769.

19. Abbott AP, Capper G, Davies DL, Rasheed RK, Tambyrajah V. Novel solvent properties of choline chloride/urea mixtures. *Chem Commun* 2003(1)**:** 70-71.

20. Abbott AP, Harris RC, Ryder KS, D'Agostino C, Gladden LF, Mantle MD. Glycerol eutectics as sustainable solvent systems. *Green Chemistry* 2011, **13**(1)**:** 82-90.

21. Lannan FM, Mamajanov I, Hud NV. Human Telomere Sequence DNA in Water-Free and High-Viscosity Solvents: G-Quadruplex Folding Governed by Kramers Rate Theory. *J Am Chem Soc* 2012, **134**(37)**:** 15324–15330.

22. Gull M, Zhou MS, Fernandez FM, Pasek MA. Prebiotic Phosphate Ester Syntheses in a Deep Eutectic Solvent. *Journal of Molecular Evolution* 2014, **78**(2)**:** 109-117.

23. Burcar B, Pasek M, Gull M, Cafferty BJ, Velasco F, Hud NV, *et al.* Darwin's Warm Little Pond: A One-Pot Reaction for Prebiotic Phosphorylation and the Mobilization of Phosphate from Minerals in a Urea-Based Solvent. *Angewandte Chemie International Edition* 2016, **55**(42)**:** 13249–13253.

24. Hagen SJ. Solvent Viscosity and Friction in Protein Folding Dynamics. *Current Protein and Peptide Science* 2010, **11**(5)**:** 385-395.

25. Sorin EJ, Rhee YM, Nakatani BJ, Pande VS. Insights into Nucleic Acid Conformational Dynamics from Massively Parallel Stochastic Simulations. *Biophysical Journal* 2003, **85**(2)**:** 790-803.

26. Doty P, Marmur J, Eigner J, Schildkraut C. STRAND SEPARATION AND SPECIFIC RECOMBINATION IN DEOXYRIBONUCLEIC ACIDS: PHYSICAL CHEMICAL STUDIES. *Proc Natl Acad Sci U S A* 1960, **46**(4)**:** 461-476.

27. Viasnoff V, Meller A, Isambert H. DNA nanomechanical switches under folding kinetics control. *Nano Lett* 2006, **6**(1)**:** 101-104.

28. Zhang X, Qu Y, Chen H, Rouzina I, Zhang S, Doyle PS, *et al.* Interconversion between Three Overstretched DNA Structures. *Journal of the American Chemical Society* 2014, **136**(45)**:** 16073-16080.

29. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research* 2003, **31**(13)**:** 3406-3415.

30. Panyutin IG, Hsieh P. The kinetics of spontaneous DNA branch migration. *Proceedings of the National Academy of Sciences* 1994, **91**(6)**:** 2021-2025.

31. Green C, Tibbetts C. Reassociation rate limited displacement of DNA strands by branch migration. *Nucleic Acids Research* 1981, **9**(8)**:** 1905-1918.

32. Radding CM, Beattie KL, Holloman WK, Wiegand RC. Uptake of homologous single-stranded fragments by superhelical DNA: IV. Branch migration. *J Mol Biol* 1977, **116**(4)**:** 825-839.

33. Engelhart AE, Cafferty BJ, Okafor CD, Chen MC, Williams LD, Lynn DG, *et al.* Nonenzymatic Ligation of DNA with a Reversible Step and a Final Linkage that Can Be Used in PCR. *ChemBioChem* 2012, **13**(8)**:** 1121-1124.

34. Powner MW, Gerland B, Sutherland JD. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* 2009, **459**(7244)**:** 239-242.

35. Schoffstall AM, Barto RJ, Ramos DL. Nucleoside and deoxynucleoside phosphorylation in formamide solutions. *Origins of life* 1982, **12**(2)**:** 143-151.

36. Mamajanov I, MacDonald PJ, Ying JY, Duncanson DM, Dowdy GR, Walker CA, *et al.* Ester Formation and Hydrolysis during Wet-Dry Cycles: Generation of Far-from-Equilibrium Polymers in a Model Prebiotic Reaction. *Macromolecules* 2014, **47**(4)**:** 1334-1343.

37. Mast CB, Schink S, Gerland U, Braun D. Escalation of polymerization in a thermal gradient. *Proceedings of the National Academy of Sciences* 2013, **110**(20)**:** 8030-8035.

38. Apel CL, Deamer DW. The Formation Of Glycerol Monodecanoate By A Dehydration Condensation Reaction: Increasing The Chemical Complexity Of Amphiphiles On The Early Earth. *Orig Life Evol Biosph* 2005, **35**(4)**:** 323-332.

39. Deamer DW, Barchfeld GL. Encapsulation of macromolecules by lipid vesicles under simulated prebiotic conditions. *Journal of Molecular Evolution* 1982, **18**(3)**:** 203-206.

40. Monnard P-A, Szostak JW. Metal-ion catalyzed polymerization in the eutectic phase in water–ice: A possible approach to template-directed RNA polymerization. *Journal of Inorganic Biochemistry* 2008, **102**(5–6)**:** 1104-1111.

41. Usher D. Early chemical evolution of nucleic acids: a theoretical model. *Science* 1977, **196**(4287)**:** 311-313.

42. Mergny JL, Lacroix L. Analysis of thermal melting curves. *Oligonucleotides* 2003, **13**(6)**:** 515-537.

CHAPTER 3 : INVESTIGATING THE LIMITS OF VISCOSITY-MEDIATED INFORMATION TRANSFER FROM A NUCLEIC ACID DUPLEX[b]

3.1. Introduction

Demonstrating a prebiotic route to the non-enzymatic, template-directed synthesis of nucleic acids has been long been stymied by the problem of strand inhibition: assembly of substrates on a template strand is inhibited in the presence of the complementary strand. We have shown that glycholine, a viscous eutectic solvent, facilitates synthesis of a complementary DNA strand from a gene-length region (>350 nt) within a longer DNA duplex (3 kb). Using this model system, we have made a proof-of-concept demonstration that viscous environments can be utilized to overcome the problem of strand inhibition and promote synthesis of a complementary strand from one strand of a template duplex.

Further exploration of the potential for nucleic acid replication in viscous environments requires characterizing the limits of our system. It is particularly important to explore which factors affect the oligonucleotide assembly process and the distribution of newly synthesized sequences produced by viscosity-mediated information transfer. Two important factors in the efficiency of the information transfer process are the lengths of the oligonucleotide substrates and the template sequence. A requirement for very long oligonucleotide substrates or template strands for effective information transfer would raise concerns about prebiotic plausibility. Additionally, substrate/template lengths affect their mobility through a viscous environment, the thermodynamic stability of substrate-template binding, and the degree of intramolecular structure formed on the template

[b] Portions of this chapter have been adapted from: He C, Gállego I, Laughlin B, Grover M, Hud NV. "A viscous solvent enables information transfer from gene-length nucleic acids in a model prebiotic replication cycle." *Nature Chemistry* 2016, In Press. It is reproduced with permission. © Nature Publishing Group.

strands. It is particularly important to consider template length in the context of understanding how a sequence sufficiently long to comprise a viable genetic system or carry out catalytic function emerged. It has been known since the 1960s that extracellular replication systems suffer from a tendency towards "survival of the shortest": since shorter sequences are replicated faster than longer sequences, genetic information is lost over multiple rounds of replication[1] unless a selection pressure is introduced that promotes accumulation of longer sequences[2, 3]. Solvent viscosity may provide a potential solution, since viscous environments should promote the amplification of longer, less mobile templates with a high degree of intramolecular structure over shorter, unstructured sequences. Demonstration that viscous environments can kinetically trap a template whose length is on the order of a ribozyme that catalyzes RNA replication—for reference, the shortest RNA-directed RNA polymerase developed is 200 nt[4]—would show that viscosity may have helped drive the selection and amplification of a replicase in the RNA World (or other similarly complex ribozymes).

We have utilized glycholine as a model viscous solvent. While glycerol can be considered a prebiotically plausible molecule[5], a chemical route for the synthesis of choline has yet to be identified. Therefore, it is important to explore the potential for viscosity-mediated information transfer in other viscous organic solvents. Examining information transfer across a range of solvents will not only allow us to determine whether viscosity is a general physical property—regardless of the specific solvent identity—that could have promoted replication, but will also allow us to more deeply explore the relationship between viscosity and efficiency of the information transfer process. It might be expected that the efficiency of information transfer increases with

increasing viscosity up to a certain critical point, after which even the diffusion of the oligonucleotide substrates is hindered, and yield decreases.

Finally, we have applied our previously demonstrated replication cycle for DNA to RNA, extending its applicability to an RNA World model. Though RNA and DNA have very similar chemical structures—differing only by a 2'-OH group on the sugar—their chemical stabilities, thermal stabilities, and reactivities are quite different. Successful demonstration of viscosity-mediated replication with both DNA and RNA would suggest that viscosity may be a general mechanism for promoting information transfer from a range of nucleic acid polymers, including the first informational polymers that may have preceded RNA[6, 7].

3.2. Simulating the effects of oligonucleotide purity on information transfer

The desired product of our viscosity-mediated information transfer process is a 352 nt strand, the maximum length product formed from assembly of all eleven oligonucleotides. However, the observed products span the range from 32 nt to 352 nt in length (Figure 2.7b). Though the yield of the 352 nt strand is the greatest (by mass), a majority of the starting oligonucleotides are incorporated into shorter strands (Figure 2.7c), resulting from incomplete assembly/ligation of the eleven oligonucleotides on the template. The wide distribution of observed product sizes may be indicative of intrinsic limits in using solvent mediated effects to overcome strand inhibition. However, another possibility is that the observed product distribution is the statistical result of the purity of our oligonucleotide substrates.

The 32 nt oligonucleotides employed in our experiments are synthetically produced, and so contain a certain number oligonucleotides which are missing either

nucleotides or phosphates at their 5' end (Figure 3.1). These oligonucleotides are “un-ligatable” by T4 DNA ligase, which forms a phosphodiester bond between a 3'-OH and a 5'-phosphate in a nicked duplex.

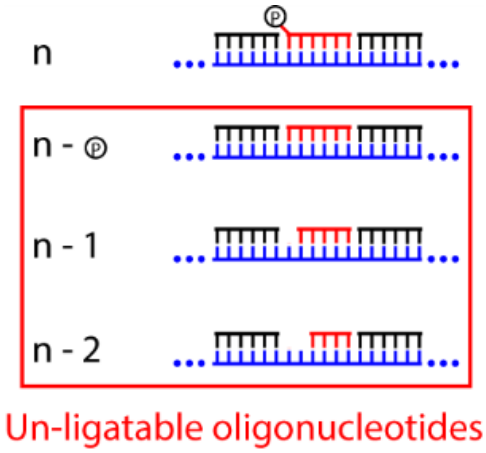

Figure 3.1. Schematic showing the structure of the desired oligonucleotide substrate (n = 32 nt, phosphorylated at the 5' end) as well as un-ligatable oligonucleotides which are either truncated in length or missing the 5'-phosphate necessary for ligation by T4 DNA ligase.

These un-ligatable oligonucleotides are side products generated during the preparation of synthetic oligonucleotides. Synthetic oligonucleotides are produced using solid phase synthesis with phosphoramidites, in a process where mononucleotides are added in individual coupling reactions to the 5' end of a growing chain (Figure 3.2). Because these coupling reactions are not 100% efficient, the result of this solid phase synthesis procedure is a set of oligonucleotides of different sizes; while the primary product is the desired length (i.e. 32 nt), there are a number of products which are 1, 2, 3…etc. nucleotides shorter at the 5' end (i.e. 31-mers, 30-mers, 29-mers…etc). These truncation products are able to compete with the 32-mers for binding to the 3 kb template, but once

bound, are too far from their 5' neighbor to be ligated by T4 DNA ligase. Additionally, synthetic oligonucleotides have a 5'-OH group rather than a 5'-phosphate group, and must be phosphorylated by a kinase prior to our experiments so that they can be ligated by T4 DNA ligase. The kinase phosphorylation reaction is not 100% efficient, and so our pool of oligonucleotide substrates also contains un-phosphorylated 32-mers which, when assembled on the template, cannot be ligated to their 5' neighbor. All of these "un-ligatable" oligonucleotides limit the size of the strands produced by our viscosity-mediated information transfer process.

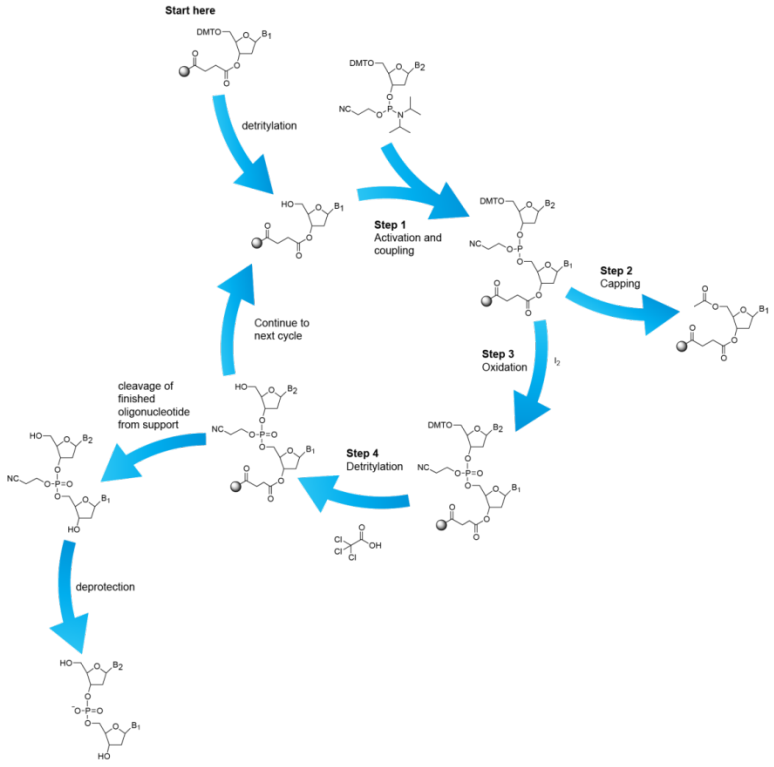

Figure 3.2. Reaction diagram showing the solid phase synthesis of oligonucleotides by addition of mononucleotides in the 3' to 5' direction of the growing chain. Image taken from atdbio's description of solid phase oligonucleotide synthesis: < http://www.atdbio.com/content/17/Solid-phase-oligonucleotide-synthesis#The-Phosphoramidite-method>.

The fraction of un-ligatable oligonucleotides in our pool of 32-mer substrates was quantified by ^{32}P-labeleing the DNA and running samples on a polyacrylamide gel (Figure 3.3). For each oligonucleotide, the band that exhibits the highest radioactivity corresponds to the full length 32-mer, while truncated oligonucleotides (31-mers, 30-mers, etc.) appear sequentially below the primary band. Densitometry analysis of the gel in Figure 3.3 reveals that the average fraction of truncation products in the set of desalted oligonucleotides is 37%, whereas the average fraction of truncation products in the PAGE purified set is 11%.

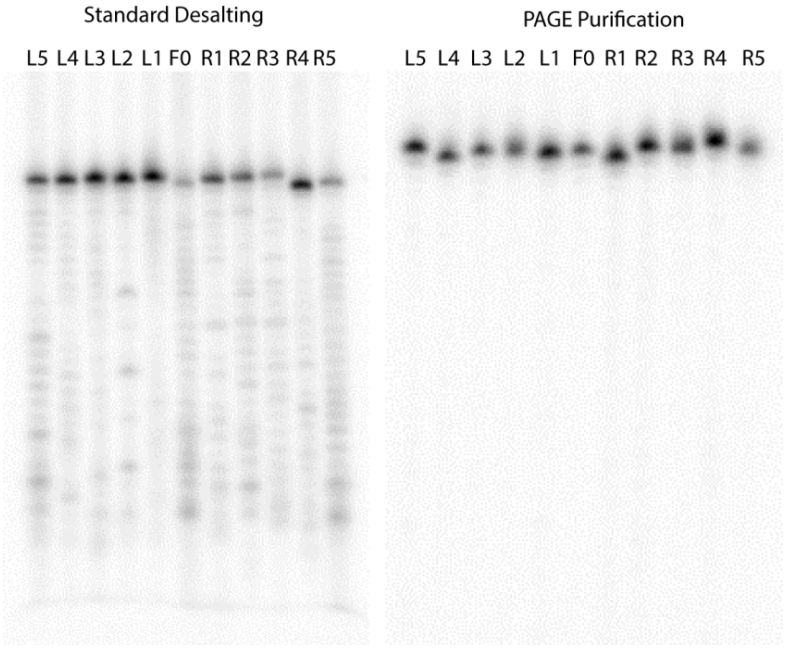

Figure 3.3. Denaturing polyacrylamide gels (10%) used to visualize and quantify truncated oligonucleotides in pools of synthetic 32-mers that underwent either standard desalting or PAGE purification. Samples were ^{32}P-labeled and visualized using autoradiography with X-ray film.

To gauge how a small percentage of un-ligatable oligonucleotides can affect the product distribution, a MATLAB simulation was employed where the percentage of un-

ligatable ends in the starting oligomer population was varied. The simulation consisted of a population of 10^4 templates with each template represented as an 11 element vector. For each binding site, the value of a randomly generated number relative to the un-ligatable percentage determined whether the site is occupied by an un-ligatable oligomer or a 32 nt, phosphorylated oligomer. Two assumptions underlying the simulation are that all binding sites are occupied (reflected experimentally in the appearance of a 2961 nt band, to which *F0* is bound, which migrates with an observably different mobility than its complementary strand; see Figure 2.6) and that the ligation reaction is much faster than the dissociation of a fully complementary 32-mer oligonucleotide from its template site at the ligation temperature of 37 °C[8].

We estimate that, even after purification by polyacrylamide gel electrophoresis, up to 11% of the oligonucleotide substrates are un-ligatable. The simulation correlates well with observed ligation product distributions for two experiments with different substrate purities, using gel purified and unpurified oligonucleotides (Figure 3.4). These results indicate that the yield of maximum length product observed in our experiments is near its theoretical limit posed by oligonucleotide purity. Additionally, the simulation indicates that at higher oligonucleotide purities (95% or higher), the yield of the 352 nt strand dramatically increases, comprising the majority of the products. From the simulation results, we can conclude that the efficiency of our viscosity-mediated information transfer process is determined by the purity of our oligonucleotide substrates rather than by intrinsic limitations in utilizing viscosity to promote template-directed assembly.

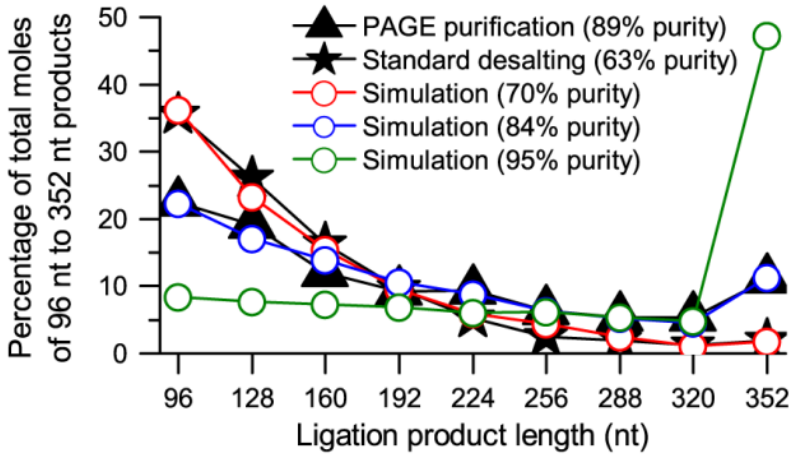

Figure 3.4. Comparison of the experimentally obtained ligation distribution and predicted distributions based on simulation. Percentages refer to the oligonucleotides with ligatable 5' ends (i.e. 32 nt long and functionalized with a 5'-phosphate) in the population.

3.3. Viscosity dependence and information transfer in a range of viscous environments

Our proof-of-concept demonstration of viscosity-mediated information transfer utilizes glycholine as a model viscous environment (CHAPTER 2). If nucleic acid hybridization is indeed a diffusion-limited process in viscous environments, then increased viscosity should promote more efficient information transfer from template strands. More viscous conditions would lead to longer kinetic trapping of intramolecular structures on the template strands, as well as a greater mobility difference between long template strands and short oligonucleotides. However, in principle, a solvent of sufficiently high viscosity will hinder the diffusion of the short oligonucleotides to the point where association with the template strands is hindered. Therefore, we expect the yield of our information transfer process to increase with increasing viscosity up to a certain point, and then decrease as the diffusion of all nucleic acid polymers is severely hindered.

To investigate the viscosity dependence of our method, we carried out thermal cycling of the 3 kb DNA duplex and oligonucleotides *L5* through *R5* in different aqueous mixtures of glycholine (Figure 3.5a) which span a viscosity range from 34 cP (80% glycholine by weight) to 437 cP (100% glycholine). After ligation by T4 DNA ligase, the resulting products were analyzed by running on a denaturing polyacrylamide gel (Figure 3.5b). Densitometry analysis was used to calculate the yield of the viscosity-mediated information transfer process, i.e. the moles of maximum length product (352 nt strand) formed per mole of template duplex (Figure 3.5c). Within aqueous mixtures of glycholine, yield is strongly correlated with the viscosity of the solvent, and decreases as the viscosity drops. While glycholine and it aqueous mixtures exhibit non-Newtonian behavior, they do exhibit a constant viscosity over a certain range of shear rates; the reported viscosity values were measured in this Newtonian range (see Section 3.9.7. Viscosity measurements).

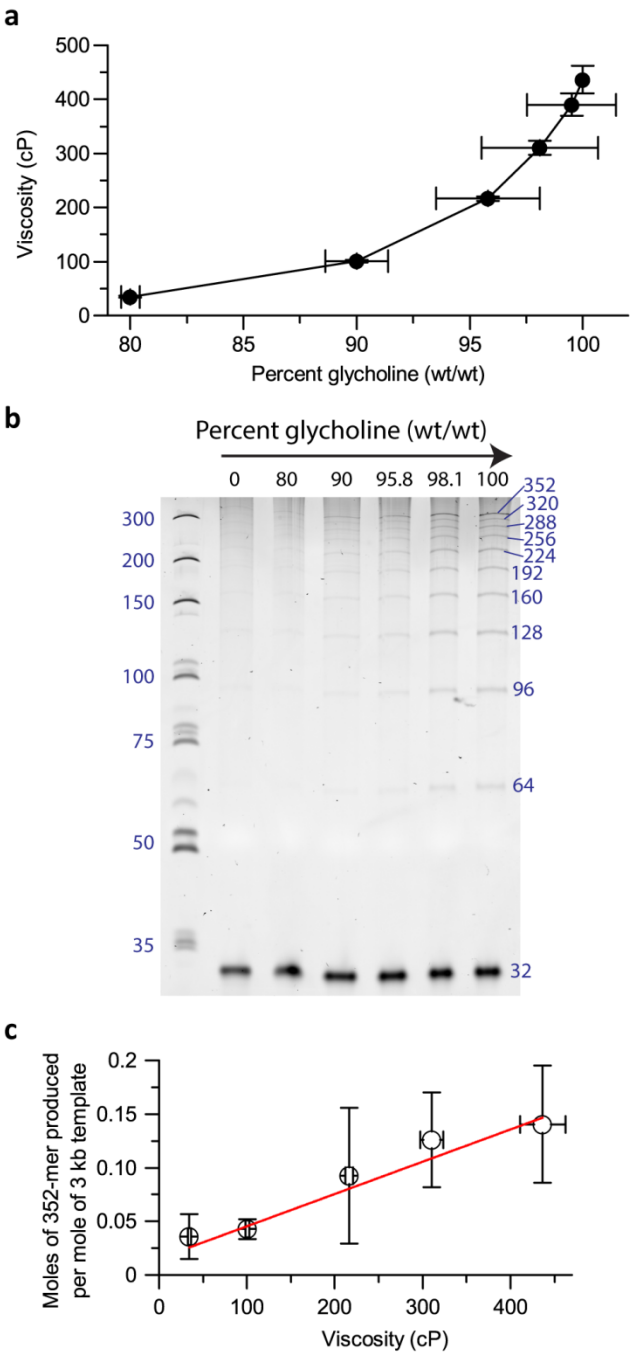

Figure 3.5.Yield of viscosity-mediated information transfer as a function of viscosity in aqueous mixtures of glycholine. **a**, Viscosity measurements of aqueous mixtures of glycholine. Data points shown are averaged from five measurements. **b**, Denaturing polyacrylamide gel (10%, 8 M urea) showing the results of heat cycling and ligating a mixture of the 3 kb DNA template and eleven 32-mer DNA oligonucleotides (*L5-R5*) in different aqueous mixtures of glycholine. **c**, Yield of 352 nt polymer produced after heat cycling in different aqueous mixtures of glycholine with varying viscosities. Yields are averaged over three or four repeated experiments, with error bars indicating standard deviations. The ladder is ds DNA (GeneRuler Ultra Low Range Ladder, Thermo Fisher).

While our results indicate that our information transfer process is indeed viscosity-mediated in aqueous mixtures of glycholine, an important question is whether other viscous environments can also promote information transfer. While glycerol is arguably a molecule which would have been available on the early Earth[5], the presence of choline on the prebiotic Earth is questionable. As such, glycholine is a model viscous environment, and a natural question is whether our information transfer process is dictated by viscosity or by other features of the solvent that may alter the kinetics of nucleic acid hybridization. It is important to note that glycholine is a small molecule solvent, whose viscosity arises from the hydrogen bonding network between the individual components that comprise the eutectic. The effects of large, macromolecular viscogens such as PEG are more distinct from that of small molecule solvents, since high viscosity is also accompanied by molecular crowding effects which increase the rate of reactions such as template duplex reformation from single strands[9], and may serve to accentuate the problem of strand inhibition. As such, in this work I focus on viscous solvents comprised of small molecules.

To explore our information transfer process in solvents beyond glycholine, we considered several additional solvents composed of small molecule viscogens: reline (a eutectic solvent composed of urea and choline chloride in a 2:1 molar ratio), glycerol, and mixtures of glycerol and choline chloride in different molar ratios. These solvents are all more viscous than glycholine (Figure 3.6).

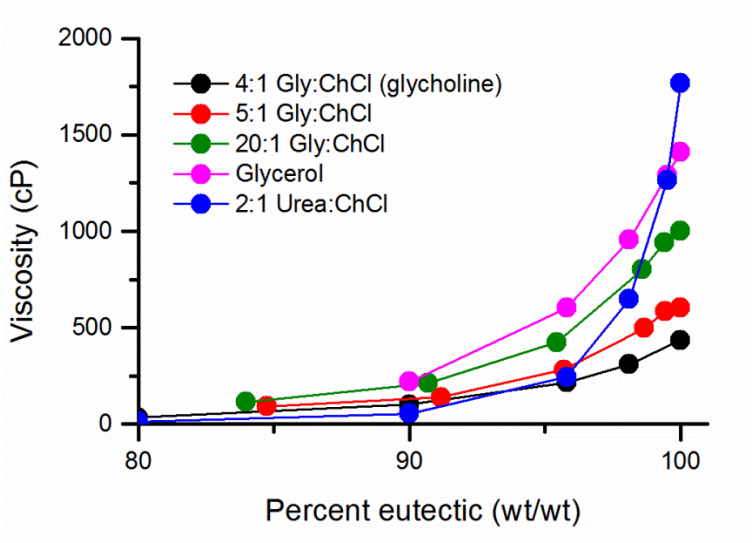

Figure 3.6. Measured viscosities of aqueous mixtures of glycerol, glycerol and choline chloride in various molar ratios, and reline (urea:choline chloride in a 2:1 molar ratio). While these solvents exhibit non-Newtonian behavior, they do exhibit a Newtonian plateau, i.e. a constant viscosity over a certain range of shear rates. The listed viscosity values were measured within the Newtonian plateau (see Section 3.9.7. Viscosity measurements).

All the viscous solvents utilized in this study depress the thermal stability of nucleic acids compared to aqueous buffer.

Table 3.1 compares the melting temperatures of a 32 bp DNA duplex (sequence listed in Section 2.9.2. DNA sequences) in glycerol (full melting curve is shown in Figure 3.20), reline[10], and glycholine[11]. None of the 32 bp duplex melting temperatures appear to be low enough to prohibit stable base pairing of a 32 nt DNA oligonucleotide to its target site on a template strand, thought differences in nucleic acid thermal stability may affect the efficiency of template-directed oligonucleotide assembly in different solvents.

Table 3.1. Melting temperatures of a 32 bp duplex in reline, glycerol, and glycholine, determined by monitoring of absorbance at 260 nm during thermal denaturation.

Solvent	T_m (^oC)
Aqueous buffer (20 mM Tris pH 8, 0.1 M NaCl)	73
Glycholine	49
Glycerol (0.1 M NaCl)	42
Reline	37

The degree to which a template duplex is kinetically trapped as single strands after thermal cycling should be dictated, in part, by the viscosity of the solvent. Figure 3.7 shows the results of thermal cycling the 3 kb DNA template duplex in glycerol and reline at different cooling rates, as well as a quantitative comparison of kinetic trapping between glycerol, reline, and glycholine. In all three solvents, a cooling rate of 4 °C/min was the most gradual cooling rate that trapped a maximal amount of template single strands, and so was chosen as the temperature profile for heat cycling.

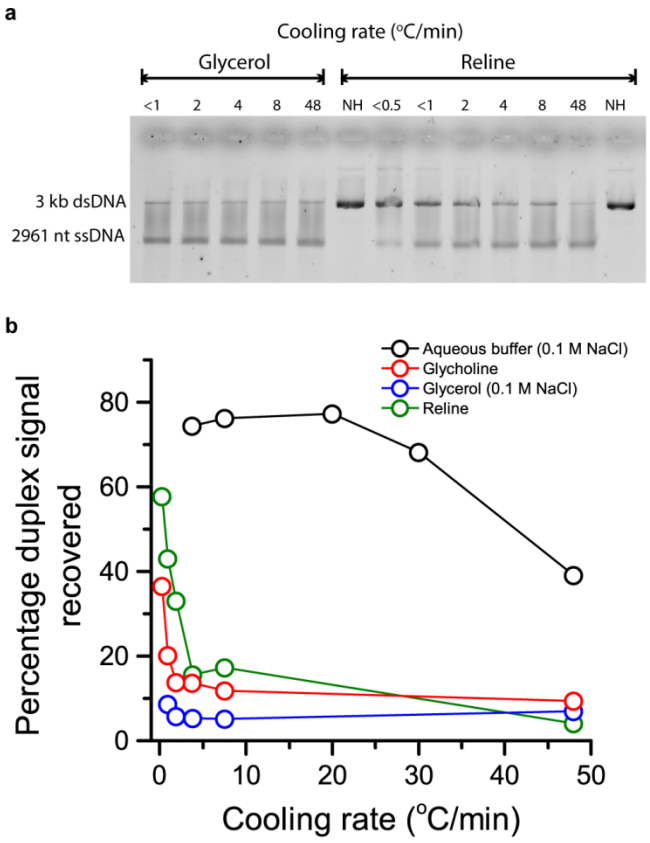

Figure 3.7. Return of 3 kb DNA to a duplex state after thermal cycling in viscous solvent. **a**, Agarose gel (2%) stained with ethidium bromide showing the separation of 3 kb DNA—after heating to denaturing temperatures and cooling at the indicated rates—into duplex (less mobile band) and single stranded (more mobile band) states. **b**, The recovery of the duplex state for the 3 kb DNA as a function of different cooling rates, as measured by densitometry analysis of the duplex band intensity on the agarose gel.

After thermal cycling the 3 kb DNA template duplex with DNA oligonucleotides *L5* through *R5*—including *F0* fluorescently tagged with carboxyfluorescein (FAM)—in glycerol and reline, we see that both viscous solvents enable oligonucleotide assembly on the DNA template, though to a lesser extent than glycholine (Figure 3.8). After sample dilution and ligation of the assembled oligonucleotides, we are able to visualize the products (Figure 3.9a). In all solvents investigated, eleven bands are observed corresponding to the eleven degrees of polymerization possible with oligonucleotides *L5*

through *R5*. In all the solvents investigated, the band containing the maximum length product—352 nt—is quite pronounced compared to the other product bands, indicating significant yields. Quantification of 352 nt product yield by densitometry analysis (Figure 3.9b) shows that yield increases with viscosity within a range of <500 cP; however, there is no clear trend at higher viscosities. These results indicate that solvent factors other than viscosity—such as the dielectric constant of the solvent[12], the reduction in thermal stability of nucleic acid duplexes by the solvent[10], or the specific interaction between choline and nucleic acids[13]—may also play a role in determining nucleic acid mobility and hybridization rates and deserve further investigation.

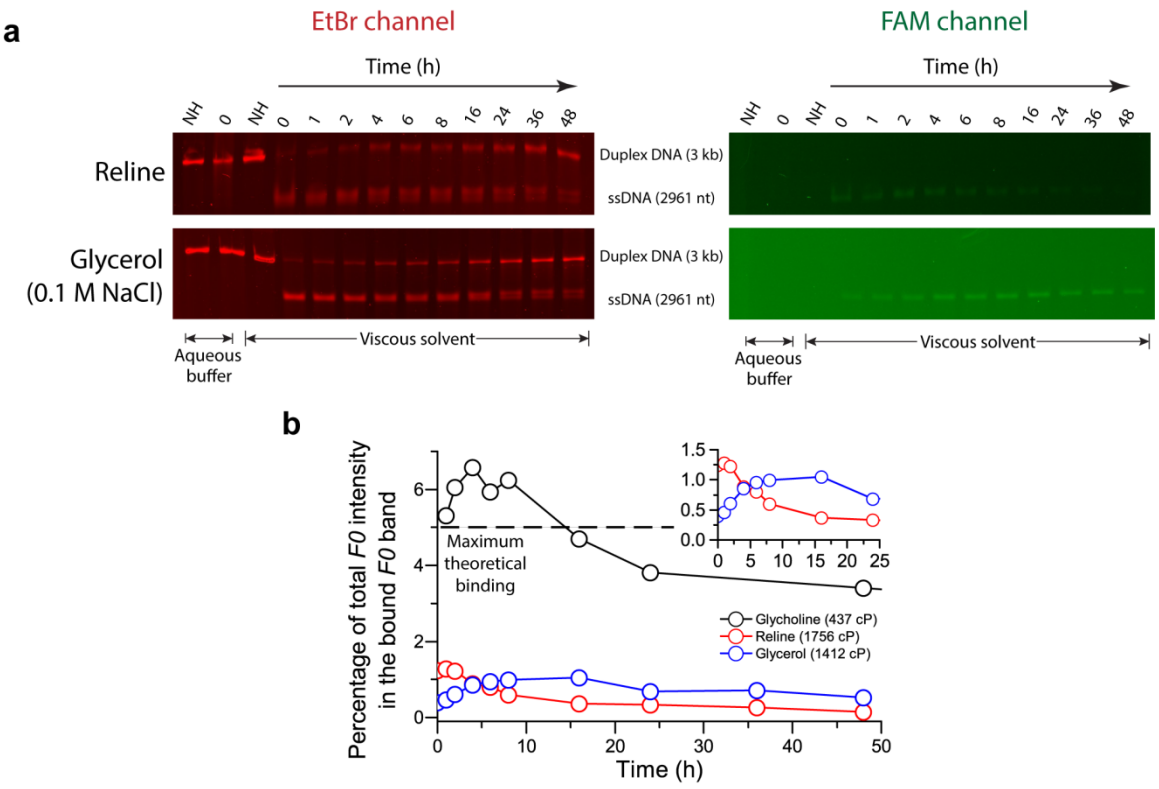

Figure 3.8. Kinetics of oligonucleotide binding to a denatured 3 kb DNA duplex. **a**, Agarose gel (2%) showing the results of thermal cycling DNA oligonuclelotides *L5* through *R5* with 3 kb duplex DNA in reline and glycerol (0.1 M NaCl). The gel was stained with ethidium bromide (EtBr), then imaged through EtBr-specific (red) and

FAM-specific (green) channels. EtBr image shows the kinetics of 3 kb DNA annealing, while FAM image shows that F0 is binding to the ssDNA. **b**, The percentage of F0 bound to the 3 kb DNA over time. Error bars account for all known sources of error. The maximum theoretical limit of 5% template-bound F0 is based on a 20:1 molar ratio of F0 to 3 kb template.

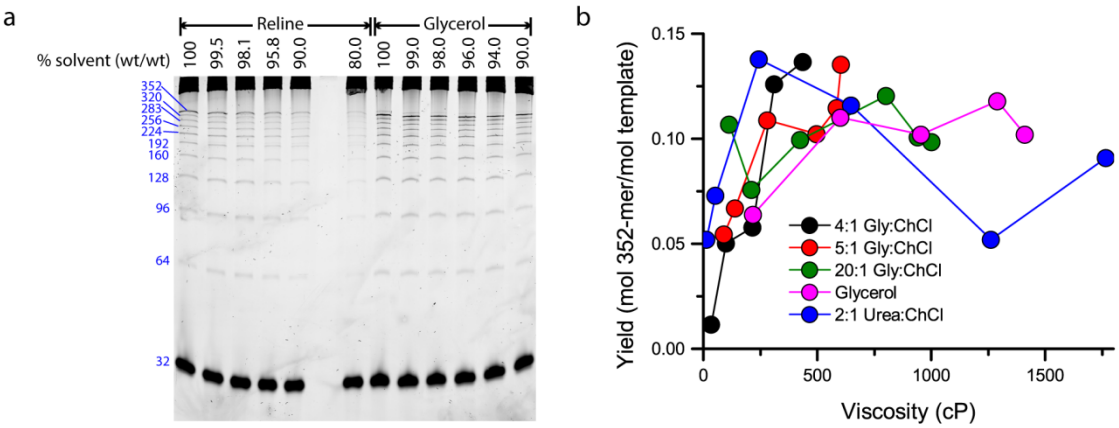

Figure 3.9. a, Denaturing polyacrylamide gel (10%) showing the products of viscosity-mediated information transfer in aqueous mixtures of glycerol and reline. **b,** Yield of maximum length product (352 nt strand) produced per mole of template strand in aqueous mixtures of glycerol, reline, and glycerol/choline chloride solvents. The ladder is ds DNA (GeneRuler Ultra Low Range Ladder, Thermo Fisher).

Importantly, these results show that our viscosity-mediated process is not limited only to glycholine or to eutectic solvents; it appears as if a range of small molecule organic solvents can served as viscous environments for promoting template-directed synthesis of nucleic acids.

3.4. Template-directed synthesis using shorter oligonucleotides

This work focuses on a relatively late stage in the chemical evolution of nucleic acids, assuming the existence of a pool of oligonucleotides and template strands. How were these oligonucleotides produced in the first place? Identifying a prebiotic route for the *de novo* (i.e. non-template directed) synthesis of oligonucleotides is another active area of

research. Studies of oligomerization aided by mineral surfaces have shown that oligonucleotides up to 50+ nt long can be polymerized from mononucleotides, though the bulk of the products are 30-50 nt long[14]. Therefore, the length of our oligonucleotide building blocks (32 nt) falls within the range of prebiotic plausibility. However, we also want to test the limits of our information transfer process using shorter oligonucleotides.

The length of the oligonucleotide building blocks used in viscosity-mediated information transfer must be long enough that the oligonucleotides form stable duplexes with their template sites at low temperatures (20 °C in our experiments). Because eutectic solvents depress the thermal stability of nucleic acid duplexes compared to aqueous buffer (

Table 2.1), thermal denaturation studies were conducted in glycholine to determine the minimum duplex length that is stable at room temperature (Figure 3.10a). Four DNA duplexes of different lengths were investigated: 10, 15, 20, and 25 bp. The sequences of these duplexes correspond to the sequence of the 32 bp DNA duplex (50% AT/GC content) utilized in previous studies (Section 2.9.2. DNA sequences), beginning from the 5' end of the top strand. The resulting UV absorbance vs. temperature curves indicate that the minimum duplex length that is stable in glycholine at 20 °C is ~15 bp (Figure 3.10a, red curve).

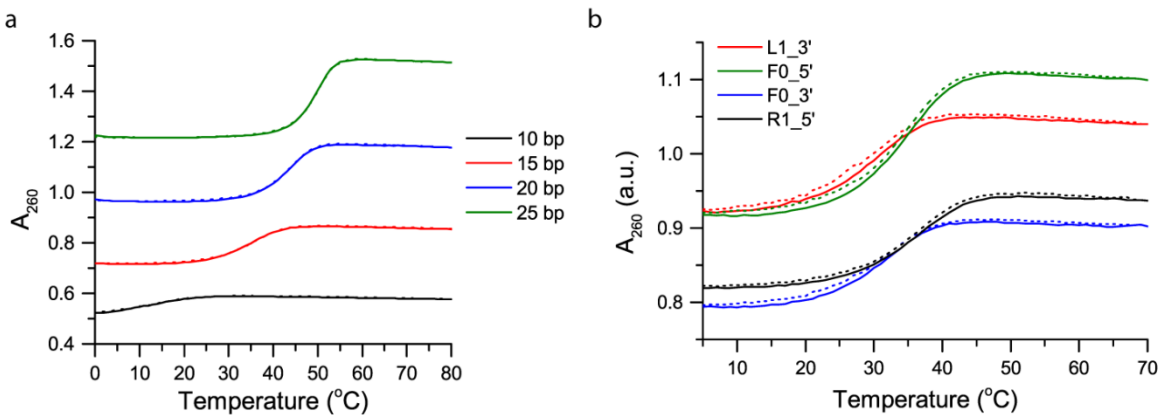

Figure 3.10. Thermal denaturation of DNA duplexes of varying lengths in glycholine, monitored by UV absorbance at 260 nm. **a,** Heating (solid lines) and cooling (dotted lines) traces of fragments of the 32 bp duplex described in Section 2.9.2. DNA sequences. **b,** Heating (solid lines) and cooling (dotted lines) traces of 16 bp duplexes whose sequences correspond to the 5' and 3' halves of the 32 nt oligonucleotides *L1*, *F0*, and *R1* (described in Section 2.9.2. DNA sequences).

Based upon a required length of ~15 bp for duplex thermal stability in glycholine, we tested our viscosity-mediated information process using oligonucleotide building blocks that are only 16 nt in length (i.e. half the size of the 32 nt oligonucleotides utilized in the previous study). The sequence of each 32 nt oligonucleotide building block (from *L5* to *R5*) was divided into half, and the corresponding 16 nt sequences were named in the following manner: *L5_5'*, *L5_3'*, …*R5_3'*. Thermal denaturation studies in glycholine were carried out on four of these 16 nt sequences and their complementary strands, confirming that the thermodynamic equilibrium favors the duplex state at 20 °C (Figure 3.10b). When a mixture of the 3 kb DNA duplex template and oligonucleotides *L5_5'* through *R5_3'* is heat cycled in glycholine and incubated with T4 DNA ligase, analysis of the sample by denaturing polyacrylamide gel electrophoresis shows 22 bands corresponding to the 22 possible product lengths formed (Figure 3.11a). These results

confirm that viscosity-mediated information transfer can be carried out with relatively short (16 nt) oligonucleotide building blocks that may have formed from spontaneous polymerization of mononucleotides (perhaps assisted by a mineral surface[14]).

Quantification of the gel bands by densitometry shows greater than expected production of strands that are 128 nt and 208 nt long (Figure 3.11b), produced by ligation of eight and thirteen 16 nt oligonucleotides together, respectively. The high yields of these specific products suggest that the template strand contains a particularly stable intramolecular duplex at the site corresponding to either the *L1_5'* or *R1_3'* (Figure 3.11c). This stable intramolecular duplex prevents efficient binding of either *L1_5'* or *R1-_3'*, leaving a "gap" in the assembled oligonculeotides that prevents the efficient production of maximum length, 352 nt strands and increases the yield of 128 nt and 208 nt strands. Interestingly, the sequences *L1_5'* and *R1_3'* do not appear particularly prone to forming hairpin structures (Figure 3.11d), suggesting that long range intramolecular structure may be inhibiting oligonucleotide binding at these sites.

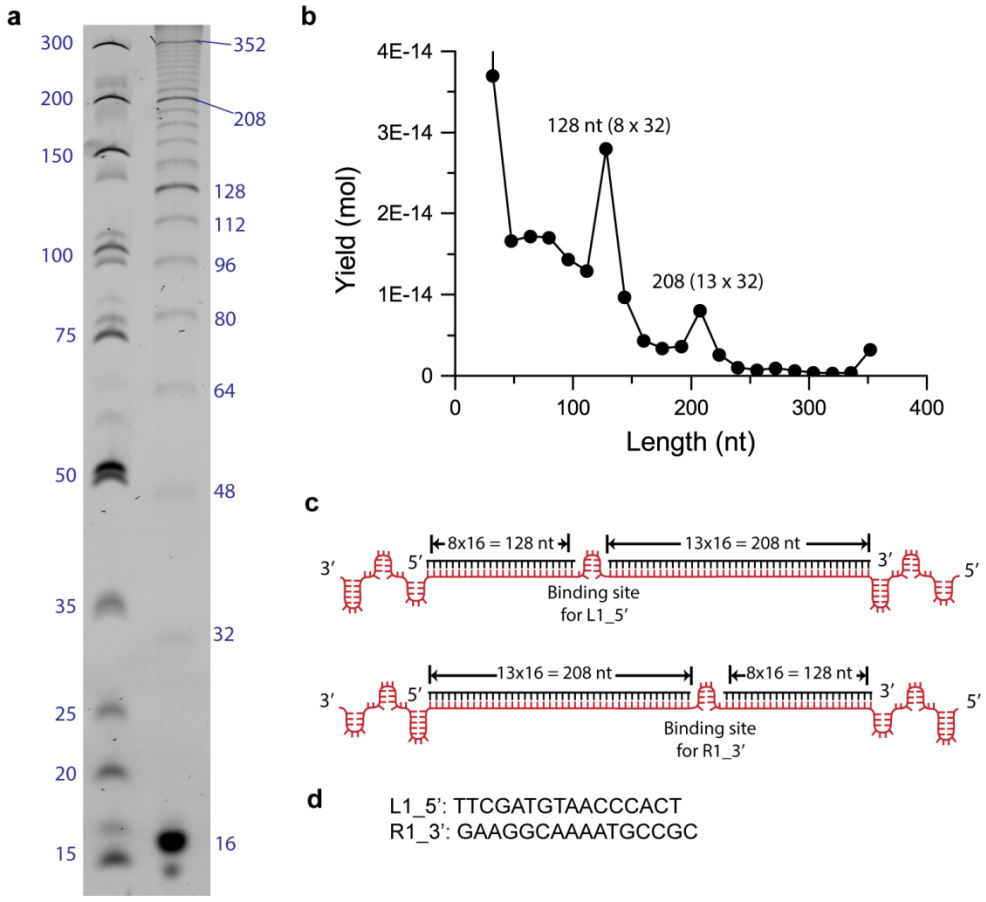

Figure 3.11. Viscosity-mediated information transfer using 16 nt DNA substrates in glycholine. **a,** Denaturing polyacrylamide gel (10%) showing the results of thermal cycling a sample containing the 3 kb DNA duplex template and 16 nt oligonucleotides *L5_5'* through *R5_3'* (1:4 molar ratio of template to oligonucleotide). Samples were ligated with T4 DNA ligase after thermal cycling. **b,** Yields of the 22 possible product strands obtained by densitometry analysis of gel **a. c,** Schematic showing the effect of an intramolecular duplex at the binding sites for either *L1_5'* or *R1_3'* on the oligonucleotide assembly pattern and ligation product distribution. **d,** Sequences of *L1_5'* and *R1_3'* oligonucleotides. The ladder is ds DNA (GeneRuler Ultra Low Range Ladder, Thermo Fisher).

3.5. Viscosity-mediated information transfer from a 545 bp DNA duplex

We have shown that thermal cycling in viscous glycholine enables information transfer from a 352 nt portion of a 3 kb DNA template duplex. While this template provides a model of a mixed, arbitrary sequence of gene length, a length of 3 kb is arguably beyond the limits of prebiotic feasibility for a template. Additionally, our viscosity-mediated information transfer process should favor the replication of long templates over short templates, since long template strands are likely to form more intramolecular duplex structure—enhancing their kinetic trapping—and are less mobile (i.e. will re-anneal with the complementary template strand more slowly) in a viscous solvent than short sequences. Theoretically, templates that are too short (i.e. lacking in intramolecular structure and/or too mobile) will not be efficiently copied in a viscous environment. These factors all motivated our desire to test viscosity-mediated information transfer with a shorter template; in particular, a template whose size is on the same order as a minimal RNA-directed RNA polymerase (~200 nt minimum length)[4].

To investigate the role of template length in viscosity-mediated information transfer, we tested our approach using a shorter DNA template duplex: a 545 bp fragment of the original 3 kb template duplex, whose sequence contains the 352 nt region for binding of DNA oligonucleotides *L5* through *R5*. After thermal cycling of the 545 bp template duplex and 32 nt oligonucleotides *L5* through *R5* (at a 4:1 molar ratio between each oligonucleotide and template), the assembled oligonucleotides were ligated using T4 DNA ligase. Analysis by PAGE shows that all eleven possible products are formed, including the full-length 352 nt product (Figure 3.12).

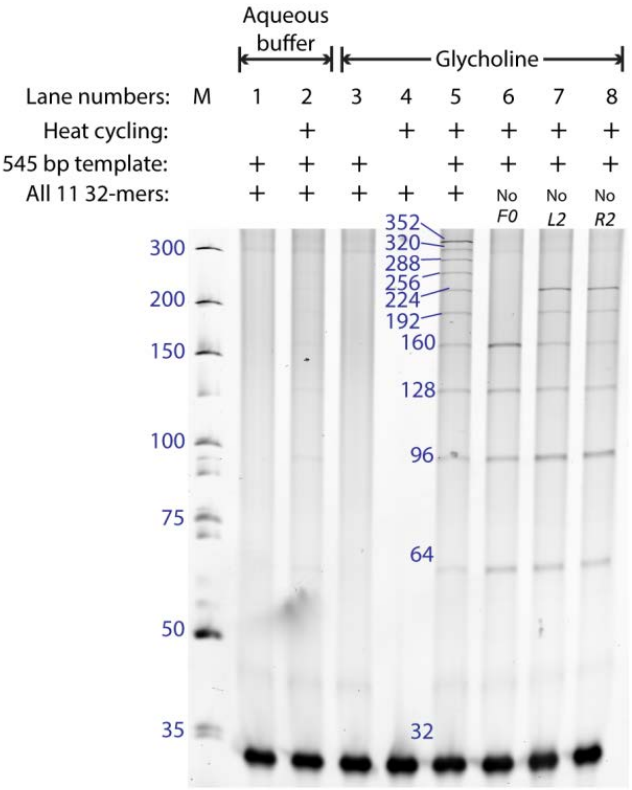

Figure 3.12. Denaturing polyacrylamide gel (10%) analysis of glycholine samples containing the 545 bp DNA template duplex and a 4:1 molar ratio of DNA oligonucleotides *L5* through *R5*. Samples were heat cycled and incubated with T4 DNA ligase. The reference marker is ds DNA (GeneRuler Ultra Low Range Ladder, Thermo Fisher).

Densitometry shows that the same yield is obtained from both the 545 bp and 3 kb DNA templates: 0.14 moles of 352 nt product are formed for each mole of template duplex. These results indicate that a template whose length is on the same order as a ribozyme (a few hundred nt) can become efficiently kinetically trapped in glycholine, enabling template-directed information transfer.

3.6. Viscosity-mediated information transfer from a 545 bp RNA duplex

For practical reasons, we utilized DNA as a model genetic polymer for our initial demonstration of viscosity-mediated information transfer from a duplex template. However, many researchers believe that, prior to the evolution of DNA and coded proteins, RNA played a more central role in both information storage and catalysis. We therefore also tested our approach with RNA by using a system composed of a RNA template duplex with the same sequence as the 545 bp DNA template duplex described above (r545 bp duplex) and eleven 32 nt RNA oligonucleotides (r*L5* through r*R5*) that have the same sequences as the eleven 32 nt DNA oligonucleotides described above.

As seen with the DNA system, thermal cycling of the r545 bp template duplex and oligonucleotides r*L5* through r*R5* in glycholine, and subsequent ligation with T4 RNA ligase 2, results in formation of all eleven possible products (Figure 3.13, lane 5). As in the DNA system, the removal of a single RNA oligonucleotide from the reaction mixture limits the length of the ligation products in a predictable manner (Figure 3.13, lanes 6-8). These results indicate that viscosity can promote information transfer from both DNA and RNA polymers, suggesting that our approach is a robust and potentially general strategy for overcoming strand inhibition for a range of informational polymers.

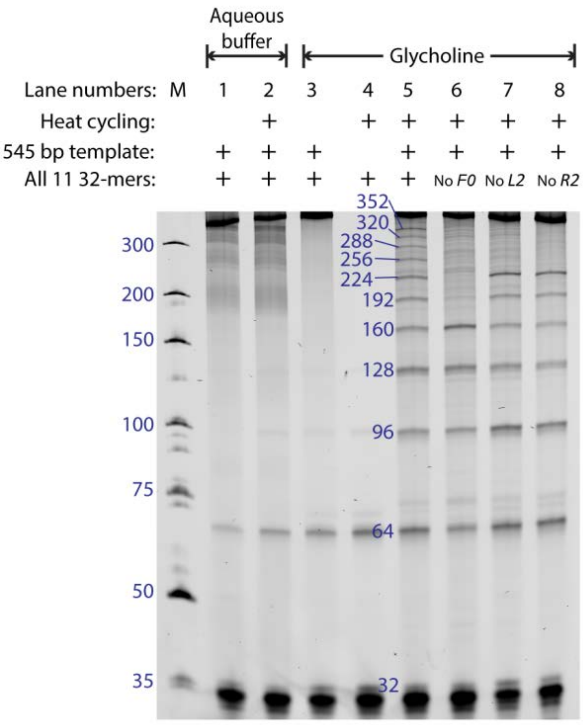

Figure 3.13. Denaturing polyacrylamide gel (10%) analysis of glycholine samples containing the 545 bp RNA template duplex and a 4:1 molar ratio of RNA oligonucleotides *L5* through *R5*. Samples were heat cycled and incubated with T4 RNA ligase 2. The ladder is ds DNA (GeneRuler Ultra Low Range Ladder, Thermo Fisher).

3.7. Information transfer between DNA and RNA

In aqueous buffer, RNA sequences generally exhibit higher thermal stabilities than the corresponding DNA sequences[15]. However, thermal denaturation studies of the 32 bp duplex formed by oligonucleotide *F0* and its complementary strand *F0'* suggest that this trend may be reversed in glycholine (Figure 3.14). The melting temperature of the DNA/DNA duplex (dF0/dF0') is nearly 12 °C lower than that of the RNA/RNA duplex (rF0/rF0'), with the two hybrid duplexes (dF0/rF0' and rF0/dF0') exhibiting thermal stabilities in between the homo-duplexes (

Table 3.2). It is unclear which interactions are responsible for differences in thermal stability between DNA and RNA in glycholine, but it has been shown that sequence-specific interactions between choline and the major/minor grooves of the nucleic acid duplex are responsible for reversing the generally positive correlation between GC content and duplex thermal stability in aqueous buffer[13]. One clue may be that the three duplexes containing at least one RNA strand appear to exhibit greater hysteresis upon cooling in glycholine than the DNA/DNA duplex (Figure 2.9), suggesting that the mobility of single stranded RNA through glycholine and/or its conformational rigidity is affecting the hybridization of ssRNA into a duplex.

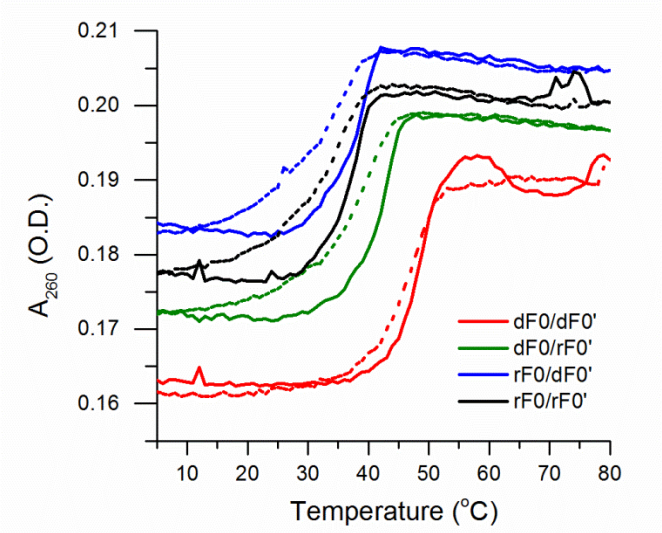

Figure 3.14. Heating (solid) and cooling (dotted) traces of the 32 bp duplex formed by oligonucleotide *F0* and its complementary strand *F0'*, for combinations of DNA/DNA, DNA/RNA, RNA/DNA, and RNA/RNA strands.

Table 3.2. Melting temperatures of the 32 bp F0/F0' duplex in glycholine, varying each strand between DNA and RNA.

Composition of F0/F0' duplex	T_m in Glycholine (^oC)
dF0/dF0'	48.1
dF0/rF0'	41.7
rF0/dF0'	37.5
rF0/rF0/	36.5

After thermal cycling of RNA oligonucleotides rL5 through rR5 with a DNA template in glycholine and incubating with T4 RNA ligase 2, ligation products were visualized with a denaturing polyacrylamide gel (Figure 3.15). With a high excess of oligonucleotide present, newly synthesized RNA strands up to 96 nt and 128 nt are seen on the 3 kb and 545 bp DNA template, respectively. These results indicate that information transfer from a DNA template to a complementary RNA strand is possible in glycholine, though it is a less efficient process than a DNA/DNA or RNA/RNA system. It would be interesting to see whether, with different template sequences, oligonucleotide sequences, or solvents, we can achieve more effective information transfer from DNA to RNA, perhaps shedding light on the transition from an RNA World (RNA-templated RNA polymerization) to a DNA/protein world (DNA-templated RNA polymerization).

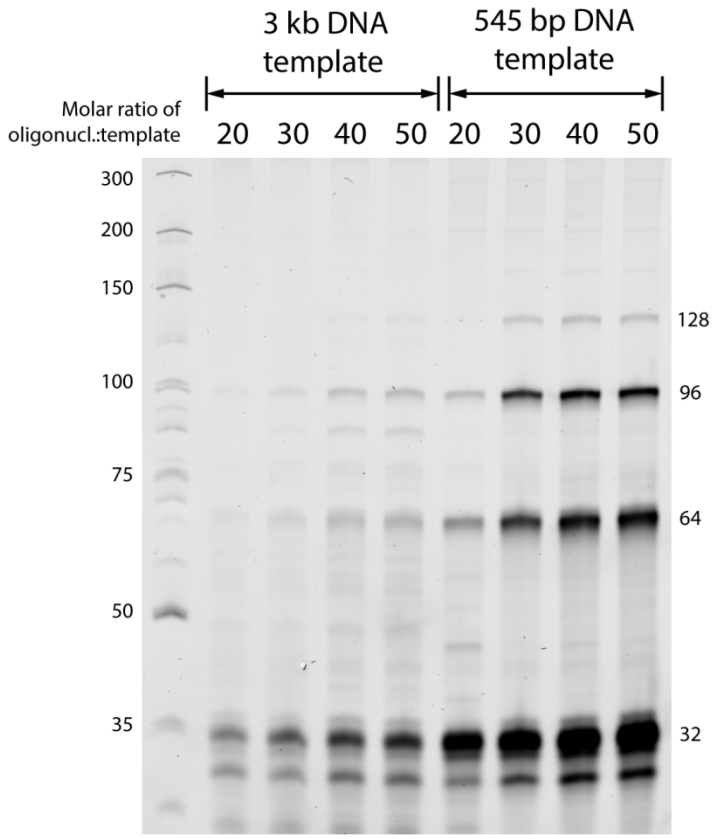

Figure 3.15. Denaturing polyacrylamide gel (10%) showing the results of thermal cycling RNA oligonucleotides rL5 through rR5 with a 3 kb and 545 bp DNA template in glycholine. Samples were incubated with T4 RNA ligase 2. The ladder is ds DNA (GeneRuler Ultra Low Range Ladder, Thermo Fisher).

3.8. Conclusions

In the previous chapter I demonstrated that glycholine promotes the template-directed assembly of 32 nt DNA oligonucleotides on a 3 kb DNA duplex template. In this chapter, I explored the effect of different parameters on viscosity-mediated information transfer from a duplex template, and explored the limits of the information transfer process. Using a kinetic Monte Carlo simulation of template-directed oligonucleotide assembly, I examined the consequence of substrate impurity on the distribution of strand lengths produced by viscosity-mediated information transfer. The simulation results, which

match reasonably well with data obtained from viscosity-mediated information transfer experiments, indicate that the yield of maximum length product (352 nt strands) is limited not necessarily by viscosity-mediated effects, but rather by the purity of our oligonucleotide substrate (i.e. the proportion of the oligonucleotide population which is 32 nt in length and contains the 5' phosphate necessary for ligation to an adjacent oligonucleotide). Though the yield of the 352 nt strands dramatically increases with higher oligonucleotide purity, appreciable yields of newly synthesized strands are achieved even when the oligonucleotide substrates used are impure (63% purity). These results are an encouraging illustration of the robustness of viscosity-mediated information transfer, as the prebiotic pool in which nucleic acid replication evolved is likely to have contained a mixture of oligonucleotides with varying lengths and sequences.

To further explore the robustness and prebiotic feasibility of viscosity-mediated information transfer, I investigated the effects of using oligonucleotide substrates and templates of different lengths. While our proof-of-principle system utilized 32 nt substrates and a 3 kb template, the prebiotic feasibility of nucleic acids of this length is questionable. However, oligonucleotide length cannot be so short that template binding is unstable, while template strands that are too short are highly mobile and may not form the necessary intramolecular structure for effective kinetic trapping. In glycholine, I have shown that information transfer from a 352 nt region of a 3 kb DNA duplex template is possible utilizing oligonucleotides that are 16 nt in length, a size which could have been feasibly produced by non-templated polymerization of activated mononucleotides on a mineral surface[14]. A length of 16 nt is approximately the minimal length needed to form a stable duplex in glycholine at 20 °C; shorter oligonucleotide lengths may be effective

substrates as well at temperatures lower than 20 °C. I have also shown that viscosity promotes information transfer from a 545 bp template duplex (both DNA and RNA), a template whose size is on the same order as the minimal RNA-directed RNA polymerase generated through *in vitro* evolution[4]. These results indicate that viscosity-mediated information transfer is feasible with templates and oligonucleotide substrates that were potentially available on the prebiotic Earth.

Additionally, I have demonstrated that information transfer is promoted in viscous environments beyond glycholine, including glycerol, mixtures of glycerol/choline chloride in a range of ratios, and reline. While yields of the information transfer process are positively correlated with viscosity within aqueous mixtures of a single solvent, there does not appear to be a strong relationship between viscosity and yield across different solvents. This result suggests that viscosity is a key factor impacting template-directed assembly, but not the sole factor. Other solvent characteristics—such as the dielectric constant of the solvent[12], the reduction in thermal stability of nucleic acid duplexes by the solvent[10], or the specific interaction between cations and nucleic acids[13]—are also important in determining nucleic acid thermodynamics and kinetics.

Finally, I have extended previous work with a model DNA system to a more prebiotically relevant informational polymer, RNA. I demonstrated that glycholine enables information transfer from a 545 bp RNA duplex template, a result which suggests that viscosity may have been a general mechanism for promoting replication from the first informational polymer, perhaps even a polymer that preceded RNA.

3.9. Materials and methods

3.9.1. Materials

Glycerol/choline chloride solvents were prepared as described previously in Section 2.9.1. Materials. Reline (urea and choline chloride in a 2:1 molar ratio) was prepared in an identical fashion as glycholine, using urea (Alfa Aesar, 99.0+%) and choline chloride (Acros Organics, 99%). Both urea and choline chloride were recrystallized from ethanol before use. Glycerol was prepared with 20 mM Tris pH 8 and 0.1 M NaCl prior to addition of any nucleic acid.

DNA templates and oligonucleotides were prepared as described previously in Section 2.9.1. Materials. RNA template was prepared as described below in Section 3.9.10. Production of 545 bp RNA duplex template. RNA oligonucleotides were purchased from GE Dharmacon and resuspended in 18.2 MΩ/cm water (Barnstead Nanopure™).

Preparation of anhydrous solvent samples containing DNA or RNA was carried out as described previously in Section 2.9.1. Materials.

3.9.2. DNA and RNA sequences

Sequences of the 3 kb DNA template duplex and DNA oligonucleotides *L5* through *R5* are listed in Section 2.9.2. DNA sequences. Sequences of RNA oligonucleotides r*L5* through *rR5* are listed below:

rL5: 5′-AGUUGCUCUUGCCCGGCGUCAAUACGGGAUAA-3′
rL4: 5′-UACCGCGCCACAUAGCAGAACUUUAAAAGUGC-3′
rL3: 5′-UCAUCAUUGGAAAACGUUCUUCGGGGCGAAAA-3′
rL2: 5′-CUCUCAAGGAUCUUACCGCUGUUGAGAUCCAG-3′
rL1: 5′-UUCGAUGUAACCCACUCGUGCACCCAACUGAU-3′
rF0: 5′-CUUCAGCAUCUUUUACUUUCACCAGCGUUUCU-3′
rR1: 5′-GGGUGAGCAAAAACAGGAAGGCAAAAUGCCGC-3′
rR2: 5′-AAAAAAGGGAAUAAGGGCGACACGGAAAUGUU-3′
rR3: 5′-GAAUACUCAUACUCUUCCUUUUUCAAUAUUAU-3′
rR4: 5′-UGAAGCAUUUAUCAGGGUUAUUGUCUCAUGAG-3′

rR5: 5′-CGGAUACAUAUUUGAAUGUAUUUAGAAAAAUA-3′

The sequence of the 545-mer duplex extends from nucleotide position 2547 and to nucleotide position 130 on the 3 kb template strand (nucleotide positions given in reference to the sense strand of pBlueScript SK II (-)).

3.9.3. Simulating the effects of oligonucleotide purity on viscosity-mediated information transfer

The following MATLAB script was written to simulate the effects of oligonucleotide purity on the product distribution from viscosity-mediated information transfer:

```
function counts_all = ligation_yields(x)

%   x = percentage of un-ligatable ends for EACH FS; should be 11x1 vector
x = x./100; % convert from percentage to fraction

figure()
hold on

counts_all = [];

% === Tell Matlab NOT to reset the random generator seed whenever it
restarts === %
rng('shuffle')

num_FS = 11; % number of oligos per templates
num_templates = 10000; % number of total templates in the population

template = {};
%   template is a cell array of matrices; each cell in the array
%   corresponds to one template molecule
pol = [];
% pol is a vector containing the lengths of every ligation product

colors = ['r'; 'g'; 'b'; 'k'; 'm'; 'c']; % colors for plotting

% === Determine what is bound to each template === %
for i = 1:num_templates
zero_indices = [];
        for j = 1:num_FS
        % For every site on every template, generate a random number
        dart = rand;
        if dart <= x(j)
```

```
            template{i,1}(1,j) = 0;
            zero_indices = [zero_indices; j];
            else
            template{i,1}(1,j) = 1;
            end
            end
            [num_zeros, ~] = size(zero_indices);
            pol_next = zeros(num_zeros+1,1);
            if num_zeros == 0
            pol_next = 11;
            else
            pol_next(1) = zero_indices(1) - 1;
            pol_next(end) = num_FS - zero_indices(end);
            if num_zeros >= 2
            for z = 2:num_zeros
            pol_next(z) = zero_indices(z) - zero_indices(z-1) - 1;
            end
            end
            end
            pol = [pol; pol_next];
end

% Delete all zero elements from pol:
pol(pol==0) = [];
% Calculate statistics
counts = hist(pol,11); % 1x11 vector
counts_all = [counts_all; counts];

for m = 1:length(colors)
plot(1:1:11, counts, strcat('o-',colors(m)), 'MarkerSize', 12);
end

x = x.*100; %convert back to percentages for legend purposes
%legend(num2str(x),'Location','northwest')
hold off

fig=gcf;
set(findall(fig,'-property','FontSize'),'FontSize',16)
title('Ligation Product Distribution With Varying Percentage of Un-Ligatable
Ends', 'FontSize', 20)
xlabel('Degree of Polymerization');
ylabel('Number of Polymers Produced');

end
```

The simulation goes through each oligonucleotide binding site on the template strand (*L5* through *R5*) and utilizes a kinetic Monte Carlo-type mechanism for determining which oligonucleotides (ligatable vs. un-ligatable) are bound to each site. For each site, a random number is generated, and the value of the random number relative to the percentage of un-ligatable oligonucleotides in the oligonucleotide population is used to decide which type of oligonucleotide is bound to that site. After the binding state of all 1000 templates has been determined, it is assumed that any bound, ligatable oligonucleotide that has a 5' neighbor is ligated to its neighbor (i.e. the ligation reaction is 100% efficient). The resulting product strands are then counted to generate a product length distribution.

3.9.4. DNA ligation procedure

Ligation of DNA oligonucleotides after assembly on a DNA template was carried out with T4 DNA ligase (New England Biolabs), using the procedure described in Section 2.9.7. DNA ligation procedure. Control experiments show that the activity of T4 DNA ligase is not appreciably decreased in 15% (wt/wt) reline or 20% (wt/wt) glycerol (Figure 3.16).

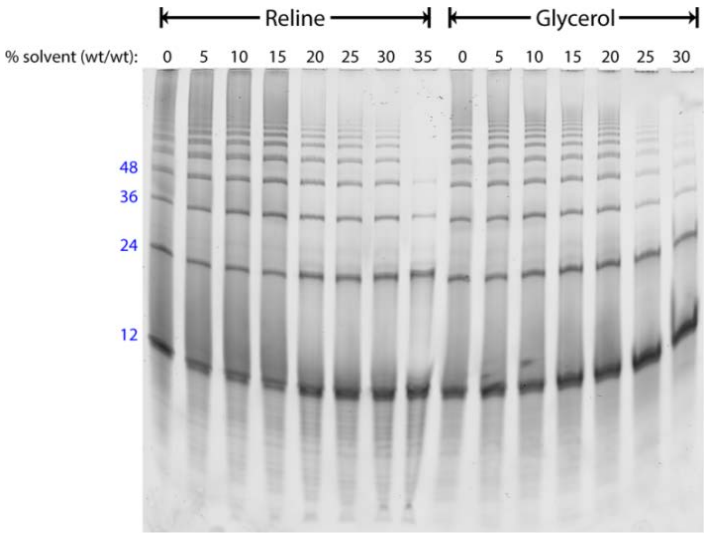

Figure 3.16. Denaturing polyacrylamide gel (10%) showing the result of using T4 DNA ligase to ligate two half-complementary DNA 12-mers in different aqueous mixtures of glycerol (0.1 M NaCl) and reline. The 12-mers are designed to tile on each other; one sequence contains a 5'-phosphate, allowing it to be ligated to its upstream neighbor, and producing strands which are 12, 24, 36…etc bp in length. Sequences of the 12-mers are: (1) 5'-phos-GAATGGGTAGAC-3'; (2) 5'-CATCTGCTTACC-3'.

Ligation of DNA oligonucleotides after assembly on an RNA template was carried out by SplintR Ligase, also known as PBCV-1 DNA Ligase or Chlorella virus DNA Ligase, an enzyme that joins DNA oligonucleotides annealed onto an RNA splint[16, 17] with high efficiency and specificity. SplintR ligase joins DNA assembled on a RNA template with roughly the same efficiency as T4 DNA ligase, which has been reported to repair nicks in some DNA/RNA hybrid duplexes. Additionally, SplintR ligase appears to exhibit a higher tolerance for glycholine than T4 DNA ligase, with significant ligation activity in up to 40% glycholine (w/w) (Figure 3.17). The ligation procedure for DNA oligonucleotides on an RNA template is identical to that described in Section 2.9.7. DNA ligation procedure, except 25 U of SplintR ligase were added to each sample after thermal cycling.

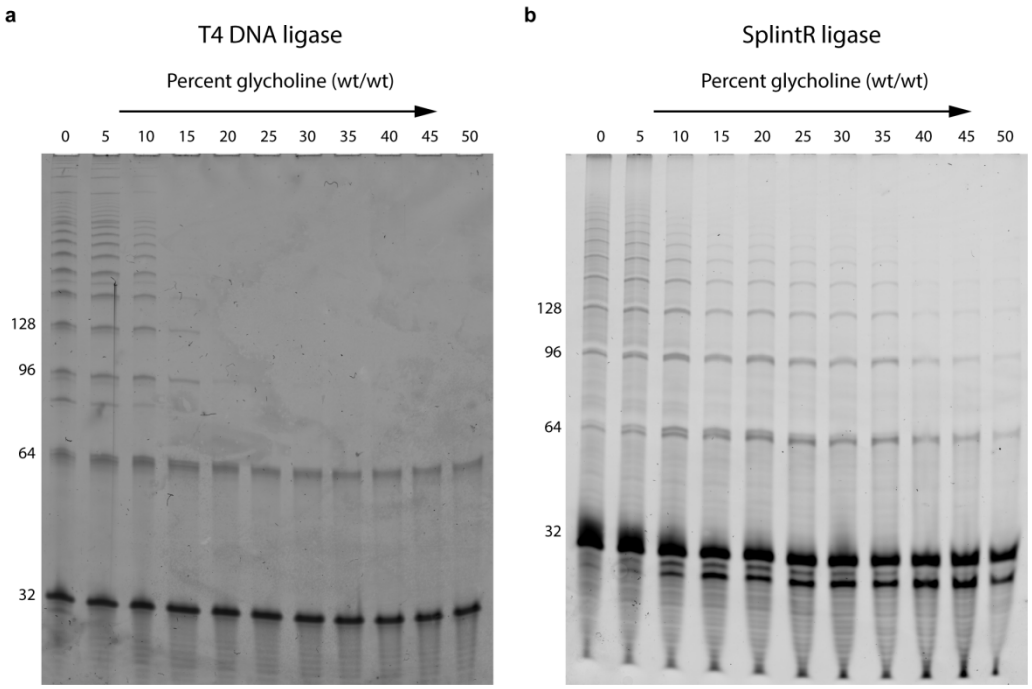

Figure 3.17. Denaturing polyacrylamide gels (10%) showing the results of ligating a mixture of two half-complementary tiling 32-mer sequences (one DNA, phosphorylated at the 5' end, and one RNA) in different aqueous mixtures of glycholine. The DNA sequence contains a 5'-phosphate, allowing it to be ligated to its upstream neighbor and generating strands that are 32, 24, 96…etc bp long. Sequences of the 32-mers are: (1) DNA: 5'-phos-CTTCAGCATCTTTTACTTTCACCAGCGTTTCT-3'; (2) RNA: 5'-GUAAAAGAUGCTGAAGAUCAGUUGGGUGCACG-3'. **a,** Ligation results using T4 DNA ligase. **b,** Ligation results using SplintR ligase.

3.9.5. ^{32}P-labeling of oligonucleotides

To visualize the purity of the oligonucleotide substrates, oligonucleotides were 5'-end labeled with ^{32}P. The phosphorylation reaction was carried out by incubating oligonucleotides (Integrated DNA Technologies) with ^{32}P-ATP (Perkin Elmer) and T4 polynucleotide kinase (New England Biolabs). After labeling, 1-2 µCi of each oligonculetoide was combined with RNA loading dye (New England Biolabs) and loaded into a 38 cm long denaturing polyacrylamide gel (6 M urea, 15%) in 1x TBE buffer. Prior to sample loading, the gel was pre-run at 75 W for ~2 hours (until the gel reached the

steady state temperature of 55 °C); after sample loading, the gel was run at 75 W for 2-2.5 hours. After running, the gel was exposed to X-ray film and the resulting autoradiograph was imaged on a Typhoon FLA 9500 laser scanner (GE Healthcare).

3.9.6. PAGE purification of oligonucleotides

DNA oligonucleotides were purified by PAGE (polyacrylamide gel electrophoresis) using the same procedure described in Section 2.9.5. PAGE purification of DNA oligonucleotides. RNA oligonucleotides were purified using an identical procedure, except samples were *not* frozen and plunged into a 95 °C water bath prior to placement on a rotary shaker, to prevent spontaneous cleavage of RNA.

3.9.7. Viscosity measurements

The viscosity of all the viscous solvents reported here were measured on a Physica MCR 300 viscometer with a 50 mm, 1° cone plate geometry. While all viscous solvents showed some non-Newtonian behavior (i.e. viscosity dependence upon the shear rate), they all exhibited a constant viscosity over a certain shear range of shear rates (a Newtonian plateau). For each sample, at least five values were measured in the Newtonian plateau (for example, shear rates of 10-300 s^{-1} in glycholine and its aqueous mixtures; see Figure 3.18) and then averaged to obtain the reported viscosities.

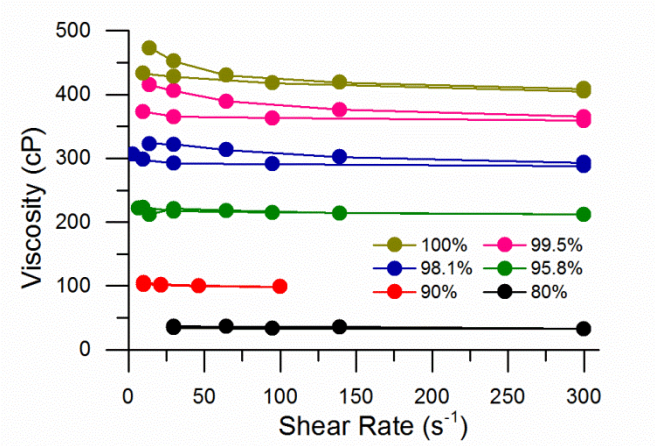

Figure 3.18. Plot of measured viscosities vs. shear rate for glycholine and its aqueous mixtures (percentages are reported as wt/wt glycholine).

3.9.8. Preparation of DNA samples in aqueous mixtures of viscous solvents

To prepare DNA samples in aqueous mixtures of viscous solvent, a solution containing the appropriate amounts of aqueous DNA solution, NaCl, and Tris pH 8 was dried in a vacuum centrifuge (Labconco). The dried DNA/salt solution was then dissolved in the desired amount of solvent.

3.9.9. Production of 545 bp DNA duplex template

The 545 bp DNA template was produced by polymerase chain reaction (PCR) to amplify a portion of the 3 kb DNA template, and then purified using a PCR purification kit (Qiagen). The following primers were used in the PCR reaction:

5′-GAATAGTGTATGCGGCGACCGAGTTGC-3′
5′-GCGAACGTGGCGAGAAAGGAAG-3′

3.9.10. Production of 545 bp RNA duplex template

An RNA sequence of 545 bp is too long to be produced by standard solid state nucleic acid synthesis, and so it was transcribed from a DNA plasmid construct. The 545 bp RNA sequence (r545-mer) was produced in two steps: 1) engineering two DNA plasmid

constructs to produce the sense and anti-sense r545 strands when transcribed, and 2) co-transcription of both DNA plasmid constructs to produce the r545 duplex.

Engineering DNA plasmid constructs required inserting a T7 promoter upstream of the 545 bp sequence to be transcribed, as well as restriction enzyme sites for EcoRI and BamHI. Separate constructs were engineered to produce the sense (S-r545) and the anti sense (AS-r545) RNA strands. The T7 promoter and restriction enzyme sequences were introduced in the 5'- and 3'- extremes using PCR with the following primers:

To produce AS-r545:
Forward 5′–
GGGGAATTCTAATACGACTCACTATAG**GGGAATAGTGTATGCGGCGACCGAGTTGC**-3′
Reverse 5′–GGGGGATCC**GCGAACGTGGCGAGAAAGGAAG**-3′

To produce S-r545:

Forward 5′–GGGGAATTC**GAATAGTGTATGCGGCGACCGAGTTGC**–3′
Reverse 5′–
GGGGGATCCTAATACGACTCACTATAG**GGGCGAACGTGGCGAGAAAGGAAG**–3′

Color code: T7 promoter, EcoRI, BamHI and **primer** sequences.

After PCR, incorporation of the T7 promoter and restriction enzyme sequences was confirmed by agarose gel electrophoresis. The sequence of each construct was confirmed by sequencing analysis. Both constructs where then co-digested with EcoRI-HF and BamHI-HF (New England Biolabs) and then inserted/ligated independently in the plasmid pUK21-NotI (GeneBank code: AF324726) after pUK21-NotI's initial T7 promoter sequence was scrambled. The S-r545 and AS-r545 plasmids were transformed in DH5α cells and then grown in the presence of kanamycin for plasmid production. The plasmid was extracted and purified using standard protocols for large-scale preparation of plasmid. The purified plasmid was then digested with proteinase K to remove RNase,

followed by phenol/chloroform purification to eliminate RNase activity. Finally the plasmids were linearized with EcoRI or BamHI. Completeness of the linearization reaction was confirmed by agarose gel electrophoresis.

The production of the double stranded r545 was achieved by co-transcription of both the sense and anti-sense constructs in a one-pot reaction. 750 ng each of the sense and anti-sense plasmid constructs were transcribed using the HiScribe T7 High Yield RNA Synthesis Kit (New England Biolabs). DNA templates were then digested with Turbo DNase (ThermoFisher Scientific) following the manufacturer's protocol. Ethanol precipitation was performed on each co-transcription reaction, and RNA was then purified using a NAP-5 column (GE Healthcare). Then ds r545-mer was purified using long sequencing PAGE in native conditions (0.5x TBE, 7.5 % acrylamide). The r545-mer band was cut and extracted from the gel using standard protocols. The r545-mer was further purified with anion-exchange chromatography (DEAE Sephadex A-25, Sigma) to remove residual acrylamide from the sample and desalted on a NAP-5 column (GE Healthcare).

3.9.11. RNA ligation procedure

To ligate RNA oligonucleotides, we utilized T4 RNA ligase 2, which efficiently repairs nicks in double stranded RNA. Samples were prepared in 2 mg of the appropriate viscous solvent containing 300 ng of the RNA template duplex and each of the 11 oligonucleotides (or a specific subset of the 11) in the appropriate molar ratio. RNA samples were heated to 95 °C for 3 minutes and cooled to 20 °C at a rate of 4 °C/min on a thermal cycler. After thermal cycling, samples were diluted with water, T4 RNA ligase 2 buffer, and 10 U of T4 RNA ligase 2 (New England Biolabs) to a final solution

concentration of 10% or 15% glycholine (wt/wt). A concentration of 10% solvent (wt/wt) was determined to be the limit above which T4 RNA ligase 2 exhibits significantly reduced activity compared with standard buffer conditions, for the solvents glycholine, reline, and glycerol (Figure 3.19). The ligation reaction was incubated at 37 °C for 1 hour before samples were mixed with 2x RNA loading dye (New England Biolabs) for analysis by gel electrophoresis.

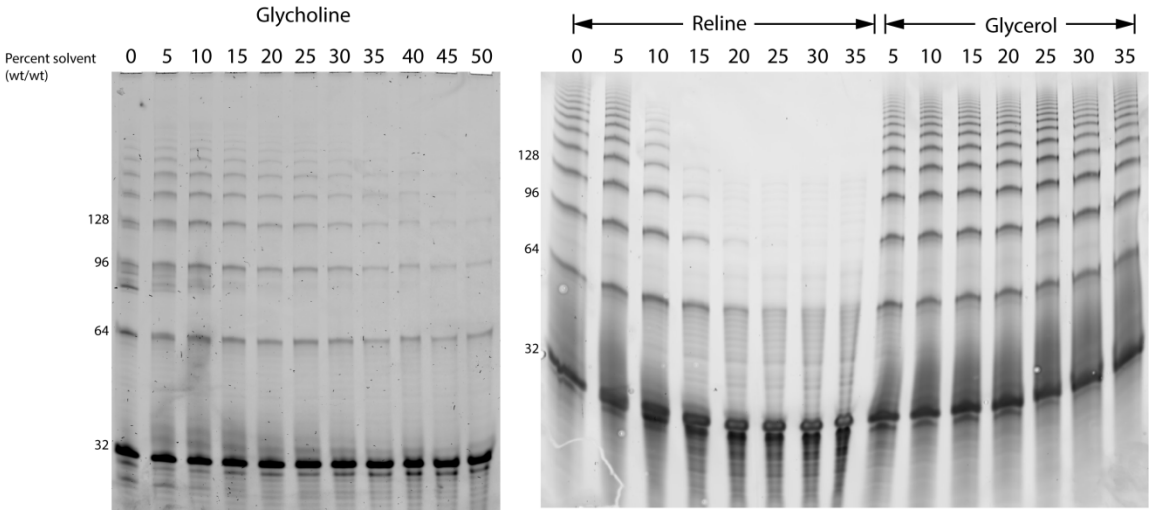

Figure 3.19. Denaturing polyacrylamide gels (10%) showing the results of using T4 RNA ligase 2 to ligate a system of half-complementary tiling 32-mers in different aqueous mixtures of glycholine, reline, and glycerol (0.1 M NaCl). The RNA sequence contains a 5'-phosphate, allowing it to be ligated to its upstream neighbor and generating strands that are 32, 64, 96…etc bp in length. Sequences of the 32-mers are: (1) DNA: 5'-CTTCAGCATCTTTTACTTTCACCAGCGTTTCT-3'; (2) RNA: 5'-phos-GUAAAAGAUGCTGAAGAUCAGUUGGGUGCACG-3'.

3.9.12. Thermal denaturation studies in glycerol

UV absorbance was used to monitor the thermal denaturation of the 32 bp duplex—described in Section 2.9.2. DNA sequences—in glycerol containing 20 mM Tris pH8 and 0.1 M NaCl. The thermal stability of the 32 bp duplex in reline[11] and glycholine[11] has

already been characterized. UV measurements were performed in a 1 mm quartz cuvette in a temperature-controlled UV-Vis spectrophotometer (Agilent 8453). To determine T_m values, heating and cooling traces were generated for each sample by recording spectra from 220-400 nm, at temperatures between 0 to 100 °C in intervals of 1 °C. Melting curves were generated using the absorbance values at 260 nm (Figure 2.9). T_m values were determined using the method described by Mergny & Lacroix[18].

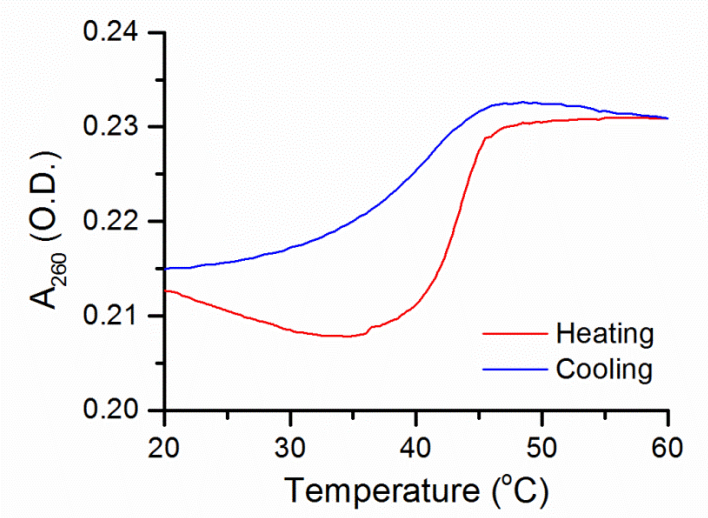

Figure 3.20. Thermal denaturation of a 32 bp duplex in glycerol (0.1 M NaCl).

3.10. References

1. Mills DR, Peterson RL, Spiegelman S. An Extracellular Darwinian Experiment with a Self-duplicating Nucleic Acid Molecule. *Proceedings of the National Academy of Sciences* 1967, **58**(1)**:** 217-224.

2. Kreysing M, Keil L, Lanzmich S, Braun D. Heat flux across an open pore enables the continuous replication and selection of oligonucleotides towards increasing length. *Nat Chem* 2015, **7**(3)**:** 203-208.

3. Matsumura S, Kun Á, Ryckelynck M, Coldren F, Szilágyi A, Jossinet F, *et al.* Transient compartmentalization of RNA replicators prevents extinction due to parasites. *Science* 2016, **354**(6317)**:** 1293-1296.

4. Wochner A, Attwater J, Coulson A, Holliger P. Ribozyme-Catalyzed Transcription of an Active Ribozyme. *Science* 2011, **332**(6026)**:** 209-212.

5. Kaiser RI, Maity S, Jones BM. Synthesis of Prebiotic Glycerol in Interstellar Ices. *Angewandte Chemie* 2014, **127**(1)**:** 197–202.

6. Hud Nicholas V, Cafferty Brian J, Krishnamurthy R, Williams Loren D. The Origin of RNA and "My Grandfather's Axe". *Chemistry & Biology*, **20**(4)**:** 466-474.

7. Engelhart AE, Hud NV. Primitive Genetic Polymers. *Cold Spring Harbor Perspect Biol* 2010, **2**(12)**:** 21.

8. Tawa K, Knoll W. Mismatching base-pair dependence of the kinetics of DNA–DNA hybridization studied by surface plasmon fluorescence spectroscopy. *Nucleic Acids Research* 2004, **32**(8)**:** 2372-2377.

9. Zhou H-X, Rivas G, Minton AP. Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annual Review of Biophysics* 2008, **37:** 375-397.

10. Mamajanov I, Engelhart AE, Bean HD, Hud NV. DNA and RNA in Anhydrous Media: Duplex, Triplex, and G-Quadruplex Secondary Structures in a Deep Eutectic Solvent. *Angewandte Chemie International Edition* 2010, **49**(36)**:** 6310-6314.

11. Gállego I, Grover MA, Hud NV. Folding and Imaging of DNA Nanostructures in Anhydrous and Hydrated Deep-Eutectic Solvents. *Angewandte Chemie International Edition* 2015, **54**(23)**:** 6765-6769.

12. Abbott AP, Capper G, Davies DL, Rasheed RK, Tambyrajah V. Novel solvent properties of choline chloride/urea mixtures. *Chem Commun* 2003(1)**:** 70-71.

13. Portella G, Germann MW, Hud NV, Orozco M. MD and NMR Analyses of Choline and TMA Binding to Duplex DNA: On the Origins of Aberrant Sequence-Dependent Stability by Alkyl Cations in Aqueous and Water-Free Solvents. *Journal of the American Chemical Society* 2014, **136**(8)**:** 3075-3086.

14. Ferris JP, Jr. ARH, Liu R, Orgel LE. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* 1996, **381:** 59-61.

15. Lesnik EA, Freier SM. Relative Thermodynamic Stability of DNA, RNA, and DNA:RNA Hybrid Duplexes: Relationship with Base Composition and Structure. *Biochemistry* 1995, **34**(34)**:** 10807-10815.

16. Lohman GJS, Zhang Y, Zhelkovsky AM, Cantor EJ, Evans TC. Efficient DNA ligation in DNA–RNA hybrid helices by Chlorella virus DNA ligase. *Nucleic Acids Research* 2013, **42**(3)**:** 1831-1844.

17. Jin J, Vaud S, Zhelkovsky AM, Posfai J, McReynolds LA. Sensitive and specific miRNA detection method using SplintR Ligase. *Nucleic Acids Research* 2016, **44**(13)**:** e116.

18. Mergny JL, Lacroix L. Analysis of thermal melting curves. *Oligonucleotides* 2003, **13**(6)**:** 515-537.

CHAPTER 4 : VISCOSITY-MEDIATED REPLICATION OF AN RNA DUPLEX CONTAINING A RIBOZYME

4.1. Introduction

While significant progress has been made in recent years towards elucidating the prebiotic synthesis of ribonucleotides and RNA polymers[1-5], we still have a poor understanding of how the next conceptual stage in the RNA World—template-directed replication of RNA—emerged. The demonstration of an RNA system that can undergo sustained cycles of enzyme-free, template-directed replication has been an important goal for decades, but remains frustratingly out of reach[6]. Existing studies of enzyme-free copying of RNA templates focus on single-stranded templates[7, 8], and rarely explicitly address the challenges associated with multiple rounds of information transfer from a nucleic acid sequence (such as strand inhibition).

A complete replication cycle, required for a viable genetic system capable of evolution, involves reciprocal synthesis of both strands of a template duplex, producing a copy of the original duplex. In the previous chapter I have shown that viscosity enables synthesis of a complementary strand from one strand of an RNA duplex, using ligation of oligonucleotides assembled on the template strand[9]. In this chapter I focus on building a complete replication cycle, utilizing viscosity to promote simultaneous assembly of oligonucleotide substrates on both strands of an RNA duplex. In doing so, I explore the challenge presented by association of oligonucleotide substrates to each other—in what may be described as a "mini strand inhibition" problem—rather than to their target sites on the template strands.

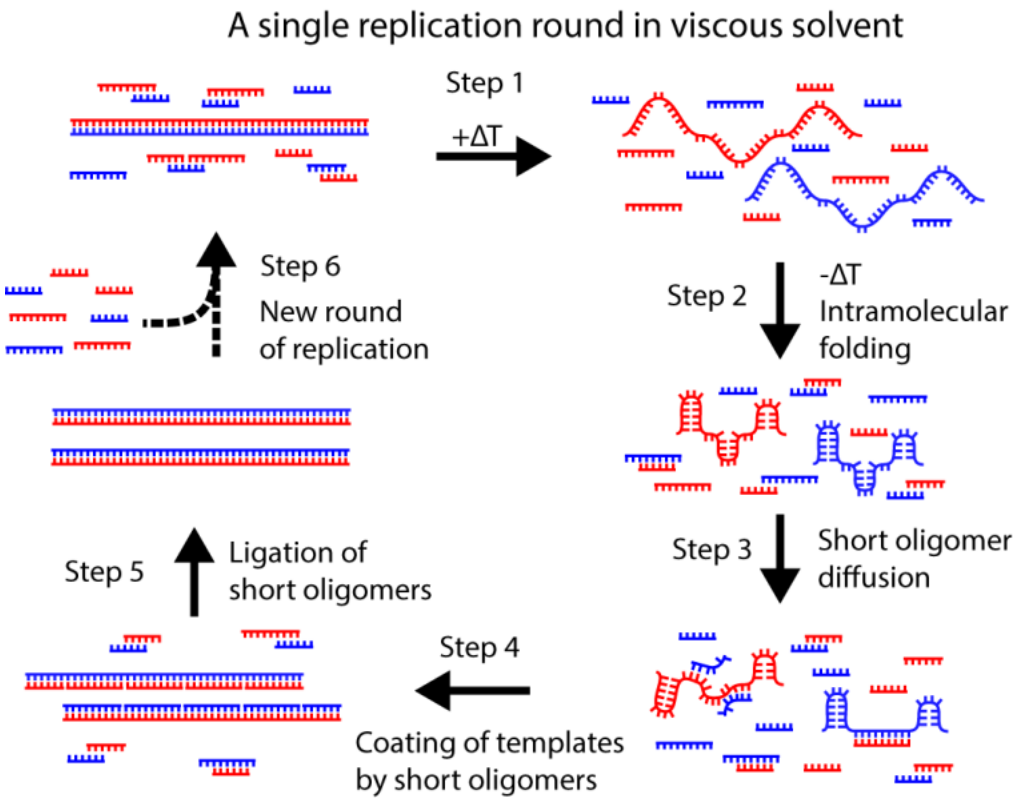

Figure 4.1. Replication of a nucleic acid duplex template in viscous solvent. After heating and cooling the duplex (steps 1 and 2), template strands become kinetically trapped as single strands. Oligonucleotide substrates associate both with each other (a "mini strand inhibition" problem) and with their targets on the template single strands (step 3). Once oligonucleotide substrates coat the entire strand (step 4) and are ligated together (step 5), a copy duplex is formed.

In addition to exploring replication of a template duplex, I aimed to explore another important feature of the RNA World model: the emergence of catalytically active RNA sequences, or ribozymes. Though the basis of the RNA World model is RNA's dual ability to transfer sequence information and catalyze reactions, these functions may be considered incompatible on a structural level. RNA sequences that are stably folded—a requirement for catalytic activity—are less accessible to mono- or oligonucleotide substrates, and therefore may serve as poor templates for replication[10]. Polymerase ribozymes generated through *in vitro* evolution are strongly biased towards the copying

of templates with no secondary structure, implying that secondary structure generally obstructs RNA-catalyzed RNA copying[11]. Viscous environments present a potential solution to this "replicator-catalyst" paradox, by promoting the kinetic trapping of nucleic acid sequences with a high degree of intramolecular structure (e.g. hairpins). Thus, in a viscous environment, stable intramolecular structure enhances a sequence's potential for catalytic activity *and* its propensity to be replicated. Viscosity provides a viable mechanism for the selective trapping and amplification of ribozyme sequences from a complex pool of sequences.

To illustrate the potential of viscous environments to promote the kinetic trapping of ribozymes, I demonstrate here the replication of a gene-length (>600 bp) RNA duplex containing a hammerhead ribozyme motif (43 nt). The hammerhead ribozyme is a naturally occurring, self-cleaving ribozyme that is extensively used as a model system for investigating RNA structure and function. I aimed to show that viscosity can promote information transfer from an RNA sequence to a complementary strand containing an active hammerhead ribozyme (Figure 4.2, step 1), resulting in a duplex which can be undergo a complete round of replication (Figure 4.2, step 2).

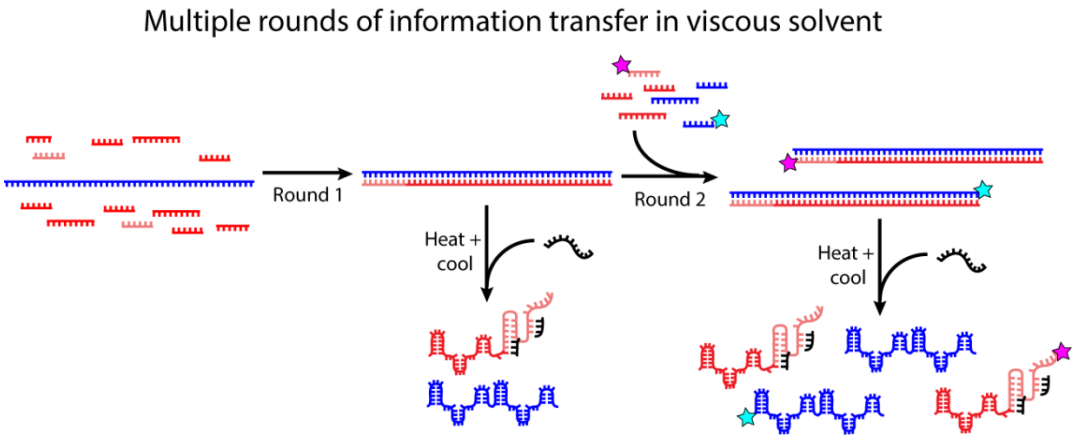

Figure 4.2. Plan for demonstrating several rounds of information transfer from an RNA duplex containing a ribozyme sequence. In this schematic, the red template strand and oligonucleotides correspond to sequences from the sense strand; the blue template strand and oligonucleotides correspond to sequences from the anti-sense strand; the pink region of the red/sense template strand corresponds to the HHR motif; and the black oligonucleotide is the HHR substrate, which binds to and is cleaved by the HHR motif. In round 1 of viscosity-enabled copying, an RNA template strand serves as a template for the assembly of complementary oligonucleotides and synthesis of its complementary strand, which contains the HHR enzyme sequence. After template-directed synthesis, the resulting RNA duplex is thermally cycled and incubated with the HHR substrate strand, to demonstrate that the newly synthesized RNA strand is catalytically active. In round 2 of viscosity-enabled copying, the RNA duplex serves as a template for the simultaneous assembly of RNA oligonucleotides on both strands, and a copy duplex is formed after ligation. Labeling of oligonucleotides in each set by fluorophores enables the visualization of each newly synthesized strand.

4.2. Promoting template-directed oligonucleotide assembly on a duplex with tiled oligonucleotide sets

While we have previously shown that viscosity enables information transfer from a single strand of a template duplex, building a system capable of multiple replication rounds requires the ability to transfer information from both strands of a template duplex. The use of oligonucleotides as substrates for duplex replication presents a conceptual “mini” strand inhibition problem: substrates for one template strand will necessarily be complementary to the substrates for the other template strand. When these two sets of complementary oligonucleotides are present in solution, it is kinetically favorable for the oligonucleotides to hybridize to each other rather than to their target sites on the much longer template strands. Since short oligonucleotides are far more mobile than long template strands in viscous solution, and kinetically trapped intramolecular structures on the template strands may present a barrier to binding of oligonucleotide targets, it is entirely possible that oligonucleotide-oligonucleotide duplex formation will kinetically “out-compete” the binding of oligonucleotides to template strands (Figure 4.3).

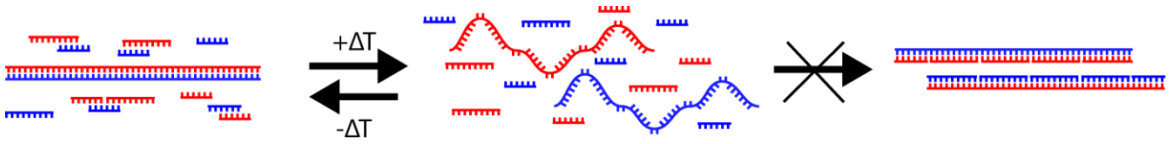

Figure 4.3. Illustration of potential competition between template-directed oligonucleotide assembly and hybridization of complementary oligonucleotides in viscosity-mediated replication.

Despite the kinetic favorability of oligonucleotides binding to other oligonucleotides rather than to their template strands, template binding may still occur if it is thermodynamically favored. For example, consider two RNA oligonucleotides that are only partially complementary to each other but fully complementary to target sites on long RNA template strands. Though it is kinetically favorable for the RNA oligonucleotides to hybridize with each other, the thermodynamically favored state is association of the RNA oligonucleotides with their target sites on the template strands. To test template-directed oligonucleotide assembly in such a system, we investigated a model system consisting of a 3 kb DNA duplex template and two sets of DNA oligonucleotides (32 nt each), designed to be complementary to a 352 nt region of the anti-sense (*L5-R5*) and the sense (*L5'-R5'*) template strands (Figure 4.4). The sequences of the two oligonucleotide sets are not perfectly complementary but offset from each other by 16 nt; i.e. *L5* forms a 16 bp duplex with *L5'*, with 8 nt single stranded overhangs at the 5' and 3' ends of the duplex. While *L5* may find and hybridize with *L5'* quickly in solution, the 8 nt single stranded overhangs provide a toehold for strand displacement of the *L5/L5'* complex by the template strands, forming a 32 bp duplex between *L5* or *L5'* and their respective template target sites.

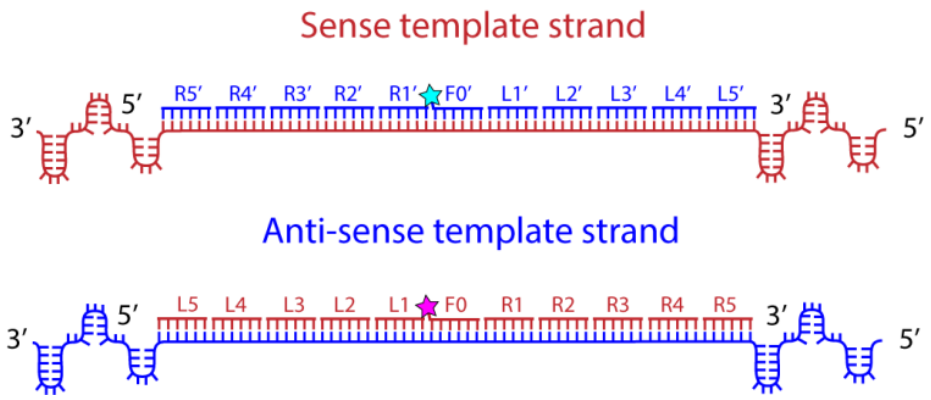

Figure 4.4. Illustration of tiled oligonucleotide sets *L5-R5* and *L5'-R5'* assembled on their respective target 3 kb DNA template strands. Sequences in each set are offset from the corresponding sequence in the other set by 16 nt; i.e. *L5* and *L5'* form a 16 bp duplex with 16 nt single stranded regions on each end. Blue star indicates a 5'-Cy5 dye; magenta star indicates a 5'-Cy3 dye.

To test the extent of template-directed oligonucleotide assembly in this system, the central oligonucleotides in each set, *F0* and *F0'*, were tagged at their 5' ends with Cy3 and Cy5 dyes, respectively, to enable tracking via fluorescence. Glycholine samples containing the 3 kb DNA duplex, oligonucleotides *L2-R2*, and oligonucleotides *L2'-R2'* (20:1 molar ratio of each oligonucleotide to the template) were thermally cycled (using conditions described in Section 2.5. Step 3: Oligonucleotide binding to kinetically trapped templates) and the hybridization behavior of *F0* and *F0'* was then monitored by agarose gel electrophoretic mobility. In Figure 4.5a, the Cy3 channel image shows the position of *F0* and reports on the binding of oligonucleotides *L5-R5* to the anti-sense template strand; the Cy5 channel reports on the position of *F0'* and reports on the binding of *L5'-R5'* to the sense template strand. Reformation of the 3 kb duplex over time was followed by the increase in the intensity of the less mobile duplex band after ethidium

bromide (EtBr) staining, and corresponding decrease in the intensity of the more mobile ssDNA band (Figure 4.5a, EtBr channel).

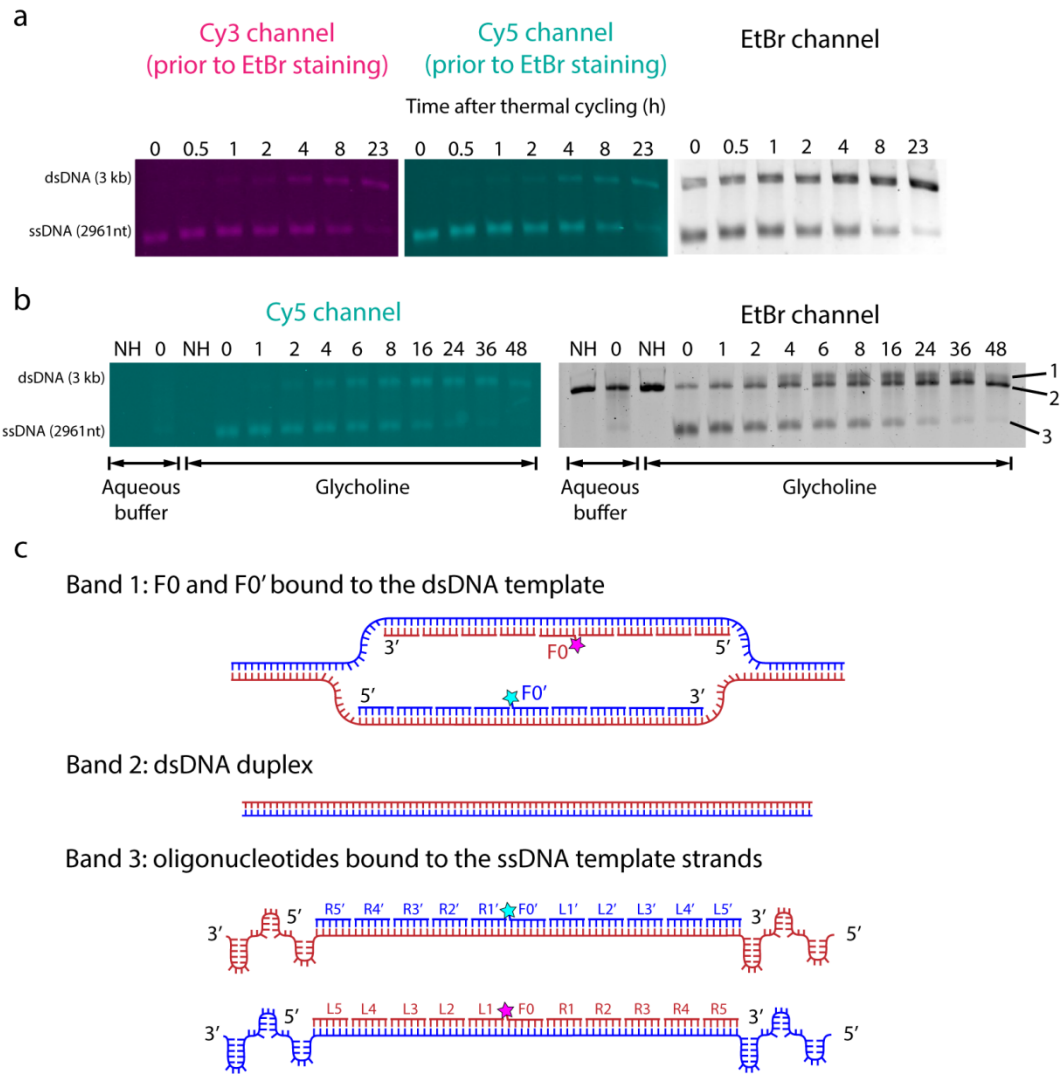

Figure 4.5. Kinetics of oligonucleotide binding (*L5-R5* and *L5'-R5')* to a denatured 3 kb DNA duplex.Sequences from the sense template strand are colored red, while sequences from the anti-sense template strand are colored blue. **a**, Agarose gel (2%) showing the results of thermal cycling DNA oligonucleotides *L2*through *R2* and *L2'* through *R2'* with 3 kb duplex DNA in glycholine. The gel was imaged for Cy3 and Cy5 fluorescence signal before staining with ethidium bromide (EtBr) and imaging for EtBr signal. EtBr image shows the kinetics of 3 kb DNA annealing, while Cy3 and Cy5 images show that both sets of oligonucleotides are binding to their respective targets on both the ssDNA and dsDNA template. **b**, Agarose gel (2%) showing the results of thermal cycling DNA oligonucleotides *L5* through *R5* and *L5'* through *R5'* with 3 kb duplex DNA in glycholine. **c**, Schematic of the anticipated oligonucleotide assembly on the ds and ssDNA templates.

Comparison of the Cy3, Cy5, and SYBR Gold images show that both sets of oligonucleotides—*L2-R2* and *L2'-R2'*—are binding to their target sites on the ssDNA templates (Figure 4.5a, more mobile band). Surprisingly, comparison of these images also reveals that both sets of oligonucleotides are bound to the dsDNA template as well (Figure 4.5a, less mobile band). When a larger set of oligonucleotides—*L5-R5* and *L5-R5'*—is thermally cycled with the 3 kb DNA duplex in glycholine, we observe two bands in the EtBr channel that migrate with mobilities similar to the dsDNA template (Figure 4.5b, EtBr channel, bands labeled 1 and 2). Direct comparison of the Cy5 and EtBr images show that the oligonucleotides are bound to the upper of these two bands (band 1). These results indicate that band 1 consists of the ds 3 kb DNA duplex which is not fully double stranded, but rather contains a bubble where both oligonucleotide sets are attached (Figure 4.5c). We propose that this double stranded bubble can form in two ways: both sets of oligonucleotides are binding to their kinetically trapped single strands, and remaining bound while the remainder of the template strands re-anneal; or the oligonucleotides are binding directly to the ds DNA duplex during periods of duplex "breathing," when the target region is transiently single stranded[12-14]. After thermal cycling in aqueous buffer, there is no observable binding of either fluorescently tagged oligonucleotide to the ss or dsDNA template (Figure 4.5b, aqueous buffer lanes). These results provide a proof of concept that glycholine enables the simultaneous assembly of oligonucleotides on both strands of a gene-length duplex template.

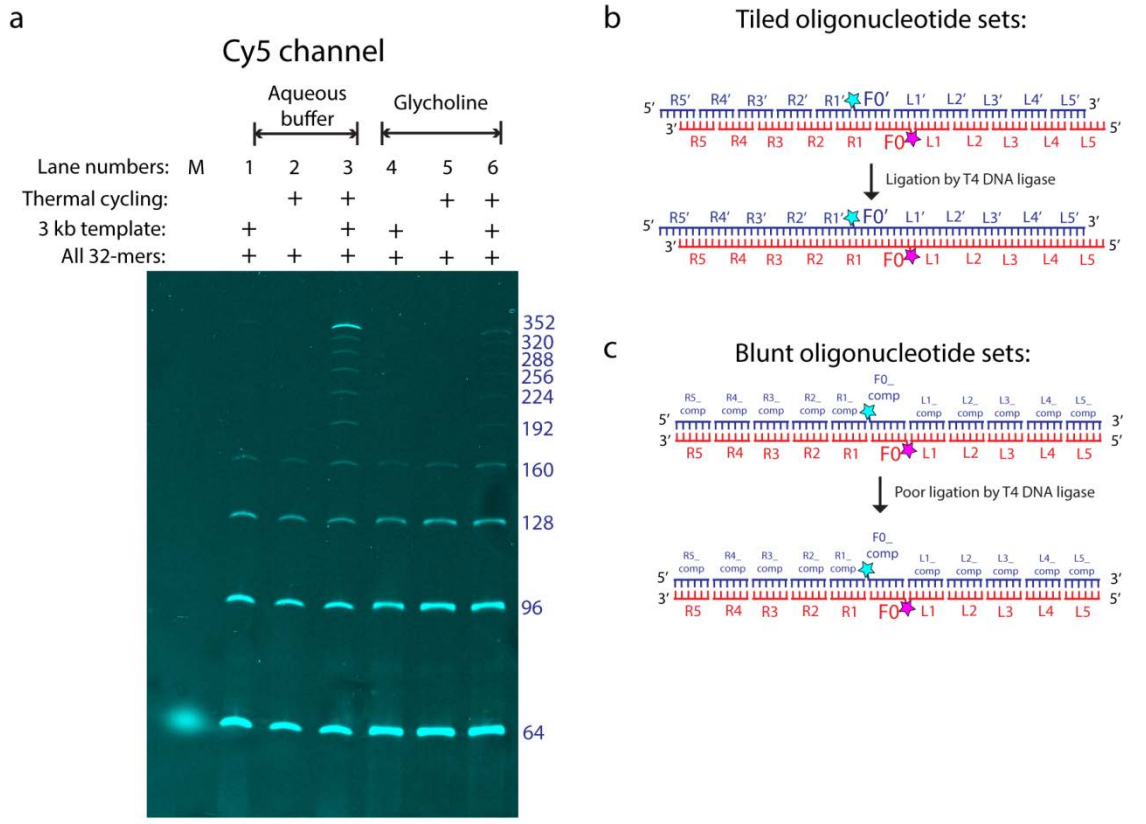

Figure 4.6. Analysis of oligonucleotide binding (*L5-R5* and *L5'-R5'*) to the 3 kb DNA duplex template. **a,** Denaturing polyacrylamide gel (10%, 8 M urea) showing the results of thermal cycling and ligating a mixture of 3 kb DNA duplex with oligonucleotides *L5-R5* and *L5'-R5'* . **b,** Schematic illustrating the association of tiled oligonucleotide sets *L5-R5* and *L5'-R5'* with each other. **C,** Schematic illustrating the association of blunt (perfectly complementary) oligonucleotide sets *L5-R5* and *L5_comp-R5-comp* with each other.

The final step in demonstrating copying from a gene-length template duplex was ligation of the oligonucleotides after assembly on the template. Samples containing the 3 kb DNA duplex, oligonucleotides *L5-R5*, and oligonucleotides *L5'-R5'* were thermal cycled and incubated with T4 DNA ligase (procedure described in Section 2.9.7. DNA ligation procedure). Analysis by denaturing polyacrylamide gel electrophoresis indicates that thermal cycling in glycholine promotes the assembly of all 11 oligonucleotides on each template strand (Figure 4.6a, lane 6). However, 11 product strands are also observed

when 3 kb template and oligonucleotides are thermal cycled in aqueous buffer (Figure 4.6a, lane 3). Additionally, product strands up to 160 nt in length are observed in all conditions, even with no heating and in the absence of 3 kb template (Figure 4.6a, lanes 1-2 and 4-5). These results indicate that the oligonucleotide sets are binding to each other rather than to their target sites on the template strand, creating a long chain of base paired oligonucleotides—a tiling system—that can then be ligated by T4 DNA ligase, which repairs nicks in double stranded DNA (Figure 4.6b). The more intense fluorescence signal from product strands in aqueous buffer compared to glycholine (Figure 4.6a, lane 3 vs. 6) is explained by the higher thermal stability of duplexes in aqueous buffer compared to glycholine. Tiling of the oligonucleotides as shown in requires formation of 16 bp duplexes, which are stable at room temperature in aqueous buffer but are minimally stable in glycholine (as shown in Figure 3.10).

4.3. Template-directed assembly of blunt oligonucleotide sets on a duplex template

While the data in Figure 4.6a definitively shows that thermal cycling in glycholine promotes reciprocal copying from both strands of a 3 kb DNA duplex, analysis of the product strands is complicated by the presence of strands synthesized by non-templated assembly, i.e. tiling of complementary oligonucleotides with each other (Figure 4.6b). To improve our “signal to noise” ratio, we have chosen to test copying with two sets of oligonucleotides that are perfectly complementary to each other (Figure 4.6c). Rather than forming a tiled system of 16 bp duplexes, fully complementary oligonucleotides will form 32 bp blunt end duplexes (no single stranded overhangs at the ends). Compared to its activity for repairing nicks in double stranded DNA, T4 DNA ligase has a lower activity for ligation of blunt ended duplexes, exhibiting almost no blunt end ligation

activity in 15% glycholine (wt/wt), the conditions during our ligation reaction (Figure 4.7).

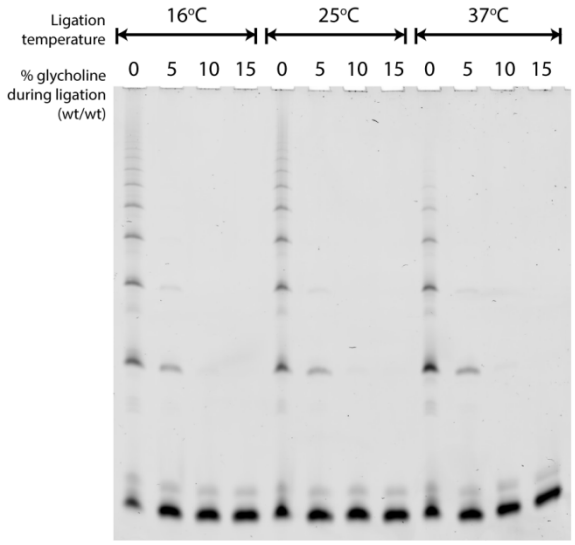

Figure 4.7. Denaturing polyacrylamide gel (10%, 8 M urea) showing the results of incubating two complementary 32 nt DNA sequences—*F0* and *F0_comp*—with T4 DNA ligase at different temperatures, and in different amounts of glycholine. Gel was stained with SYBR Gold to visualize the bands.

From these results, we can conclude that after thermal cycling a 3 kb template with two sets of fully complementary oligonucleotides, any blunt end duplexes will not be ligated together by T4 DNA ligase and will migrate as 32 nt single strands on a denaturing gel. Therefore, we propose that using perfectly complementary (blunt) oligonucleotide sets will enable us to visualize only newly synthesized strands produced through template-directed assembly of oligonucleotides. To test this proposal, I utilized two oligonucleotide sets (eleven oligonucleotides each): *L5-R5* and *L5_comp-R5_comp*, where each 32 nt sequence is perfectly complementary to its corresponding sequence in the other set.

Prior to testing oligonucleotide assembly on both strands of the 3 kb DNA template, I first analyzed the binding kinetics of each individual oligonucleotide set.

Glycholine samples containing the 3 kb DNA duplex and either *L5-R5* (with *F0* fluorescently tagged at its 5' end with Cy3) or *L5_comp-R5_comp* (with *F0_comp* fluorescently tagged at its 5' end with Cy5) were thermal cycled, and the kinetics of binding were tracked by agarose gel electrophoresis (Figure 4.8a, b). We see that each set of oligonucleotides is binding strongly to their targets on the ssDNA template, with approximately all *F0* and *F0_comp* target sites occupied at the peak of each binding reaction (Figure 4.8c; maximum theoretical binding occurs at 5% total intensity in the bound band, based upon a 20:1 molar ratio of oligonucleotide to template). There appears to be a difference in the binding efficiencies of *F0* versus *F0_comp*, perhaps indicating that slightly different intramolecular structures are formed by each strand of the DNA template, posing a different barrier to binding for each oligonucleotide set. In agarose gel analysis of both oligonucleotide sets, two bands appear at the position of the kinetically trapped ssDNA, indicating that template-directed assembly of each oligonucleotide set is opening up intramolecular structure on the DNA strands, causing a shift in gel electrophoretic mobility.

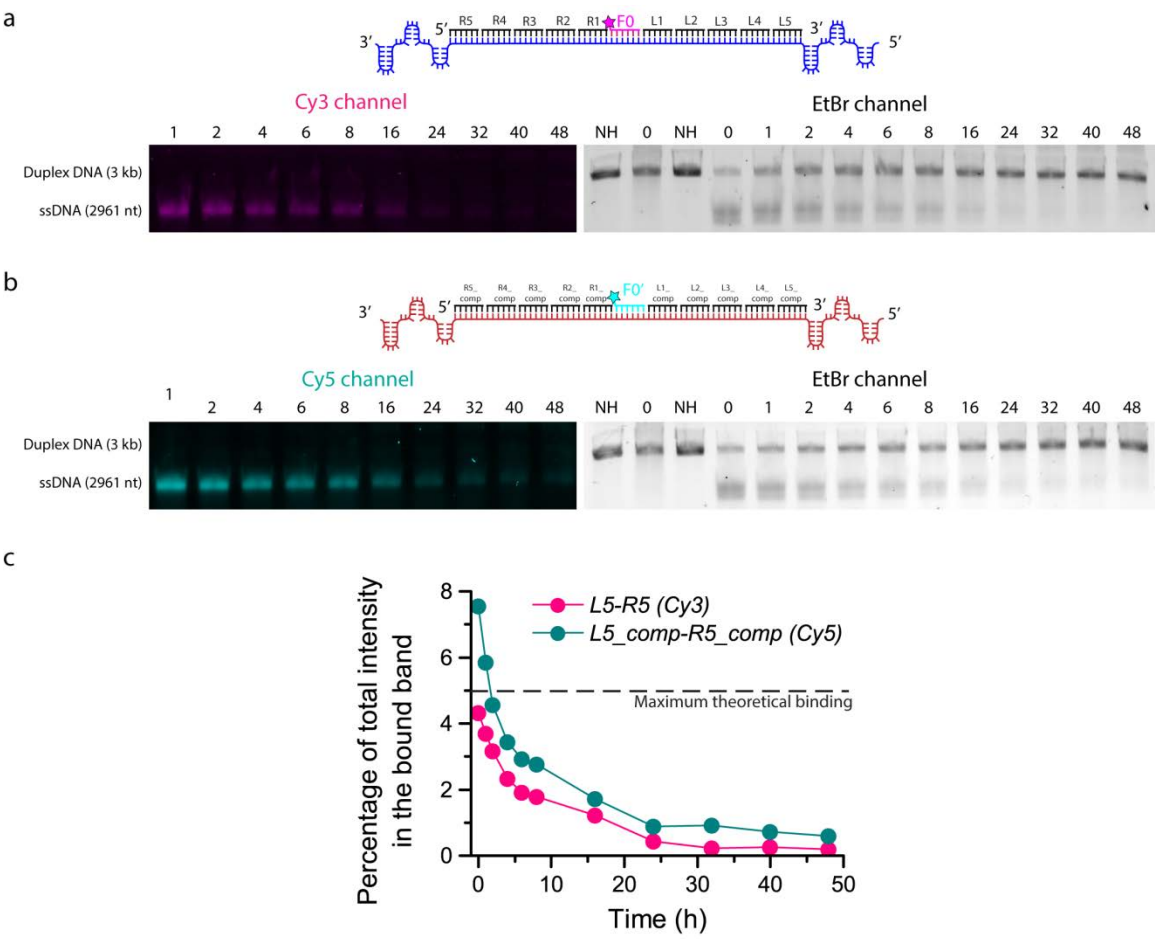

Figure 4.8. Kinetics of oligonucleotide binding (*L5-R5* or *L5_comp-R5_comp*) to a denatured 3 kb DNA duplex. **a**, Agarose gel (2%) showing the results of thermal cycling DNA oligonucleotides *L5* through *R5* (*F0* is fluorescently tagged with Cy3) with 3 kb duplex DNA in glycholine. The gel was imaged for Cy3 fluorescence signal before staining with ethidium bromide (EtBr) and imaging for EtBr signal. **b**, Agarose gel (2%) showing the results of thermal cycling DNA oligonucleotides *L5_comp* through *R5_comp* (*F0_comp* is fluorescently tagged with Cy5) with 3 kb duplex DNA in glycholine. The gel was imaged for Cy5 fluorescence signal before staining with EtBr and imaging for EtBr signal. **c**, Extent of *F0* and *F0_comp* binding over time, obtained from densitometry analysis of the gels in **a** and **b.**

When a glycholine sample containing both oligonucleotide sets and the 3 kb DNA duplex is thermal cycled, agarose gel electrophoresis shows simultaneous assembly of both oligonucleotide sets on the DNA template (Figure 4.9). As observed previously with the tiled oligonucleotide sets (Figure 4.5), the presence of two complementary

oligonucleotide sets results in oligonucleotide binding not only to the kinetically trapped ssDNA template but also to the DNA template duplex, forming a “bubble” in the duplex.

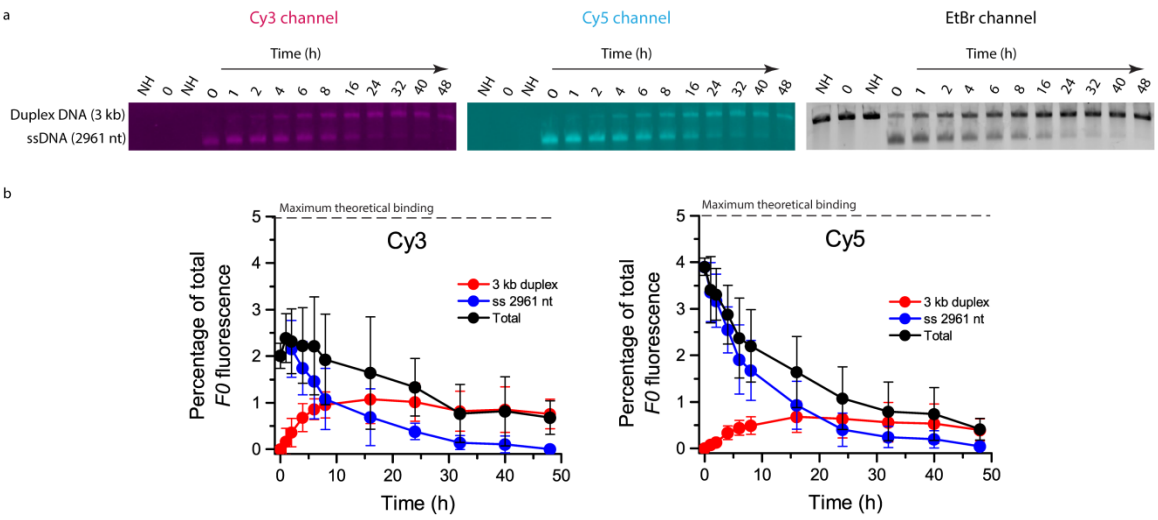

Figure 4.9. Kinetics of oligonucleotide binding (both *L5-R5* and *L5_comp-R5_comp)* to a denatured 3 kb DNA duplex. **a**, Agarose gel (2%) showing the results of thermal cycling DNA oligonucleotides *L5-R5* (*F0* is fluorescently tagged with Cy3) and *L5_comp-R5_comp* (*F0_comp* is fluorescently tagged with Cy5) with 3 kb duplex DNA in glycholine. The gel was imaged for Cy3 and Cy5 fluorescence signal before staining with ethidium bromide (EtBr) and imaging for EtBr signal. **b**, Extent of *F0* and *F0_comp* binding over time, obtained from densitometry analysis of the gel in **a**. Maximum theoretical binding at 5% total fluorescence intensity is based off stoichiometric considerations (20:1 molar ratio of each oligonucleotide to the template duplex).

Quantification of *F0* and *F0_comp* binding by densitometry analysis indicates that each set of oligonucleotides is binding to their target strands with significantly different kinetics and to different extents (Figure 4.9b). These results point to the possibility that the two complementary template strands may be forming different intramolecular structures upon kinetic trapping in glycholine. The extent of *F0* binding is significantly less than that of *F0_comp* binding (50% and 80% target site occupancy at their respective peaks), indicating that the anti-sense template strand may be forming more stable

intramolecular structure—that is less easily opened by invading oligonucleotide substrates—than the sense template strand. Additionally, the extent of both *F0* and *F0_comp* binding when both sets of oligonucleotides are present is significantly lower than the extent of binding when only one oligonucleotide set is present (Figure 4.8; approximately 100% target site occupancy by both *F0* and *F0_comp* at the peak of binding), indicating that the oligonucleotides are binding to each other in solution to form blunt end duplexes. These results show that despite oligonucleotide association competing with template-directed oligonucleotide assembly, both oligonucleotide sets are still able to simultaneously assemble on their respective target strands in glycholine.

4.4. Testing the influence of glycholine and Mg^{2+} on HHR cleavage

After demonstrating viscosity-mediated replication of a duplex sequence using a model DNA system, our next step is to extend our results to a more prebiotically relevant RNA system. Specifically, our aim is to show that viscosity can promote the replication of a gene-length, RNA duplex sequence containing a catalytically active RNA motif. The ribozyme sequence we chose to incorporate into our RNA template duplex is the hammerhead ribozyme (HHR), a small, self-cleaving RNA motif that is found in the genomes of many species and used widely as a model system for studying RNA structure and properties[15]. The self-cleavage reaction is highly site-specific, and occurs by nucleophilic attack of the 2'-oxygen of nucleotide C17 (indicated by an arrow in Figure 4.10, left image) on its adjacent phosphate, resulting in a 2',3'-cyclic phosphate and 5'-OH[16]. While the HHR is an intramolecular structure, *in vitro* studies often divide the HHR sequence into an enzyme strand (containing the 13 nt conserved core essential for catalytic activity) and substrate strand (containing nucleotide C17) to enable multiple

turnovers of the cleavage reaction. We have chosen to use separate enzyme and substrate strands, so that initiation of the cleavage reaction can be controlled by addition of the substrate strand and the required concentration of monovalent/divalent cations[17]. If we inserted the full intramolecular HHR sequence into our RNA template, cleavage would occur during transcription of the template by T7 RNA polymerase, which utilizes mM concentrations of Mg^{2+}.

The specific HHR sequences utilized in this study are the "full length" sequences from *Schistosoma mansoni*, which contains tertiary contacts that are absent in "minimal" HHR sequences and are responsible for enhancing cleavage rates by a factor of almost 1000[16]. The full length HHR sequence consists of a 43 nt enzyme strand and a 20 nt substrate strand that is cleaved into 14 nt and 6 nt. Base pairing between the enzyme and substrate strands as well as the 3D structure of the full length HHR sequences are shown in Figure 4.10.

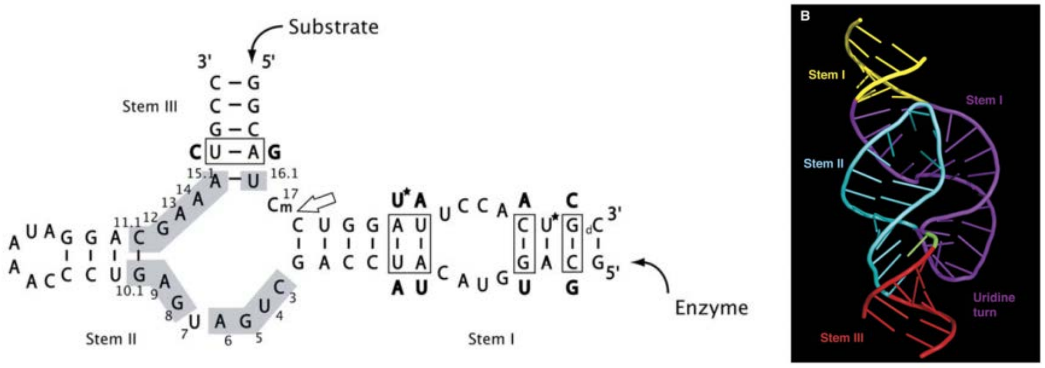

Figure 4.10. Structure of the full length *Schistosoma mansoni* hammerhead ribozyme sequence. Left: base pairing contacts between the enzyme and substrate strand. Grey background indicates the conserved catalytic core; boxes indicate base pairs that were switched to aid crystallization; large arrow points to the scissile bond. Images are taken from Figure 1B and 1C of: Martick M, Scott WG. "Tertiary Contacts Distant from the Active Site Prime a Ribozyme for Catalysis." *Cell* 2006, 126(2): 309-320.

The HHR cleavage reaction is highly dependent upon the concentration of divalent or monovalent cations, and may, like protein enzymes, be inhibited in the presence of glycholine. Prior to embedding the HHR enzyme sequence into the r545 duplex template, initial tests were performed to ensure that the HHR enzyme strand can be successfully formed from oligonucleotide pieces and can successfully cleave its substrate strand in conditions we employ for the T4 RNA ligase 2 reaction (10 % glycholine by weight, and 2 mM $MgCl_2$). First, to test whether the HHR enzyme strand can be succesfully formed from oligonucleotide pieces assembled on a splint, the HHR enzyme sequence was split into two pieces, a 20-mer and a 23-mer, and we tested whether a splint oligonucleotide is able to template the assembly and ligation of the two pieces (Figure 4.11a). In 2 mM $MgCl_2$, we see that formation of the 43 nt HHR enzyme strand during incubation with T4 RNA ligase 2 proceeds in up to 15+% glycholine (wt/wt) (Figure 4.11b).

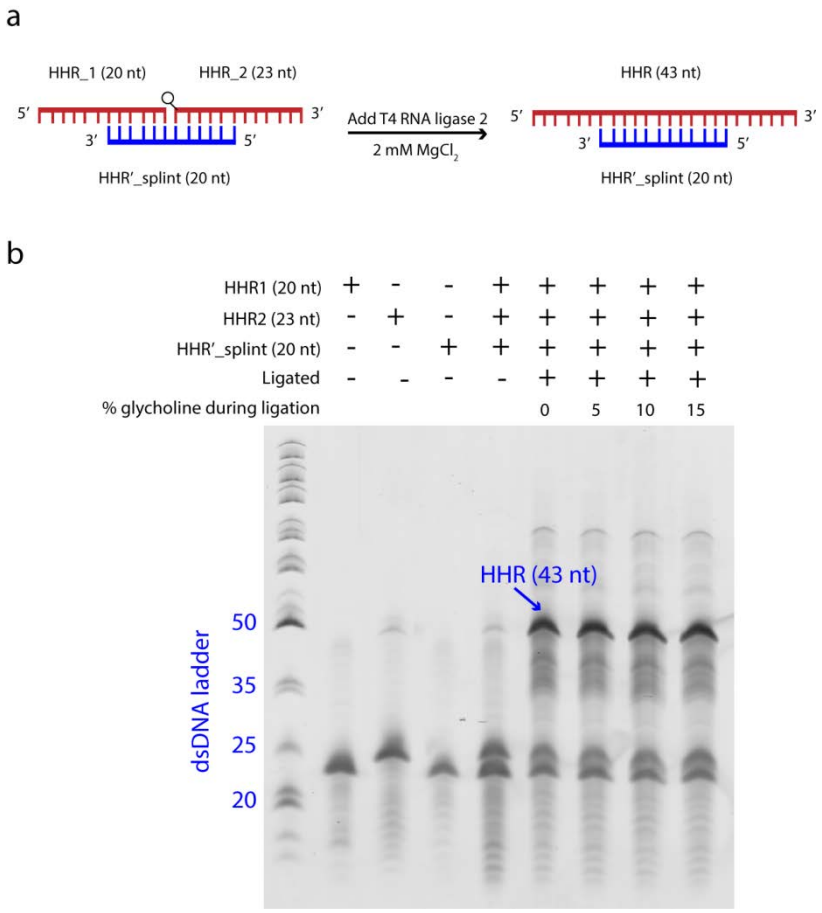

Figure 4.11. Formation of HHR enzyme sequence by ligation of two oligonucleotide sequences on a splint. **a,** Schematic illustrating the assembly of HHR_1 and HHR_2 on an RNA splint (HHR'_splint). HHR_2 is phosphorylated at its 5' end (open circle), allowing ligation to HHR_1 by T4 RNA ligase 2 in the presence of 2 mM $MgCl_2$. **b,** Denaturing polyacrylamide gel (20%, 8 M urea) showing the results of incubating HHR_1, HHR_2, and HHR'_splint with T4 RNA ligase 2 (2 mM $MgCl_2$) in different amounts of glycholine.

Additionally, the HHR enzyme strand demonstrates significant cleavage activity in 10% glycholine (wt/wt) and 2 mM $MgCl_2$ (Figure 4.11c). In these conditions, the HHR cleavage reaction plateaus after ~4 hours, and the yield is approximately half the maximum possible yield (based upon single turnover of the substrate strand, which is present in a 4:1 molar ratio with the enzyme strand) (Figure 4.11d). These results indicate that the HHR enzyme can be synthesized from oligonucleotide building blocks and can

demonstrate cleavage activity in the conditions we utilize for RNA ligation by T4 RNA ligase 2.

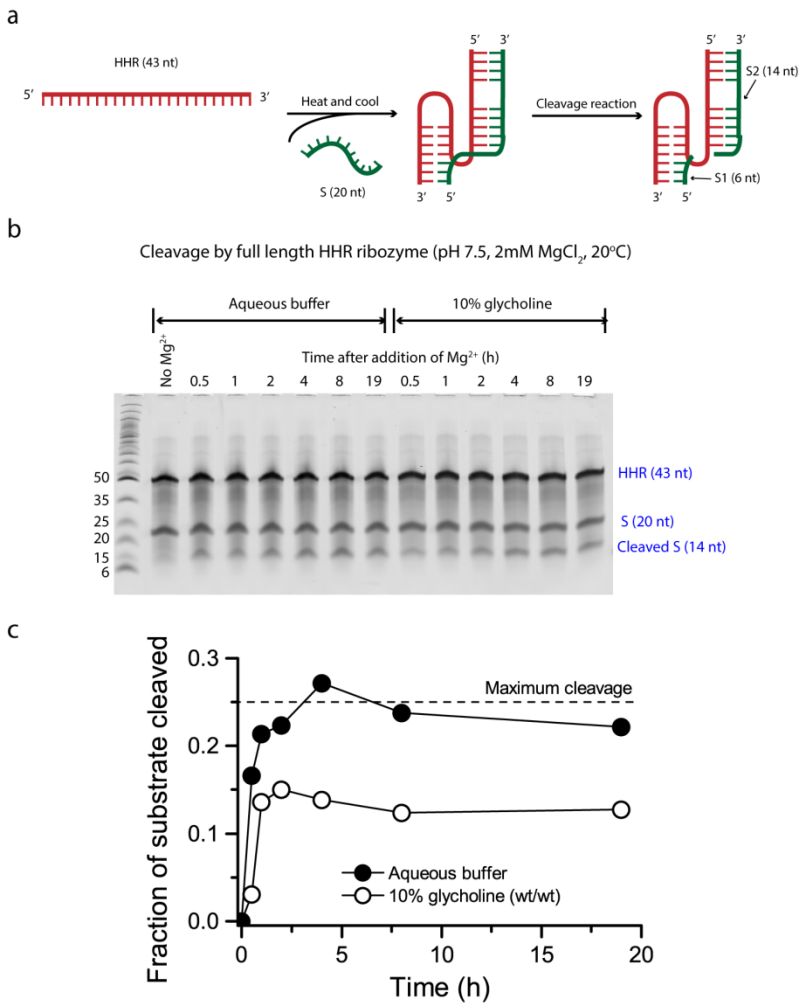

Figure 4.12. HHR cleavage activity in conditions employed for T4 RNA ligase 2 activity (2 mM $MgCl_2$ and 10% glycholine by weight). **a,** Schematic illustrating the experimental procedure for testing cleavage of S_20 by HHR_E. **b,** Denaturing polyacrylamide gel (20%, 9 M urea) showing the results of incubating HHR_E with HHR_S_20 in aqueous buffer and in 10% glycholine (wt/wt), at 2 mM $MgCl_2$. **c,** Kinetics of substrate cleavage by the HHR enzyme from densitometry analysis of gel in **d.** Maximum cleavage occurs when 25% of the substrate is cleaved due to stoichiometric considerations (4:1 molar ratio of substrate to HHR enzyme).

4.5. Information transfer both strands of a gene-length RNA duplex containing the HHR enzyme

After a successful and demonstration that the HHR enzyme sequence is catalytically active in the reaction conditions necessary for T4 RNA ligase 2 activity, we proceeded with inserting the HHR enzyme sequence (43 nt) into the 545 bp RNA duplex utilized previously in Chapter 2. This modified RNA template duplex, referred to as the r604_HHR duplex, contains the active HHR enzyme sequence in close proximity to the 5' end of the sense strand (in the middle of the *rL5* sequence). The procedure for producing the r604_HHR duplex is detailed in Section 4.8.4. Production of RNA template duplex containing the hammerhead enzyme sequence (r604_HHR).

After producing the r604_HHR template duplex, we tested the ability for viscosity to promote copying from both strands of the RNA template. Previous results with a DNA system show that viscosity enables the simultaneous binding of two complementary DNA oligonucleotide sets on both strands of a DNA template duplex (Section 4.3. Template-directed assembly of blunt oligonucleotide sets). Consistent with this proof of principle demonstration, we observe successful information transfer from both strands of the r604_HHR template duplex after glycholine samples containing the r604_HHR duplex, oligonucleotides *rL5_HHR* through *rR4* (*rF0* labeled with Cy3), and complementary oligonucleotides *RL5_HHR_comp* through *rR4_comp* (*rF0_comp* labeled with Cy5) were thermal cycled and incubated with T4 RNA ligase 2 (Figure 4.13, lane 9). The Cy3 signal is weaker than the Cy5 signal when both sets of oligonucleotides are present, particularly for the bands corresponding to the longest products. These results are consistent with previous results, which show that *L5-R5* bind to the anti-sense template strand with lower affinity than *L5_comp-R5_comp* bind to the sense template strand. Control experiments

in aqueous buffer, in glycholine without thermal cycling, and in glycholine in the absence of the r604_HHR template duplex showed no ligated products (lanes 1-4). Notably, the fluorescence signal from the ligated products appears diminished after SYBR Gold staining and imaging (Figure 4.13, right gel), indicating that quenching may be occurring through the interaction of SYBR Gold, Cy3, and Cy5 in the polyacrylamide gel; in particular, Cy3 and Cy5 are known to engage in FRET[18, 19]. While quantitative measurement of copying yields may not be possible, it is clear that glycholine enables reciprocal information transfer from both strands of a gene-length RNA duplex containing the HHR motif.

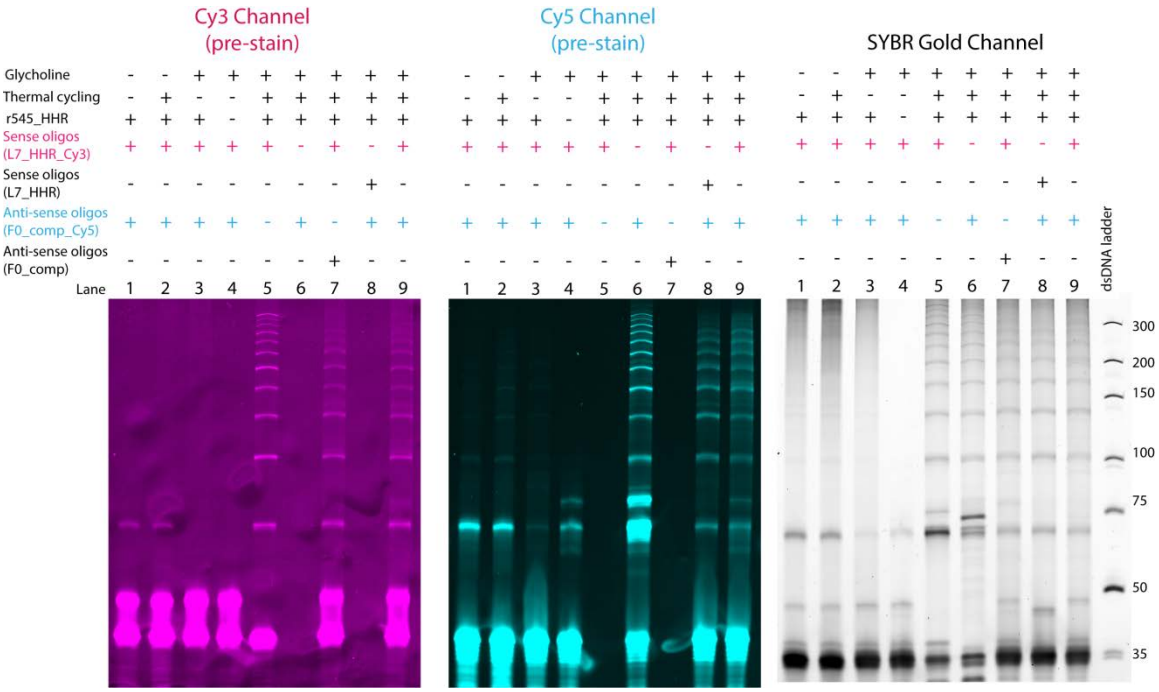

Figure 4.13. Denaturing polyacrylamide gel (10%, 8 M urea) showing the results of thermal cycling samples containing the r604_HHR template duplex and oligonucleotide sets *rL5_HHR-rR5 (rF0* tagged with Cy3) and/or *rL5_HHR_comp-rR5_comp* (*rF0_comp* tagged with Cy5) in aqueous buffer and glycholine. The gel was imaged for Cy3 and Cy5 signal prior to staining with SYBR Gold.

4.6. Cleavage of the HHR substrate by a gene-length RNA template containing the HHR enzyme sequence

While we have shown that the HHR enzyme sequence is catalytically active in our ligaiton conditions (10% glycholine, 2 mM $MgCl_2$), our next step is to show that an HHR motif embedded within a much longer sequence remains catalytically active. Intramolecular structure on the r604_HHR RNA sequence strands may present a barrier to bidning of the HHR substrate strand and correct folding of the HHR motif—which is only 43 nt—into its active conformation. To test whether intramolecular structure may block binding of the HHR substrate to the HHR enzyme sequence, short RNA templates of different sizes were synthesized, ranging from 96 nt to 176 nt, whose sequences were portions of the full r604_HHR sequence and contained the HHR enzyme sequence on the sense strand. The sense strands of these different RNA templates were incubated with the HHR substrate (radioactively labeled with ^{32}P) in aqueous buffer with 2 mM $MgCl_2$, to test HHR substrate binding and cleavage. Incubation with the HHR substrate strand resulted in detectable cleavage by template strands of all sizes, including the full r604_HHR sequence (Figure 4.14a). While the cleavage reaction by the minimal HHR enzyme sequence (43 nt) reaches completion (corresponding to 25% substrate cleavage, based on a 4:1 molar ratio of HHR substrate to enzyme), the cleavage yield drops sharply as the template length increases (Figure 4.14a, 6 nt band), indicating that longer templates possess intramolecular structure that prevents the binding of the HHR substrate to its enzyme sequence. To aid the HHR substrate in opening up intramolecular structure around the HHR enzyme sequence, additional 32 nt flanking sequences were added (corresponding to *L4_comp* to *R4_comp*) to fully coat each template (Figure 4.14a, lanes

5, 7, 9, 11, 13). Surprisingly, the presence of these flanking sequences does not measurably improve the yield of the cleavage reaction (Figure 4.14b).

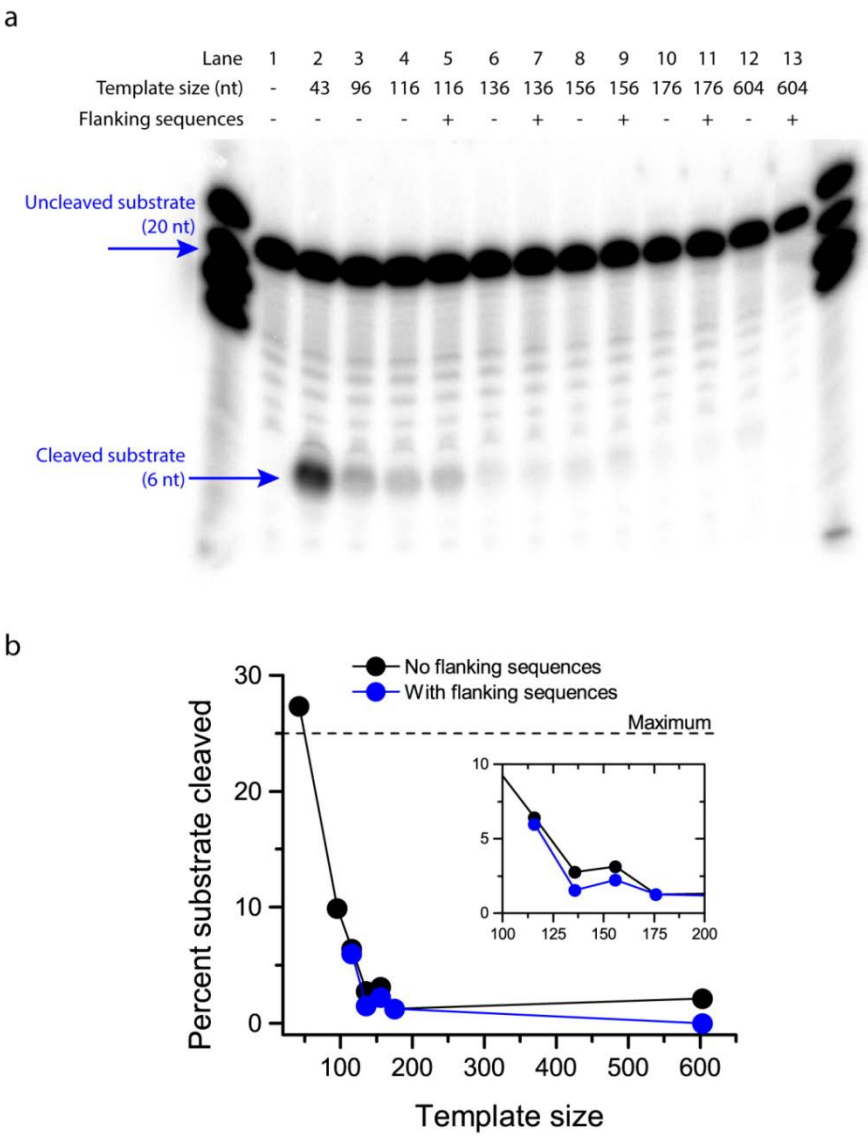

Figure 4.14. Denaturing polyacrylamide gel (20%, 8 M urea) showing the results of incubating single stranded RNA templates of different sizes with the HHR substrate, with and without the presence of complementary 32 nt oligonucleotides.Flanking sequences utilized for each lane are: lane 5, *L4_comp*; lane 7, *L4_comp* and *L3_comp*; lane 9, *L4_comp* and *L3_comp*; lane 11, *L4_comp* through *L2_comp*; lane 13, *L4_comp* through *R4_comp*.

4.7. Conclusions

In this chapter, I investigate the potential of viscous solvents to promote multiple rounds of replication, particularly of a gene-length template (604 bp) containing a hammerhead ribozyme (HHR) motif. While I have previously shown that viscous solvents promote information transfer from one strand of an RNA duplex, building a system which can undergo sustained rounds of replication requires the ability to transfer information from both strands of a template duplex. If oligonucleotide substrates are used, a problem is posed by complementarity between oligonucleotide substrates for each template strand: it is kinetically favored for oligonucleotides to bind to each other in solution, rather than to their sites on the template strands (which likely requires toehold-mediated strand displacement of intramolecular structure). Using two different configurations of oligonucleotide substrates for a 604 bp RNA template, I have demonstrated that, while significant binding of complementary oligonucleotides in free solution does occur, the majority of oligonucleotides bind to their template strands after thermal cycling in glycholine. Ligation of the assembled oligonucleotides and analysis of the product distribution shows that glycholine promotes the copying—i.e. simultaneous information transfer from both strands—of a 384 bp duplex within a larger 604 bp template. This is the first demonstration of copying from a gene-length, mixed sequence RNA template without the aid of polymerases. Additionally, I have shown that the RNA template strand containing the HHR motif can cleave the HHR substrate, demonstrating that viscous environments can be utilized to promote copying of functional (i.e. catalytically active) RNA sequences.

4.8. Materials and Methods

4.8.1. Materials

All solvents and anhydrous solvent samples containing DNA or RNA were prepared as described previously in Section 3.9.1. Materials.

4.8.2. DNA and RNA oligonucleotide sequences

Sequences for the DNA oligonucleotides *L5* through *R5* and RNA oligonucleotides *rL5* through *rR5* are listed in Sections 2.9.2. DNA sequences and 3.9.2. DNA and RNA sequences. Sequences of DNA oligonucleotides *L5'-R5'* are listed below.

L5':	5' - GCTATGTGGCGCGGTATTATCCCGTATTGACG - 3'
L4':	5' - CGTTTTCCAATGATGAGCACTTTTAAAGTTCT - 3'
L3':	5' - GTAAGATCCTTGAGAGTTTTCGCCCCGAAGAA - 3'
L2':	5' - AGTGGGTTACATCGAACTGGATCTCAACAGCG - 3'
L1':	5' - GTAAAAGATGCTGAAGATCAGTTGGGTGCACG - 3'
F0':	5' - CTGTTTTTGCTCACCCAGAAACGCTGGTGAAA - 3'
R1':	5' - CCTTATTCCCTTTTTTGCGGCATTTTGCCTTC - 3'
R2':	5' - AAGAGTATGAGTATTCAACATTTCCGTGTCGC - 3'
R3':	5' - CCTGATAAATGCTTCAATAATATTGAAAAAGG - 3'
R4':	5' - TTCAAATATGTATCCGCTCATGAGACAATAAC - 3'
R5':	5' - GGAACCCCTATTTGTTTATTTTTCTAAATACA - 3'

Sequences of anti-sense RNA oligonucleotides (target sites on the anti-sense strand of pBluescript II SK(-)) are listed below:

rL7_HHR:	5' - CGGCGACCGAGUUGCUCUUGCCCGGCGUCGCA - 3'
rL6_HHR:	5' - GGUACAUCCAGCUGAUGAGUCCCAAAUAGGAC - 3'
rL5_HHR:	5' - GAAAUGCCGAGCUCGGCGUCAAUACGGGAUAA - 3'
rF0_Cy3:	5' - CUUCAGCAUCUUUUAC(U-Cy3)UUCACCAGCGUUU CU - 3'

Sequences of sense RNA oligonucleotides (target sites on the sense strand of pBluescript II SK(-)) are listed below:

rL7_HHR_comp:	5' - UGCGACGCCGGGCAAGAGCAACUCGGUCGCCG - 3'

rL6_HHR_comp: 5' - GUCCUAUUUGGGACUCAUCAGCUGGAUGUACC - 3'
rL5_HHR_comp: 5' - UUAUCCCGUAUUGACGCCGAGCUCGGCAUUUC - 3'
rL4_comp: 5' - GCACUUUUAAAGUUCUGCUAUGUGGCGCGGUA - 3'
rL3_comp: 5' - UUUUCGCCCCGAAGAACGUUUUCCAAUGAUGA - 3'
rL2_comp: 5' - CUGGAUCUCAACAGCGGUAAGAUCCUUGAGAG - 3'
rL1_comp: 5' - AUCAGUUGGGUGCACGAGUGGGUUACAUCGAA - 3'
rF0_comp_Cy5: 5' - AGAAACGCUGGUGAAAG(U-Cy5)AAAAGAUGCUGA AG - 3'
rF0_comp: 5' - AGAAACGCUGGUGAAAGUAAAAGAUGCUGAAG - 3'
rR1_comp: 5' - GCGGCAUUUUGCCUUCCUGUUUUUGCUCACCC - 3'
rR2_comp: 5' - AACAUUUCCGUGUCGCCCUUAUUCCCUUUUUU - 3'
rR3_comp: 5' - AUAAUAUUGAAAAAGGAAGAGUAUGAGUAUUC - 3'
rR4_comp: 5' - CUCAUGAGACAAUAACCCUGAUAAAUGCUUCA - 3'
rR5_comp: 5' - UAUUUUUCUAAAUACAUUCAAAUAUGUAUCCG - 3'

The sequences of the hammerhead ribozyme enzyme and substrate strands are listed below. Sites at which the hammerhead ribozyme enzyme cleaves the phosphodiester bond on the substrate sequences are indicated by //.

HHR_E: 5'- GCAGGUACAUCCAGCUGAUGAGUCCCAAAUAGGACGAAA UGCC - 3'
HHR_E_1: 5'- GCAGGUACAUCCAGCUGAUG - 3'
HHR_E_2: 5'- AGUCCCAAAUAGGACGAAAUGCC - 3'
HHR_splint: 5'- AUUUGGGACUCAUCAGCUGG - 3'
HHR_S_20: 5' - GGCAUC//CUGGAUUCCACUGC - 3'
HHR_S_40: 5' - GCCCGGCGUCGGCAUC//CUGGAUUCCACUGCGAGCUCGG CG - 3'

The cloning vector for producing the RNA template duplex containing the hammerhead enzyme sequence was pUK21-NotI vector DNA (GenBank accession: AF324726).

4.8.3. Measurement of HHR cleavage activity

The HHR cleavage reaction was initiated by the addition of 2 µL of T4 RNA ligase 2 buffer (containing 2 mM $MgCl_2$; New England Biolabs), the desired amount of HHR substrate strand (generally a 4:1 molar ratio with the HHR enzyme sequence), and water up to 20 µL to a 150 mMolal solution of the active enzyme in either glycholine or aqueous buffer (20 mM Tris pH 7.5, 0.1 M NaCl). The final 20 µL reaction mixture was

heated and slowly annealed (95 °C for 2 minutes, cooled to 20 °C at a rate of -0.5 °C/8 seconds on a BioRad thermal cycler). Zero time was taken to occur after annealing was finished, and the reaction was allowed to proceed at 20^{o}C.

4.8.4. Production of RNA template duplex containing the hammerhead enzyme sequence (r604_HHR)

The RNA template duplex utilized in this Chapter is identical in sequence to the 545 bp RNA duplex utilized previously in Chapter 2 except for the inclusion of a hammerhead ribozyme (HHR) enzyme sequence near the 5' end of the sense strand (in the middle of the *rL5* sequence). This modified RNA template duplex (604 bp long) is referred to as r604_HHR. Production of the r604_HHR RNA duplex involved three steps: (1) removing an undesired BsaHI restriction enzyme site on the S- and AS-r545 DNA plasmid constructs; (2) insertion of the HHR enzyme sequence into the DNA plasmid constructs; and (3) co-transcription of both DNA plasmid constructs to produce the r545 duplex. Separate DNA plasmid constructs were engineered to produce the sense (S-r604_HHR) and the anti-sense (AS-r604_HHR) RNA strands. An illustration of steps 1 and 2 is shown below in Figure 4.15.

Figure 4.15. Illustration of the steps involved in engineering the S- and AS-r545 DNA plasmid constructs to include the HHR enzyme sequence. The engineered DNA plasmid constructs are referred to as S- and AS-r604_HHR.

The first step involved insertion of the HHR enzyme sequence (43 nt long; full sequence listed in Section 4.8.2. DNA and RNA oligonucleotide sequences) at the BsaHI restriction enzyme site located in the middle of the *L5* sequence within the linear r545 DNA constructs (described in Section 3.9.10. Production of 545 bp RNA duplex template). The r545 DNA constructs contain two BsaHI restriction enzyme sites, so in order to selectively insert the HHR enzyme sequence in the middle of the *L5* sequence, we mutated the sequence of the second BsaHI site so that it is no longer recognized by the BsaHI restriction enzyme (Figure 4.16).

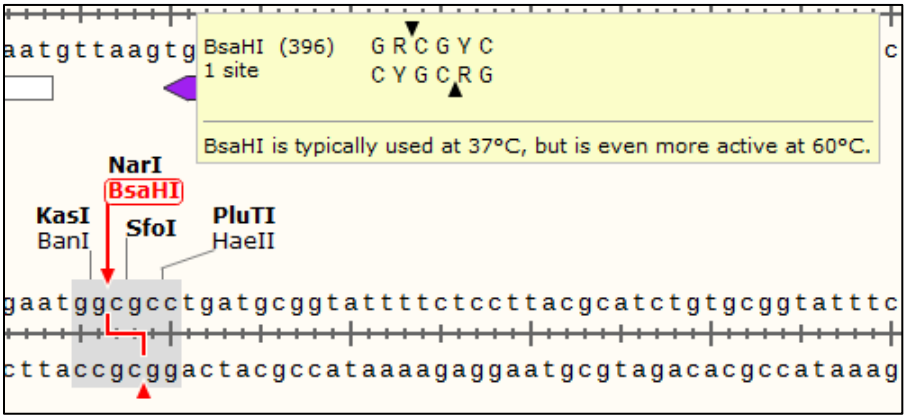

Figure 4.16. Location and sequence of the BsaHI restriction enzyme recognition sequence located at nt 396 in the pUK21-NotI DNA vector sequence. Screenshot taken from a map of the pUK21-NotI DNA vector in Snapgene Viewer.

To mutate the undesired BsaHI sequence, we performed PCR using the following primers, designed to change the restriction enzyme recognition sequence from GGCGCC to ACCGCC:

Forward: 5'- tggcgaat**ac**c**gcc**tgatgcggtattttct -3'
Reverse: 5'- ccgcatca**ggcgg**tattcgccattcaggct -3'

Color code: **Original BsaHI site; mutations introduced to BasHI site;** intended overhangs.

After the PCR reaction proceeded, the sense and anti-sense DNA constructs were treated with the kinase-ligase-DpnI (KLD) enzyme master mix provided in the Q5 site-directed mutagenesis kit (New England Biolabs) for 15 minutes at room temperature. The kinase phosphorylates the linear DNA constructs, the ligase circularizes the DNA constructs, and DpnI digests the original PCR template, leaving only the PCR product. After treatment with KLD mix, the DNA constructs were transformed into competent DH5α *E. coli* cells and grown on plates containing kanamycin to select for successful plasmid uptake. Plasmid DNA was isolated from six separate colonies, digested with BsaHI, and run on an agarose gel. Successful mutation of the undesired BsaHI restriction enzyme site was confirmed by observation of a banding pattern indicative of a single restriction enzyme cleavage reaction (rather than two).

After identification of plasmid DNA with only one BsaHI restriction enzyme site, the next step was insertion of the following sequence containing the HHR enzyme:

5' – aaagg‖cgtc**GCAGGTACATCCAGCTGATGAGTCCCAAATAGGACGAAATGCC**gagct‖cgg‖cgtcaaa – 3'
5' – tttga‖cgccgagct‖c**GGCATTTCGTCCTATTTGGGACTCATCAGCTGGATGTACCTGC**ga‖cgccttt – 3'

Code: **`HHR enzyme sequence`**; BsaHI site; SacI site; ‖ = cuts by restriction enzymes. The extra a/t's at the end are to give the restriction enzyme enough room to bind (generally 3-6 bp needed).

This HHR enzyme fragment is a 67 bp DNA duplex flanked by BsaHI sites, allowing for insertion into the middle of the *L5* sequence of the plasmid DNA construct. In addition, it contains a SacI restriction enzyme site at one end. This is because, like many restriction enzymes, BsaHI generates sticky ends which are palindromic, and so the HHR insert can

be incorporated into the plasmid in either of two orientations. We want the active HHR sequence to be specifically located on one strand, and the complementary, non-active HHR' to be located on the other strand; therefore, we will use the size of fragments generated by SacI digestion as a means of distinguishing between the two orientations. We specifically chose SacI as a restriction enzyme site because pUK21-NotI does not have SacI, resulting in only one cut.

To insert the HHR enzyme fragment, both the HHR enzyme fragment and the plasmid DNA construct were digested with BsaHI, then mixed together and incubated with KLD mix (Q5 site-directed mutagenesis kit, New England Biolabs) for 15 minutes at room temperature. The resulting circular plasmid—r604_HHR DNA plasmid—was transformed into competent DH5α *E. coli* cells. Plasmid uptake was observed by growth on media containing kanamycin, and plasmid DNA from six individual colonies was isolated. The plasmid DNA was double digested with EcoRI/SacI and run on a polyacrylamide gel. Plasmid DNA from colonies that incorporated the HHR enzyme fragment in the correct orientation produced a 114 bp fragment, while incorporation in the undesired orientation produced a 71 bp fragment. The sequence of the r604_HHR DNA plasmid was confirmed by sequencing analysis (Eurofins MWG Operon).

After successful incorporation of the HHR enzyme fragment, a large quantity of the r604_HHR DNA plasmids (both the S and AS) was isolated with a Maxiprep kit (Qiagen). The S- and AS-r604_HHR DNA plasmids were then treated with Proteinase K (New England Biolabs) and purified with phenol-chloroform extraction followed by ethanol precipitation, to remove RNase.

Transcription and purification of r604_HHR RNA duplex from the AS- and S-r604_HHR DNA plasmids was carried out using the same protocol described in Section 3.9.10. Production of 545 bp RNA duplex template.

4.8.5. Production of RNA template duplex containing the hammerhead enzyme sequence (r96_HHR through r176_HHR)

The r604_HHR RNA duplex likely contains intramolecular structure—either local hairpins or longer range tertiary interactions—that may pose a barrier to the association of the HHR substrate to the HHR enzyme sequence, and the arrangement of the HHR substrate-enzyme complex into its active conformation. Therefore, in addition to the r604_HHR RNA duplex, we wanted to test the binding and cleavage of the HHR substrate by short RNA duplex templates containing the HHR enzyme sequence. If intramolecular structure at the site of the HHR enzyme sequence is indeed blocking access by the HHR substrate, then we should see a relationship between the degree of HHR substrate binding and template length. To produce shorter RNA duplex sequences, we performed PCR on the AS-r604_HHR DNA plasmid construct using the primer combinations listed in Table 4.1.

To produce the S RNA strand:

S_FWD: 5' – TAATACGACTCACTATAGACCGAGTTGCTCTTGCCCGG – 3'
S_REV1: 5' – GGTATTATCCCGTATTGACG – 3'
S_REV2: 5' – AAAGTTCTGCTATGTGGCGC – 3'
S_REV3: 5' – TTCCAATGATGAGCACTTTT – 3'
S_REV4: 5' – TTTTCGCCCCGAAGAACGTT – 3'
S_REV5: 5' – AGCGGTAAGATCCTTGAGAG – 3'

To produce the AS RNA strand:

AS_FWD: 5' – ACCGAGTTGCTCTTGCCCGG – 3'
AS_REV1: 5' – TAATACGACTCACTATAGGGTATTATCCCGTATTGACG – 3'
AS_REV2: 5' – TAATACGACTCACTATAGAAAGTTCTGCTATGTGGCGC – 3'

AS_REV3: 5' – TAATACGACTCACTATAGTTCCAATGATGAGCACTTTT – 3'
AS_REV4: 5' – TAATACGACTCACTATAGTTTTCGCCCCGAAGAACGTT – 3'
AS_REV5: 5' – TAATACGACTCACTATAGAGCGGTAAGATCCTTGAGAG – 3'

Table 4.1. Primer combinations utilized in PCR's to produce DNA templates for transcription into short RNA duplexes containing the HHR enzyme sequence.

Template for PCR reaction	Forward/Reverse Primer to Produce the Sense Strand	Forward/Reverse Primer to Produce the Anti-sense Strand	Sequence length of RNA after co-transcription(bp)
AS-r604_HHR	S_FWD/S_REV1	AS_FWD/AS_REV1	96
AS-r604_HHR	S_FWD/S_REV2	AS_FWD/AS_REV2	116
AS-r604_HHR	S_FWD/S_REV3	AS_FWD/AS_REV3	136
AS-r604_HHR	S_FWD/S_REV4	AS_FWD/AS_REV4	156
AS-r604_HHR	S_FWD/S_REV5	AS_FWD/AS_REV5	176

After carrying out PCR, the resulting DNA templates were purified on a 7.5% polyacrylamide gel (non-denaturing, 0.5x TBE buffer) with the same procedure as described in Section 2.9.5. PAGE purification of DNA oligonucleotides. Templates were transcribed with the HiScribe T7 High Yield RNA Synthesis Kit (New England Biolabs) using conditions recommended for short transcripts: 2 µg of each DNA template to be transcribed, and incubation at 37 °C overnight. Removal of DNA from the transcription reaction and subsequent purification of the RNA were carried out as described in 3.9.10. Production of 545 bp RNA duplex template.

4.8.6. Imaging of gels containing Cy3- and Cy5-labeled oligonucleotides

Agarose and polyacrylamide gels containing samples with Cy3- and Cy5-labeled oligonucleotides were imaged on a Typhoon FLA 9500 laser scanner (GE Healthcare). Immediately after the gels finished running, they were imaged for Cy3 signal (532 nm excitation and a band pass filter at 570 ± 20 nm) and/or Cy5 signal (635 nm excitation and a long pass filter for 670+ nm). After Cy3/Cy5 imaging, the gels were stained with either ethidium bromide to visualize the position of the template (agarose gels) or SYBR

Gold to visualize all nucleic acids (polyacrylamide gels), and then imaged with the appropriate filters (described in Sections 2.9.6. Conditions for agarose gel electrophoresis of glycholine samples and 2.9.7. DNA ligation procedure). Note that the excitation/emission spectra of ethidium bromide and SYBR Gold overlap with those of Cy3, and so images of stained gels cannot be utilized to quantify nucleic acid concentration.

4.8.7. ^{32}P-labeling of the HHR substrate

The HHR substrate strands were labeled at the 5'-end with ^{32}P, to visualize binding to various RNA templates containing the HHR enzyme sequence. Prior to labeling with with ^{32}P, HHR_S_20 and HHR_S_40 were treated with Antarctic phosphatase (New England Biolabs) at 37 °C for 1+ hours to remove any 5'-phosphates present on the RNA oligonucleotides. After phosphatase treatment, the phosphorylation reaction was carried out by incubating the RNA oligonucleotides with ^{32}P-ATP (Perkin Elmer) and T4 polynucleotide kinase (New England Biolabs) for 1-4 hours. Notably, the HHR_S_40 sequence required a significantly longer kinase incubation time than the HHR_S_20 sequence to achieve comparable labeling efficiency, indicating that self-structure on the HHR_S_40 strand may be blocking access of the kinase enzyme to the 5' end. RNA oligonucleotides were then purified using a NAP-5 column (GE Healthcare) and ethanol precipitation to remove Mg^{2+} introduced in the phosphatase and kinase buffers.

4.8.8. DNA and RNA ligation procedure

DNA and RNA ligation were carried out using T4 DNA ligase and T4 RNA ligase 2, respectively, as described previously in Section 3.9.4. DNA ligation procedure and Section 3.9.11. RNA ligation procedure.

4.9. References

1. Cafferty BJ, Fialho DM, Khanam J, Krishnamurthy R, Hud NV. Spontaneous formation and base pairing of plausible prebiotic nucleotides in water. *Nature communications* 2016, **7**.

2. Cafferty BJ, Gállego I, Chen MC, Farley KI, Eritja R, Hud NV. Efficient Self-Assembly in Water of Long Noncovalent Polymers by Nucleobase Analogues. *Journal of the American Chemical Society* 2013, **135**(7)**:** 2447-2450.

3. Powner MW, Gerland B, Sutherland JD. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* 2009, **459**(7244)**:** 239-242.

4. Patel BH, Percivalle C, Ritson DJ, DuffyColm D, Sutherland JD. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat Chem* 2015, **7**(4)**:** 301-307.

5. Burcar B, Pasek M, Gull M, Cafferty BJ, Velasco F, Hud NV, *et al.* Darwin's Warm Little Pond: A One-Pot Reaction for Prebiotic Phosphorylation and the Mobilization of Phosphate from Minerals in a Urea-Based Solvent. *Angewandte Chemie International Edition* 2016, **55**(42)**:** 13249–13253.

6. Szostak JW. The eightfold path to non-enzymatic RNA replication. *Journal of Systems Chemistry* 2012, **3**(2).

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

8. Mutschler H, Wochner A, Holliger P. Freeze–thaw cycles as drivers of complex ribozyme assembly. *Nat Chem* 2015, **7**(6)**:** 502-508.

9. He C, Gállego I, Laughlin B, Grover MA, Hud NV. A viscous solvent enables information transfer from gene-length nucleic acids in a model prebiotic replication cycle. *Nat Chem* 2016, **advance online publication**.

10. Ivica NA, Obermayer B, Campbell GW, Rajamani S, Gerland U, Chen IA. The Paradox of Dual Roles in the RNA World: Resolving the Conflict Between Stable Folding and Templating Ability. *Journal of Molecular Evolution* 2013, **77**(3)**:** 55-63.

11. Attwater J, Wochner A, Holliger P. In-ice evolution of RNA polymerase ribozyme activity. *Nat Chem* 2013, **5**(12)**:** 1011-1018.

12. Dornberger U, Leijon M, Fritzsche H. High Base Pair Opening Rates in Tracts of GC Base Pairs. *Journal of Biological Chemistry* 1999, **274**(11)**:** 6957-6962.

13. Campa A. Bubble propagation in a helicoidal molecular chain. *Physical Review E* 2001, **63**(2)**:** 021901.

14. Altan-Bonnet G, Libchaber A, Krichevsky O. Bubble dynamics in double-stranded DNA. *Physical Review Letters* 2003, **90**(13)**:** 138101.

15. Seehafer C, Kalweit A, Steger G, Gräf S, Hammann C. From alpaca to zebrafish: Hammerhead ribozymes wherever you look. *RNA* 2011, **17**(1)**:** 21-26.

16. Martick M, Scott WG. Tertiary Contacts Distant from the Active Site Prime a Ribozyme for Catalysis. *Cell* 2006, **126**(2)**:** 309-320.

17. Murray JB, Seyhan AA, Walter NG, Burke JM, Scott WG. The hammerhead, hairpin and VS ribozymes are catalytically proficient in monovalent cations alone. *Chemistry & Biology* 1998, **5**(10)**:** 587-595.

18. Hohng S, Joo C, Ha T. Single-Molecule Three-Color FRET. *Biophysical Journal* 2004, **87**(2)**:** 1328-1337.

19. Lavergne T, Lamichhane R, Malyshev DA, Li Z, Li L, Sperling E, *et al.* FRET Characterization of Complex Conformational Changes in a Large 16S Ribosomal RNA Fragment Site-Specifically Labeled Using Unnatural Base Pairs. *ACS Chemical Biology* 2016, **11**(5)**:** 1347-1353.

CHAPTER 5 : CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Conclusions and Future Directions

While significant advances have been made in recent years towards elucidating a prebiotic route to mononucleotides and nucleic acid polymers[1-5], demonstrating enzyme-free replication of nucleic acid polymers remains a largely unsolved challenge[6]. The results reported here demonstrate that hot/cool cycles in viscous environments promote copying from gene-length, mixed sequence nucleic acid duplexes. I have shown that biophysical challenges associated with copying naked nucleic acids in the absence of enzymes—such as the strand inhibition problem[7]—may be overcome through the use of viscous solvents, which can dramatically alter the thermodynamics and kinetics of nucleic acid duplex formation compared to aqueous buffer[8, 9]. I have also demonstrated that viscous environments can promote the copying of RNA sequences containing a catalytically active hammerhead ribozyme motif. These results represent important progress towards the goal of understanding how the two functions that define RNA's role in the putative RNA World—the ability to serve as a template for the transfer of genetic information, and the ability to catalyze reactions—may have evolved in a prebiotic setting, driven solely by physical and chemical processes. These results also highlight the important role that geochemical factors and the physical environment likely played in the emergence and evolution of nucleic acids.

Further exploration of the prebiotic role of viscous environments may center on their ability to promote nucleic acid replication in a length- and structure-dependent manner, providing a selection pressure for the selection of certain sequence

characteristics over multiple rounds of replication. It is currently unclear how a catalytically active RNA sequence would have been selected from a complex pool of sequences and persisted in a prebiotic environment. It has been known since the 1960s, through *in vitro* experiments with viral RNA replication in an extracellular environment, that long genomic sequences are outcompeted by the faster replication of shorter sequences, resulting in shortening of the average sequence length over multiple rounds of replication[10]. Studies have demonstrated that the loss of genetic information during replication in prebiotic conditions can be avoided in specific environments that promote selective accumulation of long sequences, such as a flow system that simulates a hydrothermal vent pore[11] or repeated cycling between nucleic acid compartmentalization in oil droplets and mixing in bulk solution[12].

Viscous environments have a similar effect as compartmentalization, but rather than physically separating sequences by length or replication rate, long sequences are more readily *kinetically* trapped in the single stranded template state. Therefore, viscosity provides a means of selectively replicating long, gene-length sequences over shorter sequences. Viscous environments could have been readily generated on the early Earth by evaporation of water from terrestrial pools, and may provide a more general solution than previous approaches, which require very specific environmental conditions—such as pore geometry and flow rates[11] or manual separation of oil droplets by a microfluidic device[12]—to maintain sequences long enough to be constitute a genetically viable system or demonstrate RNA replicase activity. A demonstration of viscosity-mediated information transfer utilizing multiple nucleic acid templates of different sequences lengths—which can “compete” for oligonucleotide substrates—over multiple rounds of

replication would allow measurement of how the sequence distribution changes over time, and the relative replication rates of different sequences in a mixed pool.

Another conceptual challenge in understanding the emergence of an RNA World is the "replicator-catalyst" paradox, which describes the apparent structural incompatibility of templating and catalytic functions in a single nucleic acid sequence. Catalytic activity requires the stable folding of a nucleic acid sequence into an active conformation, which in turn presents a barrier to the binding of mono- or oligonucleotide substrates[13]. In our viscosity-mediated replication cycle, the ability of a sequence to be replicated is in fact aided by its ability to form intramolecular structure. Therefore, the problem posed by the replicator-catalyst paradox may be lessened in a viscous environment because the formation of intramolecular structure on a sequence is not necessarily at odds with its templating ability; rather, the formation of intramolecular structure on a sequence can increase the propensity for *both* efficient replication and catalytic ability. Though intramolecular structure still presents an obstacle to the binding of oligonucleotide substrates, the presence of multiple oligonucleotides helps to open up template structure through toehold-mediated strand displacement. In a viscous environment, the most readily replicated sequences are likely those which form intramolecular structure of an optimal stability—stable enough that intramolecular structure to remain kinetically trapped, but labile enough to be opened up by invading oligonucleotide substrates. An interesting question for further investigation is where the tipping point between detrimental versus beneficial intramolecular structure in a nucleic acid sequence lies.

A striking demonstration of viscosity's ability to solve the replicator-catalyst paradox would be replication of a duplex containing a catalytically active motif on one strand. Upon heating and cooling, the two strands would become kinetically trapped; one strand would serve as a template for the assembly of oligonucleotide substrates, while the other would carry out a reaction—perhaps catalyzing the replication of another sequence. In this way, a single sequence—and its complementary strand—could achieve a "division of labor" between genetic and catalytic functions. A replication system involving strand-specific functions has been proposed previously, but employs intramolecular G•U wobble pairs to generate significant differences in intramolecular folding energies between the two strands [13]. While G•U wobble pairs represent a viable means of encapsulating genetic and catalytic functions into one sequence, viscosity obviates the need for asymmetry between the folded states of each strand; intramolecular structure aids both replication and catalytic activity.

In conclusion, viscous environments produced by hot/cool cycles provide a potential route to the enzyme-free replication of nucleic acid duplexes, including those containing catalytically active motifs within their sequences. The results presented here provide insight into how structured, catalytically active RNA sequences may have been selectively amplified from a pool of sequences. Compared to previously described environments for the non-enzymatic, template-directed synthesis of nucleic acids, viscous environments are easily generated by water evaporation, and viscosity-mediated information transfer is not limited to a specific solvent. Further investigation of viscosity's role in driving replication undoubtedly requires exploration of more prebiotically feasible organic solvents[5] (such as purine based eutectic solvents[14]),

developing systems where very gradual temperature changes can promote replication (on the order of a day/night or seasonal cycle), and carrying out multiple rounds of replication. Though the ultimate goal of demonstrating a self-sustained, enzyme-free RNA replication system still remains elusive, the results described here provide important clues into the emergence of nucleic acid replication and evolution of nucleic acid polymers on the early Earth.

5.2 References

1. Powner MW, Gerland B, Sutherland JD. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* 2009, **459**(7244)**:** 239-242.

2. Patel BH, Percivalle C, Ritson DJ, DuffyColm D, Sutherland JD. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat Chem* 2015, **7**(4)**:** 301-307.

3. Cafferty BJ, Gállego I, Chen MC, Farley KI, Eritja R, Hud NV. Efficient Self-Assembly in Water of Long Noncovalent Polymers by Nucleobase Analogues. *Journal of the American Chemical Society* 2013, **135**(7)**:** 2447-2450.

4. Cafferty BJ, Fialho DM, Khanam J, Krishnamurthy R, Hud NV. Spontaneous formation and base pairing of plausible prebiotic nucleotides in water. *Nature communications* 2016, **7**.

5. Burcar B, Pasek M, Gull M, Cafferty BJ, Velasco F, Hud NV, *et al.* Darwin's Warm Little Pond: A One-Pot Reaction for Prebiotic Phosphorylation and the Mobilization of Phosphate from Minerals in a Urea-Based Solvent. *Angewandte Chemie International Edition* 2016, **55**(42)**:** 13249–13253.

6. Szostak JW. The eightfold path to non-enzymatic RNA replication. *Journal of Systems Chemistry* 2012, **3**(2).

7. He C, Gállego I, Laughlin B, Grover MA, Hud NV. A viscous solvent enables information transfer from gene-length nucleic acids in a model prebiotic replication cycle. *Nat Chem* 2016, **advance online publication**.

8. Mamajanov I, Engelhart AE, Bean HD, Hud NV. DNA and RNA in Anhydrous Media: Duplex, Triplex, and G-Quadruplex Secondary Structures in a Deep

Eutectic Solvent. *Angewandte Chemie International Edition* 2010, **49**(36)**:** 6310-6314.

9. Lannan FM, Mamajanov I, Hud NV. Human Telomere Sequence DNA in Water-Free and High-Viscosity Solvents: G-Quadruplex Folding Governed by Kramers Rate Theory. *J Am Chem Soc* 2012, **134**(37)**:** 15324–15330.

10. Mills DR, Peterson RL, Spiegelman S. An Extracellular Darwinian Experiment with a Self-duplicating Nucleic Acid Molecule. *Proceedings of the National Academy of Sciences* 1967, **58**(1)**:** 217-224.

11. Kreysing M, Keil L, Lanzmich S, Braun D. Heat flux across an open pore enables the continuous replication and selection of oligonucleotides towards increasing length. *Nat Chem* 2015, **7**(3)**:** 203-208.

12. Matsumura S, Kun Á, Ryckelynck M, Coldren F, Szilágyi A, Jossinet F, *et al.* Transient compartmentalization of RNA replicators prevents extinction due to parasites. *Science* 2016, **354**(6317)**:** 1293-1296.

13. Ivica NA, Obermayer B, Campbell GW, Rajamani S, Gerland U, Chen IA. The Paradox of Dual Roles in the RNA World: Resolving the Conflict Between Stable Folding and Templating Ability. *Journal of Molecular Evolution* 2013, **77**(3)**:** 55-63.

14. Shahbaz K, AlNashef IM, Lin RJT, Hashim MA, Mjalli FS, Farid MM. A novel calcium chloride hexahydrate-based deep eutectic solvent as a phase change materials. *Solar Energy Materials and Solar Cells* 2016, **155:** 147-154.