Spontaneous Cooperative Assembly of Replicative Catalytic RNA Systems

Nilesh Vaidya
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds

Part of the Biochemistry Commons, and the Chemistry Commons

Let us know how access to this document benefits you.

Recommended Citation
https://doi.org/10.15760/etd.934

This Dissertation is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.
Spontaneous Cooperative Assembly of Replicative Catalytic RNA Systems

by

Nilesh Vaidya

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Chemistry

Dissertation Committee:
Niles Lehman, Chair
Dirk Iwata-Reuyl
Albert S. Benight
Michael Bartlett
Gary Brodowicz

Portland State University
2012
ABSTRACT

The RNA World hypothesis proposes a period of time during the origins of life in which RNA molecules were the only source of both genotypes and phenotypes. Although a vast amount of evidence has been obtained in support of this hypothesis, a few critical demonstrations are lacking. A most crucial one is a demonstration of self-replication of RNA molecule from prebiotic soup. Previously in the Lehman laboratory, it has been demonstrated that a 198-nucleotide molecule derived from the Azoarcus group I intron can self-assemble from up to four fragments of RNA via recombination. Furthermore, the covalent full-length molecules are catalytically active and can make copies of themselves from the remaining pieces in the solution leading to their autocatalytic growth. I was able to demonstrate how this recombination system can overcome different obstacles and evolve to be an efficient replicating system. I discovered the ability of a single RNA fragment to be multifunctional in a single reaction pathway during RNA recombination events that avoids the necessity of multiple genotypes. I also confirmed the capacity of self-replicating ribozymes to form cooperative catalytic cycles and networks that would potentially prevent informational decay. Finally, I have discovered a recycling phenomenon in the RNA recombination system that exploits dynamic covalent chemistry. Recycling provides the earliest replicating system with adequate concentrations of reagents and ability to explore sequence space. Together these findings have improved our understanding of RNA recombination and bolstered the plausibility of the RNA World.
DEDICATION

To my parents Dil Chandrananda Vaidya and Amina Vaidya
ACKNOWLEDGEMENTS

Dr. Niles Lehman

Committee members:
Dr. Dirk Iwata-Reuyl
Dr. Albert S. Benight
Dr. Michael Bartlett
Dr. Gary Brodowicz

Dr. Irene A. Chen, Dr. Michael L. Manapat, and Dr. Ramon Xulvi-Brunet for modeling cooperative RNA network

Dr. Eric J. Hayden for discussions on cooperative RNA network

Dr. Sara I. Walker for discussions on recycling of RNA fragments during recombination

Lehman Lab members past and present

the Department of Chemistry at Portland State University

My family and friends
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>i</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td><strong>CHAPTER 1. INTRODUCTION</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>CHAPTER 2. ONE RNA PLAYS THREE ROLES TO PROVIDE CATALYTIC ACTIVITY TO A GROUP I INTRON LACKING AN ENDOGENOUS INTERNAL GUIDE SEQUENCE</strong></td>
<td></td>
</tr>
<tr>
<td>Abstract</td>
<td>11</td>
</tr>
<tr>
<td>Introduction</td>
<td>12</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>13</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>22</td>
</tr>
<tr>
<td><strong>CHAPTER 3: SPONTANEOUS NETWORK FORMATION AMONG COOPERATIVE RNA REPLICATORS</strong></td>
<td></td>
</tr>
<tr>
<td>Abstract</td>
<td>35</td>
</tr>
<tr>
<td>Introduction</td>
<td>36</td>
</tr>
<tr>
<td>Results</td>
<td>38</td>
</tr>
<tr>
<td>Discussion</td>
<td>54</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>57</td>
</tr>
</tbody>
</table>
CHAPTER 4: RECYCLING OF INFORMATIONAL UNITS LEADS TO SELECTION OF REPLICATORS IN A PREBIOTIC SOUP

Abstract 91
Introduction 92
Results 95
Discussion 99
Materials and Methods 101

CHAPTER 5: CONCLUSIONS 114

REFERENCES 119
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Activity assays of three clones derived from ribozymes self-assembled from RNAs lacking the canonical IGS</td>
<td>32</td>
</tr>
<tr>
<td>Table 2</td>
<td>Covalent self-assembly yields in different ratio of $V \cdot X \cdot Y$ and $h \cdot Z$ after 4 h</td>
<td>33</td>
</tr>
<tr>
<td>Table 3</td>
<td>Activity assays of reaction systems in which mutations were engineered into the exogenous IGS triplets</td>
<td>34</td>
</tr>
<tr>
<td>Table 4</td>
<td>Demonstration of cooperation through two-step (relayed) catalysis</td>
<td>88</td>
</tr>
<tr>
<td>Table 5</td>
<td>Genotype frequencies from high throughput sequence analysis of the $W \cdot X \cdot Y \cdot Z$ ribozymes formed in the randomization experiment</td>
<td>89</td>
</tr>
<tr>
<td>Table 6</td>
<td>Genotype frequencies from manual sequence analysis of the $W \cdot X \cdot Y \cdot Z$ ribozymes formed in the randomized serial transfer experiment</td>
<td>90</td>
</tr>
<tr>
<td>Table 7</td>
<td>Sequence analysis of recombined junctions from a complex RNA network</td>
<td>113</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Two-step self-splicing process in group I intron</td>
<td>7</td>
</tr>
<tr>
<td>Figure 2</td>
<td>The <em>Azoarcus</em> recombinase ribozyme</td>
<td>8</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Recombination by the <em>Azoarcus</em> ribozyme</td>
<td>9</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Self-assembly of <em>Azoarcus</em> ribozyme from inactive RNA fragments</td>
<td>10</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Schematic of the partitioning of the <em>Azoarcus</em> ribozyme into fragments</td>
<td>25</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Activity assays of ribozymes lacking a native IGS</td>
<td>26</td>
</tr>
<tr>
<td>Figure 7</td>
<td><em>Trans</em>-splicing assays of the IGS-less ribozyme $\text{V} \cdot \text{X} \cdot \text{Y} \cdot \text{Z}$</td>
<td>27</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Analysis of splice-site specificity during covalent self-assembly in the absence of a canonical IGS</td>
<td>28</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Model of utilization of an exogenous IGS</td>
<td>29</td>
</tr>
<tr>
<td>Figure 10</td>
<td>Model in which a recombinase ribozyme can covalently self-assemble from two RNA fragments despite the lack of a native IGS in the catalytic complex</td>
<td>30</td>
</tr>
<tr>
<td>Figure 11</td>
<td>Test of the ability of nucleotide triplets in $\text{h} \cdot \text{Z}$ to perform as exogenous IGSs</td>
<td>31</td>
</tr>
<tr>
<td>Figure 12</td>
<td>Cooperative covalent assembly of recombinase ribozymes</td>
<td>65</td>
</tr>
<tr>
<td>Figure 13</td>
<td>Analyses to determine the IGS-IGS target pairs to be used in the cooperative RNA network</td>
<td>66</td>
</tr>
<tr>
<td>Figure 14</td>
<td>Covalent ribozymes are more catalytically proficient than non-covalent versions</td>
<td>67</td>
</tr>
<tr>
<td>Figure 15</td>
<td>Electrophoretic observation of assemblies of $\text{E}_2$ and $\text{E}_3$</td>
<td>68</td>
</tr>
<tr>
<td>Figure 16</td>
<td>Graphical comparison of yields of assembly of full length ribozymes</td>
<td>69</td>
</tr>
</tbody>
</table>
Figure 17  Yields of individual subsystems $E_1$, $E_2$, and $E_3$ over time in the presence of all three subsystems

Figure 18  Effects of doping of covalent catalysts into system

Figure 19  Construction of mutants used to separate covalent from non-covalent catalysis

Figure 20  Characterization of non-covalent and non-productive RNAs

Figure 21  Effect of prefolding on RNA assembly yields

Figure 22  Heat cycling improves sigmoidality of the growth curves

Figure 23  Simplex plot of genotype frequencies changing over time

Figure 24  Design of selfish subsystems

Figure 25  Cooperative chemistry out-competes selfish chemistry when directly competed

Figure 26  A second example of a selfish vs. cooperative competition experiment

Figure 27  Asymmetry in subsystem assembly rates grants a competitive advantage to cooperative ribozymes in a mixed population

Figure 28  Mathematical modeling of molecular cooperation

Figure 29  The randomization experiment

Figure 30  Proposed succession from simple to complex networks using genotype frequency data

Figure 31  The potential network of RNA genotypes in the randomization experiment

Figure 32  The serial transfer experiment

Figure 33  Growth curve of a 4-piece of system

Figure 34  Example sequences resulting from the 4-piece assembly experiment
| Figure 35 | General scheme demonstrating importance of recycling | 104 |
| Figure 36 | Recycling of RNA fragments in the \textbf{W•X•Y•Z} system | 105 |
| Figure 37 | Shuffling of fragments to demonstrate the global search for a covalently contiguous ribozyme | 106 |
| Figure 38 | Shuffling of fragments during self-assembly of ribozyme | 107 |
| Figure 39 | Demonstration of recycling in a cyclical cooperative network | 108 |
| Figure 40 | Secondary structure of catalytic core of the ribozyme | 109 |
| Figure 41 | Restoration of catalytic activity after recycling | 110 |
| Figure 42 | Demonstration of recycling in recombinase ribozyme to increase local fitness | 111 |
| Figure 43 | Possible mechanism of recycling in the recombinase system | 112 |
CHAPTER 1. INTRODUCTION

The RNA molecule plays a significant role in the most fundamental cellular processes in modern biology. It can fold into a variety of complex three-dimensional structures and catalyze a significant number of diverse chemical reactions including the ribosomal translation of mRNA into protein (Ban et al., 2000; Chen et al., 2007). It is involved as a messenger that carries genetic information to the translation machinery, instructs the processing of precursor mRNA during splicing and editing, and is responsible for tRNA maturation. RNA also helps in DNA replication as a primer and in the replication of viral genomes. Moreover, individual nucleotides serve as important signaling molecules and their coenzyme derivatives participate in most of the reactions of central metabolism. All these roles of RNA in contemporary biology provide evidence for the “RNA World” hypothesis, which suggests that genetic information was stored and replicated by RNA molecules during the origins of life on the Earth (Crick, 1968; Gilbert, 1986; Joyce, 2002; Orgel, 1968). This simplifies the molecular biology necessary for the first living entities, eliminating the need for both DNA genomes and protein enzymes.

One of the key RNA molecules during an RNA World would be an RNA enzyme (ribozyme) that can catalyze replication of RNA molecules, including the replication of itself. However, such a molecule has not yet been discovered in nature or in laboratories. RNA polymers have been synthesized under prebiotic conditions via condensation of energy-rich nucleotides (Ferris et al., 1996; Monnard et al., 2003) and via polymerization of 3´-5´ cyclic nucleotides (Costanzo et al., 2009) in search of an RNA replicase.
ribozyme. However, all these experiments utilize high concentrations of activated nucleotides and result in an abundance of shorter products that do not possess the complexity and specificity necessary for catalytic RNA molecules. Therefore, these processes most likely do not favor the accumulation of biologically relevant polymers. The RNA molecules closest to an RNA replicase are the improved variants of a class I ligase originally engineered by Bartel and Szostak in 1993 using *in vitro* selection. These evolved enzymes can extend primer-template duplexes (Bartel and Szostak, 1993; Zaher and Unrau, 2007). Since they were engineered, they have undergone many modifications using different strategies. Still the number of nucleotides that can be added to the template by the most evolved version is limited to 95 nucleotides (nt), and yet the ribozyme itself is 187 nt long, and the polymerization process still requires activated nucleotides (Wochner et al., 2011).

An alternative way to create larger RNA polymers would be to recombine shorter RNA polymers obtained from prebiotic reactions. In contemporary biology, organisms exploit recombination to swap genetic information between two sources to maintain genetic diversity and repair damaged genes. Recombinase proteins, such as *RecA*, carry out this important process which involves breaking and re-forming of phosphoester bonds. The RNA World could also benefit from recombination reaction for the same reason as do modern organisms, as this type of reaction could have contributed towards the build-up of genetic diversity. Recombination could have led to quick advent of catalytic functions because it builds up longer polymers faster than other prebiotic
reactions and also could have helped in the repair of RNA polymers that are damaged frequently in the hostile environment of the early Earth (Lehman, 2003).

Recombination in the RNA World could have been achieved through recombinase enzymes. One of the ribozymes that carry out the chemistry required for recombination events is the group I intron (Figure 1). This ribozyme catalyzes its own excision from mRNA, tRNA, and rRNA precursors in a wide range of organisms, by a mechanism that involves two consecutive transesterification reactions. It results in the displacement of a RNA phosphoester bond from one dinucleotide site to another via enthalpically neutral transesterification chemistry. Also, it has been suggested that an ancient form of group I intron may date back to as early as 3.5 billion years ago (Kuhsel et al., 1990), making it a suitable candidate as a recombinase ribozyme. Overall, a ribozyme similar to the group I intron has potential to produce a polymer that is substantially larger than either of the input strands without need for high-energy intermediates, suggesting that it needed to be involved early in the origins of life on the Earth.

The recombinase ribozymes have been designed from group I introns by removing the exons, and shortening the sequences at the 5´ end so that they start with the internal guide sequence (IGS). The guanosine nucleophile required in the second step of splicing (ωG) is left on the 3´ end of the ribozyme. These shortened ribozymes can act on any exogenous substrates that contain a sequence complementary to the IGS (Riley and Lehman, 2003; Zaug and Cech, 1986). The recombinase ribozyme designed from the group I intron from the tRNA^{Ile} of the purple bacterium Azoarcus BHZ23 is shown in
The IGS sequence in this ribozyme is 5´-GUG-3´, which recognizes any RNA fragment containing the sequence 5´-CUA-3´ as substrate.

The *Azoarcus* ribozyme can efficiently catalyze the general reaction $\text{A} \cdot \text{B} + \text{C} \cdot \text{D} \leftrightarrow \text{A} \cdot \text{D} + \text{C} \cdot \text{B}$, where $\text{A}$, $\text{B}$, $\text{C}$, and $\text{D}$ are RNA sequences, and the dot indicates a scissile phosphate following a CAU trinucleotide sequence. The mechanism of recombination involves two subsequent transesterification reactions (Figure 3). The first step is a reverse splicing event, where the ribozyme binds the first substrate by base pairing between the IGS and the CAU trinucleotide, and catalyzes a transesterification reaction. This results in the covalent attachment of the 3´ portion of the substrate to the 3´ end of the ribozyme. The next step is a forward splicing event where the second substrate interacts with the ribozyme in similar manner, and the 3´ end of the new substrate attacks the scissile phosphate bond between the ribozyme and previously attached RNA fragment resulting in a recombined RNA molecule. This ribozyme has also been used to recombine substrates with secondary and tertiary structures to construct RNA molecules with functions, thereby demonstrating the viability of the recombinase ribozyme to build up genetic information with biological significance (Hayden *et al.*, 2005). Nonetheless, in order for recombinase ribozyme to have relevance to the RNA World, its origin would have to be from non-catalytic RNA molecules. This would support the notion that recombination had an important role in filling the gap between the short RNA polymers resulting from prebiotic chemistry and the more complicated ribozymes envisioned in most RNA World scenarios.
Previously, it has been demonstrated in Lehman laboratory that the IGS present in *Azoarcus* group I intron can be exploited to direct an autocatalytic assembly of the contiguous intron from four fragments of the RNA via recombination reactions (Draper *et al*., 2008; Hayden and Lehman, 2006; Hayden *et al*., 2008). The construct of the *Azoarcus* group I intron designed for its covalent assembly is shown in Figure 4A. Briefly, the ribozyme can be broken at loop regions to get 63, 39, 45, and 51 nt fragments, termed W, X, Y, and Z respectively. These four individually inactive fragments can self-assemble into a *trans* complex (using secondary and tertiary interactions) that is similar in structure to a covalently linked contiguous ribozyme (W•X•Y•Z) (Figure 4B). If these fragments have recombination tags (trinucleotide sequence CUA), at appropriate places in the molecules, then they can recombine together by means of three separate transesterification reactions leading to an autocatalytic growth of covalently linked full length molecule. This indicates the ability of recombination system derived from group I intron to utilize the short RNA fragments to construct ribozymes including itself.

Although recombination has capacity to facilitate the build-up of biologically significant molecules without much need for high-energy intermediates, there would be various obstacles that would make replication processes difficult during the early origins of life. The research discussed herein emphasizes some of the impediments that the replicating molecules could have dealt with in the RNA World and suggests how the interactions between recombining RNAs would have provided solutions to these earlier replicators. The research will be discussed in next three chapters. In Chapter 2, the ability of a single RNA fragment to play multiple roles during replication of group I intron is
demonstrated, highlighting how the limited genotypes can provide multiple phenotypes required for recombination events. In Chapter 3, a highly cooperative network of RNA replicators is demonstrated. This chapter bolsters the idea that life and evolution would be easily achieved if the RNA molecules could interact, rather than function independently and compete with each other. It stresses the benefits of cooperativity in the development of early life and helps to link a gap between selfish replicators and complex cooperative systems. And Chapter 4 demonstrates the capacity of the recombination system to perform dynamic covalent chemistry and recycle fragments of RNA. This chapter underlines the importance of recycling during the RNA World as it can avoid the waste of RNA polymers and proposes that recycling could have led to the selection of functional polymers. Overall, this research provides the RNA recombination system with different strategies to be an efficient replicating system.
Figure 1. Two-step self-splicing process in group I intron. The first step is the nucleophilic attack of the 3′-OH of an exogenous guanosine cofactor on the phosphodiester bond at the 5′ splice site, resulting in a free 3′-OH on the 5′-exon and the exogenous guanosine being attached to the 5′-end of the intron. The second step is the nucleophilic attack by the newly exposed 3′-OH of the 5′-exon on the 3′-splice site, leading to the ligation of the adjacent upstream and downstream exons and release of the catalytic intron.
**Figure 2. The Azoarcus recombinase ribozyme.** The first 8 nt has been removed from 5’ end of the ribozyme so that it begins with the IGS, which has the sequence GUG. The IGS binds RNA substrate (lower case letters and curved black line) that contains the complementary sequence CAU. Paired regions (P) are numbered as they appear 5’ to 3’ in the primary sequence. The tertiary interactions that hold the IGS and the IGS target into the active site as determined by X-ray crystallography (Adams *et al.*, 2004) are denoted using the hydrogen-bonding symbolism of (Leontis *et al.*, 2002). Nucleophilic attack by guanosine at 3’ end of the ribozyme \(\omega\) results in cleavage of the substrate directly after the “CAU”, and the 3’ end of the substrate (curved black line) gets covalently attached to the 3’ end of the ribozyme, forming the covalent intermediate necessary for recombination.
Figure 3. Recombination by the Azoarcus ribozyme. First, the IGS of free ribozyme recognizes and binds an RNA substrate \textit{A•B} (top left). The first transesterification reaction equivalent to a reverse splicing event \textit{in vivo} results in the covalent attachment of the 3′ portion of the substrate, B, to the 3′ end of the ribozyme. After the release of the 5′ piece of the substrate, A, the 5′ portion of other substrate, C, produced by the cleavage of \textit{C•D}, can bind. The next step is a forward splicing event where the 3′ end of C attacks the scissile phosphate bond between the ribozyme and attached substrate, B, and results in a recombinant RNA molecule \textit{C•D} and a free ribozyme. This ribozyme can turnover to recombine other RNA fragments.
Figure 4. Self-assembly of *Azoarcus* ribozyme from inactive RNA fragments. (A) Secondary structure of the four-piece recombination system derived from *Azoarcus* group I intron. The 198 nt ribozyme was partitioned into four fragments (W, red; X, yellow; Y, blue; Z, green) by the placement of CAU recombination tags (grey boxes) in loop regions. Substrate RNA (black line) binds via a CAU to an IGS (GUG) of the ribozyme, which catalyzes phosphotransfer of the RNA substrate to its own 3’ end. (B) Schematic of self-assembly and autocatalytic growth of a recombinase ribozyme. Four RNA fragments obtained from ribozyme, which are individually inactive, form non-covalent but catalytically active complexes through 2° and 3° interactions (bottom right). These active *trans* complexes utilize more RNA fragments as substrates to produce covalently contiguous ribozymes (top right), which then catalyze further production of like ribozymes (Hayden and Lehman, 2006).
CHAPTER 2: ONE RNA PLAYS THREE ROLES TO PROVIDE CATALYTIC ACTIVITY TO A GROUP I INTRON LACKING AN ENDOGENOUS INTERNAL GUIDE SEQUENCE

Note: This chapter is adapted from Vaidya and Lehman (2009)

ABSTRACT

Catalytic RNA molecules possess simultaneously a genotype and a phenotype. However, a single RNA genotype has the potential to adopt two or perhaps more distinct phenotypes as a result of differential folding and/or catalytic activity. Such multifunctionality would be particularly significant if the phenotypes were functionally inter-related in a common biochemical pathway. Here, this phenomenon is demonstrated by the ability of the Azoarcus group I ribozyme to function when its canonical internal guide sequence (GUG) has been removed from the 5´ end of the molecule, and added back exogenously in trans. The presence of GUG triplets in non-covalent fragments of the ribozyme allow trans-splicing to occur in both a reverse splicing assay and a covalent self-assembly assay in which the internal guide sequence (IGS)-less ribozyme can put itself together from two of its component pieces. Analysis of these reactions indicates that a single RNA fragment can perform up to three distinct roles in a reaction: behaving as a portion of a catalyst, behaving as a substrate, and providing an exogenous IGS. This property of RNA to be multifunctional in a single reaction pathway avoids the necessity of multiple genotypes and bolsters the probability that a system of self-replicating molecules could have existed in an RNA World during the origins of life on the Earth.
INTRODUCTION

Group I introns catalyze RNA phosphoester transfer reactions at specific splice sites both in vivo and in vitro. The splice site selection is precise, and relies on base-pairing interactions between the 5´ portion in P1 of the catalytic intron termed as the internal guide sequence (IGS) and a pseudo-complementary 3–6-nt region of the 3´ portion of the 5´ exon referred to as the IGS-complement (Cech, 1988; Davies et al., 1982). Base-pairing between the IGS and its complement depends on Watson–Crick pairing at most positions; however the 3´ nt of the IGS complement always forms a G•U wobble pair with the IGS to define precisely the 5´ splice site (Been and Cech, 1986; Cech, 1988; Davies et al., 1982; Michel and Westhof, 1990). For example, the IGS of the group I intron in the tRNA\textsuperscript{Ile} transcript from the purple bacterium Azoarcus can be shortened in vitro to 5´-GUG-3´, and this pairs with the complement 5´-CAU-3´ to effect splicing after the terminal U in the complement (Kuo et al., 1999; Reinhold-Hurek and Shub, 1992; Tanner and Cech, 1996) (Figure 5A).

In vitro, group I introns can behave as true enzymes, exhibiting accurate substrate specificity and multiple turnover. Notably, these enzymes often remain active when broken into non-covalent trans complexes. The work of Inoue and others has shown that the Tetrahymena, sunY, Azoarcus, Synechococcus, phage T4 td and other group I introns can retain catalytic efficiency when fragmented into several non-covalent pieces and then reconstituted in trans via base-pairing and tertiary interactions (Beaudry and Joyce, 1990; Chowrira et al., 1995; Doudna and Cech, 1995; Doudna et al., 1989; Ikawa et al., 2000a; van der Horst et al., 1991). These studies have isolated the regions of the ribozyme that
are necessary to carry out the various catalytic events of self-splicing, a multi-step process that involves two transesterification reactions plus often a hydrolytic cleavage reaction. While trans-catalysis by reconstructed group I introns—including that from Azoarcus—is possible, complete removal of the IGS or the catalytic core will abolish splicing activity. For group I introns, and the Azoarcus ribozyme in particular, deletion of the P1 and P2 elements, along with the IGS, results in a molecule that cannot self-splice (and thus recombine exogenous RNAs), even though site-specific hydrolysis at the 3′ splice site is still possible because P1 (and hence the IGS) is only required for 5′ splice site selection (Ikawa et al., 2000b; van der Horst and Inoue, 1993).

Here, we have used the Azoarcus group I intron, with 5′-GUG-3′ as its IGS as a model system to show that if the canonical IGS is missing from the ribozyme, another RNA fragment containing GUG can act in trans to rescue transesterification activity. In fact, we show that one RNA fragment can play as many as three distinct roles in the covalent self-assembly of a self-replicating ribozyme. These results have bearing on models of the origins of genetic information in which multiple RNA oligonucleotides may have had to cooperate in order to allow construction of a catalytic entity.

RESULTS AND DISCUSSION

A group I intron without an IGS

To create an IGS-less (‘blind’) version, of the L–8 Azoarcus ribozyme, the first 30 nt of the naturally occurring form of the 205-nt intron were removed, resulting in an L–30
construct lacking the first occurrence of the GUG triplet (termed GUG1), which is the canonical IGS for the ribozyme. Without this triplet, transesterification activity should be abolished. Previously it has been shown that the Azoarcus ribozyme can covalently self-assembly through recursive and autocatalytic recombination reactions after having been fragmented into four roughly equally sized pieces termed W, X, Y and Z (Hayden and Lehman, 2006). The removal of the first 22-nt in W results in a new fragment, here termed V (Figure 5A). Of interest is that the remaining 175-nt of the wild-type version of this ribozyme contain four additional GUG triplets, denoted GUG2, GUG3, GUG4 and GUG5 (Figure 5A and B). By chance, any given triplet such as GUG should appear fewer than three times in 175-nt. Regardless, these ‘exogenous’ GUGs could potentially complement the missing canonical IGS if they were to work in trans. Such activity would be similar to the manner in which fragments of the Tetrahymena ribozyme can non-covalently assemble to restore activity, although with each RNA presumably only adopting one role (Beaudry and Joyce, 1990; van der Horst et al., 1991).

The efficacy of an exogenous IGS to rescue a blind Azoarcus ribozyme was first tested using an assay in which the L–30 ribozyme V•X•Y•Z was challenged to perform trans-splicing on an external substrate, where ‘•’ signifies a covalent phosphoester bond (Figure 6A). In these reactions, specific recognition of the substrate at its IGS complement CAU would require the articulation of a GUG triplet into the active site of the V•X•Y•Z construct. Five substrates were assayed, three of which possess at least one GUG sequence (h•V, h•Y and h•Z, where h refers to a 5-nt head, GGCAU), while two of which do not (h•X, and SNL-1a=5´-GGCAU•AAAUAUUUAUAAUAUAAUAUAAUAUAAUAUAAUA-3´).
Measurable *trans*-splicing activity was observed only for the three GUG-containing substrates, with \( \text{Z} \), possessing two such sequences, exhibiting by far the most activity (Figure 6A). This activity diminishes when the GUGs are mutated to CUG in a dosage dependent manner (Figure 7). While it is conceivable that the \( \text{VXYZ} \) ribozyme itself could be supplying the exogenous IGS by partially unfolding in partnership with a folded and catalytic complex, the differential response to exogenous substrates suggests that this is not occurring under these assay conditions. The more likely scenario involves the utilization of a GUG triplet in one molecule of the substrate itself to direct splicing at the splice site of another molecule of the substrate. The case in which GUG\( \text{5} \) of \( \text{Z} \) is performing this role is diagrammed in Figure 5B. The product sizes observed during gel electrophoresis (Figure 6A) are those expected should transesterification occur following the single CAU sequence present in each substrate, and these were confirmed by RT–PCR using reverse primers specific for each substrate.

**Covalent self-assembly of a blind ribozyme by multifunctional RNA fragments**

Next, the ability of a single RNA fragment to assume multiple roles in an RNA reaction network was investigated using covalent self-assembly systems. Here, two pieces of the *Azoarcus* L–30 ribozyme were co-incubated at 2 \( \mu \text{M} \) concentrations. Covalent self-assembly requires that non-covalent complex initially catalyze a series of recombination reactions leading to the ‘full-length’ covalent ribozyme, which can then feed back in an autocatalytic fashion (Hayden and Lehman, 2006; Hayden *et al*., 2008). In the current experiments, an additional requirement exists that one fragment behave as an exogenous
IGS, with the consequence that the overall yield of covalent ribozyme is expected to be low. Nonetheless, assembly was observed in a variety of systems. When V was incubated with h•X•Y•Z, where h refers to the 5-nt sequence GGCAU, self-construction of the V•X•Y•Z ribozyme was observed in 0.1–1.3% yield, and when V•X•Y was incubated with h•Z the yield was 1.0–3.5% (Figure 6B). Trace amount of product was also seen when V•X was incubated with h•Y•Z (data not shown). The higher yields seen when the Z fragment was used in isolation is consistent with this molecule’s possession of two GUG triplets, such that either one could be serving as an exogenous IGS. Highest self-assembly yields were obtained at 42°C, as opposed to the 48°C optimum for self-assembly with the U fragment in place (Hayden and Lehman, 2006), suggesting that a multifunctional RNA fragment containing an exogenous IGS is binding to the catalytic core through a set of relatively weak hydrogen-bonding interactions.

**Mechanism of covalent self-assembly**

Recombination in these self-assembly reactions can proceed through a one-step (tF2) or a two-step (R2F2) mechanism (Draper et al., 2008), and the junctions between the two fragments were designed here in two alternative ways to favor one or the other mechanism. However, because the tF2 mechanism requires a pre-formed RNA duplex to bind to the IGS, rather than a single RNA strand as in the R2F2 mechanism, the scheme as portrayed in Figure 5B that depends on an exogenous IGS seemed less plausible with the tF2 mechanism. To test this hypothesis, the assembled RNAs at the size indicated by the arrow in Figure 6B were excised, converted to DNA by RT–PCR, and then cloned
and subjected to nucleotide sequence analysis. The two mechanisms can be distinguished by the fact that the \( tF2 \) mechanism leaves a characteristic 3–4-nt insertion, while the \( R2F2 \) does not (Draper et al., 2008; Hayden and Lehman, 2006). All clones analyzed in all experiments failed to exhibit insertions, and thus it can be concluded that two sequential recombination reactions using single-stranded substrates (\( R2F2 \) mechanism) are required when an exogenous IGS is utilized, consistent with the depiction in Figure 5B. In fact, most clones possessed deletions of 3–12 nt at the splice junction, suggesting that splicing is often sloppy, occurring a few positions 3´ of the canonical CAU target site. Some of these products are even visible in the gel shown in Figure 6B, and are further evidenced when the 5´ splicing products are examined (Figure 8).

One way in which mispairing can occur under the high salt conditions employed in self-assembly (100mM MgCl\(_2\)) is when the IGS pairs with a single-nucleotide variant of CAU (Draper et al., 2008). In Figure 9, an example of this is diagrammed, whereby the exogenous IGS is pairing with the triplet CGC of the substrate, leading to a 4-nt deletion that was commonly detected in the clones. It should be noted that the pairing between the 5´ G of the IGS and the 3´ C in this triplet is likely not a canonical \( cis \) Watson–Crick pair because the base-pairing surface of this guanosine of the IGS is participating in tertiary interactions with the \( J4/5 \) bulge (Figure 5A), at least in the wild-type \( Azoarcus \) model deduced from X-ray crystallography (Adams et al., 2004). In any event, use of an exogenous IGS leads to somewhat promiscuous splicing, as can be seen with the \( Tetrahymena \) ribozyme (Doudna et al., 1989). It also can lead to promiscuous self-assembly events, with the consequence that should autocatalytic feedback be
significant in this system, then the RNA denoted at the arrow in Figure 6B should be the result of a selection process for catalytically proficient molecules. As a rudimentary test of this notion, three clones of self-assembled ribozymes (with deletions of 3, 4 or 8 nt) were tested for splicing activity and all were found to be active for transesterification (Table 1). Of interest is that the shortest of these clones, one that exhibits an 8-nt deletion at the Y•Z junction is clearly the most active and would be expected to emerge from a selection process. The deletion in this clone disrupts the tetraloop receptor in P8 (Figure 5A), but in the absence of the P2 tetraloop that normally docks at that location, this deletion becomes less deleterious.

**Multifunctionality in catalytic RNAs**

Together, these data allow a model in which RNA multifunctionality is a requisite aspect of the construction of self-replicating catalytic RNA from its component fragments (Figure 10). In this model, one RNA sequence (h•Z in this case) is actually performing three distinct roles. First, it is operating as part of a catalytic recombinase ribozyme, either in *trans* as a non-covalent partner with the remainder of the catalyst, or in *cis*, as a covalent element within a ribozyme. Second, it is behaving as an exogenous IGS, bringing a requisite GUG into the catalytic core of a ribozyme complex. The multifunctional fragment containing the exogenous IGS is held likely in place by tertiary interactions (Chowrira *et al.*, 1995) such as those postulated in Figure 9. And third, it is being utilized as a substrate for recombination, in which its CAU sequence (or a low-error variant) is targeted by an IGS for transesterification to another RNA molecule.
A fundamental confirmation of the notion that $h\cdot Z$ is playing three distinct roles was obtained when the stoichiometry was varied between $V\cdot X\cdot Y$ and $h\cdot Z$ in self-assembly reactions (Table 2). When the concentration of $h\cdot Z$ was kept constant at 1 μM, the optimum ratio of $V\cdot X\cdot Y$ to $h\cdot Z$ was in fact 1:3, which gave a significantly better yield of $V\cdot X\cdot Y\cdot Z$ after 4 h than either lower (1:2) or higher (1:5) amounts of $h\cdot Z$ (one-tailed $t$-tests; $P=0.03$ and 0.002, respectively).

Another test of this model was provided when either the IGS, or its complement in the $h$ portion of substrates, was mutated and loss-of-activity assays were performed. One approach was to track covalent self-assembly between $V\cdot X\cdot Y$ and $h\cdot Z$ with matched and mis-matched pairings between potential exogenous IGS triplets in the $Z$ portion of $h\cdot Z$ and the IGS complement in the $h$ portion of $h\cdot Z$ (Figure 11). When $h$ was changed from $\text{GGCAU}$ to $\text{GGCUU}$, self-assembly was completely abolished. However, this required a second mutation at position 182 to alter a natural GAG to GUG to destroy a fortuitous IGS that would function with CUU. This A182U mutation by itself does not abolish activity (data not shown). Notably, when this second mutation is not made, the self-assembly activity is fully rescued, presumably by the GAG functioning as the IGS and pairing with the CUU in the $h$ of $h\cdot Z$ (Figure 11).

A second approach was simply to mutate exogenous GUG triplets, and test for loss-of-function with the canonical $\text{GGCAU}$ in $h$ (Table 3). Triplets 2, 4 and 5 were mutated to GUC, CUG and CUG, respectively. GUG3 was not mutated because it lies in J6/7 and has been shown previously to be essential for catalytic activity (Rangan et al., 2003). When GUG2, GUG4, or GUG5 was mutated, the ability of the ribozyme to self-
assemble covalently from two fragments was abolished, although it is not clear if this is a consequence of the effect this mutation has on the ribozyme as a whole. However, mutations in GUG4 and GUG5 are not inhibitory to the wild-type (W•X•Y•Z) Azoarcus ribozyme, and have minimal effect on the L−30 construct (V•X•Y•Z). Yet these two mutations severely inhibit both the covalent self-assembly reaction (Table 3) and transesterification (Figure 7), implicating them as operating as exogenous IGS triplets in a cooperative RNA–RNA interaction.

Evolutionary implications

The RNA World scenario posits the existence at some point in time on the prebiotic Earth the existence of one or a few RNA-like molecules with the capacity to carry out all catalytic and information transfer events needed for life (Crick, 1968; Gesteland, 2006; Gilbert, 1986; Orgel, 1968). To envision catalytic RNAs that are self-replicating requires an evolutionary scenario by which larger and more complex ribozymes develop from smaller and simpler structural motifs (Levy and Ellington, 2001). In the case of group I introns, which can perform RNA recombination reactions that could have been of evolutionary importance (Lehman, 2008; Riley and Lehman, 2003; Zaug and Cech, 1986), piecing together the entire catalytic complex may have required physical separation of key structural elements, including the IGS.

The trans-splicing activity of the blind Azoarcus ribozyme is in accordance with a previous study that demonstrated that if the IGS were disarticulated from the catalytic core on a separate fragment, and then an effectively full-length ribozyme could
polymerize oligonucleotide sequences on its 3’ end (Chowrira et al., 1995). In that study an exogenous IGS (termed the external guide sequence, or EGS) required the GUG from P1 plus an additional 7 nt from P10 of the natural bacterial intron. The observations in the current study that the IGS can not only be externalized but also embodied in another portion of the ribozyme itself and shortened to GUG beg the question of what the minimum possible IGS/EGS could be. It will be interesting to test whether the trinucleotide GUG itself would impart any splicing activity to a blind ribozyme. The minimum length is unlikely to be <3 nt, because then almost all specificity would be lost. Provocatively, the IGS complement used by GUG in this system is CAU, which is the biochemical precursor to the L(lysidine)AU anticodon that is used to specify isoleucine in bacterial translation (Osawa et al., 1992). The embedding of the Azoarcus self-splicing intron in an isoleucyl-tRNA gene evokes a direct relationship between an IGS/EGS and the origin of the genetic code, as alluded to by Shub upon its original discovery (Reinhold-Hurek and Shub, 1992).

The ability of RNA molecules to act multifunctionally facilitates our ability to envision the origins of complex catalytic function and lends support to the RNA World scenario. RNA molecules are excellent candidates for this, as they clearly can fold variably, often leading to inhomogeneous populations of three-dimensional shapes (Ancel and Fontana, 2000; Schmitt and Lehman, 1999; Uhlenbeck, 1995). In fact, Bartel and colleagues have previously described a single RNA sequence that can fold into two different and essentially unrelated ribozymes (Schultes and Bartel, 2000), while here we report a single sequence with three distinct functions in the same reaction pathway.
Primordial RNA molecules may have had to interact in a cooperative fashion to participate in hypercycles (Eigen and Schuster, 1977) or autocatalytic sets (Kauffman, 1993) in which the system as a whole benefits from the multiplicity of interactions among its components. The advent of such diversity would have its probability heightened if each component could play multiple roles such that fewer molecular types were required initially.

MATERIALS AND METHODS

RNA preparation

RNAs were either purchased from IDT (Coralville, IA) or prepared by run-off transcription from double-stranded DNA templates constructed through recursive gene synthesis. RNAs were gel purified and desalted prior to use. Salts and buffers were made from the highest purity available (Sigma-Aldrich) and all water used was nuclease free (Ambion). Barrier pipette tips and other strict contamination controls, including the use of dedicated rooms and isolation hoods to purify primers and set up gene synthesis and PCR reactions, were always used to ensure correct sequence identities of all RNAs employed in experimentation.

Trans-splicing assays

The ability of Azoarcus ribozyme constructs to perform trans-splicing was assayed by incubating 2 μM of an enzyme (intron only) complex with 2 μM of a substrate
oligonucleotide with or without the 5´-GUG-3´ IGS triplet somewhere within its sequence. Assays were performed at 48°C in 100mM MgCl₂ and 30mM EPPS buffer (pH 7.5). Reactions were carried out for 0–60 min in 200 μl or 600 μl microcentrifuge tubes and then quenched by the addition of an equal volume of gel-loading solution containing 8M urea, SDS, 200mM EDTA and bromophenol blue dye. The RNAs were heat denatured at 80°C for 4 min, and then immediately electrophoresed through 8% polyacrylamide / 8M urea gels. In most cases, the V (or W) fragment was 5´-end-labeled with γ[³²P]•ATP and OptiKinase (USB) and then gel-purified (Burton et al., 2009) prior to use. This allowed visualization of the products via phosphorimaging with a Typhoon 9200 instrument (GE Healthcare). In other cases the 3´ portion of the substrate molecule was radiolabeled using the method of Huang and Szostak (Huang and Szostak, 1996). In some cases no radioactivity was employed, and all oligomers were visualized by SYBR Green II staining, although when quantification was needed, this technique was not used.

Self-assembly experiments

Ribozyme covalent self-assembly from two or three oligomers was performed as described previously (Draper et al., 2008; Hayden and Lehman, 2006; Hayden et al., 2008). Briefly, RNA oligomers were incubated together at 42–48°C at a final concentration of ≤2 μM each. All reactions contained a final concentration of 100mM MgCl₂ and 30mM EPPS buffer (pH 7.5) unless otherwise indicated. Reactions were carried out for 1–6 h with visualization methods as described above.
**Genotyping**

Full-length RNA covalent constructs were identified by comparison to a bona fide *Azoarcus* RNAs (with or without the 5’ U section) run as size controls. The bands corresponding to self-assembled RNAs were carefully excised from the gel and subjected to reverse transcription using the primer T20a (5’-CCGGTTTGTGACTTTG GCC-3’), which targets the 3’ portion of the Z fragment. One-twentieth of these reactions was used to seed PCR reactions employing T20a and W3’-h (5’-TAATACGACTCAGCAAGGGATGGTG-3’) primers, the latter being specific for the 5’ portion of V. The PCR products were cloned into the vector pJET1.2 (Fermentas) and transformed into *Escherichia coli*. Individual colonies were picked as templates for colony PCR reactions employing the primers pJET1.2-F and pJET1.2-R (Fermentas), which generate products of ~310 bp [= the insert size (~192 bp) plus about 120 bp]. Products of the correct size were genotyped using BigDye (v.3) cycle sequencing chemistry and a Prism 3100 (ABI) instrument.
Figure 5. Schematic of the partitioning of the Azoarcus ribozyme into fragments. (A) The 197-nt source molecule was partitioned into five fragments (U, gray; V, red; X, yellow; Y, blue; Z, green) such that catalytic activity in trans, plus covalent self-assembly could be assayed. Removal of the U fragment from the system at the location indicated by the arrow leaves an L−30 construct. A substrate oligomer (lower-case letters=head; black line=tail) binds via a 5′-CAU-3′ to the IGS (GUG) of the ribozyme, which catalyzes phosphotransfer of the tail to its own 3′ end. The native IGS of the ribozyme is denoted by GUG₁, while the four exogenous GUG triplets that occur naturally in the remainder of the ribozyme are gray boxes denoted 2–5 (in circles). The tertiary interactions that hold the IGS and the IGS complement into the active site as determined by X-ray crystallography (Adams et al., 2004) are denoted using the hydrogen-bonding symbolism of Leontis et al. (Leontis et al., 2002). (B) Schematic of how a ribozyme complex (either as a covalently contiguous molecule or as several fragments cooperating in trans) can perform catalysis in the absence of the U fragment that contains the native IGS. Here, the use of an exogenous IGS (GUG₅) present in the h•Z fragment to bind to the IGS complement CAU is depicted.
Figure 6. Activity assays of ribozymes lacking a native IGS. (A) Splicing activity of the *Azoarcus* L–30 construct V•X•Y•Z when supplied with an exogenous substrate that either contains the IGS sequence GUG (*h•V*, *h•Y* and *h•Z*) or that does not (*h•X* and SNL-1a). Upon splicing, the 5′-radiolabeled ribozyme is expected to append the 3′ portion of the substrate to its own 3′ end, producing a product of the sizes indicated by the small arrows; no splicing is visible with *h•X* or SNL-1a, which should give products of lower molecular weight than the product with *h•V*. (B) Bipartite covalent self-assembly reactions. In each set, the V-containing fragment is 5′ radiolabeled. The R2F2 reactions were designed with splice junctions (V-X or Y-Z) that can only form a covalent bond via a two-step recombination reaction, while the tF2 reactions were designed with splice junctions that can form a covalent bond through either a two-step recombination or through a one-step recombination (Draper *et al.*, 2008). This accounts for the slight size difference between the reactants or products in each system. The covalent self-assembly product V•X•Y•Z (~175 nt) is indicated.
Figure 7. *Trans*-splicing assays of the IGS-less ribozyme V•X•Y•Z. In these reactions, 5’-end-labeled V•X•Y•Z was incubated with three versions of the h•Z fragment, which behaves both as a substrate for *trans*-splicing and also provides an exogenous IGS sequence to the ribozyme in *trans*. Reactions were run for the times indicated and then the products were electrophoresed through an 8% polyacrylamide / 8 M urea gel. Mutations of the IGS sequence from GUG to CUG in GUG4 and GUG5 (Figure 5A) greatly diminish *trans*-splicing activity.
Figure 8. Analysis of splice-site specificity during covalent self-assembly in the absence of a canonical IGS. In these reactions, the 5´end of the head portion of the h•Z fragment was radiolabeled and reacted with the IGS-less fragment V•X•Y for the time indicated, and the products were electrophoresed through a 20% polyacrylamide / 8 M urea gel. Intermediates produced during covalent self-assembly of the product V•X•Y•Z and the remaining h fragment resulting from transesterification can be observed. If splicing occurs at the typical splice site following the CAU in the 5-nt head sequence (5´-GGCAU-3´), then a 5-nt product should result (indicated). Mis-splicing produces head fragments of varying lengths. The left-hand lanes on the gel depict results from self-assembly in which the Y-Z junction was designed to favor the two-step R2F2 mechanism, while the right-hand lanes depict results from self-assembly in which the Y-Z junction was designed to favor the one-step tF2 mechanism (Draper et al., 2008). Cloned RT-PCR V•X•Y•Z products from the R2F2 reaction, if from a mis-splicing event, typically contained deletions of four nucleotides (see text), which would result in a nine-nucleotide head product (5 + 4 = 9) as a result of splicing events four nucleotides downstream of the CAU, and these can be clearly seen on the gel. Cloned RT-PCR V•X•Y•Z products from the tF2 reaction, if from a mis-splicing event, typically contained deletions of three or four nucleotides (see text), which would result in eight- or nine-nucleotide head products, seen on the gel.
Figure 9. Model of utilization of an exogenous IGS. In this example, the \( h \cdot Z \) RNA molecule is multifunctional, performing three discrete roles: acting as the 3’ end of the ribozyme, acting as a substrate for transesterification, and acting as an exogenous IGS. In the last case, the \( h \cdot Z \) molecule must be held in place in the catalytic core of the ribozyme complex \( \text{via} \) tertiary hydrogen-bonding interactions. Some potential examples of these are indicated in the diagram, although no solid evidence exists for any of these interactions. In addition, the loose interaction between the fragment supplying the exogenous IGS and the remainder of the ribozyme apparently can lead to mis-pairing between the IGS and the IGS complement (see text); here GUG5 is depicted as acting as the exogenous IGS and binding to CGC instead of CAU, which would lead to a 4-nt deletion in the splicing product.
Figure 10. Model in which a recombinase ribozyme can covalently self-assemble from two RNA fragments despite the lack of a native IGS in the catalytic complex. Here, this self-assembly requires the $h\cdot Z$ fragment to be multifunctional and exhibit three distinct phenotypes.
Figure 11. Test of the ability of nucleotide triplets in $h•Z$ to perform as exogenous IGSs. Here, covalent self-assembly was assayed when 2 μM $V•X•Y$ was incubated with 2 μM $h•Z$ for 2 h at 42°C. In the ‘wild-type’ scenario, neither the $h$ nor the $Z$ portions of $h•Z$ contained any mutations, as in Figure 6B. Self-assembly is achieved when an exogenous IGS in the $Z$ portion is utilized, such as GUG5 at positions 183–185 (depicted). In the ‘mismatch’ scenario, the $h$ was mutated from GGCAU to GGCUU, while an A182U mutation was simultaneously made to destroy a fortuitous exogenous IGS triplet of GAG. Self-assembly was abolished, but could be fully restored in the ‘rescue’ scenario when this second mutation was not present.
Table 1. Activity assays of three clones derived from ribozymes self-assembled from RNAs lacking the canonical IGS. Table values are relative transesterification yields after 1 h with the indicated substrate, normalized to that of the V•X•Y•Z construct, set at 100%. Brackets indicate a deletion relative to the wild type.

<table>
<thead>
<tr>
<th>Clone</th>
<th>Sequence at Y–Z junction</th>
<th>Relative yield using h•Z substrate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-30 wild type</td>
<td>...CUAAGGCAU•ACGCUAUUGUGAAGG...</td>
<td>100</td>
</tr>
<tr>
<td>1398: Δ4nt</td>
<td>...CUAAGGCAU[ ]UAUGUGAAGG...</td>
<td>32</td>
</tr>
<tr>
<td>1401: Δ8nt</td>
<td>...CUAAGGCA[ ]CGGUGAAGG...</td>
<td>63</td>
</tr>
<tr>
<td>1403: Δ3nt</td>
<td>...CUAAGGCACAU[ ]UAUGUGAAGG...</td>
<td>60</td>
</tr>
</tbody>
</table>
Table 2. Covalent self-assembly yields in different ratio of V•X•Y and h•Z after 4 h.

Assays were performed with V•X•Y that were 5´-end-labeled with γ[^32P]ATP. Data based on results from three independent trials with [h•Z] fixed at 1 μM.

<table>
<thead>
<tr>
<th>V•X•Y : h•Z</th>
<th>Percent yield of V•X•Y•Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 : 1</td>
<td>1.11 ± 0.01</td>
</tr>
<tr>
<td>1 : 2</td>
<td>1.24 ± 0.10</td>
</tr>
<tr>
<td>1 : 3</td>
<td>1.50 ± 0.06</td>
</tr>
<tr>
<td>1 : 5</td>
<td>1.17 ± 0.03</td>
</tr>
<tr>
<td>IGS mutations</td>
<td>Relative yield in V•X•Y + h•Z self-assembly reaction (%)</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------------------------------------------------------</td>
</tr>
<tr>
<td>(Wild-type)</td>
<td>100</td>
</tr>
<tr>
<td>GUG2</td>
<td>0</td>
</tr>
<tr>
<td>GUG4</td>
<td>0</td>
</tr>
<tr>
<td>GUG5</td>
<td>0</td>
</tr>
<tr>
<td>GUG2+4</td>
<td>0</td>
</tr>
<tr>
<td>GUG2+5</td>
<td>0</td>
</tr>
<tr>
<td>GUG4+5</td>
<td>0</td>
</tr>
<tr>
<td>GUG2+4+5</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3. Activity assays of reaction systems in which mutations were engineered into the exogenous IGS triplets. Assays were performed with V•X•Y that were 5´-end-labeled with γ[^32]P•ATP. Table values are relative bipartite covalent self-assembly yields after 3 h (second column) or splicing assays using an exogenous substrate after 1h (third or fourth columns), normalized to that of the unmutated construct, set at 100%. Trace = less than 1% product detected. 0% = no V•X•Y•Z product detected.
CHAPTER 3: SPONTANEOUS NETWORK FORMATION AMONG COOPERATIVE RNA REPLICATORS

Note: This chapter is adapted from Vaidya et al. (2012)

ABSTRACT

The origins of life on Earth required the establishment of self-replicating chemical systems capable of maintaining and evolving biological information. In an RNA World, single self-replicating RNAs would have faced the extreme challenge of possessing a mutation rate low enough both to sustain their own information and to compete successfully against parasites with limited evolvability. Thus theoretical analyses suggest that networks of interacting molecules were more likely to develop and sustain life-like behavior. Here we show that mixtures of RNA fragments that self-assemble into self-replicating ribozymes spontaneously form cooperative catalytic cycles and networks. We find that a specific 3-membered network has highly cooperative growth dynamics. When such cooperative networks are competed directly against selfish autocatalytic cycles, the former grow faster, indicating an intrinsic ability of RNA populations to evolve greater complexity through cooperation. We can observe the evolvability of networks through in vitro selection. Our experiments highlight the advantages of cooperative behavior even at the molecular stages of nascent life.
INTRODUCTION

Many lines of evidence point to an “RNA World” as a plausible stage in the development of life because RNA simultaneously possesses evolvability and catalytic function (Gilbert, 1986; Joyce, 1989; Orgel, 1968). An RNA organism that could evolve in such a fashion is prone to have been one of the Earth’s first life forms, and the search for an RNA autoreplicase molecule is underway (Johnston et al., 2001; Wochner et al., 2011; Zaher and Unrau, 2007). In fact Darwinian-type evolution relies on individual genotypes that compete for survival and reproduction in a fitness landscape. Yet the transition from a prebiotic chemical soup to this stage of life is not understood. Several authors have proposed that the most primitive life thrived less on discrete genotypes and instead on collections of molecular types more subject to systems chemistry than to straightforward selection dynamics (Eigen and Schuster, 1977; Eschenmoser, 2007; Kauffman, 1993; Levy and Ellington, 2001; Nowak and Ohtsuki, 2008; Sievers and Von Kiedrowski, 1994; Szathmáry, 2006; von Kiedrowski et al., 2010). In particular, Eigen and others suggested that webs of functionally linked, genetically related replicators were required in the earliest phases of life’s appearance to prevent informational decay (the so-called error catastrophe) (Eigen, 1971; Eigen and Schuster, 1977; Kun et al., 2005; Smith, 1979). We sought an empirical demonstration of such replicator networks that could illuminate critical features of this early stage of life. We chose ribozymes because these molecules can evolve outside of an organismal context, construct other RNAs, exhibit self-sustained reproduction, and explore sequence space in efficient ways (Doudna and Szostak, 1989; Ellington and Szostak, 1990; Joyce, 2007; Lincoln and Joyce, 2009). But
their ability to form catalytic networks capable of expanding as predicted from theory has not yet been shown, despite the observation that collections of nucleic acids have the potential to manifest complexity (Levy and Ellington, 2003; Qian and Winfree, 2011; Sievers and Von Kiedrowski, 1994). Simulations show that molecular networks should arise, evolve, and provide a population with resistance against parasitic sequences (Szathmáry, 2006). However these results are robust within structured environments such as cells or on grids, but are less so in a solution phase. Recent experimental work in vitro has been very successful at demonstrating simple ecologies (Lee et al., 1997a; Lee et al., 1997b; Voytek and Joyce, 2009), reciprocity between two species (Kim and Joyce, 2004; Sievers and Von Kiedrowski, 1994), sustained exponential growth via cross catalysis (Lincoln and Joyce, 2009), and even scalable computational prowess (Qian and Winfree, 2011). A limit to the empirical efforts to date has been expanding past reciprocal interactions between two species to prebiotically relevant systems that have the capacity to increase their complexity by expanding to three, and then more, members (Eigen and Schuster, 1978; Lee et al., 1997a). Previous efforts have been hampered by the use of systems in which the recognition domain in the catalyst and the target domain in the substrate are co-located in each replicator, preventing networks of more than two members from forming. If this molecular feature could be circumvented, larger networks could be manifest within RNA populations in the test tube and help demonstrate a potential escape from the error catastrophe problem that tends to plague selfish systems.
RESULTS

The Azoarcus ribozyme in a 3-membered cooperative cycle

The ~200 nt Azoarcus group I intron ribozyme can be broken into fragments that can then covalently self-assemble by catalyzing recombination reactions in an autocatalytic fashion (Hayden and Lehman, 2006; Hayden et al., 2008). By allowing variation in the sequence recognition mechanism by which this assembly occurs, that is provided by the 3-nt internal guide sequence (IGS) at the 5´ end of the ribozyme, many such autonomously self-assembling ribozymes become possible. We sought to determine if these ribozymes could display cooperative behavior if their IGS sequences target the assembly of other ribozymes but not themselves.

To create a cooperative network, we fragmented the Azoarcus ribozyme into two pieces in three different ways with the intent of observing how they could spontaneously reassemble via intermolecular cooperation (Figure 12). Here we manipulated the IGS (canonically GUG) and its target triplet to generate both matched and mismatched partners. For example, the canonical IGS in the Azoarcus ribozyme is 5´–GUG–3´, which matches with the complement 5´–CAU–3´ as a consequence of a requisite G-U wobble preceding the splice site. If the ribozyme is broken into two pieces, termed W•X and h•Y•Z, then these two fragments will rapidly covalently self-assemble and self-replicate autocatalytically. This requires that the triplet GUG be placed at the 5´ end of W as the IGS, while CAU be placed not only at the 3´ end of X but also in the 5-nt “head” region h prior to Y as the target (i.e., GUGW•XCAU and ggCAU•Y•Z). However, mutation of GUG to GAG would create a mismatch with target CAU, such that GAGW•XCAU and ggCAU•Y•Z.
would self-assemble only minimally. We mixed various IGS and target pairs in two-piece constructs to test the ability of mismatched pairs to promote self-assembly (Figure 13). From these data, we chose three mismatched pairs that exhibit relatively little autocatalysis: GUG/CGU, GAG/CAU, and GCG/CUU. In each case a single mutation within the IGS or the IGS target produces a new triplet that functions poorly in two-fragment self-assembly. These crippled pairs are denoted I₁, I₂, and I₃, respectively, meaning that they are informational subsystems, albeit weakly autocatalytic.

However, we chose the triplet pairs in such a way that when the three systems are mixed together, they should constitute a cyclical cooperative network (Figure 12B). In this network, the output of one subsystem can catalyze the replication of the next one in the cycle. This occurs because the IGS of one subsystem is matched to the target in the next subsystem, proceeding clockwise, and the physical separation of the IGS and its target allows for cycles of more than two members. When the six RNAs (W, h•X•Y•Z, W•X, h•Y•Z, W•X•Y, and h•Z) are allowed to fold together and be co-incubated in equimolar ratios, we expect the subsystems first to form non-covalent versions of ribozymes, and then catalyze the formation of covalent versions of the next ribozyme in the cycle.

To test whether cooperation between enzymes occurred in this system, we took several approaches. First, in order for the cycle to exhibit positive feedback (Eigen and Schuster, 1977), there should be a distinct advantage to being a covalently contiguous ribozyme (Eᵢ), as opposed to remaining fragmented (Iᵢ). In other words, once covalent ribozymes (E₁, E₂, or E₃) are formed, they should further promote synthesis of their target
ribozymes, at faster rates than would the non-covalent versions I₁, I₂, and I₃. When we tested each in isolation, we found that the Eᵢ ribozymes indeed recombined their respective target substrates into products 1.3–6.3 X more than the Iᵢ versions when assayed separately (Figure 14). Second, by examining each subsystem in isolation or in pairs, we could compare the relative strengths of autocatalysis (Eᵢ synthesizing Eᵢ), cross-catalysis (Eᵢ₊₁ synthesizing Eᵢ), and what we anticipated to be the most efficient, direct catalysis (Eᵢ synthesizing Eᵢ₊₁). When we incubated just the two RNAs from any one subsystem, such as I₂, alone, there is minimal synthesis of the corresponding ribozyme E₂; after a few hours roughly 0.1% of W•X is converted into W•X•Y•Z. This background level of autocatalytic synthesis is a consequence of residual catalytic activity available to a mismatched IGS and IGS target, e.g., GAG with CAU (Draper et al., 2008). The observed autocatalysis shows that each Iᵢ subsystem has modest information-replication potential in isolation, but it is severely limited. Likewise when we co-incubated the four RNAs of two subsystems together, the cross-catalytic synthesis of the ribozyme corresponding to the preceding subsystem in the cycle is similarly poor, again hindered by an IGS/IGS-target mismatch (Figure 15). After only one hour of incubation, the yield of E₃ from 0.5 μM I₃ is 0.10 ± 0.02 % (autocatalysis), and the yield of E₃ from 0.5 μM I₃ and 0.5μM E₁ is 0.7 ± 0.06 % (cross-catalysis), but the yield of E₃ from 0.5 μM I₃ and 0.5 μM I₂ is 13 ± 0.5 % (direct catalysis) (data not shown; errors given as s.e.m.). These differences are all statistically significant as measured by t-tests several planned comparisons (P < 0.001). From these data we could see that direct catalysis is
significantly more efficient than catalysis resulting from mismatched IGS sequences and their targets.

Next, an important consequence of cooperation should be that the synthesis of full-length RNAs rapidly escalates when we co-incubate all six RNAs of all three subsystems, and it does. The composite yield of full-length RNA after 16 h when I1, I2, and I3 are mixed is 125-fold higher than the sum of the yields of the three subsystems reproducing in isolation (Figure 16), and this enhancement can be readily visualized after shorter periods of time via electrophoresis (Figure 15). This higher yield demonstrates the advantage of a closed cycle of cooperative replicators, where the concentration of each member is increased through direct catalysis from the preceding subsystem of the cycle. Note that each subsystem grows at different rates, with E3 growing the most quickly (Figure 17). The synthesis of E3 by E2 is superior to that of the other two ribozymes, a consequence of a junction effect in that the non-covalent version of the enzyme (I2) is nearly as efficient as the covalent version (E2). It could also be because certain IGS-IGS target pairs can be more efficient (Draper et al., 2008). Importantly, we can also detect two-step (relayed) cooperativity by comparing the yields with and without the intervening enzyme. In the case of E1 for example, after 4 h the increase in yield of E1 upon addition of I2 to I1 with I3 present is 2.5%, while the increase in yield of adding I2 to I1 without I3 present is only 0.02%, showing the operation of E2 through E3 onto E1 (Table 4). Actually, doping the complete I1 + I2 + I3 reaction mixture with any Ei enzymes preferentially heightens direct catalysis in the short term, helping to underscore the
interdependency of the subsystems (Figure 18), although in a cyclical system doping with products can actually impede synthesis as time goes on.

As another way to observe the advantage of cooperation, we constructed a control system in which the $I_i$ molecules could act as catalysts, but could not be covalently assembled themselves because their target sequences were not a match for any catalyst in the system (Figure 19). Cooperation would be manifest when enzymes synthesize other enzymes, and there is some benefit to being covalent. Thus we measured the yields of $W\cdot X\cdot Y\cdot Z$ molecules at 8 h in this control system and in our normal system (i.e., Figure 12B). The yields in the control system were consistently worse, and we calculated the ratio $(E_i \text{ catalysis} + I_i \text{ catalysis})$ to $(I_i \text{ catalysis only})$ as the advantage of being covalent in each leg of the cycle. We indicated these ratios above the colored arrows in Figure 12B as 1.73, 1.02, and 1.22 for $i = 1$, 2, and 3, respectively. Assuming these values are multiplicative, that gives a cooperative benefit of about 2.2 for the entire cycle.

The only impediment to truly hyperbolic growth conceivable for such a system (Eigen and Schuster, 1977) is the occasional formation of non-productive complexes such as $W\cdots Y\cdot Z$ through partially complementary base pairing (see Figure 12A). We can detect such complexes (Figure 20) and show that when they are minimized by pre-folding each subsystem separately, the yield after 2 h increases by 25–50% (Figure 21). As shown by heat-cool regimes, reverse reactions that have the net effect of breaking down covalent ribozymes into fragments may also play a small role in preventing hyperbolic growth (Figure 22).
These networks of RNA replicators display some specific characteristics that are relevant to the origins of order in biology. First, the cooperative behavior of these three otherwise sluggish subsystems represents a sharp peak in the fitness landscape. The yield of full-length RNA jumps more than 100-fold when I₁, I₂, and I₃ are mixed. Thus, if under selection, the network of RNAs would quickly out-compete its individual subsystems if the supplies of RNA fragments were replenishing. Second, the system is self-assembling from its component RNA fragments; no full-length ribozyme is needed to seed the reaction. The *Azoarcus* ribozyme can covalently assemble from as many as four oligomer pieces averaging 50 nt in length (Hayden and Lehman, 2006), and thus collectively autocatalytic sets comprised of smaller RNA fragments than described here should be possible. And third, we observed that this network of cooperating RNAs displays an additional dynamic behavior in that the genotype frequencies change in real time. For the cycle shown in Figure 12B, this can be visualized directly on a simplex plot (Figure 23), in which the relative frequencies of the assembled E₁, E₂, and E₃ ribozymes are measured. Viewing genotypic change by tracking relative frequencies of the three sub-systems revealed a migration of the population’s composition towards an equilibrium point that is enriched in E₃ with respect to the other two covalent W•X•Y•Z ribozymes. This is likely a consequence of the superiority of the I₂/E₂ enzymes in synthesizing their product ribozyme E₃. Future studies in which mutational variability in the population exists outside the IGS will be valuable in revealing the full range of evolutionary behavior available to this system.
Cooperation vs. selfishness

Next we tested whether a 3-membered cooperative system has the potential to exhibit a higher fitness than purely autocatalytic systems when placed in direct competition. To construct “selfish” autocatalytic subsystems (S), we reverted the IGS-IGS target pairs within each subsystem so that they would match (Figure 24). To create S1 we used GUGWCAU and h•X•Y•Z, to create S2 we used GAGW•XCUU and h•Y•Z, and to create S3 we used GCAW•X•YCGU and h•Z. Each of these subsystems replicates well in isolation. Upon mixing of RNAs, we tracked selfish and cooperative ribozymes by the composition (matched or mismatched, respectively) of the W-containing fragments because these contain the IGS and hence the most crucial genetic element (Figure 25). When we compared the total yield of S1 + S2 + S3 to that of I1 + I2 + I3, the former out-performed the latter at all time points (i.e., selfishness wins in isolation). One reason for this result is that there would be less time delay in initiating covalent synthesis in the all-selfish system, while there is a delay in establishing cooperative replication because the maturation of the fastest subsystem I2/E2 is dependent on the activity of a slower subsystem I1/E1. However, when we placed all six subsystems (12 RNAs: I1 + I2 + I3 + S1 + S2 + S3) in the same reaction, the relative yields at later times are auspiciously reversed, and the growth of the enzymes resulting from the cooperative network now exceeds those from the selfish subsystems (i.e., cooperation wins in competition). These results are not dependent on the exact RNA fragments we chose, as the same result can be seen in other systems with different IGS and IGS targets (Figure 26). The yield reversal upon mixing is exhibited because the selfish enzymes now participate in – and effectively expand – the
cooperative network (Figure 27). This would be a mechanism for a network connectivity increase when the subsystems involved are competing for at least one shared resource, in this case the catalytic core (Y-Z) because all W-containing fragments can use the same 3’ fragments. And while selfish enzymes can also benefit from the network, the asymmetry in the proficiencies of the various IGS/IGS-target pairings creates potential for an asymmetry in the relative benefits of the various enzymes in the mixed environment (Figure 27). This feature would have been common in primordial genetic systems, allowing us to posit that cooperation could have been predisposed even in homogeneously mixed environments.

**Modeling**

*(The mathematical modeling were well outside the expertise of Dr. Niles Lehman and myself, and were only possible with the help of Dr. Michael L. Manapat of School of Engineering and Applied Sciences and Program for Evolutionary Dynamics of Harvard University in Cambridge MA, and Dr. Irene A. Chen and Dr. Ramon Xulvi-Brunet of FAS Center for Systems Biology, Harvard University in Cambridge MA. The kinetics experiments to obtain the rate constants for modeling were performed by myself.)*

Empirical systems such as the one described above are subject to the particularities of chemical and methodological idiosyncrasies, and thus we sought to generalize these results by constructing mathematical models that show that under a certain set of parameters, the observed laboratory results should indeed be possible. First we constructed an ordinary differential equation (ODE) model for the 3-membered network
shown in Figure 12B. We tracked the yield of each of the three E_r ribozymes separately – using three identical replicates from the same initial reaction mixture – by taking aliquots every 30 min for 16 h (Figure 17). We employed standard optimization techniques to find the rate constants of all the possible reactions in Figure 12B that produced trajectories in the ODE system closest to the observed data. We used these estimated rate constants to construct a second ODE model that would mimic the cooperative growth of the three subsystems. In general the non-covalent versions of the ribozymes form relatively tight complexes, with K_d values in the low nM range. When we built cooperative behavior into the model by relying on differential equations of type dE_j/dt = k_{ij}[I_j][E_i], the experimental data were fit very well in all three subsystems (Figure 28A). When we removed direct catalysis from the model and inserted only autocatalysis instead, the quality of the fit decayed substantially such that the RMS error was 2.4-fold greater (Figure 28B), confirming these results. These data support the contention that replication of the subsystems is indeed cooperative.

Next we constructed a toy model comparing the cooperative and selfish behaviors seen in Figure 25 using the dynamical relationships that can exist among all enzymes (Figure 25 inset). The “selfish” enzymes perform some altruistic catalysis when alternative substrates become available. The empirical data display more striking yield differences than the model perhaps because the time delays in bringing the results of the selfish catalytic events back to the selfish subsystems are exacerbated by physical processes such as diffusion. Again this result is general at least within this network topology and does not depend on the particular IGS/IGS-target pairings that we chose to
construct $S_1$, $S_2$, and $S_3$. In essence, although the selfish replicators can parasitize the cooperators, the cooperative network benefits more by incorporating the selfish RNAs. Interestingly the opposite is generally true in evolutionary dynamics: groups of cooperative individuals grow more quickly than groups of selfish individuals, but a group consisting of both types will eventually be dominated by the selfish (Nowak, 2006). One limitation to the experiment shown in Figure 24 however, is that there is only a single iteration of selection. The RNAs used to seed the experiment limit its evolutionary potential. Experiments in a serial transfer format are needed to show the selection of one strategy over the other (see below). But we can use both the data and modeling shown here to predict that cooperation would have been advantageous in the simpler chemical systems that preceded organismal biology.

**Randomization experiment**

We designed the system described in Figure 12 to explore molecular cooperation, but it is only one of a very large number of possibilities. To test the notion that cooperative networks of RNA could form spontaneously, we randomized the IGS and its target in fragments of the *Azoarcus* group I ribozyme (Figure 29A), mixed the fragments together, and sampled the population over time. Here we randomized the middle nucleotide of both the IGS (M) and its target triplet (N), generating both matched and mismatched partners within the population. Based on these sequences, we created three pools of randomized fragments containing the IGS on the 5’ end of the ribozyme: $G_{MG}W_{CNU}$, $G_{MG}W_{X_{CNU}}$, and
\( \text{GMG}W\cdot X\cdot Y\text{CNU} \), plus three fragments containing the catalytic core and the 3´ end of the ribozyme: \( X\cdot Y\cdot Z \), \( Y\cdot Z \), and \( Z \).

When we mixed these six RNA sets together in equimolar amounts and incubated them at 48°C in 100 mM MgCl\(_2\), all 48 possible full-length \( W\cdot X\cdot Y\cdot Z \) \textit{Azoarcus} ribozymes arose, and we tracked the relative frequencies of these genotypes over an eight-hour time course using a high-throughput nucleotide-sequence analysis. Four-fold variation in M and in N, combined with three-fold variation in the junction \((j)\) where recombination occurs (before \( X \), \( Y \), or \( Z \)) leads to these 48 genotypic possibilities (Figure 29A). These assembled ribozymes can thus be distinguished by three variables: \( i) \) the middle nucleotide of the IGS, \( ii) \) location of the target, which can occur before the \( X \), \( Y \) or \( Z \) fragment, and \( iii) \) the middle nucleotide of the target. We therefore denote each ribozyme with the three-letter code \( MjN \), where \( j = x, y, \) or \( z \). Each of these ribozymes can be covalently assembled by any other ribozyme, itself covalently contiguous or not, provided that \( M \) in the catalyst is complementary to \( N \) in the substrate. For example, \( G\text{AG}W\cdot X\text{CAU} \) and \( Y\cdot Z \) can anneal in solution to form a non-covalent complex with enough catalytic prowess to recombine, say, \( G\text{GG}W\cdot X\cdot Y\text{CUU} \) with \( Z \) to produce the covalent product ribozyme \( G\text{GG}W\cdot X\cdot Y\text{CUU}\cdot Z \), denoted GzU. Note that the 3´ end of the 5´ fragment must form a base-paired stem with the 5´ end of the 3´ fragment for the single-step IGS-dependent catalytic event to take place (Figure 12A) (Draper \textit{et al.}, 2008). This prevents \( W \) from being unproductively recombined with \( Y\cdot Z \) or \( Z \), for example. The growth when the IGS and targets are randomized showed dramatically greater yields (2–12X) than in our engineered 3-membered system, suggesting that in the complex mixture
of the randomization experiment, many more productive interactions among RNA species are occurring (Figure 29B). The relative frequencies of the 48 possible MjN full-length ribozymes that we recovered at each time point are shown in Table 5. In accordance with the data described so far and with published data (Draper et al., 2008), recombination at the Y-Z junction is favored, but no single genotype ever exceeded 13% of the total.

From approximately three million W•X•Y•Z genotypes sampled at each time point, we observed distinct trends that portray indirect evidence of a rapid succession from smaller to increasingly larger networks of cooperators (Figure 30). Genotypes that could easily propagate by selfish autocatalytic replication peak at or before the first time point at 30 min (Figure 30, dotted line with crosses). These are Si genotypes (e.g., those in Figure 24, 25) where M and N are complementary. A prime example is CyG, which could increase in number from the association of GCCW•XC GU and Y•Z molecules, and this genotype rose in frequency from 4.8% to 7.2% between 30 min and 2 h. The rate of ribozyme production could be autocatalytic in this case, because the product can catalyze its own synthesis. Out of the 48 possible product genotypes, twelve (25%) are of this type. After peaking early, the frequencies of genotypes of this class dropped below random expectation and then slowly climbed. Because of extremely large sample sizes, these deviations are highly significant (two-tailed G-tests of independence; \( P << 0.001 \)).

However the later increase in frequency of autocatalysts may not be a consequence of autocatalysis per se, but of the incorporation of these genotypes into higher-ordered networks, akin to the mechanism by which cooperative networks
assimilate selfish replicators (Figure 25, 27). Analyses of the frequencies of the product genotypes cannot reveal the identities of the catalysts that made them, and thus they do not provide direct evidence of replicator cycles with autocatalytic feedback. Nevertheless, we endeavored to explore if networks of two or more distinct members could be increasing over time. Some pairs of genotypes can cooperate with each other to form 2-membered cycles (e.g., AxC + GzU), while others cannot (e.g., AxC + UzG). We noticed that the global joint frequencies of the members comprising all possible 2-membered cycles peaked at 30 min, declined, and recovered, although delayed with respect to the autocatalysts (Figure 30). Support for the notion of succession from autocatalysts to these 2-membered cycles can be found in the frequencies of two possible partners for the autocatalysts GjC, which are CxG and CzG (autocatalysts themselves); the sum of these two frequencies rose monotonically between 2–8 h (3.69% to 6.10%). At roughly the 2 h time point, a succession to 3-membered cycles may have occurred. While there are hundreds of such possible assemblages, the joint frequencies of the members of diverse ones requiring synthesis at all three junctions such as UxG + AyA + CzU studied above, jump at the 2 h mark (Figure 30, solid line). Many others peak then as well; the joint frequency of the AxC + GyA + UyU trio increases nearly 20-fold after the 30 min point. At 4 h and beyond the possibility of succession to even higher-ordered networks that subsume all simpler networks obfuscates individual trends. Visualization of all possible connections among all genotypes underscores the above conclusions (Figure 31). By 8 h it appears that the network is dominated by genotypes that can only be replicated through cooperation (green circles). In fact, the variance in the
genotype frequencies drops monotonically over the course of the entire experiment, suggesting that as time goes on, all genotypes increasingly participate in the large network.

**Serial transfer of the randomized population**

The experiments depicted in Figure 29 portray the dynamical changes that occur on a kinetic time scale as a batch of RNAs approaches equilibrium. In an actual prebiotic scenario however, this effect would be iterated and perhaps magnified over several generations, as opposed to being an asymptotic value that results from mixing several RNAs in a single reaction vessel. To bring a stronger evolutionary flavor, we repeated the randomization experiment but in the serial transfer format. Starting with another aliquot of the exact same set of RNAs (i.e., products from the same *in vitro* transcription), we carried a population through eight serial transfers, taking 10% of the population each hour into a fresh tube of fragments. In this manner the W•X•Y•Z molecules that spontaneously assemble are continually being fed with new fragments, such that selection will favor those molecules and networks that grow faster and persist over iterations. Given that the assembly that occurs each round can be strongly influenced by the actions of naïve RNAs from the 90% fresh material, we opted to assay genotypic change by sampling only the most high-frequency genotypes: those present in an abundance greater than random chance (1/48). By manually sequencing the same number of genotypes (75) from transfers #1 and #8 and enumerating those genotypes present more frequently than random expectation (2/75 > 1/48), we were able to observe the amalgamation of an RNA
network over time (Figure 32). The relative frequencies of the $W \times X \times Y \times Z$ ribozymes formed at 1 hr and 8 hr time points are shown in Table 6. At the 1 hr time point, no closed network was possible and autocatalysts were relatively frequent (33%), but by 8 h a reflexively autocatalytic set was present in which every reaction is catalyzed by at least one molecule involved in any of the reactions of the set (Hordijk and Steel, 2004). This set included nine genotypes and fewer autocatalysts (25%), although the latter drop is not quite statistically significant (one-tailed $G$-test of independence; $P = 0.14$). Such expansion of the network to add additional genotypes is a more general case than the direct competition that we described in Figure 25. As another indicator of the effect of serial transfer, the outcome of this experiment differed markedly from the batch assembly experiment (Figure 31). After 8 h in the batch experiment the genotypes were dominated by pyrimidine-containing IGS’s and targets (RzR; Figure 31). By contrast, the serial transfer experiment, while also reiterating the bias for the Y-Z junction, distinctly favored IGS and target sequences containing purines (YzY, Figure 32).

**Fragmentation into four pieces**

Lastly, we tested whether increased fragmentation of the RNA could provide an additional element of complexity, while at the same time enhancing the pre-biological relevance. We did this by breaking the molecule up into four pieces instead of two, creating 4-piece versions of I$_1$, I$_2$, and I$_3$ analogously to Figure 12A, B. When we mixed the resulting 12 RNAs together, we observed two interesting phenomena (Figure 33). First, the growth curve was distinctly sigmoidal, suggesting that when more fragments
are involved, the cooperativity of the system becomes even more apparent. In the 4-piece fragmentation, $W\cdot X\cdot Y\cdot Z$ ribozymes can be made via many pathways, including those in which more than one enzyme cooperates to construct the product. For example, an $E_1$ ribozyme could recombine the $W\cdot X$ junction, an $E_2$ ribozyme could recombine the $X\cdot Y$ junction, and an $E_3$ ribozyme could recombine the $Y\cdot Z$ junction. Second, analysis of the sequences of the product $W\cdot X\cdot Y\cdot Z$ ribozymes showed that such cooperation was common (Figure 34). The example sequences of various types for 24 h time course are given. They are grouped according to whether $W\cdot X\cdot Y\cdot Z$ the molecule can be derived from only one enzyme source, i.e., the molecule could have been assembled from a single enzyme, distinguished by its IGS, (top box; $N = 19$ out of 124 total, or 15%), or derived from at least two enzyme sources (middle box; $N = 77$ out of 124 total, or 62%), or derived from at least three enzyme sources (lower box; $N = 28$ out of 124 total, or 23%). The latter two types of molecules required highly cooperative actions for their syntheses; they required multiple enzymes to operate sequentially (and/or simultaneously) on their composite RNA fragments. Overall, 85% of all sequences were of this type. In fact more examples were seen that required a minimum of three catalytic sources ($28/124 = 23\%$) than required a minimum of one catalytic source ($19/124 = 15\%$). Also, when the genotype frequencies of molecules assembled over time for the 4-piece reaction were plotted in a simplex plot, the trajectory stayed nearer to the center of the plot, suggesting a greater level of cooperation in this system (Figure 23). In contrast, for the reaction portrayed in Figure 12B, the population’s composition migrated towards an equilibrium.
point that is enriched in E\textsubscript{3} with respect to the other two covalent \( \text{W\cdotX\cdotY\cdotZ} \) ribozymes suggesting the superiority of the I\textsubscript{2}/E\textsubscript{2} enzymes in synthesizing their product ribozyme E\textsubscript{3}.

**DISCUSSION**

The historical scenario implied from our results is one in which simple and small autocatalytic cycles form easily, but are later supplanted by more complex cooperative networks that take advantage of the autocatalysts. Our system describes the short-term kinetic phenomena that provide the foundation for evolutionary behavior (Eigen, 1971) in the presence of sequence variation throughout the ribozymes analogous to that described as “prerlife” (Nowak and Ohtsuki, 2008). Over time, a transition back to purely selfish replicators based on polymerization chemistry could proceed (Levy and Ellington, 2001).

Features of the system described here that would make it a valuable scenario for early evolution are that it is comprised solely of RNA (although other polymers have the potential to display cooperative behavior (Lee \textit{et al.}, 1997a; Lee \textit{et al.}, 1997b)) and that the 3-nt IGS or IGS targets are essentially the tag sequences (Weiner and Maizels, 1987) that have been suggested as a means to form molecular coalitions that can partition genetic information in a homogeneous milieu (Levy and Ellington, 2003). The \( \text{W\cdotX\cdotY\cdotZ} \) RNAs with differing tags resemble a quasi-species containing sets of sequences with similar, but not identical, genotypes. While recombining RNA oligomers must have had their own origins in abiotic polymerization processes, a crude form of sequence similarity was available in that these sequences would have reflected any internal bias in the
chemistry of their condensation such that the sequence space was not evenly populated (Derr et al., 2012; Eigen, 1971). Closure of an autocatalytic set would have been facilitated by the cooperative aggregation of oligomers with related tags, as “the prerequisites for coexistence of precursors can be met generally only by closely related mutants” (Eigen and Schuster, 1978). Subsequent expansion of cooperative networks as shown here is possible by invasion of the network by a new set with a distinct tag sequence, e.g., moving from the 3-membered cycle UxG + [AyA] + CzU to a 4-membered cycle such as UxG + [GyA + AyC] + CzU by inclusion of the new IGS-IGS-target pair GGG/CCU (Figure 31, starred genotypes). There is potential of this network to expand well beyond even four members (Figure 32). Expansion of the network to add additional genotypes is a more general case than the direct competition that we described in Figure 25. Should this process be continued indefinitely, with enhanced selection pressure, we would expect the system to converge on a smaller, but kinetically favored, network. The evolvability of chemical networks in general predicts the emergence of “core” autocatalytic sets (Vasas et al., 2012). Moreover, the environment can influence the short-term fate of the network. Changes in the temperature or salt concentrations would lead to differing strengths of the IGS/IGS-target pairings, thereby altering the relative magnitudes of the rate constants depicted in Figure 17, leading to the establishment of new stable points in the final frequencies of each ribozyme (Figure 23). Longer-term evolutionary optimization would likely have required spatial heterogeneity (Boerlijst and Hogeweg, 1991) and/or compartmentalization (Szathmáry, 2006) to provide lasting immunity against parasitic species or short autocatalytic cycles.
In theory, an open-ended tree of direct catalytic events (i.e., many interdependencies) could be occurring in our randomization experiments, such that a true network is not forming. However, our argument that a network including many positive feedback cycles (“cooperation”) is forming is based on at least two pieces of evidence. The first is the systematic evidence of cooperation in the simple 3-membered cycle studied in Figure 12 and Figure 33. If such a dynamic can exist in isolation, the probability that it can occur in the randomization experiment is high given the enormous number of such cycles that are possible within the randomization experiment (e.g., Figure 31). Second, the growth curve from the randomization experiment shows yields that are much higher than in the simple 3-membered cycle, as shown Figure 29. The existence of positive feedback loops resulting from cycles of various sizes is the best explanation for this increase in yield given the results from study of the simple system in isolation.

Eigen and Schuster proposed highly interdependent networks of genetically related replicators as a means to circumvent the error catastrophe in nascent life (Smith, 1979). In the system we present here, we show how RNA networks have the potential to buffer informational decay. One contributor is the use of recombination for replication. While allowing for some genotypic variability, it does not lead to the accumulation of deleterious mutations in the same fashion as template-directed polymerization (Lynch et al., 1993). Another facet is the amount of information needed to specify the system. Consider an extreme case that exemplifies the network expansion potential mentioned above. Allowing not one but two positions completely free to vary within the three-nucleotide IGS tag (keeping the requisite G-U wobble at the third), a 16-member network
is possible. Specifying all 16 ribozymes independently would require roughly $16 \times 200 \times 2 = 6400$ bits of information (with 200 nucleotides per ribozyme and two bits of information per nucleotide). Yet because the network is quasi-species-like with 16 related enzymes, only slightly more than $200 \times 2 = 400$ bits of information would be needed to specify it, which, being 16-fold lower, would tend to ameliorate an error threshold. Similarly, the 3-membered cycle described here resembles a hypercycle as envisioned by Eigen and Schuster (Eigen and Schuster, 1977, 1978) but without hyperbolic growth. We prefer to focus on the observation that the cycle can be derived from simpler cycles and has the potential to expand to more complex ones as evidence that RNA molecular coalitions can display spontaneous order-producing dynamics, which already has theoretical support (Mossel and Steel, 2005). This type of molecular ecological succession is a plausible model for a bridge between selfish replicators and complex cooperative systems.

**MATERIALS AND METHODS**

*RNA preparation*

Shorter RNAs (<50 nt, such as h•Z) were purchased from TriLink Biotechnologies (San Diego, CA) and were gel purified and desalted prior to use. Longer RNAs were prepared by run-off transcription from double-stranded DNA templates, which in turn were constructed through recursive gene synthesis from single-stranded DNAs purchased from IDT (Coralville, IA), and these RNAs were gel purified and desalted prior to use. RNAs
were prepared with the 5′-IGS (GMG) or with a 5-nucleotide head containing the IGS complement \((h = \text{ggCNU}, \text{where } N \text{ represents a “tag” (target) nucleotide, and where the first two guanosines are in lowercase to denote the fact that they are there simply to improve in vitro transcription yields})\). RNAs were resuspended in 5–20 μM solutions in 0.1 mM EDTA. For subsequent quantification purposes, in most cases, a small amount of the W-containing fragment was 5′-end-labeled with \(\gamma[^{32}\text{P}]\cdot\text{ATP and OptiKinase (USB)}\). In some cases, RNA was prepared using \(\alpha[^{32}\text{P}]\cdot\text{ATP during transcription.}

**RNA covalent self-assembly assays**

Ribozyme assays or covalent self-assembly from oligomers were performed as described previously (Draper et al., 2008; Hayden and Lehman, 2006; Vaidya and Lehman, 2009). Briefly, RNA oligomers were heated to 80°C for 2 min and then cooled to 48°C with concomitant addition of reaction buffer (100 mM MgCl\(_2\) and 30 mM EPPS, pH 7.5). RNAs were then incubated together for 5 min – 16 h at 48°C at a final concentration of 0.01 – 2.0 μM each. Typically RNAs were incubated in equal ratios, except for estimation of kinetic constants and for some reactions shown in Figure 23. Reactions were quenched by the addition of enough EDTA to chelate stoichiometrically all the Mg\(^{2+}\) ions. For the serial transfer reactions (Figure 32), 10% of the reaction was removed every two hours and transferred to a new tube containing fresh RNA fragments in equimolar ratios (0.05 μM). A total of eight transfers were taken. For denaturing gel electrophoresis, an equal volume of 2X gel-loading dye containing 8 M urea, SDS, 200 mM EDTA, and bromophenol blue dye was then added. The RNAs were heat denatured
at 80˚C for 4 min, and then immediately electrophoresed through 8% polyacrylamide / 8 M urea gels for roughly 2000 V-h in 1X TBE. For native gel electrophoresis, one-fifth volume of pre-warmed 6X native-gel loading dye (40% sucrose + 0.25% bromophenol blue) was added to the reacted RNAs and the samples were immediately loaded onto an 8% polyacrylamide gel containing 11 mM MgCl₂ and 1 mM EDTA but no urea. The gels were run at a constant 48˚C temperature in a running buffer of 1X TBE and 10 mM MgCl₂ for approximately 1000 V-h. Visualization and quantification was possible via phosphorimaging on a Typhoon Trio+ variable mode phosphorimager (GE Healthcare) when W-containing fragments were 5´-end-labeled with γ[³²P]ATP prior to use. The computer program ImageQuant (GE Healthcare) was used to quantify bands stoichiometrically from the gel. Raw percent yields were calculated by computing the ratio of RNA in the product band(s) to that in the unreacted band(s). These ratios were then converted to concentration of product (in μM) for graphical representation.

**Genotyping**

For genotyping, W•X•Y•Z RNA (ca. 200 nt) was excised from a gel and subjected to RT-PCR using W (5´–TAATACGACTCACTATAGCAAGGGATGGTG–3´) and Z (5´–CCGGTTTGTGTGACTTTCGCC–3´) specific primers. Genotyping was performed using two techniques: a high throughput sequence analysis on the Illumina platform and standard dye-terminator sequencing. For high throughput sequence analysis, W•X•Y•Z RNA (ca. 200 nt) was excised from a gel by elution into 100 mM NaCl + 10 mM EDTA overnight and the adaptor sequence 5´–CCUACCAGUACCCUACCA–3´ was ligated to
the 5´ end of the RNA to allow for 5´ RACE to capture the sequence variability in the IGS of assembled ribozymes. The adapter-ligated \( W\times X\times Y\times Z \) RNA (ca. 218 nt) was subjected to RT-PCR using adapter-specific \( \text{Adap-prim1} = 5´-CCTACCAGTACCTACCAG-3´ \) and \( Z\)-specific \( \text{ZprimR-shr} = 5´-GGACTATGCTCTTCACCAG-3´ \) primers. The \( Z\)-primR-shr site lies 28 nt upstream from 3´ end of \( Z \) making RT-PCR product 193 nt long so that it can be analysed using 100-bp paired-end sequencing assay. The RT-PCR products were then prepared for paired-end sequencing using the TruSeq DNA Sample Prep kit (Illumina). (This was also done for the manual sequencing). Briefly, each time point was separately bar coded using different adapter indices from Illumina and PCR was performed with Illumina PCR master mix to enrich DNA fragments with Illumina adapters. Each time point samples were diluted to a final concentration of 100 nM and mixed in equal molar ratio to be sequenced. The samples were subjected to paired-end sequencing for 101 bases from each direction. This was done at Harvard University’s FAS Center for Systems Biology Core Facility. We found all sequences that were perfect matches to our expected 48 sequences. In addition, we classified sequences that deviated in the following possible ways (approximately 5% error rate was tolerated). For the plus strand, one error was permitted in the first 19 positions, three errors were permitted between positions 21 and 81, and one error between positions 86 and 101. For the minus strand, one error in the first 26 positions, two errors between positions 33 and 72, and one error between positions 79 and 101 were permitted. Only products that could result from the more common one-step IGS-directed “\( tF2 \)” recombination mechanism (Draper et al., 2008)
were considered, because these are ones in which the sequence in the \( h \) region \( (h = \text{GGCAU}) \) are irrelevant. The rarer, slower two-step “R2F2” recombination mechanism is unlikely to be influential over the short time of the experiment; these sequences were \( \sim 2\% \) of the total.

The same exact randomized input RNA pools \((W, W\cdot X \text{ and } W\cdot X\cdot Y)\) were also subjected to the same type of high-throughput sequence analysis on the same machine used to analyze the \( W\cdot X\cdot Y\cdot Z \) molecules that resulted from the randomization experiment to correct for bias in sequencing reactions. This required poly-A tailing of the 3´ ends of the molecules in order to perform RT-PCR. To do this, the RNA was incubated with 2 U of polyA polymerase (Ambion) for 1 h at 37˚C. The resulting RNA was subjected to reverse transcription using the primer HTSeq_polyT_2 \( (5´-\text{CACAGAATTTAATACGACTCACTATAGGTTTTTTTTTTTTTTT-3´}) \) and then to PCR using primers Adap-prim1 and 3´-RACE-OUTER PRIMER \( (5´-\text{GCGAGCACAGAATTTAATACGACT-3´}) \). The resulting DNA was then subjected to the same Illumina high-throughput sequencing protocol described above. The frequencies of the \( W\cdot X\cdot Y\cdot Z \) genotype from the randomization experiment were then corrected using the \( W, W\cdot X \) and \( W\cdot X\cdot Y \) frequencies resulting here. Using the first “block” of 16 sequences \((M_W N)\) the first block of 16 \( W\cdot X\cdot Y\cdot Z \) sequences \((MxN)\) was corrected. Using the second “block” of 16 sequences \((M_W X_N)\) the second block of 16 \( W\cdot X\cdot Y\cdot Z \) sequences \((MyN)\) was corrected. Using the third “block” of 16 sequences \((M_W X\cdot Y_N)\) the third block of 16 \( W\cdot X\cdot Y\cdot Z \) sequences \((MzN)\) was corrected. Frequency correction was done as follows. For each block of 16 control sequences, the actual reads were
summed up those numbers, and then divided each read by the sum. This gives the initial frequencies of the 16 sequences of each block. To do the correction of the $WXYZ$ genotypes resulting from the randomization experiment, the frequencies of the $WXYZ$ were each divided each by the corresponding initial frequency and by 16.

Autocatalysts were tallied as the sum of all frequencies of molecules in which the IGS matches the IGS tag, i.e., $f(AjU) + f(UjA) + f(GjC) + f(CjG)$, and the result over time appears as the dotted line with crosses in Figure 30. Potential 2-membered cycles were tallied as the sum of the joint frequencies of all pairs of genotypes that could reciprocally assemble each other, without themselves being autocatalysts (here the symbol for a genotype will refer to its frequency):

- $AjA + UjU$ (where $j$ refers to fragment 3′ of the recombination junction; $j = x, y, \text{ or } z$)
- $AjC + GjU$
- $AjG + CjU$
- $CjC + GjG$
- $CjA + UjG$
- $GjA + UjC$

To sum the joint frequencies, the following procedure was followed. For each time point, the frequency of the first member was multiplied by the frequency of the second member for every possible combination of junctions. These products were then summed, and the sum of the six pair-wise sums appears as the dotted line with squares in Figure 30. This is best described using the example of the first pair: $AjA + UjU$. For this pair, the joint frequencies are calculated as follows:
\[(AxA \cdot UxU) + (AxA \cdot UyU) + (AxA \cdot UzU) +
(AyA \cdot UxU) + (AyA \cdot UyU) + (AyA \cdot UzU) +
(AzA \cdot UxU) + (AzA \cdot UyU) + (AzA \cdot UzU)\]
\[= (AxA + AyA + AzA)(UxU + UyU + UzU).\]

For the 30-minute time point this equals \((0.0009190 + 0.0013022 + 0.0095858)(0.0038754 + 0.0060814 + 0.0951251) = 0.0012407\). Summation of the equivalent values for the other five pairs of molecules that can potentially form 2-membered cycles leads to the sum of 0.020632, shown as the first point in Figure 30 after being multiplied by 10 for ease of presentation. The same procedure was used to calculate the joint frequencies of the specific 3-membered cycles \([UxG + AyA + CzU]\) \([UyG + AzA + CxU]\) and \([UzG + AxA + CyU]\), except that here, the values reported in Figure 30 (solid line) are simply the products of the frequencies of these three genotypes, summed and then multiplied by 10,000 for ease of presentation.

For dye-terminator sequence analysis, \(W\cdotX\cdotY\cdotZ\) RNA (ca. 200 nt) was excised from a gel by elution into 100 mM NaCl + 10 mM EDTA overnight and subject to RT-PCR using \(W\)-specific (TAS 2.1a = 5´–CTGCAGAATTCTAATACGACTCACTATAGTGCCTTGCGCCGGG–3´) and \(Z\)-specific (T20a = 5´–CCGGTTTGTGTGACTTTCCG–3´) primers. Amplification of full-length \(Azoarcus\ W\cdotX\cdotY\cdotZ\) RNA with these primers gives a ~230-bp PCR product. (Note, for the serial transfer experiments in Figure 32), the \(W\)-specific primer Adap_prim1 was used instead, to capture the variation in the IGS.) The PCR products were cloned into the vector pJET1.2 using the CloneJET cloning kit (Fermentas) and
transformed into *E. coli*. Individual colonies were picked as templates for colony PCR reactions employing the primers pJET1.2-F and pJET1.2-R (Fermentas). Resulting amplicons were either subjected to nucleotide-sequence analysis or RFLP analysis. Genotyping via RFLP using the *Hpy99I* and *SphI* restriction enzymes was possible because these enzymes can discriminate between E1, E2, and E3 genotypes. The enzyme *Hpy99I* cuts only E1 PCR DNA (at recognition sequence 5´–CGWCG↓–3´) into fragments roughly 99 bp and 131 bp in size, while *SphI* cuts only E2 PCR DNA (at recognition sequence 5´–GCATG↓C–3´) into fragments roughly 135 bp and 95 bp in size, which were resolved on 2% agarose gels containing ethidium bromide. The network diagram in Figure 31 was constructed using Cytoscape v. 2.8.2. The sizes of nodes were scaled to corrected eight hour genotype abundances. Simplex plots (Figure 23) were constructed using SigmaPlot (v. 12) to represent joint frequencies of three constituents whose frequencies sum to 1.
Figure 12. Cooperative covalent assembly of recombinase ribozymes. (A) Design of recombinase ribozymes capable of spontaneous cooperative covalent assembly from RNA fragments. The sequence and secondary structure is of the foundational *Azoarcus* group I intron (Reinhold-Hurek and Shub, 1992). The 198-nt complete ribozyme can be fragmented at three loop regions to obtain four oligomers capable of self-assembling into a full length molecule (Hayden and Lehman, 2006; Hayden et al., 2008). The grey box at the 5’ end of fragment W (magenta) is the internal guide sequence (IGS). The grey boxes at the 3’ ends of fragments of the W, X (lime), and Y (blue) are recombination targets (tags) to be recognized by the IGS. Recombination occurs when an IGS from one ribozyme is complementary to the tag within two base-paired fragments, binds, and catalyzes a covalent closure (•) of the loop, often with the release of a G from the h region (h = GGCAU) (Draper et al., 2008). (B) A cooperative system comprised of three subsystems, each created from partitioning the molecule into two pieces at different junctions: I₁ (W + h•X•Y•Z; IGS = GUG), I₂ (W•X + h•Y•Z; IGS = GAG), and I₃ (W•X•Y + h•Z; IGS = GCG). The two fragments in each Iᵢ subsystem can pair up through secondary and tertiary interactions to form a non-covalent version of a ribozyme. The IGS in I₁ recognizes CAU in I₂ as its recombination tag and assembles the covalent version of I₂, namely E₂ (blue molecule) from I₂. Similarly, the non-covalent E₂ assembles E₃ (magenta molecule) from I₃, and the non-covalent I₃ assembles E₁ (orange molecule) from I₁. These newly assembled covalent versions of E₁, E₂, and E₃ ribozymes can then further enhance the assemblies of E₂, E₃, and E₁ respectively. Numbers over the colored arrows are measurements of the cooperative advantage for each step in the cycle (see text).
Figure 13. Analyses to determine the IGS-IGS target pairs to be used in the cooperative RNA network. (A) Various IGS-IGS target pairs were constructed in the $\text{GMG}_W\text{CNU} + \text{X\textcdotY\textcdotZ}$ two-piece covalent self-assembly system. The relative assay yields after one hour when 2.0 μM $W$ was incubated with 2.0 μM $\text{X\textcdotY\textcdotZ}$ were measured by denaturing polyacrylamide gel electrophoresis and phosphorimaging. The three assays named as “$S_i$” were “matched” assays where the middle nucleotide in the IGS trinucleotide GMG was complementary to the middle nucleotide in its target trinucleotide CNU. The other assays were mismatched. The three sets of pairs used in the network constructed in Figure 12B (shaded black) were chosen because they averaged the poorest yields. Yields were normalized to that of GAG-CGU (yield = 1). (B) Comparison of the effects of the junction placements of the IGS-IGS target pairs chosen in panel (A). Assays performed as in panel (A) except that the ribozyme was broken into various two-fragment arrangements (i.e., $W + \text{X\textcdotY\textcdotZ}$, $W\textcdotX + \text{Y\textcdotZ}$, or $W\textcdotX\textcdotY + Z$). The chosen junction placements (shaded black) for the network in Figure 12B exhibited the poorest yields on average. Yields were normalized to that of GUG-CGU at the $X\text{-}Y$ junction (yield = 1).
Figure 14. Covalent ribozymes are more catalytically proficient than non-covalent versions. In this assay the “cooperative advantage” of being covalently contiguous was measured by following the amount of direct catalysis (Eₐ synthesizing Eₐ₊₁) after 6 h. In each assay, 0.05 μM of each RNA was incubated and the yields were measured by electrophoresis and autoradiography. The white bars represent experiments in which a non-covalent version of the enzyme (Iₐ) was incubated with the two halves of the substrate in the subsequent system (Iₐ₊₁). For example, in the first case, the white bar represents the formation of E₁ (GUG W•X•Y•Z) by I₃ (GCG W•X•Y•Z+CUU + ggcUU Z). The gray bars represent experiments in which the covalent version of the enzyme (Eₐ) was incubated with the two halves of the substrate in the subsequent system (Iₐ₊₁). For example, in the first case, the gray bar represents the formation of E₁ (GUG W•X•Y•Z) by E₃ (GCG W•X•Y•Z). In all cases, the covalent enzymes are more proficient at catalysis, averaging about 2.2-fold better than their non-covalent counterparts. The effects of autocatalysis and cross-catalysis are minimal in all these experiments.
Figure 15. Electrophoretic observation of assemblies of E₂ and E₃. The 5′ fragments of I₂ or I₃ were specifically and independently 5′-radiolabeled with $^{32}$P (i.e., *I₂ or *I₃) to monitor the assemblies. The reactions were performed by incubating 0.5 μM (for autocatalysis) or 0.05 μM (for direct assembly, cross catalysis, and cooperation) of each fragment in 100 mM MgCl₂ and 30 mM EPPS buffer (pH 7.5) at 48°C for 8 h. In the appropriate cases, the arrows identify the subsystems being assembled by the previous subsystems in the network, where the IGS and recombination tags match.
Yields of $E_i$ were measured after 16 h in various combinations of subsystems as exemplified in Figure 15. For the autocatalysis measurements, the yields of $E_i$ were measured when 0.15 μM of each $I_i$ fragment were co-incubated. For the cross-catalysis measurements, the yields of $E_i$ were measured when 0.075 μM of each $I_i$ fragment plus 0.075 μM of each $I_i$ fragment of the preceding subsystem were co-incubated (e.g., $E_1$ was measured when $I_1$ and $I_2$ were co-incubated). For the cooperative measurements, the yields of each $E_i$ were measured separately in a single reaction in which 0.05 μM of $I_1$, $I_2$, and $I_3$ were co-incubated. When comparing the proficiencies of systems with variable numbers of subsystems, the total amount of RNA was normalized in order to ensure a conservative estimate of the yield of any one $E_i$; however, the same general trends are observed if $[I_i]$ is equalized and the total [RNA] in a reaction changes (data not shown).
Figure 17. Yields of individual subsystems $E_1$, $E_2$, and $E_3$ over time in the presence of all three subsystems. The assembly of each subsystem was tracked independently by 5′-radiolabeling $W$-containing fragments of each subsystem with $^{32}$P. The reactions were performed by incubating 0.05 μM of each fragment at 48°C and samples were quenched every 30 min for 16 h.
Figure 18. Effects of doping of covalent catalysts into system. A cooperative network was set up as in Figure 12 using an initial concentration of each RNA fragment of 0.05 μM. Three parallel reactions were initiated with the W-containing fragment of either I₁, I₂, or I₃ was 5´ radiolabeled with γ[^32P]•ATP, such that the yield of E₁, E₂, or E₃ could be monitored, as indicated along the abscissa. However in the experiments indicated by the key, in addition 0.2 μM of full length W•X•Y•Z molecule E₁, E₂, or E₃ was doped in. Reported yields are after 6 h. The results show that only the production of the enzyme that results from direct catalysis consistently increases in yield when compared to the undoped situation, although in a cyclical system doping with products can actually impede synthesis as time goes on. Production of the other two enzymes does not increase, even if production of its precursor in the cycle increases. This is likely a consequence of the fact that the magnitude of the concentration of the doped covalently-contiguous enzyme (4X over the RNA fragments) simultaneously stimulates the reverse reaction (cleavage) of covalent products back into the RNA fragments.
Figure 19. Construction of mutants used to separate covalent from non-covalent catalysis. To assess the “cooperative advantage” of being covalent, it was necessary to create versions of the ribozymes that can assemble other ribozymes via non-covalent catalysis but which could not be covalently assembled themselves. This was done by placing tags on the 3´ ends of the W-containing fragments that could not be matched by any IGS present in the system. The presence of the tags, though inert, created mutant fragments with the same lengths as the normal system (Figure 12B) but that could serve as a control for the benefit of being covalent because they could only participate in non-covalent catalysis. These tags are shown in the red circles in the diagram for each system. When these tags are used, the covalent advantage could be measured as the ratio \( E_i \) catalysis + \( I_i \) catalysis to \( I_i \) catalysis only. An example is seen for the case where \( I_2 \) is mutated on the gel image shown. These assays result in the cooperative advantage values given above the arrows in Figure 12B: 1.73, 1.02, and 1.22 for \( i = 1, 2, \) and 3, respectively.
Figure 20. Characterization of non-covalent and non-productive RNAs. (A) Electrophoretic observation of trapped RNA complexes in native (Mg$^{2+}$-containing and urea-free) polyacrylamide gel. The cooperative assembly reaction is viewed from the perspective of all six fragments, labeled one-at-a-time on their 5´ ends using [$\alpha$-$^{32}$P]•ATP. The curly brackets denote groups of RNAs that were preincubated together prior to mixing with the other groups. Each reaction was loaded in a native gel in both a denaturing condition (as negative control) and a non-denaturing condition. The bands observed in the native condition (not observed in denaturing condition) are trapped RNA complexes forming as a consequence of secondary and tertiary interactions among different fragments. (B) Tentative assignments of trapped RNA complexes. (C) Effect of trapped complexes in the cooperative assembly reaction. The fractional yield of $E_1$, $E_2$ and $E_3$ were observed separately in six different reactions by labeling the 5´ end of the 5´ fragments with $^{32}$P. In each reaction (represented by different colors), different combinations of RNA fragments were folded in assembly buffer for two hours prior to mixing of all six fragments together to start the covalent assembly reaction. When the trapped complexes observed in panel (A) were forced to form prior to the start of reaction (reaction denoted by maroon color), the fractional yields dropped drastically, demonstrating the effect of trapped complexes in the co-operative reaction.
**Figure 21. Effect of pre-folding on RNA assembly yields.** When trapped complexes are minimized by pre-folding, yield goes up by 25-50%, suggesting that more exponential growth would be possible. The data were generated as follows. For the two bars on the left, the original system of six fragments I₁ (W + h•X•Y•Z; IGS = GUG), I₂ (W•X + h•Y•Z; IGS = GAG), and I₃ (W•X•Y + h•Z; IGS = GCG) were either folded together as described in the main text (white bar) or each Iᵢ was folded separately and then mixed together to initiate the reaction. To fold each subunit separately, the two RNAs of the subsystem were heated to 80°C for 2 min, buffer was added, and the RNAs were allowed to fold at 48°C for 10 min prior to mixing with the other two similarly-treated subsystems. For the two bars on the right, the same reactions were carried out utilizing mutant versions of the Iᵢ molecules that had target triplets that matched no IGS present, such that they could not be recombined into covalent Eᵢ ribozymes, but could still provide non-covalent catalysis. In both cases the plot shows formation of E₂; similar trends were observed for E₁ and E₃ (data not shown).
Figure 22. Heat cycling improves sigmoidality of the growth curves. The formation of trapped complexes, which appear to prevent exponential or hyperbolic growth of the cooperative system (Figure 12B), can be ameliorated by a heat-cool treatment that allows melting and reformation of RNA base pairs. In each experiment, 0.05 μM of each of six RNA fragments $I_1 (W + h\bullet X\bullet Y\bullet Z; \text{IGS} = \text{GUG}), I_2 (W\bullet X + h\bullet Y\bullet Z; \text{IGS} = \text{GAG}), \text{and } I_3 (W\bullet X\bullet Y + h\bullet Z; \text{IGS} = \text{GCG})$ were co-incubated except that during incubation, the temperature was cycled between 48˚C and 53˚C every 15 min up to 8 h. Various temperature profiles were attempted, but this one was found to be the most effective. Each RNA was tracked independently by radiolabeling the 5´ ends of the W-containing fragments, followed by 8% denaturing polyacrylamide gel electrophoresis and phosphorimaging. In panel A, the yield of $E_3$ (along with total yield of all $W\bullet X\bullet Y\bullet Z$ RNAs, computed by summing $E_1$, $E_2$, and $E_3$ yields) is shown, while in panel B, the yields of $E_1$ and $E_2$ (on a different scale to highlight sigmoidality) are shown. Compare the yields and sigmoidality of these curves to that without heat cycling (Figure 17).
Figure 23. Simplex plot of genotype frequencies changing over time. The cooperative reaction portrayed in Figure 12B (UxG + AyA + CzU) was studied. First, an incubation of \( I_1, I_2, \) and \( I_3 \) for 16 h was performed taking samples of the population every 1–6 h. Three parallel reactions were performed using the same starting materials except that in each reaction the \( W \)-containing fragment in either \( I_1, I_2, \) or \( I_3 \) was radiolabeled, providing relative yields of the full-length RNAs as a function of time via denaturing gel electrophoresis. Open symbols denote frequencies of RNAs as measured by 5′-[\( ^{32}P \)]-labeled fragments as quantified by gel electrophoresis and phosphorimaging. Squares (□) are for \([I_1] = [I_2] = 0.065 \, \mu M \) and \([I_3] = 0.01 \, \mu M \). Triangles (△) are for \([I_1] = [I_2] = 0.06 \, \mu M \) and \([I_3] = 0.03 \, \mu M \). Open Circles (○) are for \([I_1] = [I_2] = [I_3] = 0.05 \, \mu M \). A second, complementary assay was performed in one such reaction by taking samples of the population every 5 min for 30 min, and then at 6 h. These samples were subjected to RT-PCR followed by manual cloning and genotyping of at least 55 full-length RNAs at each time point. Closed circles (●) are for a reaction initiated with \([I_1] = [I_2] = [I_3] = 0.05 \, \mu M \) in which genotype frequencies were determined by cloning of single molecules followed by either direct nucleotide sequencing or RFLP analysis. Hexagons depict the trajectory for a 4-piece version of the \( I_1, I_2, \) and \( I_3 \) reaction (Figure 33, 34) started with 0.5\( \mu M \) of each of the 12 RNAs, where frequencies were measured by gel electrophoresis and phosphorimaging. The population was never observed leaving the interior of the simplex plot, indicating that the equilibrium point, when reached, is a stable coexistence of the three genotypes. For the 2-piece reactions, roughly the same equilibrium point was approached (\( E_3 \) dominating), arguing against genetic drift. This is likely a consequence of the superiority of the \( I_2/E_2 \) enzymes in synthesizing their product ribozyme \( E_3 \). The trajectory for the 4-piece reaction stays nearer to the center of the plot, suggesting a greater level of cooperation in this system.
Figure 24. Design of selfish subsystems. To construct selfish autocatalytic subsystems, IGS-IGS target pairs within each subsystem of cooperative subsystems were reverted so that they would match. To create $S_1$, we used $GUG_{CAU}$ and $h\cdot X\cdot Y\cdot Z$, to create $S_2$, we used $GAG_{W\cdot X\cdot CUU}$ and $h\cdot Y\cdot Z$, and to create $S_3$ we used $GCG_{W\cdot X\cdot Y\cdot CGU}$ and $h\cdot Z$. 
Figure 25. Cooperative chemistry out-competes selfish chemistry when directly competed. Empirical results using cooperative subsystems $I_1$, $I_2$, and $I_3$ as depicted in the lower-right inset (i.e., Figure 12B). Selfish subsystems $S_1$, $S_2$, and $S_3$ were similar except that the IGS and IGS targets were changed to be matching in each subsystem such that autocatalytic replication is favored. The yields of total $W\cdot X\cdot Y\cdot Z$ RNA were monitored over time either when the cooperative (green) and selfish (red) sets of subsystems were incubated separately (dashed lines) or together in the same reaction mixture (solid lines; upper right inset). Yields were monitored over time by tracking the concentrations of cooperative (mismatched) or selfish (matched) $W$-containing RNAs (0.05 μM initial concentrations). Data points shown are averages of three independent trials. Error bars show the standard error of the mean (s.e.m.), and the yields of the cooperative trials in the mixed experiment are significantly greater than those of the selfish trials at the 10- and 16-hour time points ($P < 0.05$ by t-tests using Sidák’s correction for multiple $a$ posteriori comparisons). Simulation of growth dynamics using a toy model of the network of cooperation and selfish interaction in the upper-right inset. Cooperative enzymes ($E_1$, $E_2$, and $E_3$) fare better in competition than do selfish enzymes ($S_1$, $S_2$, and $S_3$), as demonstrated empirically.
**Figure 26. A second example of a selfish vs. cooperative competition experiment.** The experimental set-up shown for Figure 25 was repeated with a different set of competing RNA fragments. Here, $I_1 = GUGW_{CGU} + ggCGUX•Y•Z$, $I_4 = GGGW•X_{CAU} + ggCAUY•Z$, $I_5 = GCGW•X•Y_{CCU} + ggCCUZ$, $S_1 = GUGW_{CAU} + ggCAUX•Y•Z$, $S_3 = GGGW•X_{CCU} + ggCCUY•Z$, and $S_4 = GCGW•X•Y_{CGU} + ggCGUZ$. The yields of total $W•X•Y•Z$ RNA were monitored over time either when the cooperative (green) and selfish (red) sets of subsystems were incubated separately (dashed lines) or together in the same reaction mixture (solid lines). Yields were monitored over time by tracking the concentrations of cooperative (mismatched) or selfish (matched) $W$-containing RNAs (0.05 μM initial concentrations). The same general result was observed as seen with the systems described for Figure 25: cooperation does not out-compete selfishness when compared in isolation but does when placed in direct competition in a mixed system.
Figure 27. Asymmetry in subsystem assembly rates grants a competitive advantage to cooperative ribozymes in a mixed population. “Selfish” and “cooperative” ribozymes are denoted by the composition (matched or mismatched IGS/IGS-targets, respectively) of the W-containing fragments. When all 12 RNA fragments (I₁ + I₂ + I₃ + S₁ + S₂ + S₃) are mixed together, the network shown here becomes established. This diagram is analogous to Figure 12B, with the addition of the dotted line delineating the cooperative system (Figure 12B per se, rotated 120°), outside of which the selfish subsystems lie. Black arrows refer to catalytic events that benefit “like” systems (selfish catalysts catalyzing the production of selfish catalysts, or cooperative catalysts catalyzing the production of cooperative catalysts). Gray arrows refer to catalytic events that benefit “unlike” systems (selfish catalysts catalyzing the production of cooperative catalysts and vice versa). The thick arrows denote the strongest catalysts, those that utilize the GAG/CUU pair between the IGS and the IGS target. These are the S₂ and E₂ catalysts. This asymmetry results in the experimental result shown in Figure 25, in that the yield of the cooperative enzymes eventually surpasses that of the selfish enzymes when all are in the same reaction. This is because even though the cross-system benefits (gray arrows) are potentially the same for both selfish and cooperative enzymes, there is a greater time lag for the cross-system benefits of the selfish enzymes to return to the selfish system than for the parallel return for the cooperative enzymes. One way to appreciate this is to compare the route by which products made by E₂ travel. The cross-system catalysis by E₂ first makes S₂, which can quickly be routed back into the production of E₃. On the other hand, the cross-system production by S₂ takes longer to get back into selfish enzymes because the route must involve one or more of the slower reactions. In short, the catalytic activity of the selfish enzymes gets caught in the cooperative network because of unequal rate constants of the three IGS/IGS-target pairings (one is greater than the other two).
Figure 28. Mathematical modeling of molecular cooperation. (A) Comparison of empirical data from system shown in Figure 17 (solid lines) and modeling behavior of the system assuming the cooperative interactions shown in the inset (dotted lines). For the former, yields of each covalent ribozyme $E_1$, $E_2$, or $E_3$ were measured every 30 min for 16 h. Model description can be found in the Mathematical Modeling section above in the text. (B) Numerical modeling of the dynamic behavior of the RNA network without direct catalysis (dotted lines). This plot was constructed analogously to that of Figure 17 (solid lines), except that instead of assuming autocatalysis to be negligible, here direct catalysis was assumed to be negligible while autocatalysis was incorporated (i.e., pure "selfishness"). The results show that the model fits the experimental data poorly (RMS error = 0.0152 μM, compared to 0.00642 μM for panel A).
Figure 29. The randomization experiment. (A) Design of the randomization experiment. The middle nucleotide of the IGS or the IGS tag was randomized to create diverse pools of RNA fragments. A reaction of 300 pmol each (0.5 μM) of $\text{GMG}W_{\text{CNU}}$, $\text{GMG}W\cdot X_{\text{CNU}}$, $\text{GMG}W\cdot X\cdot Y_{\text{CNU}}$, $X\cdot Y\cdot Z$, $Y\cdot Z$, and $Z$ was sampled at 0.5, 2, 4, and 8 h, and approximately 3 million recombined full-length $W\cdot X\cdot Y\cdot Z$ ribozymes were genotyped by high throughput nucleotide sequence analysis (Table 5). (B) Comparison of growth curves from fixed and randomized RNA sequences. The yield as a function of time was compared for the simple 3-membered cycle (UxG + AyA + CzU) as shown in Figure 12B to that in the randomized format (panel A) when both were performed at the same RNA input concentrations (0.05 μM of each RNA pool: $W$, $W\cdot X$, $W\cdot X\cdot Y$, $h\cdot X\cdot Y\cdot Z$, $h\cdot Y\cdot Z$, and $h\cdot Z$). The curve with triangles (▲) is the sum of the three curves shown in Figure 17, and depicts the growth of the 3-membered cycle. The curve with circles (●) was obtained by tracking the progress of mixing of the randomized RNA pools.
Figure 30. Proposed succession from simple to complex networks using genotype frequency data. Simple autocatalytic cycles where M and N are complementary were directly tracked by the sum of such $W \cdot X \cdot Y \cdot Z$ molecules (dashed line with crosses; e.g., AzU). Reciprocal 2-membered cycles were tracked by the sum (x10, for ease of presentation) of the joint frequencies of all genotypes that can potentially participate in such cycles (dashed line with squares; e.g., AxA + UxU). The rise of 3-membered cycles can be seen from the sum (x10,000 for ease of presentation) of joint frequencies of three sets of genotypes: that shown in Figure 12b, plus its two permutations by junction (solid line; UxG + AyA + CzU; UyG + AzA + CxU; UzG + AxA + CyU). See materials and methods section for details on how the joint frequencies were calculated.
Figure 3. The potential network of RNA genotypes in the randomization experiment. Each node is one of the 48 possible MjN genotypes, and its size represents its relative frequency in the 8 h pool. The node colors indicate autocatalysts (red) or those that must replicate cooperatively (green). Red arrows denote autocatalysis, while grey arrows show all possible direct catalytic events. Orange arrows are the reciprocal 2-membered cycles in which the frequencies of both members at least double between 30 min and 2 h. Green arrows indicate some key 3-membered networks; thick green is the system studied in depth (Figure 12B); thin green are permutations of Figure 12B by junction; dotted green is AxC + GyA + UyU. Stars denote genotypes that can participate in a 4-membered network.
The RNA used to seed the randomization experiment (Figure 29) was subjected to a serial transfer protocol. For the first iteration, 50 pmol of \( \text{GMG}_W \text{CNU}, \text{GMG}_W \text{X}_C \text{CNU}, \text{GMG}_W \text{X}_Y \text{CNU}, \text{X}_Y \text{Z}, \text{Y}_Z, \) and \( Z \) were each incubated in a 100 \( \mu \)L volume. These RNAs were additional aliquots of RNA from the same source used in Figure 30 and Figure 31, allowing a direct comparison between the two experiments. After each 1-hour time point, 10% of the reaction mixture was transferred to a new tube containing 90% fresh RNA with a total volume of 100 \( \mu \)L. The population was sampled after the 1 h and 8 h time points via 5’ RACE and RT-PCR to capture variation in all positions of any \( W \text{X}_Y \text{Z} \) molecules present in the population. These two populations were cloned and genotype frequencies were obtained by manual sequence analysis of 75 clones each (Table 6). Any genotype present twice or more was included on this diagram (see text) where the size of the circles is scaled to relative frequencies within their respective populations. All possible catalytic interactions are shown with arrows among the non-autocatalytic genotypes (green), with autocatalytic genotypes (red) not participating in the network. Grey genotype in 1st iteration disappears. Genotypes with asterisks appear by the 8th iteration.

**Figure 32. The serial transfer experiment.** The RNA used to seed the randomization experiment (Figure 29) was subjected to a serial transfer protocol. For the first iteration, 50 pmol of \( \text{GMG}_W \text{CNU}, \text{GMG}_W \text{X}_C \text{CNU}, \text{GMG}_W \text{X}_Y \text{CNU}, \text{X}_Y \text{Z}, \text{Y}_Z, \) and \( Z \) were each incubated in a 100 \( \mu \)L volume. These RNAs were additional aliquots of RNA from the same source used in Figure 30 and Figure 31, allowing a direct comparison between the two experiments. After each 1-hour time point, 10% of the reaction mixture was transferred to a new tube containing 90% fresh RNA with a total volume of 100 \( \mu \)L. The population was sampled after the 1 h and 8 h time points via 5’ RACE and RT-PCR to capture variation in all positions of any \( W \text{X}_Y \text{Z} \) molecules present in the population. These two populations were cloned and genotype frequencies were obtained by manual sequence analysis of 75 clones each (Table 6). Any genotype present twice or more was included on this diagram (see text) where the size of the circles is scaled to relative frequencies within their respective populations. All possible catalytic interactions are shown with arrows among the non-autocatalytic genotypes (green), with autocatalytic genotypes (red) not participating in the network. Grey genotype in 1st iteration disappears. Genotypes with asterisks appear by the 8th iteration.
Figure 33. Growth curve of a 4-piece of system. A more highly fragmented system based on that shown in Figure 12B was created. Here, the *Azoarcus* ribozyme was broken into four fragments for each I₁, I₂, and I₃ subsystem. The resulting 12 RNAs were co-incubated at 0.5 μM each, and samples were removed over time for both yield analysis (plot) and nucleotide sequence analysis (frequencies). The $W\cdot X\cdot Y\cdot Z$ RNAs can be assembled from a minimum of one, two, or three IGS-bearing enzymes (examples shown with arrow diagrams), and the high frequencies of the latter two classes demonstrates the general cooperativity of the system.
Figure 34. Example sequences resulting from the 4-piece assembly experiment. Here, the ribozyme was broken into four pieces (at the W-X, X-Y, and Y-Z junctions) and allowed to reassemble (Figure 33), as opposed to all other experiments in which the ribozyme could be assembled from only two fragments. For this experiment, twelve RNAs were co-incubated in equimolar ratios at a final concentration of 0.5 μM each. These 12 RNAs correspond to the 4-piece versions of subsystems I₁ (GUGWCGU, ggCGUXCGU, ggCGUYCGU, and ggCGUZ), I₂ (GAGWCAU, ggCAUYCAU, ggCAUYCAU, and ggCAUZ), and I₃ (GCGWCUU, ggCUUXUU, ggCUUYUU, and ggCUUZ). The yield of the reaction was followed over a 24 h time course, and samples were removed periodically and subjected to RT-PCR and manual cloning and nucleotide sequence analysis. A total of 124 W•X•Y•Z molecules were analyzed for sequence variation at the IGS and junction regions to determine from which individual fragments they were derived. The IGS and target sequences are shown in the same colors as in Figure 12B easy reference. All subsystems had examples of W•X•Y•Z products from all three types, with the exception of the first example (in parentheses); the E₁ molecule synthesized entirely from I₂/E₃ sources was not detected in the sample of 124 sequences.
Table 4. Demonstration of cooperation through two-step (relayed) catalysis. Yields of E₁ in μM and percent yield from direct-catalysis (I₁ + I₃), cooperation (I₁ + I₂ + I₃), autocatalysis (I₁), and indirect-catalysis (I₁ and I₂) with 0.05 μM of each RNA fragments incubated for 2 h, 4 h and 6 h. The yield increase in the formation of E₁ when I₂ was added to I₁ and I₃ was greater than when I₂ was added to I₁ alone for all time points. Similar trends were observed when E₂ and E₃ were the tracked products (data not shown). The (*) denotes that 5′ fragment of subsystem is 5′-radiolabeled with ³²P.
<table>
<thead>
<tr>
<th>genotype</th>
<th>30 minutes</th>
<th>2 hours</th>
<th>4 hours</th>
<th>8 hours</th>
<th>(sequences)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GxG</td>
<td>0.0138900</td>
<td>0.0115186</td>
<td>0.0078272</td>
<td>0.0177096</td>
<td></td>
</tr>
<tr>
<td>GxA</td>
<td>0.0101575</td>
<td>0.0076315</td>
<td>0.0048819</td>
<td>0.0146131</td>
<td></td>
</tr>
<tr>
<td>GxC</td>
<td>0.0018055</td>
<td>0.0044971</td>
<td>0.0039851</td>
<td>0.0048323</td>
<td></td>
</tr>
<tr>
<td>GxU</td>
<td>0.0074359</td>
<td>0.0099914</td>
<td>0.0067704</td>
<td>0.0124137</td>
<td></td>
</tr>
<tr>
<td>AxG</td>
<td>0.0018726</td>
<td>0.0019855</td>
<td>0.0016019</td>
<td>0.0021291</td>
<td></td>
</tr>
<tr>
<td>AxA</td>
<td>0.0009190</td>
<td>0.0011231</td>
<td>0.0009419</td>
<td>0.0013418</td>
<td></td>
</tr>
<tr>
<td>AxC</td>
<td>0.0007226</td>
<td>0.0019278</td>
<td>0.0021050</td>
<td>0.0014582</td>
<td></td>
</tr>
<tr>
<td>AxU</td>
<td>0.0009572</td>
<td>0.0016564</td>
<td>0.0013353</td>
<td>0.0014153</td>
<td></td>
</tr>
<tr>
<td>CxG</td>
<td>0.0112859</td>
<td>0.0089159</td>
<td>0.0098774</td>
<td>0.0165233</td>
<td></td>
</tr>
<tr>
<td>CxA</td>
<td>0.0134748</td>
<td>0.0120296</td>
<td>0.0135450</td>
<td>0.0245961</td>
<td></td>
</tr>
<tr>
<td>CxC</td>
<td>0.0093911</td>
<td>0.0167631</td>
<td>0.0234030</td>
<td>0.0205188</td>
<td></td>
</tr>
<tr>
<td>CxU</td>
<td>0.0136461</td>
<td>0.0167948</td>
<td>0.0180510</td>
<td>0.0241148</td>
<td></td>
</tr>
<tr>
<td>UxG*</td>
<td>0.0095316</td>
<td>0.0083482</td>
<td>0.0086070</td>
<td>0.0119408</td>
<td></td>
</tr>
<tr>
<td>UxA</td>
<td>0.0053999</td>
<td>0.0051072</td>
<td>0.0054891</td>
<td>0.0081919</td>
<td></td>
</tr>
<tr>
<td>UxG</td>
<td>0.0027929</td>
<td>0.0060830</td>
<td>0.0084900</td>
<td>0.0064110</td>
<td></td>
</tr>
<tr>
<td>UxU</td>
<td>0.0038754</td>
<td>0.0052987</td>
<td>0.0053226</td>
<td>0.0060686</td>
<td></td>
</tr>
<tr>
<td>GyG</td>
<td>0.0031393</td>
<td>0.0057446</td>
<td>0.0050121</td>
<td>0.0049804</td>
<td></td>
</tr>
<tr>
<td>GyA</td>
<td>0.0012644</td>
<td>0.0035926</td>
<td>0.0032194</td>
<td>0.0030151</td>
<td></td>
</tr>
<tr>
<td>GyC</td>
<td>0.0017179</td>
<td>0.0053810</td>
<td>0.0057573</td>
<td>0.0032392</td>
<td></td>
</tr>
<tr>
<td>GyU</td>
<td>0.0025232</td>
<td>0.0064365</td>
<td>0.0049206</td>
<td>0.0047429</td>
<td></td>
</tr>
<tr>
<td>AyG</td>
<td>0.0043019</td>
<td>0.0094661</td>
<td>0.0078168</td>
<td>0.0054624</td>
<td></td>
</tr>
<tr>
<td>AyA</td>
<td>0.0013022</td>
<td>0.0044066</td>
<td>0.0038467</td>
<td>0.0025224</td>
<td></td>
</tr>
<tr>
<td>AyC</td>
<td>0.0016410</td>
<td>0.0060608</td>
<td>0.0063365</td>
<td>0.0024235</td>
<td></td>
</tr>
<tr>
<td>AyU</td>
<td>0.0035016</td>
<td>0.0107314</td>
<td>0.0078080</td>
<td>0.0052438</td>
<td></td>
</tr>
<tr>
<td>CyG</td>
<td>0.0484205</td>
<td>0.0718199</td>
<td>0.0867414</td>
<td>0.0764932</td>
<td></td>
</tr>
<tr>
<td>CyA</td>
<td>0.0148131</td>
<td>0.0362261</td>
<td>0.0432784</td>
<td>0.037712</td>
<td></td>
</tr>
<tr>
<td>CyC</td>
<td>0.0070897</td>
<td>0.0185869</td>
<td>0.0269449</td>
<td>0.0131831</td>
<td></td>
</tr>
<tr>
<td>CyU</td>
<td>0.0231867</td>
<td>0.0503346</td>
<td>0.0499184</td>
<td>0.0438212</td>
<td></td>
</tr>
<tr>
<td>UyG</td>
<td>0.0070274</td>
<td>0.0124152</td>
<td>0.0139490</td>
<td>0.0102433</td>
<td></td>
</tr>
<tr>
<td>UyA</td>
<td>0.0016891</td>
<td>0.0046854</td>
<td>0.0053926</td>
<td>0.0038416</td>
<td></td>
</tr>
<tr>
<td>UyC</td>
<td>0.0025078</td>
<td>0.0074463</td>
<td>0.0102462</td>
<td>0.0042336</td>
<td></td>
</tr>
<tr>
<td>UyU</td>
<td>0.0060814</td>
<td>0.0152735</td>
<td>0.0145159</td>
<td>0.0107040</td>
<td></td>
</tr>
<tr>
<td>GzG</td>
<td>0.0300383</td>
<td>0.0131832</td>
<td>0.0101212</td>
<td>0.0163321</td>
<td></td>
</tr>
<tr>
<td>GzA</td>
<td>0.0172364</td>
<td>0.0119157</td>
<td>0.0099692</td>
<td>0.0155366</td>
<td></td>
</tr>
<tr>
<td>GzC</td>
<td>0.0188210</td>
<td>0.0204125</td>
<td>0.0150289</td>
<td>0.0174034</td>
<td></td>
</tr>
<tr>
<td>GzU</td>
<td>0.0255912</td>
<td>0.0315221</td>
<td>0.0246947</td>
<td>0.0242498</td>
<td></td>
</tr>
<tr>
<td>AzG</td>
<td>0.0186831</td>
<td>0.0098387</td>
<td>0.0072592</td>
<td>0.0083752</td>
<td></td>
</tr>
<tr>
<td>AzA</td>
<td>0.0095858</td>
<td>0.0082246</td>
<td>0.0064685</td>
<td>0.0072105</td>
<td></td>
</tr>
<tr>
<td>AzC</td>
<td>0.0171920</td>
<td>0.0227244</td>
<td>0.0161284</td>
<td>0.0132234</td>
<td></td>
</tr>
<tr>
<td>AzU</td>
<td>0.0179189</td>
<td>0.0300119</td>
<td>0.0242710</td>
<td>0.0147024</td>
<td></td>
</tr>
<tr>
<td>CzG</td>
<td>0.0792077</td>
<td>0.0279951</td>
<td>0.0284082</td>
<td>0.0445223</td>
<td></td>
</tr>
<tr>
<td>CzA</td>
<td>0.0326014</td>
<td>0.0197686</td>
<td>0.0216797</td>
<td>0.0308802</td>
<td></td>
</tr>
<tr>
<td>CzC</td>
<td>0.1190263</td>
<td>0.1085128</td>
<td>0.1085050</td>
<td>0.1142560</td>
<td></td>
</tr>
<tr>
<td>CzU*</td>
<td>0.0586150</td>
<td>0.0602692</td>
<td>0.0674329</td>
<td>0.0580469</td>
<td></td>
</tr>
<tr>
<td>UzG</td>
<td>0.1276108</td>
<td>0.0550582</td>
<td>0.0522429</td>
<td>0.0650858</td>
<td></td>
</tr>
<tr>
<td>UzA</td>
<td>0.0600067</td>
<td>0.0403763</td>
<td>0.0420634</td>
<td>0.0501074</td>
<td></td>
</tr>
<tr>
<td>UzC</td>
<td>0.0354173</td>
<td>0.0369272</td>
<td>0.0335875</td>
<td>0.0300265</td>
<td></td>
</tr>
<tr>
<td>UzU</td>
<td>0.0951251</td>
<td>0.1149803</td>
<td>0.1142015</td>
<td>0.0839011</td>
<td></td>
</tr>
</tbody>
</table>

sum 1.0 1.0 1.0 1.0

Table 5. Genotype frequencies from high throughput sequence analysis of the W•X•Y•Z ribozymes formed in the randomization experiment.

89
<table>
<thead>
<tr>
<th>genotype</th>
<th>1st gen (75 sequences)</th>
<th>8th gen (75 sequences)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GxG</td>
<td>0.0133</td>
<td>0</td>
</tr>
<tr>
<td>GxA</td>
<td>0.0267</td>
<td>0.0267</td>
</tr>
<tr>
<td>GxC</td>
<td>0.0133</td>
<td>0.0267</td>
</tr>
<tr>
<td>GxU</td>
<td>0.0533</td>
<td>0.0400</td>
</tr>
<tr>
<td>AxA</td>
<td>0</td>
<td>0.0133</td>
</tr>
<tr>
<td>AxC</td>
<td>0.0133</td>
<td>0</td>
</tr>
<tr>
<td>AxU</td>
<td>0</td>
<td>0.0267</td>
</tr>
<tr>
<td>UxC</td>
<td>0</td>
<td>0.0267</td>
</tr>
<tr>
<td>GyC</td>
<td>0.0267</td>
<td>0</td>
</tr>
<tr>
<td>AyC</td>
<td>0.0267</td>
<td>0</td>
</tr>
<tr>
<td>GzG</td>
<td>0.0933</td>
<td>0.1200</td>
</tr>
<tr>
<td>GzA</td>
<td>0.0800</td>
<td>0.1333</td>
</tr>
<tr>
<td>GzC</td>
<td>0.1467</td>
<td>0.0800</td>
</tr>
<tr>
<td>GzU</td>
<td>0.1600</td>
<td>0.0667</td>
</tr>
<tr>
<td>AzG</td>
<td>0.0533</td>
<td>0.0667</td>
</tr>
<tr>
<td>AzA</td>
<td>0.0400</td>
<td>0.1333</td>
</tr>
<tr>
<td>AzC</td>
<td>0.0400</td>
<td>0.0533</td>
</tr>
<tr>
<td>AzU</td>
<td>0.1200</td>
<td>0.1067</td>
</tr>
<tr>
<td>CzA</td>
<td>0.0133</td>
<td>0</td>
</tr>
<tr>
<td>CzU</td>
<td>0.0133</td>
<td>0.0267</td>
</tr>
<tr>
<td>UzG</td>
<td>0.0267</td>
<td>0.0133</td>
</tr>
<tr>
<td>UzA</td>
<td>0.0267</td>
<td>0.0133</td>
</tr>
<tr>
<td>UzC</td>
<td>0.0133</td>
<td>0</td>
</tr>
<tr>
<td>UzU</td>
<td>0</td>
<td>0.0267</td>
</tr>
</tbody>
</table>

| sum      | 1                      | 1                      |

Table 6. Genotype frequencies from manual sequence analysis of the W•X•Y•Z ribozymes formed in the randomized serial transfer experiment (Figure 32). Grey values represent genotypes formed only once, which is lower than random expectation (1/75<1/48) and not included in Figure 32.
CHAPTER 4: RECYCLING OF INFORMATIONAL UNITS LEADS TO SELECTION OF REPLICATORS IN A PREBIOTIC SOUP

Note: This chapter is adapted from a manuscript under consideration for publication in Chemistry & Biology.

ABSTRACT

The prebiotic chemical reactions needed to foster the origins of life on the Earth would have been greatly aided by a process whereby living materials could have been recycled under conditions of limiting resources. Recombination of RNA fragments between two or more RNA species is a viable means of recycling, but the feasibility of this has not been demonstrated. An empirical laboratory investigation was used here to show the potential of recycling to select high fitness sequences, both globally and locally, from a pool of variants in which RNA fragments were scrambled. Using systems based on the Azoarcus group I intron ribozyme, the laboratory studies indicate that mixtures of scrambled and/or deleteriously mutated molecules can recycle their joint component fragments to generate fully functional recombinase ribozymes. These studies highlight the importance of recombination and recycling jointly to aid the advent of living systems and to lead to the most robust selection.
INTRODUCTION

The availability of prebiotic materials would have been an important limiting factor during the origins of life on the Earth. Attempts to find feasible prebiotic pathways to synthesize biologically relevant precursor molecules are hampered by both mixtures of products formed and product yields (Orgel, 2004; Zubay and Mui, 2001). A large variety of molecules and inadequate amounts of the right products would make it improbable for any reproducing system to arise spontaneously in a prebiotic soup (Wächtershäuser, 1988). More favorable conditions would have been possible in a recycling chemical system capable of a recurrent use of organic compounds, fed with an external energy supply (King, 1986). Because the exponential growth of the system necessitates the presence of all the reagents in adequate concentrations (King, 1982), a completely recycling system would be able to evolve by reusing improperly assembled materials that would otherwise disrupt its reproduction (Figure 35). Without recycling that could dissociate polymers, monomers that are incorporated into non-functional polymers would be wasted, and consequently there may not have been adequate material to get life started. For example, to make 1 picomole of all possible sequences of 50 nucleotide (nt) long RNA molecules (e.g., hammerhead ribozyme) by an irreversible process, it would require 1019 kg of carbon, roughly the total weight of carbon in the Earth’s crust (Engelhart and Hud, 2010).

The importance of recycling in prebiotic chemistry has been addressed in various theoretical and experimental studies. A system based on reversible reactions has been suggested to provide a driving force for the spontaneous emergence of homochirality in
closed mass reacting systems, allowing a means of correcting “mistakes” that produce the wrong enantiomer (Blackmond and Matar, 2008; Plasson et al., 2004). Environmentally driven recycling has been proposed in a theoretical study as a robust mechanism for discovery of novel functionality from a finite pool of random sequence informational polymers (Walker et al., 2012). The significance of recycling has also been established in the assembly of complex, biologically pertinent, supramolecular self-assemblies and in polymerization processes. The dynamic self-assembly of amphiphilic molecules suspended in aqueous solutions of various amphiphile concentration, pH values, and ionic contents demonstrates reuse of monomers to form different structures (Monnard and Deamer, 2002). However, the assembly of supramolecular structures does not include covalent bonds and carries information limited to reactivity and physiochemical properties (Graham et al., 2004). As such, it is not sufficient to evolve biological systems.

A fundamentally different assembly mechanism could exploit dynamic covalent chemistry (DCC) where non-covalent interactions are used to make conditionally reversible covalent linkages such that the biopolymers could be formed, broken, and re-formed without the need for continual supply of reagents (Fulton, 2008; Hickman et al., 2008; Ladame, 2008; Oh et al., 2001; Sreenivasachary et al., 2006; Whitney et al., 2004). The reversible imine condensation reaction in template-directed ligation of DNA has been shown to be significantly selective and sensitive to thermodynamics of the substrate-template association (Goodwin and Lynn, 1992; Zhan and Lynn, 1997). Similarly in peptide-based systems, template-directed reversible covalent chemistry has demonstrated the potential of recycling in sequence-selective reproduction and dynamic
error correction (Severin et al., 1998; Ura et al., 2009). Yet intermolecular template-directed recycling chemistry also carries some important limitations, such as the production of only those sequences that are complementary to the templates, the reliance on the use of high-energy leaving groups, and the involvement of informationally depauperate biopolymers.

In practice, several mechanisms of recycling could have been operational in an RNA World. The simplest possibility is the hydrolysis and subsequent re-ligation of RNA fragments, processes that could either be spontaneous or catalyzed in some fashion. Alternatively, or in addition, formal recombination of fragments could have occurred through the simultaneous swapping of blocks of genetic information. Recombination is a reversible process and thus has the potential to enable reutilization of larger biopolymer fragments rather than being limited to simpler small fragments, which is characteristic of the template-directed mechanism discussed above. Moreover, it does not rely on high-energy intermediates (Lehman, 2003). Previously, a group I intron of Azoarcus sp. BH72, which at ca. 200 nt is one of the smallest known self-splicing introns, has been designed to function as a recombinase ribozyme. When broken into four fragments (W, X, Y, and Z), the ribozyme can self-assemble into a trans complex (using secondary and tertiary interactions) that is similar in structure to a covalently linked contiguous ribozyme (W•X•Y•Z). If the fragments have 3-nt recombination tags at appropriate places in the molecules, then they can be recombined together via three sequential transesterification reactions leading to the autocatalytic growth of covalently linked full-length molecule (Draper et al., 2008; Hayden and Lehman, 2006; Hayden et al., 2008). Here, we provide
empirical evidence of the reversible nature of transesterification reaction in the four-piece self-assembling Azoarcus system that can lead to its selection (Figure 36), highlighting the importance of recycling of precursor molecules in the early stages of life on the Earth.

RESULTS

A global search for functional molecules

In the $\text{W\text{-}X\text{-}Y\text{-}Z}$ system, recyclease activity can manifest as recombinase activity. The transesterification reaction that directs the assembly of full length recombinase ribozymes can also occur in a reverse direction leading to the breakage of the $\text{W\text{-}X\text{-}Y\text{-}Z}$ molecules back to $\text{W}$, $\text{X}$, $\text{Y}$, and $\text{Z}$ fragments. It may be possible for this recombination system to reuse these fragments and execute a global search for functional molecules. To test the ability of the $\text{W\text{-}X\text{-}Y\text{-}Z}$ system to recycle fragments and to explore sequence space for functions and stabilities using a DCC, three different tetramer constructs ($\text{W\text{-}X\text{-}X\text{-}Z}$, $\text{W\text{-}Y\text{-}Y\text{-}Z}$, and $\text{W\text{-}Y\text{-}X\text{-}Z}$) were designed in the laboratory out of all possible arrangements of $\text{W}$, $\text{X}$, $\text{Y}$, and $\text{Z}$. Trans-splicing catalytic activity is not possible for $\text{W\text{-}X\text{-}X\text{-}Z}$ and $\text{W\text{-}Y\text{-}Y\text{-}Z}$ because one of the fragments (Y and X respectively) required for folding into a fully functional ribozyme is missing (Adams et al., 2004). However, the transposant $\text{W\text{-}Y\text{-}X\text{-}Z}$ retained minimal activity as a consequence of the ability of RNA fragments to function in trans and provide multiple functions (Beaudry and Joyce, 1990; Vaidya and Lehman, 2009; van der Horst et al., 1991). The contiguous $\text{W\text{-}X\text{-}Y\text{-}Z}$
molecule has an advantage over these constructs as a consequence of its stable secondary and tertiary interactions and this stability can be exploited for its selection.

First, the ability of these three constructs to cooperate/compete to shuffle fragments and to generate fully functional W•X•Y•Z molecules through fragment recycling was tested. Each of these constructs would have to be defragmented into its respective RNA oligomers, and subsequently recombined to form W•X•Y•Z. Upon incubation of W•X•X•Z with $^{32}\text{P}$-body-labeled Y, W•Y•Y•Z with $^{32}\text{P}$-body-labeled X, and W•Y•X•Z with either $^{32}\text{P}$-body-labeled X or Y in self-assembling buffer, incorporation of radiolabeled RNA into molecules similar in size to W•X•Y•Z was observed (Figure 37). Similar results were obtained when all three constructs were incubated together with either $^{32}\text{P}$-body-labeled X, or $^{32}\text{P}$-body-labeled Y, or both in separate assays. The incorporation of labeled fragments varied among W•X•X•Z, W•Y•Y•Z, and W•Y•X•Z when incubated in isolation or together, suggesting that arrangements of W, X, Y, and Z give these constructs different levels of structural stability. The variations in their stabilities would lead to their uneven defragmentation rate, and consequently their differential recycling capacity. There is also a possibility to form W•Y•X•Z (similar in length to W•X•Y•Z) because it can gain some structural stability through non-covalent trans interactions. Therefore, the confirmation for the formation of W•X•Y•Z was obtained by excising a band $\approx$212-nt in length, and performing RT-PCR for W•X, W•X•Y, and W•X•Y•Z. In the cases of W•X•X•Z and W•Y•Y•Z, the catalytic activity was also restored after recycling, because the recycled W•X•Y•Z was capable of trans-splicing activity.
Further support for a global search for function was obtained from the self-assembly reaction of \( \textbf{W} \cdot \textbf{Y} \) and \( \textbf{X} \cdot \textbf{Z} \) to form \( \textbf{W} \cdot \textbf{X} \cdot \textbf{Y} \cdot \textbf{Z} \) (Figure 38). When 5\(^\prime\)-radiolabeled \( \textbf{W} \cdot \textbf{Y} \) was incubated with \( \textbf{X} \cdot \textbf{Z} \), formation of both \( \textbf{W} \cdot \textbf{X} \cdot \textbf{Y} \cdot \textbf{Z} \) and \( \textbf{W} \cdot \textbf{Y} \cdot \textbf{X} \cdot \textbf{Z} \) was observed, confirmed again by RT-PCR for \( \textbf{W} \cdot \textbf{X} \), \( \textbf{W} \cdot \textbf{X} \cdot \textbf{Y} \), and \( \textbf{W} \cdot \textbf{X} \cdot \textbf{Y} \cdot \textbf{Z} \). Although \( \textbf{W} \cdot \textbf{Y} \cdot \textbf{X} \cdot \textbf{Z} \) is not as stable as \( \textbf{W} \cdot \textbf{X} \cdot \textbf{Y} \cdot \textbf{Z} \), \( \textbf{W} \cdot \textbf{Y} \cdot \textbf{X} \cdot \textbf{Z} \) molecules were assembled because it involves only one transesterification reaction and the assembled molecules can gain stability using intermolecular interactions (Vaidya and Lehman, 2009). On the other hand, covalent assembly of \( \textbf{W} \cdot \textbf{X} \cdot \textbf{Y} \cdot \textbf{Z} \) from \( \textbf{W} \cdot \textbf{Y} \) and \( \textbf{X} \cdot \textbf{Z} \) requires two reverse reaction leading to free fragments that eventually recombine to form \( \textbf{W} \cdot \textbf{X} \cdot \textbf{Y} \cdot \textbf{Z} \) via three transesterification reactions. Nonetheless, shuffling of fragments to form \( \textbf{W} \cdot \textbf{X} \cdot \textbf{Y} \cdot \textbf{Z} \) molecules suggests that the recombinase ribozyme system has the capability to form various complexes (both non-covalent and covalent) that can recycle fragments to search for functions globally. Lastly, the recombination guide sequence and tags of \( \textbf{W} \cdot \textbf{X} \cdot \textbf{Y} \cdot \textbf{Z} \) system were demonstrated to be modifiable to assemble in cyclical cooperative network (Chapter 3; Vaidya \textit{et al.} 2012). Recycling was observed even in the complex cooperative assembly of three versions of \( \textbf{W} \cdot \textbf{X} \cdot \textbf{Y} \cdot \textbf{Z} \) network (Figure 39).

\textit{Local Search for Fitness}

Another important facet of recycling would be to increase the local fitness of a system by exploring different closely related sequences. Here, the ability of the \( \textbf{W} \cdot \textbf{X} \cdot \textbf{Y} \cdot \textbf{Z} \) recombination system to increase local fitness was investigated by reacting \( \textbf{W} \cdot \textbf{X} \cdot \textbf{Y} \cdot \textbf{Z} \) with mutations in \( \textbf{Z} \) fragment (\( \textbf{W} \cdot \textbf{X} \cdot \textbf{Y} \cdot \textbf{Z}^\dagger \)) with wild-type \( \textbf{Z} \) fragment (Figure 40). The
mutations in \( Z \) fragment lie in the catalytic core of the molecule and are lethal for its recombinase activity (Figure 41). When \( W \cdot X \cdot Y \cdot Z \) was incubated with wild-type \(^{32}P\)-body-labeled \( Z \), the incorporation of \(^{32}P\)-body-labeled \( Z \) in \( W \cdot X \cdot Y \cdot Z \) was observed demonstrating recycling of \( W \cdot X \cdot Y \) from \( W \cdot X \cdot Y \cdot Z \) to form fully functional \( W \cdot X \cdot Y \cdot Z \) (Figure 42). When tested for catalytic activity using \textit{trans}-splicing assay, recycled \( W \cdot X \cdot Y \cdot Z \) fully restored its activity (Figure 41), indicating the ability of a recombinase system to explore various sequences in a quasispecies cloud to increase local fitness.

\textit{Mechanism of Recycling in the Recombinase Ribozyme}

Recycling in the recombinase system is possible because of dynamic covalent transesterification chemistry. In the \( W \cdot X \cdot Y \cdot Z \) system, this can proceed through one of the two mechanisms – \( tF2 \) or \( R2F2 \) – that will result in either a sequential loss of nucleotides from 5’ end of recombination tag, or the entire tag “GGCAU” respectively (Draper \textit{et al.}, 2008). Once the entire tag has been removed, the fragments cannot be recognized by the system and thus cannot be recycled. The systems used in this study were designed to favor the \( tF2 \) mechanism (Draper \textit{et al.}, 2008). In a \( tF2 \) construct, if a removal of single G, then two Gs, and finally the entire recombination tag (GGCAU) occurs over time, it would demonstrate that the \( W \cdot X \cdot Y \cdot Z \) system is capable of at least three recycling events (Figure 43). Evidence for this phenomenon comes from additional analysis of sequences from previously studied complex RNA network system (Vaidya \textit{et al.}, 2012). When \( W \cdot X \cdot Y \cdot Z \) RNAs assembled at different time points were excised, converted to DNA by RT-PCR, and then subjected to nucleotide sequence analysis, a
trend of removal of one G, then two Gs, and lastly the entire tag from the recombination tag in recombined junctions was observed to occur in chronological order (Table 7). It is however possible for reactions to proceed without following the sequential order mentioned above, resulting in fewer than three recycling events (Figure 43).

DISCUSSION

Prebiotic scenarios, especially those advocating some form of a primordial soup, have been often the subject of a criticism that organics in an aqueous milieu would have been too dilute to sustain life’s origins (Schwartz, 1996; Wächtershäuser, 1988). In this study, we demonstrate recycling as a driving force for the search and selection of evolutionary function, by the use of W•X•Y•Z recombination system. We employ experimental data to bolster the idea that DCC can be manifested in the RNA World for both the global and local search for high fitness genotypes. It has been demonstrated earlier that RNA ligase ribozymes can be used to tie together external RNA molecules (Bartel and Szostak, 1993; Ekland et al., 1995; McGinness and Joyce, 2002), and that they can be designed to perform cross-catalytic replication of the ribozymes themselves (Kim and Joyce, 2004; Ohmori et al., 2011). Thus, hydrolysis and the subsequent ligation by a ligase ribozyme could have been a potent means to recycle fragments of RNA in early life. However, our findings suggest that recycling can be achieved via recombination reactions, avoiding some of the limitations of ligation reactions. In fact, the studied recombinase system
obtained from the group I intron suggests that recombination and recycling might have gone hand-in-hand during the origins of life.

The experimental data with the *Azoarcus* ribozyme extends the potency of recycling during global searches for function. Here, covalently contiguous $W\cdot X\cdot Y\cdot Z$ molecules can self-assemble and cross-assemble by shuffling the blocks of RNA fragments from three constructs – $W\cdot X\cdot X\cdot Z$, $W\cdot Y\cdot Y\cdot Z$, and $W\cdot Y\cdot X\cdot Z$. Rearrangement of other genotypes through cooperative group-level activities should also be possible. This finding, when considered in the light of the fact that single RNA sequences have previously been shown to assume multiple structures and/or catalyze multiple reactions (Lau and Unrau, 2009; Schultes and Bartel, 2000; Vaidya and Lehman, 2009), heightens the power of recycling as an evolutionary force. As such, recycling could have provided RNA with an additional tool to explore different functions by shuffling motifs among extant RNA sequences to search for varieties of functions. In addition, recycling can occur locally to optimize functions once motifs are established. A recycling of $W\cdot X\cdot Y$ from $W\cdot X\cdot Y\cdot Z$ to form fully functional $W\cdot X\cdot Y\cdot Z$ illustrates the local search for higher fitness. In fact, it may also be possible for a global search and a local search to occur simultaneously so that they can have synergistic interplay (Wagner, 2008). The studies described here were supported with kinetic Monte Carlo simulations performed by Dr. Sara I. Walker (Arizona State University). The simulations emphasized the significance of recycling in exploration of sequence space to select for stability and function.

Recycling can lead to an exponential growth of a replicating system as it allows systems to shuffle building blocks to reach a global optimum, where the threshold
concentrations of all component reagents are achieved (King, 1982). It prevents the proliferation of improperly assembled products through defragmentation enabling a build-up of reagents. The recyclase activity of $W \cdot X \cdot Y \cdot Z$ allows defragmentation of all molecules, preferentially defragmenting other non-functional unstable molecules, that can result in homogenization of the distribution of $W$, $X$, $Y$, and $Z$ fragments, thus favoring the rapid growth of $W \cdot X \cdot Y \cdot Z$ molecules. If the $W \cdot X \cdot Y \cdot Z$ molecules also possess a recombinase activity, then the system can have an additional advantage leading to shorter fixation time and rapid growth. Recycling could have therefore played critical role in controlling concentrations of reagents and thus providing replicative systems with adequate concentration of reactants.

Overall, it would be highly advantageous for the earliest reproducing system to exploit recycling to ameliorate the need for constant influx of new materials. This would have been true for the initial establishment of autocatalytic reactions of simple organic and inorganic compounds (Wächtershäuser, 1988), and/or later for the advent of self-reproducing polymers as targeted in the current work (King, 1982). Our study supports the notion that recycling could have been critical for exploring sequence space for both discovering novel functions and fine-tuning those functions once they are discovered. We posit that recycling makes the origins of life on the Earth (or elsewhere) more probable, and was likely a key feature in the regulation of resource availability in the first living systems.
MATERIALS AND METHODS

RNA preparation

Shorter RNAs (<50 nt, such as X) were purchased from TriLink Biotechnologies (San Diego, CA) and were gel purified and desalted prior to use. Longer RNAs were prepared by run-off transcription from double-stranded DNA templates constructed from recursive gene synthesis, and were gel purified and desalted prior to use. Strict contamination controls were always used to ensure correct sequence identities of all RNAs employed in the experiments.

Trans-splicing and Self-assembling Assays

The ability of Azoarcus ribozyme to perform trans-splicing, shuffling of fragments to form contiguous ribozyme, and covalent self-assemblies of ribozyme were assayed as described previously (Draper et al., 2008; Hayden and Lehman, 2006; Vaidya and Lehman, 2009; Vaidya et al., 2012). Briefly, RNA oligomers were incubated together at 48°C at a final concentration of 1 – 2 μM. All reactions contained a final concentration of 100 mM MgCl₂ and 30 mM EPPS buffer (pH 7.5). Reactions were carried out in 200 μl or 600 μl microcentrifuge tubes and were quenched by the addition of equal volume of gel-loading solution containing 8M urea, 200 mM EDTA, and bromophenol blue. RNAs were either 5´-end labeled with γ[^32P]·ATP or body-labeled with α[^32P]·ATP to allow visualization of the products via phosphorimaging on a Typhoon-Trio Plus instrument (GE Healthcare).
**Genotyping**

Full-length RNA covalent products were identified by comparison to *Azoarcus* ribozymes run as size control. For genotyping, the bands corresponding to recycled and self-assembled RNAs were carefully excised from the gel and subjected to reverse transcription using primers for X, Y, and Z, which targets the 3´ portions of X, Y, and Z respectively (Hayden and Lehman, 2006). These reactions were used as template to seed PCR to determine the arrangements of fragments in assembled products.
Figure 35. General scheme demonstrating importance of recycling. Each color block represents a different substrate or information unit. These blocks assemble to form polymers of different lengths and compositions as shown in the middle of the schematic. Some of these polymers are more stable than others, and their stabilities determine the rate at which they undergo spontaneous hydrolysis to be converted back to monomers. The most stable one is shown in the box. If the stable molecule can catalyze recycling, shown as a curved grey arrow, it can enhance hydrolysis of all polymers, preferentially hydrolyzing those which are less stable than itself. Further, if it has potential to replicate itself, it can utilize recycled monomers for replication.
Figure 36. Recycling of RNA fragments in the W•X•Y•Z system. Four fragments W, X, Y, and Z can recombine in various arrangements in a “global search” for a stable and functional polymer. Few of the possible outcomes of different recombination events are shown in middle. The secondary and tertiary interactions confer greater stability to W•X•Y•Z as compared to other polymers. As a result, other polymers are preferentially converted back to monomers, leaving more fragments to form W•X•Y•Z. It is also possible to exchange a fragment from W•X•Y•Z to increase the fitness of the molecule. The cartoon of “local search” for improved function is shown in the grey box. In the presence of multiple Z genotypes, the system can search for the fittest Z to improve the structural stability or activity of W•X•Y•Z. Here, a swapping between two genotypes Z and Z‡ is shown as an example.
Figure 37. Shuffling of fragments to demonstrate the global search for a covalently contiguous ribozyme. W•X•X•Z was incubated with $^{32}$P-body-labeled Y. Formation of contiguous W•X•Y•Z molecules was observed along with other intermediates. One of the intermediates is pointed with an arrow.
Figure 38. Shuffling of fragments during self-assembly of ribozyme. W•Y molecule, 5’- radiolabeled with $^{32}$P, was incubated with X•Z and formation of W•X•Y•Z along with W•Y•X•Z was observed.
Figure 39. Demonstration of recycling in a cyclical cooperative network. The internal guide sequence (IGS) and IGS recognition tag of the recombinase system can be modified to create a cyclical cooperative network of three replicating RNA species (Vaidya et al., 2012). Briefly, the E₁ ribozyme recognizes RNA fragments from a second system E₂, and assembles the fragments into a covalently linked E₂. The E₂ ribozyme recognizes RNA fragments from a third system E₃, and assembles the fragments into a covalently linked E₃. And, the E₃ ribozyme recognizes RNA fragments from a first system E₁, and assembles the fragments into a covalently linked E₁. Here, two reactions are initiated where W-containing fragment of E₃ (W•X•Y) that was 5´ radiolabeled with γ[^32P]·ATP are incubated either with E₁ only or with E₁, E₂, and E₃ together. In former case, incorporation of W•X•Y into a full-length E₃ was not observed. While in the later case, incorporation of W•X•Y in E₃ was detected after 5 h as a result of E₂ recycling Z from E₃ to make covalently contiguous E₃. Similar results were observed for E₁ and E₂.
Figure 40. Secondary structure of catalytic core of the ribozyme. Fragments Y and Z form the catalytic core of the ribozyme shown in the grey box. Four nucleotides in Z, GUCC (shown in black box), were mutated to CGUA to get a construct \( W \cdot X \cdot Y \cdot Z^\dagger \) that is catalytically inactive. CGUA is also a recognition site for the restriction endonuclease SnaBI that allows a differentiation between \( W \cdot X \cdot Y \cdot Z \) and the recycled \( W \cdot X \cdot Y \cdot Z \) that is catalytically active.
Figure 41. Restoration of catalytic activity after recycling. When the W•X•Y•Z ribozyme with mutation in the catalytic core (W•X•Y•Z‡) was tested for its trans-splicing activity using the substrate SNL1a (5’-GGCAU•AAAUAUUUAAUA-3’), with a recombination tag CAU on the 5’ end, a transesterification product was not detected (left lanes). On the other hand, trans-splicing activity was fully restored on the recycled W•X•Y•Z (right lanes).
Figure 42. Demonstration of recycling in recombinase ribozyme to increase local fitness. $W \cdot X \cdot Y \cdot Z^\ddagger$ was incubated with $^{32}$P-body-labeled $Z$ in self-assembling buffer. Incorporation of $^{32}$P-body-labeled $Z$ into full length $W \cdot X \cdot Y \cdot Z$ molecules was observed to increase with time demonstrating the ability of system to recycle portions of molecules. Various intermediate products were observed (arrows), especially during earlier time points, on the way to form recycled $W \cdot X \cdot Y \cdot Z$. 
Figure 43. Possible mechanism of recycling in the recombinase system. In this model, recombination and recycling events occurring at the Y–Z junction is diagrammed. The first recombination event leads to a covalent assembly of W•X•Y and Z to form W•X•Y•Z (top left). Here, the IGS of the ribozyme recognizes the recombination tag on the 3′ end of W•X•Y (GGCAU; CAU in oval) folded together with Z and catalyzes a transesterification reaction, where 3′ hydroxyl of W•X•Y attacks an ester bond between the alpha (α) and beta (β) phosphates in 5′ end of Z to form covalently linked W•X•Y•Z molecule with the release of pyrophosphate. The resultant W•X•Y•Z has an insertion of GGCAU in the Y–Z junction. The newly formed W•X•Y•Z can undergo the first recycling event shown in grey dotted box in the upper left. The initial step of recycling involves catalyzed hydrolysis (this could also happen spontaneously), shown as Ia, to regenerate fragments, W•X•Y and Z with GGCAU in the 5′ end. After the formation of enzyme-substrate (E•S) complex (Ib), the same or different enzyme can catalyze another transesterification reaction (Ic), resulting in W•X•Y•Z with GCAU in the Y–Z junction and the release of a guanosine. The newly formed W•X•Y•Z can enter second recycling event depicted in the lower left dotted grey box. Similar to the first recycling event, subsequent hydrolysis (IIa) and covalent assembly (IIc) leads to a formation of W•X•Y•Z, with only CAU in Y–Z junction and the release of second guanosine. The assembled W•X•Y•Z can enter another round of hydrolysis to regenerate W•X•Y and Z. But the Z fragment formed here contains only CAU and it is more favorable to undergo R2F2 mechanism (Draper et al., 2008). This results in the formation of Y–Z junction with no tag. As a consequence, further hydrolysis will result in the Z fragment with no recombination tag and hence cannot be utilized further for recombination. It is also possible for the reaction to proceed without following the sequential order shown in the diagram, resulting in fewer than three recycling events.
Table 7. Sequence analysis of recombined junctions from a complex RNA network (Vaidya et al., 2012). W•X•Y•Z RNAs were self-assembled from either two or four fragments, and the nucleotide sequence of the resulting assembled RNAs were analyzed using high throughput sequencing or manual sequencing. Tag sequences used were either CAU or CNU, where the middle nucleotide was randomized for network formation (Vaidya et al., 2012). $N_{ave}$ indicates average number of sequences analyzed for each time point. In all cases, the frequencies of assembled molecules with two G insertions and single G insertions decrease over time. In contrast, the frequencies of molecules with CAUs and no tags (–) increase over time, demonstrating the recycling phenomenon (Figure 43). Because the trinucleotide CAU gives additional stability to the loop region, the assembled molecules do not tend to recycle unless they are incubated in the assembly buffer for a longer period.
CHAPTER 5: CONCLUSION

A robust, prebiotically plausible route to functional RNA oligomers, if achieved, will dramatically reinforce the case for the RNA world hypothesis. The efforts of prebiotic chemists and nucleic acid biochemists in last half century have lent much support for the idea of the RNA world. Various potential chemical routes to RNA from prebiotic molecules have been suggested (Anastasi et al., 2007). The synthesis of activated pyrimidine nucleotides from cyanamide, cyanoacetylene, glycoaldehyde, glyceraldehydes, and inorganic phosphates – all plausible prebiotic molecules – has been demonstrated (Powner et al., 2009) eliminating the requirement of difficult selective formation of ribose (Ricardo et al., 2004) and the requirement of inefficient addition of nucleobases to ribose (Fuller et al., 1972). Once activated nucleotides are formed, they can be polymerized to form RNA oligomers of approximately 50 nt in length (Ferris et al., 1996; Monnard et al., 2003; Verlander et al., 1973). These oligomers in turn can be recombined to form longer replicating RNA systems with complex functions (Hayden et al., 2005; Hayden and Lehman, 2006; Lehman, 2008; Striggles et al., 2006).

Such RNA systems could then develop a crude replication ability and be subjected to natural selection. Yet these replicating systems would still have to deal with various obstructions during the early RNA world. The hostile environment in the early Earth would not only make the synthesis and polymerization of activated nucleotides difficult, but the RNA polymers, once formed, would also be susceptible to random hydrolysis. In this thesis, I characterized the RNA recombination system derived from group I intron and demonstrated how its novel features would provide the recombination
system with advantages to deal with the various potential obstacles in the early evolution of life.

The first complication, discussed in Chapter 2, deals with limited genotypic variability in a prebiotic pool. The polymerization of RNA oligomers in prebiotic world had to rely on chemical processes that were not as efficient as the processes in contemporary biology. Specifically, there would have been limitations in the availability of precursor molecules. Hence, the prebiotic pool would include only few of all the possible RNA sequences. In contrast, the origin and evolution of the first replicating systems would benefit from the exploration of many sequences. If these systems were capable of utilizing the available limited genotypes for achieving complex functions, the replication and eventually evolution of RNAs would be rendered more plausible. In Chapter 2, I demonstrated that a single RNA genotype can achieve up to three distinct phenotypes in a single reaction during recombination events. I designed a ribozyme construct without an IGS by deleting 22 nt from 5′ end of the ribozyme. This “blind” ribozyme utilized a differential folding potential of a single RNA fragment – as a portion of a catalyst, as a substrate, and provide an exogenous IGS, to be a fully functional ribozyme. The removal of 22 nt also made the recombination system shorter, thereby narrowing the gap between the limitation in length of RNA that can be polymerized in prebiotic scenario and the length of self-replicating RNA molecules.

The second obstacle would be vulnerability of early replicators to mutations, a topic that I highlighted in Chapter 3. The lack of complex error-correcting enzymes (and also exposure to harsh conditions) would introduce mutations in the earlier replicating
molecules, inevitably leading to informational decay. It has been suggested that webs of functionally and genetically related replicators could provide the earliest replicators with a possible solution to mutational meltdown, because mutations in one component of the network can be compensated by the other component(s) of the network (Eigen, 1971; Kun et al., 2005). However, networks of replicating molecules with more than two members (Kim and Joyce, 2004; Lee et al., 1997) have never been empirically demonstrated. In Chapter 3, I exploited recombination system from group I intron to demonstrate that complex networks of replicating RNA molecules are indeed possible. Further, I was able to show that a cooperative network of RNA systems can out-compete selfish replicators, and moreover that the simpler cycles of replicators has the potential to expand to more complex ones.

The third obstacle, which I discussed in Chapter 4, focuses on inadequate concentrations of precursor molecules required for the exponential growth of replicating systems during the early stages of life. If the precursor molecules were recycled from non-informative and non-functional polymers that were trapping these valuable reagents, it would avoid the necessity of constant supply of building blocks (King, 1982). In addition, if the larger fragments of polymers (rather than monomers) can be recycled, it would not only circumvent the need for new monomer sources, but also would bypass the inefficient polymerization processes. I demonstrated the ability of RNA recombination system to recycle fragments of RNA to explore sequence space to search for function. I was also able to show that the system can increase the fitness by exchanging the RNA motifs once the function was established. From simulation results, we saw that the
recombination system was also able to exhibit a dynamic threshold phenomenon where we observed explosive growth of system after recycling resulted in adequate concentrations of all the components of the system.

Overall, this thesis demonstrates how an RNA recombination system can not only lead to the build-up of complexity by increasing the length of RNA polymers, but also has capacity to tackle at least three major obstacles during the origins of life: a limited genetic variability, an error threshold, and an inadequate resource of precursor molecules. All these characteristics of an RNA recombination system could have been synergistic, maybe at the same time, and led to a relatively rapid evolution of complex, self-replicating ribozymes. All this suggests that a simple RNA organism that originated from recombination of RNA fragments could have been the Earth’s first life form and these newly discovered characteristics discussed in my thesis could have assisted them to evolve into more complex RNA organisms.

A key RNA molecule in the early history of life would be a minimal ribozyme that originated from prebiotic chemistry, and once formed, could catalyze the replication of RNA molecules, including copies of itself. This type of molecule could have taken advantage of recombination reactions because recombination events are enthalpically energy neutral. The RNA system that I studied in this thesis is approximately 200 nt long and requires various structural domain to be stable. The assembly of the full-length ribozyme so far has been designed to occur from four RNA oligomers that range from 39 – 63 nt in length. One of the challenges would be to make this system assemble from more fragments that are smaller and are significantly easier to make in the prebiotic pool.
This should be experimentally accessible through identification of fragmentation points in the ribozyme in the regions that have no direct interaction with the catalytic core.

Another important challenge is to design a single RNA system that can recombine without a requirement of multiple fragments. The catalytic core of the ribozyme resides on P7 helix (Adams et al., 2004) and comparative sequence analysis of other group I introns has suggested that some structural domains could be eliminated. Thus, a single RNA recombining fragment could be designed by removing stability loop regions of the molecule and randomizing certain regions of the molecule that have a capacity to form a scaffold for a catalytic core. This single pool of RNA could be selected in vitro to form a recombination system that utilizes the novel characteristics discussed in this thesis (formation of multiple phenotypes from a single genotype, recycling and cooperation among fragments), and thus to form a minimal key RNA molecule.
REFERENCES


Vaidya N, Walker SI and Lehman N. Recycling of informational units leads to selection of replicators in a prebiotic soup. Submitted to *Chem Biol*.


