Recycling of Informational Units Leads to Selection of Replicators in a Prebiotic Soup

Nilesh Vaidya,1,3,4 Sara Imari Walker,2,4 and Niles Lehman1,*

1Department of Chemistry, Portland State University, P.O. Box 751, Portland, OR 97207, USA
2BEYOND Center for Fundamental Concepts in Science, Arizona State University, P.O. Box 871504, Tempe, AZ 85287, USA
3Present address: Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08544, USA
4These authors contributed equally to this work
*Correspondence: niles@pdx.edu
http://dx.doi.org/10.1016/j.chembiol.2013.01.007

SUMMARY

Prebiotic chemical reactions would have been greatly aided by a process whereby living materials could have been recycled under conditions of limiting resources. Recombination of RNA fragments is a viable means of recycling but has not been demonstrated. Using systems based on the Azoarcus group I intron ribozyme, computational Monte Carlo studies indicate that a moderate level of recycling activity, spontaneous or catalyzed, leads to the most robust selection scenarios. It is interesting that recycling leads to a threshold effect where a dominant species suddenly jumps to fixation. In conjunction, laboratory studies with the Azoarcus ribozyme corroborate these results, showing that mixtures of scrambled and/or deleteriously mutated molecules can recycle their component fragments to generate fully functional recombinase ribozymes. These studies highlight the importance of recombination and recycling jointly in the advent of living systems.

INTRODUCTION

The availability of prebiotic materials would have been an key limiting factor during the origins of life on Earth. Feasible prebiotic pathways to synthesize biologically relevant precursor molecules were impeded both by mixtures of products formed and by product yields (Zubay and Mui, 2001; Orgel, 2004). 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 an 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 1A). In short, without recycling that could dissociate polymers and reuse monomers, there may not have been adequate material to get life started.

For example, the synthesis of 1 pmole of all possible sequences of 50 nucleotide (nt)-long RNA molecules (e.g., hammerhead ribozymes) by an irreversible process requires roughly the total weight of carbon (10^{19} kg) in the Earth’s crust (Engelhart and Hud, 2010).

The importance of recycling in prebiotic chemistry has been addressed in a few prior 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 (Plisson et al., 2004; Blackmond and Matar, 2008). Environmentally driven recycling was 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 such as those of amphiphilic molecules suspended in aqueous solutions (Monnard and Deamer, 2002). However, the assembly of supramolecular structures does not include covalent bonds and carries information limited to reactivity and physicochemical properties (Graham et al., 2004) and may be insufficient to evolve biological systems.

A fundamentally different assembly mechanism could exploit dynamic covalent chemistry (DCC) where noncovalent interactions are used to make conditionally reversible covalent linkages so that the biopolymers could be formed, broken, and reformed without the need for continual supply of reagents (Oh et al., 2001; Whitney et al., 2004; Fulton, 2008; Ladame, 2008; Hickman et al., 2008). 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 (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). However, 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.

Several mechanisms of recycling could have been operational in an RNA world. The simplest possibility is the hydrolysis and subsequent religation 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 molecules that contain four fragments as final replication and selection in the early stages of life on Earth. We also provide empirical evidence of the reversible nature of trans-esterification reaction in the four-piece self-assembling Azoarcus system that can lead to its selection (Figure 1C).

RESULTS
Modeling Recycling and Selection with Spontaneous Hydrolysis
First, we performed a kinetic Monte Carlo simulation of recycling and selection, roughly based on the Azoarcus ribozyme. Here, the initial computational analysis of the $W \cdot X \cdot Y \cdot Z$ self-assembling system was performed with spontaneous hydrolysis as the only means to defragment assembled products. In the experimental system, $W$ can be present only at the 5' end and $Z$ only at the 3' end of the molecules (Hayden and Lehman, 2006), yet all possible combinations of $W, X, Y,$ and $Z$ were included in our modeling here. Also, despite a possibility that polymers of various lengths can form in the laboratory, the modeling only accommodated molecules that contain four fragments as final products. We only allowed tetramers to be formed from the assembly of four monomers, and the tetramers are directly hydrolyzed to produce four monomer fragments. Thus, there

Figure 1. Schematic of Recycling
(A) General recycling scheme. Each color block represents a different substrate or information unit; the most stable set is shown in the box. If the stable molecule can catalyze recycling, shown as a curved gray arrow, it can enhance hydrolysis of all polymers, preferentially hydrolyzing those that are less stable than itself. Moreover, it does not rely on high-energy intermediates like the template-directed mechanism discussed earlier. It thus has the potential to enable reutilization of larger biopolymer molecules that contain four fragments as final.

(B) Design of recombinase ribozyme from Azoarcus group I intron. The 198 nt ribozyme can be fragmented to obtain four oligomers, $W, X, Y,$ and $Z$. These fragments are capable of self-assembling into a full-length molecule, $W \cdot X \cdot Y \cdot Z$; when they are coincubated in 100 mM MgCl$_2$ they form noncovalent but catalytically active trans complexes through secondary and tertiary interactions (top right) that can use other RNA fragments as substrates and reassemble the covalently contiguous molecule (bottom right). Both catalysts can also catalyze reverse reaction to recycle fragments.

(C) Recycling of RNA fragments in the $W \cdot X \cdot Y \cdot Z$ system. Fragments $W, X, Y,$ and $Z$ can recombine in various arrangements in a global search for a stable and functional polymer. Some possible outcomes of different recombination events are shown in the middle. A cartoon of a local search for improved function is shown. In the presence of multiple $Z$ genotypes, the system can search for the fittest $Z$ to improve the structural stability or activity of $W \cdot X \cdot Y \cdot Z$. An example swapping between two genotypes $Z$ and $Z_1$ is shown.

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 molecules that contain four fragments as final.
are 256 possible arrangements of fragments (including combinations with replacement). The assembly events to form tetramers can be considered as a “global search” over extant assemblies for the most stable assembled structures, W- X- Y- Z in this case (Figure 1C). In the experimental system, the W- X- Y- Z molecule has an advantage over other sequences as a consequence of its stable secondary and tertiary interactions (Hayden and Lehman, 2006). Therefore, for the computational study, the stabilities of other tetramers in the library were ranked based on their bond structure similarity to W- X- Y- Z, where the distance from W- X- Y- Z was calculated by determining how many bonds a given tetramer has in common with W- X- Y- Z, regardless of ordering or redundancy. The stability landscape is thus dependent on the bonds formed between unit pairs in the assembled tetramers rather than on sequence similarity. Using these calculated distances, bond strengths were assumed to follow a Gaussian distribution for zero-error (W- X- Y- Z), one-error, two-error, and three-error molecules with relative susceptibilities to hydrolysis calculated as \( \exp(n^2/2)k_h \), where \( n \) is the number of bond “errors,” and \( k_h \) is the spontaneous hydrolysis rate constant. For \( n = 0, 1, 2, \) and 3 error molecules, this yields values of 1.0 \( k_h \), 1.65 \( k_h \), 7.39 \( k_h \), and 90.0 \( k_h \) respectively. “Error” (\( \psi \)) here refers to the number of bonds which are not W- X, X- Y, or Y- Z. The spontaneous hydrolysis defragments polymers into oligomers, recycling them for reuse in subsequent rounds of assembly, thus enabling selection for the most stable configurations. The replicative rate constant was set to \( k_r = 0.0001 \), and the rate constant for spontaneous assembly was set to \( k_s = 10^{-9} \), so that both rates are sequence independent (all bond formation rates are equal). However, replication was assumed to rely on the availability of W, X, Y, and Z, introducing a dynamic dependence of tetramer replication rates on resource availability (see Experimental Procedures).

Simulations to illustrate the role of recycling via spontaneous hydrolysis were carried out for a fixed system size of 20,000 informational units, comprising 5,000 each of the four fragment species (Figure 2). Populations were initialized with a small population of 50 tetramers drawn at random from all possible arrangements of monomers so that approximately one in five of the 256 possible tetramers will be represented in a given simulation run (chosen to be small enough so that W- X- Y- Z does not dominate the selection dynamics for all runs, because it is not present in all simulation instantiations). The system was then permitted to evolve until a steady-state population was achieved. Data were averaged over 100 simulations runs for each value of \( k_h \).

Selection, corresponding to the total fixation of only one tetramer species, was observed to be strongly dependent on \( k_h \), where an increase in the rate constant yields stronger selection for more stable structures and to the total extinction of populations dominated by low-stability sequences (Figure S1 available online). In the absence of spontaneous hydrolysis \( (k_h = 0) \), i.e., the limit of no recycling, all tetramers present in the initial population would clearly survive, resulting in a diverse random population. For low values of \( k_h \), all possible two-error, one-error, and zero-error tetramer species are stable against hydrolysis. Selection probabilities for populations of \( \psi \)-error mutants therefore map the stability landscape and are proportionate to the total number of \( \psi \)-error mutants (i.e., at low \( k_h \), survival is stochastically determined and the initial random populations lead to selection of two-error, one-error, or zero-error mutants based on the initial abundances because there is little competitive advantage for stability) (Figure 2A). Increasing \( k_h \) to 0.1 leads to increased competition and the total recycling of two-error mutants. At lower hydrolysis rates, populations that would have selected two-error mutants instead experienced recycled and requisitioning to support the growth of populations of one-error or zero-error mutants or otherwise become totally extinct. When \( k_h \) was increased to 0.25, the spontaneous recycling rate is sufficiently high so that populations supporting successful selection of one-error mutants begin to decline in
**Mathematical Modeling of Catalyzed Recycling**

(A) Time evolution of extant species population for catalyzed recycling with rate constant $k_{ch} = 0.00001$. Rate constants for replication and spontaneous assembly are set as $k_r = 0.0001$ and $k_s = 10^{-5}$, respectively. The legend highlights only those sequences for which the population achieves $>10\%$ of the total tetramer population. The inset shows the time evolution of the Shannon diversity of the extant sequence population (normalized to the total size of the sequence space).

See Figure S2 for simulation runs for varying $k_{ch}$ values.

(B) Time evolution of monomer abundances for catalyzed recycling with rate constant $k_{ch} = 0.00001$. Rate constants for replication and spontaneous assembly are set as $k_r = 0.0001$ and $k_s = 10^{-5}$, respectively. See Figure S2 for simulation runs for varying $k_{ch}$ values.

(C) Average fixation time (triangles) and average steady-state W·X·Y·Z abundance (circles) for catalyzed recycling. Rate constants for replication and spontaneous assembly are set as $k_r = 0.0001$ and $k_s = 10^{-5}$, respectively, with 256 possible tetraperomers. The inset shows the time evolution of the inset shows the time evolution of the sequence space. Increasing $k_h$ still further, W·X·Y·Z is no longer stable against hydrolysis and all populations go extinct, indicating that hydrolysis is important for recycling precursor molecules and for exploring sequence space to discover functionality. However, a greater hydrolysis rate is lethal for the reproducing system; this rate must fall within a fixed range of values to support selection of the most stable sequence W·X·Y·Z.

**Modeling Recycling and Selection with Catalyzed Recycling**

The results of the previous section suggest that a stable W·X·Y·Z molecule capable of catalyzing hydrolysis could dynamically modulate the rate of recycling and perhaps even lead to its own selection. Because the total recycling rate would depend on the catalyst abundance, a catalytic recyclase would need to control the recycling rate to stay below the lethality threshold. Recyclase activity could therefore be selected for as an important functional role in the earliest stages of life. To test this hypothesis, recyclase activity for W·X·Y·Z was introduced to the simulation model (Figure 3). In contrast to the previous model where the reverse reaction was carried out only via spontaneous hydrolysis, here there is no spontaneous hydrolysis ($k_h = 0$), and all recycling is catalytically driven by the activity of W·X·Y·Z. Other sequences have no recyclase activity. The recycling rate is therefore dependent on the catalytic recycling rate constant $k_{ch}$ and the total abundance of W·X·Y·Z. The stability landscape, resource-dependent replication, and the total fixed population size of W, X, Y, and Z were set up identically to the model described earlier. However, to demonstrate how W·X·Y·Z competes all other sequences in the pool, and to connect with experimental systems that are initialized with large populations of tetraperomers, the catalytic recycling simulations were initialized with a randomized population of 1,000 tetraperomers, so that all of the 256 possible tetraperomers are likely to be represented in a given simulation run (some with multiple copies). Data were collected for 100 simulations for each of the catalytic rate constants explored in Figure 3.

The initial dynamics of the extent population of tetraperomers is marked by competition and rapid consumption of resources, yielding a high initial extinction rate (Figure 3A, inset). The system then transitions to resource-limited dynamics, where the resource availability is driven by the global recycling rate, dynamically determined by the stability of the extent population of sequences and the total abundance of the catalyst W·X·Y·Z. The initial selection favors one-error mutants with close bond
similarity to $W \cdot X \cdot Y \cdot Z$ with the five most abundant sequences highlighted in the legend of Figure 3A, along with $W \cdot X \cdot Y \cdot Z$. During this period, the population maintains a nearly constant, and relatively high, diversity. Although $W \cdot X \cdot Y \cdot Z$ is initially very low in abundance as a consequence of an early selective disadvantage under resource-limited conditions (discussed later), its catalytic activity drives the dynamic reshuffling of the extant sequence library. Higher rates of fragmentation of less stable structures and the recycling of resources to support replication of extant sequences dynamically change the selection landscape over time. Once resource abundances matched the resources distribution of $W \cdot X \cdot Y \cdot Z$ (i.e., equal amounts of $W$, $X$, $Y$, and $Z$ monomers), $W \cdot X \cdot Y \cdot Z$ gains replicative advantage and rapidly takes over the population as a result of a stochastically determined “threshold effect,” observed in the explosive growth of the $W \cdot X \cdot Y \cdot Z$ population (black curve in Figure 3A). The stability and recyclease activity of $W \cdot X \cdot Y \cdot Z$ result in its selection causing the population diversity to drop to zero, indicating total fixation of $W \cdot X \cdot Y \cdot Z$.

The simulations imply significant resource dependence in the early selection of sequences prior to fixation of $W \cdot X \cdot Y \cdot Z$ (Figure 3B), where the distribution of monomers correlates with the formation and selection of $W \cdot X \cdot Y \cdot Z$. Because $X$ and $Y$ are each part of only one bonded pair, sequences with close bond structure similarity to $W \cdot X \cdot Y \cdot Z$ are twice as likely to contain $X$ and $Y$ oligomers as $W$ and $Z$. In general, more stable sequences are therefore enriched in $X$ and $Y$. This leads to strong competition for the monomer species $X$ and $Y$, which become depleted early, being consumed by a competitive population of relatively stable sequences with close bond structure similarity to $W \cdot X \cdot Y \cdot Z$ (Figure 3A). The rapid depletion of $X$ and $Y$ yields a mismatch between $W \cdot X \cdot Y \cdot Z$ and the available resources, causing $W \cdot X \cdot Y \cdot Z$ replication to be less competitive as compared to other sequences in the pool. As the catalytic activity of $W \cdot X \cdot Y \cdot Z$ recycles monomer resources, the resource abundances slowly approach an equal distribution among all four species of monomer. Once a uniform distribution of monomer population frequencies is achieved, replication of the $W \cdot X \cdot Y \cdot Z$ is favored over other closely related sequences, resulting in growth of both species through recycling of other extant sequences that no longer have the advantage of enhanced resource availability. The greater stability of $W \cdot X \cdot Y \cdot Z$ allows it to outcompete mutants efficiently during this period of rapid recycling. Selection of $W \cdot X \cdot Y \cdot Z$ and extinction of closely related competitors fixes homogeneously distributed resources in the system, leading to the permanent fixation of $W \cdot X \cdot Y \cdot Z$ within the extant population.

The fixation time for $W \cdot X \cdot Y \cdot Z$ is stochastically determined by the interpopulation dynamics of the extant sequence library and monomer resource distributions, yet the fixation time and catalytic recycling rate constant demonstrate a direct correlation (Figure 3C). Selection timescales are, on average, longer for low catalytic efficiencies and decrease with increasing recycling rates. However, there is a tradeoff between fixation time and the final $W \cdot X \cdot Y \cdot Z$ abundance (Figure 3C; Figure S2). On the other hand, the monomer abundance increases with the increase in catalytic rate constant as the polymers are more rapidly disassembled (Figure S2). This suggests that the optimal scenario for replicating systems would be to possess a moderate recycling rate so that the fixation time is low but fixation yields a relatively high abundance.

**Empirical Demonstration of Recycling**

Next, a global search for functional molecules was demonstrated experimentally in the $W \cdot X \cdot Y \cdot Z$ system, where recyclease activity is manifest as recombinase activity. Three different tetrameric constructs ($W \cdot X \cdot X \cdot Z$, $W \cdot Y \cdot Y \cdot Z$, and $W \cdot Y \cdot X \cdot Z$) were designed in the laboratory out of all possible arrangements. Trans-splicing catalytic activity is not possible for $W \cdot X \cdot X \cdot Z$ or $W \cdot Y \cdot Y \cdot Z$ because one of the fragments ($Y$ and $X$, respectively) required for folding into a fully functional ribozyme is missing (cf. Vaidya and Lehman, 2009). However, the “transposant” $W \cdot Y \cdot X \cdot Z$ retained minimal activity (data not shown) as a consequence of the ability of RNA fragments to function in trans and provide multiple functions (Beaudry and Joyce, 1990; van der Horst et al., 1991; Vaidya and Lehman, 2009). Here, the ability of these three constructs to cooperate/compete to shuffle fragments and to generate fully functional $W \cdot X \cdot Y \cdot Z$ molecules through fragment recycling was tested. Each of these constructs would have to be fragmented into its respective RNA oligomers, and subsequently recombined to form $W \cdot X \cdot Y \cdot Z$.

Upon incubation of $W \cdot X \cdot X \cdot Z$ with $^{32}$P-body-labeled $Y$, incubation of $W \cdot Y \cdot Y \cdot Z$ with $^{32}$P-body-labeled $X$, and incubation of $W \cdot Y \cdot X \cdot Z$ with either $^{32}$P-body-labeled $X$ or $Y$ in self-assembling buffer, incorporation of radiolabeled RNA into $W \cdot X \cdot Y \cdot Z$ product molecules was observed (Figure 4). Of note is that absolutely no reaction was observed when $W \cdot X \cdot X \cdot Z$ was incubated with $^{32}$P-body-labeled $X$ or when $W \cdot Y \cdot Y \cdot Z$ was incubated with $^{32}$P-body-labeled $Y$ (Figure 4A), as these negative controls each lacked one of the fragments needed to compose a functional enzyme in trans. A confirmation of the formation of $W \cdot X \cdot Y \cdot Z$ was obtained by excising the putative product band $\approx 200$ nt in length and assaying the RNA for trans-splicing activity; in each case, the RNA could splice an exogenous RNA oligonucleotide (data not shown).

The extents of incorporation of labeled fragments varied among $W \cdot X \cdot X \cdot Z$, $W \cdot Y \cdot Y \cdot Z$, and $W \cdot Y \cdot X \cdot Z$ when incubated in isolation or together (in the order $W \cdot Y \cdot X \cdot Z > W \cdot X \cdot X \cdot Z > W \cdot Y \cdot Y \cdot Z$), 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, features highlighted in the computational studies described earlier. In fact, a small but significant size difference was seen in the $W \cdot Y \cdot Y \cdot Z$ reaction compared to either the $W \cdot X \cdot X \cdot Z$ or the $W \cdot Y \cdot X \cdot Z$ reactions. The $W \cdot Y \cdot Y \cdot Z$ reaction product migrated at the same rate as a marker band (~202 nt) produced by reacting $W \cdot X \cdot Y$ with $Z$ (Hayden et al., 2008), while the $W \cdot X \cdot X \cdot Z$ or the $W \cdot Y \cdot X \cdot Z$ reactions produced products that were slightly larger (Figure 4). This suggests that two different synthesis pathways are possible in this system, and accordingly, both products were observed in a reaction where both $W \cdot X \cdot X \cdot Z$ and $W \cdot Y \cdot Y \cdot Z$ were incubated in the presence of $^{32}$P-body-labeled $X$ and $Y$ in self-assembling buffer (Figure 4B). Plausible mechanisms for this are described later.

Further support for a global search for function was obtained from the self-assembly reaction of $W \cdot Y$ and $X \cdot Z$ to form
Figure 4. Empirical Demonstration of Global Search via Recycling

(A) Shuffling of fragments to demonstrate the global search for a covalently contiguous ribozyme. W-X-X-Z (2 μM) was incubated with 32P-body-labeled Y (2 μM), or W-Y-Y-Z (2 μM) was incubated with 32P-body-labeled X (2 μM) at 48°C and analyzed by 8% polyacrylamide-8 M urea gel electrophoresis. Time points taken were at 0, 2, 4, and 6 hr. The product band W-X-Y-Z is indicated with an arrow; RNA from this position in the gel was excised and assayed for trans-splicing activity (see text). The marker lane (M) contained the product of a control reaction where 5’-end-labeled W-X was reacted with Z to give W-X-Y-Z at a size of 202 nt (Hayden et al., 2008). Alternative larger products such as W-Y-Y-Z-X (232 nt) that can be produced by various means are indicated with the carat symbol (†). Negative controls were performed in which W-X-Z (2 μM) was incubated with 32P-body-labeled X (2 μM) or in which W-Y-Y-Z (2 μM) was incubated with 32P-body-labeled Y (2 μM). The gel image intensity was enhanced for the W-X-Z reaction compared to the W-Y-Y-Z reaction; otherwise, no alterations were made to the image.

(B) The same experiment as in (A) was performed, except W-X-X-Z (2 μM) was incubated with W-Y-Y-Z (2 μM), 32P-body-labeled Y (2 μM), and 32P-body-labeled X (2 μM) together. Here, two product bands can be observed, demonstrating that the product initiating from W-X-X-Z is slightly larger than that from W-Y-Y-Z, implicating two slightly different synthesis pathways (see text).
Chemistry & Biology
Recycling in Catalytic RNA

Figure 5. Demonstration of Local Search via Recycling
(A) Base pairing in the catalytic core of the ribozyme between Y and Z (gray box). Four nucleotides in Z, GUCC (black box), were mutated to CGUA to make catalytically inactive construct W·X·Y·Z*. GUCC is a recognition site for the restriction endonuclease SnaBI, allowing a differentiation between W·X·Y·Z* and the active recycled W·X·Y·Z.
(B) Demonstration of recycling by electrophoresis. Unlabeled W·X·Y·Z* was incubated with 32P-body-labeled Z. Incorporation of 32P into full-length W·X·Y·Z molecules increased with time, demonstrating the ability of system to recycle portions of molecules. Upon gel excision, these molecules are catalytically active. Intermediate products of recycled W·X·Y·Z were observed (arrows), especially during earlier time points (see text). See also Figure S4.
(C) Mathematical modeling of local searches for fitness through spontaneous recycling. Shown are the steady-state abundances of W·X·Y·Z, (n = 0, 1,…,4) variant tetramer species, demonstrating a selection bias for the wild-type W·X·Y·Z, with the final abundances of other variants decreasing with increased lethality of Z, Z, Z, Z mutant species. Rate constants for replication and spontaneous assembly are set as k· = 0.0001 and k· = 10^{-3} respectively, with varying catalytic rate constant kµ. See Figure S5 for runs under other values of kµ.

W·X·Y·Z (Figure 4D). When S’-radiolabeled W·Y was incubated with X·Z, formation of bands corresponding in size to both W·X·Y·Z (≈ 150 nt) and the full-length RNA (≈ 200 nt) was observed, increasing over time and peaking at 4 hr (Figure 4D). To confirm the identities of these bands, the RNA was excised from the gel and subjected to RT-PCR, using reverse-transcription primers for Y and Z followed by PCR using primer pairs for W·Y or W·Z. Because no RNA larger than 110 nt was used to seed these experiments, PCR products obtained in this manner must be products of recycling assembly. The shorter band was confirmed as W·X·Y, while the larger band was confirmed as a mixture of W·Y·X·Z and W·X·Y·Z, with perhaps more of the former, with the caveat that the PCR was not quantitative data not shown). Here also, the shuffling of fragments to form W·X·Y·Z requires multiple steps and again demonstrates that the recombinase ribozyme system has the capability to form various complexes (both noncovalent and covalent) that can recycle fragments to search globally for fitness. Final support for global recycling comes from a recent variation of this system (Vaidya et al., 2012), where the recombination guide sequence and tags of W·X·Y·Z were shown to be modifiable to create cyclic cooperative networks. In the current study, recycling was observed even in the complex cooperative assembly of three versions of the W·X·Y·Z network (Figure S3).

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 W·X·Y·Z recombination system to increase local fitness was investigated by reacting W·X·Y·Z with mutations in Z fragment (W·X·Y·Z1) with wild-type Z fragment (Figure 5A). The mutations in Z fragment lie in the catalytic core of the molecule and are lethal for its recombinase activity (Figure S4). When W·X·Y·Z1 was incubated with wild-type 32P-body-labeled Z, the incorporation of 32P-body-labeled Z in W·X·Y·Z was observed demonstrating recycling of W·X·Y from W·X·Y·Z1 to form fully functional W·X·Y·Z (Figure 5B). When tested for catalytic activity using trans-splicing assay, recycled W·X·Y·Z fully restored its activity (Figure S4), indicating the ability of a recombinase system to explore various sequences in a quasispecies cloud to increase local fitness.

The simulations described previously demonstrate the fecundity of global searches for fitness over all 256 possible tetramers composed of W, X, Y, and Z. To investigate a local search for

(C) The same experiment as in (A) was performed, except W·Y·X·Z (2 μM) was incubated with 32P-body-labeled X (2 μM), 32P-body-labeled Y (2 μM), or the combination of both. Here, the product band was very strong and roughly the same size as that from the W·X·Y·Z reaction (A); the marker lane (M) contained the same product used in (A).

(D) A more extreme shuffling of fragments during self-assembly of ribozyme. Here the W·Y molecule, S’- radiolabeled with 32P, was incubated with X·Z (both at 2 μM), and the formation of W·X·Y (lower arrow) and W·X·Y·Z (along with W·Y·X·Z, upper arrow) was observed. RNA from these positions in the gel was excised and assayed for genotype via RT-PCR (see text). Time points taken were at 0 min, 15 min, 30 min, 1 hr, 2 hr, 4 hr, and 6 hr.
fitness, we also introduced mutant variants of the Z monomer species to the simulation model for recycling via spontaneous hydrolysis, implementing simulations with five variants of Z: Z0 (the wild-type), Z1, Z2, Z3, and Z4. In the experimental system, Z mutants are lethal. In the computational study, we therefore modeled nonlethal Z mutants to explore how selection leads to a distribution of W·X·Y·Zn variants. The stability of a W·X·Y·Zn variant is thus dependent on which variant of Z is present in the fully assembled tetramer, with relative susceptibilities to hydrolysis computed in a multiplicative progression of 1.0, 1.25, 1.67, 2.5, and 5.0 for Z0 (the wild-type), Z1, Z2, Z3, and Z4, respectively, so that a wild-type sequence W·X·Y·Z has a susceptibility to hydrolysis, which is 80% of W·X·Y·Z1, 60% of W·X·Y·Z2, 40% of W·X·Y·Z3, and 20% of W·X·Y·Z4. The results (Figure 5C) bear out the expectations from the empirical study with the lethal Z1 mutant. Recycling was demonstrated to lead to growth of the wild-type W·X·Y·Z population, despite not being present in the initial population. W, X, and Y fragments initially sequestered in the mutant variants W·X·Y·Zn were observed to be recycled and reincorporated in wild-type W·X·Y·Z tetramers. For all hydrolysis rates explored, a dominant fraction of W, X, and Y were reincorporated into wild-type W·X·Y·Z due to its enhanced stability over all mutant variants (Figure 5S). However, unlike in the experimental system, here W·X·Y·Zn mutants are not lethal and recycling, therefore, does not lead to complete conversion. The final equilibrium distribution of W·X·Y·Zn variant abundances instead reflects the relative stability of these sequences, with W·X·Y·Z achieving the largest equilibrium population size and the most unstable mutant W·X·Y·Z4 achieving the smallest equilibrium abundance.

**Mechanism of Recycling in the Recombinase Ribozyme**

Recycling in the recombinase system is possible because of dynamic covalent trans-esterification chemistry. The first step must be the formation of an intact recombinase ribozyme (E), by a combination of cis and trans interactions. Following that, recycling can occur by one of two mechanisms: R2F2, a two-step sequence of trans-esterification reactions that does not involve nucleotide insertion at the splice site, or tF2, a one-step reaction that typically proceeds with the insertion of 3–5 nt in the product (Draper et al., 2008). In either case, recombination depends on the existence of a 3-nucleotide tag sequence (CAU) that is on the 5′ ends of fragments beginning with X, Y, or Z (Draper et al., 2008). Initially, these fragments contain the tag in an appended head (h) sequence GGCAU. The CAU tag also exists at the 3′ end of fragments ending in W, X, or Y. In the tF2 case, an initial reaction at a given splice site results in an insertion of 3–5 nt; for example, WCAU + GGCAUX → WCAU+GGCAUX + G (Draper et al., 2008). Subsequent reactions at this same site will result in either a sequential loss of the G nucleotides from the 5′ end of the recombination tag or a loss of the entire tag (Figure 6). Once the entire tag has been removed, the fragments cannot be recognized by the system and thus cannot be recycled further at that splice site. In the R2F2 case, there is no original insertion, only the exchange of the h sequence; for example, WCAU + GGCAUX → WCAU+X + GGCAU. This reaction is the one that can remove a CAU tag after a series of tF2 reactions (Figure 6, last reaction). The systems used in this study were designed to favor the tF2 mechanism (Draper et al., 2008), although the slower R2F2 mechanism can always occur. Both mechanisms are reversible and can hydrolyze RNA at any CAU-containing site (cf. Figure 6), yet this reaction is inhibited by the secondary structure that forms in the loops of a “proper” junction (i.e., W·X·X·Y, or Y·Z; Riley and Lehman, 2003), rationalizing the hydrolysis rate parameters used in the computational studies discussed earlier.

There exists various pathways for the assembly of W·X·Y·Z from scrambled precursors and fragments. For example, from the experimental results (Figure 4), we can predict a plausible pathway by which the reaction W·Y·Y·Z + 32P-labeled X → W·X·Y·Z + Y proceeds:

1. \( W·Y·Y·Z + X \rightarrow W·Y·Y·Z−X \) (\( = E \), trans complex formation; dash indicates a noncovalent bond);
2. \( W·Y·Y·Z + X \rightarrow W·Y·Y·Z−X \) (a 238-nt covalent side product via R2F2);
3. \( W·Y·Y·Z → W + Y + Y·Z \) (hydrolysis); and
4. \( W + X + Y·Z \rightarrow W·X·Y·Z \) (the 202-nt autocatalytic product with a 5-nt insertion via tF2).

Reactions 2–4 are all catalyzed by E, and the underline indicates a 32P-body-labeled RNA that is visible on the gel (Figure 4A). The rapidly formed side product is clearly seen on the gel, marked with the caret (\( ^\ddagger \)) in Figure 4, and derives from the “pick-up-the-tail” (reverse splicing) aspect of the R2F2 reaction (Riley and Lehman, 2003). Analogous reaction schemes exist for the reactions initiated with W·X·X·Z or W·Y·X·Z, and these must involve one or two more tF2 steps because the final product is 5–10 nt larger. Analogous reactions must occur for the local search recycling as demonstrated in Figure 5. Here, the shorter intermediates marked with arrows are likely pick-up-the-tail reactions between the active Z fragment and other fragments generated in the reaction, e.g., Z·X (104 nt), Z·Z (106 nt), and Z·Y (110 nt). The larger intermediates likewise are appended ribozymes such as W·X·Y·Z.

In a tF2 construct, if a removal of a singlet G, then two Gs, and finally the entire recombination tag (GGCAU) occurs over time, it would demonstrate that the W·X·Y·Z system is capable of at least three recycling events (Figure 6). Evidence for this phenomenon comes from additional analysis of sequences from a previously studied complex RNA network system (Vaidya et al., 2012). When W·X·Y·Z RNAs assembled at different time points were excised, converted to DNA by RT-PCR, and then subjected to nt 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 with time (Table 1). It is, however, possible for reactions to proceed without following the aforementioned sequential order, resulting in fewer than three recycling events (cf. Figure 6).

**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 (Wächtershäuser, 1988). In this study, we demonstrate recycling as a driving force for the search and selection of evolutionary function by the use of both
Figure 6. Possible Mechanism of Recycling in the Recombinase System

In this model, recombination and recycling events occurring at the Y-Z junction are 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 internal guide sequence (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 trans-esterification reaction, where the 3’ hydroxyl of W-X-Y attacks an ester bond between the α and β phosphates in the 5’ end of Z to form a 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 the gray 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 trans-esterification 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 the second recycling event, depicted in the lower left dotted gray 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 the Y-Z junction and the release of the second guanosine. The assembled W-X-Y-Z can enter another round of hydrolysis to regenerate W-X-Y and Z. However, 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 a 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.

computational simulations and empirical data. Our modeling data strongly suggest the importance of DCC in the global search for functional molecules. Our experimental data further 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; McGinnness and Joyce, 2002) and that they can even perform cross-catalytic replication of the ribozymes themselves (Kim and Joyce, 2004). Thus, hydrolysis and the subsequent ligation by a ligase ribozyme could have been a potent means to recycle fragments of RNA. However, our findings suggest that recycling can be achieved via recombination reactions, avoiding some of the limitations of ligation reactions. Recombination and recycling might have gone hand in hand during the origins of life.

Recycling requires both a breakdown and a synthesis step. For RNA, the former can be an uncatalyzed hydrolysis via water and/or a catalyzed fragmentation by an enzyme. At neutral pH and room temperature, the rate constant for RNA hydrolysis in the absence of metals or an enzyme is approximately $10^{-8}$ min$^{-1}$, while the trans-esterification activities of ribozymes can typically enhance this rate by up to $10^8$-fold (Emilsson et al., 2003). While our experimental studies were performed over a time span (<8 hr) in which spontaneous hydrolysis plays a relatively minor role in fragment generation, our computational
Table 1. Sequence Analysis of Recombined Junctions from a Complex RNA Network

<table>
<thead>
<tr>
<th>Sequencing and Time</th>
<th>GGCAU</th>
<th>GCAU</th>
<th>CAU</th>
<th>(-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Throughput Sequencing (Two-Piece Assembly)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N_{ave} = 3,086,731</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 min</td>
<td>0.066</td>
<td>0.685</td>
<td>0.249</td>
<td>ND</td>
</tr>
<tr>
<td>2 hr</td>
<td>0.055</td>
<td>0.555</td>
<td>0.390</td>
<td>ND</td>
</tr>
<tr>
<td>4 hr</td>
<td>0.031</td>
<td>0.403</td>
<td>0.566</td>
<td>ND</td>
</tr>
<tr>
<td>8 hr</td>
<td>0.026</td>
<td>0.303</td>
<td>0.671</td>
<td>ND</td>
</tr>
<tr>
<td>Manual Sequencing (Two-Piece Assembly)</td>
<td>N_{ave} = 21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 min</td>
<td>0.053</td>
<td>0.842</td>
<td>0.053</td>
<td>0.053</td>
</tr>
<tr>
<td>6 hr</td>
<td>0.091</td>
<td>0.591</td>
<td>0.227</td>
<td>0.091</td>
</tr>
<tr>
<td>24 hr</td>
<td>0.000</td>
<td>0.632</td>
<td>0.105</td>
<td>0.263</td>
</tr>
<tr>
<td>Manual Sequencing (Four-Piece Assembly)</td>
<td>N_{ave} = 132</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 hr</td>
<td>0.046</td>
<td>0.769</td>
<td>0.130</td>
<td>0.056</td>
</tr>
<tr>
<td>8 hr</td>
<td>0.038</td>
<td>0.535</td>
<td>0.306</td>
<td>0.121</td>
</tr>
<tr>
<td>24 hr</td>
<td>0.025</td>
<td>0.449</td>
<td>0.407</td>
<td>0.119</td>
</tr>
</tbody>
</table>

W-X·Y·Z RNAs were self-assembled from either two or four fragments, and the nt sequence of the resulting assembled RNAs were analyzed using high throughput sequencing or manual sequencing (Vaidya et al., 2012). Tag sequences used were either CAU or CNU, where the middle nt was randomized for network formation. The frequencies of assembled molecules with two-G and single-G insertions decrease over time, while those of molecules with CAUs and no tags (-) increase over time, demonstrating recycling (Figure 8). N_{ave} = average number of sequences analyzed for each time point. ND, not determined.

See Figure S3 for additional supporting data.

The experimental data with the Azoarcus ribozyme extend the realism of recycling during global searches for function. Here, covalently contiguous W·X·Y·Z molecules can self-assemble and cross-assemble by shuffling the blocks of RNA fragments from various constructs. 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 (Schultes and Bartel, 2000; Vaidya and Lehman, 2009; Lau and Unrau, 2009), heightens the power of recycling as an evolutionary force. In addition, recycling can occur locally to optimize functions once motifs are established. A recycling of W·X·Y from W·X·Y·Z' to form fully functional W·X·Y·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).

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 shown here (King, 1986). 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 Earth (or elsewhere) more probable and that it was likely a key feature in the regulation of resource availability in the first living systems.

**SIGNIFICANCE**

In this work, we show the potential key role that fragment recycling plays in the transition from chemistry to biology. Most previous scenarios for the origins of life invoke either a high concentration of available reagents for chemical systems or a flowthrough mechanism that could continually supply the reaction locale(s). While these latter situations may have indeed been realized under certain environmental conditions, the opportunity to recycle materials opens up abiogenesis to a wider range of conditions, such as harsher scenarios in which resource limitation was a key obstacle. In truth, many authors have raised concerns about the inefficiencies of prebiotic chemical reactions (Wächtershäuser, 1988; Shapiro, 2006). If a system has the intrinsic ability to repurpose its component parts on a continual basis, then these concerns are somewhat ameliorated, at least to the point that certain prebiotic reactors could be added to the list of possibilities.

Our studies take two complementary approaches, one computational and one empirical. In that regard, the results presented here are relatively uncommon. The computational approach allowed a thorough search of parameter space and concluded that catalytic recycling could lead to the selection of high-fitness genotypes, especially when rate constants were in an intermediate range. This result mirrors a common feature of biology: that reaction rates need to be attenuated to be neither too rapid nor too slow. The empirical approach, while unable to explore as many...
combinations as could be done on the computer, allowed an imposition of chemical reality to the recycling system. These studies show that recycling with a known recombinate of biological origin not only is facile but can lead to evolutionary self-selection.

EXPERIMENTAL PROCEDURES

Computational Modeling

Simulations were implemented using a kinetic Monte Carlo algorithm that included two kinetic processes: the spontaneous or catalyzed hydrolysis of tetramers to monomers and the replicative assembly of monomers into tetramers. Each of the 256 possible tetrameric arrangements of W, X, Y, and Z was assigned a relative stability based on its bond structure similarity to W-X-Y-Z. The reaction probabilities for the spontaneous hydrolysis of tetramers were then calculated as $P_i = k_{ch} S_i X_i$, where $k_{ch}$ is the rate constant for spontaneous hydrolysis, $X_i$ is the total number of tetramers of species $i$ (where runs over all 256 possible tetramer species), and $S_i$ is the susceptibility to hydrolysis for species $i$ calculated by its distance from W-X-Y-Z as outlined earlier. For simulations including catalyzed hydrolysis due to functional activity of W-X-Y-Z, the reaction probabilities for catalyzed hydrolysis of tetramers were calculated as $P_i = k_{ch} S_i X_i W_{XYX} X_i$, where $S_i$ and $X_i$ are as defined earlier, $k_{ch}$ is the rate constant for catalyzed hydrolysis, and $W_{XYX}$ is the total number of assembled W-X-Y-Z molecules in the system. The replicative rate constant was set to be sequence independent. However, the probability for replication of a given tetramer was set to be dependent on the availability of resources necessary to assemble that tetramer. To capture sequence dependence, the reaction probability for replication was calculated as $P_i = k_i (M_1 + M_2 + M_3 + M_4)$, where $M_i$ is the total abundance of the monomer species located at Position $i$ in a tetramer sequence (i.e., W for $i = 1$, X for $i = 2$, etc., for W-X-Y-Z). Resource dependence was thus modeled as a sum over the formation rates of bond pairs in the fully assembled tetramer and is dependent on the total availability of resources necessary to assemble a specific sequence.

To investigate the role of recycling via spontaneous hydrolysis in driving local searches for fitness, simulations were carried out for a fixed system size of 40,000 informational units, comprising 5,000 each of the four (wild-type) fragment species W, X, Y, and Z and 5,000 each of the four mutant Z species $Z_1$, $Z_2$, $Z_3$, and $Z_4$. Populations were initialized with a large population of 5,000 tetramers, composed of 1,250 of each of the mutant variants W-X-Y-Z, W-X-Y-Z, W-X-Y-Z, and W-X-Y-Z, and no tetramers containing the wild-type Z. The initial population of tetramers therefore contained all W, X, and Y fragments, with none in monomer. As such, formation of wild-type W-X-Y-Z required the recycling of W, X, and Y from mutant variants of the sequence. The system was permitted to evolve until a steady-state population was achieved. Data were averaged over 100 simulations runs for each value of the spontaneous hydrolysis rate constant, $k_{ch}$.

Computational Data Analysis

A data set size of 100 simulation runs was collected for each hydrolysis rate constant explored (spontaneous or catalytic), where each run was initialized with a different random seed. For simulations with spontaneous hydrolysis, the distribution of runs leading to selection of $\mu$-error mutants was determined from the steady-state distribution of tetramers, where the steady state was defined to be achieved when one species achieved total (100%) fixation. Not all runs achieved steady state or the total fixation of one species within the timescale of the simulation; however, a minimal set of 90 out of the 100 total simulation runs was used for each data set presented. All sequence abundances remained within 5% of a fixed value (i.e., when the abundance of each species of tetramer achieved a final fixed value within small stochastic fluctuations). Typically, one to three sequences survived the initial selection period and became fixed at steady state. Runs were binned by the number of errors of the sequence that achieved fixation (i.e., 0 for W-X-Y-Z). Runs were binned by the average distance of all extant sequences in the final population from W-X-Y-Z into whole number intervals (i.e., a run with survival of 2 one-error and 1 two-error mutants would yield an average bond-structure distance of 1.33 and thus be binned in the population of runs leading to selection of one-error mutants). Fractional values capturing the fraction of total runs leading to selection of $\mu$-error mutants were then calculated by dividing by the total number runs in each bin by the total ensemble size of 100 simulation runs. For runs with catalyzed recycling, the final steady-state abundance was determined by time-averaging the abundance of W-X-Y-Z over 100 time steps after 100% fixation was achieved (i.e., identified when W-X-Y-Z was the only remaining extant sequence). Time-averaged abundances were then ensemble-averaged over the 100 experimental runs for each value of $k_{ch}$ explored, and the standard deviation calculated based on the distribution of abundances in the 100-run ensemble. The fixation time was defined as the time when the W-X-Y-Z abundance achieved 95% of its final steady-state abundance. Fixation times were then likewise ensemble-averaged over all 100 runs for each value of $k_{ch}$ explored, and SDs were calculated based on the distribution of fixation times in the 100-run ensemble.

Preparation of RNA

RNAs were prepared by runoff 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 used 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 (Hayden and Lehman, 2006; Draper et al., 2008; 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 microcentrifuge tubes and were quenched by the addition of an equal volume of gel-loading solution containing 8 M 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+ 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 Y and Z, which target the 3' portions of Y and Z, respectively (Hayden and Lehman, 2006). These reactions were used as a template to seed PCR to determine the arrangements of fragments in assembled products. Pairs of primers specific for W-X, W-X-Y, and W-X-Y-Z were used to reveal the existence of W-X-Y-Z, while pairs of primers specific for W-Y and W-Y-X were used to reveal the existence of W-Y-X-Z; for example, use of W- and Y-specific primers in the PCR will generate a product ca.144 base pairs (bp) if W-X-Y-Z is present and will generate a product ca. 108 bp if W-Y-X-Z is present.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and can be found with this article online at http://dx.doi.org/10.1016/j.chembiol.2013.01.007.

ACKNOWLEDGMENTS

We thank E.J. Hayden and J.A.M. Yeates for useful discussions during the preparation of the manuscript. This work was funded by the National Aeronautics and Space Administration (NASA; NNX10AR15G to N.L. and a NASA Post-doctoral Program Fellow through the NASA Astrobiology Institute to S.I.W.).


