

Self-Assembly of a Group I Intron from Inactive Oligonucleotide Fragments

Eric J. Hayden¹ and Niles Lehman^{1,*}

¹Department of Chemistry
Portland State University
P.O. Box 751
Portland, Oregon 97207

Summary

The *Azoarcus* group I ribozyme was broken into four fragments, 39–63 nucleotides long, that can self-assemble into covalently contiguous ribozymes via RNA-directed recombination events. The fragments have no activity individually yet can cooperate through base pairing and tertiary interactions to produce stable *trans* complexes at 48°C. These complexes can then catalyze a sequence of energy-neutral recombination reactions utilizing other oligomers as substrates, assembling covalent versions of the ribozyme. Up to 17% of the original fragments are converted into ~200 nucleotide products in 8 hr. Assembly occurs primarily by only one of many possible pathways, and the reaction is driven in the correct and forward direction by the burial of key base-pairing regions in stems after recombination. Autocatalysis, and hence self-replication, is inferred by a reaction rate increase upon doping the reaction with full-length RNA.

Introduction

The origins of life on the Earth required the emergence of molecules that had the capacity to maintain genetic information and to evolve new phenotypes. Among the many obstacles to reconstructing this process some four billion years later are demonstrations of molecular self-assembly, self-replication, and chemical autocatalysis, all of which are features of popular models of biogenesis [1–9]. The RNA World scenario exploits RNA's duality of genotype and phenotype in positing that RNA-like molecules could have been Earth's first living entity [1, 10]. However, this model still lacks a plausible means by which chains of ribonucleotides long enough to support catalytic activity relevant to evolutionary progression could have assembled from pools of random oligomers. Polymerization of activated nucleotides on inorganic substrates such as clay is promising; oligomers as long as 55 nucleotides have been observed in the lab [11]. Yet such oligomers would not likely possess enough catalytic prowess to replicate themselves. This is true in part because complex functionality, such as phosphoanhydride energy (e.g., ATP) utilization, appears only to be found in longer ribozymes. Even if an RNA replicase could arise via selection from random sequences, this molecule likely would make Watson-Crick complementary copies of itself, posing additional constraints on self-replication. Moreover, primitive RNAs are faced with the Catch-22 problem of Eigen's error

threshold, in which higher fidelity is needed to maintain information in longer sequences, and yet longer sequences are needed to improve fidelity [12, 13].

Previous efforts to isolate a self-replicating RNA species have met with some success by focusing on template-directed ligation reactions. For example, von Kiedrowski has shown that oligomers such as hexanucleotides can direct the ligation of short complementary oligomers when the latter are charged with potent condensing agents such as carbodiimides [2, 14]. The products in these reactions serve as additional templates, such that the systems can display autocatalytic behavior. Joyce has devised elegant schemes involving ligase ribozymes, in which reactions of the general scheme $A + B \rightarrow T$ can be demonstrated [15, 16]. In this case the template (T) ribozyme can be 50–70 nucleotides in length, autocatalysis is possible, and cross catalysis (i.e., $A + B \rightarrow T'$ coupled with $A' + B' \rightarrow T$) can result in the ultimate amplification of a specific sequence. Variants of both types of systems can self-assemble; that is, templates can be produced from component fragments at least to some degree. However, these schemes are still somewhat constrained. They can suffer from a high degree of product inhibition as a result of the need for the diffusion of the ligated product off of the template. The longer and more exact the complementary region, the more severe the problem. They also have a tendency either to show self-assembly or to exhibit autocatalytic behavior, but rarely both; if there is a high rate of ligation in the absence of added template, it is difficult to show autocatalysis because the nontemplated ligation rate is dominant. Lastly, they produce at best small catalytic entities from at most two fragments.

RNA recombination, as opposed to template-directed ligation, has the potential to alleviate the problems of complementarity, product inhibition, autocatalysis, and limited product scope in a fashion that does not require high-energy leaving groups [17, 18]. Here, we investigate the possibility that recombination played a direct role in the origins of life by providing a mechanism by which complex phenotypes could have arisen from a collection of much smaller RNA fragments. The underpinnings of our approach were provided in the early 1990's when it was shown that the *Tetrahymena* and *sunY* group I ribozymes could be broken into as many as three pieces and retain catalytic activity through intermolecular *trans* interactions involving secondary and tertiary structure formation without a covalently contiguous RNA chain [19–22]. Such *trans* complexes could perform ligation and splicing reactions on exogenous RNA substrates and could thereby promote the assembly of fragments complementary to the ribozymes themselves [20, 22]. However, these ribozyme complements possessed no catalytic activity, and consequently autocatalytic self-assembly was not possible. We approached this problem from a different perspective, focusing on generalized RNA-directed recombination events rather than on template-directed complementary strand syntheses, with the aim of demonstrating a mechanism by which covalently contiguous group I ribozymes can self-assemble

*Correspondence: niles@pdx.edu

from inactive RNA fragments and then catalyze further assembly reactions.

We chose the *Azoarcus* group I intron as a candidate for self-assembly (Figure 1). This molecule has several attractive features in this regard. It is the smallest naturally occurring group I intron, and its high 71% G+C content allows it to remain functional at temperatures of 60°C or more [23, 24]. These properties contribute to its effectiveness as an RNA recombinase, in which it catalyzes reactions of general formula $A \bullet B + C \bullet D \leftrightarrow C \bullet B + A \bullet D$, where A and C are 5' "head" portions of RNA substrate oligonucleotides, while B and D are 3' "tail" portions [25, 26]. The only strict requirement for recombination is the presence of a 5'-CAU-3' trinucleotide in the substrate, because this sequence is bound by the internal guide sequence of the *Azoarcus* ribozyme, which then cleaves 3' of the U (i.e., at the \bullet) in the oligomers and effects a swapping of heads through a two-step mechanism [26]. The small size and exquisite stability of *Azoarcus* RNA have also permitted recent crystal structures of the complete intron to be obtained [27, 28] that are useful in engineering mutations with desired effects. Here, we show that the *Azoarcus* ribozyme can be broken into as many as four nonfunctional fragments that can self-assemble by promoting synthesis of full-length ribozymes via these RNA-directed recombination events.

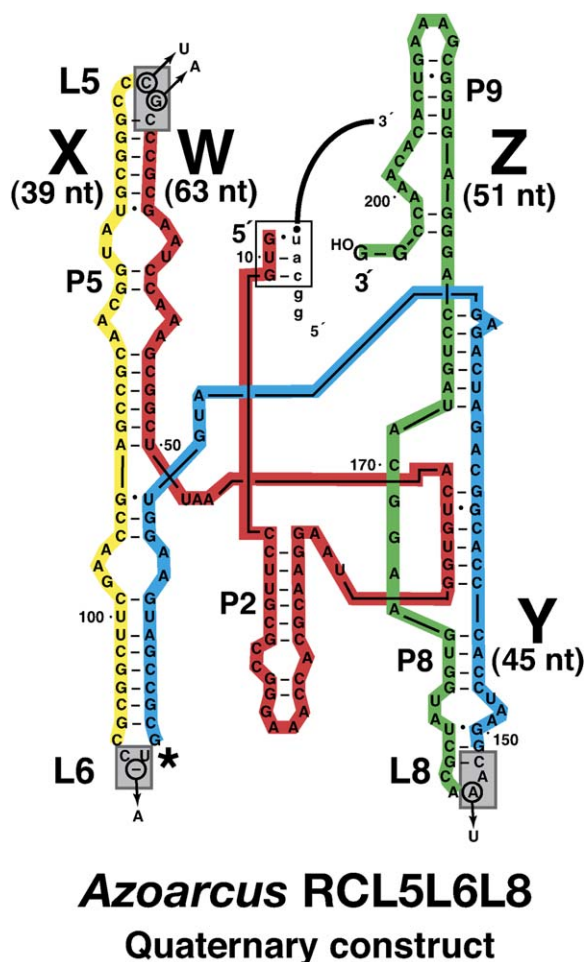
Results and Discussion

Logic of Self-Assembly through Recombination

Guided by structural data available for *Azoarcus* RNA, and by the operative breakpoints in other group I introns, we conceptually broke the *Azoarcus* sequence into four fragments of 63, 39, 45, and 51 nucleotides in size by placing CAU triplets in the L5, L6, and L8 loop regions (Figure 1). We speculated that a contiguous *Azoarcus* RNA of 198 nucleotides could be formed through a series of recombination reactions by using CAU-containing oligomers (Figure 2A). These fragments, denoted W, X, Y, and Z, each would be derived from CAU-containing RNAs with short 5 nt head (h = GGCAU) and tail (t = AAAUA) segments to allow for recombination. For example, *Azoarcus*-RNA-guided recombination of $WX \bullet t$ and $h \bullet YZ$ should generate the full-length *Azoarcus* ribozyme $WX \bullet YZ$ plus the 10-mer $h \bullet t$. The t -tails are not actually needed for recombination, and their omission simplifies the process if one is focusing on the $WX \bullet YZ$ product and not the $h \bullet t$ product. More importantly, if the oligomers themselves could coalesce in the absence of added *Azoarcus* RNA through 2° and 3° structural interactions to form a catalytic entity, then self-assembly would be possible (Figure 2B).

Self-Assembly from Four Inactive RNA Fragments

When we incubated the four oligomers W, $h \bullet X$, $h \bullet Y$, and $h \bullet Z$ together in a 1:1:1:1 molar ratio, we in fact observed self-assembly of covalently contiguous RNA approximately 200 nucleotides in length, accompanied by a series of intermediates (Figure 3). Because no full-length *Azoarcus* RNA was added to the reaction, and no reaction occurs when any one of the four pieces is left out of the milieu, we propose that the tetramolecular *trans* complex is catalyzing assembly, at least initially. This



Azoarcus RCL5L6L8 Quaternary construct

Figure 1. Partition and Autoreassembly of the *Azoarcus* Ribozyme

The 198 nt source molecule was partitioned into four fragments (W, red; X, yellow; Y, blue; Z, green) by the placement of CAU recombination signals (gray boxes) in loop regions. An intrinsic CAU at positions 170–172 was not removed, despite its potential as an internal cleavage site, because this mutation causes a ~30% drop in splicing activity. Substrate oligomers (lowercase letters, head; black line, tail) bind via a CAU (boxed) to the internal guide sequence (GUG) of the ribozyme, which catalyzes phosphotransfer of the tail to its own 3' end. The asterisk denotes (G)CAU insertion site at the X-Y junction upon self-assembly.

process is nevertheless quite distinct from previous demonstrations of group I intron *trans* activities that were only able to catalyze the assembly of portions of full-length ribozymes and then only complementary sequences [20, 22].

In Figure 3A, only the W-containing molecules are labeled, and examination of the reaction over time shows that the system produces $W \bullet X$, $W \bullet X \bullet Y$, and then $W \bullet X \bullet Y \bullet Z$ molecules in turn through a series of successive recombination events (cf. Figure 2B). This result requires the following events to occur. First, the W, X, Y, and Z RNA elements must assemble *in trans* to form a labile, but catalytically active ribozyme. The presence of the short h sequences must not inhibit this activity. Next, this complex must bind another oligomer via base pairing to its CAU sequence (in h) and catalyze a phosphotransfer reaction in which the nucleotides 3'

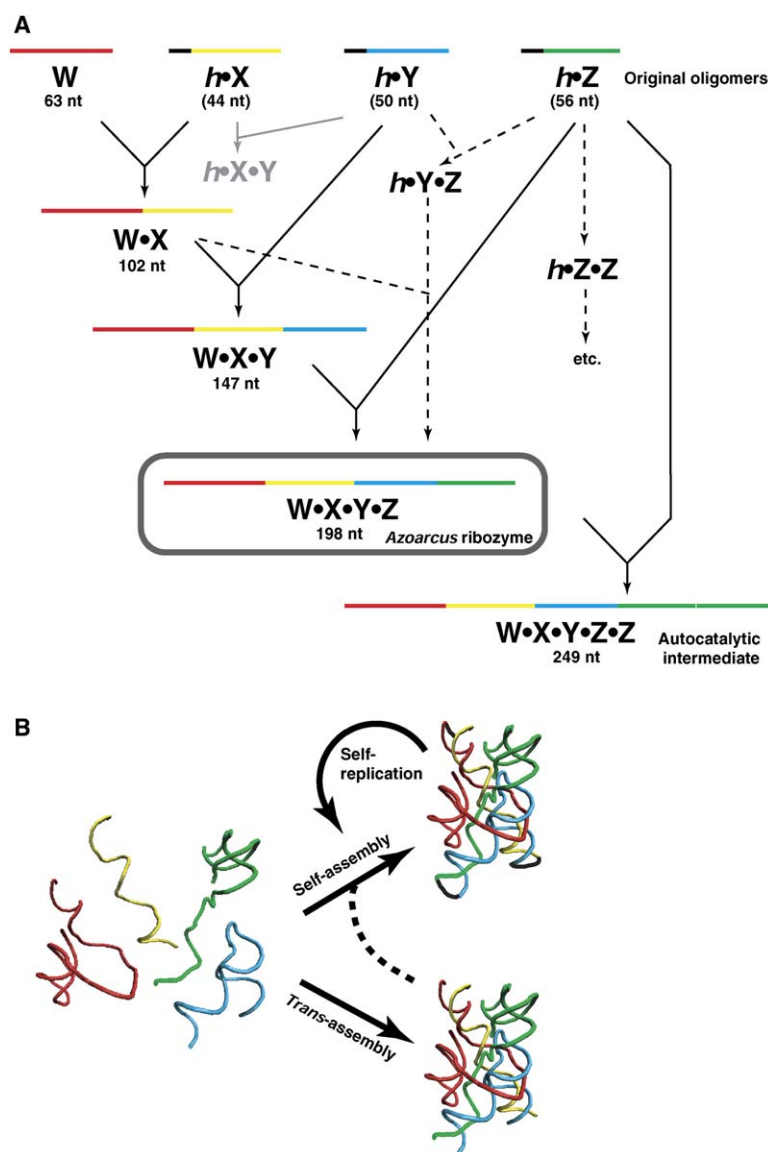


Figure 2. Proposed Self-Assembly of the *Azoarcus* Group I Intron through Recombination Events among Individually Inactive Oligomers

(A) Scheme of recombination events that lead from four inactive oligomers to a full-length covalently contiguous recombinase ribozyme. Recombination events in solid black arrows are productive and were experimentally predominant (Figure 3). Events in dashed lines are side reactions or alternative pathways that were observed, while the one in gray arrows is an example of a pathway that was theoretically possible but not observed within limits of detection (Figure 3). The production of a transposed W•Y•X•Z molecule would be another such example. RNA sizes are those predicted a priori; an insertion at the X-Y junction (see text) results in actual sizes 3–4 nt larger.

(B) Schematic of self-assembly and self-reproduction of a complex ribozyme. Individually inactive oligomers (left) associate through 2° and 3° *trans* interactions to produce noncovalent but catalytically active complexes (bottom), which utilize more oligomers as substrates in recombination reaction to produce covalently contiguous RNAs (top), which may then promote further production of like RNAs.

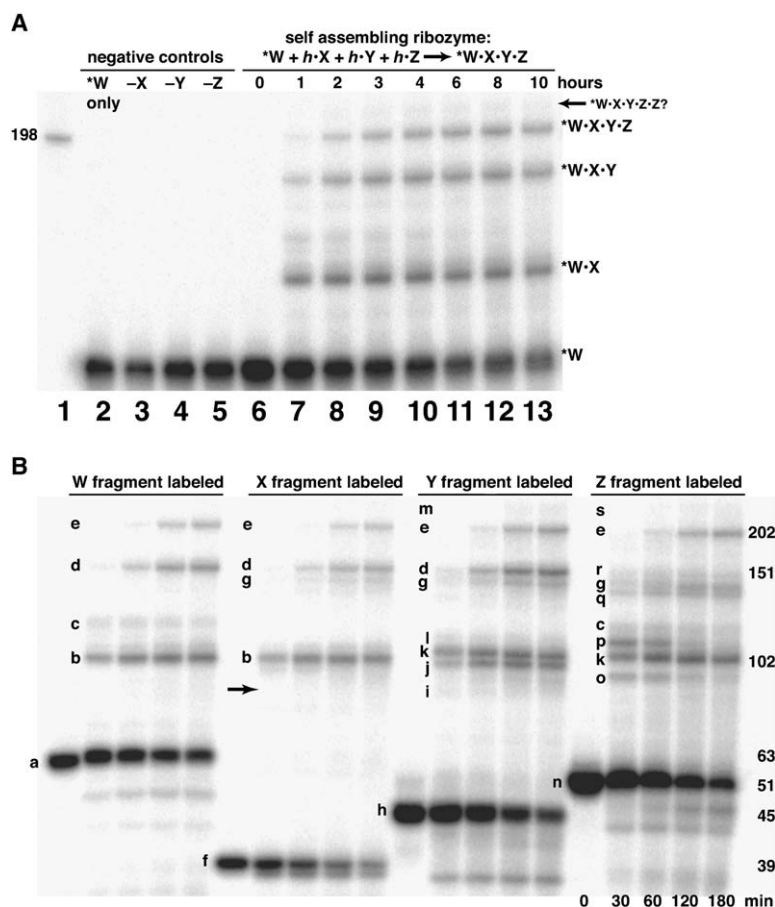
of the CAU are transferred to the 3' end of the Z element in the catalyst, while the *h* nucleotides 5' of the splice site are released. Similar reactions must occur with other oligomers. Lastly, 5' elements must diffuse to 3'-extended catalytic complexes, bind, and promote additional phosphotransfer reactions to accomplish recombination. These recombinations must proceed in one of the pathways that produce the W•X•Y•Z construct. Conceivably, newly made W•X•Y•Z products could themselves catalyze further recombination reactions in an autocatalytic manner.

We were able to detect as much as 17% of the 5'-radiolabeled W molecules converted to W•X•Y•Z after 8 hr under optimal conditions. We found that an equimolar ratio of oligomers (2 μ M each in Figure 3A) input into reactions run isothermally at 48°C produced the highest yield of full-length RNA (data not shown). In general, as oligomer concentrations are raised, the reaction is enhanced, presumably because the recombination pathway depends on sufficient numbers both of oligomers behaving as catalytic complexes and of oligomers

behaving as substrates. Temperatures above or below 48°C have an adverse effect on the reaction; higher temperatures likely disfavor formation of *trans* complexes, while lower temperatures likely disfavor their catalysis of phosphotransfer reactions. The four-piece self-assembly reactions also work when 5 nt *t* tails were included downstream of the CAU splice site at the 3' ends of the W, X, and Y fragments, but these reactions are slower because of the need to have the *t* fragments cleaved off prior to recombination.

Self-Assembly from Three or Two Inactive RNA Fragments

Reactions in which the *Azoarcus* molecule was broken into arrangements of three pieces ($WX + hY + hZ \rightarrow WX\bullet Y\bullet Z$) and of two pieces ($WX + hYZ \rightarrow WX\bullet YZ$) gave higher proportions of full-length product: 20% and 66%, respectively, after 8 hr (Figure 4). The ternary and binary self-assembly reactions are more productive because the initial *trans* formation of a catalytic assembly is less challenging and because fewer recombination



c, W•Z; d, W•X•Y; e, W•X•Y•Z (full-length product); f, h•X; g, h•X•Y•Z; h, h•Y; i, h•X•Y (barely seen and not seen well in the X-labeled lanes: arrow); j, h•Y•Y (side product); k, h•Y•Z + h•Z•Y; l, h•W•Y (side product); m, W•X•Y•Z•Y (putative autocatalytic intermediate); n, h•Z; o, h•X•Z + h•Z•X; p, h•Z•Z; q, h•Z•Z•X?; r, h•Z•Y•Z; s, W•X•Y•Z•Z (putative autocatalytic intermediate).

reactions are needed to obtain full-length products. Nevertheless the quaternary reaction is surprisingly robust (17% yield) and portends that the *Azoarcus* ribozyme could be fragmented even further.

Genotyping of Full-Length Products

To ensure that the reactions were proceeding as outlined in Figure 2A, we confirmed the identity of the full-

length product by RT-PCR. Following incubation of W, h•X, h•Y, and h•Z and denaturing gel electrophoresis, the ~200 nt band was excised from the gel, reverse transcribed, and PCR amplified with primers that yield amplicons of any constructs containing both W and Z. We detected a strong DNA product of the correct size, but this in itself was not assurance of the existence of W•X•Y•Z because W•Y•X•Z is also a potential result

Figure 3. Molecular Self-Assembly of the *Azoarcus* Ribozyme

(A) The self-assembly of the full-length molecule when the ribozyme is broken into four pieces (W + h•X + h•Y + h•Z) and reassembled via recombination. Lane 1 is a 198 nt size control; in all other lanes only the W-containing fragments are 5' radiolabeled (asterisk). Negative controls include reactions where one fragment is left out of the reaction to demonstrate lack of activity of individual oligomers (lanes 2–5). Lanes 6–13 are a time course of incubation of 2 μM of each RNA oligomer at 48°C, showing the progressive self-assembly of W•X, then W•X•Y, and then W•X•Y•Z RNAs. Faint bands above the W•X•Y•Z RNA are likely the result of splicing activity of newly assembled W•X•Y•Z molecules on other RNAs such as h•Z.

(B) The W + h•X + h•Y + h•Z self-assembly reaction viewed from the vantage of all four fragments being labeled, one at a time. Fragments here are [α -³²P]ATP labeled, so band intensities do not scale with molar amounts because of differing numbers of adenosine residues in each fragment (17 in W, seven in X, 13 in Y, and 16 in Z). Sizes on right-hand side correspond to expected major W-labeled bands, plus h•X, h•Y, and h•Z, except that the h•Y and h•Z fragments used in this experiment were each 1 nt larger than the sizes indicated, to resolve several predicted comigrations. Identities of bands were confirmed in a separate experiment where pair-wise combinations of oligomers were recombined by using wild-type *Azoarcus* RNA [26] into distinct products and compared against reactions performed here (not shown). Assignments are: a, W; b, W•X; c, W•Z; d, W•X•Y; e, W•X•Y•Z (full-length product); f, h•X; g, h•X•Y•Z; h, h•Y; i, h•X•Y (barely seen and not seen well in the X-labeled lanes: arrow); j, h•Y•Y (side product); k, h•Y•Z + h•Z•Y; l, h•W•Y (side product); m, W•X•Y•Z•Y (putative autocatalytic intermediate); n, h•Z; o, h•X•Z + h•Z•X; p, h•Z•Z; q, h•Z•Z•X?; r, h•Z•Y•Z; s, W•X•Y•Z•Z (putative autocatalytic intermediate).

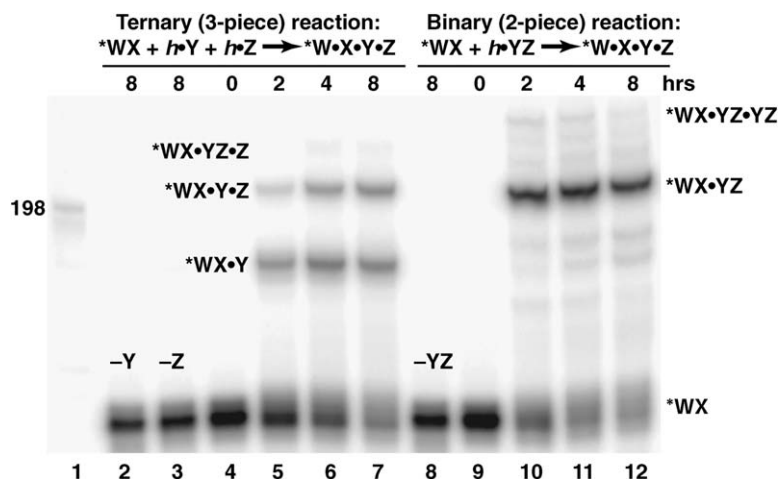


Figure 4. Ternary and Binary Self-Assembly and Self-Replication

The ternary reactions were performed by incubating 2 μM each of WX, h•Y, and h•Z in 100 mM MgCl₂ at 48°C. The binary reactions were performed by incubating 2 μM each of WX and h•YZ under the same conditions. Negative controls (lanes 2, 3, and 8) were the same reactions in which one of the component oligomers was left out of the reaction as indicated. Predicted RNA identities of the major bands are denoted on the gel.

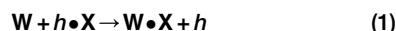
of recombination. We thus genotyped the DNA products by both RFLP and sequence analysis and found that all of them within our detection limits were indeed **W•X•Y•Z**. Predominance of the desired **W•X•Y•Z** construct is very provocative, given that both it and **W•Y•X•Z** can result from as few as three recombination events (e.g., Figure 2A). But the **W•X•Y•Z** construct is the only entity that should possess recombinase activity, and thus its accumulation might result from autocatalysis.

Nucleotide sequencing revealed that all full-length RNAs did contain a short insertion of three or four nucleotides at the X-Y junction. The insertion is always CAU or GCAU positioned 3' of the intrinsic CAU forming the junction. This suggests that a different recombination mechanism is occurring at this junction, one shown previously with the *Anabaena* group I intron [29] in which transesterification at its 3' splice site is facilitated by two adjacent short base-pairing regions (Figure 5). This reaction is only seen at the X-Y junction, and the X-Y hybrid complex is held into place by the IGS. We placed the CAU more centrally in the X-Y junction than in the other two because other placements proved to be less active, with the consequence that, uniquely in the X-Y hybridization: (1) the 5' G of the head of Y is correctly aligned to bind in the G binding site, (2) the cytidine of the CAU of X is not paired with Y and thus free to bind to the IGS (Figure 5, box), and (3) the 3'-OH of X falls in position to catalyze a transesterification leading to insertion of CAU or GCAU, with roughly equal frequency.

Pathway to the Covalently Contiguous Product Ribozyme

The 17% yield of the **W•X•Y•Z** product ribozyme after 8 hr is surprisingly high, given that all of the recombination reactions are energy neutral and that many competing side reactions (such as the route to **W•Y•X•Z**) are conceivable. There may then be a driving force that favors the correct assembly pathway. To begin to address this phenomenon, we examined the four-piece assembly reaction by tracing not only 5'-labeled W fragments as in Figure 3A but also by tracking ³²P[A]-body-labeled W, X, Y, and Z fragments over time (Figure 3B). Viewing the reaction from the vantage of all four fragments allows a comparison of the existences of intermediate products and their orders of appearance and hence a deduction of the most populated pathway to the product.

With three "correct" recombination junctions in the full-length ribozyme (Figure 1), there are four pathways that could lead to **W•Y•X•Z** RNA. The self-assembly could occur 5' to 3' in this sequence:



Conversely, the assembly could occur 3' to 5' in this sequence:

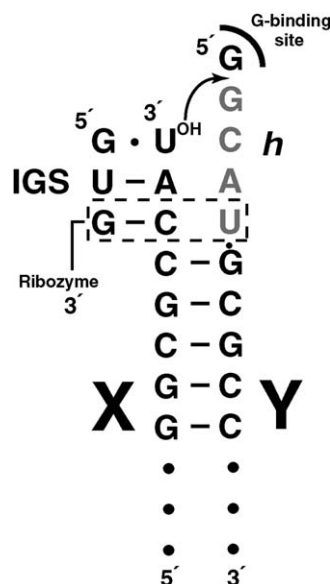
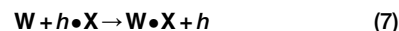


Figure 5. Proposed Mechanism for Recombination at the X-Y Junction

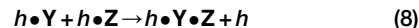
The short insertion of three (CAU) or four (GCAU) nucleotides (gray) at the X-Y junction revealed by nucleotide sequence analysis of the self-assembled ~200-mer RNA can be explained by a reaction analogous to that at the 3' splice site during in vivo splicing of *Anabaena* pre-tRNA [29]. Unlike the other two recombination junctions, here, the cytidine of the CAU of X is not paired with Y and thus free to bind to the IGS (dashed box).



Or the two outer junctions could be made first, followed by the joining of the two resultant halves at the X-Y junction:



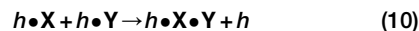
and



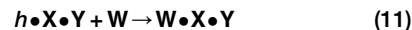
then



Finally, the center X-Y junction could form first, followed by the addition of the W and Z fragments, in either order:



then



or



The data in Figures 3A and 3B are most consistent with the first alternative, with reactions 1–3 being prevalent. From the two-step chemical mechanism, a Z molecule with a covalent 3' extension should appear first for recombinations at the W-X and Y-Z junctions because

the 3' portion (Z) of the *trans* complex becomes initially covalently attached to another substrate oligomer but will dissociate from the rest of the complex on a denaturing gel. (Z-tailed intermediates would not be expected for the X-Y junction [Figure 5].) Thus, for a 5' → 3' assembly, one should expect bands corresponding to $h\bullet Z\bullet X$ and $W\bullet X$ to appear early, bands for $h\bullet Z\bullet Y$ and $W\bullet X\bullet Y$ to appear later, and bands for $h\bullet Z\bullet Z$ and $W\bullet X\bullet Y\bullet Z$ to appear later still. This pattern is observed in both the W- and X-labeled portions of Figure 3B as well as in Figure 3A. Additional support for reaction 1 being fast is the observation that the ternary reaction, in which the W-X junction is already formed (Figure 4), is only slightly more productive than the quaternary reaction: 20% versus 17% after 8 hr.

On the other hand, for a 3' → 5' assembly, one should expect a strong band for $h\bullet X\bullet Y\bullet Z$ to appear prior to the appearance of $W\bullet X\bullet Y\bullet Z$, while $W\bullet X\bullet Y$ should be skipped over. Although an $h\bullet X\bullet Y\bullet Z$ band can be seen in the X- and Y-labeled experiments, it is weak compared to $W\bullet X\bullet Y$ and does not precede the final product. Moreover, bands for $h\bullet Z\bullet Y\bullet Z$ and $h\bullet Z\bullet X\bullet Y\bullet Z$ should appear as the reaction progresses, but they do not or are very weak. Similarly, bands for $h\bullet Z\bullet Y\bullet Z$ and $h\bullet X\bullet Y$ should appear if $W\bullet X\bullet Y\bullet Z$ were being made from the ends to the middle (reactions 7–9) or from the middle to the ends (reactions 10–14), respectively. The paucity of these bands suggests that these pathways are not well populated. We must note that with these data, the relative band intensities can be deceiving, in large part because an unequal amount of adenosine residues in the oligomers means that bands will not be evenly labeled (Figure 3B).

The assembly $W \rightarrow W\bullet X \rightarrow W\bullet X\bullet Y \rightarrow W\bullet X\bullet Y\bullet Z$ is clearly the dominant pathway, although some nonproductive side reactions are seen. For example, a small amount of $W\bullet Z$ is seen in the W- and Z-labeled experiments of Figure 3B, and $h\bullet Y\bullet Y$ is seen in the Y-labeled experiment. Nevertheless, it does appear as though the correct $W\bullet X\bullet Y\bullet Z$ product is being preferentially constructed. We suggest that CAU accessibility helps explain this observation: these recognition sequences are only available for catalyst binding when they are in the oligomer substrates but not when they are bridging adjacent fragments in the intermediates or the product. The $W\bullet X\bullet Y\bullet Z$ molecule buries all of its CAU triplets in stems and short loops, a phenomenon that can reduce subsequent cleavages by other active ribozymes [25]. In effect, the correct recombination pathway benefits from a series of “snap” reactions in which base pairing of complementary sequences removes the CAU recombination signals from accessibility and thus drives the reactions in the direction of self-assembly. Notably, with the exception of $h\bullet Y\bullet Y$ and Z-tailed intermediates, all the major bands in Figure 3 can be attributed to recombination events between adjacent RNA regions in the intact ribozyme ($W \leftrightarrow X \leftrightarrow Y \leftrightarrow Z$), indicating that base pairing indeed favors pushing the overall equilibrium to the final product.

trans Assembly of a Catalytic Complex

For covalent self-assembly to occur via RNA recombination reactions as proposed in Figure 2, the oligomers must interact to form a catalytic complex as seen with

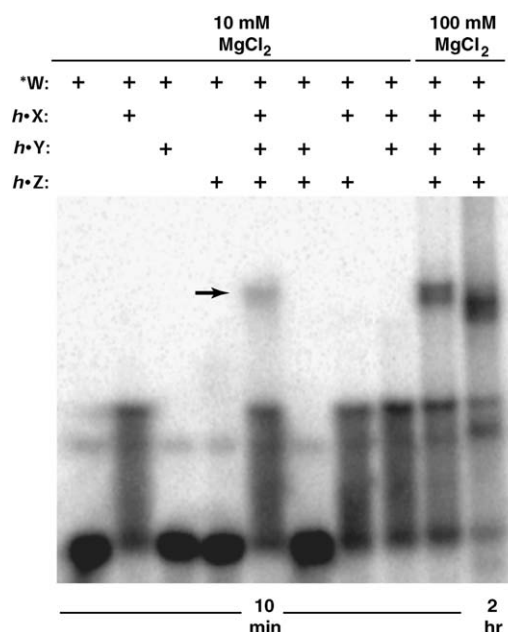


Figure 6. Native Gel Electrophoresis of Oligomer Mixtures

Equimolar ratios of oligomers as indicated were incubated in either 10 or 100 mM $MgCl_2$ at 48°C for 10 min (except last lane) prior to loading on a 10% native polyacrylamide gel containing 10 mM $MgCl_2$. The putative catalytically active *trans* complex (arrow) only forms when all four oligomers are present. Note that no detectable covalent self-assembly can be seen on denaturing gels after only 10 min, even in 100 mM $MgCl_2$. In the last lane, the self-assembly reaction was allowed to occur for 2 hr prior to electrophoresis, resulting in a slight increase in mobility of the product.

other group I introns [19–21]. We thus corroborated the existence of the tetramolecular *trans* complex by native gel electrophoresis (Figure 6). By incubating the four oligomers W, $h\bullet X$, $h\bullet Y$, and $h\bullet Z$ in either 10 or 100 mM $MgCl_2$ for 10 min and then immediately loading the mixture on a 10% nondenaturing polyacrylamide gel, we hoped to capture the *trans* complex prior to the formation of any significant amount of covalent $W\bullet X\bullet Y\bullet Z$ product, which begins to occur at about 30 min in 100 mM $MgCl_2$ (Figure 3B). We ran the gel at 48°C in a running buffer containing 10 mM $MgCl_2$ to maintain the folded conformation of the complex. When the W fragment is labeled, retardation of gel mobility can be seen as this RNA complexes with other RNAs (Figure 6). Of interest is the fact that W makes a strong complex with X, but not with Y or Z, and that a slow-moving band that we postulate is the four-piece *trans* complex only appears when all four fragments are present (Figure 6, lanes 5, 9, and 10). It has been shown previously that the *Azoarcus* ribozyme folds rapidly and cooperatively through strong tertiary interactions, between P2–P8 and P5–P9 for example [30], and our data are in strong agreement. The percentage of total radioactivity contained in the complete *trans* complex rises from 4.6% following incubation in 10 mM $MgCl_2$ to 16% following incubation in 100 mM $MgCl_2$. Moreover, if the oligomers are allowed to react for 2 hr prior to electrophoresis, a slight gain in mobility is seen, as would be expected if the covalent $W\bullet X\bullet Y\bullet Z$ product is being formed and its gel mobility is less impeded without the GGCAU heads on X, Y, and Z

(Figure 6, lane 10). The native gel not only confirms the existence of the *trans* complex but also suggests that base-pairing interactions between neighboring fragments such as W and X are an important process in complex formation.

Autocatalytic Potential of the System

A critical aspect of this system is that it has the potential to demonstrate strong autocatalytic behavior. When no product RNA is originally present in the reaction (e.g., Figures 3A and 3B), the assembly of a covalently contiguous *Azoarcus* ribozyme from inactive fragments permits the products themselves to enhance the rate of their own synthesis. Clearly if this reaction pathway is catalyzed by a noncovalent *trans* complex, then it stands to reason that covalently contiguous molecules can perform the same operation, provided they fold correctly upon recombination. Importantly, because autocatalytically made full-length products would effectively be exact copies of the catalyst, as opposed to complementary copies, autocatalysis in this system could circumvent one of the major drawbacks of replication via a hypothetical RNA replicase ribozyme.

The system described here though, is a three-step cascade in which the product, **W•X•Y•Z**, potentially feeds back to catalyze any and all of the reactions that precede its formation. Moreover, any covalent products in the pathway, such as **W•X** or **W•X•Y**, can participate in catalytic *trans* assemblages comprised of fewer than four molecules. Consequently our system would be better described as collectively autocatalytic, provided it could be demonstrated that any or all products enhance the rates of their own production. Here, we focus simply on one aspect of feedback, and that is whether production of the full-length **W•X•Y•Z** molecule has any influence on the rate of its own formation.

One indication that the reaction may be autocatalytic would be a sigmoidal shape of the quaternary reaction profile. However, the complex multi-step nature of the reaction may contribute to a sigmoidal shape through time lags. In fact, although many autocatalytic reactions do display sigmoidal product formation, such a result is not sufficient to demonstrate autocatalysis [31]. Moreover, the basic sigmoidal or logistic equations are strictly applied to a single-step reaction of type **A** → **B**, where **B** is the autocatalyst. Nevertheless, the reaction-time profiles in both the presence and absence of added **W•X•Y•Z** are distinctly sigmoidal. Under either of two experimental conditions without added **W•X•Y•Z** RNA, the data fit sigmoidal functions derived from an autocatalytic rate law for **A** → **B** with R^2 values of 0.99 (Figure 7, circles and dashed lines).

We acknowledge that fitting the data to a sigmoidal function is a statistical test of little power because alternative functions (such as a third-order polynomial) fit the data nearly as well. A rate increase upon addition of product would be a more robust test of autocatalytic behavior. We thus assayed whether the initial rate of the quaternary self-assembly reaction would be bolstered by the doping in of the full-length autocatalytic product prior to reaction initiation. To do this, we transcribed RNA from the RT-PCR product of the ~200 nt gel extract as in Figure 3. This ensured that the RNA used for doping was identical to the reaction products. When

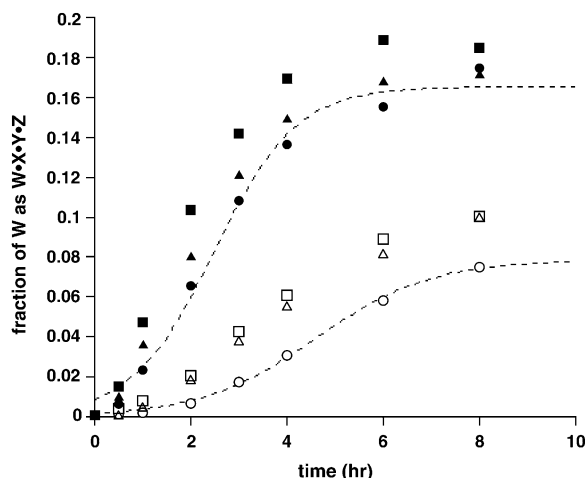


Figure 7. Kinetic Traces Demonstrating Autocatalysis

The oligomers **W**, **h•X**, **h•Y**, and **h•Z** were incubated at 48°C either without (circles) or in the presence of 0.5 μM (triangles) or 1.0 μM (squares) full-length **W•X•Y•Z** RNA. The reactions were performed either under optimal conditions of 2.0 μM each oligomer in 100 mM MgCl_2 (closed symbols) or under suboptimal conditions of 1.0 μM each oligomer in 25 mM MgCl_2 (open symbols). The full-length RNA was derived from a transcription of the RT-PCR amplicon of the ~200 nt bands such as shown in Figure 3A to best represent the product of the quaternary self-assembly reaction. The dashed lines represent fits of the nondoped experiments (circles) to sigmoidal curves.

this is done, the initial rate of product formation increases (Figure 7). We first examined the reaction under suboptimal conditions (1 μM each oligomer in 25 mM MgCl_2) that retard the net reaction to spotlight the influence of added product. When 0.5 μM product is doped in to a mixture of 1 μM each of **W**, **h•X**, **h•Y**, and **h•Z**, the initial velocity of the reaction jumps 2.5-fold, and when 1.0 μM product is doped in, it jumps 4.7-fold. Incidentally, the total yield of product after 8 hr, measured as the percentage of radiolabeled **W** that ends up as **W•X•Y•Z**, also increases upon doping. This value rises from 7.5% in the absence of added **W•X•Y•Z** to 10% with 0.5 μM or 1.0 μM added **W•X•Y•Z**. In these experiments, the doped **W•X•Y•Z** is not labeled, and thus these yields reflect only those of self-assembled RNA. In addition, we also observed rate enhancement upon doping under optimal reaction conditions (2 μM each oligomer in 100 mM MgCl_2 , as in Figure 3), where the initial velocity of the reaction jumps 3.5-fold when 0.5 μM product is doped into the reaction or 4.6-fold when 1.0 μM product is doped in. Therefore it was not necessary to minimize the nonautocatalytic rate (i.e., the *trans*-only activity) in order to get accurate estimates of the autocatalytic rate enhancement. In both cases, the initial rates of the reactions show a linear dependence on the initial concentration of added product. These rates are 0.00172 hr^{-1} , 0.00426 hr^{-1} , and 0.00802 hr^{-1} with 0 μM , 0.5 μM , and 1 μM added **W•X•Y•Z** RNA for suboptimal conditions, and 0.0233 hr^{-1} , 0.0369 hr^{-1} , and 0.0473 hr^{-1} for optimal conditions (linear regressions: $R^2 = 0.99$ for each condition).

We considered the possibility that the folded conformation of exogenously added **W•X•Y•Z** RNA may differ from that of the nascent recombined molecule.

Auspiciously, the faint band migrating at ~250 nt, seen in both **Figures 3A** and **3B**, is indicative that the **W•X•Y•Z** recombinants can in fact attain active folded state in situ, binding and cleaving an additional substrate, and transiently generating molecules such as **W•X•Y•Z•Z** en route to autocatalysis. For confirmation, we also excised the full-length RNA from the gel following the quaternary reaction, directly incubated it with **W** and **h•X** substrates for 1 hr, and observed recombination to **W•X**, providing further evidence that the full-length products are active **W•X•Y•Z** constructs that could be autocatalytic (data not shown). However, formal determination of the extent of autocatalysis in this system will require an exhaustive examination of each possible feedback reaction in detail.

Spontaneous Self-Assembly of Complex Ribozymes

The self-assembling and potentially self-replicating system described here differs from replicase-based systems [6, 7, 32] or template-directed ligation systems [2, 15, 16, 22, 33–35] in at least two important regards. First, our system can self-replicate because there are no competing choices for analogous substrates, such as those with GAU triplets in their 5' heads. The CAU triplets in substrate oligomers provide the only sequence recognition for the ribozyme. These CAUs thus serve as recombination signals and distinguish RNA molecules that can be recombined from those that cannot, in much the same way that recombination signals for contemporary protein enzymes work. In a heterogeneous pool of oligomers, recombination should produce and sustain quasispecies of recombinases, propagating each other as cooperative autocatalytic sets. Later, with the emergence of true RNA replicase activity, genotypic (selfish) selection would take over, as predicted by Kauffman [3]. Second, recombination by its very nature is an energy-neutral process. It has been noted that replication based on disproportionation reactions would be a “losing proposition” because the gain to one polymer would be offset by the loss to another [36]. This is certainly true, and would clearly be problematic for selection operating at the level of individuals, as would be expected for a RNA replicase whose fitness depends on accurate self-copying. The events described here though are intended to model the original advent of phenotypic activity. We view this as a necessarily cooperative process in which the proliferation of a coalition of molecules with catalytic activity needs to be established before Darwinian-type selection for individual genotypes or even facultatively cooperative hypercyclical systems (e.g., [33]) could be effective. From our experiments, it can be seen that complex catalytic RNAs can spontaneously self-assemble from coordinated fragments. The information transfer in this system is manifest in the CAU-IGS interaction, the matching base pairing among the fragments, and in the formation of catalytic pockets by complementary tertiary structures. Evolution to and from the assemblage engineered here should be possible by a variety of selection mechanisms.

Significance

As many as four inactive oligomers can cooperate through base pairing and tertiary interactions to pro-

duce a catalytically active variant of the *Azoarcus* group I intron. This coalition of molecules can then catalyze the production of covalently contiguous versions of the ribozyme, thereby articulating complex genotypes from simple ones. Likewise, this system provides a means for the advent of complex phenotypes other than the spontaneous assembly of an RNA replicase, which is currently an oft-sought goal of prebiotic chemistry. Although the necessary base-pairing interactions were engineered into this system, the bootstrapping of inactive fragments into a powerful catalytic entity provides a proof-of-concept for a plausible chemistry-to-biology transition during the RNA World. It should now be possible to produce mutagenized versions of the four fragments and ask which cooperative assemblages best self-assemble and reproduce. Replication by recombination requires cooperation among molecular fragments, giving support to the idea that life's beginnings may have had an altruistic essence prior to takeover by selfish replicators.

Experimental Procedures

RNA Synthesis

RNA fragments of the *Azoarcus* ribozyme were prepared by run-off transcription from double-stranded DNA derived from recursive gene synthesis as described previously [25]. All RNAs were gel purified prior to recombination experiments to insure length homogeneity. Radiolabeled fragments were either labeled on the 5' end after transcription with polynucleotide kinase and [γ - 32 P]ATP or internally with [α - 32 P]ATP in the transcription milieu. RNAs were prepared with a five-nucleotide head (h = GGCAU) and in earlier trials, with a tail (t = AAAUA) to permit them to be recombination substrates. Substrates for the binary reaction were **WX** and **h•YZ**, which recombine to generate **WX•YZ** with one CAU in L6; substrates for the ternary reaction were **WX**, **h•Y**, and **h•Z**, which recombine in two steps to generate **WX•Y•Z** with CAUs in L6 and L8; substrates for the quaternary reaction were **W**, **h•X**, **h•Y**, and **h•Z**, which recombine in three steps to generate **W•X•Y•Z** with CAUs in L5, L6, and L8 (**Figure 2A**).

Self-Assembly through Recombination

RNAs were mixed in desired ratios to a final concentration of 2.0 μ M each in water. In some experiments, only the **W** fragment was (5'-end) labeled. In other experiments, one of the four fragments was internally labeled with [α - 32 P]ATP. In all cases however, the total concentration of each oligomer was 2.0 μ M in the final reaction. Self-assembly and recombination was initiated by adding buffer to a final concentration of 100 mM MgCl₂ + 30 mM EPPS (pH 7.5) and incubating the reactions in 5 μ l volumes in 200 μ l thin-walled centrifuge tubes in a thermocycler at 48°C with a heated lid. After various times, reactions were quenched with EDTA and diluted in polyacrylamide gel-loading buffer containing SDS and urea. Samples were heated to 80°C prior to loading on 8% polyacrylamide/8 M urea gels and electrophoresed for approximately 2000 V hr in 1× TBE. For native gel electrophoresis, [α - 32 P]ATP-labeled fragments were incubated at 2 μ M concentrations in 30 mM EPPS (pH 7.5) containing either 10 mM or 100 mM MgCl₂ for 10 min at 48°C. One-fifth volume of prewarmed 6× native-gel loading dye (40% sucrose + 0.25% bromphenol blue) was added, and the samples were immediately loaded onto a 10% polyacrylamide gel containing 11 mM MgCl₂ and 1 mM EDTA but no urea. The gel was run at a constant 48°C temperature in a running buffer of 1× TBE and 10 mM MgCl₂ for approximately 1000 V hr. Gel images and product quantification were obtained by phosphorimaging.

Genotyping

Following self-assembly, bands corresponding to the ~200 nt RNA were excised from the acrylamide gel by elution into 100 mM NaCl + 10 mM EDTA. The eluted RNA was subject to RT-PCR amplification with an RT primer (5'-CCGGTTTGTGTGACTTTGCC-3') complementary to the **Z** fragment and second primer (5'-CTGCAGAAATTCTA

ATACGACTCACTATAGTCCTTGCCTGCGCCGGGAA-3') for the PCR that coincides with the 5' portion of the *W* fragment plus a promoter sequence for T7 RNA polymerase. Amplification of *Azoarcus* RNA with these primers gives a ~230 bp PCR product. This product was cut with *AluI*, *HindIII*, and *HhaI*, which generate clear RFLP patterns between *W•X•Y•Z* and *W•Y•X•Z*, and the fraction of each was quantified by fluorescence imaging following agarose gel electrophoresis and ethidium bromide staining. The PCR DNA was also subjected to analysis on an ABI 3100 automated sequencer, as well as transcribed into *W•X•Y•Z* RNA for analysis of autocatalysis.

Autocatalytic Analysis

The quaternary *W* (5' radiolabeled), *h•X*, *h•Y*, and *h•Z* mixture was incubated at either 2.0 μ M each in 100 mM MgCl_2 or at 1.0 μ M each in 25 mM MgCl_2 . These incubations were done in the absence of any covalently contiguous ribozyme (as above) and additionally in the presence of either 0.5 or 1.0 μ M RNA resulting from the RT-PCR and transcription of gel-excised full-length quaternary product. The four or five RNAs were incubated for 0.5, 1, 2, 3, 4, 6, and 8 hr at 48°C. The percentage of full-length product as a function of time was quantified by electrophoresis and phosphorimaging. The initial velocity was measured as the slope of the percent product versus time curves at 30 min, at which point all reactions had generated detectable *W•X•Y•Z* RNA but had achieved <8% of their maximal yields. At this time point, this calculation of initial velocity was identical to a differential method suggested by others for autocatalytic reactions in which rate constants cannot be easily modeled [37]. Data corresponding to no added *W•X•Y•Z* RNA were fit to sigmoidal functions of form:

$$y = (A_0 + B_0) / (1 + [A_0/B_0] \exp[-(A_0 + B_0) \times kx])$$

applied to the reaction formula $A \rightarrow B$ with rate law: $\text{rate} = k[A][B]$, by using Kaleidagraph software.

Contamination Controls

Strict measures were employed to ensure that individual RNA fragments were pure and uncontaminated by any full-length *Azoarcus* RNA during self-assembly experiments. These measures included use of dedicated rooms for DNA oligonucleotides preparation, use of self-contained hoods for PCR and RT-PCR set up, use of dedicated pipettes for various manipulations, and the use of barrier pipette tips at all times [38]. Both the full-length product (following gel excision) and the complete reaction mixtures themselves for both the quaternary and the tertiary self-assembly experiments were subject to RT-PCR analysis as described in *Genotyping* above. After RT-PCR amplification of the reaction mixtures themselves, only those reactions in which full-length products were seen in Figure 3A gave full-length (~230 bp) PCR products. Negative controls, in which one or more of the substrate oligomers were left out of the reaction, as well as RNA excised from blank regions of the gel failed to generate PCR products, confirming that all oligomer preparations and buffers were free of any *Azoarcus* ribozyme contamination.

Acknowledgments

We thank K. Rusterholtz for technical support and R. Barr, A. Burton, C. Riley, G. von Kiedrowski, and M. Zwick for comments on the manuscript. This work was supported by research grant NNG04-GM20G from the National Aeronautics and Space Administration.

Received: May 11, 2006

Revised: June 16, 2006

Accepted: June 26, 2006

Published: August 25, 2006

References

- Joyce, G.F., and Orgel, L.E. (1999). Prospects for understanding the origin of the RNA world. In *The RNA World* Second Edition, R.F. Gesteland, T.R. Cech, and J.F. Atkins, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Press), pp. 49–77.
- Sievers, D., and von Kiedrowski, G. (1994). Self-replication of complementary nucleotide-based oligomers. *Nature* 369, 221–224.
- Kauffman, S.A. (1993). *The Origins of Order* (New York: Oxford University Press).
- Orgel, L.E. (1992). Molecular replication. *Nature* 358, 203–209.
- Maynard Smith, J., and Szathmáry, E. (1995). *The Major Transitions in Evolution* (New York: Oxford University Press).
- McGinness, K.E., and Joyce, G.F. (2003). In search of an RNA replicase ribozyme. *Chem. Biol.* 10, 5–14.
- Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E., and Bartel, D.P. (2001). RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* 292, 1319–1325.
- Yarus, M. (1999). Boundaries for an RNA world. *Curr. Opin. Chem. Biol.* 3, 260–267.
- Joyce, G.F. (1991). The rise and fall of the RNA world. *New Biol.* 3, 399–407.
- Gilbert, W. (1986). The RNA world. *Nature* 319, 618.
- Ferris, J.P., Hill, A.R., Jr., Liu, R., and Orgel, L.E. (1996). Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* 381, 59–61.
- Maynard Smith, J. (1983). Models of evolution. *Proc. R. Soc. Lond. B. Biol. Sci.* 219, 315–325.
- Eigen, M. (1971). Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58, 465–523.
- Luther, A., Brandsch, R., and von Kiedrowski, G. (1998). Surface-promoted replication and exponential amplification of DNA analogues. *Nature* 396, 245–248.
- Kim, D.E., and Joyce, G.F. (2004). Cross-catalytic replication of an RNA ligase ribozyme. *Chem. Biol.* 11, 1505–1512.
- Paul, N., and Joyce, G.F. (2002). A self-replicating ligase ribozyme. *Proc. Natl. Acad. Sci. USA* 99, 12733–12740.
- Lehman, N. (2003). A case for the extreme antiquity of recombination. *J. Mol. Evol.* 56, 770–777.
- Santos, M., Zintzaras, E., and Szathmáry, E. (2004). Recombination in primeval genomes: a step forward but still a long leap from maintaining a sizeable genome. *J. Mol. Evol.* 59, 507–519.
- Doudna, J.A., and Cech, T.R. (1995). Self-assembly of a group I intron active site from its component tertiary structural domains. *RNA* 1, 36–45.
- Doudna, J.A., Couture, S., and Szostak, J.W. (1991). A multisubunit ribozyme that is a catalyst of and template for complementary strand RNA synthesis. *Science* 251, 1605–1608.
- van der Horst, G., Christian, A., and Inoue, T. (1991). Reconstitution of a group I intron self-splicing reaction with an activator RNA. *Proc. Natl. Acad. Sci. USA* 88, 184–188.
- Doudna, J.A., and Szostak, J.W. (1989). RNA-catalysed synthesis of complementary-strand RNA. *Nature* 339, 519–522.
- Kuo, L.Y., Davidson, L.A., and Pico, S. (1999). Characterization of the *Azoarcus* ribozyme: tight binding to guanosine and substrate by an unusually small group I ribozyme. *Biochim. Biophys. Acta* 1489, 281–292.
- Reinhold-Hurek, B., and Shub, D.A. (1992). Self-splicing introns in tRNA genes of widely divergent bacteria. *Nature* 357, 173–176.
- Hayden, E.J., Riley, C.A., Burton, A.S., and Lehman, N. (2005). RNA-directed construction of structurally complex and active ligase ribozymes through recombination. *RNA* 11, 1678–1687.
- Riley, C.A., and Lehman, N. (2003). Generalized RNA-directed recombination of RNA. *Chem. Biol.* 10, 1233–1243.
- Stahley, M.R., and Strobel, S.A. (2005). Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science* 309, 1587–1590.
- Adams, P.L., Stahley, M.R., Kosek, A.B., Wang, J., and Strobel, S.A. (2004). Crystal structure of a self-splicing group I intron with both exons. *Nature* 430, 45–50.
- Zaug, A.J., McEvoy, M.M., and Cech, T.R. (1993). Self-splicing of the group I intron from *Anabaena* pre-tRNA: requirement for base-pairing of the exons in the anticodon stem. *Biochemistry* 32, 7946–7953.
- Rangan, P., Masquida, B., Westhof, E., and Woodson, S.A. (2003). Assembly of core helices and rapid tertiary folding of a small bacterial group I ribozyme. *Proc. Natl. Acad. Sci. USA* 100, 1574–1579.

31. Sievers, D., Achilles, T., Burmeister, J., Jordan, S., Terfort, A., and von Kiedrowski, G. (1994). Molecular replication: from minimal to complex systems. In *Self-Production of Supramolecular Structures*, G.R. Fleischaker, S. Colonna, and P.L. Luisi, eds. (Dordrecht, The Netherlands: Kluwer), pp. 45–64.
32. McGinness, K.E., and Joyce, G.F. (2002). RNA-catalyzed RNA ligation on an external RNA template. *Chem. Biol.* 9, 297–307.
33. Lee, D.H., Severin, K., Yokobayashi, Y., and Ghadiri, M.R. (1997). Emergence of symbiosis in peptide self-replication through a hypercyclic network. *Nature* 390, 591–594.
34. Giver, L., Lato, S., and Ellington, A.D. (1994). Models for the autocatalytic replication of RNA. In *Self-Production of Supramolecular Structures*, G.R. Fleischaker, S. Colonna, and P.L. Luisi, eds. (Dordrecht, The Netherlands: Kluwer), pp. 137–146.
35. Tjivikua, T., Ballester, P., and Rebek, J. (1990). A self-replicating system. *J. Am. Chem. Soc.* 112, 1249–1250.
36. Joyce, G.F. (1996). Ribozymes: building the RNA world. *Curr. Biol.* 6, 965–967.
37. Mata-Perez, F., and Perez-Benito, J.F. (1987). The kinetic rate law for autocatalytic reactions. *J. Chem. Educ.* 64, 925–927.
38. Weber, D.S., Stewart, B.S., and Lehman, N. (2004). Genetic consequences of a severe population bottleneck in the Guadalupe fur seal (*Arctocephalus townsendi*). *J. Hered.* 95, 144–153.