BIOCHEMISTRY

Assembly of catalytic complexes from randomized oligonucleotides

Xu Han and Ulrich F. Müller*

Copyright © 2025 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

The early evolution of life relied on catalytic RNAs (ribozymes) for central functions. To test whether early catalysts could have assembled from multiple short nucleic acid fragments in random sequence environments, we performed an in vitro selection from a short RNA library in the presence of 256 different DNA 20-nucleotide oligomers. High-throughput sequencing and biochemical analysis showed that most of the selected 1331 RNA sequences required at least one DNA for activity. Representatives for four of six RNA clusters that depended on DNA cofactors were active even when the 256 DNAs were replaced by completely random DNA 20-nucleotide oligomers. The formation of these catalytic complexes and the recruitment of oligonucleotide cofactors from completely random libraries demonstrate an important principle for the emergence of the earliest oligonucleotide catalysts.

INTRODUCTION

Before the emergence of a translation apparatus, early life forms relied on catalytic RNAs (ribozymes) for central functions in metabolism and replication (*1*–*5*). This hypothesis is supported by the ribosome being a highly evolved catalytic RNA (*6*, *7*), by RNA's ability to catalyze a wide range of chemical reactions (*8*), and by many cofactors being derived from nucleotides (*9*). Unfortunately, macromolecules from that time, more than 3.5 billion years ago, will likely never be found because they would long have degraded. Instead, researchers use hints from commonalities of all known life forms (*10*), chemical principles (*11*), and experiments that generate model systems for parts of hypothesized RNA organisms. Such experiments usually aim to recapitulate aspects of an RNA organism with highly sophisticated ribozymes (*12*, *13*) or focus on the simplest self-replicating systems that could have emerged in a prebiotic environment (*14*, *15*). The current study tries to identify a plausible way in which some of the earliest nucleic acid catalysts could have emerged in a prebiotic setting.

An important constraint on the prebiotic emergence of catalytic RNAs is the apparent requirement for long RNA polymers. Libraries of RNAs longer than 100 nucleotides (nt) have a much higher chance of containing catalytically active structures because a longer-than-necessary RNA library allows a specific RNA structure to be assembled in increasingly many different registers, thereby markedly increasing the chance of finding a specific RNA structure within the library (*16*). This idea is supported by the success of libraries with more than 100 randomized positions to identify highly active and large numbers of different ribozymes (*17*–*22*). However, most prebiotic model reactions form oligomers of less than 50 nt, depending on the incubation conditions and time, whether activated monomers are added at multiple times, and whether homopolymers or heteropolymers are formed (*23*–*27*). While the formation of longer polymers is possible [(*28*, *29*), but see (*30*)], RNA oligomers below 50 nt were likely more abundant under prebiotic conditions. One possible solution to circumvent the requirement for long RNA polymers is to assemble large ribozymes from short RNA fragments, which is possible to test by fragmenting existing ribozymes (*31*–*35*). Some of these catalytic complexes catalyze the ligation of their own fragments into a continuous strand as a form of self-replication (*36*, *37*). However, all of these fragmented ribozymes were developed from existing longer ribozymes by design and optimization. To test a prebiotically more plausible scenario, here, we used an in vitro selection approach to test whether fragmented catalysts can assemble from random libraries containing short oligonucleotides.

As model system to test the assembly of catalytic complexes from random libraries, we used the in vitro selection for RNAs that catalyze the triphosphorylation of their 5′-hydroxyl groups with cyclic trimetaphosphate (cTmp) (*22*). This system selected RNAs catalyzing the reaction between their 5′-hydroxyl groups and cTmp with a k_{cat}/k_{uncat} of about 25,000 compared to the Mg^{2+}-catalyzed background reaction [compare (*22*) to (*38*)]. The reagent cTmp is prebiotically plausible because it can be generated by the condensation of phosphates or the reaction of phosphite with pyrophosphate (*39*–*42*). The idea that cTmp could have been used as energy source in early life forms is consistent with the formation of ATP from adenosine and cTmp (*38*, *43*) and the use of ATP as central energy currency in biology (*44*). The in vitro selection system for self-triphosphorylation ribozymes is suitable to serve as model system also because of its robust performance in six separate selection experiments, the extremely low false positive rate [at least 34 of 36 arbitrarily chosen cloned sequences showed self-triphosphorylation activity; (*22*)], and the discovery of several hundred different ribozyme clusters (*22*, *45*–*50*).

Here, we show the in vitro selection of catalytic oligonucleotide complexes from randomized oligonucleotide libraries. Using an RNA library with only 25 randomized positions and a secondary library composed of 256 different DNA 20-nucleotide (nt) oligomers, at least 14 different clusters of RNA triphosphorylation complexes were selected. Seven representatives were characterized biochemically, including one representative that did not require a DNA for activity. The six RNA/DNA complexes required between one and three DNAs for catalytic activity and showed different secondary structures. Four of the six RNAs were catalytically active with a library of completely randomized DNAs $(dN)_{20}$, suggesting that the assembly of catalytic complexes from short oligonucleotides was also plausible in a prebiotic environment.

Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA 92093, USA.
*Corresponding author. Email: ufmuller@ucsd.edu

Downloaded from https://www.science.org at University of California San Diego on April 30, 2025

RESULTS

To test whether catalytic complexes can assemble from short, randomized libraries we used an in vitro selection of ribozymes from randomized sequences that were thought to be too short to have activity on their own (Fig. 1). Specifically, we analyzed in a previous study what were the shortest active sequences in a selected library containing more than 300 different self-triphosphorylation ribozymes (*22*, *46*). The shortest active RNA had sequence constraints over a range of about 25 positions (*47*), which means that a new library containing 25 randomized positions would be expected to contain few or no examples of self-triphosphorylation ribozymes. Therefore, the primary library of the current study contained only 25 randomized nucleotides and was incubated with a secondary library to allow the formation of catalytically active complexes from both libraries. The secondary library was composed of 256 DNA 20-nt oligomers. By truncating the randomized region to 25 nt and giving access to the secondary library, we hypothesized that the randomized RNA in the primary library would contain RNAs that recruit DNA cofactors from the secondary library and, thereby, gain activity. DNA was used in the secondary library instead of RNA because obtaining 256 RNA 20-nt oligomers would have been cost prohibitive and because DNAs can be generated prebiotically via the same pathways as RNAs in the presence of H_2S (*51*, *52*). The number 256 was chosen as the number of DNA sequences because this made it possible to incubate 100 nM of the RNA library with 25.6 μM of the DNA library such that each of the 256 different DNAs was present at 100 nM concentration. This allowed every member of the RNA library to be saturated with any individual member of the DNA library. Additionally, this DNA concentration was low enough to prevent nonspecific inhibition of RNA catalysis, and the number of 256 DNAs made it feasible to identify activity-mediating DNAs for a given RNA after the selection by testing the DNAs in 16 sets of 16 different DNAs.

The selection system was based on our previously developed selection for ribozymes catalyzing the triphosphorylation of their own 5′-hydroxyl groups (Fig. 1A) (*22*). The library molecules were transcribed in the presence of a 5′-terminal hammerhead ribozyme, which cleaved itself off co-transcriptionally, resulting in a 5′-terminal hydroxyl group. These library molecules were purified by denaturing polyacrylamide gel electrophoresis (PAGE). In the selection described here, the RNA library was then incubated with 256 different DNA oligonucleotides (each at 100 nM, a total of 25.6 μM), 50 mM cTmp (Na_3cTmp), 100 mM $MgCl_2$, and 50 mM tris-HCl (pH 8.5). The sequences of the DNA cofactors were designed to represent random distributions with respect to nucleotide bias and the tendency to form self-structures (Fig. 1, B and C; sequences in table S1). During the incubation, the mixture was exposed to temperature cycling for 8 hours between 10° and 40°C and then kept at 4°C for about 8 hours (fig. S1). The first round of the selection covered an effective complexity of about 2×10^{14}, which covered about 18% of all possible sequences in the 25-nt-long random region (4^{25} ~11×10^{11}). Reacted library RNAs were then ligated to a biotinylated oligonucleotide using the R3C ligase ribozyme (*53*). The biotin-conjugated molecules were isolated by streptavidin-coated magnetic beads, reverse transcribed, and polymerase chain reaction (PCR) amplified to complete one round of selection and regenerate the library for the next round of selection.

Progress of the selection and HTS analysis

During the selection rounds, the number of PCR cycles required for PCR amplification of the reverse transcription products was monitored (fig. S2). During the first selection rounds, about 18 PCR cycles were required. After five rounds of selection, the number dropped to

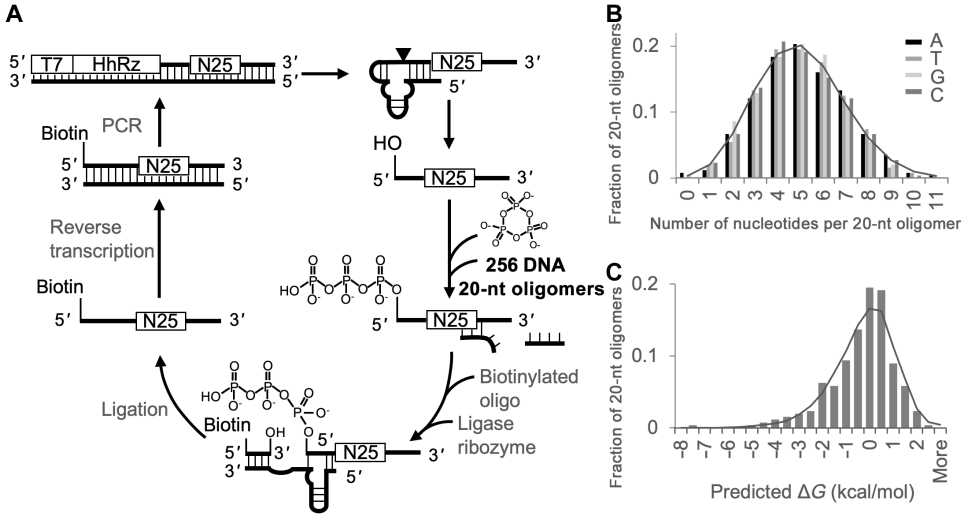


Fig. 1. Schematic for the in vitro selection of catalytic complexes. (**A**) Schematic for the selection procedure, modified after (*22*). The primary library was transcribed using a T7 promoter, starting with a hammerhead ribozyme (HhRz), which cleaved itself off co-transcriptionally and thereby generated a 5′-hydroxyl group at the library 5′-terminus. This primary library was incubated with cTmp and the secondary library (256 different DNA 20-nt oligomers) to allow the formation of RNA/DNA complexes that may catalyze the reaction of the 5′-hydroxyl group with cTmp. Sequences containing a 5′-triphosphate were ligated to a biotinylated oligonucleotide, isolated by capture on streptavidin-coated magnetic beads, reverse transcribed, and PCR amplified to regenerate the primary library for the next selection round. (**B**) Histogram showing the nucleotide bias of the 256 DNAs in the secondary library. Each of the four nucleotide distributions (columns in different gray scales) closely matched the distribution expected from an ideally distributed random library (gray line). (**C**) Histogram showing the self-folding energies of the 256 DNAs in the secondary library (gray columns). The distribution was close to the distribution of 30,000 randomly generated sequences (gray curve), suggesting that the 256 DNAs closely represented a random distribution.

Downloaded from https://www.science.org at University of California San Diego on April 30, 2025

about eight cycles, suggesting that the population was then dominated by active sequences. The sixth round of selection was performed in 18 variations, to allow high-throughput sequencing (HTS) analysis to identify the DNA sets that may have resulted in activity. This analysis compared the frequency of sequence clusters in the absence of the DNA library, in the presence of all 256 DNAs, and in the presence of each of the 16 sets (A to P) of DNAs. A total of 1331 RNA sequences had at least 500 reads in the sixth round of the selection (Fig. 2A). The enrichment of each RNA in the 18 sub-rounds of selection round 6 was calculated to identify the effects from different DNA sets (Fig. 2B). The peaks in this enrichment plot suggested that many RNAs were enriched up to 1000-fold by specific DNA sets.

Forty-five RNA sequences were chosen for biochemical testing, in two steps. In the first step, the HTS results were analyzed for the enrichment of RNA sequences by a specific DNA set as shown in Fig. 2B. For each of the 16 DNA sets, the two most enriched sequences were chosen. To avoid statistical problems from calculating the enrichment by dividing with a low frequency in the absence of DNA, this choice was made two times: for RNAs with more than 50 reads without DNA and for RNAs with 6 to 50 reads without DNA. Due to overlaps of sequences between these groups, this resulted in 39 RNA sequences. Additionally, 25 RNA sequences were chosen that showed 0 to 5 reads without DNA but more than 100 reads with the 256 DNAs, 12 RNA sequences were chosen where Fig. 2B suggested a strong discrimination between DNA sets, and three RNAs from cloning arbitrary sequences and preliminary biochemical testing were added. This resulted in 79 RNA sequences that were promising to cover different behavior with respect to the DNAs. From these 79 sequences, 34 sequences were removed to reduce redundancy of very similar sequences. This resulted in 45 RNAs that were used for biochemical analysis (Fig. 2, lines in dark gray and in color). More details are given in the Supplementary Materials.

Identification of DNA cofactors and clustering of 45 RNAs

To test whether the identified RNA sequences would benefit from DNA cofactors as suggested by HTS analysis, a biochemical assay was carried out in two steps. The RNAs were prepared with a 5′-hydroxyl terminus and incubated with cTmp under selection conditions, in the presence of specific DNAs. Possible self-triphosphorylation was then analyzed by ligating 5′-triphosphorylated RNAs to a 5′-[^{32}P]–labeled oligonucleotide using the R3C ligase ribozyme and separating by denaturing PAGE. The signal intensities of the gel shifted bands reported on the extent of the self-triphosphorylation reaction (Fig. 3A, left). In the first step of the assay, DNAs were added as 16 sets of 16 DNAs, which covered all 256 DNAs used in the selection. In the second step of the assay, the same procedure was carried out by adding the 16 individual DNAs from the DNA sets that were found to mediate activity (Fig. 3A. right). In cases where no single DNA set was able to replicate the activity shown with all 256 DNAs, we assumed that multiple DNA sets were necessary for activity. These DNAs were identified by first adding a mixture containing 15 of the 16 DNA sets to the reaction mixture and omitting each DNA set in turn, thereby identifying the DNA sets required for activity (Fig. 3B, top). In the second step, individual DNAs were omitted from the DNAs of the identified two to three DNA sets, to determine the individual DNAs mediating activity (Fig. 3B, bottom). Last, the results were tested by measuring activity with only the two to three chosen DNAs. Therefore, the identity of the beneficial DNA(s) was determined for each of the 45 RNAs (Fig. 3 and fig. S3A).

The biochemical analysis showed that 24 of the 35 RNAs with clear biochemical activity (69%) were dependent on specific DNAs. The activity of a given RNA with multiple DNA sets showed a complex pattern of RNA activity dependence on DNAs (fig. S3B). The comparison of results from HTS analysis and biochemical experiments also showed that the HTS analysis was a good predictor for the DNA benefit on biochemical activity (fig. S3C): Of the 35 biochemically tested RNAs with an HTS read increase of more than 10-fold in the presence of DNAs, 28 showed at least 10-fold activity enhancement in the presence of their most beneficial DNA set. In contrast, of the 10 biochemically tested RNAs with a less than 10-fold HTS read increase in the presence of DNAs, only two showed more than a 10-fold increase in biochemical activity.

The sequence alignment and biochemical testing of the 45 RNA sequences allowed estimating the fraction of the 1331 sequences

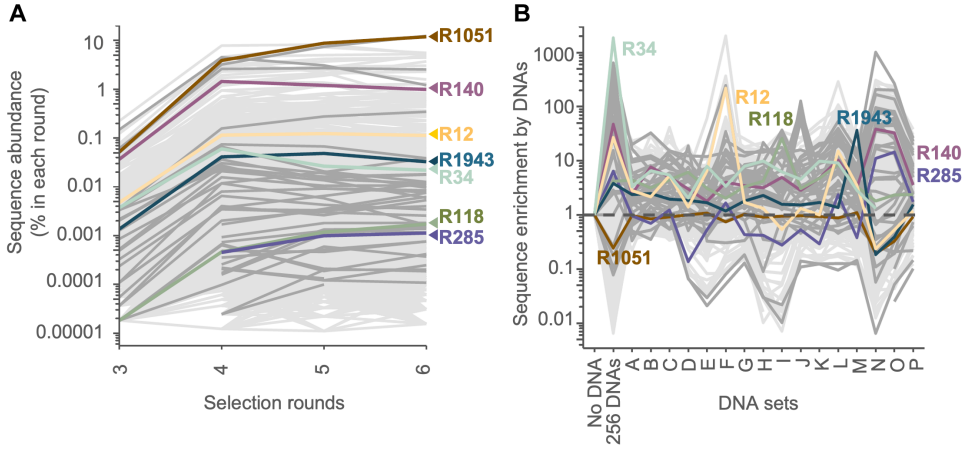


Fig. 2. HTS analysis of the selected RNAs. (**A**) Enrichment of sequences over selection rounds 3 to 6. The abundance of individual sequences is plotted as a function of the selection rounds. The 1331 sequences with at least 500 reads in selection round 6 are shown in light/dark gray or color. The 45 sequences that were chosen for biochemical analysis are in dark gray or color. The seven sequences that were chosen for more detailed analysis are in color, with the RNA sequence indicated. (**B**) Analysis of sequence enrichment in the presence of DNAs. Shown is the abundance of a given sequence in the presence of DNAs (either all 256 DNAs or 1 of the 16 DNA sets A to P) divided by the sequence abundance in the absence of DNAs. The color coding is the same as in (A).

Downloaded from https://www.science.org at University of California San Diego on April 30, 2025

that required DNA cofactor use. This consideration assumed that all sequences within any given cluster behaved the same with respect to DNA cofactor use, which seemed to be the case based on the analysis of 45 sequences (Fig. 4A). In total, nine clusters containing 611 sequences (C2, C4, C5, C6, C7, C9, C10, C13, and C14) depended on DNAs for catalytic activity. One cluster (C11) with 150 sequences did not require a DNA for activity. Four clusters with a total of 388 sequences (C1, C3, C8, and C12) showed no detectable activity under all tested conditions. A total of 182 sequences were not captured by these 14 clusters. Under the assumption that DNA cofactor use reflected the behavior of the 10 tested clusters with detectable activity (611 + 150 sequences), around 80% of the selected RNA sequences required a DNA cofactor for activity (see fig. S4 for details).

The sequence alignment tree of the 45 tested RNA sequences closely reflected their use of DNA cofactors, together delineating at least 14 clusters (Fig. 4A). Clusters 13 and 14 contained most of the tested sequences and formed two large branches, although they recruited the same DNA cofactors. These two branches were still designated into two clusters because they were distant in sequence and appeared to reflect different secondary structures. The distance between the sequences was plotted as a network, illustrating the clustering distribution (Fig. 4B). The average distance between clusters was about 12 edits, and the shortest distance between sequences of

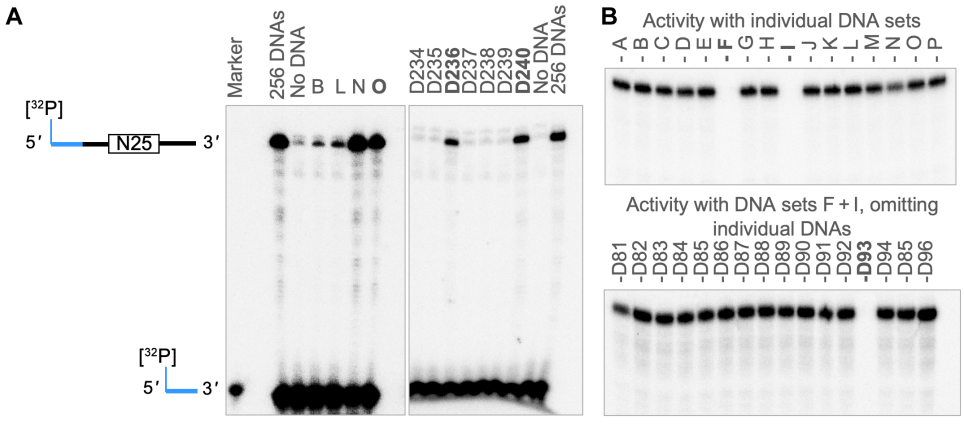


Fig. 3. Biochemical identification of DNA cofactors for 45 selected RNAs. (**A**) RNA self-triphosphorylation activity was detected by ligation to an equimolar amount of 5′-[^{32}P]–radiolabeled oligonucleotide (blue) using the R3C ligase ribozyme and quantifying the portion of gel shifted signal. In the first step of identifying required DNAs (left), RNA activity was tested in the presence of either of 16 sets of 16 DNAs (the shown example contains DNA sets B, L, N, and O). In the second step (right), individual DNAs of the activity mediating DNA sets were tested (here, DNAs D236 and D240 mediated activity). (**B**) In cases where multiple DNA sets were required to mediate activity, RNA activity was tested in the presence of 15 of the 16 DNA sets, omitting every DNA set in turn (top). The absence of the required DNA sets resulted in the loss of RNA activity (here, DNA sets F and I). In the second step (bottom), RNA activity was tested in the absence of each individual DNA from the two required DNA sets. Loss of activity indicated the requirement for that specific DNA (here, D93). Data for all tested RNAs are shown in fig. S3.

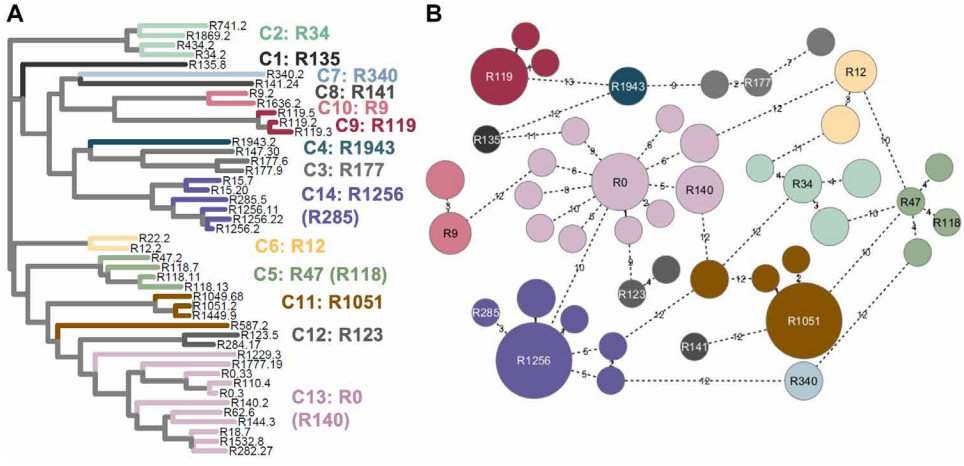


Fig. 4. Clustering analysis for 45 selected RNAs. (**A**) Sequence alignment tree of the 45 biochemically tested RNAs. Clustering of the RNAs by sequence alignment coincided by the RNA's use of specific DNA cofactors as coded by different colors. Clusters were determined by both RNA sequence distances and the DNA cofactors they recruited. The name of each of the 14 clusters and its most abundant sequence are given on the right. Sequences given in parentheses name the representative RNAs that were tested biochemically. The detailed tree with all 1331 sequences is shown in fig. S4. (**B**) Network illustrating the edit distances between the biochemically tested 45 RNAs (circles). The size of the circles represents the relative abundance of the RNA in selection round 6. The color of the circles reflects DNA cofactor use and is the same as the 14 clusters in panel A and Fig. 2. The numbers on the lines are the edit distances between the two connected RNAs.

Downloaded from https://www.science.org at University of California San Diego on April 30, 2025

different clusters was 9 edits. This suggested the absence of neutral pathways (*54*) between these clusters because the selection covered about 18% of the theoretical sequence space, and more bridging sequences would be expected if neutral paths existed.

Biochemical analysis of six catalytic complexes

Six catalytic complexes and one RNA that did not require DNA cofactors were characterized biochemically in more detail. The choice of these RNAs was made to cover different requirements of DNA cofactors. Because the RNAs within each cluster showed the same requirement for DNAs, one representative was chosen for each cluster of interest. Ten of the 14 clusters showed biochemical activity. These 10 required none, one, or two DNAs or tolerated different alternative DNAs. Cluster 11 (representative R1051) did not require a DNA; cluster 4 (R1943) required one DNA; cluster 2 (R34) required two DNAs; and clusters 5 (R118), 6 (R12), 13 (R140), and 14 (R285) tolerated multiple different DNAs. These seven representative RNAs were characterized with respect to the minimal number of DNAs that were required for their maximum activity, the minimal lengths of DNAs and RNAs, the minimal concentration of DNAs required for maximum activity, their optimum reaction temperatures, the reaction kinetics, and their secondary structures. Four representative complexes that require three, two, or one DNA cofactor are described below. Two additional complexes that require one cofactor and one ribozyme that did not require DNA for activity are described in the supplementary figures (figs. S9 to S11 and table S2).

Catalytic complex of RNA R34 with DNAs D93, D132, and D221

RNA R34 required three different DNAs for maximum activity (Fig. 5). During the last selection round that tested different DNA sets, the R34 sequence enriched only in the presence of 256 DNAs, but not in the presence of any individual DNA set (Fig. 2B, teal curve). This suggested that multiple DNA sets were required for activity. Accordingly, the biochemical assays found that the DNA D93 (DNA set F) as well as DNA D132 (DNA set I) were required for the RNA to be triphosphorylated and that DNA D221 (DNA set N) increased activity by about twofold (Fig. 5A and fig. S5A). Because this was the minimal number of DNAs that mediated maximal activity of R34 (fig. S5B), the following experiments were conducted with the combination of R34/D93/D132/D221. Because no activity was detected without DNAs, the DNAs appear essential to form the active structure or may be directly involved in catalysis. RNA R34 and its DNAs were then truncated to their smallest size that mediated full activity. While each DNA could be truncated by 3 to 6 nt (fig. S5, C and D), the activity was affected when all three DNAs were truncated at the same time. Therefore, D93, D132, and D221 were eventually truncated by 4, 2, and 0 nt, respectively (fig. S5E). RNA R34 could not be truncated without activity loss (Fig. 5D and fig. S5F). To determine

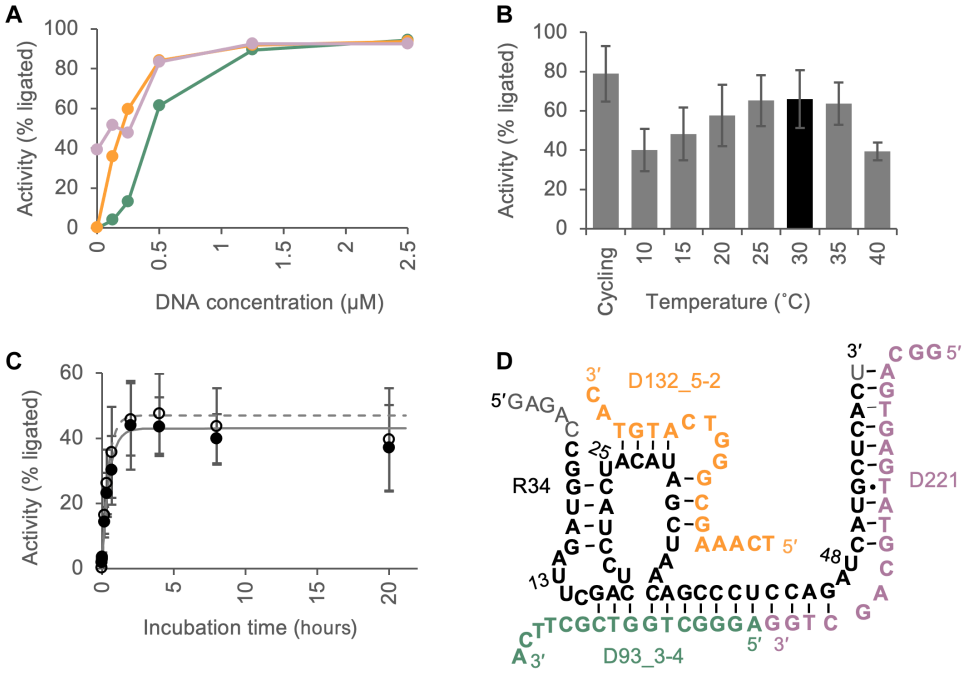


Fig. 5. Biochemical characterization of the catalytic complex of RNA R34 with DNAs D93, D132, and D221. (**A**) Activity of the complex as a function of DNA cofactor concentration. With the R34 concentration at 0.25 μM, each DNA was titrated from 0.125 to 2.5 μM, and the other two DNAs at 1.75 μM. Colors represent D93 (green), D132 (yellow), and D221 (pink). (**B**) Activity of the complex as a function of temperature. The complex with optimized DNA concentrations (1.25 μM D221, 2.5 μM D93, and 2.5 μM D132) was incubated under the same cycling temperature used in the selection. The black column shows the temperature which was used in later assays. Error bars are SDs from triplicate experiments. (**C**) Kinetics of the catalyzed reaction after preheating to 40°C and then dropping to 30°C (solid gray line and filled circles), compared to constant temperature at 30°C (dashed gray line and empty circles). The gray curves are least-square fitted first-order exponential functions. Error bars are SDs from triplicate experiments. (**D**) Secondary structure model of the R34 complex with D93, D132, and D221. The model was generated on the basis of in-line probing data in the absence and presence of different DNAs (fig. S5, J to L). Positions in gray did not yield secondary structure information from inline probing because of their closeness to the termini. The labels of the DNAs include information on their truncation, with _5-2 corresponding to two truncated nucleotides at the 5′-terminus and _3-4 corresponding to four truncated nucleotides at the 3′-terminus.

Downloaded from https://www.science.org at University of California San Diego on April 30, 2025

the lowest concentration of each DNA cofactor required for maximum activity, the truncated DNAs were titrated individually in the presence of 250 nM RNA R34 (Fig. 5A). The activity increased from 26% when each DNA was equimolar with R34 (fig. S5E) to 94% when DNAs D93, D132, and D221 were used in 10-fold, 10-fold, and 5-fold molar excess over R34, respectively (Fig. 5A). This suggested that active complex formation relied on weak interactions between R34 and its DNA cofactors. The activity of this complex showed a weak temperature preference for 25° to 35°C (Fig. 5B and fig. S5H). The reaction kinetics at 30°C were similar with and without preheating to 40°C (0.035 versus 0.040 min^{-1}); therefore, the preheating steps used during the selection procedure were eliminated from biochemical analysis of R34 (Fig. 5C and fig. S5I). The secondary structure analysis of R34 by in-line probing (Fig. 5D and fig. S5J) found strong protection in the presence of all three DNAs, with flexibility only at positions 13 to 15, 25, and 48 (fig. S5K). In-line probing of the RNA without DNAs led to extensive cleavage throughout the sequence, suggesting that the RNA alone formed little secondary structure (fig. S5L). The addition of any two DNAs reduced the cleavage at specific regions that informed about their interaction with the RNA (fig. S5J). Together, these data suggested a secondary structure model of the complex R34/D93/D132/D221 where the DNA cofactors paired to most RNA positions in a well-protected structure (Fig. 5D). The cluster represented by R34 (cluster 2) contained only 5 of the 1331 sequences (fig. S4), which suggests a narrow fitness peak and that only few mutations were tolerated by this RNA.

The catalytic complex of RNA R1943 and DNA D198

RNA R1943 appeared to form a complex with DNA D198 in a 1:2 stoichiometry, as judged by the DNA titration and secondary structure analysis (Fig. 6). This stoichiometry is supported by the shape of the titration curve between 250 nM R1943 and increasing concentrations of D198, which is fit well by two straight lines that suggest a tight interaction of 250 nM R1943 with 500 nM D198 (Fig. 6A). Because no activity was detected without DNAs, the DNAs appeared essential for formation of the active structure or may be directly involved in catalysis. Additionally, the inline probing data were best described with a secondary structure model containing two molecules of D198 (Fig. 6 and fig. S6). To obtain the minimal size of the complex with maximum activity, the RNA R1943 was truncated by 7 nt at its 3′-terminus, while D198 required its full length for maximum activity (fig. S6, A and B), and the DNA was used at a fivefold molar excess over R1943 (Fig. 6A and fig. S6C). The optimal temperature was 35°C (Fig. 6B and fig. S6D). The preheating to 40°C was omitted because, at 35°C temperature, the 40°C preincubation generated only an insignificant benefit for reaction kinetics (0.018 versus 0.012 min^{-1}) (Fig. 6, B and C, and fig. S6E). While the secondary structure model explained a large part of the inline probing data (Fig. 6D and fig. S6, F to H), a total of 12 nt within RNA positions 15 to 28 appeared to be protected by tertiary contacts that were not captured in the secondary structure model. The cluster represented by R1943 (cluster 4) contained only 4 of the 1331 sequences (fig. S4), which suggests a narrow fitness peak and that only few mutations were tolerated by this RNA.

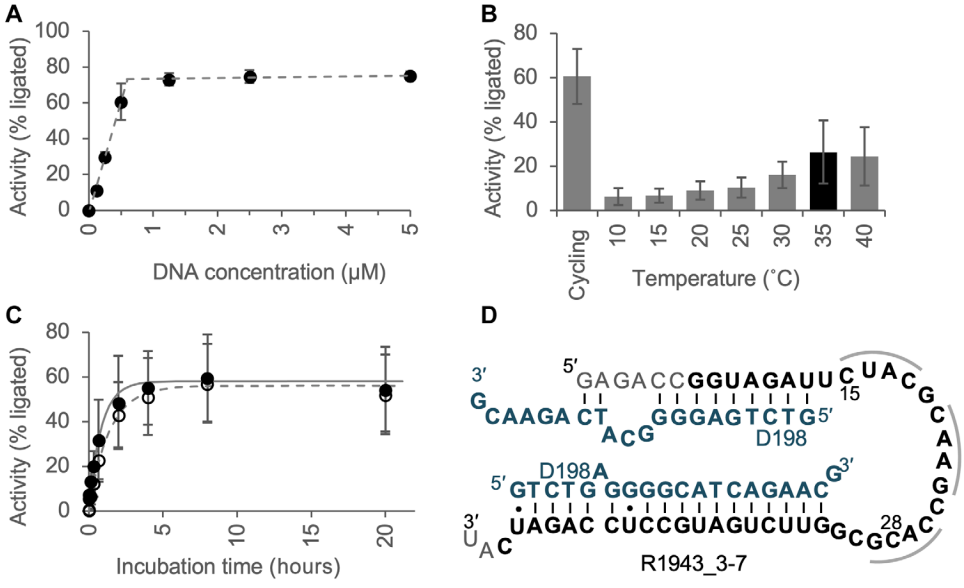


Fig. 6. Biochemical characterization of the catalytic complex between RNA R1943 with DNA D198. (**A**) Titration of D198 to determine the concentration that can achieve maximum activity. With R1943 at 250 nM, D198 was titrated from 0.125 to 5 μM. Dashed lines show the linear fit of the unsaturated phase and the saturated phase, with an intercept at 613 nM D198. Error bars are SDs from triplicate experiments. (**B**) Histogram showing the temperature preference of R1943/D198 complex activity. The complex with 500 nM R1943 and 1.25 μM D198 was incubated with the same temperature cycling profile used in the selection or one of the 1-hour cycles at the given temperature. The black column indicates the temperature that was used in the later assays. Error bars are SDs from triplicate experiments. (**C**) Kinetics of the catalyzed reaction when preheating to 40°C and then incubation at 35°C (solid gray line and filled circles) or constant temperature at 35°C (dashed gray line and empty circles). The gray curves are first-order exponential functions least-square fitted to the data. Error bars are SDs from six replicate experiments. (**D**) Secondary structure model of complex R1943 with two molecules of D198. The model was generated on the basis of in-line probing (fig. S6). Thick gray lines indicate protection from in-line probing that is not captured by the secondary structure. Positions in gray did not yield secondary structure information from inline probing because of their closeness to the termini. The label of the RNA includes information on its truncation, with _3-7 corresponding to seven truncated nucleotides at the 3′-terminus.

Downloaded from https://www.science.org at University of California San Diego on April 30, 2025

The catalytic complex of RNA R140 with DNA D221

The RNAs R140 and R285 represented the two most abundant clusters in the selection, clusters 13 and 14, respectively. The characterization of R140 is shown below (Fig. 7), while the results for RNA R285 are shown in the supplement (fig. S10). Both RNAs benefited from the same DNAs D221, D240, and D236 with most benefit from D221 (DNA set N) and D240 (DNA set O), consistent between HTS analysis and biochemical analysis (Fig. 2B and figs. S7A and S10). Because R140 showed detectable activity without DNA, the DNA does not appear to be involved in catalysis. During biochemical analysis, the two DNAs D221 and D240 were truncated by 6 and 1 nt at their 5′-terminus, respectively, while maintaining full activity (fig. S7B). In the absence of D240, RNA R140 was still able to achieve maximum activity when D221 was used at a fourfold excess over R140 (Fig. 7A and fig. S7D); therefore, only D221 was used as cofactor. The shape of the RNA activity as function of DNA concentration suggested that R140 and D221 bound tightly to each other in a 1:1 stoichiometry: The data were best described by two straight lines that intersected at 250 nM D221 concentration when 250 nM R140 was used. This 1:1 stoichiometry was consistent with the secondary structure model based on inline probing data, where two duplexes, 7 and 4 base pairs long, formed between R140 and D221 (Fig. 7D and fig. S7, G to I). The catalytic activity of the complex appeared highest at 30°C (Fig. 7B and fig. S7E), with kinetics of 0.009 min^{-1} with or without preheating to 40°C (Fig. 7C and fig. S7F). RNA R285 from cluster 13 also appeared to form a 1:1 complex with DNA D221 but with a different secondary structure (fig. S10I) and with faster reaction kinetics of 0.023 min^{-1}, at the optimal temperature of 25°C (fig. S10F). The dominance of the R140/D221 and R285/D221 clusters in the selected population of RNAs can be explained by DNA D221 having some complementarity to the 3′ conserved region of the RNA library, helping the recruitment of D221. The cluster represented by R140 (cluster 13) contained 414 of the 1331 sequences (fig. S4), which suggests a broad fitness peak and that many mutations were tolerated by this RNA.

The catalytic complex of RNA R12 and DNA D88

The catalytic complex of RNA R12 appeared to form from two molecules of RNA and one molecule of DNA D88 (Fig. 8). Because a very low level of activity was detected without DNA, the DNA did not seem to be directly involved in catalysis. RNA R12 was truncated by 1 nt at its 3′-terminus, and D88 was truncated by 1 nt at 3′- and 5′-termini while maintaining full activity (fig. S8, B and C). When the concentration of DNA D88 was increased at a constant concentration of R12 at 250 nM, the catalytic activity was best described by two straight lines that intersected close to 125 nM DNA. This suggested that one D88 molecule was sufficient to activate two R12 molecules (Fig. 8A and fig. S8D). The complex had a broad temperature optimum around 25°C (Fig. 8B and fig. S8E), with reaction kinetics of 0.005 and 0.004 min^{-1} with or without preheating to 40°C (Fig. 8C and fig. S8F). Consistent with the DNA titration, the inline probing data supported a secondary structure formed by two R12 molecules and one D88 molecule (Fig. 8D and fig. S8, G to I). It is unclear how such a complex could have formed during the first selection rounds, where it would be unlikely for two R12 molecules to find each other. The cluster represented by R12 (cluster 6) contained only 6 of the 1331 sequences (fig. S4), which suggests a narrow fitness peak and that only few mutations were tolerated by this RNA.

Downloaded from https://www.science.org at University of California San Diego on April 30, 2025

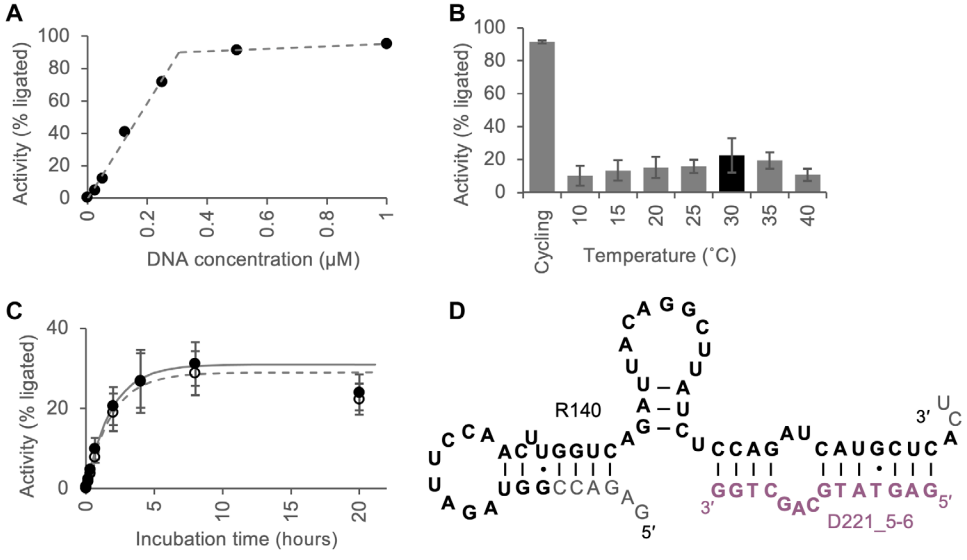


Fig. 7. Biochemical characterization of the catalytic complex between RNA R140 with DNA D221. (**A**) Catalytic activity as function of the DNA D221 concentration, at a fixed concentration of 250 nM R140. Dashed lines show a linear fit for the unsaturated phase and another linear fit for the saturated phase, with an intercept at 305 nM D221_5-6. (**B**) Catalytic activity at different temperatures. The complex with 1 μM D221_5-6 was incubated with the same temperature cycling profile as used in the selection (left column) or for one 1-hour cycle at the given temperatures. The black column indicates the temperature that was used in the later assays. Error bars represent SDs from triplicate experiments. (**C**) Reaction kinetics of the R140/D221complex with preheating to 40°C and then incubation at 30°C (solid gray curve and filled circles) or without preheating and incubation at 30°C (dashed gray curve and empty circles). The gray curves are first-order exponential functions least-square fitted to the data. Error bars represent SDs from triplicate experiments. (**D**) Secondary structure model of complex R140 with D221. The model was generated with the help of the mfold software, based on data from in-line probing. The raw image and analytical data are in fig. S7. Positions in gray did not yield secondary structure information from inline probing because of their closeness to the termini. The label of DNA D221 includes information on its truncation, with _5-6 corresponding to six truncated nucleotides at the 5′-terminus.

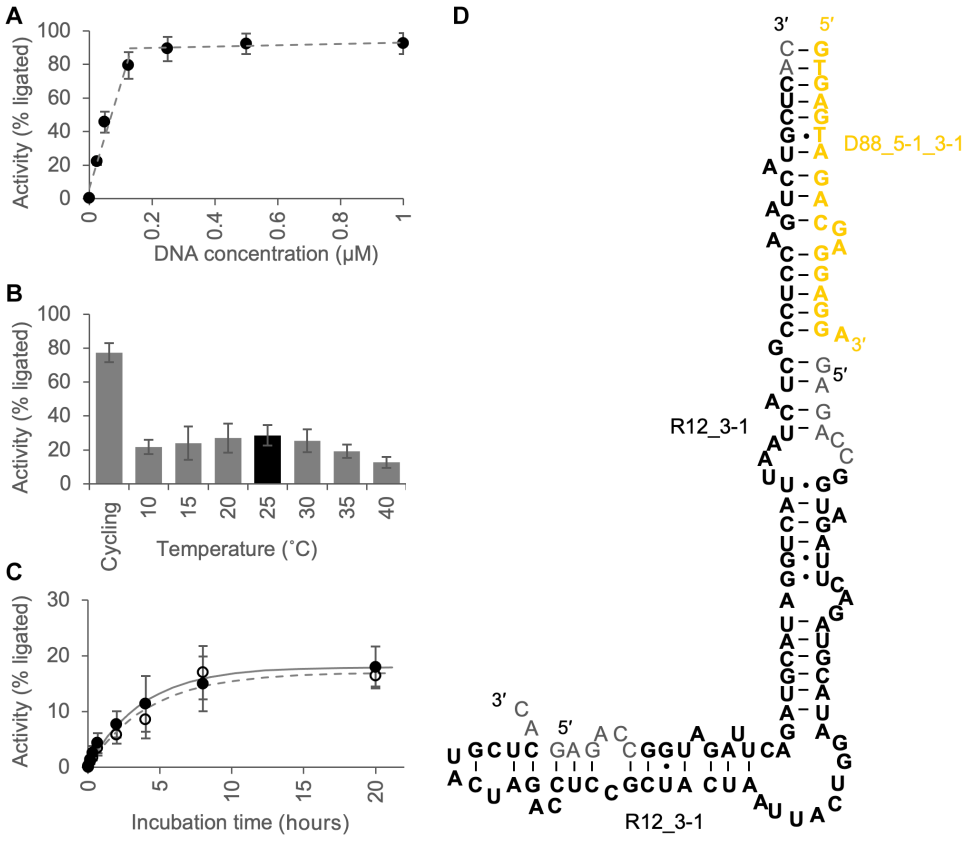


Fig. 8. Biochemical characterization of the catalytic complex between RNA R12 with DNA D88. (**A**) Titration of D88 with 250 nM R12. Dashed lines show the linear fit of the unsaturated phase and the saturated phase, with the intersection at 136 nM DNA. Error bars are SDs from triplicate experiments. (**B**) Temperature dependence of the R12/D88 catalytic activity. The complex with 250 nM DNA D88_5-1_3-1 was incubated under the same temperature cycling profile as used in the selection (left column) or for one 1-hour cycle at the given temperature. The black column indicates the temperature that was used in the later assays. Error bars represent SDs from triplicate experiments. (**C**) Reaction kinetics with the R12/D88 complex after preheating to 40°C and then incubation at 25°C (solid gray line and filled circles) or without preheating and incubation at 25°C (dashed gray line and empty circles). The gray curves are first-order exponential functions least-square fitted to the data. Error bars represent SDs from triplicate experiments. (**D**) Secondary structure model of the complex formed by two molecules of R12 and one molecule of D88. The model is generated on the basis of in-line probing assisted by the mfold software. The raw image and analytical data are in fig. S8. Positions in gray did not yield secondary structure information from inline probing because of their closeness to the termini. The labels of the RNAs and DNA include information on their truncation, with _5-1 corresponding to one truncated nucleotide at the 5′-terminus and _3-1 corresponding to one truncated nucleotide at the 3′-terminus.

A combinatorial experiment in a prebiotically more relevant scenario

To test whether the characterized, catalytic complexes would be affected by each other's presence, the catalysis of all seven RNAs was monitored in a one-pot reaction. This included six RNAs that required DNA cofactor(s) and one RNA that did not require a DNA cofactor (R1051). To allow the simultaneous monitoring of all seven reactions in the same reaction mixture, the 3′-termini of the seven RNAs were extended or truncated such that the assay for the triphosphorylation reaction generated a different gel shift for each RNA's product (Fig. 9A, fig. S12A, and table S2). When the seven RNAs were incubated with all 256 DNAs used in the selection procedure, all seven RNAs showed activity as expected (Fig. 9B, left and fig. S12, J and K).

To simulate a more complex scenario, the 256 DNAs were replaced with the same total concentration (25.6 μM) of randomized DNA 20-nt oligomers $(dN)_{20}$. We were delighted to find that four of the six RNAs that required DNA cofactors were active with $(dN)_{20}$ (Fig. 9B, right). The same pattern of activity resulted with three batches of separately synthesized, randomized DNA 20-nt oligomers (fig. S12L). This means that the four RNAs—R140, R12, R118, and R285, all of which recruited a single DNA—were able to recruit activity-mediating DNAs from the mixture containing 4^{20} or about 10^{12} different DNA sequences. The two RNAs that remained inactive were R34 that formed a complex with three DNAs (Fig. 5) and R1943 that apparently recruited two molecules of D198 (Fig. 6). Because each of the DNA sequences existed at a concentration of around 2×10^{-17} M (25.6 μM/4^{20} ~2×10^{-17} M) and around 10 nM of each RNA reacted (4% of 250 nM = 10 nM), each of the DNA sequences existed at about 4×10^{8}–fold lower concentration than the reacted RNA. This means that the RNA molecules tolerated a large number of mutated positions in the recruited DNAs (see Discussion). Such a high sequence tolerance may have been important for the formation of the first catalysts from mixtures of short oligonucleotides in a complex prebiotic environment.

Downloaded from https://www.science.org at University of California San Diego on April 30, 2025

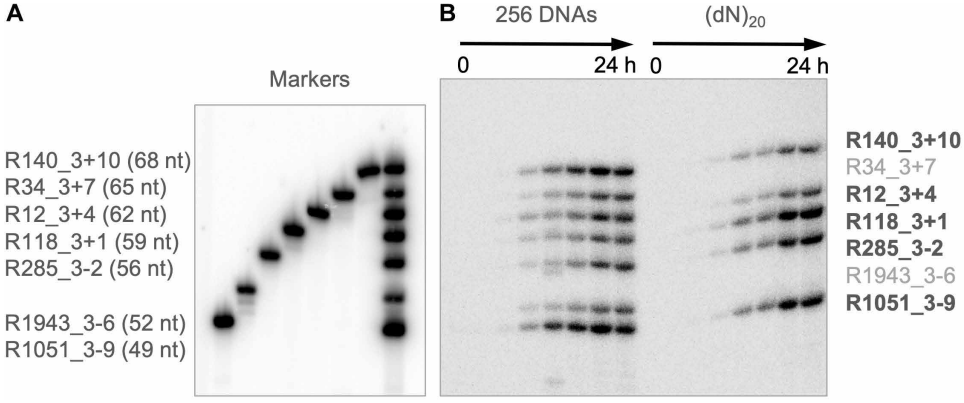


Fig. 9. Simultaneous reaction of seven RNAs and their catalytic complexes. (**A**) PAGE separation of seven radiolabeled RNAs with seven different lengths. The seven left lanes show the individual loading of these RNAs, and the right lane shows that these RNAs are well-separated by PAGE within the same sample. (**B**) One-pot reaction time courses of the seven RNAs with cTmp and 256 DNAs that were used in the selection (left) and randomized DNA 20-nt oligomers (right), each at a total RNA concentration of 250 nM and a total DNA concentration of 25.6 μM. The names of the RNAs are shown on the right, together with their length modifications. Complete data are shown in fig. S12. h, hours.

DISCUSSION

This study addressed the question whether the assembly of oligonucleotide complexes from randomized libraries could have facilitated encoded oligonucleotide catalysts to emerge in a prebiotic environment that contained only short oligonucleotides. To do that, an in vitro selection of RNA sequences with a 25-nt-long random region was conducted in the presence of 256 DNA oligomers representing a random distribution. The selection identified multiple different complexes that were active by RNA/DNA assembly. Most (~80%) of the selected RNA sequences depended on DNAs for catalytic function. Four of the six biochemically studied RNAs were able to recruit cognate DNA(s) from a completely random DNA library $(dN)_{20}$, which makes the principle of oligonucleotide recruitment suitable for a prebiotic environment.

Two previous studies have identified the emergence of molecular cooperation during ribozyme in vitro evolution experiments. One study evolved fragments of the *Azoarcus* group I intron ribozymes that reconstituted the ~200-nt-long full-length strand through cooperation between the fragment termini (*55*). Networks of cooperating variants emerged with different numbers of cooperators for self-replication. The second study (*56*) evolved a polymerase ribozyme for extension with trinucleotide 5′-triphosphates. This study witnessed the emergence of structural diversification of the ~200-nt-long polymerase ribozyme into a catalytic module and a module specialized for binding of the primer-template substrate. While the catalytic module in complex with the substrate binding module showed better overall activity than the catalytic module alone, each of the two modules folded independently. These two studies were conceptually different from this study because they started with highly evolved ribozymes: The *Azoarcus* ribozyme evolved in nature over several billion years (*57*), and the polymerase ribozyme was developed by a series of at least four consecutive in vitro selection experiments (*12*, *17*, *58*, *59*). This long evolutionary optimization established robust RNA folding pathways and stable tertiary structures that displayed some catalytic activity already before the selection. In contrast, the study presented here started de novo from randomized sequence without an evolutionary history and with a sequence length (25 randomized positions within a ~58-nt-long RNA) that was in most cases too short to independently fold into a catalytic structure. Therefore, this study generated the need for trans-acting cofactors to achieve a stable catalytic structure. Therefore, these three studies illustrate different forms of molecular cooperation to emerge between nucleic acids during evolution. While the two previous studies model cooperation in stages of RNA evolution that already contain highly evolved ribozymes, the study described here models the formation of catalytic nucleic acid complexes in what may have been the earliest forms of nucleic acid catalysis.

The selected oligonucleotide complexes were diverse regarding temperature preference, reaction kinetics, and DNA dependence. Therefore, different types of complexes may be selected under different prebiotic conditions. For example, complexes that require lower DNA concentration (R12) or require only one DNA cofactor (R140, R285, and R118) would carry an evolutionary benefit in an environment containing many different sequences because they remained active in the presence of the $(dN)_{20}$ DNA oligomers (Fig. 9B). In contrast, complexes that require more than one DNA (R34 and R1943; Figs. 5 and 6) may carry an evolutionary advantage in an environment enriched for specific DNA sequences because, based on the seven complexes studied here, they reacted with faster kinetics (fig. S13). It would be interesting to see whether this correlation holds true for ribozymes catalyzing other reactions than self-triphosphorylation.

The 18 different DNA cofactors recruited by the 45 RNAs (Fig. 4A and fig. S14) showed a nucleotide bias toward a high-G and low-C content (32% G, 27%A, 23% T, and 18% C), which was not present in the library of 256 DNAs. However, this nucleotide bias can be explained by several of the DNAs pairing to the 3′-constant region of the RNA sequences, which is rich in C and low in G (39% C, 28% T, 22% A, and 11% G). For example, RNAs from the clusters C13 (R140), C14 (R285), and C7 (R340) were active with the G-rich DNAs D221 (45% G) or D240 (30% G). The DNAs D198 (40% G), D88 (45% G), and D138 (30% G) mediated activity for the RNAs R12, R118, and R1943, all of them apparently base pairing to the RNA 3′-constant regions. The C-rich 3′-terminal region could also have been a preferred target for DNA recruitment because the increased strength of G:C base pairs over A:U base pairs reduces the

Downloaded from https://www.science.org at University of California San Diego on April 30, 2025

number of bases that need to be complementary between DNA and RNA, thereby increasing the frequency of DNAs that are able to base pair with the RNA. Future experiments may show whether a nucleotide bias in the oligonucleotide libraries is able to increase the abundance of catalytic complexes in combinatorial mixtures.

The unexpected result that a completely randomized $(dN)_{20}$ library was able to support the catalytic function for four of the six biochemically characterized RNA/DNA complexes showed a large tolerance of these RNAs for the sequence of the recruited DNA sequences. Given that each of the $(dN)_{20}$ sequences existed at around 2×10^{-17} M (25.6 μM/4^{20} ~2×10^{-17} M) and that each of the reacted RNAs had a concentration of ~10 nM (~4% reacted of 250 nM RNA), the concentration of reacted RNA was about 4×10^8–fold higher than that of any specific DNA sequence. To explain how 10 nM RNA could have formed an active complex with this $(dN)_{20}$ mixture, we suggest two factors, sequence tolerance and possibly some DNA cofactor turnover. Because the self-triphosphorylation is a single-turnover reaction, a reacted RNA is free to release the DNA after catalysis so that other RNAs would be able to benefit from the same DNA molecule. While this turnover may have played a role, it is unlikely that a given DNA molecule mediated the activity of ~4×10^8 separate RNA molecules. Instead, the major reason for the observed activity with $(dN)_{20}$ must be a high mutation tolerance in the recruited DNA sequences: For each position in the DNA where the RNA molecule can tolerate the three other nucleotides, the frequency of beneficial DNAs increases by fourfold. If an RNA molecule could tolerate 14 mutated positions in its DNA cofactor, then the concentration of suitable DNAs would rise by the required factor (4^{15} ~3×10^8). We speculate that a combination of both effects, DNA turnover and DNA sequence tolerance, allowed four of the six characterized DNA-dependent RNAs to be active in the presence of the $(dN)_{20}$ library and that mutation tolerance was the dominant factor. Several RNAs showed tolerance to mutations in their DNA, for example, R140 was active with D221, D236, and D240 (fig. S7). In contrast, RNAs requiring more than one DNA (R34 and R1943) were not active with $(dN)_{20}$. The reason could lie in the constraints of a multimeric assembly: If a given DNA needs to make multiple sequence contacts (with the RNA and/or another DNA), then the DNA sequence is more constrained and may not tolerate as many mutations as when a single DNA is recruited. Similar rules may have guided the formation of the first catalysts from mixtures of short oligonucleotides in a complex prebiotic environment.

The results of the present study could have been very different under different selection conditions. The present study analyzed the effects of a 25.6 μM 20-nt oligomer DNA secondary library with 256 different sequences on a primary RNA library with 25-nt randomized region that was challenged to catalyze self-triphosphorylation and found that most of the selected RNA sequences (~80%) required DNA cofactors for activity (fig. S4). It is unclear how this fraction changes with shorter or longer oligonucleotides in the secondary library, at higher or lower concentrations of the secondary library, or with shorter randomized regions in the RNA library. A previous selection for self-aminoacylation ribozymes (*60*) found six different RNA sequence clusters from a library with 21 randomized nucleotides, showing that, for certain catalytic activities, shorter randomized regions are sufficient to mediate catalysis without a secondary library. It would be interesting whether many more active RNA/DNA complexes could be found within this library for the same self-aminoacylation activity if the selection was conducted in the presence of the DNA 20-nt oligomer library used here. Another constraint that defines the search for catalytic activity at the "lower end" of complexity is the limit of catalytic rate enhancement: A total size of only 24 nt was sufficient for self-aminoacylating ribozymes to achieve a 100-fold catalytic rate enhancement (*61*). This illustrates that an exploration of low catalytic rate enhancements, and a correspondingly low information content (*62*) may be a promising route to answer how the earliest catalysts could have emerged in a prebiotic environment (*63*).

MATERIALS AND METHODS

Generating the DNA oligonucleotide library

The secondary library of 256 DNA 20-nt oligomers was generated with the randomization function in the excel software. To identify a set of 256 random sequences that showed a distribution in the frequency of A, G, C, and T close to that of an ideal random distribution (0.8 of the 256 sequences without a given nucleotide, 5.3 with one occurrence of that nucleotide, 17 with 2 nt, 35 with 3 nt, 48 with 4 nt, 52 with 5 nt, 43 with 6 nt, 29 with 7 nt, 16 with 8 nt, 6.9 with 9 nt, 2.5 with 10 nt, 0.8 with 11 nt, 0.2 with 12 nt, and none with 13 nt or more), a total of 100,000 sequences of 20-nt oligomers were scanned using a sliding window of 256 sequences. As expected, the occurrence in the 256 randomized sequences was significantly higher or lower in a few instances, and these differences between ideal and random sets were reduced by rejecting one or more sequences that caused the difference and replacing it with a different random sequence that was not part of the initial 256. This process was repeated about 10 times. The final set of 256 sequences was chosen by calculating a factor F for all sequences, which measured the deviation from ideal random behavior (Fig. 1B). The equation was $F = (x - x_0)^2 * \mathrm{ABS}(x - x_0)/x_0$, where x is the occurrence of a given nucleotide in the random sequence of 20-nt oligomers, x_0 is the ideal value, and ABS is the absolute value, and included all 4 nt. The quadratic portion reduced occurrences of large differences between the ideal values and the random values, while the non-quadratic portion reduced occurrences of the significant absolute deviations from small values, respectively. The final factor F was 48 for the chosen set of 256 sequences. All DNA oligonucleotides in this study were purchased from Integrated DNA Technologies (IDT). The DNA library was generated by mixing equal volumes of a 0.1 mM solution of each of the 256 DNAs. The sequences are listed in table S1.

In vitro selection

The selection was carried out essentially as described, with modifications for the presence of the DNA 20-nt oligomer library during the self-triphosphorylation incubation (Fig. 1A). The DNA library was generated by PCR from the DNA oligomer 5′-AATTTAATACGACTCACTATA*GGGCGGTCTCCTGACGAGCTAAGCGAAACTGCGGAAACGCAGTC*GAGACCGGTAGATTC-N_{25}-CTCCAGATCATGCTCACT-3′ containing the promoter for T7 RNA polymerase (underlined), a hammerhead ribozyme (italics), and 25 randomized positions (N_{25}) flanked by constant regions for PCR amplification. The DNA oligomer was amplified with 5′-PCR primer AATTTAATACGACTCACTATAG and 3′-PCR primer 5′-AGTGAGCATGATCTGGAG-3′ in nine PCR cycles using Taq polymerase. A total of 732 pmol of template was used for this PCR, corresponding to a complexity of 4.4×10^{14}. A total of 2.52 nmol of PCR product were used for transcription of the RNA, thereby representing 97% of the

Downloaded from https://www.science.org at University of California San Diego on April 30, 2025

complexity or 4.3×10^{14}. The RNA library was generated by in vitro transcription using T7 RNA polymerase and purified by denaturing 7 M urea 15% PAGE.

To allow the self-triphosphorylation of catalytically active RNAs, a concentration of 100 nM RNA library was incubated with 25.6 μM DNA, 50 mM Na_3cTmp, 100 mM $MgCl_2$, and 50 mM tris-HCl (pH 8.5). The reaction was incubated for 8 hours with temperature cycling as shown in fig. S1, at temperatures no more than 40°C because RNA degradation was detected when the cycling temperature reached 50°C for 8 hours. The RNA was recovered from this reaction by ethanol precipitation. The ethanol pellet contained large amounts of salt, which was reduced by elution with small aliquots of water. The absorbance at 260 nm of each fraction was measured, and only fractions with detectable concentration of nucleic acids were collected. The initial fractions did not contain RNA. In the first four rounds of selection, small amounts (1/20,000 of library RNA) of 5′-triphosphorylated library RNAs were added to these collected fractions. These triphosphorylated RNAs reduced the number of PCR cycles and avoided PCR artifacts. The fractions were desalted by size exclusion chromatography (Bio-Spin columns P-30, Bio-Rad) and concentrated by ethanol precipitation. The DNA in the mixture was digested by deoxyribonuclease RQ1 (Promega) and then removed by denaturing 7 M urea and 15% PAGE purification and ethanol precipitation of the RNA. A total of 400 nM purified RNA was mixed with 500 nM R3C ligase ribozyme, 600 nM biotinylated oligonucleotide (5′-biotin-d(GTAGAG-ACGTAA)rU-3′), 80 mM KCl, and 100 mM tris-HCl (pH 8.5); heated for 10 min at 50°C; and cooled down to 30°C before adding an equal volume of 2× reaction buffer containing 40% PEG-8000 (polyethylene glycol, molecular weight 8000), 100 mM $MgCl_2$, and 5 mM spermidine. The ligation was incubated at 30°C for 3 hours and quenched by adding 3 M equivalents of Na_2EDTA over Mg^{2+}.

The ligated RNAs were prepared for capture by adjusting the buffer concentration to 50 mM tris-HCl (pH 8.5), 50 mM KCl, and 0.01% Triton X-100 and adding a final concentration of 200 nM DNA that was complementary to the ligase ribozyme to free non-ligated library RNAs from the biotinylated oligonucleotide. The capture mixture was incubated at 50°C for 10 min and slowly cooled to room temperature. The magnetic streptavidin-coated beads (Promega, Z5481) were washed three times with buffer containing 20 mM Hepes/KOH (pH 7.2), 0.01% Triton X-100, and 50 mM KCl. Each 25 pmol of RNA library was incubated with 100 μl of beads by rotating for 30 to 60 min at room temperature. The beads were then washed twice with a buffer containing 0.01% Triton X-100 and 20 mM NaOH. The bound RNA was eluted in 25 mM tris-HCl, 1 mM Na_2EDTA, and 95% formamide at 65°C for 3 min and ethanol precipitated from the supernatant.

The extracted library RNA was reverse transcribed by Super-Script III (Thermo Fisher Scientific) at 55°C for 1 hour and denatured at 70°C for 15 min. The reverse transcription product was purified by ethanol precipitation and subjected to PCR. Ten percent of the RT product was used in pilot PCR reactions to determine the number of PCR cycles to amplify to a clear but not oversaturated band on ethidium bromide stained agarose gels. The PCR cycle number that generated visible products but had not reached saturation was used for the rest of the material. The recovery rate of RNA starting from the cTmp incubation to the PCR was about 11% of an initial 5.0 nmol (550 pmol), thereby reducing the effective complexity from 4.3×10^{14} to 2.4×10^{14} in the first round of the selection. The PCR products were extended with 12 more PCR cycles to add the sequences of T7 promoter and hammerhead ribozyme as shown in the sequence of the initial library above, completing one round of selection. This library was used as input for the next round of in vitro selection.

Selection round 6 was performed in 18 sub-rounds. One sub-round was performed in the absence of DNA oligonucleotides, one sub-round in the presence of all DNA oligonucleotides, and 16 sub-rounds with each one set of 16 DNA oligonucleotides (A to P), thereby covering all 256 DNAs used in the DNA library.

HTS analysis

After attaching Illumina adapter and barcode primers to the selected libraries, HTS was done on an Illumina NovaSeq S4 platform, generating a total of 300 million reads for all selection rounds. From the reads of selection round 6, a total of the 1331 sequences existed in at least 500 copies each in the sixth round (all 18 sub-rounds). For sequence enrichment between selection rounds, the sequence abundance was calculated by the number of reads divided by the total reads of a given round.

To analyze the effects of DNA sets on the activity of individual sequences (Fig. 2B), the number of sequences with DNAs in a given DNA set (or in the set with all DNAs) was divided by the number of sequences in the sub-round without DNA. To reduce statistical artifacts, the RNAs were designated into four groups based on the number of reads when the selection step was performed without DNA: 780 RNAs with more than 50 reads (group a), 519 RNAs with 6 to 50 reads (group b), 30 RNAs with 1 to 5 reads (group c), and two RNAs with no reads (group d). From group a, the top two enriched RNAs were chosen for each of the 16 DNA sets, leading to 21 unique RNAs. From group b, the top two enriched RNAs were also chosen for each of the 16 DNA sets, leading to 18 unique RNAs. Both numbers are less than 32 because several RNAs appeared in multiple groups. As comparison, six RNAs with the lowest enrichments among all DNA sets (sometimes, zero reads with DNAs) were also chosen from groups a and b; note that all of these RNAs were enriched by other DNA sets between 1.1-fold and 30-fold and, therefore, displayed discrimination between DNAs. From group c (1 to 5 reads without DNA), 23 RNAs that had more than a total sum of 100 reads in the selection sub-round with 256 DNAs was chosen. From group d (no reads without DNA), both RNAs in this category were chosen. These RNAs from HTS analysis, together with three most active RNAs identified from cloning of the selected pool, resulted in 79 RNAs and were analyzed by clustering. The 79 sequences were aligned using Clustal X. To reduce the redundancy of very similar RNAs, sequences were removed mostly from clusters 13 and 14, with 11 of the 27 RNAs chosen from cluster 13 and 6 of 16 from cluster 14 (Fig. 3C). This resulted in 45 RNA sequences that were used for biochemical testing.

Self-triphosphorylation activity assays

The 45 chosen sequences were purchased as single-stranded DNA from IDT, PCR amplified with primers pairing to the constant regions, in vitro transcribed, and purified by denaturing 7 M urea 15% PAGE. Only the RNAs for the seven complexes characterized in more detail (table S2) the DNA templates were purchased in the form of gBlock fragments (IDT) containing T7 promoter, hammerhead ribozyme, RNA sequence, and a 3′-terminal HDV ribozyme to generate a precise 3′ end (*59*). The appropriate DNA sub-libraries were prepared by mixing the corresponding DNA oligonucleotides in equimolar ratios (table S1). The triphosphorylation reaction was

Downloaded from https://www.science.org at University of California San Diego on April 30, 2025

set up with 250 nM RNA, each DNA at 250 nM concentration (up to 256 DNAs), 50 mM Na_3cTmp, 100 mM $MgCl_2$, and tris-HCl (pH 8.5) with the same temperature conditions as in the selection (fig. S1), or as described for seven characterized complexes. The reaction was quenched by adding a 1.2-fold molar excess of Na_2EDTA. The mixture was directly added to annealing buffer containing 100 nM ligase ribozyme, 100 nM 5′-[^{32}P]–radiolabeled oligonucleotide, 80 mM KCl, and 100 mM tris-HCl (pH 8.5). After annealing at 50°C for 10 min, the mixture was cooled to 30°C and mixed with the same volume of buffer containing 40% PEG-8000, 100 mM $MgCl_2$, and 5 mM spermidine. The reaction was incubated for 3 hours at 30°C, and the reaction products were ethanol precipitated and separated by denaturing 7 M urea 10% PAGE. The gels were exposed to phosphorimager screens and scanned by a Typhoon phosphorimager (GE). The activity was quantified by the Quantity One software (Bio-Rad) and calculated as the ratio for the signal intensity of the ligated band divided by the sum of the signal intensities for ligated and unligated bands. The three batches of randomized DNA 20-nt oligomers $(dN)_{20}$ were purchased on three different days from IDT with machine-mixing for the randomized region. The elongations or truncations used in the combinatorial assay are given in table S2.

Clustering analysis

The phylogenetic tree of the 45 RNAs (Fig. 4A) was generated in Clustal X by the neighbor-joining method, and the plot was generated with R. The 14 clusters were assigned on the basis of the use of DNA cofactors. The network plot was generated with the software Cytoscape (Fig. 4B). The size of each node was assigned to six scales according to the abundance of the RNA in HTS analysis. The edit distances between any two sequences shown in the plot were given in the figure if the distance was less than 13. One distance of 13 (cluster 9 to cluster 4) was shown because the closest distance of cluster 9 to other clusters was 13.

In-line probing

A total of 250 nM RNA, traces of the same 5′-[^{32}P]–radiolabeled RNA, and the optimum concentration of DNAs, 50 mM Na_3cTmp, 100 mM $MgCl_2$, and 50 mM tris-HCl (pH 8.5) were incubated at the optimum temperature for 24 hours. The reaction was quenched by adding an 1.2-fold molar excess of Na_2EDTA. Samples were separated by denaturing 7 M urea 10% PAGE. Each band's intensity was quantified by Quantity One (Bio-Rad). The intensity of each band was determined by calculating the percentage of each band in the same lane and normalized by subtracting the corresponding position of the untreated RNA. The threshold between single-stranded and double-stranded regions was determined based on the predicted structure by the software mFold and RNAstructure. The nucleotide identify of bands was determined using ribonuclease T1 (Thermo Fisher Scientific) and an alkaline hydrolysis ladder. A sample of untreated RNA was used for background subtraction. The final structures were modified on the basis of the predicted structures or redrawn based on the in-line probing data.

Supplementary Materials

This PDF file includes:

Figs. S1 to S14
Tables S1 to S3

REFERENCES AND NOTES

1. A. Rich, "On the problems of evolution and biochemical information transfer" in *Horizons in Biochemistry* (Academic Press, 1962), pp. 103–126.
2. C. R. Woese, The fundamental nature of the genetic code: Prebiotic interactions between polynucleotides and polyamino acids or their derivatives. *Proc. Natl. Acad. Sci. U.S.A.* **59**, 110–117 (1968).
3. L. E. Orgel, Evolution of the genetic apparatus. *J. Mol. Biol.* **38**, 381–393 (1968).
4. F. H. C. Crick, The origin of the genetic code. *J. Mol. Biol.* **38**, 367–379 (1968).
5. M. Neveu, H.-J. Kim, S. A. Benner, The "strong" RNA world hypothesis: Fifty years old. *Astrobiology* **13**, 391–403 (2013).
6. H. F. Noller, V. Hoffarth, L. Zimniak, Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* **256**, 1416–1419 (1992).
7. P. Nissen, J. Hansen, N. Ban, P. B. Moore, T. A. Steitz, The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).
8. X. Chen, N. Li, A. D. Ellington, Ribozyme catalysis of metabolism in the RNA world. *Chem. Biodivers.* **4**, 633–655 (2007).
9. H. B. White III, Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol.* **7**, 101–104 (1976).
10. A. J. Crapitto, A. Campbell, A. Harris, A. D. Goldman, A consensus view of the proteome of the last universal common ancestor. *Ecol. Evol.* **12**, e8930 (2022).
11. R. Krishnamurthy, Role of pK(a) of nucleobases in the origins of chemical evolution. *Acc. Chem. Res.* **45**, 2035–2044 (2012).
12. W. K. Johnston, P. J. Unrau, M. S. Lawrence, M. E. Glasner, D. P. Bartel, RNA-catalyzed RNA polymerization: Accurate and general RNA-templated primer extension. *Science* **292**, 1319–1325 (2001).
13. A. Akoopie, J. T. Arriola, D. Magde, U. F. Müller, A GTP-synthesizing ribozyme selected by metabolic coupling to an RNA polymerase ribozyme. *Sci. Adv.* **7**, eabj7487 (2021).
14. G. Von Kiedrowski, Ein selbstreplizierendes hexanukleotid. *Angew. Chem.* **98**, 932–934 (1986).
15. H. Park, S. Parshotam, S. C. Hales, A. K. Mittermaier, J. M. Gibbs, Minimizing product inhibition in DNA self-replication: Insights for prebiotic replication from the role of the enzyme in lesion-induced DNA amplification. *Chemistry* **29**, e202300080 (2023).
16. P. C. Sabeti, P. J. Unrau, D. P. Bartel, Accessing rare activities from random RNA sequences: The importance of the length of molecules in the starting pool. *Chem. Biol.* **4**, 767–774 (1997).
17. D. P. Bartel, J. W. Szostak, Isolation of new ribozymes from a large pool of random sequences. *Science* **261**, 1411–1418 (1993).
18. B. Zhang, T. R. Cech, Peptide bond formation by in vitro selected ribozymes. *Nature* **390**, 96–100 (1997).
19. P. J. Unrau, D. P. Bartel, RNA-catalysed nucleotide synthesis. *Nature* **395**, 260–263 (1998).
20. B. Seelig, A. Jaschke, A small catalytic RNA motif with Diels-Alderase activity. *Chem. Biol.* **6**, 167–176 (1999).
21. Y. Ryu, K.-J. Kim, C. A. Roessner, A. I. Scott, Decarboxylative Claisen condensation catalyzed by in vitro selected ribozymes. *Chem. Commun. (Camb.)* , 1439–1441 (2006).
22. J. E. Moretti, U. F. Muller, A ribozyme that triphosphorylates RNA 5′-hydroxyl groups. *Nucleic Acids Res.* **42**, 4767–4778 (2014).
23. J. P. Ferris, A. R. Hill, R. Liu, L. E. Orgel, Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* **381**, 59–61 (1996).
24. W. Huang, J. P. Ferris, Synthesis of 35–40 mers of RNA oligomers from unblocked monomers. A simple approach to the RNA world. *Chem. Commun. (Camb)* **9**, 1458–1459 (2003).
25. P. A. Monnard, A. Kanavarioti, D. W. Deamer, Eutectic phase polymerization of activated ribonucleotide mixtures yields quasi-equimolar incorporation of purine and pyrimidine nucleobases. *J. Am. Chem. Soc.* **125**, 13734–13740 (2003).
26. L. Li, N. Prywes, C. P. Tam, D. K. O'Flaherty, V. S. Lelyveld, E. C. Izgu, A. Pal, J. W. Szostak, Enhanced nonenzymatic RNA copying with 2-aminoimidazole activated nucleotides. *J. Am. Chem. Soc.* **139**, 1810–1813 (2017).
27. C. F. Dirscherl, A. Ianeselli, D. Tetiker, T. Matreux, R. M. Queener, C. B. Mast, D. Braun, A heated rock crack captures and polymerizes primordial DNA and RNA. *Phys. Chem. Chem. Phys.* **25**, 3375–3386 (2023).
28. S. Rajamani, A. Vlassov, S. Benner, A. Coombs, F. Olasagasti, D. Deamer, Lipid-assisted synthesis of RNA-like polymers from mononucleotides. *Orig. Life Evol. Biosph.* **38**, 57–74 (2008).
29. C. A. Jerome, H.-J. Kim, S. J. Mojzsis, S. A. Benner, E. Biondi, Catalytic synthesis of polyribonucleic acid on prebiotic rock glasses. *Astrobiology* **22**, 629–636 (2022).
30. C. V. Mungi, S. Rajamani, Characterization of RNA-like oligomers from lipid-assisted nonenzymatic synthesis: Implications for origin of informational molecules on early Earth. *Life (Basel)* **5**, 65–84 (2015).
31. J. A. Doudna, S. Couture, J. W. Szostak, A multisubunit ribozyme that is a catalyst of and template for complementary strand RNA synthesis. *Science* **251**, 1605–1608 (1991).
32. J. A. Doudna, T. R. Cech, Self-assembly of a group I intron active site from its component tertiary structural domains. *RNA* **1**, 36–45 (1995).
33. K. Adamala, A. E. Engelhart, J. W. Szostak, Generation of functional RNAs from inactive oligonucleotide complexes by non-enzymatic primer extension. *J. Am. Chem. Soc.* **137**, 483–489 (2015).

Downloaded from https://www.science.org at University of California San Diego on April 30, 2025

34. H. Mutschler, A. Wochner, P. Holliger, Freeze-thaw cycles as drivers of complex ribozyme assembly. *Nat. Chem.* **7**, 502–508 (2015).
35. A. Akoopie, U. F. Muller, Lower temperature optimum of a smaller, fragmented triphosphorylation ribozyme. *Phys. Chem. Chem. Phys.* **18**, 20118–20125 (2016).
36. E. J. Hayden, N. Lehman, Self-assembly of a group I intron from inactive oligonucleotide fragments. *Chem. Biol.* **13**, 909–918 (2006).
37. T. A. Lincoln, G. F. Joyce, Self-sustained replication of an RNA enzyme. *Science* **323**, 1229–1232 (2009).
38. F. Chizzolini, A. D. Kent, L. F. M. Passalacqua, A. Lupták, Enzymatic RNA production from NTPs synthesized from nucleosides and trimetaphosphate. *Chembiochem* **22**, 2098–2101 (2021).
39. Y. Yamagata, H. Watanabe, M. Saitoh, T. Namba, Volcanic production of polyphosphates and its relevance to prebiotic evolution. *Nature* **352**, 516–519 (1991).
40. M. A. Pasek, D. S. Lauretta, Aqueous corrosion of phosphide minerals from iron meteorites: A highly reactive source of prebiotic phosphorus on the surface of the early Earth. *Astrobiology* **5**, 515–535 (2005).
41. M. Pasek, K. Block, Lightning-induced reduction of phosphorus oxidation state. *Nat. Geosci.* **2**, 553–556 (2009).
42. B. Herschy, S. J. Chang, R. Blake, A. Lepland, H. Abbott-Lyon, J. Sampson, Z. Atlas, T. P. Kee, M. A. Pasek, Archean phosphorus liberation induced by iron redox geochemistry. *Nat. Commun.* **9**, 1346 (2018).
43. E. Etaix, L. E. Orgel, Phosphorylation of nucleosides in aqueous solution using trimetaphosphate: Formation of nucleoside triphosphates. *J. Carbohydr. Nucleosides Nucleotides* **5**, 91–110 (1978).
44. K. Lohmann, Über die pyrophosphatfraktion im muskel. *Naturwissenschaften* **17**, 624–625 (1929).
45. G. F. Dolan, A. Akoopie, U. F. Muller, A faster triphosphorylation ribozyme. *PLOS ONE* **10**, e0142559 (2015).
46. A. Pressman, J. E. Moretti, G. W. Campbell, U. F. Muller, I. A. Chen, Analysis of in vitro evolution reveals the underlying distribution of catalytic activity among random sequences. *Nucleic Acids Res.* **45**, 8167–8179 (2017).
47. J. T. Arriola, U. F. Müller, A combinatorial method to isolate short ribozymes from complex ribozyme libraries. *Nucleic Acids Res.* **48**, –e116 (2020).
48. K. J. Sweeney, X. Han, U. F. Müller, A ribozyme that uses lanthanides as cofactor. *Nucleic Acids Res.* **51**, 7163–7173 (2023).
49. K. J. Sweeney, T. Le, M. Z. Jorge, J. G. Schellinger, L. J. Leman, U. F. Müller, Peptide conjugates with polyaromatic hydrocarbons can benefit the activity of catalytic RNAs. *Chem. Sci.* **14**, 10318–10328 (2023).
50. J. T. Arriola, S. Poordian, E. M. Valdivia, T. Le, L. J. Leman, J. G. Schellinger, U. F. Müller, Weak effects of prebiotically plausible peptides on self-triphosphorylation ribozyme function. *RSC Chem. Biol.* **5**, 1122–1131 (2024).
51. M. W. Powner, S. L. Zheng, J. W. Szostak, Multicomponent assembly of proposed DNA precursors in water. *J. Am. Chem. Soc.* **134**, 13889–13895 (2012).
52. D. J. Ritson, J. D. Sutherland, Conversion of biosynthetic precursors of RNA to those of DNA by photoredox chemistry. *J. Mol. Evol.* **78**, 245–250 (2014).
53. J. Rogers, G. F. Joyce, The effect of cytidine on the structure and function of an RNA ligase ribozyme. *RNA* **7**, 395–404 (2001).
54. E. A. Schultes, D. P. Bartel, One sequence, two ribozymes: Implications for the emergence of new ribozyme folds. *Science* **289**, 448–452 (2000).
55. N. Vaidya, M. L. Manapat, I. A. Chen, R. Xulvi-Brunet, E. J. Hayden, N. Lehman, Spontaneous network formation among cooperative RNA replicators. *Nature* **491**, 72–77 (2012).
56. J. Attwater, A. Raguram, A. S. Morgunov, E. Gianni, P. Holliger, Ribozyme-catalysed RNA synthesis using triplet building blocks. *eLife* **7**, e35255 (2018).
57. S. O. Rogers, Integrated evolution of ribosomal RNAs, introns, and intron nurseries. *Genetica* **147**, 103–119 (2019).
58. E. H. Ekland, D. P. Bartel, The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res.* **23**, 3231–3238 (1995).
59. A. Wochner, J. Attwater, A. Coulson, P. Holliger, Ribozyme-catalyzed transcription of an active ribozyme. *Science* **332**, 209–212 (2011).
60. A. D. Pressman, Z. Liu, E. Janzen, C. Blanco, U. F. Müller, G. F. Joyce, R. Pascal, I. A. Chen, Mapping a systematic ribozyme fitness landscape reveals a frustrated evolutionary network for self-aminoacylating RNA. *J. Am. Chem. Soc.* **141**, 6213–6223 (2019).
61. N. V. Chumachenko, Y. Novikov, M. Yarus, Rapid and simple ribozymic aminoacylation using three conserved nucleotides. *J. Am. Chem. Soc.* **131**, 5257–5263 (2009).
62. J. M. Carothers, S. C. Oestreich, J. H. Davis, J. W. Szostak, Informational complexity and functional activity of RNA structures. *J. Am. Chem. Soc.* **126**, 5130–5137 (2004).
63. R. Knight, M. Yarus, Finding specific RNA motifs: Function in a zeptomole world? *RNA* **9**, 218–230 (2003).

Acknowledgments: We thank A. Akoopie for the design of the DNA cofactor library. We thank K. J. Sweeney and J. T. Arriola for help on HTS analysis. **Funding:** This work was supported by NASA Exobiology grant 80NSSC19K0467 to U.F.M. **Author contributions:** Conceptualization: U.F.M. Methodology: U.F.M. and X.H. Investigation: X.H. Visualization: X.H. and U.F.M. Funding acquisition: U.F.M. Project administration: U.F.M. Supervision: U.F.M. Writing—original draft: U.F.M. and X.H. Writing—review and editing: U.F.M. and X.H. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 30 October 2024
Accepted 25 March 2025
Published 30 April 2025
10.1126/sciadv.adu2647

Downloaded from https://www.science.org at University of California San Diego on April 30, 2025