RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension

Wendy K. Johnston, Peter J. Unrau,* Michael S. Lawrence, Margaret E. Glasner, David P. Bartel†

The RNA world hypothesis regarding the early evolution of life relies on the premise that some RNA sequences can catalyze RNA replication. In support of this conjecture, we describe here an RNA molecule that catalyzes the type of polymerization needed for RNA replication. The ribozyme uses nucleoside triphosphates and the coding information of an RNA template to extend an RNA primer by the successive addition of up to 14 nucleotides—more than a complete turn of an RNA helix. Its polymerization activity is general in terms of the sequence and the length of the primer and template RNAs, provided that the 3′ terminus of the primer pairs with the template. Its polymerization is also quite accurate: when primers extended by 11 nucleotides were cloned and sequenced, 1088 of 1100 sequenced nucleotides matched the template.

The RNA world hypothesis states that early life forms lacked protein enzymes and depended instead on enzymes composed of RNA (*1*). Much of the appeal of this hypothesis comes from the realization that ribozymes would have been far easier to duplicate than proteinaceous enzymes (*2–5*). Whereas coded protein replication requires numerous macromolecular components [including mRNA, transfer RNAs (tRNAs), aminoacyl-tRNA synthetases, and the ribosome], replication of a ribozyme requires only a single macromolecular activity: an RNA-dependent RNA polymerase that synthesizes first a complement, and then a copy of the ribozyme. If this RNA polymerase were itself a ribozyme, then a simple ensemble of molecules might be capable of self-replication and eventually, in the course of evolution, give rise to the protein-nucleic acid world of contemporary biology. Finding a ribozyme that can efficiently catalyze general RNA polymerization would support the idea of the RNA world (*1, 6, 7*) and would provide a key component for the laboratory synthesis of minimal life forms based on RNA (*8, 9*).

Although progress has been made in finding ribozymes capable of template-directed RNA synthesis, none of these ribozymes has the sophisticated substrate-binding properties needed for general polymerization (*7*). Derivatives of self-splicing introns are able to

Whitehead Institute for Biomedical Research, and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA.

*Present address: Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada.
†To whom correspondence should be addressed. E-mail: dbartel@wi.mit.edu

join oligonucleotides assembled on a template (*10–12*). However, the templates that can be copied are limited to those that match the oligonucleotide substrates, and it has not been possible to include sufficient concentrations of all the oligonucleotide substrates needed for a general copying reaction. More recently, efforts have shifted to derivatives of an RNA-ligase ribozyme that was isolated from a large pool of random RNA sequences (*13–15*). Some derivatives are capable of template-directed primer extension using nucleoside triphosphate (NTP) substrates, but their reaction is also limited to a small subset of possible template RNAs (*15*). These ligase derivatives recognize the primer-template complex by hybridizing to a particular unpaired segment of the template (Fig. 1A). In using a short segment of a special template to direct primer extension, these ribozymes resemble telomerases more than they resemble the enzymes that replicate RNA and DNA by means of general polymerization.

Polymerase isolation. We have used the catalytic core of the ligase ribozyme (*14, 16*) as a starting point for the generation of a ribozyme with general RNA polymerization activity. The new polymerase ribozyme was isolated from a pool of over 10^{15} different RNA sequences. Sequences in the starting pool contained a mutagenized version of the parental ligase (Fig. 1B). To sample a broad distribution of mutagenesis levels, the starting pool comprised four subpools in which the core residues of the ligase domain were mutagenized at levels averaging either 0, 3, 10, or 20% (*17*). Two loops within the ligase domain, both unimportant for ligase function, were replaced with 8-nucleotide (nt) random-

sequence segments (Fig. 1B). The 5′ terminus of the ligase domain was covalently attached to an RNA primer so that molecules able to catalyze primer extension could be selected by virtue of their attachment to the primer that they extended.

In contrast to the parental ribozyme, which hybridizes to the template, a ribozyme capable of general polymerization must recognize the primer-template complex without relying on sequence-specific interactions. Therefore, the template RNA was designed to be too short for extensive hybridization with the ribozyme (Fig. 1B). For the parental ribozyme (Fig. 1A), the pairing between the ribozyme and the template also comprised a stem known to be necessary for ligase function (*16*). To restore this stem, a short RNA, GGCACCA (purple RNA in Fig. 1B), was introduced to hybridize to the segment of the ligase domain that formerly paired with the template. Finally, because a more general mode of primer-template recognition might require the participation of an additional RNA domain, a 76-nt random-sequence segment was appended to the 3′ terminus of the degenerate ligase domain (Fig. 1B).

Sequence variants able to recognize the primer-template in this new configuration and then extend the primer with tagged nucleotides were enriched by repeated rounds of in vitro selection and amplification (Table 1). RNAs that extended their primer by using 4-thioUTP were isolated on APM gels (urea denaturing gels cast with a small amount of *N*-acryloylaminophenylmercuric acetate, which impedes migration of RNA containing 4-thioU) (*18*). To favor variants that recognize generic rather than sequence-specific features of the primer-template, different primer-template sequences and lengths were used in different rounds of selection (Table 1). To favor the more efficient ribozyme variants, the stringency of the selection was increased in later rounds by requiring addition of two tagged nucleotides, such as biotinylated A and 4-thioU (*19*), and by decreasing the time of incubation with the tagged NTPs (Table 1).

After eight rounds of selection and amplification, desirable variants had increased in abundance to the point that a detectable fraction of the pool molecules could extend their primer by using both 4-thioUTP and radiolabeled ATP in a template-dependent fashion (Fig. 2). Other variants able to tag themselves were detected as early as round four, but most of these ribozymes added tagged nucleotides in the absence of the template oligonucleotide, or they decorated themselves at sites other than the 3′ terminus of the primer (*20*). Seventy-four variants from rounds 8 through 10 were cloned and found to represent 23 sequence families, each family having descended from a different ancestral sequence of the starting pool. Ribozymes from two families extended their

primer by using both 4-thioUTP and radiolabeled ATP in a template-dependent fashion. These two families are represented by isolates 9.1 and 10.2 (Fig. 2). Isolate 10.2 (Fig. 1C), from round 10, was from the more prevalent and active of these two families and was chosen for further study.

Polymerization with multiple turnover. Having isolated a ribozyme that did not rely on forming base pairs with the template RNA during primer extension, we next determined whether it instead recognized the particular sequence used to link the primer to the ribozyme. We were pleased to discover that the round-10 ribozyme did not require this sequence. In fact, it did not require any covalent attachment to the primer. When incubated with a 10-fold excess of both a 6-nt primer and a 9-nt template, as well as the appropriate

nucleoside triphosphate, the round 10 polymerase fully extended the primer, with multiple turnover (Fig. 3). The primer and template sequences in this experiment were designed to differ from those used most frequently during the selection (Table 1, aligning the sequences relative to the primer 3′ termini). Despite the complete change in the sequence of the primer-template complex, the polymerase isolate was able to recognize the complex and extend the primer (21). This indicates that the ribozyme binds the primer-template without relying upon recognition of a particular sequence.

The new mode of primer-template recognition appears to be conferred by the accessory domain derived from the 76-nt random-sequence segment and the 3′-terminal segment that binds the primer used for the reverse tran-

scription–polymerase chain reaction (RT-PCR). Without the accessory domain, the ligase domain of the round-10 ribozyme, like the parental ligase itself (Fig. 3B), displayed no activity in polymerization assays requiring general primer-template recognition. Indeed, deleting only 9 nt from the 3′ terminus of the round-10 ribozyme severely diminished activity (20). It is interesting that the core ligase residues emerged unchanged in this round-10 isolate (compare Fig. 1, B and C), and the GGCACCA oligonucleotide designed to complete the ligase domain proved to be necessary for polymerase function (20). This suggests that the parental ligase did not need to adapt in order to accommodate the more general primer-template recognition afforded by the accessory domain.

The round-10 ribozyme was tested with numerous other primer-template pairs. In all cases

Fig. 1. Secondary structure models of the ribozymes. Short dashes indicate base pairs. (**A**) A ribozyme (black strand) able to promote limited RNA polymerization (15). It extends an RNA primer (orange strand) by using nucleoside triphosphates and coding information from an appropriate RNA template (red strand). The ribozyme can accommodate any of the four RNA nucleotides at residues indicated by an X, provided that the primer pairs with the template. However, the 5′ portion of the template must pair with the ribozyme. The depicted ribozyme was derived from an RNA ligase ribozyme; black uppercase residues and defined residues of the template comprise the core of the ligase ribozyme. (**B**) A pool of RNA sequences based on the ligase ribozyme (17). Colors differentiate residues representing the ligase core (black, purple), random sequence (blue), primer (orange), template (red), and RT-PCR primer-binding sites (green). Residues prefixed by "d" are DNA; all others are RNA.
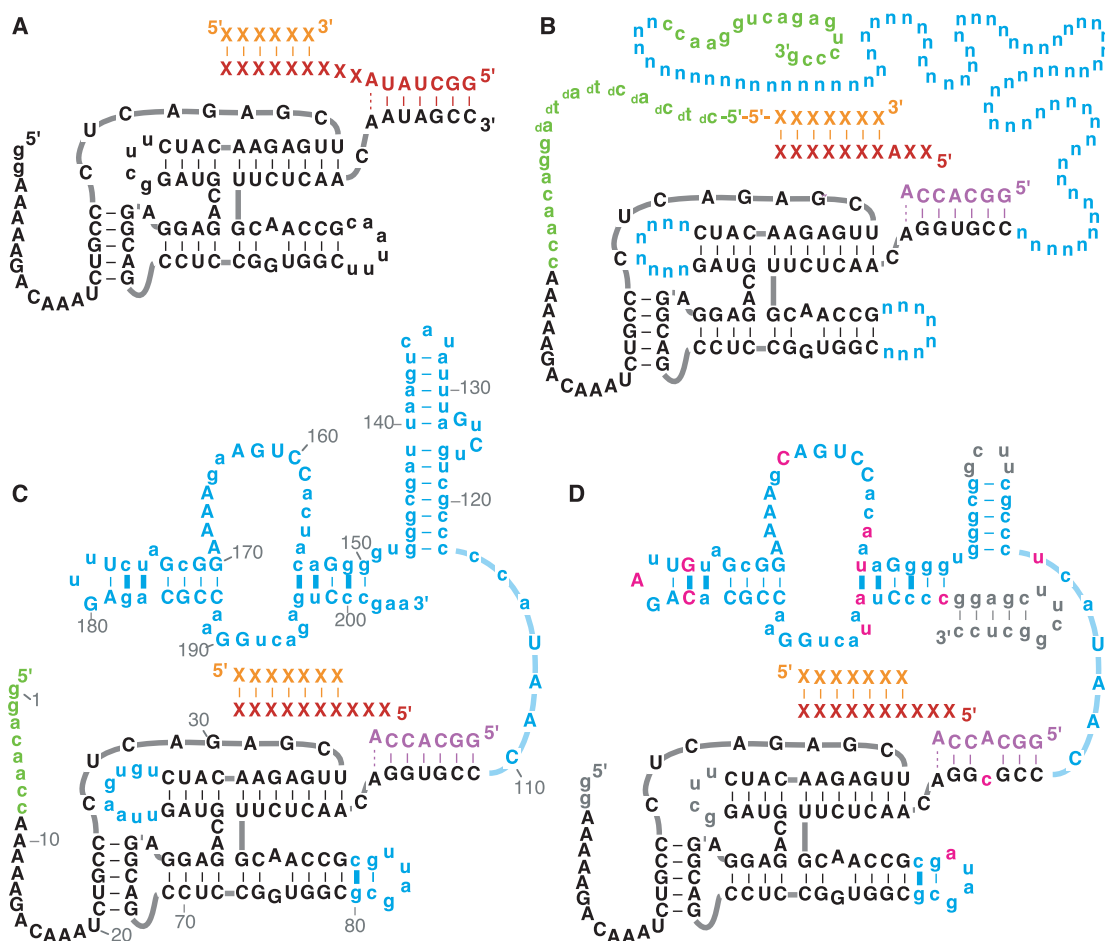
The 5′ end of the RNA primer is covalently joined to the 5′ end of each pool molecule via a phosphodiester linkage (-5′-5′-) (38). The sequence of the primer-template (X) in a given round usually differed from that of the previous round (Table 1). (**C**) The round-10 ribozyme (isolate 10.2). Residues derived from the random-sequence segments or the 3′ RT-PCR primer-binding site of the starting pool are colored blue; other drawing conventions are as in (B). Comparative sequence analysis of improved isolates from rounds 14 and 18 (23) supports the importance, as well as the proposed secondary structure, of the accessory domain (residues 110 to 204), particularly within the 3′ region of this domain (residues 150 to 201). Blue uppercase residues were invariant among all 22 improved isolates. Because the chance conservation of a residue not important for activity is low ($P = 0.0074$ for conservation in 22 of 22 isolates), nearly all 29 of these residues must be important for ribozyme function. Thick blue dashes mark covarying pairs, five of which (G151:C200, A153:U198, C154:G197, U175:A183, and C176:G182) support the proposed pairing within the 3′ region of the accessory domain. (**D**) The round-18 ribozyme, a shortened derivative of an improved isolate from round 18. Nucleotide changes from the round-10 isolate that arose from combinatorial mutagenesis are in pink; changes engineered when reducing the ribozyme's size are in gray (23). The four changes consistently found among the improved round-14 and round-18 isolates are in uppercase pink. Other drawing conventions are as in (C).

it was able to recognize the primer-template complex and to extend the primer by the Watson-Crick match to the template. Extension across from C or G template residues was usually more efficient and accurate than extension across from A or U. Extension was also much more efficient when the unpaired portion of the template was shorter than 5 nt (Fig. 4B).
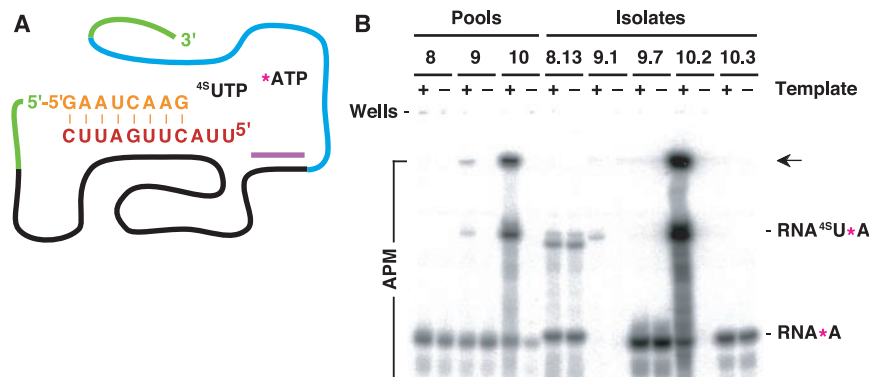
Sequence optimization of the polymerase. To find improved polymerase variants, the in vitro selection procedure was continued for another eight rounds of selection and amplification, starting with a newly synthesized pool of variants based on the round-10 isolate (22). In this pool, the accessory domain, as well as the two 8-nt segments at the

Fig. 2. Detection of ribozymes able to extend an attached primer by two nucleotides in a template-dependent manner. (**A**) Schematic of the RNAs in this experiment. Ribozymes attached to an RNA primer (orange) were incubated with 1 mM 4-thioUTP (^{4S}UTP) and trace [α-^{32}P]ATP (*ATP), in the presence or absence of an RNA template (red) that codes for the addition of U and A. (**B**) Activities of ribozyme pools and isolates after 8 to 10 rounds of selection. Extension reactions were for 12 hours, under the conditions used during the rounds of selection (18). Shown is a PhosphorImager scan of an APM denaturing gel separating RNAs extended with radiolabeled A (RNA*A) from those extended with both 4-thioU and radiolabeled A (RNA^{4S}U*A). The arrow points to RNA^{4S}U*A extended by a second 4-thioU, which did not migrate into the APM portion of the gel. Note that addition of the second ^{4S}U could not have been directed by an A in the template because only one of the template coding residues is an A; some misincorporation of ^{4S}U was expected in this experiment because of the very large excess of ^{4S}UTP over ATP. The sequence families represented by 9.1 and 10.2 add both 4-thioU and radiolabeled A in a template-dependent manner. The round-10 isolate (10.2) was chosen for further analysis and is shown in Fig. 1C.

Table 1. Parameters and substrates for in vitro selection. For each round of selection, pool RNA with covalently attached primer (38) was incubated with the indicated template RNA and NTPs for the indicated time (18). Nucleotide analogs 4-thioU, biotin-N6-A, and 2-aminopurine (39) are abbreviated ^{4S}U, ^{B}A, and ^{2N}P, respectively. The primer attached to the pool molecules was complementary to the underlined segment of the template. Variants with polymerase activity were selected based on their primer being extended with the tagged nucleotides indicated in the Selection criteria column (18). In late rounds, 2 mM ATP, 2 mM CTP, and 2 mM GTP were included as competitor NTPs (Comp. NTPs). Pool mutagenesis was either during chemical synthesis (Synthesis) or during error-prone amplification (PCR) of the template DNA (17, 22, 40).

Round	Mutagenesis	Template RNA	NTPs	Time (hour)	Selection criteria
1	Synthesis	3′-<u>GGUCAGAUU</u>	^{4S}UTP (2 mM)	36	^{4S}U
2	None	3′-<u>GGUCAGAACC</u>	^{4S}UTP (2 mM)	20	^{4S}U
3	None	3′-<u>GGUCAGAA</u>	^{4S}UTP (2 mM)	20	^{4S}U
4	None	3′-<u>CUUAGUUCAUU</u>	^{4S}UTP (2 mM)	19	^{4S}U
5	None	3′-<u>CUUAGUUCAUU</u>	^{4S}UTP (2 mM)	1	^{4S}U
6	None	3′-<u>GGUCAGAUU</u>	^{4S}UTP, ^{B}ATP (1 mM each)	14	^{B}A, ^{4S}U
7	None	3′-<u>CUUAGUUCAUU</u>	^{4S}UTP, ^{B}ATP (1 mM each)	17	^{B}A, ^{4S}U
8	None	3′-<u>GGUCAGAUU</u>	^{4S}UTP, ^{B}ATP (1 mM each)	17	^{B}A, ^{4S}U
9	None	3′-<u>GGUCAGAUU</u>	^{4S}UTP, ^{B}ATP (1 mM each)	4	^{B}A, ^{4S}U
10	None	3′-<u>CUUAGUUCAUU</u>	^{4S}UTP (1 mM)	20	^{4S}U
11	Synthesis	3′-<u>UCGACGG</u>AACC	^{4S}UTP (1 mM)	4	2 ^{4S}U
12	None	3′-<u>ACCUGAG</u>AAGG	^{4S}UTP (1 mM)	4	2 ^{4S}U
13	None	3′-<u>CAAGUCC</u>AACC	^{4S}UTP (1 mM)	0.2	2 ^{4S}U
14	None	3′-<u>UCGACGG</u>AACC	^{4S}UTP (1 mM)	0.2	2 ^{4S}U
15	PCR	3′-<u>UCGACGG</u>^{2N}P^{2N}PCCUGCGUC	^{4S}UTP (0.1 mM), Comp. NTPs	20	2 ^{4S}U
16	PCR	3′-<u>CAAGUCC</u>^{2N}P^{2N}PUGAUCGUA	^{4S}UTP (0.1 mM), Comp. NTPs	4	2 ^{4S}U
17	PCR	3′-<u>ACCUGAG</u>^{2N}P^{2N}PGUGUAUGU	^{4S}UTP (0.1 mM), Comp. NTPs	2	2 ^{4S}U
18	None	3′-<u>UCGACGG</u>^{2N}P^{2N}PCCUGCGUC	^{4S}UTP (0.1 mM), Comp. NTPs	0.1	2 ^{4S}U

loops within the ligase domain, were mutagenized at an average level of 20%. Because this mutagenesis included the former RT-PCR primer-binding sequence, a new RT-PCR primer-binding sequence was appended to each pool molecule. In half the pool molecules, most of the ligase core residues were mutagenized at an average level of 3%, whereas in the other half, they were not intentionally mutagenized.

The additional rounds of selection-amplification were performed with three noteworthy modifications to the protocol of the first 10 rounds (Table 1). First, longer template RNAs were used to favor those variants better able to recognize primer-template complexes with long, unpaired template segments. Second, selection was based on the ability to add two tagged U's rather than one tagged U and one tagged A. This change was implemented after learning that the round-10 ribozyme uses biotinylated ATP much less efficiently than unmodified ATP. Third, high concentrations of unmodified ATP, CTP, and GTP were included to disfavor those variants prone to incorporating these competitor nucleotides instead of the tagged Watson-Crick match to the template.

Isolates from rounds 14 to 18 were screened for the ability to fully extend a 10-nt primer (GAAUCAAGGG) on an 18-nt template (3′-CUUAGUUCCCGCCCGGCC, underline indicates segment that pairs with the primer). Most isolates from round 18 had disrupted ligase domains and showed no sign of polymerase activity when assayed individually. They presumably were selected because of a parasitic ability to deliver their primer to the active site of a different molecule. Other isolates had polymerase activity and were much more active than the round-10 parental ribozyme. Comparative sequence analysis of the 22 most active isolates (23) identified conserved residues and structural features, which clustered in the 3′-terminal half of the accessory domain (Fig. 1C) and are likely to be critical for its function. This analysis also suggested a model for the secondary structure of the accessory domain (Fig. 1C) and identified four residues in the domain that consistently differed from the round-10 ribozyme (23). These mutations likely conferred increased polymerase activity.

One isolate from round 18 was particularly adept at using long templates. To investigate features of the ribozyme needed for activity, derivatives of this isolate were constructed and tested (23). A 189-nt derivative (Fig. 1D) that retains the polymerization activity of the full-length round-18 isolate has been most extensively characterized. This derivative (hereafter referred to as the round-18 ribozyme) has all the features of the accessory domain that were conserved among the 25 most active isolates, including the four mutations thought to confer improved activity (Fig. 1D). Additionally, it has

a U-to-C mutation within the ligase domain in the segment designed to pair with the 7-nt RNA that completes the ligase domain (Fig. 1D). Reversing this point mutation diminished activity, and omitting the 7-nt RNA abolished activity altogether. We therefore speculate that, although the 7-nt RNA must still pair to this segment, a non–Watson-Crick distortion of the helix better accommodates long template RNAs. It is noteworthy that the four other active round-18 isolates also had point mutations within the segment designed to pair with the 7-nt RNA (23).

Extensive and accurate RNA polymerization. Although the round-18 ribozyme was only marginally improved over the round-10 ribozyme when short templates were used (Fig. 3), it was much better when longer templates were used (Fig. 4). With templates coding for 4, 8, 11, and 14 nucleotides, the round-18 ribozyme extended the primer by the corresponding number of residues (Fig. 4B). Normal RNA linkages were synthesized, as determined by nuclease analysis of the extension product (23). Furthermore, extension was predominantly by the Watson-Crick match to the template. When primers that were fully extended using the template coding for 11 nt were cloned and sequenced, 89 of 100 sequences precisely matched the template. Of the 1100 residues sequenced, only 12 were mismatches (Fig. 4C), implying an overall Watson-Crick error rate of 0.011 per nucleotide. Thus, the round-18 ribozyme can accurately use information from an RNA template and all four nucleoside triphosphates to extend an RNA primer by a complete turn of an RNA helix.

To examine the accuracy of polymerization more systematically, we measured the efficiency of matched and mismatched extension using four templates that differed only at the first coding nucleotide. For each template, the Watson-Crick match was added most efficiently (Table 2). The best fidelity was with the -C- template, for which the error rates ranged from 0.00004 to 0.0002. Fidelity was lower for the -G- and -U- templates, primarily because extension by the two G:U wobble mismatches had error frequencies of 0.044 and 0.085. The overall fidelity was 0.967 per residue. In other words, with all four NTPs supplied at equimolar concentrations, extension by the matched nucleotide typically would be 96.7% of the total extension.

A fidelity of 96.7% (Table 2) is somewhat lower than the 99% fidelity inferred from sequencing fully extended primer molecules (Fig. 4C). Two factors contribute to the higher fidelity observed in Fig. 4C. The first is the influence of sequence context on fidelity (24). The second arises from the fact that, after a mismatch was incorporated, further extension of the growing chain was less efficient because the 3′ terminus of the primer no longer paired with

the template (25). Thus, the full-length product of Fig. 4C was enriched in molecules with few misincorporated nucleotides. Mismatch incorporation also reduces the extension efficiency of proteinaceous polymerases, a property particularly important for certain DNA polymerases because it facilitates exonucleolytic proofreading (26).

Polymerase fidelity is most simply expressed by assuming that all four NTPs are present at equal concentration, even though cellular NTP concentrations are not equimolar (27). For the round-18 ribozyme, certain asymmetric NTP ratios would produce observed fidelities significantly greater than 0.967. For example, lowering the GTP concentration to one-tenth that of the other NTPs would decrease G misincorporation by 10-fold, while

only lowering the fidelity of extension across from C from 0.9996 to 0.996. Because G misincorporation was the major source of error, this would increase the overall fidelity from 0.967 to 0.985 with the templates in Table 2.

A Watson-Crick fidelity of 0.985 is still lower than the ≥0.996 fidelity seen with viral polymerases that replicate RNA by using RNA templates (28, 29), and it is much lower than that seen for polymerases that replicate DNA (30). Nevertheless, the Watson-Crick fidelity of the round-18 ribozyme compares favorably to that of other ribozymes. Previously, the best ribozyme fidelity had been obtained with the engineered ligase derivative (Fig. 1A), which has an overall fidelity of 0.85 with equimolar NTPs and observed fidelities of 0.88 to 0.92 with more favorable
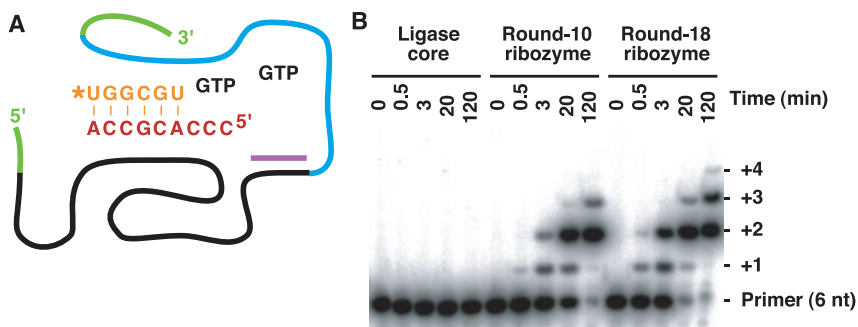
Fig. 3. Intermolecular primer extension using a short primer-template. (**A**) Schematic of the RNAs used in these polymerization reactions. Drawing conventions are as in Fig. 2A. Note that the primer is ^{32}P end-labeled and that neither the primer RNA nor the template RNA is tethered or hybridized to the ribozyme. (**B**) PhosphorImager scan of a denaturing gel separating primer-extension products of the indicated ribozymes. Reactions included 1 μM ribozyme, 10 μM primer, 10.5 μM template, and 4 mM GTP, and were incubated for the indicated time in polymerization assay conditions (33). "Ligase core" refers to a ribozyme identical to that of Fig. 1A (black strand), except that its 3′ terminus was modified to pair with the 7-nt RNA (GGCACCA) that completes the ligase core; no extension was observed with this ribozyme. The round-10 and round-18 ribozymes are depicted in Figs. 1C and D, respectively. After long incubation times, some of the primer was extended with three templated residues plus one nontemplated residue. Many proteinaceous polymerases, including Qβ replicase (42) and *Taq* DNA polymerase (43), also tend to add an extra, nontemplated residue.

Table 2. Watson-Crick fidelity of RNA polymerization. For each template-NTP combination, the efficiency of extension by at least 1 nt was determined. For each template, the four efficiencies were normalized to that of the matching NTP, yielding the relative efficiencies of extension. The relative efficiency of extension for a mismatch is the same as its error rate (27) and misinsertion ratio (26). Fidelities were calculated as the efficiency for the match, divided by the sum of the efficiencies for all four NTPs. The average fidelity is the geometric average of the fidelities for each template (41). For each Watson-Crick match, the absolute efficiency (per molar per minute) is also shown in parentheses. It is reported as the observed rate constant of primer extension divided by NTP concentration, from polymerase assays (33) using 5 μM ribozyme, 2 μM primer (CUGCCAACCG), and 2.5 μM template (3′-GACGGUUGGCXCGCUUCG, where X is the indicated template residue). In these assays, NTPs were supplied at concentrations well below half-saturating.

Template	Relative efficiency of extension				Fidelity
	ATP	CTP	GTP	UTP	
-A-	0.0034	0.0014	0.0043	1.0 (5.3)	0.991
-C-	0.0002	0.0002	1.0 (5.4)	0.00004	0.9996
-G-	0.0002	1.0 (41)	0.0006	0.044	0.957
-U-	1.0 (87)	0.0001	0.085	0.0002	0.921
				Average =	0.967

NTP ratios (*15*). Moreover, the fidelity of the round-18 ribozyme approaches that of one proteinaceous polymerase, pol η, a eukaryotic polymerase needed for accurate replication of UV-damaged DNA (*31*). Yeast pol η has an overall fidelity of 0.984, which would increase to 0.989 with an optimal NTP ratio (*32*).

General RNA-dependent RNA polymerization. The round-18 ribozyme worked with every primer-template tested. As the primer-templates have no sequence features in common, the ribozyme does not rely on any sequence-specific contacts. Additionally, because the primer-template complex must shift in register relative to the polymerase active site each time another nucleotide is added, every polymerization experiment actually examines primer extension in a series of differing sequence contexts, demonstrating further that the polymerization is general with respect to nucleotide sequence. Granted, the

efficiency of nucleotide addition varied depending on the sequence context, as evidenced by an uneven distribution of extension intermediates (Fig. 4), but this phenomenon is also observed with protein polymerases (*26*, *27*).

All templates used heretofore were less than 21 nt long, leaving open the question of whether the ribozyme could accommodate longer primer-template substrates, as would be required of an RNA replicase. To address this question, three related substrates were tested. The first was a short substrate, with a 10–base pair primer-template duplex and a 10-nt template coding region. The second substrate was the same, except its template coding region was lengthened from 10 to 100 nt. The ribozyme extended this substrate by as many as 9 nt in 23 hours, although somewhat less efficiently than it extended the short version (Fig. 5). The third substrate was the same as the second, except that its primer-template duplex was lengthened

from 10 to 60 base pairs. The ribozyme extended this substrate just as efficiently as the second substrate. Thus, the ribozyme is free from steric constraints that would preclude polymerization using long templates or long primer-template helices.

Given this general recognition of primer-templates, the range for primer extension, currently just beyond one helical turn, is limited merely by the ribozyme's efficiency. Polymerization is too slow for more extension to be observed within 24 hours, and longer incubations yield limiting returns, because buffer and ionic conditions optimal for polymerization (*33*) also promote ribozyme and template degradation. Reactions would have to be supplemented periodically with fresh ribozyme to achieve polymerization substantially beyond one helical turn. Nonetheless, in initiating extension of a long primer hybridized to a long template, the round-18 ribozyme demonstrates polymerization that is
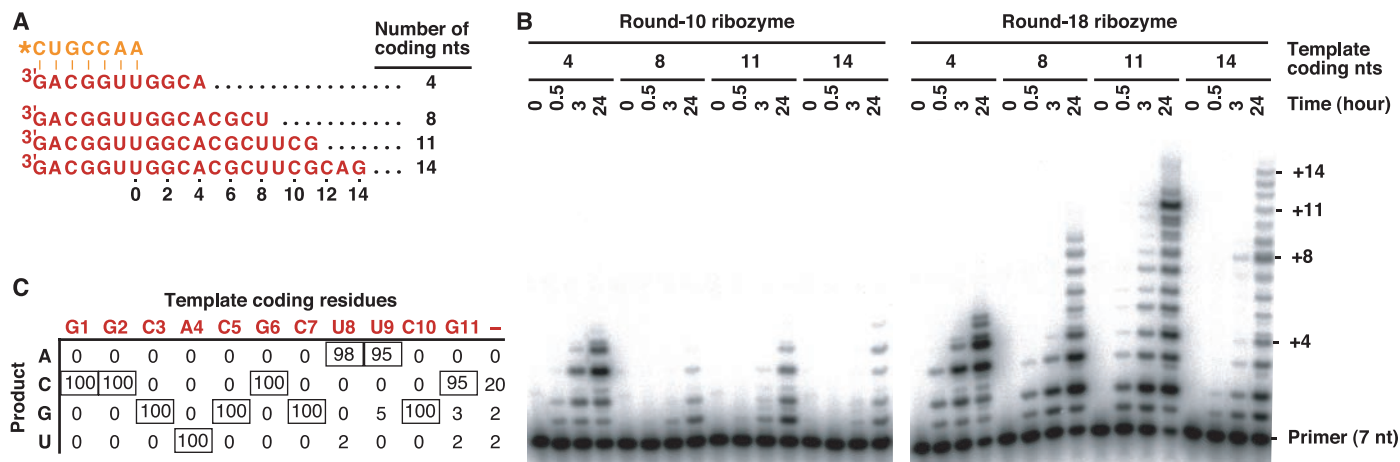
Fig. 4. Improved RNA polymerization. (**A**) Schematic of the 7-nt primer (orange) and the templates (red) with 4 to 14 coding residues used in these polymerization reactions. The ribozymes (not depicted) were as in Fig. 3. (**B**) PhosphorImager scan of a denaturing gel separating primer-extension products of the indicated ribozymes. Reactions included 5 μM ribozyme, 2 μM primer, 2.5 μM template, and 4 mM each NTP, and were incubated for the indicated time in polymerization assay conditions (*33*). (**C**) Tabulation of nucleotides inserted across from template residues during primer-extension by the round-18 ribozyme. Full-length primer-extension products encoded by the template with 11 coding residues were cloned, and 100 full-length clones were sequenced (*23*). For each template residue (red), the number of clones that had an A, C, G, or U at the corresponding position was tabulated. Tallies representing Watson-Crick matches to the template are boxed. The column below the red dash reports the identity of the nontemplated nucleotide (Fig. 3B, legend) added to the 3′ terminus of 24 of the sequenced primer-extension products. Coding residues are numbered as in (A).
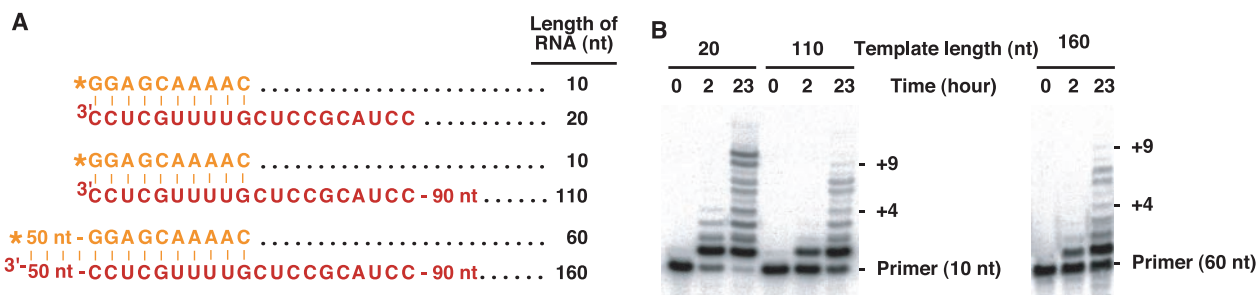
Fig. 5. RNA polymerization using long primers and long templates. (**A**) Schematic of the three primer-template combinations examined (*23*). Each combination included a 10- or 60-nt primer (orange) and a 20-, 110-, or 160-nt template (red). All three combinations have the same sequence near the site of polymerization (residues defined). (**B**) PhosphorImager scan of denaturing gels separating primer-extension products. Reactions included 5 μM round-18 ribozyme, 0.5 μM primer, 1 μM template, and 4 mM each NTP, and were incubated for the indicated time in polymerization assay conditions (*33*).

general not only with respect to the nucleotide sequence but also to the length of the primer-template complex.

General template-directed RNA polymerization requires recognition of the generic features of a primer-template complex in addition to ever-changing NTP specificity, as dictated by the next template residue. It is a complex reaction—one of the more sophisticated reactions catalyzed by single polypeptides. The demonstration that such an activity can be generated de novo, without reference to any biological ribozyme or structure, is a testament both to the catalytic abilities of RNA, as well as to modern combinatorial and engineering methodology. Key to this success may have been the stepwise procedure of first, isolating from random sequences an appropriate catalytic core in the context of a simple reaction (13, 14); second, optimizing the sequence of the catalytic core (16); third, determining the limits of the core activity (15, 34, 35); and fourth, flanking the core with additional random sequence and selecting for more sophisticated substrate binding. Thus far, efforts to select for polymerization activity in a single step directly from random-sequence RNA have yielded only ribozymes that decorate themselves inappropriately with tagged nucleotides (36).

How could general polymerase activity have arisen on early Earth? If emergence of the first RNA replicase ribozyme coincided with the origin of life, it would have had to arise in a single step from prebiotically synthesized RNA, without the benefit of Darwinian evolution. Our shortest construct retaining activity was 165 nt, with about 90 nt involved in important Watson-Crick pairing and at least another 30 critical nucleotides (23). Ribozymes with the efficiency, accuracy, and other attributes of an RNA replicase might have to be even larger than this. However, current understanding of prebiotic chemistry argues against the emergence of meaningful amounts of RNA molecules even a tenth this length (1). This difficulty is anticipated by those who propose that life, and Darwinian evolution, began before RNA. Some speculate that in this "pre-RNA world," life was based on an RNA-like polymer, yet to be identified, that possessed the catalytic and templating features of RNA but also a more plausible prebiotic synthesis (1). The pre-RNA life forms presumably later developed the ability to synthesize RNA, facilitating the emergence of an RNA replicase ribozyme, which in turn enabled the transition to the RNA world.

It will be interesting to examine the extent to which continued mutation and selection can improve the activity of the polymerase ribozyme. Perhaps ribozymes with accuracy and efficiency sufficient for self-replication can be generated. The requisite fidelity may be close at hand, possibly only requiring a reduction of the overall error rate to one-third its current value,

thereby increasing fidelity observed with unequal NTP concentrations from 0.985 to 0.995 (37). The increase in polymerization efficiency would need to be more substantial (at least 100-fold), although not beyond the degree of optimization achieved previously with in vitro evolution experiments. Other important issues will need to be addressed, including strand dissociation after polymerization. Nevertheless, the general polymerization activity of the round-18 ribozyme offers support for the idea of autocatalytic RNA replication in the distant past, as well as a new starting point for its demonstration in the not-so-distant future.

References and Notes

1. G. F. Joyce, L. E. Orgel, in *The RNA World*, R. F. Gesteland, T. R. Cech, J. F. Atkins, Eds. (Cold Spring Harbor Laboratory Press, New York, 1999), pp. 49–77.
2. N. R. Pace, T. L. Marsh, *Origins Life* **16**, 97 (1985).
3. P. A. Sharp, *Cell* **42**, 397 (1985).
4. T. R. Cech, *Proc. Natl. Acad. Sci. USA* **83**, 4360 (1986).
5. L. E. Orgel, *J. Theor. Biol.* **123**, 127 (1986).
6. A. J. Hager, J. D. Pollard, J. W. Szostak, *Chem. Biol.* **3**, 717 (1996).
7. D. P. Bartel, in *The RNA World* R. F. Gesteland, T. R. Cech, J. F. Atkins, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1999), pp. 143–162.
8. D. P. Bartel, P. J. Unrau, *Trends Cell Biol.* **9**, M9 (1999).
9. J. W. Szostak, D. P. Bartel, P. L. Luisi, *Nature* **409**, 387 (2001).
10. J. A. Doudna, J. W. Szostak, *Nature* **339**, 519 (1989).
11. J. A. Doudna, S. Couture, J. W. Szostak, *Science* **251**, 1605 (1991).
12. R. Green, J. W. Szostak, *Science* **258**, 1910 (1992).
13. D. P. Bartel, J. W. Szostak, *Science* **261**, 1411 (1993).
14. E. H. Ekland, J. W. Szostak, D. P. Bartel, *Science* **269**, 364 (1995).
15. E. H. Ekland, D. P. Bartel, *Nature* **382**, 373 (1996).
16. ———, *Nucleic Acids Res.* **23**, 3231 (1995).
17. Each subpool had the structure shown in Fig. 1B but a different percent mutagenesis of ligase core residues. For a residue mutagenized at 20%, the nucleotide precursor reservoir was doped such that the parental base would be added at a frequency of 0.8, whereas each of the other three bases would be added at a frequency of 0.067. Each subpool was constructed starting with two synthetic single-stranded pools (16). The only black residues of Fig. 1B that were not mutagenized were the six residues at the 3' terminus of the ligase core domain. These six residues (GGUGCC) corresponded to the Ban I restriction site used to construct the subpools and were designed to pair with the 7-nt RNA that completes the ligase domain (GGCACCA, purple RNA in Fig. 1B). An average of three RNA copies of each subpool were combined to generate a starting pool with a sequence complexity of 2 × 10¹⁵.
18. Pool extension reactions typically included 0.2 to 0.5 µM pool RNA linked to primer [Fig. 1B; (38)], 1.0 µM template RNA (Table 1), 1.0 µM GGCACCA RNA (Fig. 1B), and 0.1 to 2 mM tagged NTPs (Table 1) in pool extension buffer (60 mM MgCl₂, 200 mM KCl, 50 mM EPPS, pH 8.0, 22°C). In rounds 10 to 18, incubations also included 1.0 µM reverse-transcription primer (TTCAGATTGTAGCCTTC) and lacked KCl. The pool RNA and oligonucleotides were mixed together in water, incubated at 80°C for 3 min, then allowed to cool to room temperature for 10 min. Pool extension reactions were started by the simultaneous addition of pool extension buffer and NTPs. Reactions were stopped with addition of 80 mM EDTA after 0.1 to 36 hours (Table 1). Excess 4-thioUTP was removed with a Centricon YM-10 centrifugal filter device (Millipore), and molecules that had been tagged with 4-thioU were isolated on APM gels (44). APM gels were prepared by using 10 to 80 µM APM in the lower portion of the gel, no APM in the top portion of the gel, and 8M urea per 5% acrylamide

throughout the gel (Fig. 2). For isolation of RNA extended with at least one 4-thioU (rounds 1 to 10, Table 1), RNA was eluted from a gel slice extending from the APM interface to the migration of pool RNA with a single 4-thioU. For isolation of RNA extended by at least two 4-thioUs (Table 1), RNA was eluted from a slice containing only the APM interface. In rounds 2 to 5, 9, 12 to 14, and 17, RNA was further purified on a second APM gel. Eluted RNA was precipitated and reverse-transcribed (16) by using primer CGGGACTCTGACCT-TGG (rounds 1 to 3, 6) AAACGGGACTCTGACCTTG (rounds 4, 5, 7, 9), TTCGGGACTCTGACCTT (rounds 8, 10), or TTCAGATTGTAGCCTTC (rounds 11 to 18). RT primers were alternated in rounds 4 to 10 to vary the 3'-terminal residues of the pool RNA, disfavoring the selection of molecules that extend their 3' terminus rather than extending the attached primer. In rounds including biotinylated ATP (biotin-N6-ATP, New England Nuclear), RNA-cDNA duplex molecules were also purified by capture on streptavidin magnetic beads (16). PCR amplification of the cDNA used a primer that completed the T7 RNA polymerase promoter sequence, (TTC*TAATACGA*C*TCACTATA*GGACAACC, italics, T7 promoter sequence; underline, 5' primer-binding site, Fig. 1B). PCR DNA was transcribed (16) to generate RNA for the next round of selection or cloned (Topo cloning kit, Invitrogen) for sequencing and further analysis.
19. Throughout this manuscript, added nucleotide residues (pA, pC, pG, and pU) are abbreviated by the corresponding nucleoside symbols (A, C, G, and U).
20. W. K. Johnston, D. P. Bartel, data not shown.
21. To examine whether the ribozyme preferred substrate sequences used during its selection, the primer-template from selection rounds 1, 6, 8, and 9 (Table 1) was also tested and found not to be used any more efficiently than the substrate of Fig. 3.
22. The starting pool for round 11 was constructed as in (17), except: (i) it was based on the round-10 ribozyme; (ii) it was a combination of two subpools, with the ligase core positions mutagenized at 0 or 3%; (iii) blue positions of Fig. 1C were mutagenized at 20%; and (iv) a new RT-PCR primer-binding site (GAAGGCUACAAUCUGAA) was appended to the 3' end of the pool molecules. About three RNA copies of each subpool were combined to generate a starting pool with a sequence complexity of 10¹⁵.
23. Supplemental material describing comparative sequence analysis of ribozyme variants, engineering of shorter and more active polymerase constructs, nuclease analysis of the primer-extension product, methods for Fig. 4C, and RNAs of Fig. 5 is available on *Science* online at www.sciencemag.org/cgi/content/full/292/5520/1319/DC1
24. The contribution of sequence context [as opposed to sequence composition (45)] to the higher fidelity in Fig. 4C is particularly apparent with the extension across from G1, G2, and G6, which was by C at all three positions in all 100 products sequenced. In contrast, Table 2 shows that when coding residue 4 was changed to G, 4% of the extension across from that G was by the U mismatch. To confirm this apparent effect of sequence context, the relative efficiencies of matched and mismatched extension across from the G6 coding residue were determined as in Table 2, by using the primer CUGCCAACCGUG and the template, ribozyme, and conditions of Fig. 4C. The efficiency with the CTP match was 27 M⁻¹ min⁻¹, and relative efficiencies were 0.0007, 1.0, 0.002, and 0.017 with ATP, CTP, GTP, and UTP, respectively, for a fidelity at G6 of 0.981—somewhat better than the 0.957 fidelity seen across from a G at coding residue four (Table 2). Thus, sequence context can influence accuracy of polymerization, a phenomenon also observed with protein polymerases (26, 27).
25. For example, in one context, relative extension efficiencies were 1.0, 0.044, 0.78, or 0.056 when the previous nucleotide was a U incorporated across from an A, C, G, or U template residue, respectively.
26. H. Echols, M. F. Goodman, *Annu. Rev. Biochem.* **60**, 477 (1991).

27. T. A. Kunkel, *Bioessays* **14**, 303 (1992).
28. D. A. Steinhauer, J. J. Holland, *J. Virol.* **57**, 219 (1986).
29. C. D. Ward, J. B. Flanegan, *J. Virol.* **66**, 3784 (1992).
30. L. A. Loeb, T. A. Kunkel, *Annu. Rev. Biochem.* **51**, 429 (1982).
31. M. T. Washington, R. E. Johnson, S. Prakash, L. Prakash, *J. Biol. Chem.* **274**, 36835 (1999).
32. When using k_{cat}/K_m values from (*31*), the optimal ratio of dNTP concentrations for yeast pol η would be 100:23:90:35 for dATP, dCTP, dGTP, and TTP, respectively. When using the values from Table 2 for the round-18 ribozyme, a ratio of 50:100:4:25 for ATP, CTP, GTP, and UTP would result in an observed fidelity of 0.990.
33. Ribozyme RNAs were transcribed from PCR-generated DNA templates, which were produced by using plasmid DNA and the appropriate primers. The DNA template for the round-18 ribozyme had penultimate 2′-methoxyl substitutions to reduce the addition of nontemplated residues at the ribozyme 3′ terminus (*46*). Standard polymerization assays were in 200 mM MgCl₂, 50 mM tris, pH 8.5, at 22°C. Although polymerization using an intermolecular primer was readily observed in the conditions of the selection (60 mM MgCl₂, 200 mM KCl, 50 mM EPPS, pH 8.0, 22°C), the standard assay conditions were more optimal. Ribozyme, primer, template, and NTPs were included at the concentrations indicated, and the GGCACCA RNA that completes the ligase domain was in 1.25-fold molar excess over the ribozyme concentration. These RNAs were mixed together in water, incubated at 80°C for 3 min, then incubated at 22°C for at least 5 min before starting the reaction by simultaneous addition of buffer, salt, and NTPs. Heating the ribozyme separately from the primer and template RNAs did not affect the polymerization reaction kinetics. Reactions were stopped by addition of 1.6 volumes of 145 mM EDTA/6 M urea, heated (90°C, 5 min) in the presence of competitor RNA designed to hybridize to the template RNA, then analyzed on sequencing gels. Incubation with competitor RNA, added in 10-fold molar excess over template RNA, led to better gel resolution because it prevented reassociation of the extended primer with the template.
34. N. H. Bergman, W. K. Johnston, D. P. Bartel, *Biochemistry* **39**, 3115 (2000).
35. M. E. Glasner, C. C. Yen, E. H. Ekland, D. P. Bartel, *Biochemistry* **39**, 15556 (2000).
36. P. J. Unrau, D. P. Bartel, unpublished results.
37. Consider hypothetical replicases about the size of the round-18 ribozyme (∼200 nt) that synthesize an average of five 200-nt RNA strands within their lifetime. For replication through complementary-strand then second-strand synthesis, two of the five strands produced by each polymerase must have the correct residues at all of the positions that contribute to function. If the identities of effectively 80% of the 200 nucleotides contribute to function, then these replicases would require a fidelity of at least 0.995 ($0.995^{160} = 0.45 ≈ 2/5$). Note that this scenario does not account for the dilution of productive ribozyme that would occur as active ribozyme variants replicate an increasing number of inactive variants; compartmentalization and selection would be needed to achieve sustainable replication (*7, 9*).
38. The 5′ terminus of gel-purified pool RNA was ligated (*47, 48*) to the DNA portion of a DNA-RNA chimeric oligonucleotide, 3′-ATATCACTC-5′-5′-X$_{(6−8)}$ (residues defined, deoxynucleotides synthesized by using "reverse" phosphoramidites from Glenn Research; 5′-5′, phosphodiester linkage joining the 5′ terminus of the DNA with the 5′ terminus of the RNA; X$_{(6−8)}$, RNA primer segment 6 to 8 nt in length complementary to the template indicated in Table 1). The pool was then gel-purified to remove the splint oligonucleotide needed for ligation.
39. An A analog, 2-aminopurine pairs with U and 4-thioU, but unlike A, its Watson-Crick pairing with 4-thioU does not involve the sulfur atom. Because the sulfur atom subtly distorts pairing geometry, templates with 2-aminopurine were used in later rounds of selection.
40. Repeated error-prone PCR was performed as described (*13*), except serial dilutions were after every 6 cycles of PCR. RNA from four subpools with expected mutagenesis levels of 0, 2, 4, and 6%, was mixed and used for round 15. RNA from two subpools with expected mutagenesis levels of 0 and 1% was used for round 16. RNA from two subpools with expected mutagenesis levels of 0 and 2% was used for round 17.
41. Because the net accuracy of RNA polymerization is represented by the product of the fidelities for each added nucleotide, geometric averages are reported throughout this study when describing overall fidelity.
42. J. N. Bausch, F. R. Kramer, E. A. Miele, C. Dobkin, D. R. Mills, *J. Biol. Chem.* **258**, 1978 (1983).
43. J. M. Clark, *Nucleic Acids Res.* **16**, 9677 (1988).
44. P. J. Unrau, D. P. Bartel, *Nature* **395**, 260 (1998).
45. Sequence composition of the template explains practically none of the higher fidelity in Fig. 4C. The template of Fig. 4C has one A, four C, four G, and two U coding residues. Thus, the expected fidelity per nucleotide, calculated by using the values of Table 2, is $[0.991 × (0.9996)^4 × (0.957)^4 × (0.921)^2]^{(1/11)} = 0.969$, a value very close to the generic overall fidelity of 0.967 calculated in Table 2.
46. C. Kao, M. Zheng, S. Rudisser, *RNA* **5**, 1268 (1999).
47. M. J. Moore, P. A. Sharp, *Science* **256**, 992 (1992).
48. T. Tuschl, M. P. Ng, W. Pieken, F. Benseler, F. Eckstein, *Biochemistry* **32**, 11658 (1993).
49. We thank members of the laboratory and F. Solomon for helpful comments on this manuscript. Supported by grants from the NIH (D.P.B.), a Medical Research Council (Canada) postdoctoral fellowship (P.J.U.), and a Howard Hughes Medical Institute predoctoral fellowship (M.E.G.).

16 March 2001; accepted 10 April 2001