A polymerase ribozyme increases copying fidelity through pyrophosphate-mediated RNA repair

Alexandra D. Kent,[1,2] Lucy L. Yang,[1] Kyle H. Cole,[3] Ivee A. Lee,[2] and Andrej Lupták[1, 2, 3]*

[1] Department of Chemistry, University of California at Irvine; Irvine, CA 92617

[2] Department of Pharmaceutical Sciences, University of California at Irvine; Irvine, CA 92617

[3] Department of Molecular Biology and Biochemistry, University of California at Irvine; Irvine CA 92617

Abstract: Prior to the emergence of the contemporary biosphere, the first replicating systems are thought to have progressed through an RNA-based stage. Such an evolving world would likely have transferred heritable information during replication using RNA polymerase ribozymes. Though substantial effort has been put forth towards evolving RNA polymerases, many variants suffer from premature termination and low fidelity, resulting in low yields of full-length or active sequences. Replication of longer sequences requires a sufficiently high fidelity to lend an evolutionary advantage to an evolvable system. Here we demonstrate ribozyme-mediated repair of mismatched and damaged RNA sequences. Under conditions of saturating pyrophosphate concentrations, we show that a polymerase ribozyme can repair RNA sequences terminated in a mismatch, a non-extendable 2′-3′ cyclic phosphate, or both, to generate a triphosphorylated nucleotide. This repair step increases the fidelity and allows polymerization along an extended template. This increase of copying fidelity advances the longstanding goal of developing a self-replicating polymerase ribozyme.

Main Text

Because the cellular functions of ribonucleic acids (RNAs) include critical roles in catalysis, translation, and metabolite-dependent regulation of gene expression, RNAs may have been the first biopolymers to store information and catalyze chemical reactions on an early Earth.[1–5] The replication process is fundamental to life and is the mechanism by which heritable information is stored and propagated. In the extant biosphere, this is achieved through protein-based polymerase enzymes, but in a hypothetical RNA World—prior to the advent of genetically encoded proteins—replication would likely have been carried out by catalytic RNAs termed polymerase ribozymes.[6,7] These ribozymes could serve as the central molecule linking the phenotype of an early protocell to its genotype, but sufficient activity and fidelity would be critical to build up an information storage system.[8,9]

The central reaction catalyzed by enzymes replicating phosphodiester-based informational polymers is the transfer of the α-phosphate of a triphosphorylated nucleoside to a 3′-hydroxyl of another nucleotide, with concomitant release of pyrophosphate (Fig. 1). The reverse reaction is pyrophosphorolysis of polynucleotides, yielding a nucleoside triphosphate and a one-residue shorter strand. Beginning with the first *in vitro* selection of a ligase ribozyme, a large body of work using molecular evolution and engineering has sought to isolate ribozymes that catalyze phosphodiester formation.[10–14] The ligase ribozyme has been evolved into progressively more sophisticated polymerases capable of promoter-based polymerization, production of other functional nucleic acids, reverse transcription, and transcription of non-canonical (xeno-) nucleic acids, among other functions.[15–21] However, most of the advanced polymerase ribozymes suffer from both mutation burdens that reduce the activity of produced transcripts and low yields of full-length products due to premature termination and degradation.[22–24] Recent reports demonstrate that evolving the polymerase ribozymes further under selection pressure to produce active ribozyme sequences increased the fidelity enough to enable successive rounds of amplification

of an active, short self-cleaving ribozyme.[25] Even with these advances, replicating active polymerase ribozymes remains an unsolved challenge, highlighting a need for novel approaches to increasing replication fidelity. An early process for RNA repair could serve to increase replication fidelity, but very few mechanisms of RNA repair have been reported to date and none have been catalyzed by ribozymes. A mechanism for repairing both degraded and mismatched sequences would lend an evolutionary advantage to an evolving system and is an unexplored approach to increasing polymerase ribozyme fidelity.[26–28]

We demonstrate that ribozyme-catalyzed pyrophosphorolysis of mismatched RNA sequences yields a nucleoside triphosphate, repairs the mutation, and increases the fidelity of the subsequent extension. Using sequencing analysis of mismatched and repaired RNAs, we show that repair of terminal mismatches improves downstream replication fidelity. Sequencing of RNAs generated from the polymerization of long templates reveal an increase in overall length and fidelity in the presence of pyrophosphate. We further demonstrate that addition of pyrophosphate and the polymerase ribozyme to 2′-3′ cyclic phosphate-terminated (damaged) RNA sequences also facilitate damage repair and subsequent RNA polymerization. Thus, we present a mechanism for increasing the copying fidelity of long RNAs by a single polymerase ribozyme bringing us one step closer to a fully self-replicating RNA system.

Results

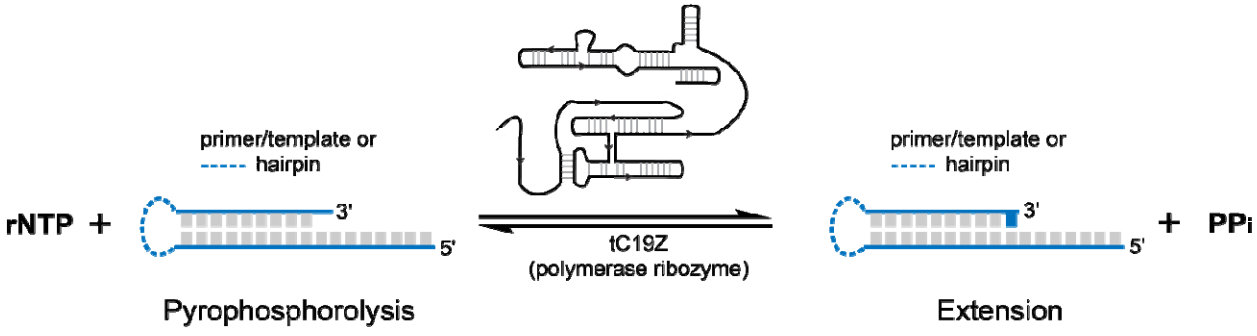

Figure 1. Reversible RNA synthesis by the polymerase ribozyme. The RNA polymerase ribozyme, tC19Z, catalyzes templated addition of a nucleotide to the 3′-end of a primer–template RNA with pyrophosphate (PPi) as a byproduct of the reaction. The reverse reaction, pyrophosphorolysis of RNA, removes the last nucleotide, yielding a nucleoside triphosphate. Dashed line indicates a linkage between the primer and the template, forming a single-molecule RNA hairpin used in this study.

One of the higher-fidelity RNA polymerases ribozymes, tC19Z, was selected *in vitro* from an earlier polymerase (R18), which in turn originated from the class I RNA ligase [16,18]. The tC19Z sequence was evolved to have improved sequence generality and polymerase activity, although the template sequence also has a significant influence on replication efficiency. To extend a sequence, the 5′ end of the ribozyme base-pairs with the 5′ end of the template sequence and extends a primer (Fig. 1). Extension of the primer generates pyrophosphate, a byproduct of polymerization that has also been suggested as a source of available phosphate on an early Earth, which would have been critical due to the role phosphate plays in energy metabolism and genome replication.[29–33] We hypothesized that addition of pyrophosphate to the polymerization reaction would shift the equilibrium of the reaction and induce a low level of pyrophosphorolysis (Fig. 1) that could increase replication fidelity by selectively removing mismatched and damaged RNA.

Pyrophosphate-mediated mismatch repair

To maintain fidelity of genome copying at non-deleterious levels, modern replication machinery uses distinct active sites or entire protein complexes to correct mismatched base-pairs, but at the onset of evolvable systems, it is unlikely that mismatch repair was available as an independent biochemical activity.[34,35] Thus, a simpler process for increasing replication fidelity would have been a desirable feature of an early replicating polymerase. To test whether ribozyme-mediated pyrophosphorolysis removes a mismatched nucleotide, we prepared hairpin sequences with 3′-terminal mismatches. We fused the primer and template to form a hairpin (Fig. 1) that serves both as a substrate for the tC19Z ribozyme at exact stoichiometry and as a practical construct for downstream analysis using high-throughput sequencing. We generated the hairpins *via* an *in vitro* transcription reaction of a self-cleaving HDV-like ribozyme (drz-Mbac-1) fused to the 3′ terminus of the target sequence.[36] The ribozyme self-cleaves, yielding a 2′-3′-cyclic phosphate (2′-3′-cP) on the 3′ end of the primer segment. The cyclic phosphate can be removed enzymatically by T4 polynucleotide kinase (PNK) (Fig. S1).[37,38]

We began by characterizing the reactivity of a terminal A/C mismatch in the primer/template sequence with the cyclic phosphate removed from the 3′ end. RNA polymerization by the ribozyme was not observed upon addition of nucleoside triphosphates (NTPs), indicating that an A/C mismatch acts as a chain-terminator (Fig. 2a,b). After addition of saturating concentrations of pyrophosphate (10 mM, with clearly visible Mg-PPi precipitate) and NTPs, we observed ribozyme-mediated extension of the sequence (Fig. 2), but we did not observe the −1 product by denaturing polyacrylamide gel electrophoresis (PAGE), likely because the equilibrium highly favors polymerization (SI Note 1, Fig. 2b, and Fig. S2). When removed through pyrophosphorolysis, the terminal nucleotide is converted into a nucleoside triphosphate, and an A/C-terminated hairpin would produce a molecule of ATP. To obtain independent evidence of pyrophosphorolysis, we detected this critical metabolite using an ultrasensitive firefly-luciferase-based assay, which yields a photon for each ATP molecule present in the reaction.[39] Because

pyrophosphate inhibits the luciferase reaction, insoluble pyrophosphate was first removed from the samples by centrifugation, and the remaining pyrophosphate was subsequently hydrolyzed by addition of pyrophosphatase enzyme. The resulting samples were incubated with the firefly luciferase and D-luciferin and were then imaged using an ultracooled EM-CCD camera. The integrated luminescence was measured for both the pyrophosphorolysis reaction and control reactions lacking one of the components. In the absence of either pyrophosphate, the ribozyme, or the hairpin substrate, we observed a baseline level of luminescence. In reactions where all components were present, the luminescence was significantly greater, indicating pyrophosphate-mediated synthesis of ATP (Fig. 2c and Fig. S3). These experiments provide evidence that the mechanism of repair occurs by a pyrophosphorolysis reaction and demonstrate the first instance of ribozyme-catalyzed production of ATP.

To further analyze the repair reaction, we used high-throughput sequencing (HTS) of the mismatched-terminated hairpin reactions (Fig. 2d,e). HTS analysis revealed that pyrophosphorolysis of the substrate yields approximately 100-times higher concentrations of a hairpin lacking the terminal nucleotide than the no-pyrophosphate controls (Fig. S4), providing experimental confirmation for the pyrophosphorolysis of the terminal nucleotide.

We repeated this analysis for each of the eleven other possible combinations of terminal mismatches (Figs. S2 and S5). Sequence analysis of the extension reactions correlated well with the gel electrophoresis analysis. In addition, the HTS data provided a deeper insight into the reactions, revealing the fractions of the sequences that were repaired and subsequently extended. PAGE analysis was consistent over replicate experiments. We found that the efficiency of repair and extension depends on the identity of the mismatch: purine/pyrimidine mismatches were repaired the most efficiently, with the exception of the U/G wobble. Purine/purine mismatches were repaired and extended at a low frequency upon the addition of pyrophosphate but were not extendable by the ribozyme in the absence of pyrophosphate (Fig.

S5). In contrast, pyrimidine/pyrimidine mismatches were extended well by the ribozyme, but most of the extension products retained the mismatch (Fig. S5). Sequences containing terminal C/A or G/U mismatches behave like the A/C mismatch and only showed extension (with varying efficiencies) after pyrophosphorolysis of the mismatch.

In addition to PAGE analyses, we obtained quantitative results for the fidelity of extension products from the sequencing data. We measured the fidelity at each position of the extension and found that for all mismatches, repair of the mismatched base led to an increase in the fidelity of downstream extensions (Fig. 2f,g). In some cases, such as the pyrimidine/pyrimidine mismatches, the effect is strongly pronounced: although the frequency of repair is low, when a hairpin is repaired, the extension occurs with high fidelity. Specifically, in the case of the U/C mismatch, extension of the terminal mismatch preferentially adds a second mismatch (C/A) which then acts as a chain terminator (fig. S6). Therefore, we conclude that mismatched sequences that do not terminate the extension are more likely to acquire additional mutations, as has been observed in enzyme-free polymerization of nucleic acid sequences, implying a mechanism for error propagation that must involve an allosteric modulation of the fidelity of incoming nucleotide incorporation by upstream mismatches.[40]

We observed that addition of pyrophosphate inhibited the forward polymerization reaction in the positive control (Fig. S7). To quantify the pyrophosphate reaction inhibition, we used a non-hairpin, primer–template complex to measure reaction kinetics of the forward extension reaction. Initial rates were plotted as a function of substrate concentration and revealed an approximate K_M of 4 mM with respect to NTPs (Fig. S8A–C). We then measured the reaction kinetics in the presence of pyrophosphate and plotted the initial rates of tC19Z-mediated polymerization as a function of pyrophosphate concentration (Fig. S8D,E). At millimolar PPi concentrations, the forward reaction was inhibited approximately five-fold, compared to when no PPi is present.

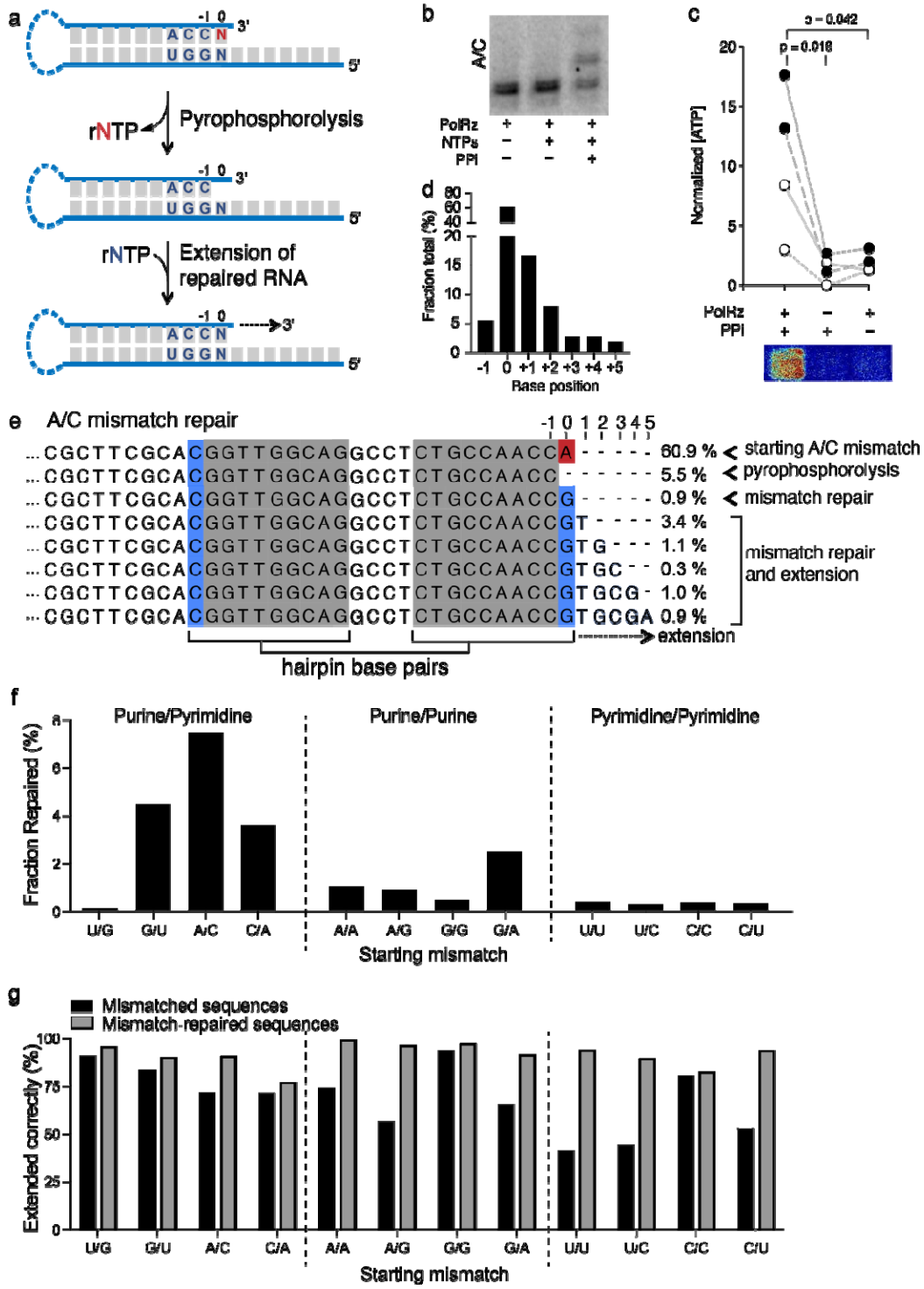

Figure 2. Pyrophosphorolysis-mediated mismatch repair and ATP generation by a polymerase ribozyme. (a) Graphical representation of a 3′ terminal mismatch (shown in red) undergoing ribozyme-catalyzed pyrophosphorolysis and subsequent repair. (b) High-resolution PAGE analysis of a ^{32}P-labeled A/C mismatched hairpin. The left two lanes show no extension of the hairpin in the absence of nucleotides or pyrophosphate. Upon addition of pyrophosphate, the sequence can be extended (right lane). (c) ATP generated by the removal of the terminal AMP of an A/C-mismatched

hairpin. Reactions lacking either the polymerase ribozyme or pyrophosphate show only background levels of luminescence by firefly luciferase. Open and filled circles represent two independent RNA preparations. Gray lines link experiments performed side-by-side from the same component stocks. Representative luminescence image is shown below. (d) Distribution of sequences extended by the ribozyme in the presence of pyrophosphate (mismatched and repaired), determined by HTS. (e) HTS analysis of sequences of the A/C-mismatched hairpin after incubation with the polymerase ribozyme and PPi. The sequences show the starting A/C mismatch (red A), the removal of the mismatched nucleotide by phosphorolysis (second row sequence, comprising 5.5% of total hairpin reads), the addition of the correct nucleotide (blue G at position 0), and the extension of the corrected sequence (light-blue boxes at positions 1–5). Total percentage of sequences extended are shown in d. (f) Fraction of each mismatch repaired through ribozyme-catalyzed pyrophosphorolysis, grouped by base-pair type. (g) Fidelity of extensions, comparing the fractions of the correctly extended sequences in sequences containing a mismatch (black bars) and the correctly extended sequences that are extended after mismatch repair (gray bars), indicating the fidelity of all sequences that follow the mismatch-repaired position. In all cases, pyrophosphate-mediated mismatch repair leads to an increase in the fidelity of extended sequences. The large differences between the dark and light bars imply that unrepaired mismatches lead to further errors in RNA polymerization.

Addition of pyrophosphate to extension reactions of a long template results in increased length and fidelity

Based on our demonstration that a polymerase ribozyme can repair hairpin sequences terminating in a mismatch and the mechanism for that repair, we hypothesized that pyrophosphate-mediated repair would suppress mutations in long extension reactions of primers, leading to increased fidelity in the polymerization of long templates. To do this, we used a primer–template system that can result in a templated polymerization of 117 nucleotides (Fig. 3a). We performed the ribozyme–catalyzed primer extension reactions for either three hours or four days at varying concentrations of pyrophosphate, and the RNAs were analyzed using HTS. As the concentration of pyrophosphate increased, we observed an increase in the average length of the primer extensions, although the total number of sequences decreased overall (Fig. 3b), consistent with pyrophosphate inhibition of polymerization. As hypothesized, the sequencing data showed an increase in fidelity at saturating concentrations of pyrophosphate (Fig. 3c,d). Further analysis revealed that for most extensions, sequences resulting from early termination had mismatches at the 3′-termini, suggesting that the mismatches caused the

polymerization to abort (Fig. S9). We observed a negative correlation between the number of sequences and their length, suggesting a general trade-off between copying fidelity and the number of extension products.

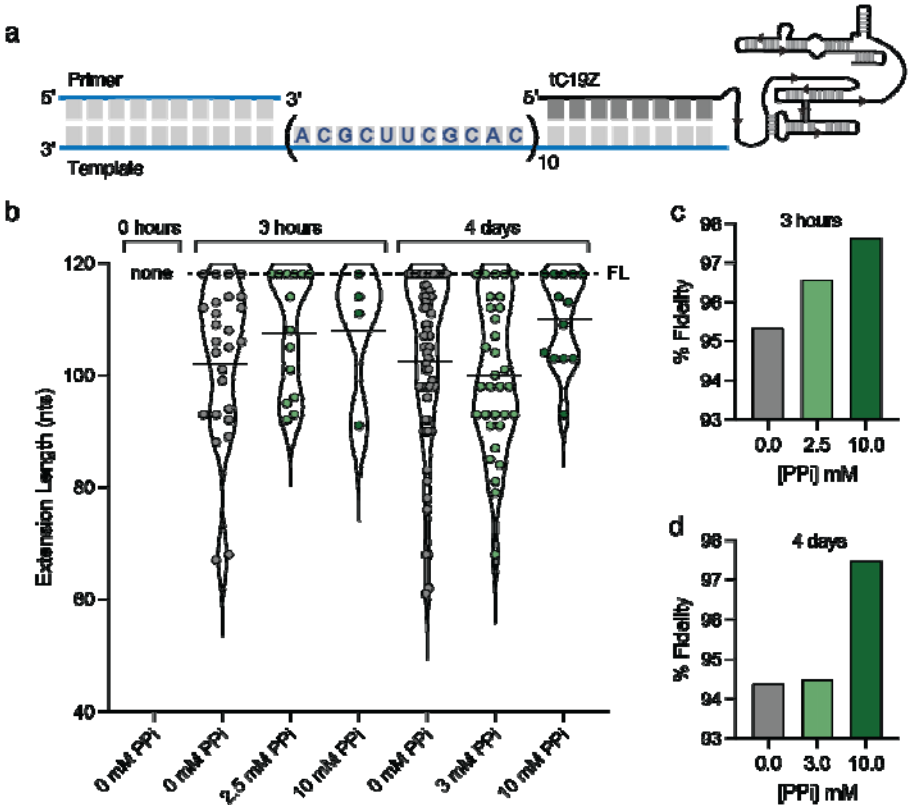

Figure 3. Effect of pyrophosphate on RNA polymerization along a long template. (a) The primer–template construct used for long extension experiments. (b) Plots showing the length of sequences resulting from polymerase-mediated extension of the primer at varying concentrations of pyrophosphate at either three hours or four days. Black bar indicates average length of sequences. FL indicates full length sequences. (c,d) Graphs showing the overall fidelity of extension products at varying concentrations of pyrophosphate at either three hours or four days. All analyses are based on HTS data. At ~10 mM, Mg-PPi is above saturating concentration and is partly precipitated.

Pyrophosphate-mediated repair of 2′-3′-cyclic-phosphate-terminated RNA sequences

When the phosphodiester of the RNA backbone undergoes nucleophilic attack by an adjacent 2′ oxyanion, the RNA strand is cleaved, generating sequences terminated with a 2′-3′-cyclic phosphate (2′-3′-cP); this damage is the most common type of RNA degradation. This cyclization reaction also occurs in self-cleaving ribozymes, many RNases, and base-catalyzed

RNA degradation,[41–43] and is responsible for the relatively low stability of RNA, when compared with DNA or other nucleic acids that lack a phosphate-adjacent hydroxyl group. Because the 2′-3′-cP acts as a chain terminator and prevents the 3′ hydroxyl from participating in extension during polymerization, it represents a replication dead end (Fig. 4a,b). Thus, we tested the polymerase ribozyme for pyrophosphate-mediated damage repair of an RNA hairpin terminated with a 2′-3′-cP. The cyclic phosphate can be removed enzymatically by T4 polynucleotide kinase (PNK), which we used as a positive control for testing the extension efficiencies of the hairpin substrates (Fig. S1).[37,38]

We prepared RNA hairpins terminated with each of the four canonical nucleotides base-paired to the matching nucleotide on the template strand and terminated with a 2′-3′-cP. As expected, no polymerization by the ribozyme was observed when extension of the 2′-3′-cP-damaged sequence was attempted. However, upon addition of pyrophosphate, the terminal damaged nucleotide was removed, and each of the sequences could be extended by the polymerase ribozyme (Fig. 4c-g, and Figs. S10–S12). The sequences could also be extended after removal of the 2′-3′-cP by PNK. Both extension and repair were largely dependent on the sequence, with the U/A hairpin undergoing repair the most efficiently.

We hypothesized that the mechanism of damage repair proceeds through pyrophosphorolysis of the 3′ terminal 2′-3′-cP-damaged nucleotide, as described earlier in the context of mismatch repair. Therefore, we analyzed the pyrophosphate-mediated removal of the terminal nucleotide by high-resolution PAGE but did not observe a gel mobility change that could be unequivocally attributed to removal of a single nucleotide from the 2′-3′-cP-terminated hairpin (see SI note 1). To better understand pyrophosphate-mediated repair of damaged RNAs, we began by performing repair experiments that could indirectly indicate the removal of the terminal nucleotide. First, we omitted the rNTP that corresponded to the damaged nucleotide in the extension reaction. For example, in the case of an A/U hairpin, we omitted ATP from the rNTPs

mixture, because following excision of the terminal AMP-2′-3′-cP, the ribozyme would need to replace it with undamaged AMP to continue the extension reaction. As expected, we found that if ATP was not present in the reaction, the sequence was not extended (Fig. S13). To gather additional evidence of damage repair, we transcribed hairpins without ^{32}P and performed the polymerization reaction with α-^{32}P-NTPs to incorporate the ^{32}P tracer. We observed that the α-^{32}P-NTP is only utilized in sequences that have been repaired, and although the ribozyme can incorporate it as a mismatched nucleotide during the extension, this occurs with lower efficiency (Fig. S14). Finally, we observed no reaction in the presence of only uridine and pyrophosphate, indicating that a non-canonical reaction, such as opening of the 2′-3′-cP by pyrophosphate to form a triphosphate that can subsequently react with unphosphorylated uridine, did not take place (Fig. S15).

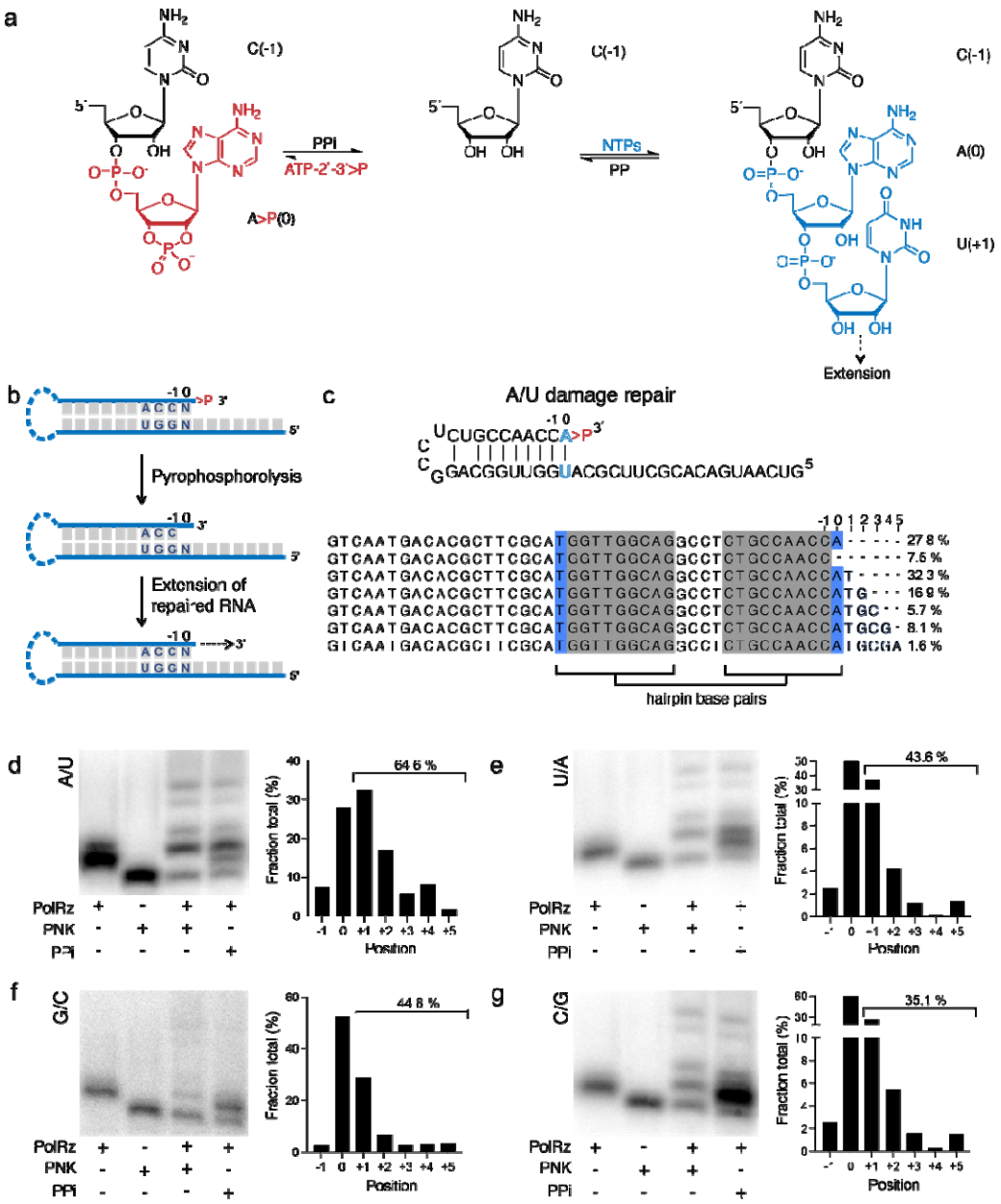

Figure 4. Pyrophosphate-facilitated repair of an RNA damage site by a polymerase ribozyme. (a) A hairpin substrate terminated with a 2′-3′-cyclic phosphate (2′-3′-cP; red nucleotide) cannot be extended by the polymerase ribozyme. However, upon removal of the terminal adenosine by ribozyme-mediated pyrophosphorolysis, the revealed 3′-OH becomes available for further addition of a nucleoside (blue) from a "correct" NTP. (b) Graphical representation of a hairpin with 2′-3′-cP damage and subsequent repair *via* pyrophosphorolysis. (c) Full sequence of an A/U hairpin and aligned reads from high-throughput sequencing analysis showing removal of the terminal adenosine (dark blue), and subsequent extension up to five bases (light blue). Percentages on the right indicate the fraction of each sequence in the dataset. The numbering above the alignment represents terminal nucleotides defined in a. (d-g) Analysis of ^{32}P-ATP-labeled 2′-3′-cP-damaged hairpins terminated in all four base-pairs; terminal base-pairs are indicated alongside each PAGE analysis (left-hand image of each set). Graphs on the right of each panel show the percentage of each sequence as determined

from HTS data, starting with sequences in which the damaged terminal nucleotide has been removed (fraction of sequences ending at position –1) and proceeding to extensions of the repaired RNA (0–5). The starting RNA containing the 2′-3′-cP cannot be ligated to the primer used for sequencing and is therefore not included in this analysis.

Repair of both cyclic-phosphate damage and mismatches

For certain hairpins, we observed both damage and mismatch repair upon addition of pyrophosphate. Specifically, the purine/purine mismatches (A/A, G/A, and A/G) and the G·U wobble pairs (G/U and U/G) could be repaired by the ribozyme when pyrophosphate was added to the extension reaction of these 2′-3′-cP-terminated mismatches (Fig. S16). These reactions exemplify a repair and copying of RNA strands that were mismatched, cleaved, or degraded and formed imperfect primer–template pairs either in *cis* or in *trans*, enabling replication of novel segments or duplication of existing segments as a pathway in genetic innovation.

Discussion

Modern biological systems possess distinct complexes and pathways dedicated to repairing different types of genome damage that could lead to degradation of the genetic information. Early replicating systems in an RNA world would not have had those functionalities or enzymatic diversity available to them, but damage and mismatch repair would likely have been necessary to replicate long genomes. It has been widely hypothesized that for known polymerase ribozymes to truly achieve efficient self-replication capability, an increase in the current standard of fidelity of replication is necessary. Though early studies on the first polymerase ribozymes hypothesized that pyrophosphorolysis could improve fidelity, this reactivity was not detected.[44] We address this gap by demonstrating repair activity in the presence of pyrophosphate by an evolved polymerase ribozyme originally selected to extend a primer along a template in an effort to generate a system of self-sustained replication.[18] We show that the tC19Z polymerase ribozyme can repair mismatched sequences, which presents a possible solution to the fidelity problems faced by early replication ribozymes: in a reaction with a long extension product, we

observed an increase in the average length of sequences and an overall higher fidelity in the presence of millimolar pyrophosphate. In addition, tC19Z can use pyrophosphate to repair 2′-3′-cyclic-phosphate-damaged RNA sequences that are formed *via* base-catalyzed RNA cleavage. Although many ribozymes exhibit varying levels of replication efficiency, under saturating pyrophosphate conditions, tC19Z is the first ribozyme with a demonstrated capability of repairing RNA.

Although our findings established that tC19Z can repair terminal mismatched sequences, the extent of ribozyme-mediated repair is highly dependent on mismatch identity, and distinct trends are present. Previously published findings suggest that less-evolved polymerase ribozymes (more closely resembling the class I ligase) also exhibit nucleotide preferences and add mismatches or extend mismatches with varying efficiencies.[15] Our observations on single nucleotide incorporation and patterns of extended sequences show that tC19Z has preferences for some nucleotides and sequences over others. For example, the ribozyme will add to uridine nucleotides in preference to other nucleotides, and similarly will repair damaged uridine nucleotides with greater frequency. Based on the mismatch repair trends, we hypothesize that there is a structural foundation for the trends in ribozyme-mediated extension of mismatched sequences and the effect of pyrophosphate on extension. Specifically, the width of the terminal base-pair appears to affect the recognition and polymerization capabilities of the ribozyme [45–49]. Although a recent cryo-EM structure has provided new insight into the structural basis for the function of polymerase ribozymes, additional structural data for this ribozyme-hairpin RNA complex are needed to definitively conclude what the ramifications of mismatches are on the ribozyme's polymerization reactivity.[50]

The relationship of sequence length and replication fidelity was originally outlined by Eigen, stating that as conserved sequence length increases so must replication fidelity to maintain functionality over successive rounds of copying, and hypothesizing that a mechanism of repair

would be critical in an evolving world.[51,52] Our sequencing analysis reveals a connection between downstream fidelity and mismatch repair: a substantial increase the downstream fidelity of extended sequences is observed when mismatches are corrected. Some mismatches negatively affect the fidelity of downstream polymerization more prominently than others, indicating that if the ribozyme inserts one incorrect base, it is more likely to add additional mismatches. For example, if the U/C mismatched hairpin is not repaired, over 60% of subsequently extended sequences have a second mismatch in the +1 position, and the majority involve addition of a cytosine to make a C/A mismatch. Furthermore, in the long extension experiments, shorter sequences present in the data primarily exhibited a varying number of mismatches at their 3′-end. Ultimately, compounding downstream mismatches leads to termination of replication, which would result in genome replication catastrophe in an evolving world. By improving fidelity through a ribozyme-mediated repair mechanism, we show a new functionality accessible to an evolving system that can directly address the theoretical limits on replicating longer sequences that maintain their function.

Our study establishes that a single ribozyme can perform several distinct biochemical functions, likely with the same active site.[16,19,23] Such economy of function may be a hallmark of emerging genome-encoded functional macromolecules before dedicated synthesis and repair machinery evolved. These results bring us one step closer to generating truly self-replicating systems that could have evolved in an RNA world.

Materials and Methods

<u>Cloning tC19Z into a pUC19 vector.</u>

The insert, tC19Z, was synthesized and obtained from Integrated DNA Technologies and then was subsequently amplified with primers containing the restriction enzyme sites BamHI and EcoRI (all enzymes were obtained from New England BioLabs) (**Table S1**; sequence #1 is the ribozyme construct).[18] The tC19Z insert and pUC19 vector were double

digested with BamHI and EcoRI restriction enzymes for 4 hours. Additionally, the pUC19 vector was further treated with calf-intestinal, alkaline phosphatase (CIP) for 1 hour following the double digestion. The insert was ligated into the vector using T4 DNA ligase and incubating at 16 ºC for 16 hours. The resulting ligation was transformed into DH5α competent cells and subsequently plated on ampicillin-containing agar plates which were grown at 37 ºC for 14 hours. Single colonies were picked and grown in 3 mL of ampicillin containing LB broth for 14 hours. Cultures were subjected to alkaline lysis and the plasmids were purified. The presence of tC19Z in the vector was confirmed by Sanger sequencing. Future polymerase chain reactions (PCR) of tC19Z were performed fresh from the plasmid for each batch of ribozyme.

<u>PCR and transcription of tC19Z and generation of hairpin constructs</u>.

PCR was performed to isolate tC19Z from the plasmid prepared above. The plasmid (5 ng), each primer (0.5 µM) (**Table S1**; primers # 4 and 5), and *Taq* DNA polymerase were used to amplify the sequence at an annealing temperature of 49 ºC. PCR products of tC19Z from the plasmid prepared above were used for *in vitro* transcription. Transcription was performed under the following conditions: 50 mM Tris HCl, 2 mM spermidine, 12 mM $MgCl_2$, 8 mM DTT, 2 mM each rNTP, and 0.01% Triton X-100 in a total volume of 100 µL. T7 RNA polymerase was added to the reaction and allowed to react at 37 ºC for 3 hours. The transcriptions were desalted with G15 Sephadex and purified using PAGE [18% w/v acrylamide/bisacrylamide 19:1 solution, 7 M urea, and 0.5x TBE Buffer (50 mM tris(hydroxymethyl)aminomethane hydrochloride, 45 mM boric acid, and 0.5 mM EDTA, pH 8.4)]. RNA was detected by UV shadowing and then was (i) eluted in 300 µL of 300 mM KCl from gel pieces, (ii) precipitated by adding 700 µL of ethanol, and (iii) incubated at –20 ºC overnight. The samples were then centrifuged to pellet the precipitate, which was subsequently dried in air and resuspended in water.

To prepare the RNA hairpin substrates for the tC19Z polymerase ribozyme (**Table S1**), the hairpin sequences (0.5 µM primers #7–#22; **Table S1**) were generated as fusions with the drz-Mbac-1 HDV-like ribozyme (0.5 µM; #6 construct; **Table S1**) using primer extension.[36] Following amplification, the sequence were transcribed in the presence of α-^{32}P[ATP] (PerkinElmer) by adding 50 mM Tris HCl, 2 mM spermidine, 12 mM $MgCl_2$, 8 mM DTT, 2 mM GTP, 2 mM CTP, 2 mM UTP, 200 µM ATP, and 0.01% Triton X-100 to a 50 µL reaction containing T7 RNA polymerase, and the mixture was allowed to react at 37 ºC for 2 hours. The HDV-like ribozyme self-cleaves co-transcriptionally to yield cyclic-phosphate-terminated hairpins. The transcriptions were desalted with Sephadex G10 spin columns (10 mM Tris, 1 mM EDTA, pH 7.2) purified using PAGE, and worked up as described above to generate internally ^{32}P-labeled polymerase substrates.

<u>Adding ^{32}P-phosphate to an RNA primer.</u>

An RNA primer 10 nucleotides in length was obtained from Dharmacon (primer #2; **Table S1**). A ^{32}P radiolabel was added to the 5′-end of the sequence by adding the RNA primer at a concentration of 5 µM to a solution containing γ-^{32}P[ATP], 1x T4 polynucleotide kinase buffer, and 5 units T4 polynucleotide kinase. The reaction was incubated at 37 °C for 2 hours, purified using PAGE, and worked up as described above.

<u>Kinetic analysis of pyrophosphate inhibition of RNA primer extension.</u>

Kinetic data for pyrophosphate inhibition were obtained by monitoring the extension of a ^{32}P-labeled primer by tC19Z over time and varying concentrations of pyrophosphate. Equimolar concentrations (0.5 µM) of tC19Z, primer, and template were mixed and heated to 82 ºC for 2 minutes, then cooled slowly to 17 ºC for 10 minutes. Varying concentrations of pyrophosphate were then added to the reactions followed by buffer containing 200 mM $MgCl_2$, 50 mM Tris HCl buffer (pH 8.3), and 4 mM rNTPs, and the solution was then allowed to incubate at 17 ºC.

Aliquots (5 µL) were removed at various time points and were quenched with 250 mM EDTA and 7 M urea gel-loading buffer. The reactions were analyzed using an 18% denaturing PAGE to obtain single-nucleotide-resolution images. The gels were imaged on a Typhoon 9410 Variable Mode Imager, and the pixel density of the bands was analyzed using ImageJ to generate the kinetic data. Initial rates were calculated using the first four time points and used to demonstrate pyrophosphate inhibition.

Repair of cyclic phosphate damaged RNA hairpins.

Hairpin substrate repair was visualized and quantified by imaging the extension of an internally ^{32}P-labeled hairpin by tC19Z. For negative controls, equimolar concentrations (0.5 µM) of tC19Z and 2′-3′-cyclic-phosphate-terminated hairpin were mixed and heated to 82 ºC for 2 minutes, and then cooled slowly to 17 ºC for 10 minutes. At this point, one of two buffers was added (1st buffer: 200 mM $MgCl_2$ and 50 mM Tris buffer; 2nd buffer: 200 mM $MgCl_2$, 50 mM Tris buffer, and 4 mM rNTPs), and the reactions were incubated at 17 ºC for 24 hours. The reactions were quenched with 250 mM EDTA and 7 M urea gel-loading buffer, and then were analyzed on a 15% PAGE gel. For the repair reaction, equimolar concentrations (0.5 µM) of tC19Z and cyclic-phosphate-terminated hairpin were mixed and heated to 82 ºC for 2 minutes, followed by slow cooling to 17 ºC for 10 minutes and subsequent addition of 10 mM pyrophosphate. A buffer containing 200 mM $MgCl_2$, 50 mM Tris buffer, and 4 mM rNTPs was added, and the reaction was allowed to incubate at 17 ºC for 24 hours. The reactions were quenched with 250 mM EDTA and 7 M urea gel-loading buffer and were analyzed on a 15% denaturing polyacrylamide gel. To generate a positive control sample, the cyclic phosphate was removed from the hairpin by treating with polynucleotide kinase in 1x PNK buffer (NEB) for 1 hour at 37 ºC then inactivated at 65 ºC for 10 minutes. The PNK-repaired hairpin substrate was then added to the tC19Z polymerase ribozyme, both at concentrations of 0.5 µM, and the solution was heated to 82 ºC for 2 minutes and then cooled slowly to 17 ºC for 10 minutes. Next, one of two buffers

was added (1st buffer: 200 mM $MgCl_2$ and 50 mM Tris buffer; 2nd buffer: 200 mM $MgCl_2$, 50 mM Tris buffer, and 4 mM rNTPs), and the reactions were allowed to incubate at 17 ºC for 24 hours. The reactions were quenched with 250 mM EDTA and 7 M urea gel-loading buffer and were analyzed on a 15% denaturing polyacrylamide gel. The gels were imaged on a Typhoon 9410 Variable Mode Imager.

Kinetic analysis of cyclic-phosphate damage repair.

The kinetics of tC19Z-mediated extension of pyrophosphate-repaired hairpin, PNK-repaired hairpin, and pyrophosphate inhibited PNK-repaired hairpin were obtained under similar conditions as those described above. Either hairpin or PNK-repaired hairpin was incubated with tC19Z at 82 ºC for 2 minutes and then cooled slowly to 17 ºC for 10 minutes. For the reactions containing pyrophosphate-repaired hairpin and pyrophosphate-inhibited hairpin, pyrophosphate was added at a concentration of 10 mM. A buffer containing 200 mM $MgCl_2$, 50 mM Tris buffer, and 4 mM rNTPs was added to each reaction, and the solutions were allowed to incubate at 17 ºC. Time points were removed from the reactions in 5 µL aliquots and quenched with 250 mM EDTA and 7 M urea gel-loading buffer. The reactions were analyzed on a 15% denaturing polyacrylamide gel that was imaged on a Typhoon 9410 Variable Mode Imager. ImageJ was used to quantify band density and generate the kinetic data for tC19Z extension of hairpins under varying pyrophosphate and repair conditions.

Hairpin migration and extension under various conditions.

To better understand hairpin migration and reactivity, an A/U hairpin was either left untreated or allowed to react with an alkaline phosphatase [5 µM hairpin, 1x Cutsmart buffer (NEB), and shrimp alkaline phosphatase (NEB)] for 30 mins at 37 ºC. The two conditions were then either left untreated, treated with PNK as described above, or incubated with acid (25 mM HCl) followed by quenching. Each solution was then analyzed on a 15% denaturing

polyacrylamide gel. Furthermore, each treated hairpin (0.5 µM) was heated with tC19Z (0.5 µM) for 2 minutes at 85 ºC and then cooled to 17 ºC over 10 minutes. A buffer containing 200 mM $MgCl_2$ and 50 mM Tris HCl (and, in some reactions, 4 mM rNTPs) was added. Additionally, 4 mM uridine and/or 10 mM pyrophosphate were added to some reactions. The mixtures were incubated at 17 ºC for 24 hours followed by quenching with 250 mM EDTA and 7 M urea gel-loading buffer. The reactions were analyzed on a 15% denaturing polyacrylamide gel that was imaged on a Typhoon 9410 Variable Mode Imager.

<u>Incorporation of ^{32}P-rNMPs into non-labeled, repaired hairpins.</u>

To further demonstrate repair, an A/U hairpin was transcribed with an a-^{32}P[ATP] or without (cold). Some of the hairpins were then treated with PNK as described above. The ^{32}P-labeled hairpin (0.5 µM) was incubated with tC19Z (0.5 µM) at 82 ºC for 2 mins and cooled to 17 ºC over 10 minutes. A buffer containing 200 mM $MgCl_2$ and 50 mM Tris HCl (and, in some reactions, 4 mM rNTPs) was added. To some reactions was added either 10 mM pyrophosphate or a-^{32}P[ATP] and 10 mM pyrophosphate. The cold hairpin was treated in the same way. The mixtures were incubated at 17 ºC for 24 hours followed by quenching with 250 mM EDTA and 7 M urea gel-loading buffer. The reactions were analyzed on a 15% denaturing polyacrylamide gel that was imaged on a Typhoon 9410 Variable Mode Imager.

A U/A hairpin substrate was transcribed without α -^{32}P-ATP and either left untreated or allowed to react with PNK as described above. The hairpin (0.5 µM) was heated with tC19Z (0.5 µM) for 2 minutes at 85 ºC and then cooled to 17 ºC over 10 minutes. A buffer containing 200 mM $MgCl_2$, 50 mM Tris HCl and, in some reactions, 10 mM pyrophosphate was added. Additionally, either α -^{32}P[ATP], α -^{32}P[UTP], or α -^{32}P[CTP] was added. The reactions were incubated at 17 ºC for 24 hours followed by quenching with 250 mM EDTA and 7 M urea gel-loading buffer. The reactions were analyzed on a 15% denaturing polyacrylamide gel that was imaged on a Typhoon 9410 Variable Mode Imager.

Extension of mismatched RNA hairpin sequences.

Mismatched RNA hairpins were extended by the tC19Z polymerase ribozyme under conditions of PNK repair or pyrophosphate. Pyrophosphate facilitated extension of both a cyclic-phosphate-damaged hairpin and a mismatched hairpin when 0.5 µM hairpin was added to 0.5 µM tC19Z. The solution was incubated at 82 ºC for 2 minutes and then cooled slowly to 17 ºC for 10 minutes. Pyrophosphate was added at a final concentration of 10 mM, followed by a buffer containing 200 mM $MgCl_2$, 50 mM Tris buffer, and 4 mM rNTPs, and the solution was incubated at 17 ºC for 24 hours. A reaction containing a PNK-repaired mismatched hairpin and tC19Z at concentrations of 0.5 µM each was incubated at 82 ºC for 2 minutes and then cooled slowly to 17 ºC for 10 minutes. Pyrophosphate was added at a concentration of 10 mM, followed by a buffer containing 200 mM $MgCl_2$, 50 mM Tris buffer, and 4 mM rNTPs, and the reaction was allowed to proceed at 17 ºC for 24 hours. All reactions were quenched with 250 mM EDTA and 7 M urea gel-loading buffer and then analyzed on a 15% denaturing polyacrylamide gel that was imaged on a Typhoon 9410 Variable Mode Imager. ImageJ was used to quantify band density to generate fraction extended graphs.

High-throughput sequencing of RNA hairpins.

RNA hairpins were analyzed by Illumina Sequencing to determine sequence identity. Cyclic phosphate repair and mismatch extension reactions, as described above, were performed and fractionated by PAGE. The images obtained were analyzed for extension products, and the corresponding bands were excised from the gel, eluted, and precipitated with ethanol. The RNA sequences were resuspended in water and then ligated to a DNA sequence containing an Illumina forward adapter (#24 and 25, **Table S1**). The DNA adapter sequence was designed with a 5′-phosphate, followed by four random nucleotides to reduce ligation bias at the 3′-RNA terminus. The adapter also contained a 3′-amine to prevent self-ligation. The oligo (10 µM) was

added to a mixture containing PEG 8000, 1x T4 RNA ligase buffer, 15 µM ATP, and T4 RNA ligase 1 to a total volume of 100 µL in order to adenylylate the 5′-end of the DNA sequence (5′-App oligo). G15 Sephadex was used to desalt and remove residual ATP from the 5′-App oligo.

To anneal the sequences, the 5′-App oligo was added in approximately three times excess of the RNA hairpin, and the resulting solution was heated to 95 ºC for 3 mins and then cooled to 25 ºC at a rate of 1 ºC/min. To ligate, the annealed oligos were added to 1x T4 RNA ligase buffer, PEG 8000, and T4 RNA ligase 2, truncated KQ to a total volume of 20 µL, and allowed to react at 25 ºC for 18 hours and then at 4 ºC for 2 hours.

To a 5 µL aliquot of the ligation reaction was added 2.5 µM of the RT primer (#29 in **Table S1**), which was annealed at 75 ºC for 3 mins, 37 ºC for 10 mins, and 25 ºC for 10 mins. Following annealing, 1x ProtoScript II buffer, 0.5 mM each dNTP, 5 mM $MgCl_2$, 10 mM DTT, and ProtoScript II were added and the reverse transcription reaction was heated to 50 ºC for 2 hours to generate cDNA.

The Illumina forward adapter was added to the cDNA through a PCR reaction (using primer #31 in **Table S1**). Forward and reverse primers (#31 and #30; 1 µM) were added to the cDNA, as well as 1x Q5 reaction buffer, 5 mM of each dNTP, and Q5 DNA polymerase. The reaction was annealed at 52 ºC and extended at 72 ºC. The Illumina barcodes were added in a second PCR reaction containing 1 µM of each primer, 1x Q5 buffer, 5 mM each dNTP, and Q5 DNA polymerase. The reaction to install the Illumina barcodes was annealed at 65 ºC and extended at 72 ºC, and the reaction was purified via agarose gel. The libraries were submitted for Illumina Sequencing. PhiX was removed from the resulting reads using bowtie2, and the remaining reads were trimmed using CutAdapt to remove adapter regions. The 5′-adapter (TCTACACTCTTTCCCTACACGACGCTCTTCCGATCT) was removed first, then the 3′-adapter (GATCGGAAGAGCACACGTCTGAACTCCAGTCAC) with up to at most 10% error in single-end mode. Sequences were counted in FastAptamer using fastaptamer_count. Starting and

extended sequences were extracted from data for analysis in FastAptamer using fastaptamer_search for each sequence with unique identity and length.[53] Specifically, the two nucleotide tag (AC or GT) incorporated from sequences #24 or #25, followed by four random N nucleotides, and ending with the sequence of interest were searched for and compiled. These sequences were then curated visually in jalview. Sequence identity was analyzed at each position and the number of sequences in each case were used to calculate extension efficiency, pyrophosphorolysis, and fidelity (SI Note 2). The sequencing was performed on one representative reaction for each condition, as is convention in the field when determining polymerase ribozyme fidelity.

Long-read primer extension and sequencing.

For sequencing the long-read experiments, we prepared an RNA template corresponding to a sequence previously used with the tC19Z rRNA polymerase ribozyme .[18] The sequence was transcribed from the long-read DNA template (sequence #23 in **Table S1**) prepared as a double-stranded DNA using the template strand and a T7 RNAP primer-extended as described above, and PAGE-purified, as described above. The template RNA was combined with the RNA primer (sequence #2 in **Table S1**) and the tC19Z ribozyme at 1 µM concentration (each) and this mixture was incubated at various concentrations of pyrophosphate for 4 days at 17 ºC as described above. The reactions were desalted using Sephadex G10. The desalted samples were not further purified. The samples were then ligated to the ligation primer for sequencing (sequences #24 and #25 in **Table S1**) and reverse-transcribed as described above. The cDNAs were then ligated to the cDNA ligation hairpin (sequence #26 in **Table S1**), which contains a 7-nt random sequence to allow splint-ligation by a DNA ligase independent of the 3′ terminal sequence of the cDNA and allow amplification of all RNAs present in the samples, as described previously.[54] To suppress self-ligation, the cDNA ligation hairpin DNAs contained 3′

terminal carbon spacer. The samples were incubated in T4 DNA ligase buffer (NEB) with 1 mM ATP, 0.5 µL of T4 DNA ligase, 0.5 M betaine, 20% PEG 8000, and 0.5 µL of RNaseH (NEB). To preferentially amplify the ribozyme-extended RNA primers, a DNA primer that overlapped with the RNA primer sequence (CUGCCAACCG) extended by the ligated hairpin sequence was designed and used for PCR amplification (sequence #27 in **Table S1**) together with the RT primer for sequencing (#29 in **Table S1**). A nested PCR reaction was performed to further enrich for sequences of the RNA primer–template ribozyme extension reaction and introduced the Illumina reverse primer sequence (#28 in **Table S1**) and the reactions were sequenced as described above. Whereas reactions were designed to bias towards amplification of the RNA primer-extensions, we observed many other amplicons, including many sequences that corresponded to products of primer or hairpin self-ligations.

The libraries were sequenced by Illumina Sequencing, and the resulting reads were trimmed using CutAdapt to remove primer regions. Bowtie2 was used to align resulting sequences to the long template and samtools was used to convert the output file to a sorted .bam file that could be opened using Integrated Genome Viewer. Sequence lengths and fidelity were obtained using Jalview (SI Note 2). The sequencing was performed on one representative reaction for each condition, as is convention in the field when determining polymerase ribozyme fidelity.

<u>Measuring ATP concentrations using a firefly luciferase assay.</u>

A solution containing a 2-µM PNK-repaired A/C hairpin and 0.5-µM tC19Z polymerase ribozyme was incubated at 82 ºC for 2 minutes and then cooled slowly to 17 ºC for 10 minutes. Pyrophosphate was added at a concentration of 10 mM followed by a buffer containing 200 mM $MgCl_2$ and 50 mM Tris buffer, and the reaction was allowed to proceed at 17 ºC for 3 days. Control reactions were incubated as above, but in the absence of individual components, as indicated. After 3 days, the reactions were centrifuged, and 0.5 µL of the supernatant was

added to 4.5 µL of a solution containing inorganic pyrophosphatase and firefly luciferase (preincubated for 3 h in presence of D-luciferin to deplete any ATP co-purified with the enzyme). Luminescence was generated using the StayBrite Highly Stable ATP Bioluminescence Assay Kit (BioVision), as described previously.[39] The reaction contained 50 µM D-luciferin, 1 µL of 4x StayBrite reaction buffer, and 1 µL StayBrite luciferase enzyme, in addition to 0.5 µL of inorganic pyrophosphatase (NEB). Each reaction was loaded in a well of a black 384-well plate and imaged using a high numerical aperture lens (Voigtländer 29 mm F/0.8) mounted on an EMCCD camera (iXon Ultra 888, Andor). The temperature of the camera electronics was maintained at –100 ºC during detection using a thermoelectric recirculating chiller (Thermocube 300, Solid State Cooling Systems). During imaging, 10 one-minute exposures were collected. The images were then stacked, a Z-projection was generated, and relative luminescence was measured using ImageJ. To generate a standard curve, different ATP concentrations in nM range were incubate in the same buffer system, including pyrophosphate. For the no-pyrophosphate control, pyrophosphate was added to the samples at the end of the experiment and the samples were treated so that the same amount of soluble pyrophosphate was present in each sample when incubated with the pyrophosphatase and luciferase enzymes.

References

1. Leslie E., O. Prebiotic Chemistry and the Origin of the RNA World. *Crit. Rev. Biochem. Mol. Biol.* **39**, 99–123 (2004).
2. Robertson, M. P. & Joyce, G. F. The origins of the RNA world. *Cold Spring Harb. Perspect. Biol.* **4**, a003608 (2012).
3. Schrum, J. P., Zhu, T. F. & Szostak, J. W. The Origins of Cellular Life. *Cold Spring Harb. Perspect. Biol.* **2**, (2010).
4. Bartel, D. P. & Unrau, P. J. Constructing an RNA world. *Trends Cell Biol.* **9**, 9–13 (1999).
5. Joyce, G. F. The antiquity of RNA-based evolution. *Nature* **418**, 214–221 (2002).
6. Ma, W. & Yu, C. Intramolecular RNA replicase: Possibly the first self-replicating molecule in the RNA world. *Orig. Life Evol. Biosph.* **36**, 413–420 (2006).
7. Chen, X., Li, N. & Ellington, A. D. Ribozyme Catalysis of Metabolism in the RNA World. *Chem. Biodivers.* **4**, 633–655 (2007).
8. Benner, S. A., Ellington, A. D. & Tauer, A. Modern metabolism as a palimpsest of the

RNA world. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 7054–7058 (1989).

9. Neveu, M., Kim, H.-J. & Benner, S. A. The “Strong” RNA World Hypothesis: Fifty Years Old. *Astrobiology* **13**, 391–403 (2013).

10. Ekland, E. H., Szostak, J. W. & Bartel, D. P. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science* **269**, 364–370 (1995).

11. Bartel, D. P. & Szostak, J. W. Isolation of new ribozymes from a large pool of random sequences [see comment]. *Science* **261**, 1411–1418 (1993).

12. Ekland, E. H. & Bartel, D. P. The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res.* **23**, 3231–3238 (1995).

13. Jaeger, L., Wright, M. C. & Joyce, G. F. A complex ligase ribozyme evolved in vitro from a group I ribozyme domain. *Proc. Natl. Acad. Sci.* **96**, 14712–14717 (1999).

14. Shechner, D. M. *et al.* Crystal Structure of the Catalytic Core of an RNA-Polymerase Ribozyme. *Science* **326**, 1271–1275 (2009).

15. Ekland, E. H. & Bartel, D. P. RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* **382**, 373–376 (1996).

16. Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E. & Bartel, D. P. RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension. *Science* **292**, 1319–1325 (2001).

17. Samanta, B. & Joyce, G. F. A reverse transcriptase ribozyme. *Elife* **6**, e31153 (2017).

18. Wochner, A., Attwater, J., Coulson, A. & Holliger, P. Ribozyme-Catalyzed Transcription of an Active Ribozyme. *Science* **332**, 209–212 (2011).

19. Horning, D. P. & Joyce, G. F. Amplification of RNA by an RNA polymerase ribozyme. *Proc. Natl. Acad. Sci.* **113**, 9786–9791 (2016).

20. Horning, D. P., Bala, S., Chaput, J. C. & Joyce, G. F. RNA-Catalyzed Polymerization of Deoxyribose, Threose, and Arabinose Nucleic Acids. *ACS Synth. Biol.* **8**, 955–961 (2019).

21. Razvan, C. & J., U. P. Processive RNA polymerization and promoter recognition in an RNA World. *Science* **371**, 1225–1232 (2021).

22. Tjhung, K. F., Shokhirev, M. N., Horning, D. P. & Joyce, G. F. An RNA polymerase ribozyme that synthesizes its own ancestor. *Proc. Natl. Acad. Sci.* **117**, 2906–2913 (2020).

23. Zaher, H. S. & Unrau, P. J. Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA* **13**, 1017–1026 (2007).

24. Attwater, J. *et al.* Chemical fidelity of an RNA polymerase ribozyme. *Chem. Sci.* **4**, 2804–2814 (2013).

25. Papastavrou, N., Horning, D. P. & Joyce, G. F. RNA-catalyzed evolution of catalytic RNA. *Proc. Natl. Acad. Sci.* **121**, e2321592121 (2024).

26. Beate, S., Rana, S., Kiong, H. C. & Stewart, S. Portability and fidelity of RNA-repair systems. *Proc. Natl. Acad. Sci.* **101**, 2788–2793 (2004).

27. Burroughs, A. M. & Aravind, L. RNA damage in biological conflicts and the diversity of responding RNA repair systems. *Nucleic Acids Res.* **44**, 8525–8555 (2016).

28. Nandakumar, J., Schwer, B., Schaffrath, R. & Shuman, S. RNA Repair: An Antidote to Cytotoxic Eukaryal RNA Damage. *Mol. Cell* **31**, 278–286 (2008).

29. Kamerlin, S. C. L., Sharma, P. K., Prasad, R. B. & Warshel, A. Why nature really chose phosphate. *Q. Rev. Biophys.* **46**, 1–132 (2013).

30. Pasek, M. A. Rethinking early Earth phosphorus geochemistry. *Proc. Natl. Acad. Sci.* **105**, 853–858 (2008).

31. Miller, S. L. & Parris, M. Synthesis of pyrophosphate under primitive earth conditions. *Nature* **204**, 1248–1250 (1964).

32. Hørder, M. Complex formation of inorganic pyrophosphate with magnesium: The influence of ionic strength, supporting medium and temperature. *Biochim. Biophys. Acta - Enzymol.* **358**, 319–328 (1974).

33. Holm, N. G. & Baltscheffsky, H. Links Between Hydrothermal Environments, Pyrophosphate, Na+, and Early Evolution. *Orig. Life Evol. Biosph.* **41**, 483–493 (2011).

34. Li, G.-M. Mechanisms and functions of DNA mismatch repair. *Cell Res.* **18**, 85–98 (2008).

35. Iyer, R. R., Pluciennik, A., Burdett, V. & Modrich, P. L. DNA mismatch repair: functions and mechanisms. *Chem. Rev.* **106**, 302–323 (2006).

36. Riccitelli, N. J., Delwart, E. & Lupták, A. Identification of Minimal HDV-Like Ribozymes with Unique Divalent Metal Ion Dependence in the Human Microbiome. *Biochemistry* **53**, 1616–1626 (2014).

37. Wang, L. K., Lima, C. D. & Shuman, S. Structure and mechanism of T4 polynucleotide kinase: an RNA repair enzyme. *EMBO J.* **21**, 3873–3880 (2002).

38. Das, U. & Shuman, S. Mechanism of RNA 2',3'-cyclic phosphate end healing by T4 polynucleotide kinase-phosphatase. *Nucleic Acids Res.* **41**, 355–365 (2013).

39. Chizzolini, F., Kent, A. D., Passalacqua, L. F. M. & Lupták, A. Enzymatic RNA Production from NTPs Synthesized from Nucleosides and Trimetaphosphate. *ChemBioChem* **22**, 2098–2101 (2021).

40. Leu, K. *et al.* Cascade of Reduced Speed and Accuracy after Errors in Enzyme-Free Copying of Nucleic Acid Sequences. *J. Am. Chem. Soc.* **135**, 354–366 (2013).

41. Jimenez, R. M., Polanco, J. A. & Lupták, A. Chemistry and Biology of Self-Cleaving Ribozymes. *Trends Biochem. Sci.* **40**, 648–661 (2015).

42. Li, Y. & Breaker, R. R. Kinetics of RNA Degradation by Specific Base Catalysis of Transesterification Involving the 2'-Hydroxyl Group. *J. Am. Chem. Soc.* **121**, 5364–5372 (1999).

43. Seith, D. D., Bingaman, J. L., Veenis, A. J., Button, A. C. & Bevilacqua, P. C. Elucidation of Catalytic Strategies of Small Nucleolytic Ribozymes from Comparative Analysis of Active Sites. *ACS Catal.* **8**, 314–327 (2018).

44. Glasner, M. E., Yen, C. C., Ekland, E. H. & Bartel, D. P. Recognition of Nucleoside Triphosphates during RNA-Catalyzed Primer Extension. *Biochemistry* **39**, 15556–15562 (2000).

45. Cruse, W. B. *et al.* Structure of a mispaired RNA double helix at 1.6-A resolution and implications for the prediction of RNA secondary structure. *Proc. Natl. Acad. Sci.* **91**, 4160–4164 (1994).

46. Pan, B., Mitra, S. N. & Sundaralingam, M. Crystal Structure of an RNA 16-mer Duplex R(GCAGAGUUAAAUCUGC)2 with Nonadjacent G(Syn)·A+(Anti) Mispairs. *Biochemistry* **38**, 2826–2831 (1999).

47. Sugimoto, N., Kierzek, R. & Turner, D. H. Sequence dependence for the energetics of terminal mismatches in ribooligonucleotides. *Biochemistry* **26**, 4559–4562 (1987).

48. Sydow, J. F. *et al.* Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol. Cell* **34**, 710–721 (2009).

49. Kahn, J. D. & Hearst, J. E. Reversibility of nucleotide incorporation by Escherichia coli RNA polymerase, and its effect on fidelity. *J. Mol. Biol.* **205**, 291–314 (1989).

50. McRae, E. K. S. *et al.* Cryo-EM structure and functional landscape of an RNA polymerase ribozyme. *Proc. Natl. Acad. Sci.* **121**, e2313332121 (2024).

51. Eigen, M. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**, 465–523 (1971).

52. Eigen, M., McCaskill, J. & Schuster, P. Molecular quasi-species. *J. Phys. Chem.* **92**, 6881–6891 (1988).

53. Alam, K. K., Chang J. L. & Burke, D. H. FASTAptamer: A bioinformatic toolkit for high-throughput sequence analysis of combinatorial selections. *Mol. Ther. Nucleic Acids* **4**, e230 (2015).

54. Latifi, B., Cole, K. H., Vu, M. M. K., Lupták, A. Rapid discovery of functional RNA domains. *Nucleic Acids Res.* **53**, gkaf307 (2025).

Acknowledgements

We thank the staff at the Genomics High-Throughput Facility (GHTF) Shared Resource of the Cancer Center, Support Grant (P30CA-062203) at the University of California, Irvine and NIH shared instrumentation grants 1S10RR025496-01, 1S10OD010794-01, and 1S10OD021718-01; and S. Flynn for assistance with sequencing data. We thank C. Langeberg and D. Treco for comments on the manuscript, and J. H. D. Cate for his support. This work was supported by the John Templeton Foundation through a grant to the Foundation for Applied Molecular Evolution (to A. L.), NSF CBET Biophotonics 1804220 (to A. L.), NASA Exobiology 80NSSC21K0488 (to A. L.), NASA ICAR 80NSSC21K0596 (to A.L.), NSF GRFP (to A. K.), Goldwater Scholarship (to L. Y.), and UCI Undergraduate Research Opportunities Program awards (to L. Y. and I. L.).

Affiliations

Department of Chemistry, University of California, Irvine, Irvine, CA, USA
Alexandra D. Kent, Lucy L. Yang, and Andrej Lupták

Department of Pharmaceutical Sciences, University of California, Irvine, Irvine, CA, USA
Alexandra D. Kent and Andrej Lupták

Department of Molecular Biology and Biochemistry, University of California, Irvine, Irvine, CA, USA
Kyle H. Cole and Andrej Lupták

Current Affiliations

California Institute for Quantitative Biosciences, University of California at Berkeley, Berkeley, CA, USA

Alexandra D. Kent

Department of Chemistry, Harvard University; Cambridge, MA, USA

Lucy L. Yang

Amgen; Thousand Oaks, CA, USA

Kyle H. Cole

Author contributions

A.D.K. and A.L. conceived and designed the study. A.D.K. performed the experiments with assistance from L.L.Y., K.H.C. and A. L. The data were analyzed by A.D.K. and A.L. The manuscript was prepared by A.D.K and A.L.

Data and materials availability: All data are available in the main text, the supplementary materials, or deposited in a sequence repository.