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, including the ribozyme itself. 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, and even include copies of the polymerase ribozyme. 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

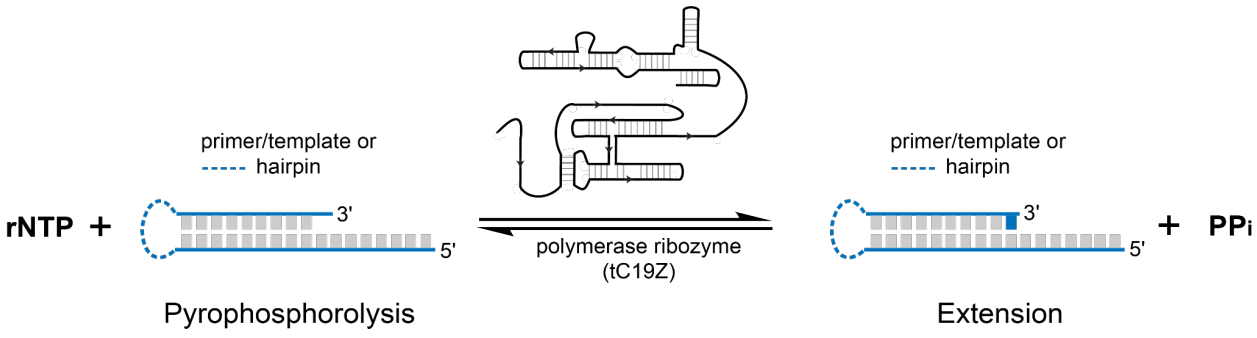


Figure 1. Reversible RNA synthesis by the polymerase ribozyme. The RNA polymerase ribozyme, tC19Z, catalyzes templated addition of a nucleotide to the 3′-end of a primer–template RNA with pyrophosphate (PPi) as a byproduct of the reaction. The reverse reaction, pyrophosphorolysis of RNA, removes the 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] Inspired by a previous study that proposed pyrophosphorolysis was possible with a class I ligase derived ribozyme, 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.[34]

Pyrophosphate-mediated mismatch repair

91 To maintain fidelity of genome copying at non-deleterious levels, modern replication machinery
92 uses distinct active sites or entire protein complexes to correct mismatched base-pairs, but at the
93 onset of evolvable systems, it is unlikely that mismatch repair was available as an independent
94 biochemical activity.[35,36] Thus, a simpler process for increasing replication fidelity would have
95 been a desirable feature of an early replicating polymerase. To test whether ribozyme-mediated
96 pyrophosphorolysis removes a mismatched nucleotide, we prepared hairpin sequences with 3′-
97 terminal mismatches. We fused the primer and template to form a hairpin (Fig. 1) that serves both
98 as a substrate for the tC19Z ribozyme at exact stoichiometry and as a practical construct for
99 downstream analysis using high-throughput sequencing. We generated the hairpins *via* an *in vitro*
100 transcription reaction of a self-cleaving HDV-like ribozyme (drz-Mbac-1) fused to the 3′ terminus
101 of the target sequence.[37] The ribozyme self-cleaves, yielding a 2′-3′-cyclic phosphate (2′-3′-cP)
102 on the 3′ end of the primer segment. The cyclic phosphate can be removed enzymatically by T4
103 polynucleotide kinase (PNK) (Extended Data Fig. 1).[38,39]

104 We began by characterizing the reactivity of a terminal A/C mismatch in the primer/template
105 sequence with the cyclic phosphate removed from the 3′ end. RNA polymerization by the
106 ribozyme was not observed upon addition of nucleoside triphosphates (NTPs), indicating that an
107 A/C mismatch acts as a chain-terminator (Fig. 2a,b). After addition of saturating concentrations
108 of pyrophosphate (10 mM, with clearly visible Mg-PPi precipitate) and NTPs, we observed
109 ribozyme-mediated extension of the sequence (Fig. 2), but we did not observe the -1 product by
110 denaturing polyacrylamide gel electrophoresis (PAGE), likely because the equilibrium highly
111 favors polymerization (SI Note 1, Fig. 2b, and Extended Data Fig. 2). When removed through
112 pyrophosphorolysis, the terminal nucleotide is converted into a nucleoside triphosphate, and an
113 A/C-terminated hairpin would produce a molecule of ATP. To obtain independent evidence of
114 pyrophosphorolysis, we detected this critical metabolite using an ultrasensitive firefly-luciferase-

115 based assay, which yields a photon for each ATP molecule present in the reaction.[40] Because
116 pyrophosphate inhibits the luciferase reaction, insoluble pyrophosphate was first removed from
117 the samples by centrifugation, and the remaining pyrophosphate was subsequently hydrolyzed
118 by addition of pyrophosphatase enzyme. The resulting samples were incubated with the firefly
119 luciferase and D-luciferin and were then imaged using an ultracooled EM-CCD camera. The
120 integrated luminescence was measured for both the pyrophosphorolysis reaction and control
121 reactions lacking one of the components. In the absence of either pyrophosphate, the ribozyme,
122 or the hairpin substrate, we observed a baseline level of luminescence. In reactions where all
123 components were present, the luminescence was significantly greater, indicating pyrophosphate-
124 mediated synthesis of ATP (Fig. 2c and Extended Data Fig. 3). These experiments provide
125 evidence that the mechanism of repair occurs by a pyrophosphorolysis reaction and demonstrate
126 the first instance of ribozyme-catalyzed production of ATP.

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

132 We repeated this analysis for each of the eleven other possible combinations of terminal
133 mismatches (Fig. 2 and Extended Data Figs. 2, 4, and 5). Sequence analysis of the extension
134 reactions correlated well with the gel electrophoresis analysis. In addition, the HTS data provided
135 a deeper insight into the reactions, revealing the fractions of the sequences that were repaired
136 and subsequently extended. PAGE analysis was consistent over replicate experiments. We found
137 that the efficiency of repair and extension depends on the identity of the mismatch:
138 purine/pyrimidine mismatches were repaired the most efficiently, with the exception of the U/G
139 wobble. Purine/purine mismatches were repaired and extended at a low frequency upon the

140 addition of pyrophosphate but were not extendable by the ribozyme in the absence of
141 pyrophosphate (Fig. 2f,g and Extended Data Fig. 5). In contrast, pyrimidine/pyrimidine
142 mismatches were extended well by the ribozyme, but most of the extension products retained the
143 mismatch (Fig. 2f,g and Extended Data Fig. 5). Sequences containing terminal C/A or G/U
144 mismatches behave like the A/C mismatch and only showed extension (with varying efficiencies)
145 after pyrophosphorolysis of the mismatch.

146 In addition to PAGE analyses, we obtained quantitative results for the fidelity of extension
147 products from the sequencing data. We measured the fidelity at each position of the extension
148 and found that for all mismatches, repair of the mismatched base led to an increase in the fidelity
149 of downstream extensions (Fig. 2f,g). In some cases, such as the pyrimidine/pyrimidine
150 mismatches, the effect is strongly pronounced: although the frequency of repair is low, when a
151 hairpin is repaired, the extension occurs with high fidelity. Specifically, in the case of the U/C
152 mismatch, extension of the terminal mismatch preferentially adds a second mismatch (C/A) which
153 then acts as a chain terminator (Extended Data Fig. 6). Therefore, we conclude that mismatched
154 sequences that do not terminate the extension are more likely to acquire additional mutations, as
155 has been observed in enzyme-free polymerization of nucleic acid sequences, implying a
156 mechanism for error propagation that must involve an allosteric modulation of the fidelity of
157 incoming nucleotide incorporation by upstream mismatches.[41]

158 We observed that addition of pyrophosphate inhibited the forward polymerization reaction in the
159 positive control (Supplementary Fig. 1). To quantify the pyrophosphate reaction inhibition, we
160 used a non-hairpin, primer-template complex to measure reaction kinetics of the forward
161 extension reaction. Initial rates were plotted as a function of substrate concentration and revealed
162 an approximate K_M of 4 mM with respect to NTPs (Extended Data Fig. 7A,B,C). We then
163 measured the reaction kinetics in the presence of pyrophosphate and plotted the initial rates of

tC19Z-mediated polymerization as a function of pyrophosphate concentration (Extended Data Fig. 7D,E). At millimolar PPi concentrations, the forward reaction was inhibited approximately five-fold, compared to when no PPi is present.

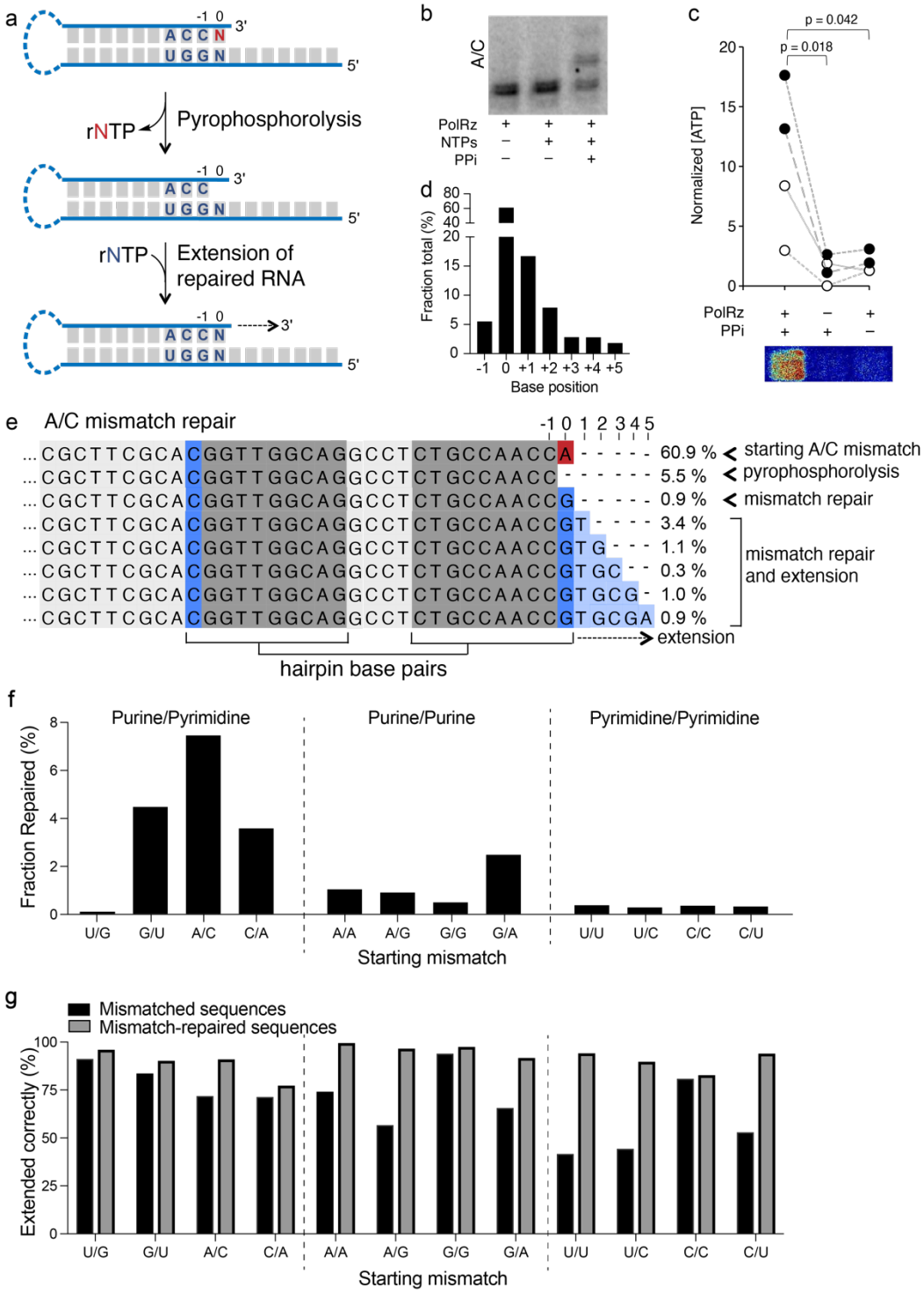


Figure 2. Pyrophosphorolysis-mediated mismatch repair and ATP generation by a polymerase ribozyme. (a) Graphical representation of a 3′ terminal mismatch (shown in red) undergoing ribozyme-catalyzed pyrophosphorolysis and subsequent repair. (b) High-resolution PAGE analysis of a ^{32}P-labeled A/C mismatched 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 118 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, Supplementary Table 3), 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 many sequences resulting from early termination had mismatches at the 3′-termini, suggesting that the mismatches caused the polymerization to abort (Supplementary Fig. 2). We observed a negative correlation between the number of sequences and their length in the three-hour experiment, suggesting a trade-off between copying fidelity and the number of extension products. We also observed complements of tC19Z sequences in the HTS data with strand-specific ligation products from sequencing library preparation, indicating that the sequences originated from off-target priming and extension of the polymerase ribozyme itself. Nearly full-length (98%) ribozyme sequences were present in samples containing pyrophosphate and exhibited higher fidelity, whereas no long-read RNA segments antisense to the ribozyme were detected in the no-pyrophosphate HTS dataset (Supplementary Fig. 3).

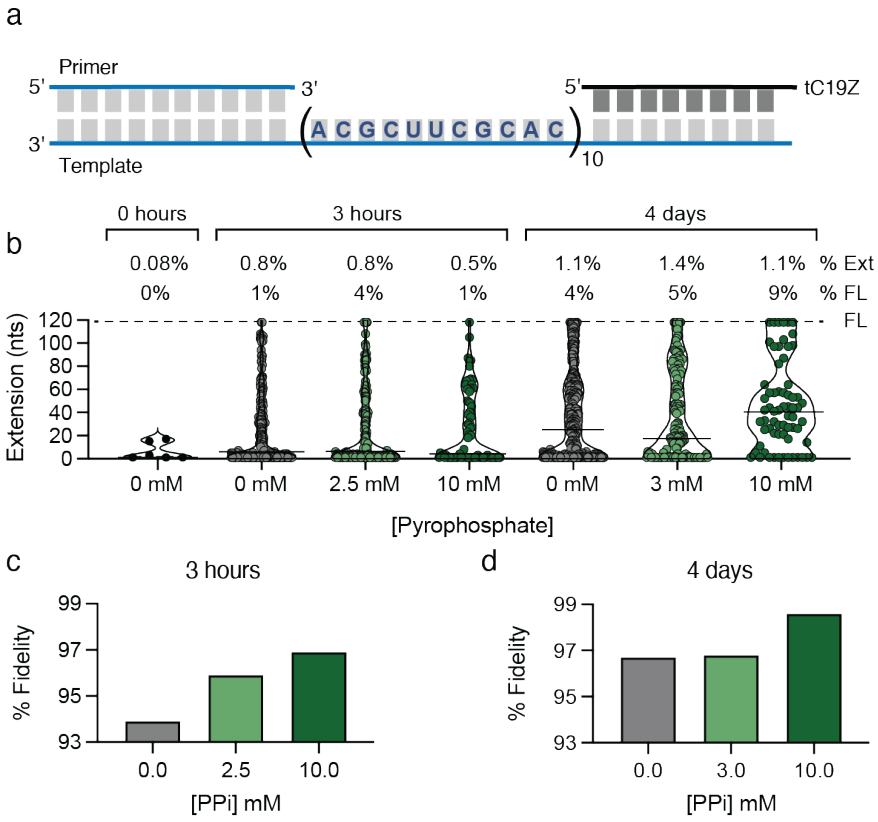


Figure 3. Effect of pyrophosphate on RNA polymerization along a long template. (a) The primer–template construct used for long extension experiments. (b) Plots showing the length of sequences resulting from polymerase-mediated extension of the primer at varying concentrations of pyrophosphate at either three hours or four days. Black horizontal bar indicates average length of sequences. FL dashed line indicates full length sequences. "% Ext" shows the percentage of sequences that were extended from primers, and % FL indicates the percentage of extended sequences that are full length. (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,[42–44] 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 (Extended Data Fig. 1).[38,39]

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, Extended Data Fig. 8–9, and Supplementary Fig. 4). The sequences could also be extended after removal of the 2′-3′-cP by PNK. Both extension and repair were largely dependent on the

247 sequence, with the U/A hairpin undergoing repair the most efficiently.

248 We hypothesized that the mechanism of damage repair proceeds through pyrophosphorolysis of
249 the 3′ terminal 2′-3′-cP-damaged nucleotide, as described earlier in the context of mismatch
250 repair. Therefore, we analyzed the pyrophosphate-mediated removal of the terminal nucleotide
251 by high-resolution PAGE but did not observe a gel mobility change that could be unequivocally
252 attributed to removal of a single nucleotide from the 2′-3′-cP-terminated hairpin (see SI note 1).
253 To better understand pyrophosphate-mediated repair of damaged RNAs, we began by performing
254 repair experiments that could indirectly indicate the removal of the terminal nucleotide. First, we
255 omitted the rNTP that corresponded to the damaged nucleotide in the extension reaction. For
256 example, in the case of an A/U hairpin, we omitted ATP from the rNTPs mixture, because following
257 excision of the terminal AMP-2′-3′-cP, the ribozyme would need to replace it with undamaged
258 AMP to continue the extension reaction. As expected, we found that if ATP was not present in the
259 reaction, the sequence was not extended (Supplementary Fig. 5). To gather additional evidence
260 of damage repair, we transcribed hairpins without ^{32}P and performed the polymerization reaction
261 with α-^{32}P-NTPs to incorporate the ^{32}P tracer. We observed that the α-^{32}P-NTP is only utilized in
262 sequences that have been repaired, and although the ribozyme can incorporate it as a
263 mismatched nucleotide during the extension, this occurs with lower efficiency (Supplementary
264 Fig. 6). Finally, we observed no reaction in the presence of only uridine and pyrophosphate,
265 indicating that a non-canonical reaction, such as opening of the 2′-3′-cP by pyrophosphate to form
266 a triphosphate that can subsequently react with unphosphorylated uridine, did not take place
267 (Supplementary Fig. 7).

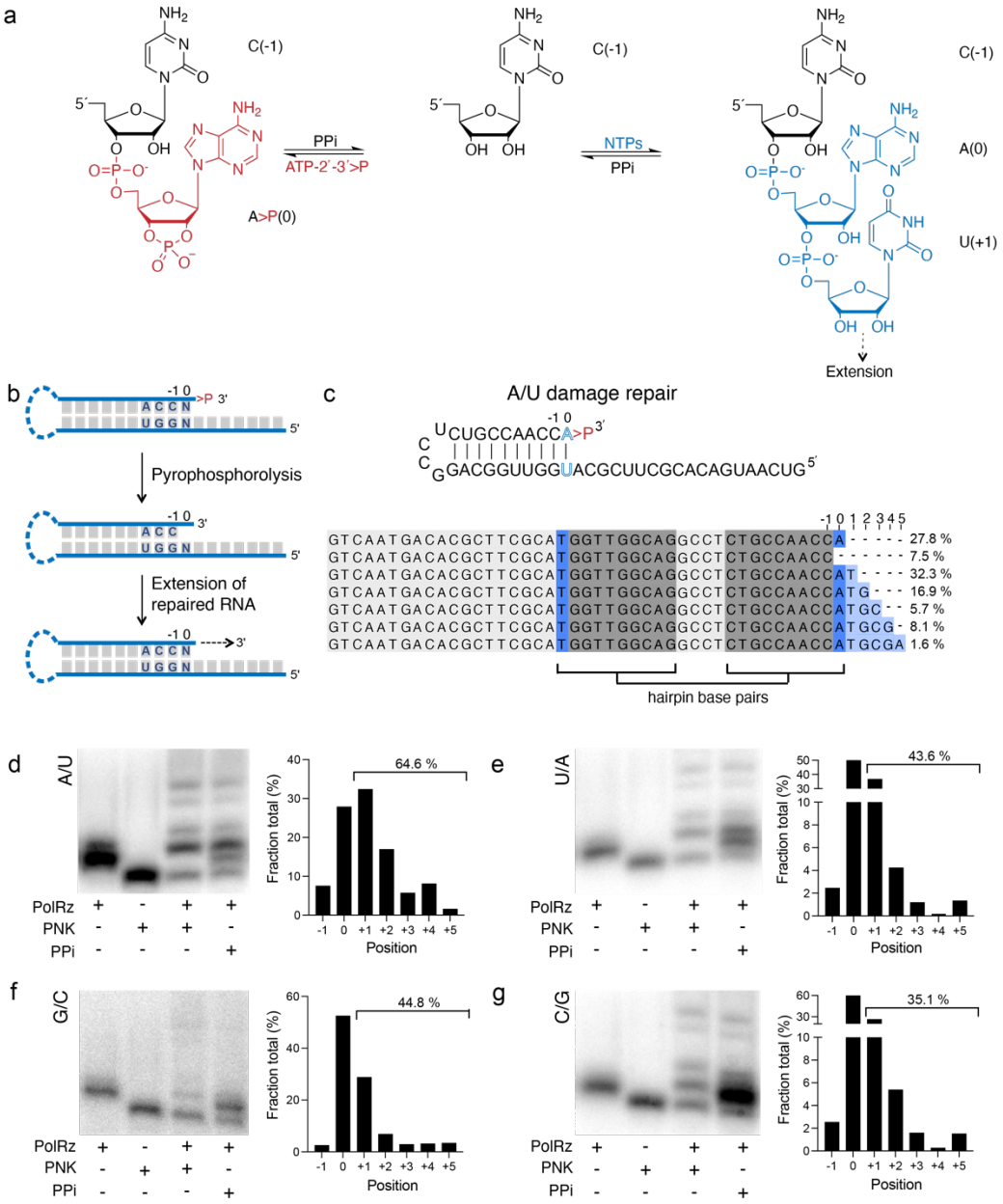


Figure 4. Pyrophosphate-facilitated repair of an RNA damage site by a polymerase ribozyme. (a) A hairpin substrate terminated with a 2′-3′-cyclic phosphate (2′-3′-cP; red nucleotide) cannot be extended by the polymerase ribozyme. However, upon removal of the terminal adenosine by ribozyme-mediated pyrophosphorolysis, 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

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

285 **Repair of both cyclic-phosphate damage and mismatches**

286 For certain hairpins, we observed both damage and mismatch repair upon addition of
287 pyrophosphate. Specifically, the purine/purine mismatches (A/A, G/A, and A/G) and the G·U
288 wobble pairs (G/U and U/G) could be repaired by the ribozyme when pyrophosphate was added
289 to the extension reaction of these 2′-3′-cP-terminated mismatches (Extended Data Fig. 10). These
290 reactions exemplify a repair and copying of RNA strands that were mismatched, cleaved, or
291 degraded and formed imperfect primer-template pairs either in *cis* or in *trans*, enabling replication
292 of novel segments or duplication of existing segments as a pathway in genetic innovation.

293 **Discussion**

294 Modern biological systems possess distinct complexes and pathways dedicated to repairing
295 different types of genome damage that could lead to degradation of the genetic information. Early
296 replicating systems in an RNA world would not have had those functionalities or enzymatic
297 diversity available to them, but damage and mismatch repair would likely have been necessary
298 to replicate long genomes. It has been widely hypothesized that for known polymerase ribozymes
299 to truly achieve efficient self-replication capability, an increase in the current standard of fidelity
300 of replication is necessary. Though early studies on the first polymerase ribozymes hypothesized
301 that pyrophosphorolysis could improve fidelity, this reactivity was not detected.[34] We address this
302 gap by demonstrating repair activity in the presence of pyrophosphate by an evolved polymerase
303 ribozyme originally selected to extend a primer along a template in an effort to generate a system
304 of self-sustained replication.[18] We show that the tC19Z polymerase ribozyme can repair
305 mismatched sequences, which presents a possible solution to the fidelity problems faced by early

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

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

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

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

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

350

351 **Materials and Methods**

352 Cloning tC19Z into a pUC19 vector.

353 The insert, tC19Z, was synthesized and obtained from Integrated DNA Technologies and
354 then was subsequently amplified with primers containing the restriction enzyme sites BamHI and
355 EcoRI (all enzymes were obtained from New England BioLabs) (**Supplementary Table 1**;
356 sequence #1 is the ribozyme construct).[18] The tC19Z insert and pUC19 vector were double
357 digested with BamHI and EcoRI restriction enzymes for 4 hours. Additionally, the pUC19 vector
358 was further treated with calf-intestinal, alkaline phosphatase (CIP) for 1 hour following the double
359 digestion. The insert was ligated into the vector using T4 DNA ligase and incubating at 16 ºC for
360 16 hours. The resulting ligation was transformed into DH5α competent cells and subsequently
361 plated on ampicillin-containing agar plates which were grown at 37 ºC for 14 hours. Single
362 colonies were picked and grown in 3 mL of ampicillin containing LB broth for 14 hours. Cultures
363 were subjected to alkaline lysis and the plasmids were purified. The presence of tC19Z in the
364 vector was confirmed by Sanger sequencing. Future polymerase chain reactions (PCR) of tC19Z
365 were performed fresh from the plasmid for each batch of ribozyme.

366

367 PCR and transcription of tC19Z and generation of hairpin constructs.

368 PCR was performed to isolate tC19Z from the plasmid prepared above. The plasmid (5
369 ng), each primer (0.5 μM) (**Supplementary Table 1**; primers # 4 and 5), and *Taq* DNA polymerase
370 were used to amplify the sequence at an annealing temperature of 49 ºC. PCR products of tC19Z
371 from the plasmid prepared above were used for *in vitro* transcription. Transcription was performed
372 under the following conditions: 50 mM Tris HCl, 2 mM spermidine, 12 mM $MgCl_2$, 8 mM DTT, 2
373 mM each rNTP, and 0.01% Triton X-100 in a total volume of 100 μL. T7 RNA polymerase was
374 added to the reaction and allowed to react at 37 ºC for 3 hours. The transcriptions were desalted
375 with G15 Sephadex and purified using PAGE [18% w/v acrylamide/bisacrylamide 19:1 solution,

376 7 M urea, and 0.5x TBE Buffer (50 mM tris(hydroxymethyl)aminomethane hydrochloride, 45 mM
377 boric acid, and 0.5 mM EDTA, pH 8.4)]. RNA was detected by UV shadowing and then was (i)
378 eluted in 300 μL of 300 mM KCl from gel pieces, (ii) precipitated by adding 700 μL of ethanol, and
379 (iii) incubated at –20 ºC overnight. The samples were then centrifuged to pellet the precipitate,
380 which was subsequently dried in air and resuspended in water.

381 To prepare the RNA hairpin substrates for the tC19Z polymerase ribozyme
382 (**Supplementary Table 1**), the hairpin sequences (0.5 μM primers #7–#22; **Supplementary**
383 **Table 1**) were generated as fusions with the drz-Mbac-1 HDV-like ribozyme (0.5 μM; #6 construct;
384 **Supplementary Table 1**) using primer extension.[37] Following amplification, the sequence were
385 transcribed in the presence of α-^{32}P[ATP] (PerkinElmer) by adding 50 mM Tris HCl, 2 mM
386 spermidine, 12 mM $MgCl_2$, 8 mM DTT, 2 mM GTP, 2 mM CTP, 2 mM UTP, 200 μM ATP, and
387 0.01% Triton X-100 to a 50 μL reaction containing T7 RNA polymerase, and the mixture was
388 allowed to react at 37 ºC for 2 hours. The HDV-like ribozyme self-cleaves co-transcriptionally to
389 yield cyclic-phosphate-terminated hairpins. The transcriptions were desalted with Sephadex G10
390 spin columns (10 mM Tris, 1 mM EDTA, pH 7.2) purified using PAGE, and worked up as described
391 above to generate internally ^{32}P-labeled polymerase substrates.

392

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

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

399

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

401 Kinetic data for pyrophosphate inhibition were obtained by monitoring the extension of a
402 ^{32}P-labeled primer by tC19Z over time and varying concentrations of pyrophosphate. Equimolar
403 concentrations (0.5 μM) of tC19Z, primer, and template were mixed and heated to 82 ºC for 2
404 minutes, then cooled slowly to 17 ºC for 10 minutes. Varying concentrations of pyrophosphate
405 were then added to the reactions followed by buffer containing 200 mM $MgCl_2$, 50 mM Tris HCl
406 buffer (pH 8.3), and 4 mM rNTPs, and the solution was then allowed to incubate at 17 ºC. Aliquots
407 (5 μL) were removed at various time points and were quenched with 250 mM EDTA and 7 M urea
408 gel-loading buffer. The reactions were analyzed using an 18% denaturing PAGE to obtain single-
409 nucleotide-resolution images. The gels were imaged on a Typhoon 9410 Variable Mode Imager,
410 and the pixel density of the bands was analyzed using ImageJ to generate the kinetic data. Initial
411 rates were calculated using the first four time points and used to demonstrate pyrophosphate
412 inhibition.
413
414 Repair of cyclic phosphate damaged RNA hairpins.
415 Hairpin substrate repair was visualized and quantified by imaging the extension of an
416 internally ^{32}P-labeled hairpin by tC19Z. For negative controls, equimolar concentrations (0.5 μM)
417 of tC19Z and 2′-3′-cyclic-phosphate-terminated hairpins were mixed and heated to 82 ºC for 2
418 minutes, and then cooled slowly to 17 ºC for 10 minutes. At this point, one of two buffers was
419 added (1st buffer: 200 mM $MgCl_2$ and 50 mM Tris buffer; 2nd buffer: 200 mM $MgCl_2$, 50 mM Tris
420 buffer, and 4 mM rNTPs), and the reactions were incubated at 17 ºC for 24 hours. The reactions
421 were quenched with 250 mM EDTA and 7 M urea gel-loading buffer, and then were analyzed on
422 a 15% PAGE gel. For the repair reaction, equimolar concentrations (0.5 μM) of tC19Z and cyclic-
423 phosphate-terminated hairpin were mixed and heated to 82 ºC for 2 minutes, followed by slow
424 cooling to 17 ºC for 10 minutes and subsequent addition of 10 mM pyrophosphate. A buffer
425 containing 200 mM $MgCl_2$, 50 mM Tris buffer, and 4 mM rNTPs was added, and the reaction was
426 allowed to incubate at 17 ºC for 24 hours. The reactions were quenched with 250 mM EDTA and

427 7 M urea gel-loading buffer and were analyzed on a 15% denaturing polyacrylamide gel. To
428 generate a positive control sample, the cyclic phosphate was removed from the hairpin by treating
429 with polynucleotide kinase in 1x PNK buffer (NEB) for 1 hour at 37 ºC then inactivated at 65 ºC
430 for 10 minutes. The PNK-repaired hairpin substrate was then added to the tC19Z polymerase
431 ribozyme, both at concentrations of 0.5 μM, and the solution was heated to 82 ºC for 2 minutes
432 and then cooled slowly to 17 ºC for 10 minutes. Next, one of two buffers was added (1st buffer:
433 200 mM $MgCl_2$ and 50 mM Tris buffer; 2nd buffer: 200 mM $MgCl_2$, 50 mM Tris buffer, and 4 mM
434 rNTPs), and the reactions were allowed to incubate at 17 ºC for 24 hours. The reactions were
435 quenched with 250 mM EDTA and 7 M urea gel-loading buffer and were analyzed on a 15%
436 denaturing polyacrylamide gel. The gels were imaged on a Typhoon 9410 Variable Mode Imager.
437
438 Kinetic analysis of cyclic-phosphate damage repair.
439 The kinetics of tC19Z-mediated extension of pyrophosphate-repaired hairpin, PNK-
440 repaired hairpin, and pyrophosphate inhibited PNK-repaired hairpin were obtained under similar
441 conditions as those described above. Either hairpin or PNK-repaired hairpin was incubated with
442 tC19Z at 82 ºC for 2 minutes and then cooled slowly to 17 ºC for 10 minutes. For the reactions
443 containing pyrophosphate-repaired hairpin and pyrophosphate-inhibited hairpin, pyrophosphate
444 was added at a concentration of 10 mM. A buffer containing 200 mM $MgCl_2$, 50 mM Tris buffer,
445 and 4 mM rNTPs was added to each reaction, and the solutions were allowed to incubate at 17
446 ºC. Time points were removed from the reactions in 5 μL aliquots and quenched with 250 mM
447 EDTA and 7 M urea gel-loading buffer. The reactions were analyzed on a 15% denaturing
448 polyacrylamide gel that was imaged on a Typhoon 9410 Variable Mode Imager. ImageJ was used
449 to quantify band density and generate the kinetic data for tC19Z extension of hairpins under
450 varying pyrophosphate and repair conditions.
451
452 Hairpin migration and extension under various conditions.

453 To better understand hairpin migration and reactivity, an A/U hairpin was either left
454 untreated or allowed to react with an alkaline phosphatase [5 μM hairpin, 1x Cutsmart buffer
455 (NEB), and shrimp alkaline phosphatase (NEB)] for 30 mins at 37 ºC. The two conditions were
456 then either left untreated, treated with PNK as described above, or incubated with acid (25 mM
457 HCl) followed by quenching. Each solution was then analyzed on a 15% denaturing
458 polyacrylamide gel. Furthermore, each treated hairpin (0.5 μM) was heated with tC19Z (0.5 μM)
459 for 2 minutes at 85 ºC and then cooled to 17 ºC over 10 minutes. A buffer containing 200 mM
460 $MgCl_2$ and 50 mM Tris HCl (and, in some reactions, 4 mM rNTPs) was added. Additionally, 4 mM
461 uridine and/or 10 mM pyrophosphate were added to some reactions. The mixtures were incubated
462 at 17 ºC for 24 hours followed by quenching with 250 mM EDTA and 7 M urea gel-loading buffer.
463 The reactions were analyzed on a 15% denaturing polyacrylamide gel that was imaged on a
464 Typhoon 9410 Variable Mode Imager.
465
466 Incorporation of ^{32}P-rNMPs into non-labeled, repaired hairpins.
467 To further demonstrate repair, an A/U hairpin was transcribed with an α-^{32}P[ATP] or
468 without (cold). Some of the hairpins were then treated with PNK as described above. The ^{32}P-
469 labeled hairpin (0.5 μM) was incubated with tC19Z (0.5 μM) at 82 ºC for 2 mins and cooled to 17
470 ºC over 10 minutes. A buffer containing 200 mM $MgCl_2$ and 50 mM Tris HCl (and, in some
471 reactions, 4 mM rNTPs) was added. To some reactions was added either 10 mM pyrophosphate
472 or α-^{32}P[ATP] and 10 mM pyrophosphate. The cold hairpin was treated in the same way. The
473 mixtures were incubated at 17 ºC for 24 hours followed by quenching with 250 mM EDTA and 7
474 M urea gel-loading buffer. The reactions were analyzed on a 15% denaturing polyacrylamide gel
475 that was imaged on a Typhoon 9410 Variable Mode Imager.
476 A U/A hairpin substrate was transcribed without α-^{32}P-ATP and either left untreated or
477 allowed to react with PNK as described above. The hairpin (0.5 μM) was heated with tC19Z (0.5
478 μM) for 2 minutes at 85 ºC and then cooled to 17 ºC over 10 minutes. A buffer containing 200 mM

479 $MgCl_2$, 50 mM Tris HCl and, in some reactions, 10 mM pyrophosphate was added. Additionally,
480 either α -^{32}P[ATP], α -^{32}P[UTP], or α -^{32}P[CTP] was added. The reactions were incubated at 17 ºC
481 for 24 hours followed by quenching with 250 mM EDTA and 7 M urea gel-loading buffer. The
482 reactions were analyzed on a 15% denaturing polyacrylamide gel that was imaged on a Typhoon
483 9410 Variable Mode Imager.
484
485 Extension of mismatched RNA hairpin sequences.
486 Mismatched RNA hairpins were extended by the tC19Z polymerase ribozyme under
487 conditions of PNK repair or pyrophosphate. Pyrophosphate facilitated extension of both a cyclic-
488 phosphate-damaged hairpin and a mismatched hairpin when 0.5 μM hairpin was added to 0.5 μM
489 tC19Z. The solution was incubated at 82 ºC for 2 minutes and then cooled slowly to 17 ºC for 10
490 minutes. Pyrophosphate was added at a final concentration of 10 mM, followed by a buffer
491 containing 200 mM $MgCl_2$, 50 mM Tris buffer, and 4 mM rNTPs, and the solution was incubated
492 at 17 ºC for 24 hours. A reaction containing a PNK-repaired mismatched hairpin and tC19Z at
493 concentrations of 0.5 μM each was incubated at 82 ºC for 2 minutes and then cooled slowly to 17
494 ºC for 10 minutes. Pyrophosphate was added at a concentration of 10 mM, followed by a buffer
495 containing 200 mM $MgCl_2$, 50 mM Tris buffer, and 4 mM rNTPs, and the reaction was allowed to
496 proceed at 17 ºC for 24 hours. All reactions were quenched with 250 mM EDTA and 7 M urea
497 gel-loading buffer and then analyzed on a 15% denaturing polyacrylamide gel that was imaged
498 on a Typhoon 9410 Variable Mode Imager. ImageJ was used to quantify band density to generate
499 fraction extended graphs.
500
501 High-throughput sequencing of RNA hairpins.
502 RNA hairpins were analyzed by Illumina Sequencing to determine sequence identity. Cyclic
503 phosphate repair and mismatch extension reactions, as described above, were performed and
504 fractionated by PAGE. The images obtained were analyzed for extension products, and the

505 corresponding bands were excised from the gel, eluted, and precipitated with ethanol. The RNA
506 sequences were resuspended in water and then ligated to a DNA sequence containing an Illumina
507 forward adapter (#24 and 25, **Supplementary Table 1**). The DNA adapter sequence was
508 designed with a 5′-phosphate, followed by four random nucleotides to reduce ligation bias at the
509 3′-RNA terminus. The adapter also contained a 3′-amine to prevent self-ligation. The oligo (10
510 μM) was added to a mixture containing PEG 8000, 1x T4 RNA ligase buffer, 15 μM ATP, and T4
511 RNA ligase 1 to a total volume of 100 μL to adenylylate the 5′-end of the DNA sequence (5′-App
512 oligo). G15 Sephadex was used to desalt and remove residual ATP from the 5′-App oligo.

513 To anneal the sequences, the 5′-App oligo was added in approximately three times excess
514 of the RNA hairpin, and the resulting solution was heated to 95 ºC for 3 mins and then cooled to
515 25 ºC at a rate of 1 ºC/min. To ligate, the annealed oligos were added to 1x T4 RNA ligase buffer,
516 PEG 8000, and T4 RNA ligase 2, truncated KQ to a total volume of 20 μL, and allowed to react
517 at 25 ºC for 18 hours and then at 4 ºC for 2 hours.

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

523 The Illumina forward adapter was added to the cDNA through a PCR reaction (using
524 primer #31 in **Supplementary Table 1**). Forward and reverse primers (#31 and #30; 1 μM) were
525 added to the cDNA, as well as 1x Q5 reaction buffer, 5 mM of each dNTP, and Q5 DNA
526 polymerase. The reaction was annealed at 52 ºC and extended at 72 ºC. The Illumina barcodes
527 were added in a second PCR reaction containing 1 μM of each primer, 1x Q5 buffer, 5 mM each
528 dNTP, and Q5 DNA polymerase. The reaction to install the Illumina barcodes was annealed at 65
529 ºC and extended at 72 ºC, and the reaction was purified via agarose gel. The libraries were
530 submitted for Illumina Sequencing. PhiX was removed from the resulting reads using bowtie2,

531 and the remaining reads were trimmed using CutAdapt to remove adapter regions. The 5′-adapter
532 (TCTACACTCTTTCCCTACACGACGCTCTTCCGATCT) was removed first, then the 3′-adapter
533 (GATCGGAAGAGCACACGTCTGAACTCCAGTCAC), with at most 10% error in single-end
534 mode. Sequences were counted in FastAptamer using fastaptamer_count. Starting and extended
535 sequences were extracted from data for analysis in FastAptamer using fastaptamer_search for
536 each sequence with unique identity and length.[53] Specifically, the two-nucleotide tag (AC or GT)
537 incorporated from sequences #24 or #25, followed by four random N nucleotides, and ending with
538 the sequence of interest were searched for and compiled. These sequences were then curated
539 visually in Jalview. Sequence identity was analyzed at each position and the number of
540 sequences in each case were used to calculate extension efficiency, pyrophosphorolysis, and
541 fidelity (SI Note 2). The sequencing was performed on one representative reaction for each
542 condition, as is convention in the field when determining polymerase ribozyme fidelity.
543
544 Long-read primer extension and sequencing.
545 For sequencing the long-read experiments, we prepared an RNA template corresponding
546 to a sequence previously used with the tC19Z rRNA polymerase ribozyme .[18] The sequence was
547 transcribed from the long-read DNA template (sequence #23 in **Supplementary Table 1**)
548 prepared as a double-stranded DNA using the template strand and a T7 RNAP primer-extended
549 as described above, and PAGE-purified, as described above. The template RNA was combined
550 with the RNA primer (sequence #2 in **Supplementary Table 1**) and the tC19Z ribozyme at 1 μM
551 concentration (each) and this mixture was incubated at various concentrations of pyrophosphate
552 for 4 days at 17 ºC as described above. The reactions were desalted using Sephadex G10. The
553 desalted samples were not further purified. The samples were then ligated to the ligation primer
554 for sequencing (sequences #24 and #25 in **Supplementary Table 1**) and reverse-transcribed as
555 described above. The cDNAs were then ligated to the cDNA ligation hairpin (sequence #26 in
556 **Supplementary Table 1**), which contains a 7-nt random sequence to allow splint-ligation by a

557 DNA ligase independent of the 3′ terminal sequence of the cDNA and allow amplification of all
558 RNAs present in the samples, as described previously.[54] To suppress self-ligation, the cDNA
559 ligation hairpin DNAs contained 3′ terminal carbon spacer. The samples were incubated in T4
560 DNA ligase buffer (NEB) with 1 mM ATP, 0.5 µL of T4 DNA ligase, 0.5 M betaine, 20% PEG 8000,
561 and 0.5 µL of RNaseH (NEB). To preferentially amplify the ribozyme-extended RNA primers, a
562 DNA primer that overlapped with the RNA primer sequence (CUGCCAACCG) extended by the
563 ligated hairpin sequence was designed and used for PCR amplification (sequence #27 in
564 **Supplementary Table 1**) together with the RT primer for sequencing (#29 in **Supplementary**
565 **Table 1**). A nested PCR reaction was performed to further enrich for sequences of the RNA
566 primer–template ribozyme extension reaction and introduced the Illumina reverse primer
567 sequence (#28 in **Supplementary Table 1**) and the reactions were sequenced as described
568 above. Whereas reactions were designed to bias towards amplification of the RNA primer-
569 extensions, we observed many other amplicons, including many sequences that corresponded to
570 products of primer or hairpin self-ligations.

571 The libraries were sequenced by Illumina Sequencing, and the resulting reads were
572 trimmed using CutAdapt to remove primer regions. Sequences were counted in FastAptamer
573 using fastaptamer_count. Sequences containing the primer sequence were extracted from the
574 data for analysis in FastAptamer using fastaptamer_search for each sequence and subsequently
575 visually aligned and curated using Jalview. Sequence lengths and fidelity were obtained using
576 Jalview (SI Note 2). The sequencing was performed on one representative reaction for each
577 condition, as is convention in the field when determining polymerase ribozyme fidelity.

578

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

580 A solution containing a 2-µM PNK-repaired A/C hairpin and 0.5-µM tC19Z polymerase
581 ribozyme was incubated at 82 ºC for 2 minutes and then cooled slowly to 17 ºC for 10 minutes.
582 Pyrophosphate was added at a concentration of 10 mM followed by a buffer containing 200 mM

583 $MgCl_2$ and 50 mM Tris buffer, and the reaction was allowed to proceed at 17 ºC for 3 days. Control
584 reactions were incubated as above, but in the absence of individual components, as indicated.
585 After 3 days, the reactions were centrifuged, and 0.5 µL of the supernatant was added to 4.5 µL
586 of a solution containing inorganic pyrophosphatase and firefly luciferase (preincubated for 3 h in
587 presence of D-luciferin to deplete any ATP co-purified with the enzyme). Luminescence was
588 generated using the StayBrite Highly Stable ATP Bioluminescence Assay Kit (BioVision), as
589 described previously.[40] The reaction contained 50 µM D-luciferin, 1 µL of 4x StayBrite reaction
590 buffer, and 1 µL StayBrite luciferase enzyme, in addition to 0.5 µL of inorganic pyrophosphatase
591 (NEB). Each reaction was loaded in a well of a black 384-well plate and imaged using a high
592 numerical aperture lens (Voigtländer 29 mm F/0.8) mounted on an EMCCD camera (iXon Ultra
593 888, Andor). The temperature of the camera electronics was maintained at –100 ºC during
594 detection using a thermoelectric recirculating chiller (Thermocube 300, Solid State Cooling
595 Systems). During imaging, 10 one-minute exposures were collected. The images were then
596 stacked, a Z-projection was generated, and relative luminescence was measured using ImageJ.
597 To generate a standard curve, different ATP concentrations in nM range were incubated in the
598 same buffer system, including pyrophosphate. For the no-pyrophosphate control, pyrophosphate
599 was added to the samples at the end of the experiment and the samples were treated so that the
600 same amount of soluble pyrophosphate was present in each sample when incubated with the
601 pyrophosphatase and luciferase enzymes.

602 **References**

603 1. Leslie E., O. Prebiotic Chemistry and the Origin of the RNA World. *Crit. Rev. Biochem.*
604 *Mol. Biol.* **39**, 99–123 (2004).

605 2. Robertson, M. P. & Joyce, G. F. The origins of the RNA world. *Cold Spring Harb.*
606 *Perspect. Biol.* **4**, a003608 (2012).

607 3. Schrum, J. P., Zhu, T. F. & Szostak, J. W. The Origins of Cellular Life. *Cold Spring Harb.*
608 *Perspect. Biol.* **2**, (2010).

609 4. Bartel, D. P. & Unrau, P. J. Constructing an RNA world. *Trends Cell Biol.* **9**, 9–13 (1999).

610 5. Joyce, G. F. The antiquity of RNA-based evolution. *Nature* **418**, 214–221 (2002).

611 6. Ma, W. & Yu, C. Intramolecular RNA replicase: Possibly the first self-replicating molecule
612 in the RNA world. *Orig. Life Evol. Biosph.* **36**, 413–420 (2006).

613 7. Chen, X., Li, N. & Ellington, A. D. Ribozyme Catalysis of Metabolism in the RNA World.
614 *Chem. Biodivers.* **4**, 633–655 (2007).

615 8. Benner, S. A., Ellington, A. D. & Tauer, A. Modern metabolism as a palimpsest of the
616 RNA world. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 7054–7058 (1989).

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

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

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

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

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

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

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

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

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

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

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

639 20. Horning, D. P., Bala, S., Chaput, J. C. & Joyce, G. F. RNA-Catalyzed Polymerization of
640 Deoxyribose, Threose, and Arabinose Nucleic Acids. *ACS Synth. Biol.* **8**, 955–961
641 (2019).

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

644 22. Tjhung, K. F., Shokhirev, M. N., Horning, D. P. & Joyce, G. F. An RNA polymerase
645 ribozyme that synthesizes its own ancestor. *Proc. Natl. Acad. Sci.* **117**, 2906–2913
646 (2020).

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

649 24. Attwater, J. *et al.* Chemical fidelity of an RNA polymerase ribozyme. *Chem. Sci.* **4**, 2804–
650 2814 (2013).

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

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

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

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

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

661 30. Pasek, M. A. Rethinking early Earth phosphorus geochemistry. *Proc. Natl. Acad. Sci.*
662 **105**, 853–858 (2008).

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

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

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

670 34. Glasner, M. E., Yen, C. C., Ekland, E. H. & Bartel, D. P. Recognition of
671 Nucleoside Triphosphates during RNA-Catalyzed Primer Extension. *Biochemistry* **39**,
672 15556–15562 (2000).

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

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

676 37. Riccitelli, N. J., Delwart, E. & Lupták, A. Identification of Minimal HDV-Like Ribozymes
677 with Unique Divalent Metal Ion Dependence in the Human Microbiome. *Biochemistry* **53**,
678 1616–1626 (2014).

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

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

683 40. Chizzolini, F., Kent, A. D., Passalacqua, L. F. M. & Lupták, A. Enzymatic RNA Production
684 from NTPs Synthesized from Nucleosides and Trimetaphosphate. *ChemBioChem* **22**,
685 2098–2101 (2021).

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

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

690 43. Li, Y. & Breaker, R. R. Kinetics of RNA Degradation by Specific Base Catalysis of
691 Transesterification Involving the 2'-Hydroxyl Group. *J. Am. Chem. Soc.* **121**, 5364–5372
692 (1999).

693 44. Seith, D. D., Bingaman, J. L., Veenis, A. J., Button, A. C. & Bevilacqua, P. C. Elucidation
694 of Catalytic Strategies of Small Nucleolytic Ribozymes from Comparative Analysis of

Active Sites. *ACS Catal.* **8**, 314–327 (2018).

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

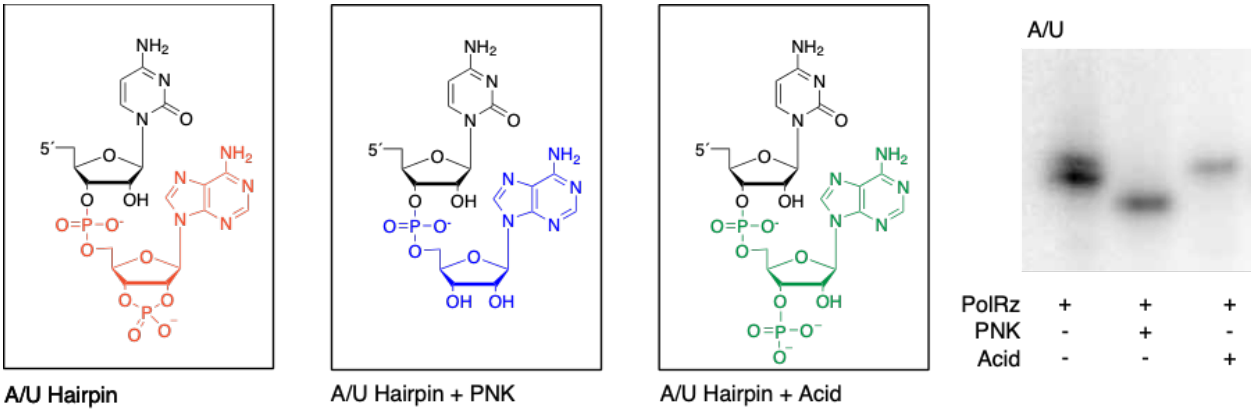
Kyle H. Cole

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.

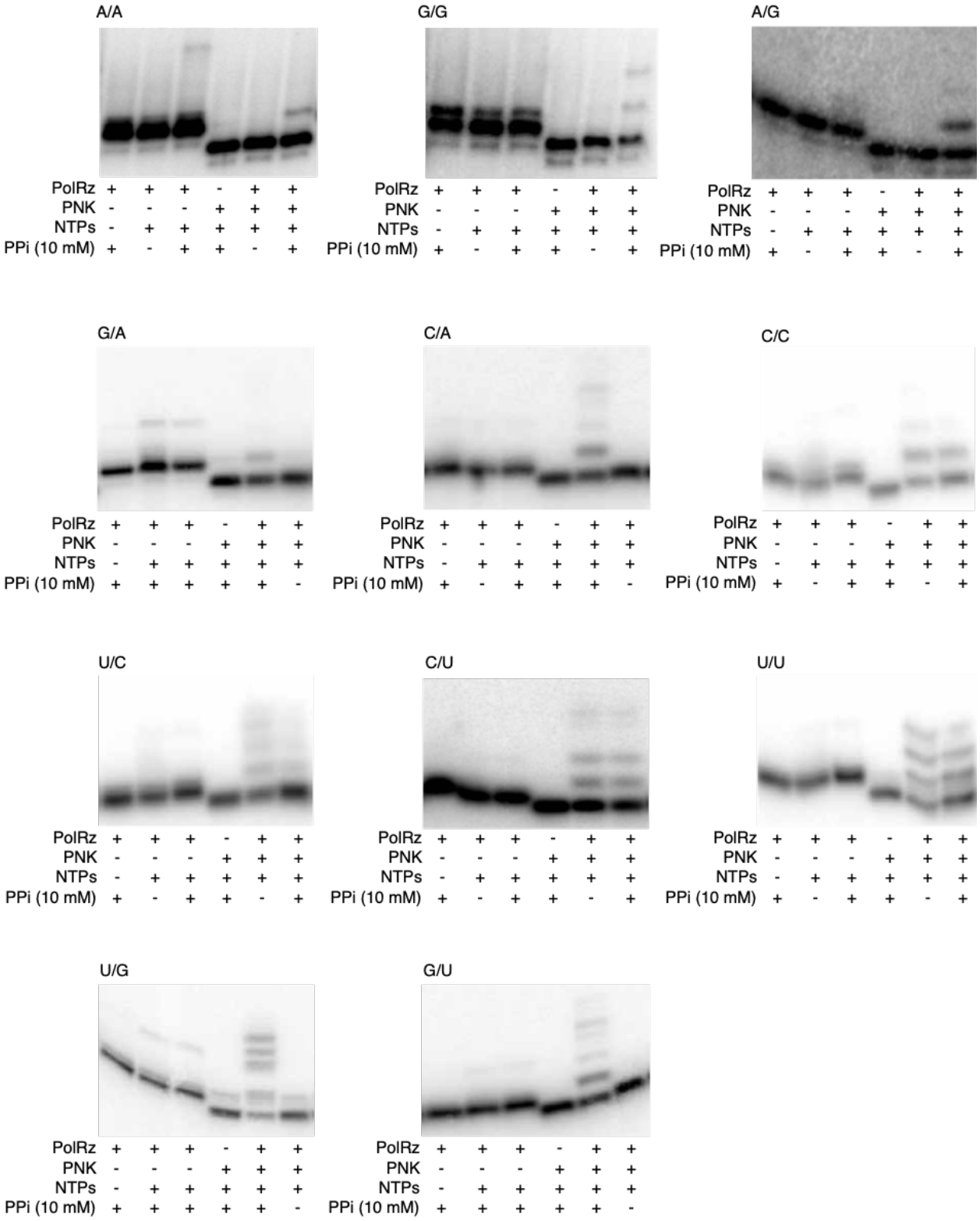
762 **Extended Data Figures**



763

764 **Extended Data Fig. 1. Gel mobility of A/U hairpin.** An A/U hairpin has a 2′-3′-cyclic phosphate
765 resulting from drz-Mbac-1 ribozyme self-scission (red and left lane on PAGE gel). After repair with
766 PNK, the cyclic phosphate is removed (blue), and the corresponding band migrates faster on the
767 PAGE gel (middle lane). When the 2′-3′-cyclic-phosphate-terminated hairpin is treated with
768 hydrochloric acid, the cyclic phosphate is partially hydrolyzed, yielding a 3′ monophosphate
769 (green) or 2′ monophosphate (not shown), and the mobility of the corresponding PAGE band is
770 retarded (right lane).

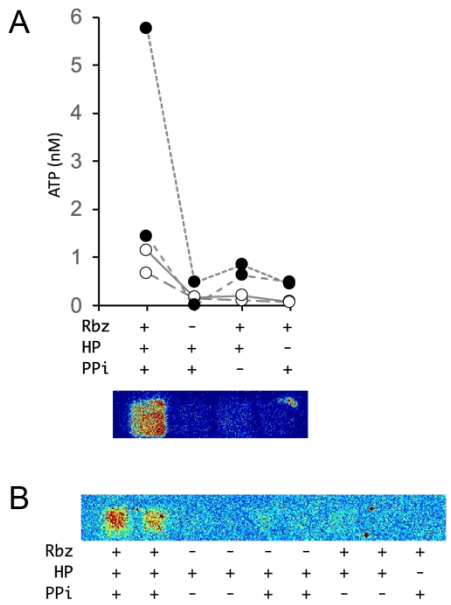
771



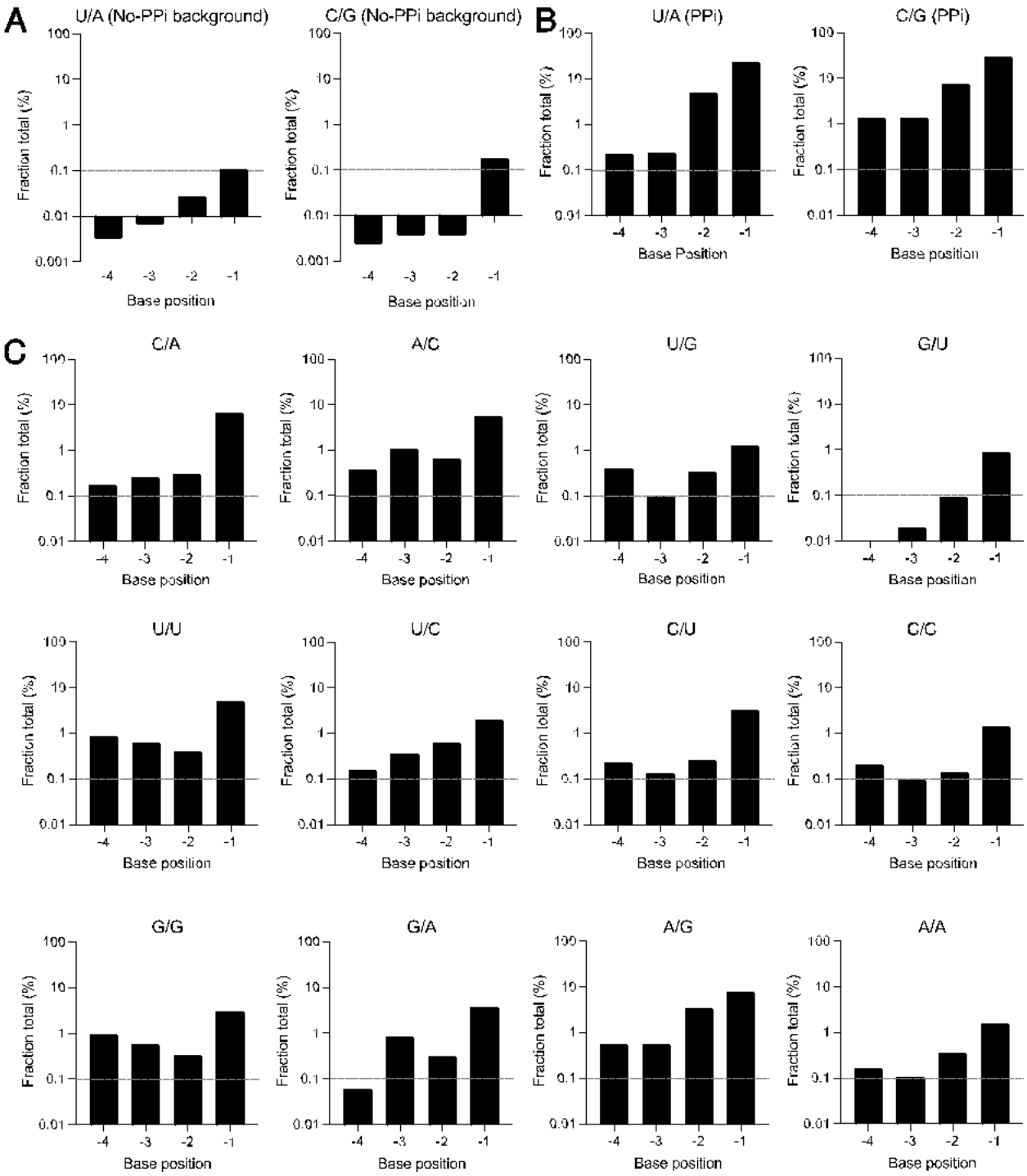
Extended Data Fig. 2. Ribozyme-catalyzed, pyrophosphate-facilitated repair reactions for

774 **all mismatch combinations.** All mismatches were ^{32}P-labeled and reacted under standard
775 conditions with variations indicated with + and - symbols. The reactions were then analyzed by
776 high-resolution PAGE.

777

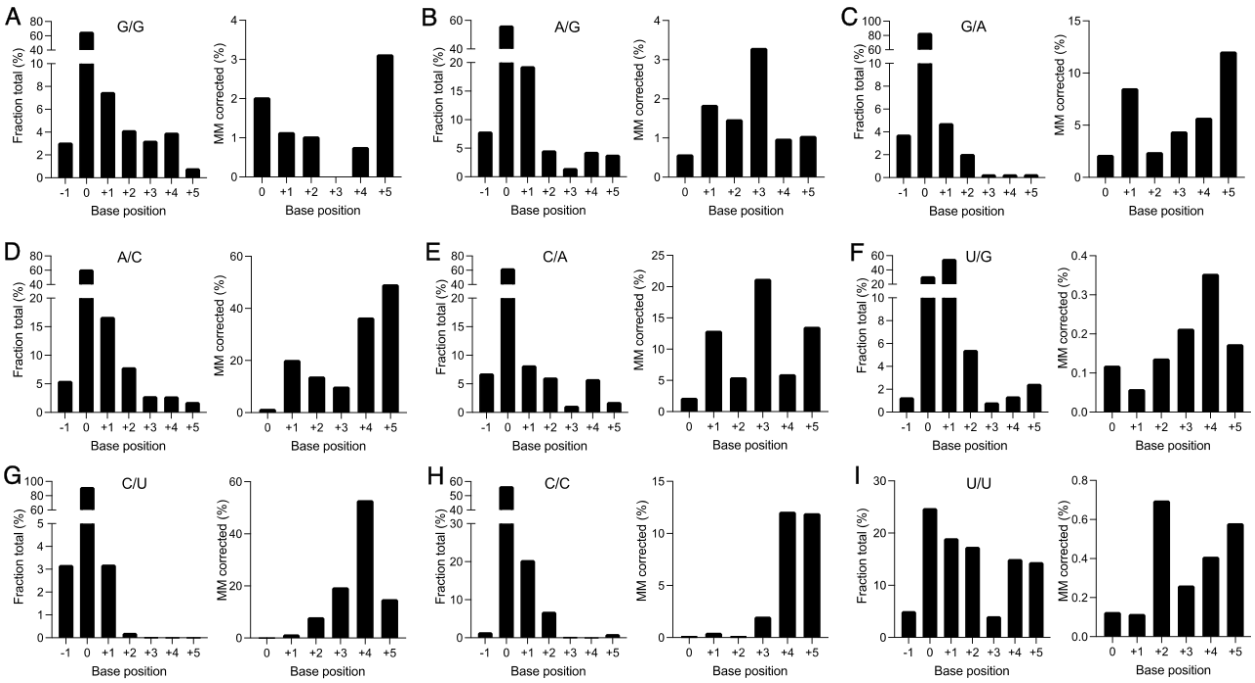


Extended Data Fig. 3. Luciferase-based assay measures ATP produced during ribozyme-catalyzed pyrophosphorolysis. (A) Firefly luciferase assay measuring the concentration of ATP in nM generated by the removal of the terminal adenosine monophosphate of an A/C mismatched hairpin, as calculated by use of a standard curve. Reactions lacking the polymerase ribozyme, hairpin, or pyrophosphate show only background levels of luminescence. Higher concentrations of ATP are present in reactions containing all three components. Open and filled circles represent two independent preparations of the polymerase ribozyme and the hairpin substrate, with filled circles showing higher background ATP levels apparently co-purified with the polymerase ribozyme. Lines connect individual experiments performed in parallel on the same day and from the same stocks. A representative luminescence image is shown below the graph. (B) Additional luminescence image used to generate the values shown in A and showing additional control reactions lacking both the ribozyme and pyrophosphate.



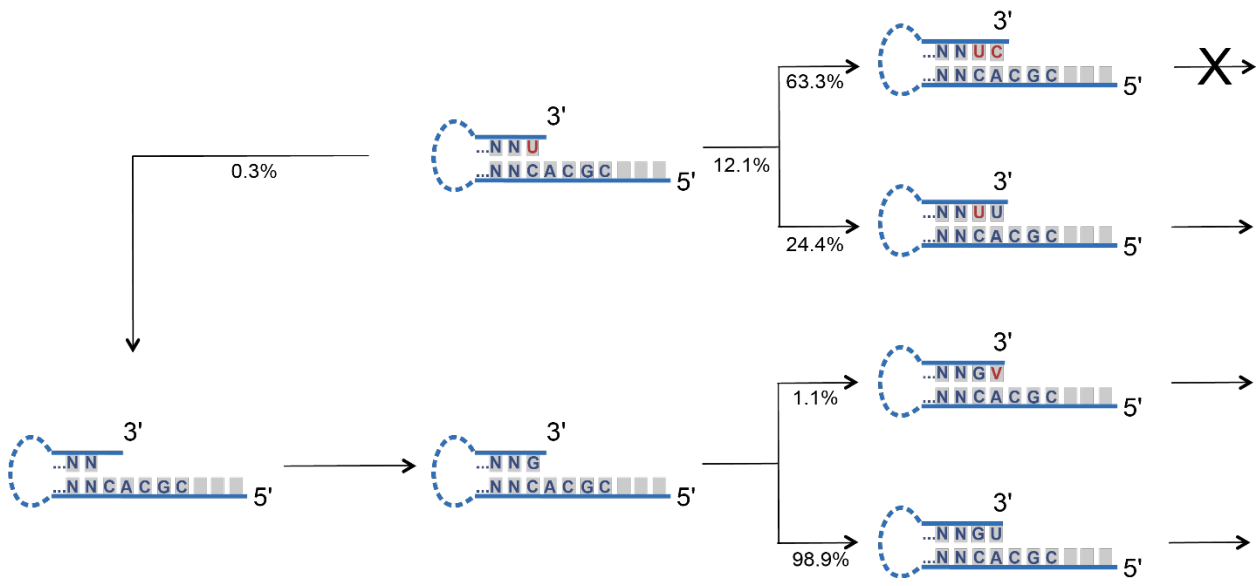
Extended Data Fig. 4. Removal of terminal nucleotides in the presence and absence of pyrophosphate. Addition of pyrophosphate increases the fraction of shorter sequences as determined by high-throughput sequencing. (A) Matched U/A and C/G hairpins terminated with a

2′-3′-cyclic phosphate and incubated in absence of pyrophosphate show very low fractions of shorter sequences, indicating low background of these sequences in the samples. (B) Matched hairpins terminated with a 2′-3′-cyclic phosphate yield a fraction of shorter sequences that is about 2 orders of magnitude higher, particularly in the –1 position, after treatment with ribozyme and pyrophosphate. These results provide direct evidence for ribozyme-catalyzed pyrophosphorolysis of RNA. (C) All mismatched sequences following incubation with pyrophosphate and ribozyme. Some sequences exhibit higher percentages of pyrophosphorolysis than others, largely following the trends described in Figure 4. To ease comparison, all graphs are plotted on the same scale, with the horizontal axis placed at 0.01% and a gray dashed line placed at 0.1%.

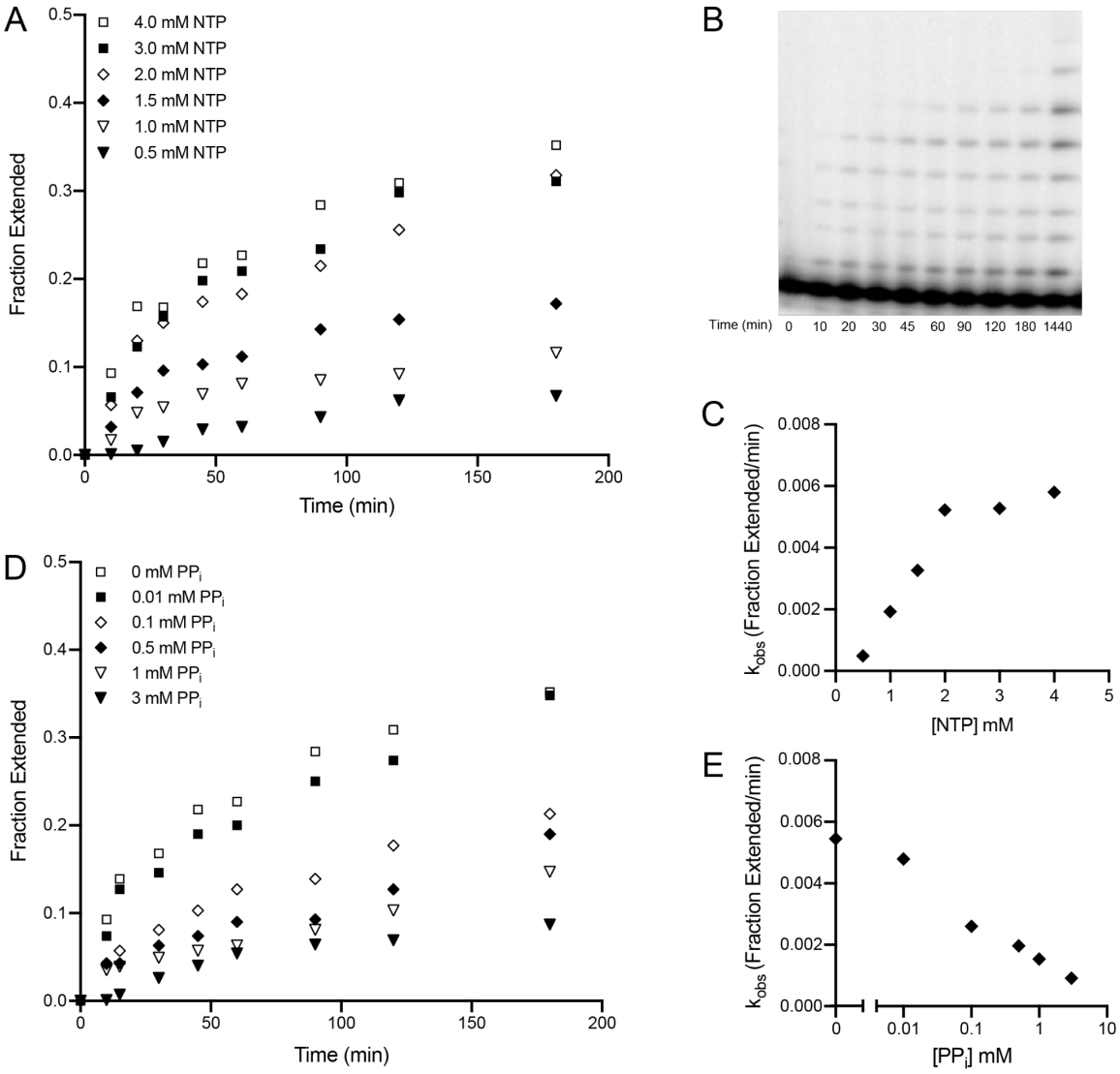


Extended Data Fig. 5. Extension and repair of mismatches. Mismatched hairpins repaired and extended with distinct efficiencies, as determined by high-throughput sequencing. The left graph of each panel shows the fraction of sequences terminated at each position along the template relative to the starting length (0) after incubation with the polymerase ribozyme, pyrophosphate, and rNTPs. All hairpins were treated with PNK to remove the 2′-3′-cyclic phosphate. "–1" corresponds to the fraction of sequences terminated at nucleotides corresponding to the products of pyrophosphorolysis of the mismatch-terminated hairpin. Hairpins terminated in A/G (B), A/C (D), C/A (E), and U/U (I) show particularly high fractions of sequences with the terminal mismatches removed. The right graph of each panel shows the percentage of each extended segment with the starting mismatch repaired. (A-C) Purine/purine mismatches are repaired and extended at low efficiency, but longer products contain high fraction of repaired sequences. (D–F) Purine/pyrimidine mismatches are extended well and have high percentages of repair, except the U/G wobble pair. For example, a hairpin that started with an A/C mismatch (D), showed more than 15% of sequences extended by one nucleotide (+1, left graph) and ~20 % of these 1-nt extensions had been repaired (+1, right graph). (G–I) Pyrimidine/pyrimidine mismatches are

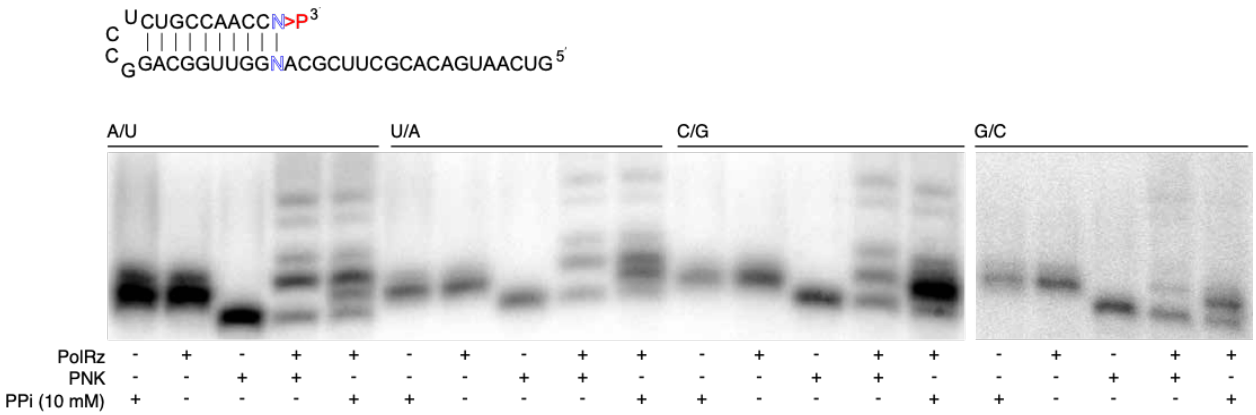
extended with high efficiencies but are not repaired well.



Extended Data Fig. 6. One mismatch can induce addition of more mutations. When a hairpin having a terminal U/C mismatch is incubated with a polymerase ribozyme and tC19Z it can either undergo pyrophosphorolysis or be extended by one nucleotide. In only a low percentage is the mismatch removed, but if it is, the subsequent extension proceeds with high fidelity. If the mismatch is not corrected, the ribozyme will add a second mismatch preferentially which results in chain termination.



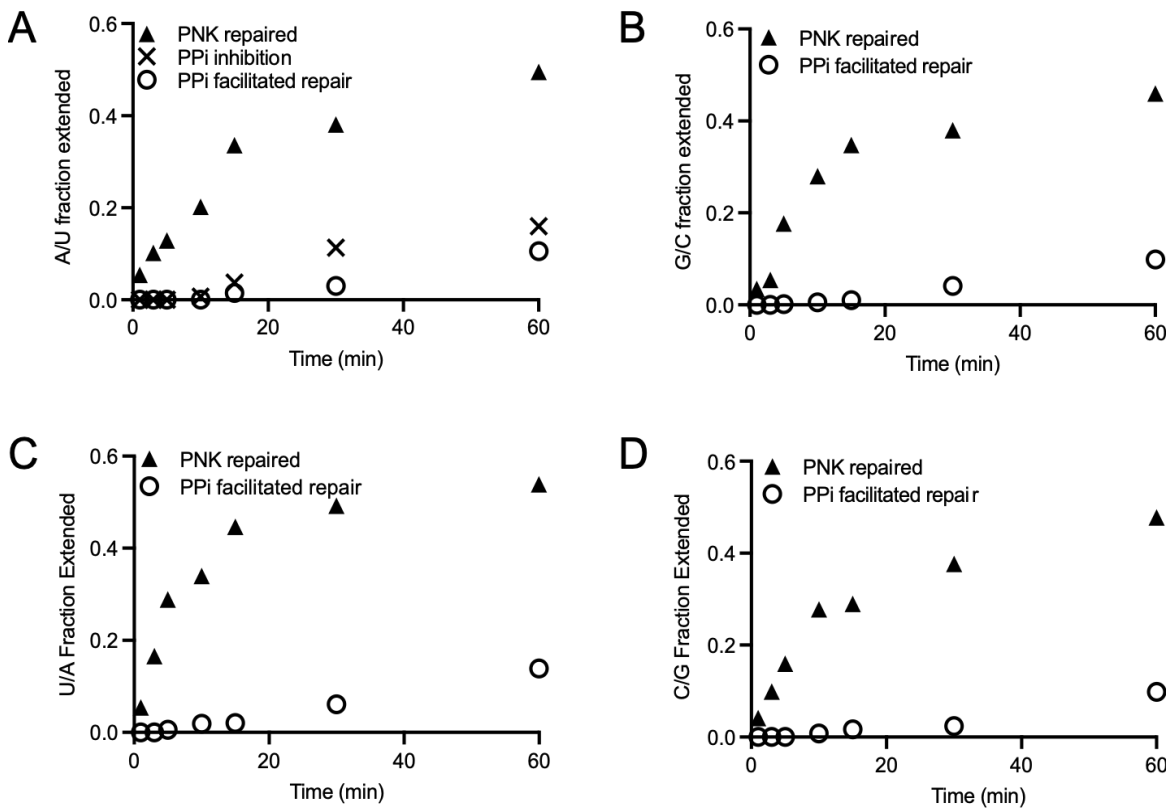
Extended Data Fig. 7. Reaction kinetics of primer extension and pyrophosphate inhibition. (A) The fraction extended of the ^{32}P-labeled primer is quantified as a function of time at varying NTP concentrations by high resolution PAGE. (B) High-resolution PAGE of the primer extension at 4 mM rNTPs which is plotted on the graph in A. (C) The fraction of primer extended per minute is plotted as a function of NTP concentration and shows an estimated K_M of ~4 mM. (D) The fraction of the ^{32}P-labeled primer extended quantified as a function of time at varying pyrophosphate conditions. (E) The fraction extended per minute is plotted as a function of the log of pyrophosphate concentration.



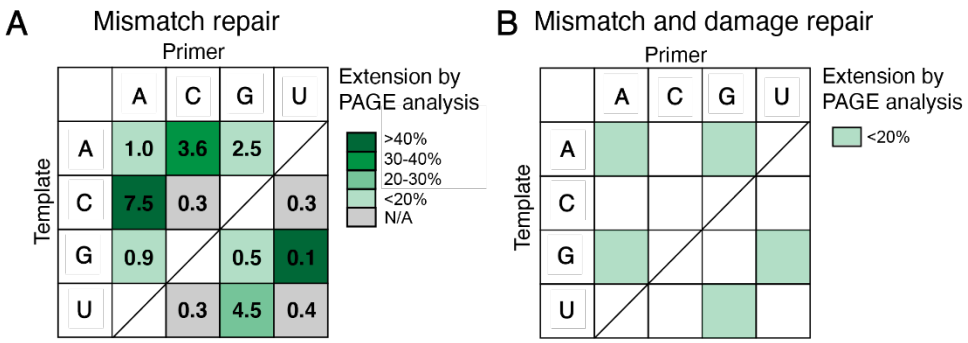
841

842 **Extended Data Fig. 8. Damage repair reactions for all matched hairpins.** The general
843 sequence of the hairpins is shown above the PAGE gels, with terminal base-pair shown as N-N
844 in blue and specific terminal base-pairs indicated above each set of experiments. The site of RNA
845 damage, corresponding to the 2′-3′-cyclic phosphate is indicated as >P in red. All hairpins were
846 allowed to react under standard conditions with reaction conditions variations shown by the + and
847 - symbols. Hairpins were analyzed by high-resolution PAGE. The A/U, U/A, and C/G hairpins
848 were all resolved on the same gel, and the G/C hairpin was resolved on a separate PAGE gel.
849 PNK indicates removal of the terminal 2′-3′-cyclic phosphate by the polynucleotide kinase enzyme
850 to yield canonical 2′ and 3′ hydroxyls, which results in faster migration. All hairpins were
851 extendable by the tC19Z polymerase ribozyme when either treated by the PNK enzyme or
852 pyrophosphate (right two lanes of each set of experiments).

853



Extended Data Fig. 9. Kinetics of RNA polymerization of damage-repaired substrates. A ^{32}P-labeled hairpin terminated with 2′-3′-cyclic phosphate was either treated with PNK or pyrophosphate and then subjected to standard polymerase-ribozyme extension conditions. Nine time points were collected and analyzed by high-resolution PAGE. (A) For an A/U hairpin, triangles indicate the fraction extended of the PNK-repaired hairpin, crosses indicate the fraction extended of the PNK repaired hairpin in the presence of 10 mM PPi to determine the inhibition of the reaction by pyrophosphate, and open circles indicate the fraction extended of the 2′-3′-cyclic-phosphate-terminated hairpin that was repaired by the ribozyme in presence of 10 mM PPi. For hairpins terminated in G/C (B), U/A (C), or C/G (D), triangles indicate the fraction extended of the PNK-repaired hairpin, and open circles indicate the fraction extended of the 2′-3′-cyclic-phosphate-terminated hairpin that was repaired by addition of 10 mM PPi.



867

868 **Extended Data Fig. 10. Trends in mismatch repair and damage and mismatch repair.** (A)
869 Grid showing the correlation between mismatch repair and overall mismatch extension. The box
870 colors indicate what percentage of each mismatch is extended when repaired by the ribozyme in
871 the presence of pyrophosphate, compared to a no-pyrophosphate control, as determined by
872 PAGE analysis. The gray color indicates no significant difference observed in extension reactions
873 in the presence/absence of pyrophosphate. The numbers correspond to the total fraction of
874 mismatch repair derived from HTS data. (B) Grid showing mismatches in which both 2′-3′-cP
875 phosphate damage and the mismatch could be repaired *via* ribozyme-catalyzed
876 pyrophosphorolysis of the terminal nucleotide.

877

878 **Supplementary Information**

879 **SI Note 1. Gel migration of cyclic-phosphate–terminated hairpins**

880 Previous reports showed that treatment of 2′-3′-cP-terminated oligonucleotide sequences
881 with PNK to remove the cyclic phosphate leads to a gel mobility retardation, and treatment of a
882 2′-3′-cyclic-phosphate-terminated sequence with acid to generate a linear monophosphate leads
883 to a faster-migrating species.[52–54] We consistently observed the opposite effect and have
884 concluded that the anomalous migration is a property of the hairpin sequence, which differs
885 structurally from the linear RNA sequences used in previously reported experiments. Following
886 treatment with PNK, all 2′-3′-cP-terminated hairpin sequences exhibit faster gel migration. The
887 relative gel mobility of the 2′-3′-cP-terminated hairpins treated with acid (which initially hydrolyzes
888 just one of the phosphodiesters of the cyclic phosphate) shows a gel mobility retardation. This
889 data confirmed that PNK treated hairpin sequences indeed migrate faster upon removal of the 2′-
890 3′-cP (Extended Data Fig. 1). Many of the hairpin sequences when visualized by PAGE show two
891 bands depending on the resolution of the experiment. Based on the gel migration of the acid
892 treated 2′-3′-cP, we determined that these two bands arise from the difference between closed
893 and open states of the terminal cyclic phosphate.

894 **SI Note 2. Analysis of HTS sequencing data for each relevant figure**

895 Additional details on the HTS analysis for each figure:
896 All libraries were submitted for Illumina sequencing. See Supplementary Table 2 for explanation
897 of datasets.
898 Illumina MiSeq Single Read 150 sequencing was used for tC19Z_MM_1, tC19Z_MM_2,
899 tC19Z_MM_3, tC19Z_MM_4, tC19Z_MM_5, tC10Z_MM_6, tC19Z_MM_7, tC19Z_MM_8,
900 tC19Z_MM_9, tC19Z_MM_10, tC19Z_MM_11, tC19Z_MM_12, tC19Z_MA_1, tC19Z_MA_2,
901 tC19Z_MA_3, tC19Z_MA_4, tC19Z_MA_NoPPi_1, tC19Z_MA_NoPPi_2

902 Illumina MiSeq Paired End sequencing was used for tC19Z_LT_3_0, tC19Z_LT_3_2.5,
903 tC19Z_LT_3_10, tC19Z_LT_4_0, tC19Z_LT_4_3, tC19Z_LT_4_10, tC19Z_LT_5_10,
904 tC19Z_LT_5_2.5, tC19Z_LT_5_1, tC19Z_LT_5_0
905 A PhiX library was added into all samples and its sequences were first removed from all
906 sequencing files using Bowtie2.
907
908 **Analysis for Figure 2d:**
909 CutAdapt was used to remove the adapters on first the 5′ and then the 3′-termini of the sequences.
910 FastAptamer was used to count the sequences. FastAptamer search was then used to find all
911 sequences that had the full hairpin present, including sequences in which the 3′-ends were either
912 extended or shortened by up to 5 nucleotides. The total counts of these sequences were
913 tabulated. FastAptamer search was then used to find sequences that corresponded to the starting
914 hairpin, sequences that were -1, sequences that were extended by one nucleotide correctly,
915 sequences that were extended by one nucleotide incorrectly, sequences that were extended by
916 two nucleotides correctly, sequences that were extended by one nucleotide correctly and the
917 second nucleotide incorrectly, etc., up to five nucleotides. These sequences were then visually
918 inspected and manually curated. The counts of the remaining sequences were tabulated. The
919 extension percentages shown in Figure 2d were calculated by generating percentages of the total
920 sequences for each extension length regardless of extended nucleotide identity.
921
922 **Analysis for Figure 2e:**
923 CutAdapt was used to remove the adapters on first the 5′ and then the 3′-termini of the sequences.
924 FastAptamer was used to count the sequences. FastAptamer search was then used to find all
925 sequences that had the hairpin present, including sequences in which the 3′-ends were either
926 extended or shortened by up to 5 nucleotides. The total counts of these sequences were

927 tabulated. FastAptamer search was then used to find sequences that corresponded to the starting
928 hairpin, sequences that were -1, sequences that were extended by one nucleotide correctly,
929 sequences that were extended by one nucleotide incorrectly, sequences that were extended by
930 two nucleotides correctly, sequences that were extended by one nucleotide correctly and the
931 second nucleotide incorrectly, etc., up to five nucleotides. These sequences were then visually
932 inspected and manually curated. The counts of the remaining sequences were tabulated. The
933 extension percentages shown in Figure 2e were calculated by generating percentages of just
934 sequences, in which the nucleotides were extended correctly as shown in the figure. The same
935 normalization factor as used in Figure 2e was applied.
936
937 **Figure 2f:**
938 CutAdapt was used to remove the adapters on first the 5′ and then the 3′-termini of the sequences.
939 FastAptamer was used to count the sequences. FastAptamer search was then used to find all
940 sequences that had the hairpin present, including sequences in which the 3′-ends were either
941 extended or shortened by up to 5 nucleotides. The total counts of these sequences were
942 tabulated. FastAptamer search was then used to find sequences that corresponded to the starting
943 hairpin, sequences that were -1, sequences that were extended by one nucleotide correctly,
944 sequences that were extended by one nucleotide incorrectly, sequences that were extended by
945 two nucleotides correctly, sequences that were extended by one nucleotide correctly and the
946 second nucleotide incorrectly, etc., up to five nucleotides. These sequences were then visually
947 inspected and manually curated. The counts of the remaining sequences were tabulated. The
948 percentages shown in Figure 2f were calculated by tabulating all sequences that had the
949 mismatched nucleotide repaired and extended from zero to five nucleotides for each mismatched
950 sequence.
951

952 **Figure 2g:**

953 CutAdapt was used to remove the adapters on first the 5′ and then the 3′-termini of the sequences.
954 FastAptamer was used to count the sequences. FastAptamer search was then used to find all
955 sequences that had the hairpin present, including sequences in which the 3′-ends were either
956 extended or shortened by up to 5 nucleotides. The total counts of these sequences were
957 tabulated. FastAptamer search was then used to find sequences that corresponded to the starting
958 hairpin, sequences that were -1, sequences that were extended by one nucleotide correctly,
959 sequences that were extended by one nucleotide incorrectly, sequences that were extended by
960 two nucleotides correctly, sequences that were extended by one nucleotide correctly and the
961 second nucleotide incorrectly, etc., up to five nucleotides. These sequences were then visually
962 inspected and manually curated. The counts of the remaining sequences were tabulated. The
963 percentages shown in the black bars in Figure 2g were calculated by tabulating the sequences
964 that were not repaired and the percentage that were subsequently extended correctly was
965 calculated. The grey bars in Figure 2g were calculated by tabulating the sequences that were
966 repaired and the percentage that were subsequently extended correctly was calculated.

967

968 **Figure 3b and 3c:**

969 CutAdapt was used to remove the adapters and primer regions on first the 5′ and then the 3′-end
970 of the sequences. Sequences were counted in FastAptamer using fastaptamer_count.
971 Sequences containing the primer were extracted from data for analysis in FastAptamer using
972 fastaptamer_search for each sequence and subsequently visually aligned and curated using
973 Jalview. Sequences that were not extended, ligation artifacts, and sequencing artifacts were
974 removed. Jalview was used to view the aligned sequences. The sequence lengths were verified
975 visually. This procedure was repeated for each condition shown in the figure.

976

977 **Figure 3d and 3e:**

978 CutAdapt was used to remove the adapters and primer regions on first the 5′ and then the 3′-end
979 of the sequences. Sequences were counted in FastAptamer using fastaptamer_count.
980 Sequences containing the primer were extracted from data for analysis in FastAptamer using
981 fastaptamer_search for each sequence and subsequently visually aligned and curated using
982 Jalview. Sequences that were not extended, ligation artifacts, and sequencing artifacts were
983 removed. The fidelity was determined by visually inspecting the sequences in Jalview and
984 tabulating the number of incorrectly incorporated nucleotides. The percentage fidelity was
985 determined as the average of the fidelity at each position.

986

987 **Figure 4c-g:**

988 CutAdapt was used to remove the adapters on first the 5′ and then the 3′-termini of the sequences.
989 FastAptamer was used to count the sequences. FastAptamer search was then used to find all
990 sequences that had the hairpin present, including sequences in which the 3′-ends were either
991 extended or shortened by up to 5 nucleotides. The total counts of these sequences were
992 tabulated. FastAptamer search was then used to find sequences that corresponded to the starting
993 hairpin, sequences that were -1, sequences that were extended by one nucleotide correctly,
994 sequences that were extended by one nucleotide incorrectly, sequences that were extended by
995 two nucleotides correctly, sequences that were extended by one nucleotide correctly and the
996 second nucleotide incorrectly, etc., up to five nucleotides. These sequences were then visually
997 inspected and manually curated. The counts of the remaining sequences were tabulated. The
998 extension percentages shown in Figure 4c-g were calculated by generating percentages of the
999 total sequences for each extension length.

1000

1001 **Extended Data Fig. 4:**

1002 CutAdapt was used to remove the adapters on first the 5′ and then the 3′-end of the sequences.
1003 FastAptamer was used to count the sequences. FastAptamer search was then used to find all
1004 sequences that had the hairpin present including sequences where the 3′-end was either
1005 extended or shortened up to 5 nucleotides. The total counts of these sequences were tabulated.
1006 FastAptamer search was then used to find sequences that corresponded to sequences that were
1007 -1, -2, -3, and -4. These sequences were then visually inspected and manually curated. The
1008 counts of the remaining sequences were tabulated. The percentages shown in Extended Data
1009 Fig. 4 were calculated by generating percentages of sequences that had 1-4 nucleotides
1010 removed.
1011
1012 **Extended Data Fig. 5:**
1013 CutAdapt was used to remove the adapters on first the 5′ and then the 3′-termini of the sequences.
1014 FastAptamer was used to count the sequences. FastAptamer search was then used to find all
1015 sequences that had the hairpin present, including sequences in which the 3′-ends were either
1016 extended or shortened by up to 5 nucleotides. The total counts of these sequences were
1017 tabulated. FastAptamer search was then used to find sequences that corresponded to the starting
1018 hairpin, sequences that were -1, sequences that were extended by one nucleotide correctly,
1019 sequences that were extended by one nucleotide incorrectly, sequences that were extended by
1020 two nucleotides correctly, sequences that were extended by one nucleotide correctly and the
1021 second nucleotide incorrectly, etc., up to five nucleotides. These sequences were then visually
1022 inspected and manually curated. The counts of the remaining sequences were tabulated. The
1023 extension percentages shown in the left graph of each panel in Extended Data Fig. 5 were
1024 calculated by generating percentages of the total sequences for each extension length regardless
1025 of extended nucleotide identity. The percentages were then normalized to add to 100%. The right

graph of each panel in Extended Data Fig. 5 shows the percentage of the sequences in the left panel at each length where the mismatch was corrected.

Extended Data Fig. 6:

CutAdapt was used to remove the adapters on first the 5′ and then the 3′-termini of the sequences. FastAptamer was used to count the sequences. FastAptamer search was then used to find all sequences that had the hairpin present, including sequences in which the 3′-ends were either extended or shortened by up to 5 nucleotides. The total counts of these sequences were tabulated. FastAptamer search was then used to find sequences that corresponded to the starting hairpin, sequences that were -1, sequences that were extended by one nucleotide correctly, sequences that were extended by one nucleotide incorrectly, sequences that were extended by two nucleotides correctly, sequences that were extended by one nucleotide correctly and the second nucleotide incorrectly, etc., up to five nucleotides. These sequences were then visually inspected and manually curated. The counts of the remaining sequences were tabulated. Extension percentages shown in Extended Data Fig. 6 were calculated for each sequence shown as a fraction of the total hairpin containing sequences.

Supplementary Fig. 2:

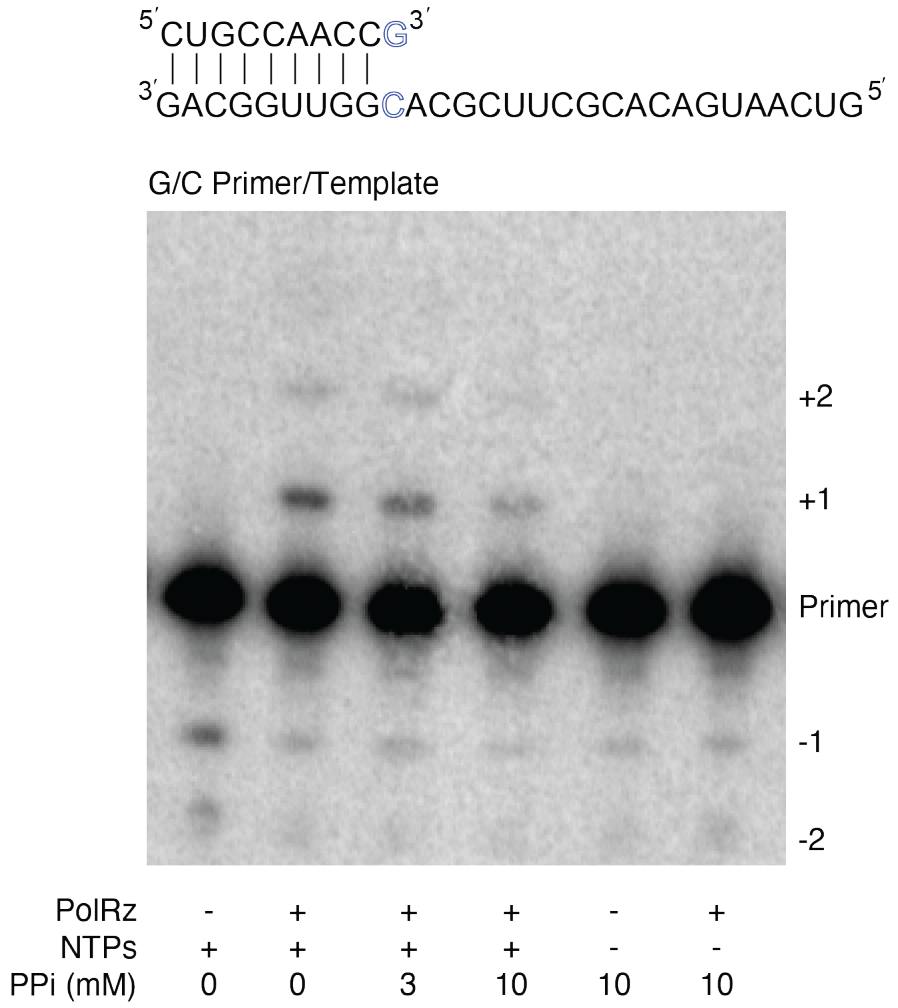
CutAdapt was used to remove the adapters and primer regions on first the 5′ and then the 3′-end of the sequences, including the seven random nucleotides. Sequences were counted in FastAptamer using fastaptamer_count. Sequences containing the primer were extracted from data for analysis in FastAptamer using fastaptamer_search for each sequence and subsequently visually aligned and curated using Jalview. Sequences that were not extended, ligation artifacts, and sequencing artifacts were removed. The sequences were visually inspected, manually aligned, and sequences that did not align to the start of the sequence were removed. The

1051 sequences were exported to Adobe Illustrator, the reference sequence was colored green and

1052 mismatches were colored blue.

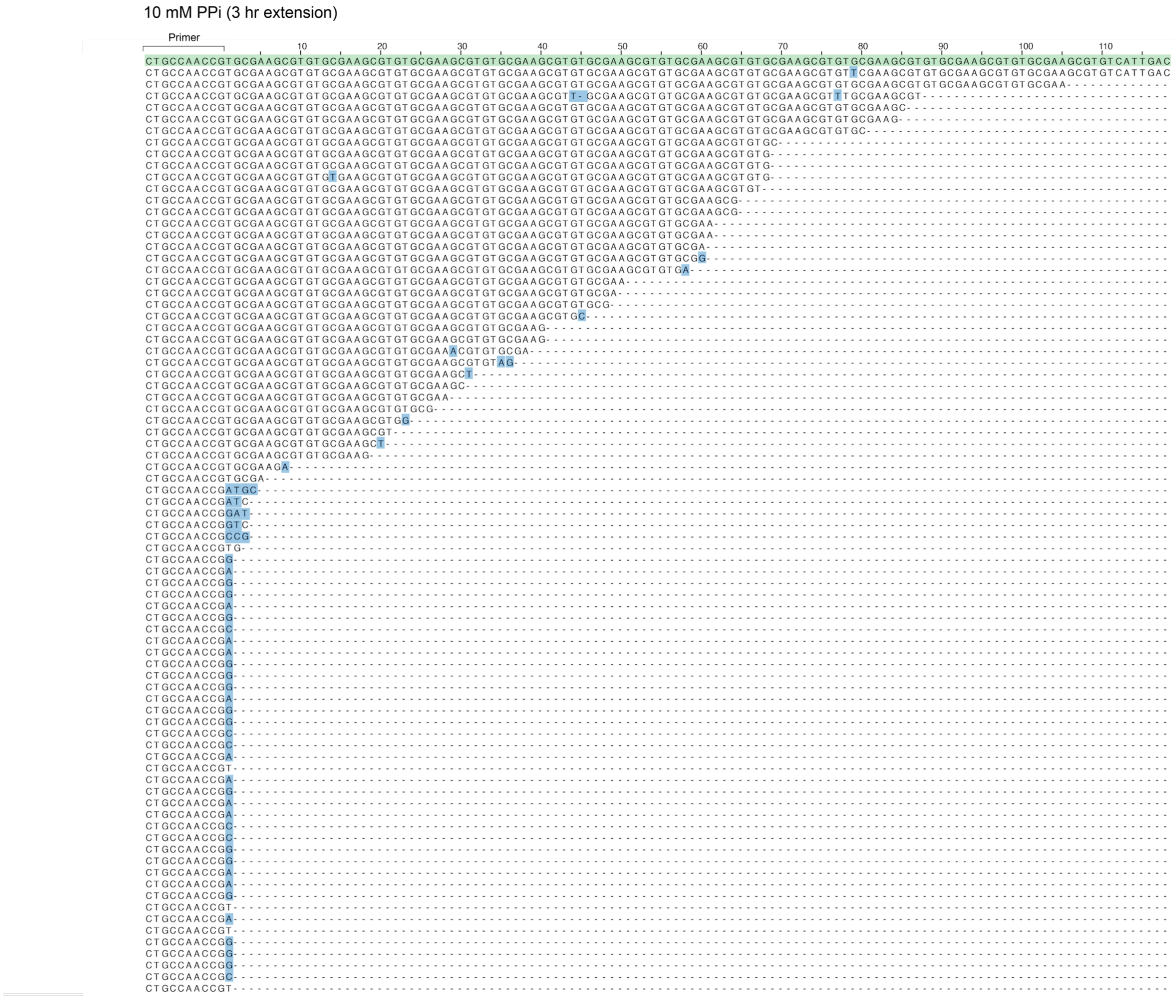
1053

1054

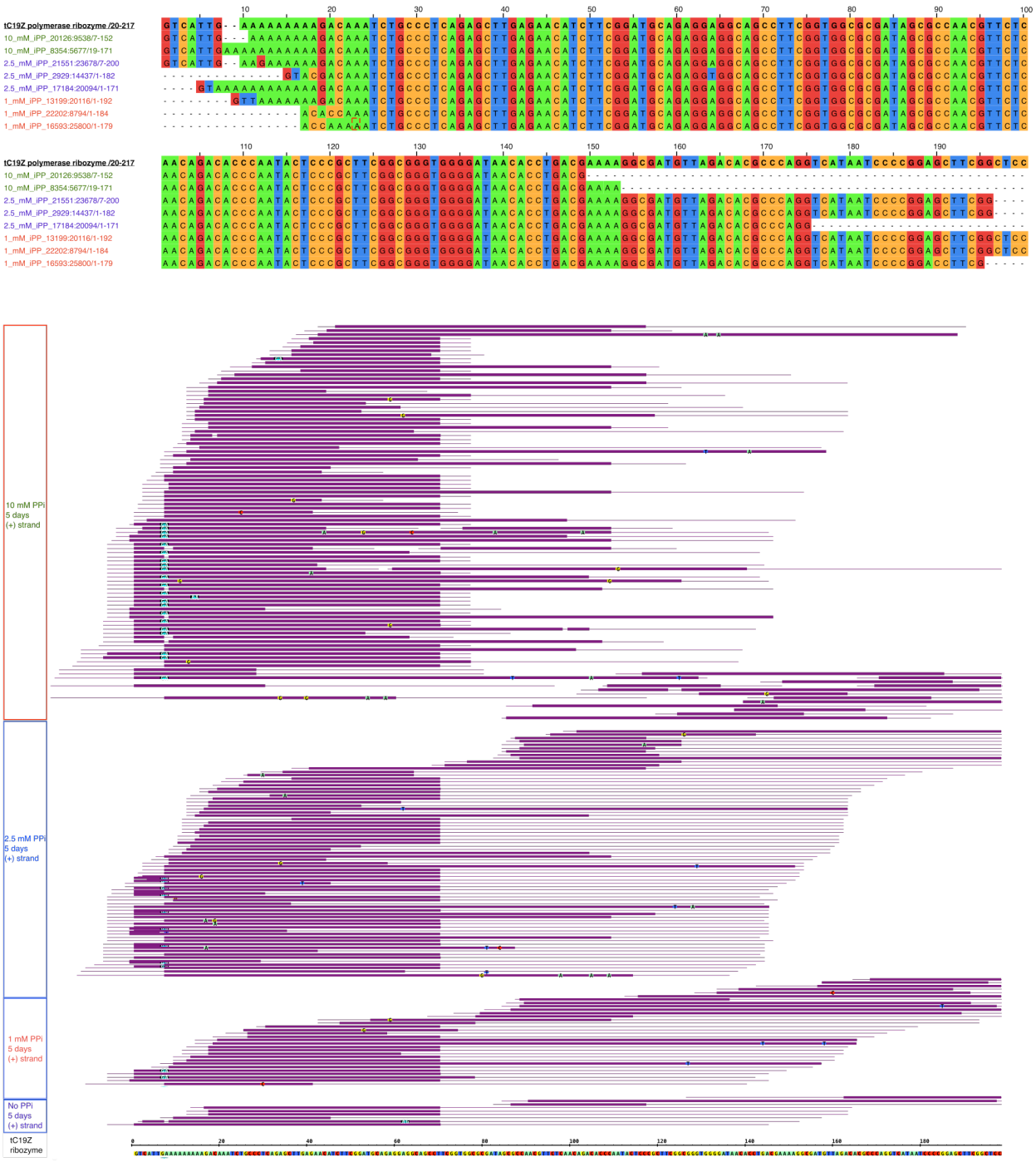


1055

1056 **Supplementary Fig. 1. Pyrophosphate inhibits ribozyme mediated extension.** A
1057 primer/template sequence is extended by the ribozyme under standard conditions. The extension
1058 is inhibited when pyrophosphate is also added to the reaction (see also Extended Data Fig. 3).
1059 However, addition of pyrophosphate either with or without the tC19Z polymerase ribozyme does
1060 not result in any additional bands corresponding to RNA degradation or significant observable
1061 pyrophosphorolysis.



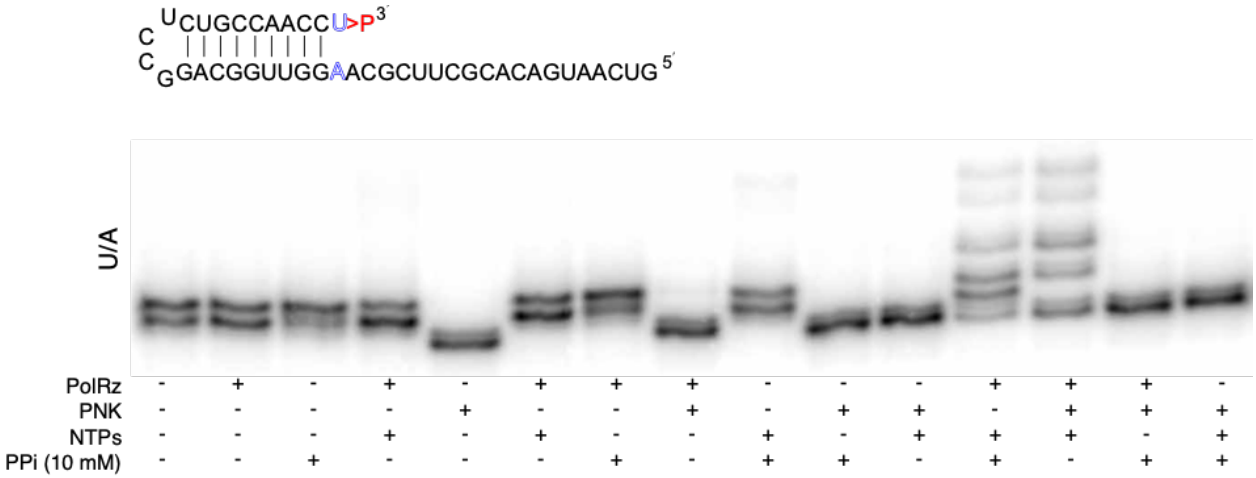
Supplementary Fig. 2. Alignment of long extension sequences from HTS. Sequences corresponding to extensions present in the HTS data generated from the three-hour polymerization reaction and subsequent sequencing described in Figure 3B. Many of the sequences that are not full length terminate in one or more mismatches. The primer region is shown in the bracket, and the expected full-length sequence the HTS data was aligned to is highlighted in green. Mismatches are highlighted in blue.



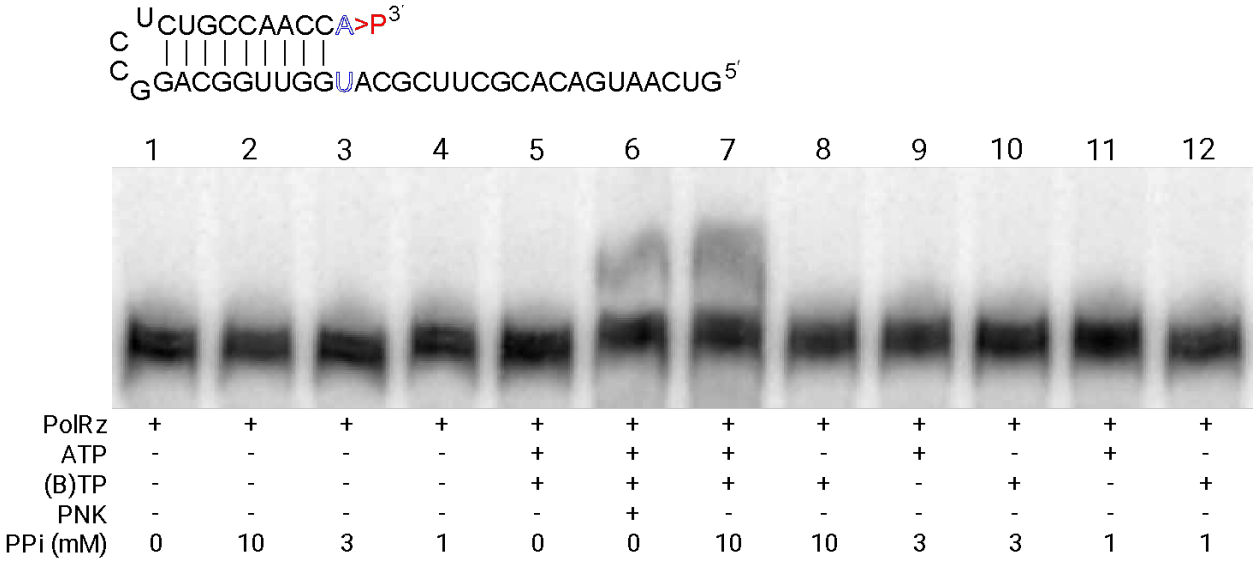
Supplementary Fig. 3. tC19Z sequences present in HTS analysis of long extension reactions. (Top) Sequences of cDNAs generated by 3′–terminal ligation of the RNA strand and, after reverse transcription, 3′–terminal ligation of the cDNA strand. The high-throughput

1075 sequencing forward adapter was ligated to the 3′–terminus of the RNA. The cDNAs shown
1076 matching the tC19Z ribozyme (top sequence) therefore report generation of RNAs antisense to
1077 the ribozyme during the extension reactions (data shown are for analysis of 5-day experiments).
1078 RNAs corresponding to long segments of tC19Z were observed only in experiments containing
1079 pyrophosphate. The pyrophosphate concentrations are annotated on the left. The longest copies
1080 of the ribozyme lack just 4 nucleotides (2%) from the ribozyme termini. (Bottom) All cDNA
1081 sequences matching the sense strand of the tC19Z ribozyme generated under 0-, 1-, 2.5-, and
1082 10-mM pyrophosphate conditions. No long sequences were observed in samples lacking
1083 pyrophosphate. Thick lines represent the ribozyme matching segments, whereas the thin lines
1084 represent segments of other sequences, such as ligation and sequencing adapters.

1085

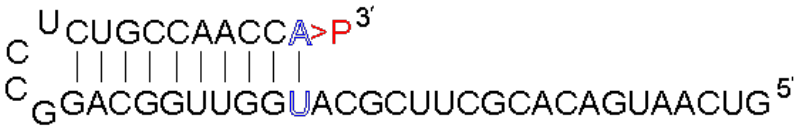


Supplementary Fig. 4. Damage repair reactions with a full set of controls for the U/A-terminated hairpin substrate. A U/A hairpin was allowed to react under standard extension conditions with variations shown by + and - symbols below the gel image.



Supplementary Fig. 5. Omission of a single nucleotide prevents extension of a repaired A/U hairpin. A ^{32}P-labeled A/U hairpin was extended under standard reaction conditions with variations indicated by the + and – symbols below each reaction resolved by PAGE. The left 5 lanes are negative controls, that lack any NTP; lane 6 is a positive control (PNK-repaired hairpin extension); and lane 7 is the pyrophosphate-facilitated repair reaction at 10 mM pyrophosphate. Lane 8 indicates that upon addition of all nucleotides except ATP (indicated by (B)TP (all, except ATP)), no extension is observed, strongly suggesting that the pyrophosphate-facilitated repair removes adenosine 5′ monophosphate-2′-3′-cyclic phosphate from the damaged terminus and requires ATP to incorporate the first nucleotide. Lanes 9–12 show that the repair does not occur at subsaturating concentrations of pyrophosphate.

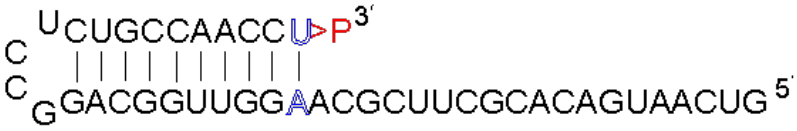
A



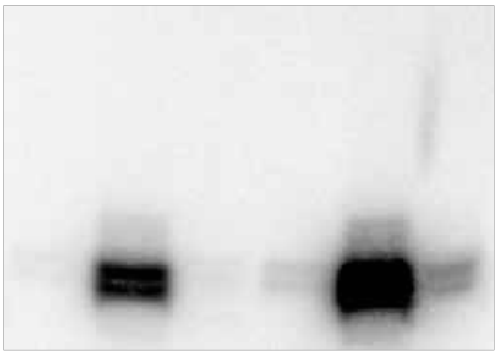
A/U α-32P hairpin | A/U non-radiolabeled hairpin

PolRz	+	+	+	+	+	+	+	+	+	+	+
PNK	-	-	-	+	+	+	-	-	-	+	+
α-32P ATP	-	+	-	-	+	-	+	+	+	+	+
NTPs	+	-	+	-	-	+	-	-	+	-	+
PPi (10 mM)	-	+	+	-	-	-	-	+	+	-	-

B



U/A non-radiolabeled hairpin



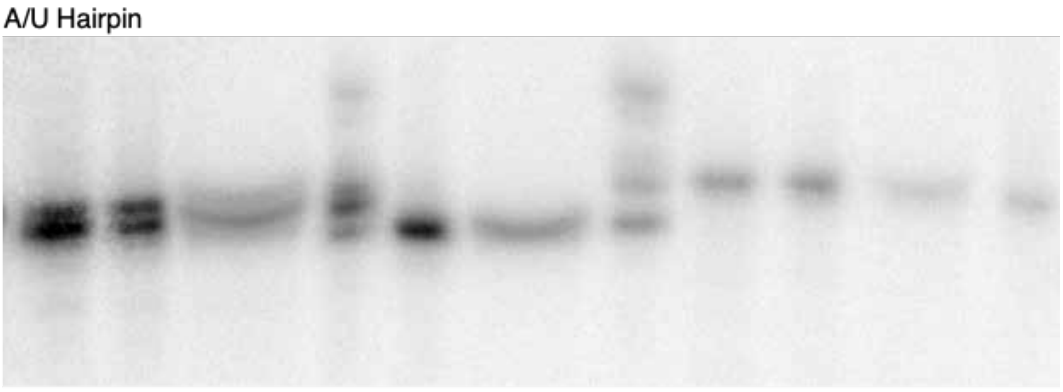
tC19Z	+	+	+	+	+	+
PNK	-	-	-	+	+	+
α-32P ATP	+	-	-	+	-	-
α-32P UTP	-	+	-	-	+	-
α-32P CTP	-	-	+	-	-	+
PPi (10 mM)	+	+	+	-	-	-

1104

1105 **Supplementary Fig. 6. Incorporation of α-^{32}P-labeled nucleotides into repaired, non-labeled**
1106 **hairpins.** (A) An A/U hairpin was prepared with or without internal ^{32}P, as indicated above the
1107 PAGE gel image. The hairpin was then subjected to repair and extension under standard reaction
1108 conditions, except that α-^{32}P-ATP was added to some reactions. In the case of the non-labeled
1109 hairpin, the ^{32}P-AMP is only incorporated if the sequence has been repaired either by
1110 pyrophosphate-facilitated reaction or by PNK. (B) A U/A hairpin was transcribed without α-^{32}P-
1111 ATP and was subjected to standard extension conditions but using α-^{32}P-rNTP (indicated) in place
1112 of the standard NTPs. Different α-^{32}P-rNTPs are utilized by the tC19Z polymerase ribozyme with
1113 different efficiencies, as shown by the varying intensities of the bands. ^{32}P-UMP is incorporated
1114 most efficiently because the first nucleotide on the template sequence is an adenosine. However,
1115 misincorporation is seen for the other nucleotides, and a second nucleotide appears to be added
1116 in presence of UTP. No ^{32}P labeling is observed in control reactions that lack PNK or
1117 pyrophosphate (as shown in panel A).

1118

1119



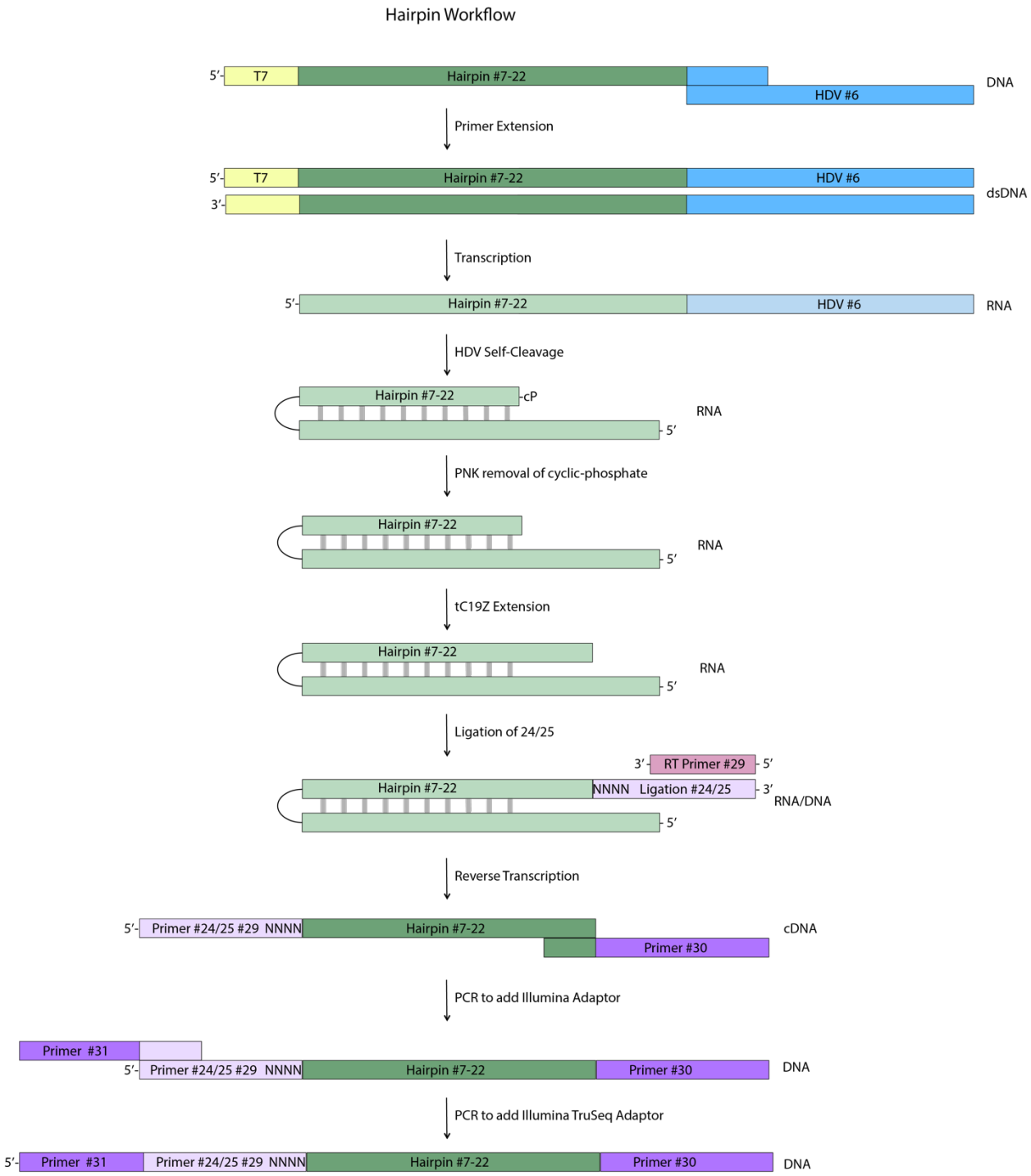
	Hairpin				PNK			Acid			
Uridine (4 mM)	-	-	+	-	-	+	-	-	-	+	-
NTP	-	-	-	+	-	-	+	-	+	-	+
PPi (10 mM)	-	+	+	+	-	-	-	-	+	-	-

1120

1121 **Supplementary Fig. 7. The mechanism of repair does not involve cyclic-phosphate opening**
1122 **by pyrophosphate.** Pyrophosphate and uridine were added to a damaged A/U hairpin to
1123 determine whether the cyclic phosphate could be opened by pyrophosphate to form a
1124 triphosphate and allow for subsequent addition of uridine (lane 3). No additional bands were
1125 observed on the PAGE gel, indicating that this is not the mechanism of repair. Uridine was added
1126 to PNK-repaired or acid-treated hairpins as positive and negative controls, respectively.

1127

1128



1129
1130 **Supplementary Fig. 8. Workflow for HTS analysis of hairpin substrates of the polymerase**
1131 **ribozyme.** The hairpins were first prepared by primer extension of an drz-Mbac-1 HDV-like
1132 ribozyme sequence (#6) and a hairpin sequence (#7–#22). This DNA construct with the ribozyme

1133 on the 3′-end of the hairpin was transcribed into RNA where the HDV-like ribozyme co-
1134 transcriptionally cleaves, leaving a hairpin with a 2′-3′-cyclic phosphate. For mismatch
1135 experiments, that cyclic phosphate can be removed during incubation with PNK. The hairpins are
1136 then extended by the tC19Z polymerase ribozyme. Following this reaction, the RNA is ligated to
1137 a DNA oligo with a randomized 5′-end (#24/#25) and the resulting construct is reverse-transcribed
1138 with RT primer #29. The Illumina adapters (#30 and #31) are then added via PCR prior to
1139 sequencing.

1140 **Supplementary Table 1. DNA and RNA sequences.**

1141 T7 RNA polymerase promoter sequences are underlined. HDV overlap sequences in the hairpin

1142 constructs are italicized. The terminal match or mismatch in the hairpin sequences is bolded.

1	tC19Z Ribozyme with T7 RNAP promoter sequence (underlined)	GCTAATACGACTCACTATA GTCATTGAAAAAAAAAGACAAATCTGCCCTCAGAGCTTGAGAACATC TTCGGATGCAGAGGAGGCAGCCTTCGGTGGCGCGATAGCGCCAAC GTTCTCAACAGACACCCAATACTCCCGCTTCGGCGGGTGGGGATAA CACCTGACGAAAAGGCGATGTTAGACACGCCCAGGTCATAATCCCC GGAGCTTCGGCTCC
2	RNA Primer	CUGCCAACCG
3	Short RNA Template	GUCAAUGACACGCUUCGCACGGUUGGCAG
4	tC19Z forward primer	GCTAATACGACTCACTATAGTCATTG
5	tC19Z reverse primer	GGAGCCGAAGCTCCGGGGATTATGAC
6	drz-Mbac-1 HDV-like ribozyme template (italicized segment overlaps with individual hairpin RNA constructs)	ATACGTATAACTATTAGCATGACCCATCGTCAACCACGAGGA*GGTTG* *TTATACTATGACGAC*
7	G/C Hairpin with drz-Mbac-1 HDV-like ribozyme overlap (italicized)	GCTAATACGACTCACTATA GTCAATGACACGCTTCGCA**C**GGTTGGCAGGCCTCTGCCAACC**G** *GTCGTCATAGTATAACAACC*
8	C/G Hairpin with drz-Mbac-1 HDV-like ribozyme overlap	GCTAATACGACTCACTATA GTCAATGACACGCTTCGCA**G**GGTTGGCAGGCCTCTGCCAACC**C***GTC* *GTCATAGTATAACAACC*
9	A/U Hairpinwith drz-Mbac-1 HDV-like ribozyme overlap	GCTAATACGACTCACTATA GTCAATGACACGCTTCGCA**T**GGTTGGCAGGCCTCTGCCAACC**A***GTC* *GTCATAGTATAACAACC*
10	U/A Hairpin with drz-Mbac-1 HDV-like ribozyme overlap	GCTAATACGACTCACTATA GTCAATGACACGCTTCGCA**A**GGTTGGCAGGCCTCTGCCAACC**T***GTC* *GTCATAGTATAACAACC*
11	G/G Hairpin with drz-Mbac-1 HDV-like ribozyme overlap	GCTAATACGACTCACTATA GTCAATGACACGCTTCGCA**G**GGTTGGCAGGCCTCTGCCAACC**G***GTC* *GTCATAGTATAACAACC*
12	C/C Hairpin with drz-Mbac-1 HDV-like ribozyme overlap	GCTAATACGACTCACTATA GTCAATGACACGCTTCGCA**C**GGTTGGCAGGCCTCTGCCAACC**C***GTC* *GTCATAGTATAACAACC*
13	A/A Hairpin with drz-Mbac-1 HDV-like ribozyme overlap	GCTAATACGACTCACTATA GTCAATGACACGCTTCGCA**A**GGTTGGCAGGCCTCTGCCAACC**A***GTC* *GTCATAGTATAACAACC*
14	U/U Hairpin with drz-Mbac-1 HDV-like ribozyme overlap	GCTAATACGACTCACTATA GTCAATGACACGCTTCGCA**T**GGTTGGCAGGCCTCTGCCAACC**T***GTC* *GTCATAGTATAACAACC*
15	G/A Hairpin with drz-Mbac-1 HDV-like ribozyme overlap	GCTAATACGACTCACTATA GTCAATGACACGCTTCGCA**A**GGTTGGCAGGCCTCTGCCAACC**G***GTC* *GTCATAGTATAACAACC*
16	G/U Hairpin with drz-Mbac-1 HDV-like ribozyme overlap	GCTAATACGACTCACTATA GTCAATGACACGCTTCGCA**T**GGTTGGCAGGCCTCTGCCAACC**G***GTC* *GTCATAGTATAACAACC*
17	C/A Hairpin with drz-Mbac-1 HDV-like ribozyme overlap	GCTAATACGACTCACTATA GTCAATGACACGCTTCGCA**A**GGTTGGCAGGCCTCTGCCAACC**C***GTC* *GTCATAGTATAACAACC*
18	C/U Hairpin with drz-Mbac-1 HDV-like ribozyme overlap	GCTAATACGACTCACTATA GTCAATGACACGCTTCGCA**T**GGTTGGCAGGCCTCTGCCAACC**C***GTC* *GTCATAGTATAACAACC*
19	A/G Hairpin with drz-Mbac-1 HDV-like ribozyme overlap	GCTAATACGACTCACTATA GTCAATGACACGCTTCGCA**G**GGTTGGCAGGCCTCTGCCAACC**A***GTC* *GTCATAGTATAACAACC*
20	A/C Hairpin with drz-Mbac-1 HDV-like ribozyme overlap	GCTAATACGACTCACTATA GTCAATGACACGCTTCGCA**C**GGTTGGCAGGCCTCTGCCAACC**A***GTC* *GTCATAGTATAACAACC*

21	U/C Hairpin with drz-Mbac-1 HDV-like ribozyme overlap	GCTAATACGACTCACTATA GTCAATGACACGCTTCGCA**C**GGTTGGCAGGCCTCTGCCAACC**T***GTC GTCATAGTATAACAACC*
22	U/G Hairpin with drz-Mbac-1 HDV-like ribozyme overlap	GCTAATACGACTCACTATA GTCAATGACACGCTTCGCA**G**GGTTGGCAGGCCTCTGCCAACC**T***GTC GTCATAGTATAACAACC*
23	Long read template with T7 RNAP promoter sequence (underlined)	CTGCCAACCGTGCGAAGCGTGTGCGAAGCGTGTGCGAAGCGTGTG CGAAGCGTGTGCGAAGCGTGTGCGAAGCGTGTGCGAAGCGTGTGC GAAGCGTGTGCGAAGCGTGTGCGAAGCGTGTCATTGAC TATAGTGAGTCGTATTAGC
24	Ligation primer for sequencing	5Phos/NNNNACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAG ATC/3AmMO
25	Ligation primer for sequencing	5Phos/NNNNGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAG ATC/3AmMO
26	cDNA Ligation Primer for Sequencing of long reads	5Phos/CCTATAGTGAGTCGTATTAAAAAAATATAGGNNNNNNN/3SpC3
27	Primer for Long Read Sequencing overlapping sequences #26 and #2	CGACTCACTATAGG CTGCCAA
28	Illumina reverse primer to Extend from primer 27 and overlap primer #2	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCT CACTATA GGCTGCCAACCG
29	RT primer for used after ligation of RNA with #24 and #25	GATCTACACTCTTTCCCTACACG
30	Illumina (R) adapter for hairpin sequencing overlapping the 5' terminus of the hairpin template strand (green)	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC GTCAATGACACGCTTCGC
31	TrueSeq Universal Fwd to amplify after RT with #29 from ligated oligos #24&25	AATGATACGGCGACCACCGA GATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

1143

1144 **Supplementary Table 2. Correlation of figures to deposited sequencing data.**

1145

Dataset	Main Figures	Extended Data and SI Figures
tC19Z_MM_1	2f (U/G), 2g (U/G)	Extended Data Fig. 4C (U/G), Extended Data Fig. 5F
tC19Z_MM_2	2f (G/U), 2g (G/U)	Extended Data Fig. 4C (G/U)
tC19Z_MM_3	2d, 2e, 2f (A/C), 2g (A/C)	Extended Data Fig. 4C (A/C), Extended Data Fig. 5D
tC19Z_MM_4	2f (C/A), 2g (C/A)	Extended Data Fig. 4C (C/A), Extended Data Fig. 5E
tC19Z_MM_5	2f (A/A), 2g (A/A)	Extended Data Fig. 4C (A/A)
tC19Z_MM_6	2f (A/G), 2g (A/A)	Extended Data Fig. 4C (A/G), Extended Data Fig. 5B
tC19Z_MM_7	2f (G/G), 2g (G/G)	Extended Data Fig. 4C (G/G), Extended Data Fig. 5A
tC19Z_MM_8	2f (G/A), 2g (G/A)	Extended Data Fig. 4C (G/G), Extended Data Fig. 5C
tC19Z_MM_9	2f (U/U), 2g (U/U)	Extended Data Fig. 4C (U/U)
tC19Z_MM_10	2f (U/C), 2g (U/C)	Extended Data Fig. 4C (U/C), Extended Data Fig. 6
tC19Z_MM_11	2f (C/C), 2g (C/C)	Extended Data Fig. 4C (C/C), Extended Data Fig. 5H
tC19Z_MM_12	2f (C/U), 2g (C/U)	Extended Data Fig. 4C (C/U), Extended Data Fig. 5G
tC19Z_MA_1	4e	Extended Data Fig. 4B (U/A)
tC19Z_MA_2	4c, 4d	
tC19Z_MA_3	4f	
tC19Z_MA_4	4g	Extended Data Fig. 4B (C/G)
tC19Z_MA_NoPPi_1		Extended Data Fig. 4A (U/A)
tC19Z_MA_NoPPi_2		Extended Data Fig. 4A (C/G)
tC19Z_LT_3_0	3b, 3c (0 mM PPi)	Supplementary Fig. 2
tC19Z_LT_3_2.5	3b, 3c (2.5 mM PPi)	
tC19Z_LT_3_10	3b, 3c (10 mM PPi)	Supplementary Fig. 2
tC19Z_LT_4_0	3b, 3d (0 mM PPi)	
tC19Z_LT_4_3	3b, 3d (3 mM PPi)	
tC19Z_LT_4_10	3b, 3d (10 mM PPi)	
tC19Z_LT_0_0	3b (0 mM PPi)	

1146 These datasets are available in NLM under SRA PRJNA1174777.

1147

1148 **Supplementary Table 3: Sequencing data in Figure 3.**

1149

Dataset	Total seq.	% Seq. with primer	% Extended
0 hours, 0 mM PPi	50825	14.7	0.08
3 hours, 0 mM PPi	101862	30.9	0.83
3 hours, 2.5 mM PPi	56588	37.9	0.77
3 hours, 10 mM PPi	42930	35.6	0.52
4 days, 0 mM PPi	140423	25.2	1.11
4 days, 3 mM PPi	65769	17.1	1.39
4 days, 10 mM PPi	26457	22.3	1.13

1150

Supplementary Information: Full Length Gels

Boxes have been drawn around relevant lanes.

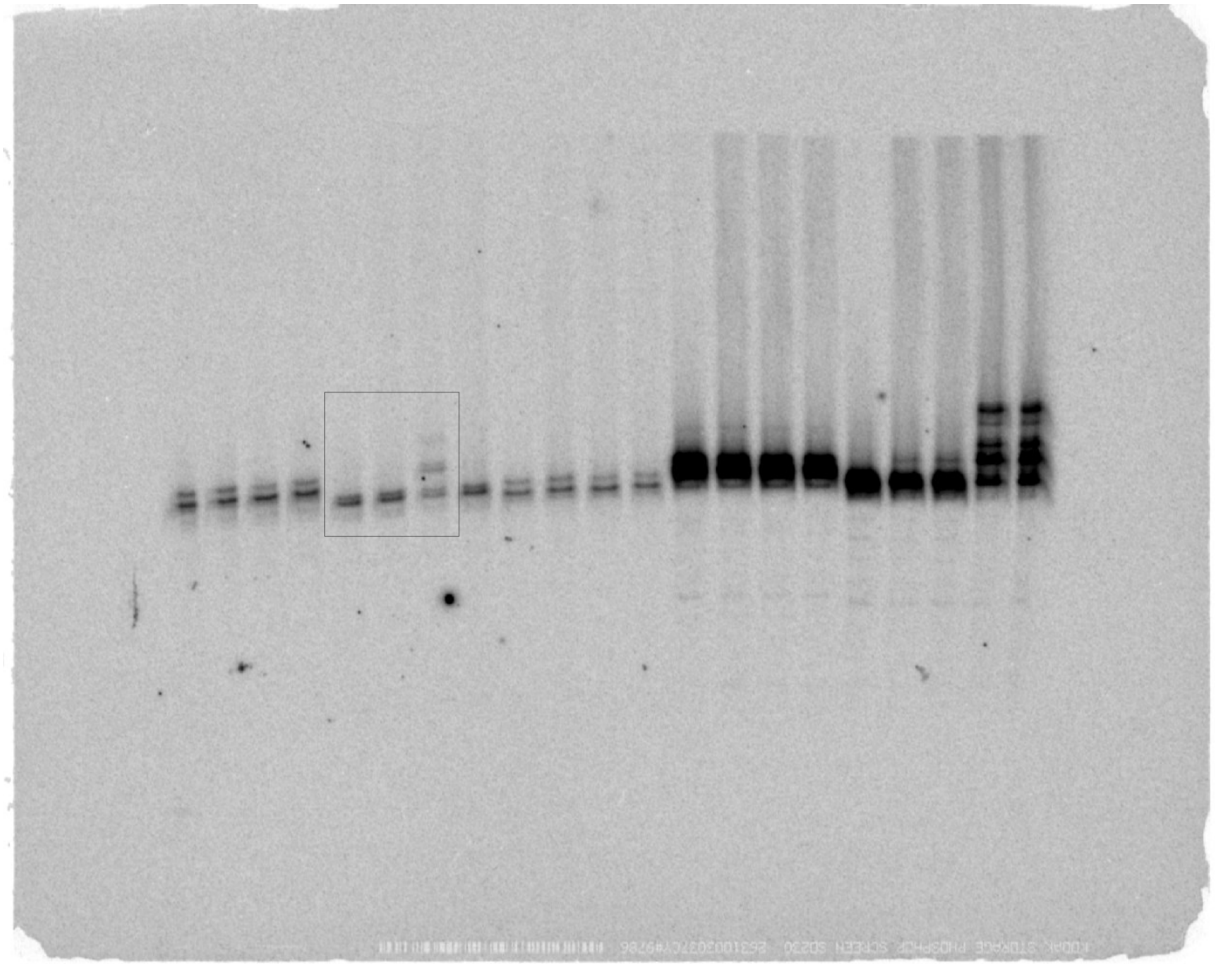
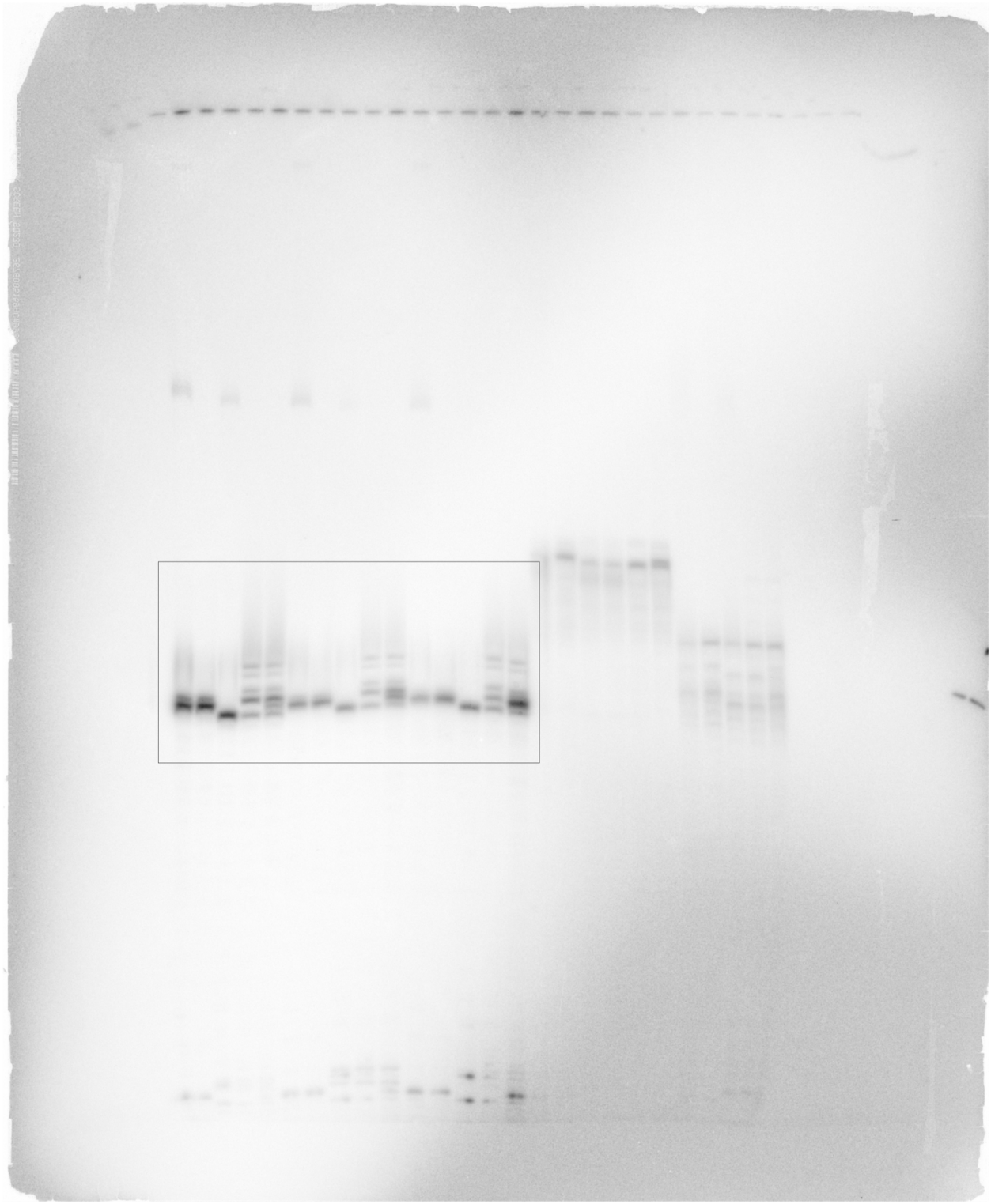
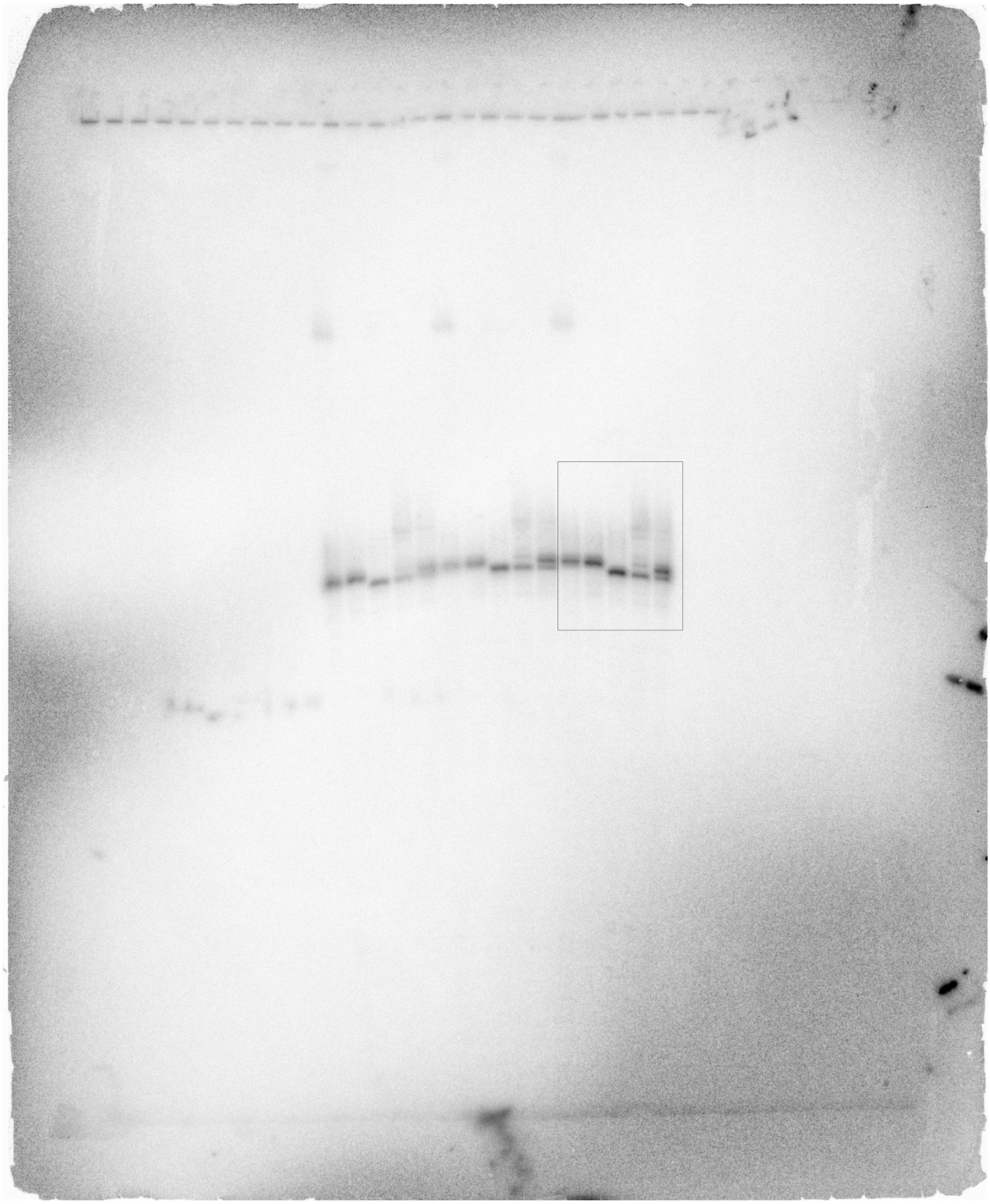


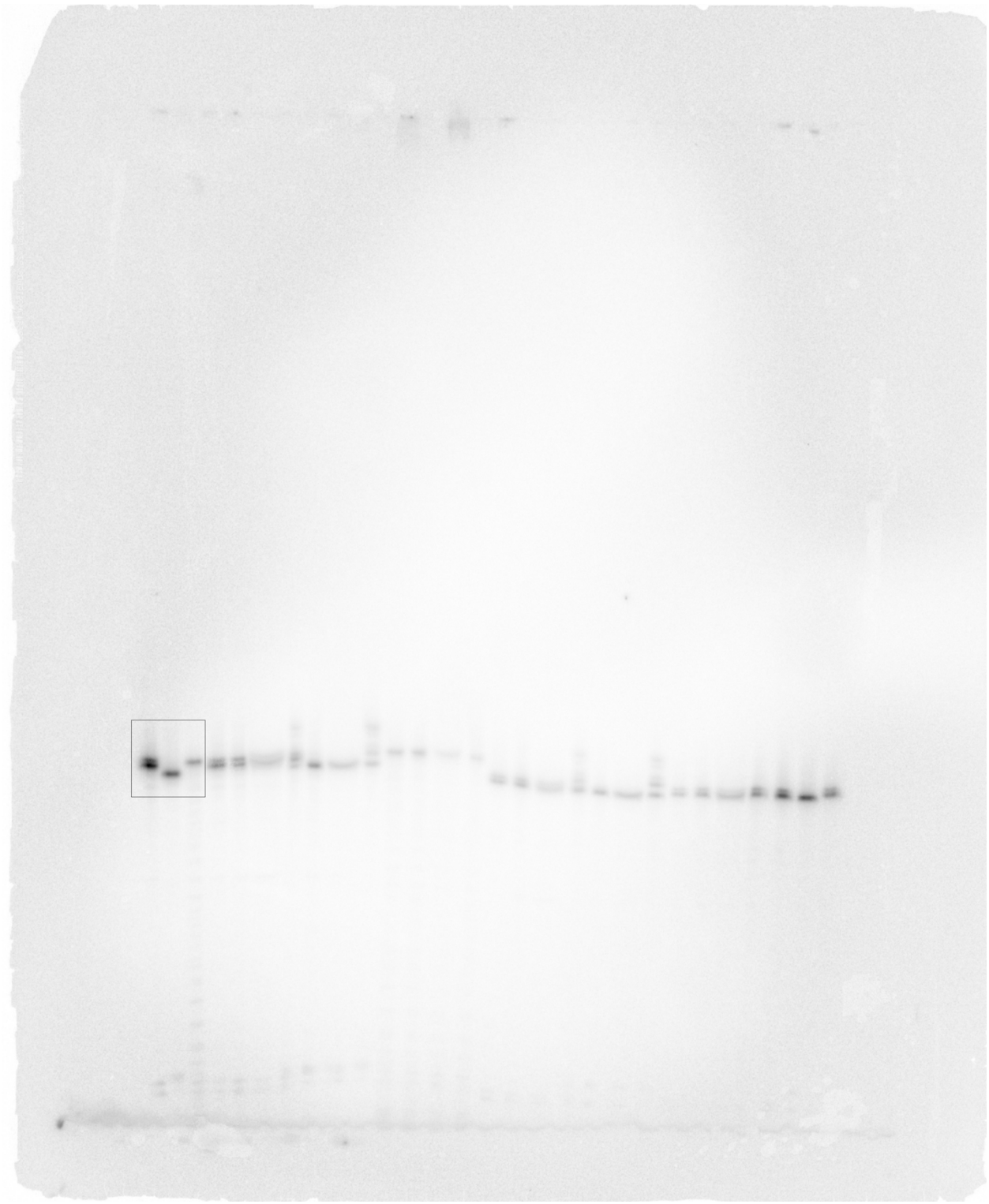
Figure 2b



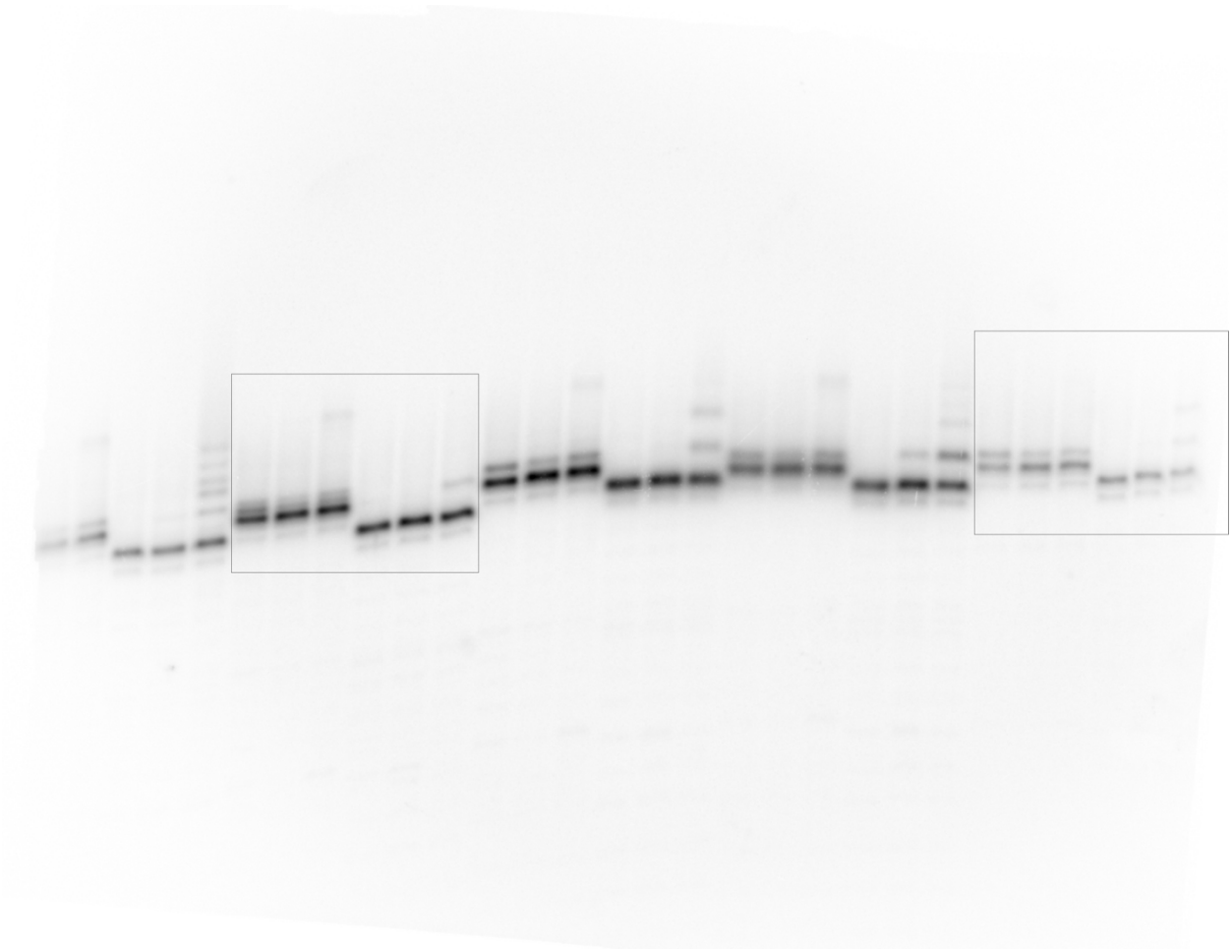
Figures 4d, 4e, 4g and Extended Data Fig. 8 (A/U, U/A, C/G)



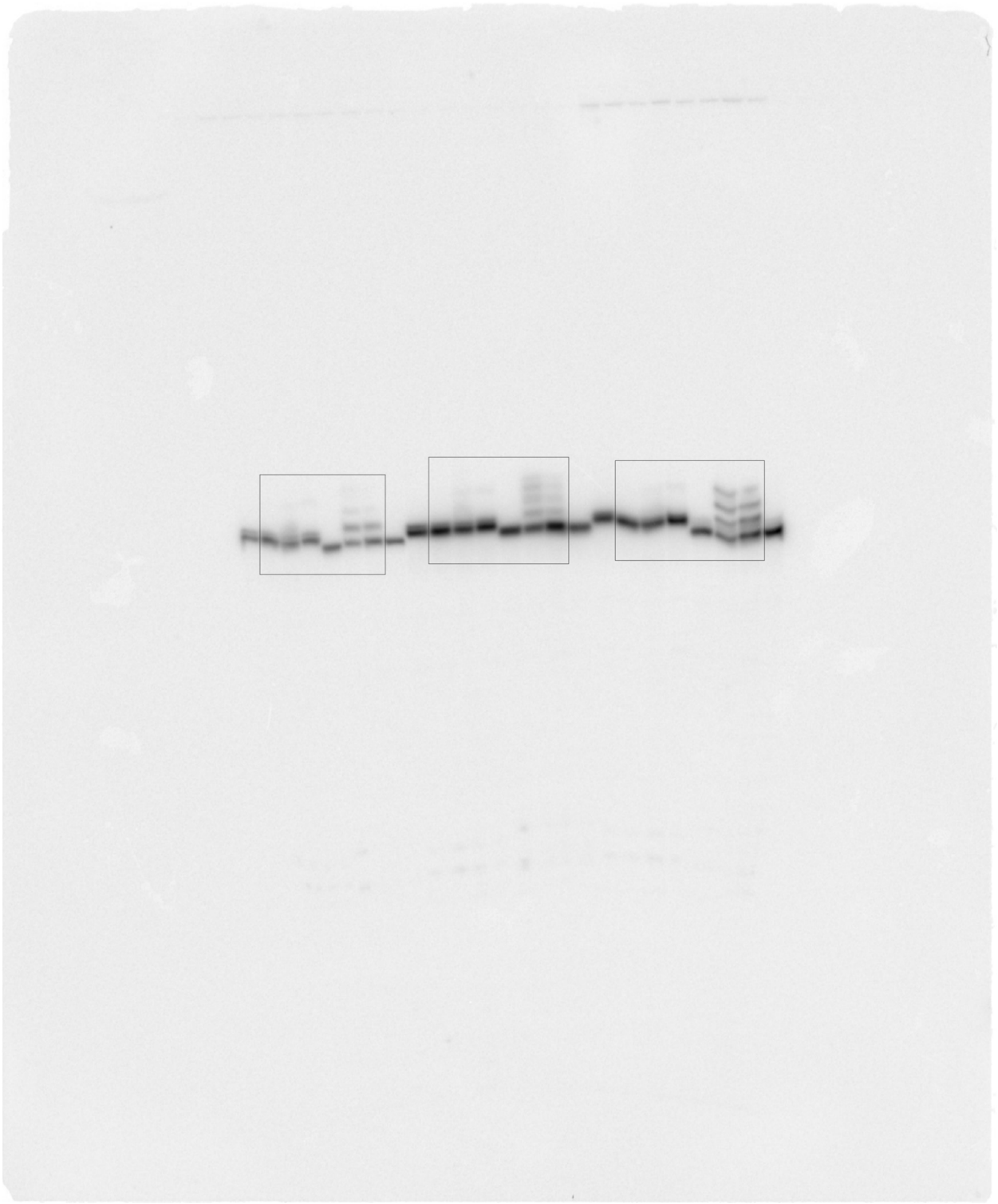
Extended Data Fig. 8 (G/C)



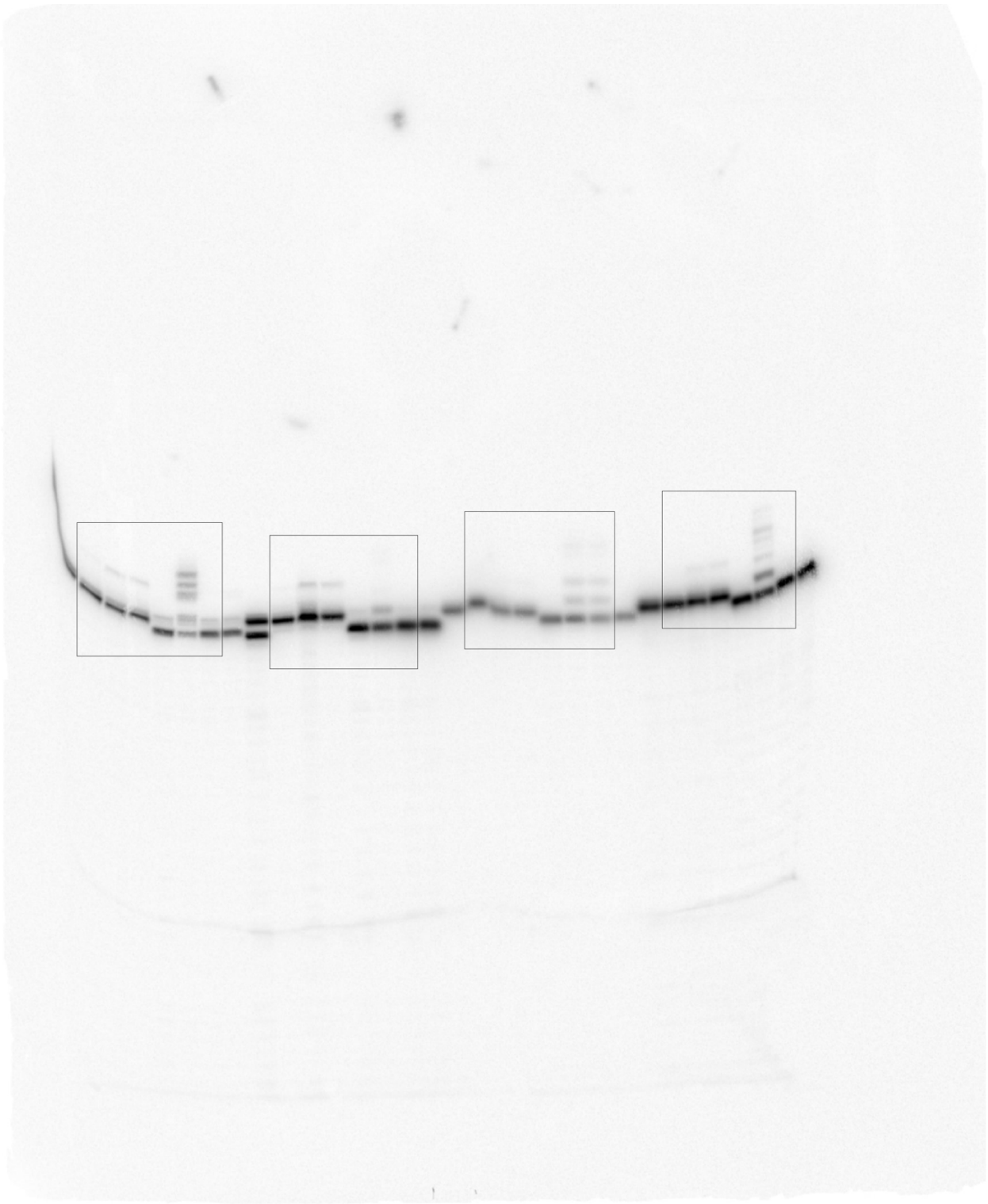
Extended Data Fig. 1



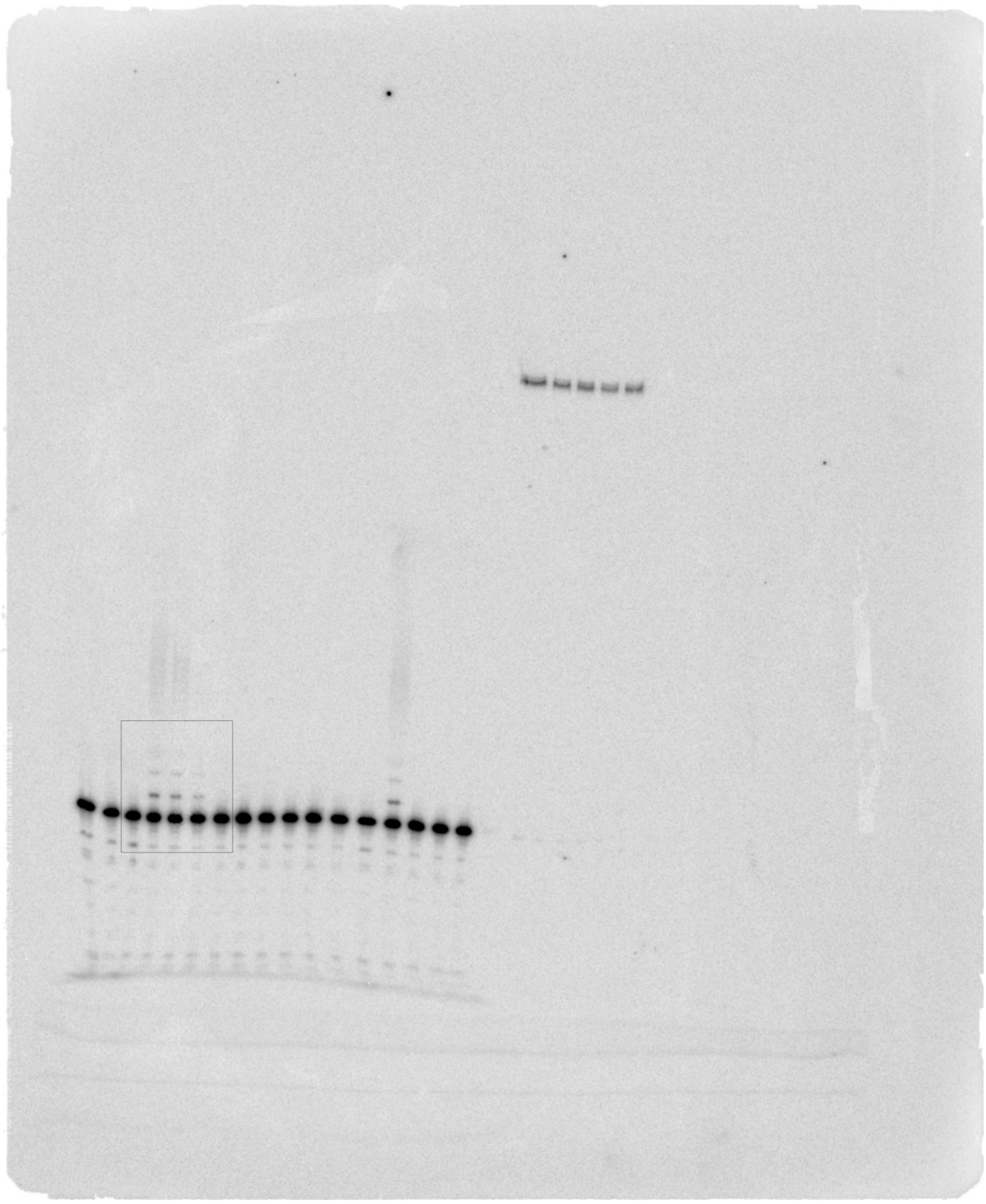
Extended Data Fig. 2: A/A left, G/G right



Extended Data Fig. 2: C/C left, U/C middle, U/U left

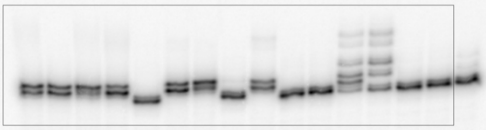


Extended Data Fig. 2: U/G left, G/A left middle, C/U right middle, G/U right

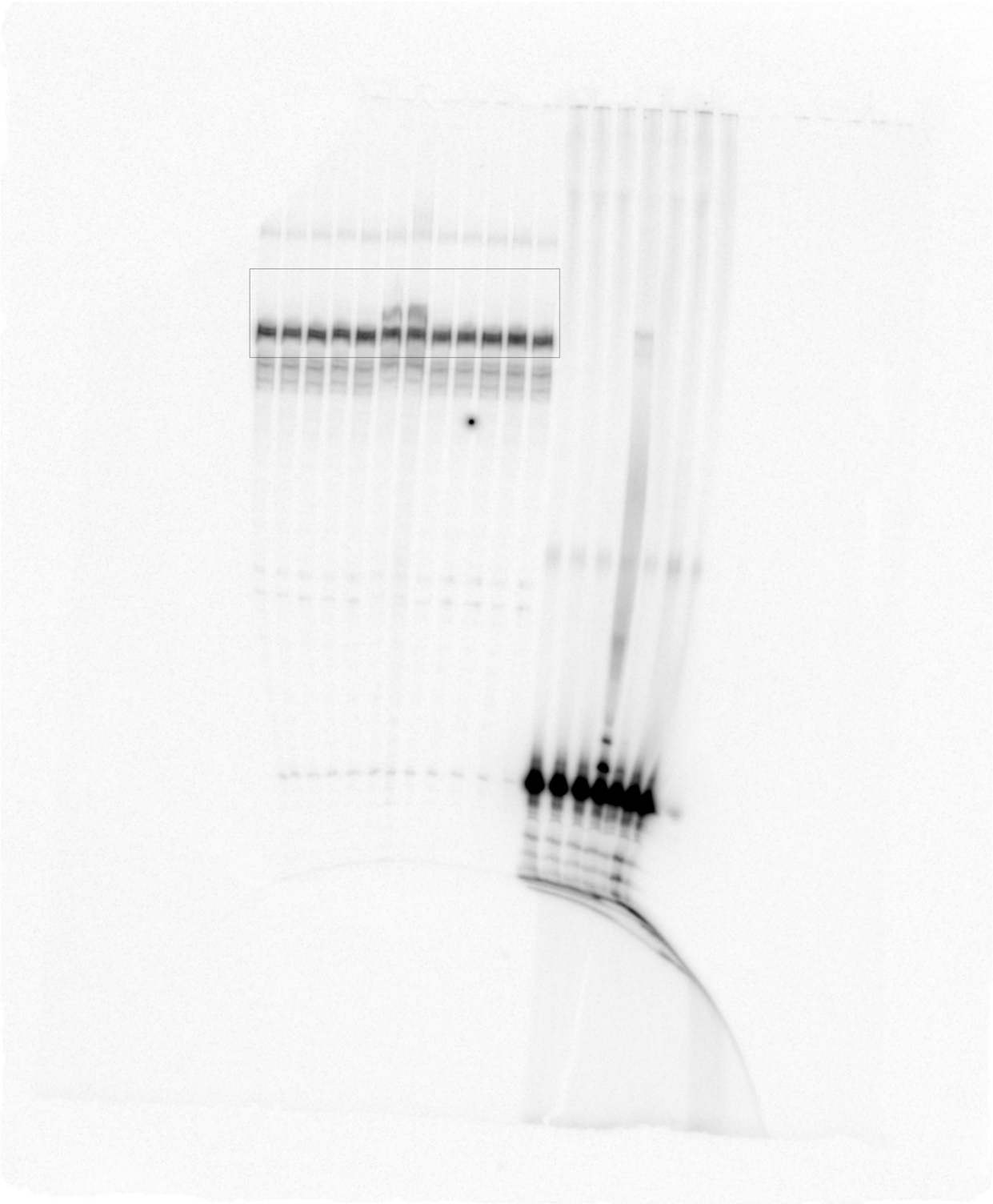


Supplementary Fig. 1

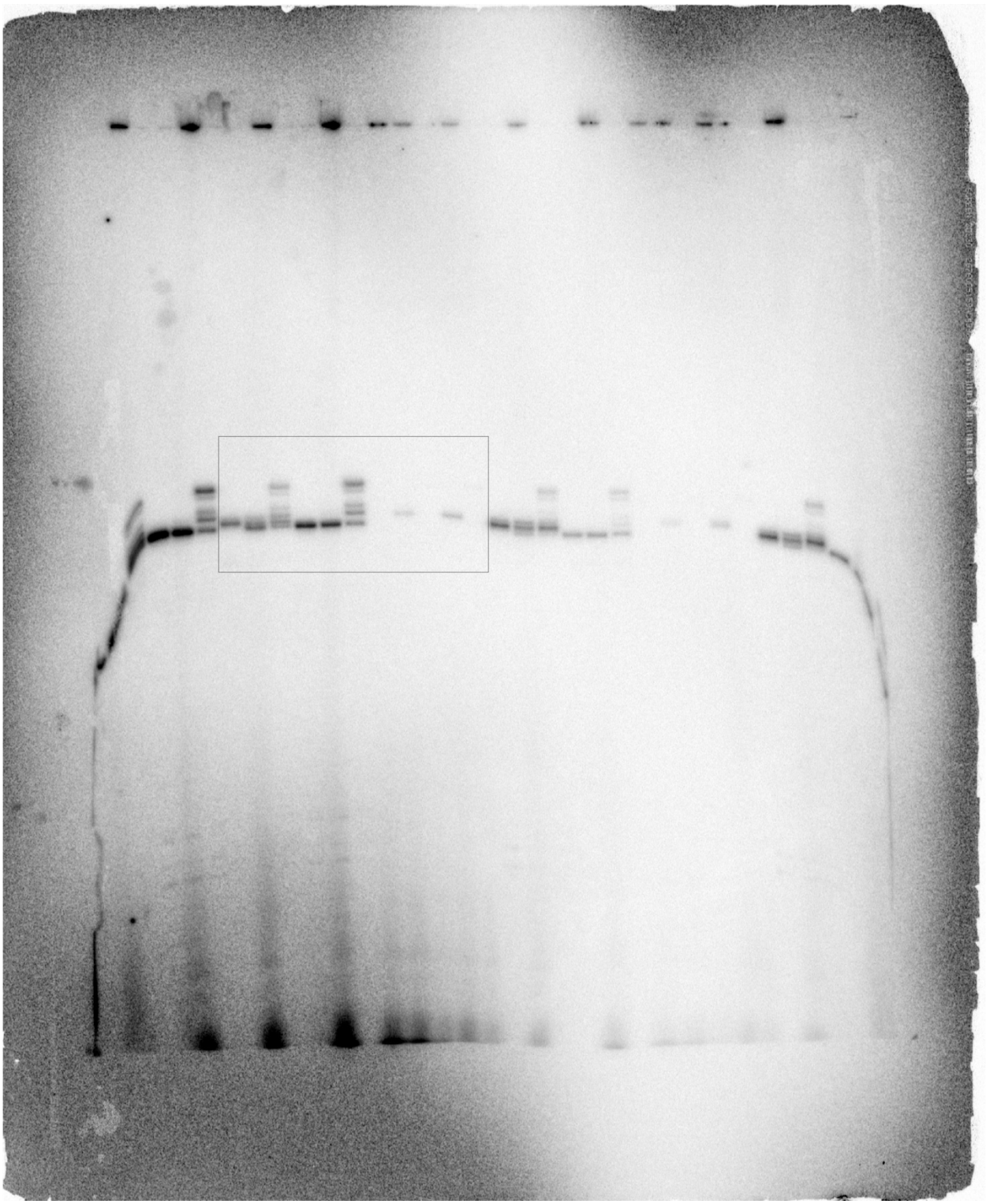
Extended Data Fig. 7



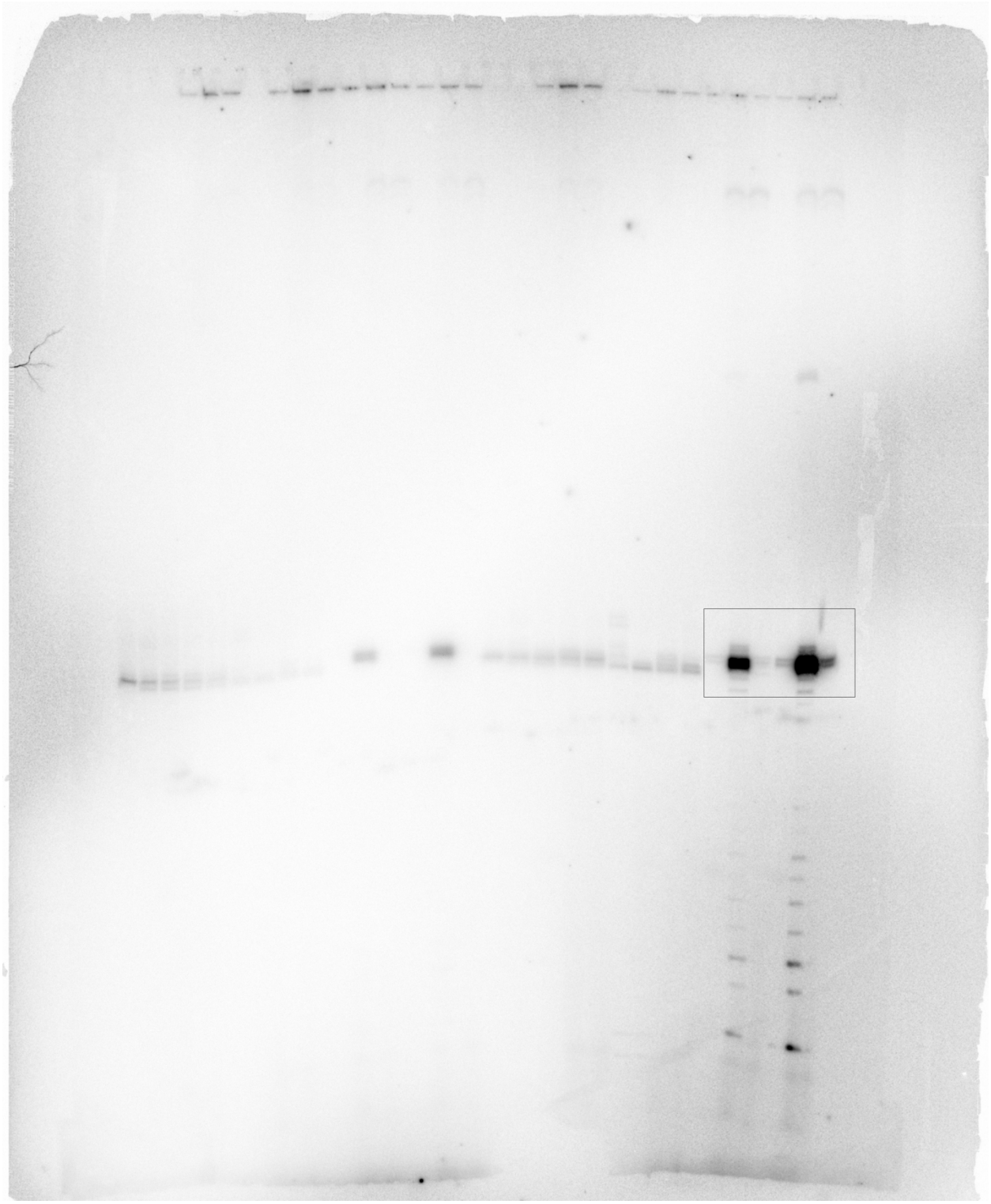
Supplementary Fig. 4



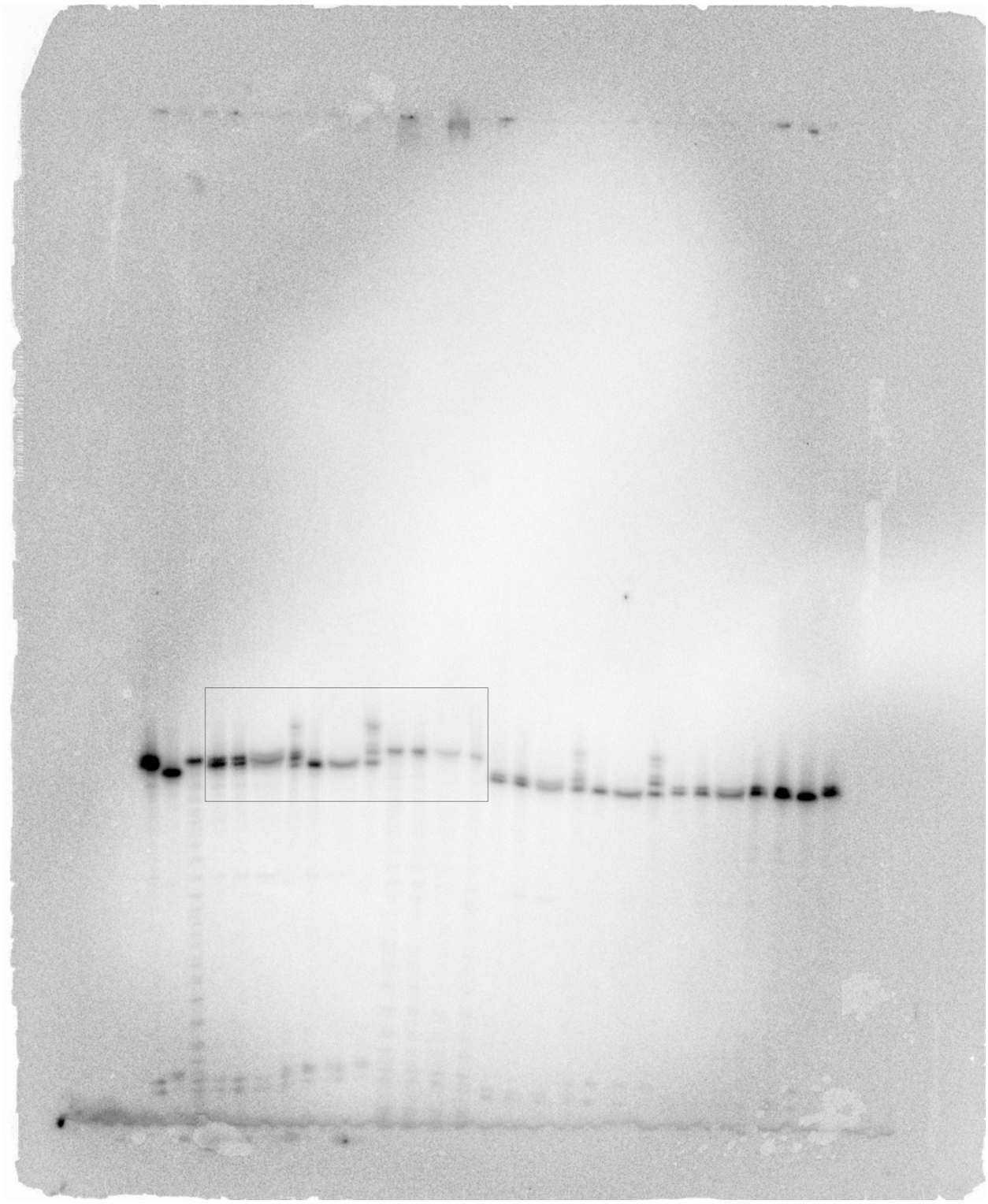
Supplementary Fig. 5



Supplementary Fig. 6A



Supplementary Fig. 6B



Supplementary Fig. 7