Ribozyme-catalysed RNA synthesis using triplet building blocks

Authors: James Attwater¹, Aditya Raguram†, Alexey S. Morgunov¹, Edoardo Gianni¹ & Philipp Holliger¹*

Affiliations:
1 MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge, CB2 0QH, UK.
† Current address: Harvard University, Department of Chemistry and Chemical Biology, 14 Oxford Street, Cambridge, MA 02138, USA.
*Correspondence to: Philipp Holliger (ph1@mrc-lmb.cam.ac.uk).
RNA-catalyzed RNA replication is widely believed to have supported a primordial biology. However, RNA catalysis is dependent upon RNA folding, and this yields structures that can block replication of such RNAs. To address this apparent paradox we have re-examined the building blocks used for RNA replication. We report RNA-catalysed RNA synthesis on structured templates when using trinucleotide triphosphates (triplets) as substrates, catalysed by a general and accurate triplet polymerase ribozyme that emerged from in vitro evolution as a mutualistic RNA heterodimer. The triplets cooperatively invaded and unraveled even highly stable RNA secondary structures, and support non-canonical primer-free and bidirectional modes of RNA synthesis and replication. Triplet substrates thus resolve a central incongruity of RNA replication, and here allow the ribozyme to synthesise its own catalytic subunit ‘+’ and ‘−’ strands in segments and assemble them into a new active ribozyme.
Introduction

The premise that some RNA sequences can catalyse and template their own replication - synthesizing both their own ‘+’ and ‘−’ strands - underpins current thinking about early genetic systems (Crick, 1968; Orgel, 1968; Szostak et al., 2001). Any ancient ribozyme with such RNA replicase capability seems to be lost, but efforts to recreate RNA self-replication in the laboratory are ongoing (Martin et al., 2015) as a critical test of the “RNA world” hypothesis (Gilbert, 1986).

Early on, derivatives of naturally occurring self-splicing introns (Doudna et al., 1991; Green & Szostak, 1992; Hayden & Lehman, 2006) as well as later in vitro evolved ligase ribozymes (Lincoln & Joyce, 2009; Szczepanski & Joyce, 2014) were shown to be able to assemble one of their own strands from cognate constituent RNA segments. A critical drawback of such systems is their use of specific preformed building blocks of at least 8 nucleotides (nt) average length, limiting their potential for open ended evolution, and precluding their replication from pools of random-sequence oligonucleotide substrates (Green & Szostak, 1992; Doudna et al., 1993).

In a contrasting approach, RNA polymerase ribozymes (RPRs) have been developed that can use general monomer building blocks (ribonucleoside 5′ triphosphates (NTPs)) in RNA-templated RNA synthesis (Johnston et al., 2001; Zaher & Unrau, 2007; Wochner et al., 2011; Attwater et al., 2013b; Horning & Joyce, 2016), akin to the activity of modern proteinaceous polymerases. However, even the most highly-evolved RPRs (Horning & Joyce, 2016) are substantially impeded by template secondary structures. Such structures are ubiquitous in larger, functional RNAs (including the RPRs themselves) and generally indispensable for function. The strong inhibitory role of this central feature of RNA leads to an antagonism between the degree to which an RNA sequence is able to fold into a defined three-dimensional structure to encode function (such as catalysis) and the ease with which it can be replicated (Boza et al., 2014). This ostensible “structure vs. replication” paradox would have placed stringent probability constraints...
on the emergence of an RNA replicase and generally impeded the ability of RNA to function as an early genetic polymer.

We wondered whether this paradox might be evaded through a re-consideration of plausible building blocks for early RNA replication. Models of non-enzymatic polymerisation of all four activated ribonucleotides – the presumed source of the first RNA sequences – yield pools of di-, tri- and tetranucleotide- etc. length oligonucleotides (in decreasing abundance) dominating the population alongside longer products (Monnard et al., 2003). Here we have examined whether substrates of such lengths can support RNA-catalyzed RNA replication, by developing a ribozyme capable of iterative templated ligation of 5’-triphasphorylated RNA trinucleotides (henceforth called triplets). This heterodimeric triplet polymerase ribozyme demonstrated a striking capacity to copy a wide range of RNA sequences, including highly structured, previously intractable RNA templates, as well as its own catalytic domain and encoding template in segments. Its characterization revealed emergent properties of triplet-based RNA synthesis, including cooperative invasion and unraveling of stable RNA structures by triplet substrates, bi-directional (both 5’-3’ and 3’-5’) and primer-free (triplet-initiated) RNA synthesis, and fidelity augmented by systemic properties of the random triplet pools.
Results

**In vitro evolution of triplet polymerase activity**

We set out to explore the potential of short RNA oligonucleotides as substrates for RNA-catalyzed RNA replication. To do this we required a ribozyme capable of general, iterative RNA-templated oligonucleotide ligation. Previously-described RNA polymerase ribozymes such as the ‘Z’ RPR (Wochner et al., 2011) can use NTPs to iteratively extend a primer hybridized to an RNA template, but do not accommodate oligonucleotides bound downstream of the primer or accept them as substrates. However, we detected a weak templated ligation activity in a truncated version of the Z RPR comprising its catalytic core domain (Zcore) (Figure 1a), which supported incorporation of oligonucleotide substrates as short as 3 nt (Figure 1 - figure supplement 1), when incubated in the eutectic phase of water ice (Attwater et al., 2010; Mutschler et al., 2015).

To be able to properly examine such RNA trinucleotide triphosphates (triplets) as replication substrates, we first sought to convert Zcore into an effective triplet polymerase ribozyme using *in vitro* evolution. We devised a selection strategy that required iterative templated triplet ligation by ribozymes to achieve their *in cis* covalent linkage to a tagged primer (Figure 1 - figure supplement 2). This enables their recovery, amplification and mutagenesis before further rounds of selection to enrich the selection pool in improved triplet polymerase ribozyme variants.

We initiated selections from a library of $1.5 \times 10^{15}$ Zcore variants with a new random $3' N_{30}$ region under eutectic phase conditions that increase RNA half-life and enhance ribozyme activity (Attwater et al., 2010). After 7 rounds of in-ice evolution (Attwater et al., 2013b) one-quarter of the selection pool comprised an improved ribozyme (type 0). Its core domain (0core, Figure 1a) could catalyse the iterative polymerization of multiple triplets allowing us to begin to investigate the properties of triplet-based RNA replication.
Significantly, we found that 0core could catalyze triplet polymerisation on a series of structured templates, which had proven intractable to the parental Z RPR (Figure 1b). Here, primer extension exhibited a steep sigmoidal dependence upon triplet concentrations (Figure 1c), suggestive of a cooperative invasion and unraveling of template secondary structures by the triplet substrates themselves. Although still inefficient, the fact that the nascent activity of the 0core ribozyme could already copy templates that had confounded an established RPR encouraged us to continue to seek improved triplet polymerase ribozymes to leverage this substrate behaviour.

Emergence of cooperativity and characterisation of a ribozyme heterodimer

We continued selections for a further 14 rounds. At this point, the type 0 ribozyme had gone extinct, replaced by six new types of RNA each characterised by a unique 3’ domain (Figure 2a, Figure 2 - figure supplement 1). Type 1 RNAs were most abundant, comprising ~50% of pool sequences but mysteriously were catalytically inactive with diverse mutations in their core domains. In contrast, the type 2-6 RNAs all displayed triplet polymerase activity, but fell short of the polyclonal activity of the selection pool (Figure 2 - figure supplement 2). To attempt to explain this discrepancy, we explored potential interactions among the different pool lineages, and found that addition of an equimolar amount of type 1 RNA substantially enhanced triplet polymerase activity of all the other ribozyme types 2-6 (Figure 2b).

Dissecting type 1 RNA function, we found that 5’ truncation of the region that previously contacted the primer/template duplex (Shechner et al., 2009) did not affect its cofactor activity (Figure 3a, Figure 3 - figure supplement 1). As judged by gel mobility shift (Figure 3b) and activity enhancement (Figure 3 - figure supplement 1), type 1 RNA appears to form a 1:1 heterodimeric complex directly with active triplet polymerase ribozymes. Our attention was drawn to their selection construct-derived 5’ hairpin elements, which differed between active
triplet polymerases (‘cap+’, Figure 3a, Figure 2 - figure supplement 1) and the most common type 1 variants in the selection pool where this hairpin had acquired a mutation (yielding ‘cap–’, Figure 3a). ‘cap–’ was dispensible for type 1’s cofactor activity, but when replacing ‘cap+’ in active triplet polymerases it abolished both their activity enhancement by type 1 (Figure 3 - figure supplement 1) and complex formation (Figure 3b). This points to the ‘cap+’ hairpin as the critical site of interaction with type 1; ‘cap–’ in type 1 presumably served to deter its homodimerisation.

Indeed, transplanting the ‘cap+’ element could make the parental ribozymes (Zcore & Z RPR) receptive to activity enhancement by type 1 RNA (Figure 2b, Figure 3 - figure supplement 2). The catalytically inert type 1 RNA thus represents a general, mutualistic RNA species. This molecular symbiont appears to have emerged spontaneously during in vitro evolution by forming a heterodimeric holoenzyme with triplet polymerase ribozymes, enhancing their activity and their joint recovery prospects.

In complex with type 5 (the fastest enriching triplet polymerase ribozyme in the final selection pool), type 1 boosts polymerization of triplets (or longer oligonucleotides) to enable synthesis of long RNAs (Figure 3c). Here it became apparent that type 1 also obviates the need for ribozyme-template tethering. Due to their poor affinity for primer/template duplex (Lawrence & Bartel, 2003), RPRs generally depend upon such tethering to template (Attwater et al., 2010; Wochner et al., 2011; Horning & Joyce, 2016), which enhances local ribozyme concentration and promotes formation of the RPR-primer/template holoenzyme (Attwater et al., 2010; Attwater et al., 2013a). In contrast, the triplet polymerase heterodimer appears to have a capacity for true intermolecular, sequence-general interaction with primer-template duplexes, which enables holoenzyme formation and copying of RNA templates without requiring specific ribozyme-template hybridization sites.

Secondary structure invasion by triplet substrates
We performed an additional five rounds of in vitro evolution to further evolve the type 5 triplet polymerase ribozyme (now in the presence of truncated type 1 RNA), diversifying the previously-fixed 3’ domain reverse transcription primer binding sequence. This reselection yielded a shorter final heterodimeric triplet polymerase holoenzyme, hereafter termed ‘t5+1’, (Figure 4). This robust triplet polymerase activity now proved suitable for exploring the scope and potential of triplet-based RNA replication.

As a first examination of t5+1 activity, we revisited triplet-based RNA synthesis on structured templates. To provide a stringent test of template structure inhibition, we now examined hairpin-containing templates (4S, 6S, 8S) with increasing RNA hairpin stability and estimated TMs of up to 93°C (8S). The latter had previously strongly arrested even the most advanced mononucleotide RPRs at higher temperatures (Horning & Joyce, 2016). However, using triplets as substrates t5+1 robustly copied all of these (Figure 5a), even when templates were pre-folded allowing RNA secondary structures to form prior to triplet addition (Figure 5 - figure supplement 1). The triplet concentration-dependent cooperative structure invasion and unraveling (previously observed with the simpler 0core domain and partly wobble-paired RNA template structures (Figure 1b, c)) was recapitulated with t5+1 and the highly stable 8S hairpin template (Figure 5 - figure supplement 2). In contrast, dinucleotide triphosphate substrates yielded extension only up to the structured region (Figure 5 - figure supplement 2).

We began to explore whether triplet-based RNA synthesis by t5+1 might exhibit the generality required not just for synthesis of arbitrary structured sequences, but for replication of functional sequences (requiring synthesis of both ‘+’ and ‘−’ strands). Encouragingly, t5+1 could synthesise both a functional fluorescent ‘+’ strand of the 52 nt Broccoli RNA aptamer (Filonov et al., 2014) and its encoding ‘−’ strand template from their 13 (+) & 12 (−) different constitutive triplets (Figure 5b).
Ribozyme sequence self-synthesis and assembly

We next turned to the critical test of generality: could triplet substrates allow self-synthesis? As t5\textsuperscript{+1} currently lacks the efficiency to synthesize RNAs its own length, we divided the catalytic t5 ribozyme into five segments \( \alpha, \beta, \gamma, \delta \), and \( \varepsilon \). This segmentation strategy (akin to that used by some RNA viruses e.g. influenza) could reduce tertiary structures (Doudna et al., 1991; Mutschler et al., 2015) and ease product separation during RNA replication (Szostak, 2012). Starting from ~8 nt RNA primers, t5\textsuperscript{+1} achieved synthesis of the \( \beta^+, \gamma^+, \) and \( \delta^+ \) segments from their constitutive triplets as well as all of the ‘–’ strand segments \( \alpha^-, \beta^-, \gamma^-, \delta^- \), and \( \varepsilon^- \), but required some triplets pre-linked (as e.g. hexanucleotides) for synthesis of full-length \( \alpha^+ \) and \( \varepsilon^+ \) segments (Figure 6a).

Operating across 70 distinct ligation junctions in these reactions including AU-rich sequences, t5\textsuperscript{+1} using triplet substrates demonstrates the sequence generality for self-synthesis. Notably, the average extent of ligation per junction during synthesis of t5 ‘+’ and ‘–’ strands (78%) was similar to that observed when t5\textsuperscript{+1} used an unstructured model template (74%, Figure 3c) upon which the parental Z and other RPRs excel (Attwater et al., 2013b; Horning & Joyce, 2016).

At this point we tested whether the broad oligonucleotide ligation capacity of t5\textsuperscript{+1} (Figure 3c) might allow assembly of synthesised ‘+’ strand segments. Indeed, t5\textsuperscript{+1} could assemble these into \( \alpha\beta^+ \) and \( \gamma\delta\varepsilon^+ \) fragments, guided only by partially overlapping ‘–’ strands (Figure 6b, Figure 6 - figure supplement 1). Through non-covalent association (Vaish et al., 2003; Mutschler et al., 2015), the ribozyme-synthesised \( \alpha\beta^+ \) and \( \gamma\delta\varepsilon^+ \) fragments spontaneously reconstituted a new catalytically active triplet polymerase ribozyme (with \textit{in vitro} transcribed type 1 RNA). We found that this synthesis product could regenerate fresh \( \delta^- \) segment using t5\textsuperscript{+1} ribozyme-synthesised \( \delta^+ \) (left over from ribozyme assembly) as a template (Figure 6c), recapitulating elements of a self-
replication cycle. However, while the t5+1 ribozyme displays a nascent capacity for templated
synthesis of its own catalytic domain ‘+’ strands (and ‘−’ strands), efficiency of both segment
synthesis and assembly will need to be increased significantly to realise a full self-replication
cycle (which would also require synthesis and replication of the type 1 subunit).

Primer-free RNA synthesis

Templated ‘+’ strand self-synthesis is a central element of ribozyme self-replication. However, a limitation of our above strategy in the context of triplet-based self-replication is the
continued requirement for some pre-synthesized longer oligonucleotides to act as primers and
occasional substrates (together providing here the equivalent of ~1/4 of triplet junctions pre-
ligated). Specific oligonucleotide substrates were used in places to compete out inhibitory mutual
hybridisation between ‘−’ strand template and corresponding ‘+’ strand unstructured elements in
the t5 ribozyme (Figure 6 - figure supplement 2). In vitro selections that stabilise the ribozyme
tertiary structure (Figure 6 - figure supplement 3) may contribute to attenuating this requirement.
Additionally, more concentrated triplet substrates can successfully compete with ribozyme
unstructured elements for hybridization to ‘−’ strand templates (Figure 6 - figure supplement 2).

The majority of specific oligonucleotides, however, were provided as primers to initiate
syntheses, as required by all RPRs akin to the activity of replicative polymerases in biology. As a
consequence of this, the priming sequence would effectively be excluded from evolution.
Furthermore, RNA oligonucleotides able to act as specific primers are unlikely to be prevalent in
prebiotic substrate pools, and their depletion during successive replication cycles could lead to
sequence loss at genome ends. This ‘primer problem’ has previously been noted in the context of
nonenzymatic replication (Szostak, 2012) as one of the fundamental obstacles to RNA self-
replication.
Unexpectedly, triplet substrates provide a route to bypass the ‘primer problem’. We observed that t5\(^{+1}\) can extend primers bidirectionally, in both the canonical 5’-3’ as well as the reverse 3’-5’ directions (Figure 7a). This not only allows completion of RNA synthesis from either template end but initiation from anywhere along a template potentially allowing non-classical hierarchical or distributive RNA replication schemes as previously proposed (Szostak, 2011, 2012). Given this flexibility, we wondered if t5\(^{+1}\) even had a requirement for a primer oligonucleotide. Indeed, the t5\(^{+1}\) ribozyme could achieve ‘primer free’ RNA synthesis (whereby synthesis is presumably initiated by ligation of adjacent triplets anywhere on the template), as exemplified here for the β\(^{+}\) segment (Figure 7b), as well as ‘primer free’ RNA replication as shown for the ‘+’ and ‘−’ strands of the γ segment, which can be replicated using triplets alone (Figure 7c).

Thus the capacity of triplet substrates to pre-organise themselves on a template not only enables replication of structured templates but allows complete copying of some RNA sequences exclusively from triplet building blocks, suggesting an alternative to the canonical end-primed replication strategies inspired by PCR. Such a ribozyme operating in a more distributive polymerisation mode might be able to replicate RNA sequences directly from the putative pools of short random RNA oligonucleotides furnished by prebiotic chemistry.

Fidelity of triplet-based RNA synthesis

Next we investigated the consequences of using defined modern analogues of such prebiotic pools as a source of substrates for the t5\(^{+1}\) triplet polymerase ribozyme. Random sequence triplet pools (\(^{\text{ppp}}\)NNN, comprising equimolar amounts of all 64 triplets) could be used as substrates by t5\(^{+1}\) in segment syntheses in place of defined triplet sets (Figure 6 – figure
supplement 4). Furthermore, extension activity remained robust upon pool supplementation with noncanonical dinucleotide and mononucleotide substrates (Figure 6 – figure supplement 5).

However, a replicase must incorporate the correct template-complementary substrate from random sequence pools, or genetic information may become irretrievably corrupted during replication (Eigen, 1971). Sequence fidelity is therefore a critical parameter of RNA replication. The fidelity challenge is exacerbated in triplet-based RNA replication by the need to discriminate between 64 distinct substrates; indeed, a previous investigation into the incorporation of individual trinucleotides indicated that misincorporations could outstrip cognate incorporation for some triplets (Doudna et al., 1993).

In order to assess the fidelity of triplet polymerase ribozymes of widely differing activity, we identified the triplets incorporated from random $\text{ppp}NNN$ triplet pools using 12 different compositionally representative $\text{N}^{\prime}\text{N}^{\prime}\text{N}^{\prime}$ triplet sequences as templates in a consistent sequence context (5'-GGG-N'N'-GGG-3'), whose collation allowed an estimation of ribozyme misincorporation tendencies. On average, the starting Zcore ribozyme exhibited ~91% fidelity per position (Figure 8a), lower than that described for RPRs (92% (Horning & Joyce, 2016) – 97% (Attwater et al., 2013b)). Furthermore, its accuracy exhibited a pronounced downward gradient from the 1\textsuperscript{st} (5') to the 3\textsuperscript{rd} (3') triplet position, highlighting escalating risks to fidelity of synthesis founded on longer building blocks.

To investigate if ribozymes could exhibit higher triplet incorporation fidelity, we had included a persistent adaptive pressure for fidelity during \textit{in vitro} evolution, spiking in an excess of mispairing 3'-deoxy ‘terminator’ triplets from round 9 onwards, precluding recovery of ribozymes that incorporated these mispairs (Figure 2 - source data 1, Figure 4 - source data 1, Figure 2 - figure supplement 2). This yielded reshaped and improved fidelity profiles in the ‘surviving’ type 2-6 ribozymes (Figure 8a). Notably, the final t5$^{+1}$ ribozyme achieves an average positional fidelity of 97.4% using $\text{ppp}NNN$ in this sequence context, higher than the best RPR.
fidelity with NTPs under comparable eutectic conditions (Attwater et al., 2013b). Deep sequencing of internal triplet positions of a defined sequence (β+ segment) synthesised by t5+1 using pppNNN indicated similar fidelity could be achieved during longer product synthesis (Table 1).

**Molecular basis of triplet polymerase ribozyme fidelity**

Having established that accurate triplet-based copying is possible (in at least some sequence contexts), we sought to understand how the triplet polymerase ribozyme achieves it. Investigating the fidelity contributions of different t5+1 ribozyme components, we found that the type 1 RNA cofactor did not contribute; rather, fidelity gains appeared to be mediated by the newly-evolved t5 ‘ε’ 3’-domain, as its deletion (yielding the truncated ‘αβγδ’ ribozyme) reverted its fidelity profile towards that of Zcore (Figure 8 - figure supplement 1). Presence of the ε domain did not uniformly increase fidelity, but selectively reduced the most acute errors at the 2nd and 3rd triplet positions (with over 10-fold reductions for some errors, Figure 8b, Figure 8 - figure supplement 2). Overall error rates at the 2nd and 3rd triplet positions were reduced by 4-fold and 9-fold compared to Zcore (Figure 8a), though slightly increased (1.3-fold) at the 1st triplet position due to a localised asymmetric tolerance of G:U wobble pairing (Figure 8b). The ε domain fidelity function is contingent upon the presence of a downstream triplet, operating only with basal fidelity for final triplet incorporation (Figure 8 - figure supplement 3).

Dissecting the molecular determinants of the fidelity phenotype, we found that using triplet substrates modified at the 3rd position with a 2-thiouracil in place of a uracil (disrupting minor groove hydrogen bonding) rendered the ε fidelity domain unable to discriminate mismatches (Figure 8c, Figure 8 - figure supplement 3). Previously, a similar replacement of a uracil 2-keto group with a 2-thio modification had been shown to impair Z RPR activity when
present upstream in the primer/template region (Attwater et al., 2013a), where Z is thought to rely upon sequence-general minor groove contacts through an ‘A-minor’ motif (Shechner et al., 2009). Modification at the 3rd triplet position reverts ε’s divergent effects on fidelity at the adjacent 2nd and the distal 1st triplet positions (Figure 8 - figure supplement 3); disruption of this minor groove contact site thus abolishes overall ε fidelity domain operation. ε sensitivity to minor groove composition may be critical to its recognition of cognate Watson-Crick base pairs, reminiscent of Tetrahymena group I intron folding (Battle & Doudna, 2002) and the decoding centre of the ribosome (which also tolerates wobble pairing at the analogous (5’) triplet position) (Ogle et al., 2001).

Systems-level properties of triplet pools

An important contribution to triplet fidelity also appears to arise from unexpected behaviours of the triplet substrates themselves. We observed that in some direct pair-mispair triplet contests, inclusion of their complementary triplets caused a striking (~3-fold) drop in misincorporation errors (Figure 9). A potential explanation may arise from differential formation of triplet:anti-triplet dimers in the reaction: for example, more extensive pppGCC:pppGGC (than pppACC:pppGGU) dimer formation would selectively reduce the effective concentration of free pppGCC vs. pppACC upon inclusion of their complementary pppGGC and pppGGU.

These pairwise reductions were recapitulated in the presence of random pppNNN substrate pools (Figure 9). Indeed, counterintuitively, raising pppNNN concentrations from 0.5 to 5 μM each almost halved the overall error rate (Figure 9 - figure supplement 1). Although diverse effects upon individual misincorporations were observed, this fidelity enhancement was driven by pronounced reductions in errors where the mismatched triplet has a high GC content compared to the cognate triplet, including common G-U wobble mispairs (Figure 9 - figure supplement 1). Dimer formation among pppNNN substrate pools would be expected to selectively buffer the free
concentrations of the more strongly-pairing GC-rich triplets, which could promote both fidelity and sequence generality through normalization of triplet availability against template (and complementary triplet) binding strength.

Indeed, more efficient, higher fidelity segment synthesis was observed when partially mimicking this outcome using an \( \text{pppNNN} \) pool formulated with a reduced G content (Figure 6 – figure supplement 4, Table 1). In a prebiotic scenario, substrate pool composition would have been determined by the abundance and nontemplated polymerization tendencies of the different nucleotides; large biases in these could skew triplet compositions or deplete a triplet (resulting in mismatch incorporation). However, the potential for replication to proceed in different triplet registers may provide a degree of resilience towards such biases.
Discussion

Here we describe the discovery and characterization of a ribozyme (t5⁺1) with a robust ability to polymerize RNA trinucleotide triphosphate (triplet) substrates. Unusually, this triplet polymerase ribozyme comprises a heterodimer of a catalytic triplet polymerase subunit (t5) and a non-catalytic RNA cofactor (type 1), which enhances triplet polymerase activity and abrogates the need for template tethering. Such a quaternary structure - involving a heterodimer of a full-length and a truncated subunit - is reminiscent of the processivity factors of proteinaceous polymerases such as the heterodimeric p66/p51 HIV reverse transcriptase holoenzyme (Huang et al., 1992). There are multiple examples of dimerization in RNA evolution - such as the VS ribozyme (Suslov et al., 2015), retroviral RNA genome dimerization (Paillart et al., 2004), in vitro evolved heterodimeric RNA liposome binders (Vlassov et al., 2001), and recently the homodimeric CORN fluorescent RNA aptamer (Warner et al., 2017). However, the spontaneous emergence of a general, mutualistic RNA cofactor has not previously been observed for ribozymes and may suggest an underappreciated dimension to the evolutionary dynamics of ribozyme pools under stringent adaptive pressures. Indeed, the extinction of previously dominant species in the selection that were unable to benefit from type 1 enhancement (e.g. type 0, see Figure 3 - figure supplement 2) and succession with cooperative RNA species (Vaidya et al., 2012) illustrates the potential for such symbioses to shape RNA molecular ecologies.

The t5⁺1 ribozyme’s principal current shortcoming is its low catalytic efficiency. In the optimal context for mononucleotide polymerase ribozymes, this triplet polymerase heterodimer yields ~4-fold more unligated junctions than the RPR tC19Z (Attwater et al., 2013b), which itself is 240-fold slower than the currently most advanced RPR 24-3 (Horning & Joyce, 2016). Yet despite this modest catalytic power, t5⁺1 displays much enhanced generality in RNA synthesis and now achieves both copying of previously intractable structured RNA templates, and templated
synthesis and assembly of an active ‘+’ strand copy of its catalytic domain, suggesting key contributions of the triplet substrates themselves.

Indeed, one of the main findings of our work are the compelling advantages triplet substrates appear to offer for sequence general RNA replication. For instance, when binding templates, triplets incur a lower entropic cost per position compared to canonical mononucleotides (thus aiding copying of sequences rich in weakly pairing A and U bases), with particularly helpful stability contributions from intra-triplet base stacking (Eigen, 1971). Furthermore, energetically favourable inter-triplet stacking interactions appear to instigate cooperative binding and unfolding of even highly stable RNA template structures (Figures 1b, 5a) upon reaching the required substrate concentration threshold. In our work, this process is aided by the cold temperature and solute concentration effect of eutectic ice phase formation (Attwater et al., 2010; Mutschler et al., 2015). Counterintuitively, a general solution to the copying of structured RNAs arises not from conditions that disfavour base-pairing (which would also hinder substrate binding), but rather from conditions that promote it.

Together these favourable molecular traits serve to pre-organize the template towards a double-stranded RNA duplex with triplet junctions poised for ligation. A triplet/template duplex presents a more ordered, regular target for sequence-general ribozyme docking (by e.g. the ε domain) than a single stranded template (variably prone to secondary structure formation or sequence-specific interactions with the ribozyme (Wochner et al., 2011)). Such general duplex interactions also underlie other notable features observed in our triplet-based RNA synthesis such as in trans template binding (Figure 3c) as well as the capacity for bidirectional (5’-3’ / 3’-5’) and primer-free RNA synthesis (Figure 7).

Contrary to expectations RNA-catalyzed triplet polymerisation can proceed with a fidelity matching or exceeding even the best mononucleotide RNA polymerase ribozymes (Attwater et al., 2013b; Horning & Joyce, 2016). t5⁺¹ ribozyme fidelity is due to both a readout of cognate
minor groove interactions by the ribozyme ε domain (Figure 8) and an unanticipated fidelity
boost arising from systems-level properties of triplet pools, that appear to normalize the
availability of free triplet (and potentially longer oligonucleotide) substrates against their base-
pairing strength (Figure 9). Though further work will be required to characterize triplet pool
properties, they likely involve formation of cognate or near-cognate triplet:anti-triplet interaction
networks, as formation of tRNA dimers via cognate anticodon:anticodon interactions has been
observed in a similar concentration range (Eisinger & Gross, 1975).

While phylogenetically unrelated, mechanistic analogies between the triplet polymerase
ribozyme and the ribosome are apparent. Both are RNA heterodimers that operate in a triplet
register along a single-stranded RNA template, whilst enforcing a minor-groove mediated pattern
of triplet / anticodon readout (including tolerance of 5’ wobble pairing), suggestive of convergent
adaptive solutions to the challenges of replication and decoding. It has long been speculated that
the decoding centre of the small ribosomal subunit might have had its origins in an ancestral RNA
replicase, but the implied triplet-based character of such a replicase was conspicuously discordant
with modern mononucleotide-based replication (Weiss & Cherry, 1993; Poole et al., 1998;
Noller, 2012). The utility of triplets as substrates for RNA synthesis and self-synthesis described
herein suggests that these early ideas deserve to be reconsidered. In the context of initial
uncorrelated evolution of the small and large ribosomal subunits (Petrov et al., 2015), an early
reliance upon triplets in RNA replication could have inadvertently supplied a decoding center for
translation.

In conclusion, the unexpected emergent properties of triplets – including cooperative
binding and unfolding of structured RNA templates, enhanced incorporation of AU-rich
substrates, and error attenuation (resulting from triplet pool interaction networks) – argue that
short RNA oligonucleotides represent predisposed substrates for RNA-catalyzed RNA
replication. Some of these benefits might also extend to anticodon behaviour in early translation,
and to the non-enzymatic replication of RNA (Szostak, 2012), where downstream trinucleotides have recently been shown to enhance incorporation of preceding activated mononucleotides both through stacking and positioning effects (Vogel et al., 2005; Zhang et al., 2018) and the formation of a highly reactive intermediate (Prywes et al., 2016; O'Flaherty et al., 2018). Taken together, the interaction of triplet substrate pools with RNA templates promotes uncoupling of an RNA’s sequence (i.e. information content, and associated folding tendencies) from its replicability, thereby enhancing RNA’s capacity to serve as an informational polymer.
Materials and Methods

Templated RNA-catalysed RNA synthesis

Standard ribozyme activity assays (modified where specified) comprise 5 pmol of each ribozyme annealed in 2.5 μl water (80°C 2 min, 17°C 10 min), with 2 μl of 1 M MgCl₂ and 0.5 μl of 1 M tris•HCl pH 8.3 (at 25°C, pH raised to 9.2 at −7°C) then added on ice, and left for >5 min to ensure folding. This was added to 5 pmol each of primer and template and 50 pmol of each triplet pre-annealed in 5 μl water, then frozen on dry ice (10 min) and incubated at −7°C in a R4 series TC120 refrigerated cooling bath (Grant) to allow eutectic phase formation and reaction.

Final pre-freezing concentrations of components are displayed throughout (in this example, yielding 0.5 μM ribozyme/primer/template, 5 μM each triplet, 200 mM MgCl₂, 50 mM tris•HCl pH 8.3). Supercooled reactions (Figure 1c) remained liquid by omitting the dry-ice freezing step, maintaining these concentrations. Ice crystal formation upon eutectic phase equilibration, however, concentrates all solutes ~ 4-5 fold (Attwater et al., 2010) to their final operational levels and cooling elevates tris-buffered pH to ~9.2.

Some substrate mixes (e.g. pppNNN) led to a higher final reaction volume, but eutectic phase equilibration restored standard operational concentrations, also applicable to the four-fold-diluted extensions with the fragmented ribozyme (Figure 6c). These used 2 pmol each ribozyme/fragment annealed in 3.25 μl 62 mM MgCl₂, 15 mM tris•HCl pH 8.3 (37°C 5 min, ramped to 4°C at 0.1°C/s, 4°C 10 min), with pre-annealed primer/template/substrate (0.5/0.5/5 pmol) added in 0.75 μl water. These reactions, and preparative syntheses (Figure 6 - figure supplement 1, Figure 7c), were supercooled at −7°C followed by ice crystal addition for quick freezing and optimal activity.

Figure 1c extensions were set up by adding buffer, then RNAs (preannealed together, 0.1 μM final concentrations) to triplets. RNAs for ε⁺ syntheses were chilled on ice instead of annealing, with ribozyme/MgCl₂/tris•HCl pH 8.3 mixed with the other RNAs at −7°C.
Oligonucleotide substrates were added equimolar to template binding sites in the primer/template/substrate anneal. NTPs, on the other hand, were added with the MgCl$_2$/tris•HCl pH 8.3 to the ribozyme polymerase.

**Extension product separation**

At the end of standard incubations, reactions were thawed and 2 μl aliquots added to stop buffer (1 μl 0.44 M EDTA (pH 7.4), with urea to a 6 M final concentration and a 10-20 fold molar excess over template of complementary competing oligonucleotide (see Supplementary file 3) to prevent long product/template reannealing). Samples were denatured (94˚C 5 min) and RNAs separated by 8 M urea 1× TBE denaturing PAGE.

To avoid using potentially confounding competing oligonucleotide when purifying extension products, reactions with a biotinylated primer or template (stopped as above) could be purified by bead capture using MyOne C1 (Invitrogen) streptavidin-coated paramagnetic microbeads (using 5 μg pre-washed beads per pmol biotinylated RNA) in 0.5×-0.8× bead buffer (BB: 200 mM NaCl, 10 mM tris•HCl pH 7.4 (at 25˚C), 1 mM EDTA, 0.1% Tween-20). After washing twice in BB to remove unbound components, beads were incubated (1 min) in 25 mM NaOH, 1 mM EDTA, 0.05% Tween-20 to denature the duplexes (Horning & Joyce, 2016). To recover biotinylated extension products (e.g. Figure 5b left panels, Figure 6 - figure supplement 1 αβ$^+$/γδε$^+$, Figure 7c left panel) the supernatant was discarded, and beads were washed first in BB with 200 mM tris•HCl pH 7.4, then in BB, then heated (94˚C 4 min) in 95% formamide, 10 mM EDTA to release primers for urea-PAGE. To recover extension products bound to biotinylated templates (e.g. Figure 6 - figure supplement 1 β$^+$, δ$^+$, ε$^+$, Figure 7c right panel) the supernatant was removed, neutralized with 500 mM tris•HCl pH 7.4, spin-concentrated using Ultracel 3K filters (Amicon), recovered and denatured in 6M urea/10 mM EDTA before urea-PAGE. β$^+$ synthesised in Figure 6 - figure supplement 1 was not spin-concentrated, leading to a lower
recovery yield; δε⁺ synthesis was denatured directly from the ligation reaction in 60% formamide
with excess EDTA.

For gel mobility shift assays (Figure 3b), ribozymes were mixed at 0.5 μM, pre-annealed
and buffer added on ice as for extension reactions, then mixed with 5× loading buffer (50%
glycerol, 250 mM tris•HCl pH 8.3, 125 mM MgCl₂) for separation by native PAGE (0.5× TB, 8%
59:1 acrylamide:bisacrylamide, 25 mM MgCl₂, run in a Hoefer SE600 Chroma (upper chamber:
0.5×TB 50 mM NaOAc, lower chamber: 0.5×TB 25 mM Mg(OAc)₂) kept at 4˚C in a circulator
bath for 6-8 h at 10 W), then SYBR Gold stained.

**Extension product detection, quantification and purification**

Fluorescent primer extension products were detected using the appropriate laser
wavelength on a Typhoon Trio scanner (GE); gel densitometry allowed quantification of RNA
synthesis efficiency. Gel contrasts in figures were linearly adjusted to optimize display of bands
of differing intensities.

The gel in Figure 5b (middle panel) was washed thrice (5 min) in water, incubated with
10 μM DFHBI-1T ligand in buffer for 20 min to fold full-length broccoli aptamer (as in (Filonov
et al., 2015)) and scanned. The ligand was then eluted in three 1× TBE washes (leaving negligible
background fluorescence), and stained in 1× TBE with SYBR Gold (1:10000), washed again, and
re-scanned to detect all RNA products (left panel); scans were aligned via an adjacent Cy5-
labelled primer extension lane (not shown).

Full-length product yields in the Figure 6 - figure supplement 1 plus-strand syntheses were
calculated by running samples of bead-eluted products (or raw reaction for δε⁺) alongside known
amounts of the positive controls indicated, followed by SYBR-Gold staining. To purify, bead-
eluted products were run similarly, and excised using UV shadowing. Products were then eluted
from the gel fragments in 10 mM tris•HCl pH 7.4, and Spin-X column filtrate (Costar)
precipitated in 75% ethanol with 1 μl 1% glycogen carrier (omitted for β+). Recovered full-length product yields were calculated similarly to reaction yields for αβ+/δε*/γδε*, or using A_{260} for β+, δ*, ε*.

**Fragment sequencing**

Purified ribozyme- and TGK-synthesized αβ*/γδε* fragments were sequenced by first ligating a 3’ adaptor (10 U/μl T4 RNA Ligase 2 truncated KQ in 1× RNA ligase buffer (NEB) with 15% PEG-8000 and 2 μM AdeHDVlig at 10°C overnight). These reactions were bound to MyOne C1 microbeads, washed with BB to remove unligated adaptor, and reverse transcribed (50°C 30 min) with 1 μM HDVrec primer using Superscript III (Invitrogen). Beads were washed again then PCR amplified (5 cycles with a 40°C annealing step, then 20 cycles with a 50°C annealing step) using GoTaq HotStart master mix (Promega) and 0.8 μM each of primers P3HDV, and P5Xα8 or P5Xγ7, applying tags for high-throughput sequencing (Illumina MiSeq or HiSeq) after PCR product agarose gel purification. β+ syntheses’ cDNAs were amplified with P3HDV and P5Xβ6.

**Fidelity assay**

To estimate RNA synthesis fidelity, ribozymes extended primers using pppNNN on templates encoding CCC-XXX-CCC, where XXX were 12 different triplet sequences evenly exploring base composition and distribution (see Supplementary file 3; for XXX = ACC, template encodes CCC-ACC-UCC to avoid a terminal run of Gs).

Each primer/template pair (0.45/0.525 pmol per reaction) was annealed in 4 mM MgCl₂, 1 mM tris•HCl pH 7.4 (80°C 2 min, ramped to 4°C at 0.1°C/s, then kept on ice). The 12 pairs were combined in 0.27M MgCl₂/67 mM tris•HCl pH 8.3 on ice to discourage primer-template assortment (of which sequencing later revealed negligible levels). 36 pmol of each triplet
(equivalent to 5 μM final concentration after considering eutectic phase equilibration effects upon this more dilute reaction) in pppNNN were added to a reaction vessel in 10.8 μl water, to which 5.4 μl of the primer/template/buffer mix was added followed by 7.2 pmol of ribozyme pre-annealed (80°C 2 min 17°C 10 min, ice >5 min) in 1.8 μl water (f.c. equivalent 1 μM, in excess over the 0.875 μM template to which some ribozymes could tether to enhance extension).

Reactions were frozen and incubated (7 d at −7°C) as described above.

Reactions were stopped with 3.6 μl 0.44 M EDTA and 10.5 pmol of each template competing oligonucleotide (migrating above product, with marker mutations to ensure exclusion), denatured with 6 M urea, and urea-PAGE separated. After alignment with a fluorescence scan of the gel, a region of the sample lane corresponding to primers extended by +4 to +14 nt was excised (encompassing 2-4 triplet additions), and extension products were eluted, precipitated in 77% ethanol with 1 μl 1% glycogen carrier, washed in 85% ethanol and resuspended in water.

These extension products were 3’ adaptor ligated as for fragment sequencing. Products were reverse transcribed (0.2× adaptor ligation reaction, 1 μM HDVrec primer in Superscript III reaction, 50°C 30 min) and then PCR-amplified (1/30th reverse transcription mix, 0.8 μM each or primers P3HDV and P5GGGX) for sequencing as above (yielding 2×10^5-4×10^6 sequences per ribozyme assay).

After processing and 3’ adaptor trimming, sequences corresponding to primer extended by CCC + 1-3 additional triplets were collated for analysis. Variations in upstream primer sequences (see Supplementary file 3) allowed the partner template to be identified for each sequenced product; the triplet incorporated after the first CCC was counted. Separately, 10 μl extensions by t5+1 of each primer/template alone with its encoded triplet and pppCCC (and pppUCC for the ACC pair) were combined for purification and sequencing as above, to allow isolation of the ribozyme-mediated errors resulting from inclusion of the other 62 (61 for ACC) triplets in the reaction (versus errors from sequencing, recombination etc.). The counts of cognate triplet (C) & each
error triplet \((E)\) in the positive control \((p)\) reduced error counts in the experimental samples \((x)\) to yield ribozyme-mediated error counts \((E_r)\) thusly: \(E_r = E_x - E_p \times (C_x/C_p)\) (not reducing \(E_x\) below 0, and reallocating all reductions to \(C_r\); \(\text{pppCCC}\) counts (and \(\text{pppUCC}\) for the ACC template) remained uncorrected).

For each template, counts were then collated at the 1\(^{\text{st}}/2\text{nd/3}^{\text{rd}}\) positions to yield base-specific mutation rates for each position (Figure 8 - figure supplement 2, Figure 8 - source data 1). Across the 12 triplets, A, C, G, and U were encoded at each position three times; linear averages were calculated to map the position’s error profile (Figure 8b) and geometric means of the four nucleobases yielded the position’s overall fidelity (Figure 8a, Figure 8 - figure supplement 1).

**Triphosphorylated triplet synthesis**

Triplets (and some other short oligonucleotides) were prepared from NTPs by T7 RNA polymerase run-off transcription of a 5’ single-stranded DNA overhang downstream of a DNA duplex T7 promoter sequence. In most cases the 5’ overhang encoded (was the reverse complement of) the desired oligonucleotide. These oligonucleotides were short enough to synthesise during the abortive initiation stage of transcription, attenuating sequence constraints on the first bases of the transcript. However, T7 RNA polymerase exhibited tendencies to skip the first (or even second) base (most severe for U>C>A>G before 2\(^{\text{nd}}\) position purines: encoding CGU yielded some \(\text{pppGU}\), encoding UAC yielded just \(\text{pppAC}\) or use oligonucleotides generated during transcription to re-initiate (e.g. encoding GAG yielded \(\text{pppGAGAG}\), encoding AAA yielded \(\text{pppA}_{6-9}\), encoding UCC yielded \(\text{pppCCC}\), encoding CGC yielded some \(\text{pppGCGC}\); this tendency was most severe when the oligonucleotide could be accommodated opposite the final template bases of the promoter).
These tendencies could be subverted by encoding additional first bases (usually without providing the corresponding NTP). This initiated the oligonucleotide at the second position where skipping tendencies were lower (e.g. encoding CUAG without CTP yielded \( \text{ppp} \text{UAG} \), encoding UUAC yielded some \( \text{ppp} \text{UAC} \)), and reduced recruitment as initiators of products with bases not complementary to the introduced first position template base (e.g. encoding CGAG without CTP yielded \( \text{ppp} \text{GAG} \), encoding CAA without CTP yielded \( \text{ppp} \text{AA} \) and \( \text{ppp} \text{AAA} \), encoding AUCC without ATP yielded \( \text{ppp} \text{UCC} \), encoding UCGC without UTP yielded \( \text{ppp} \text{CGC} \)).

Each 30 μl transcription reaction contained 72 nmol of each desired product base as an NTP (Roche) (e.g. for \( \text{ppp} \text{UCC} \), 72 nmol UTP, 144 nmol CTP) in 1× MegaShortScript kit buffer with 1.5 μl MegaShortScript T7 enzyme (ThermoFisher). Also present were 15 pmol of each DNA oligonucleotide forming the transcription duplex target (see Supplementary file 2). The reactions were incubated overnight at 37°C, stopped with 3 μl 0.44 M EDTA and 17 μl 10 M urea, and separated by electrophoresis (35 W, 4.5 h) on a 35×18×0.15 cm 30% 19:1 acrylamide:bis-acrylamide 3 M urea tris-borate gel. Products were identified through their relative migrations (reflecting overall composition, fastest to slowest: C>U≈A>G) by UV shadowing. Triplet bands were excised and eluted overnight in 10 mM tris•HCl pH 7.4, and filtrate (Spin-X) precipitated with 0.3 M sodium acetate pH 5.5 in 85% ethanol. Pellets were washed in 85% ethanol, resuspended in water, and UV absorbances measured with a Nanodrop ND-1000 spectrophotometer. Oligocalc (Kibbe, 2007) was used to calculate sequence-specific concentrations and yields. \( \text{ppp} \text{NNN} \) was generated by combination of equal amounts of each of the 64 triplet stocks in a lo-bind microcentrifuge tube (Eppendorf).

3′-deoxy triphosphorylated ‘terminator’ triplets were transcribed as above but using a 3′ deoxynucleoside 5′ triphosphate (Trilink biotechnologies) for the last position, migrating faster during PAGE than the equivalent all-RNA triplet. Triplets with 2-thiouridine residues were transcribed as for their corresponding U, replacing UTP with U\(^{2S}\)TP (Jena Bioscience);
incorporation and migration were similar between the two, and their concentrations were calculated from $A_{260\text{nm}}$ by comparison to the $A_{260\text{nm}}$ of mixtures of the component ribonucleotides with UTP vs. $U^{2S}\text{TP}$. Triplets with 2’-fluoro, 2’-deoxy positions could also be transcribed, with lower efficiency, by substituting the corresponding triphosphate (Trilink Biotechnologies). The biotinylated $\text{pppGAU}^{\beta'}\text{B}$ triplet used in $\gamma$ segment synthesis (Figure 7c) was transcribed as for $\text{pppGAU}$, replacing UTP with biotin-16-aminoallylxuridine-5’-triphosphate (Trilink Biotechnologies), quantified via by comparison to the $A_{290\text{nm}}$ of mixtures of the component ribonucleotides.

Longer triphosphorylated oligonucleotides used in ribozyme self-synthesis were generated similarly, but using ~200 ng of fully double stranded DNA as a template. Candidate product bands were purified and the desired oligonucleotide identified by ribozyme-catalysed in-frame incorporation and, for some, fragment sequencing.

RNA oligonucleotide/ribozyme preparation

Transcriptions were performed on ~ 15 ng/μl dsDNA using MegaShortScript enzyme and buffer (ThermoFisher) with 7.8 mM of each NTP, or, to yield a 5’ monophosphate on the product to avoid aberrant ligation (‘GMP transcription’), 10 mM GMP and 2 mM of each NTP.

dsDNA templates for some of these (in Supplementary file 3) were generated (‘fill-in’) using three cycles of mutual extension (GoTaq HotStart, Promega) between the associated DNA oligonucleotide and 5T7 (or, where indicated, HDVrt for defined 3’ terminus formation (Schurer et al., 2002)) followed by column purification (QiaQuick, Qiagen).

Some 5’ biotinylated RNAs were synthesized using the TGK polymerase (Cozens et al., 2012) (56 μg/ml, in 1× Thermopol buffer (NEB) supplemented with 3 mM MgCl$_2$) to extend 5’ biotinylated RNA primers (0.75 μM) on DNA templates (1 μM) using 2.5 mM of each NTP
Biotinylated products were bead-purified as above.

3’ biotinylation of RNAs was achieved in two stages: 3’ azidylation (at 2 μM with 25 U/μl yeast poly-A polymerase (ThermoFisher) and 0.5 mM 2’-azido-2’-deoxycytidine triphosphate (Trilink Biotechnologies) for 1 h at 37°C) with subsequent acidic phenol/chloroform extraction and 75% ethanol precipitation, then copper-catalysed biotin-(PEG)$_4$-alkyne (ThermoFisher) cycloaddition (Winz et al., 2012) with subsequent 75% ethanol precipitation followed by resuspension and buffer exchange in Ultracel 3K filters (Amicon) to remove residual biotin-alkyne.

**Selection library synthesis**

Round 1 libraries were synthesised by mutual extension of 4 nmol of oligonucleotides 1baN30 and 1GMPfo or 1GTPfo at 1 μM each in 1× isothermal amplification buffer (NEB) with 250 μM each dNTP, annealed (80°C 3 min, 65 °C 5 min) before addition of 0.4 U/μl Bst 2.0 (NEB) and 30 min 65°C incubation.

After purification, 375 μg of each DNA (~1.5×10$^{15}$ molecules) were transcribed in 5 ml transcription reactions (36 mM tris•HCl pH 7.9 (at 25°C), 1.8 mM spermidine, 9 mM DTT, 10.8 mM MgCl$_2$, 2 mM each NTP, 1% 10× MegaShortScript buffer, 2% 1:9 MegaShortScript:NEB T7 RNA polymerase, 37°C overnight). These were treated with DNase, acid phenol/chloroform extracted and 73% ethanol precipitated prior to urea-PAGE purification, elution, filtering (Spin-X) and re-precipitation, yielding the 1GTP Zcore selection construct (Supplementary file 1). 10 mM GMP was present in transcriptions of the 1GMP construct, and for future transcriptions of the GMP construct selection branch and rounds 8-18; round 19-21 and reselection libraries were transcribed without GMP. Most subsequent selection rounds were
transcribed in 1/10th scale transcriptions with 15 μg of DNA (≈6×10^{13} molecules) derived from amplification of recovered PCR products (see later).

For round 8, 700 pmol DNA was formed with Tri3CUUQ, amplifying round 7 merged output (50 pmol), round 7 merged output recombined by StEP (Zhao & Zha, 2006) (200 pmol), and 0core ribozyme with the starting 3’ 30N library domain added (450 pmol, but extinct at the end of selection). DNA encoding type 5s amplified with AACAt5s was used to generate reselection libraries by PCR amplification using primers TriGAA7GAAM and T5ba13N/T5ba20N/T5ba28N; 5 pmol of the three dsDNA products were transcribed to generate reselection constructs.

**In vitro evolution cycle**

An outline of the selection strategy is shown in Figure 1 - figure supplement 2, with detailed lists of selection oligonucleotides and extension parameters in Figure 1 - source data 1, Figure 2 - source data 1, Figure 4 - source data 1 and Supplementary file 3. Selection construct was first annealed with equimolar dual-5’ biotinylated primer in water (80˚C 2-4 min, 17˚C 10 min), then chilled extension buffer and triplets were added before freezing and −7˚C incubation.

At the end of incubation the reaction was thawed on ice. To link the primer 3’ hydroxyl to the 5’ monophosphate of GMP constructs, selection constructs were buffer-exchanged directly after thawing using a PD-10 column (GE) in a cold room, into 3 ml ligation mix (optimised to prevent ligation over gaps) (2 mM MgCl₂, 50 mM tris•HCl pH 7.4, 0.1 mM ATP, 1 mM DTT, 2 μM HO-GCG (Rounds 1-7) or 2 μM HO-CUG (Rounds 8-18) with 30 U/ml T4 RNA Ligase 2 (NEB)). After incubation at 4˚C for 1 h, these were stopped with 2.2 mM EDTA and acid phenol/chloroform treated.

Constructs were then precipitated with glycogen carrier and 0.3 M sodium acetate in isopropanol (55%) before resuspension and denaturation (94˚C 4 min, in 6M urea 10 mM EDTA
with a 3× excess of competing oligonucleotide against the primer). For the reselection rounds, constructs were then treated with polynucleotide kinase (NEB) before denaturation to resolve the HDV-derived 2’, 3’-cyclic phosphates and allow later adaptor ligation.

Constructs were urea-PAGE separated alongside FITC-labelled RNAs equivalent to successfully ligated constructs. The gel region in the construct lane alongside these was excised to exclude the bulk unreacted construct (judged by UV shadowing). Biotinylated (primer-linked) constructs were eluted overnight into BB with 100 μg MyOne C1 beads. After 30 μm filtering (Partec Celltrics) of the supernatant to remove gel fragments, the beads were washed in BB then denaturing buffer (8 M urea, 50 mM tris•HCl pH 7.4, 1 mM EDTA, 0.1% Tween-20, 10 μM competing oligonucleotide, 60°C 2 min) to confirm covalent linkage of construct to primer, before further BB washing and transfer (to a fresh microcentrifuge tube to minimize downstream contamination). At this stage in the reselection, 3’ adaptors were then ligated to bead-bound constructs as above for 2 h (with buffer/enzyme added after bead resuspension in other reaction components including 0.04% Tween-20), and beads BB washed and transferred again.

Bead-bound constructs were now reverse-transcribed using 1 μM RTri (or HDVRec for the reselection) by resuspension in a Superscript III reaction with added 0.02% Tween-20 (50°C 30 min). Beads were BB washed and the RNA-bound cDNA 3’ end blocked by incubation with terminal deoxynucleotidyl transferase (ThermoFisher) and 0.2 mM dideoxy-ATP (TriLink) with 0.02% added Tween-20 (37°C 30 min), and beads were BB washed and transferred again.

cDNAs were eluted (10 μl 0.1 M NaOH 0.1% Tween-20 20 min), neutralized and plus strands regenerated with 0.2 μM rescue oligonucleotide in an IsoAmp II universal tHDA kit (NEB) reaction (65°C 60 min) to read through the structured product region. Whilst at this temperature, reactions were stopped with 5 mM EDTA and 1 vol BB with 50-100 μg beads to bind the nascent biotinylated plus strands at room temperature. These beads were then BB
washed, NaOH washed again to discard cDNAs (and recover only correctly-primed plus strands), washed and transferred again.

Each 50 μg beads were then subjected to plus strand recovery PCR in a 100 μl GoTaq HotStart reaction with 0.5 μM each RTri (or HDVrec for reselection) and RecInt (rounds 1-5) / RecIntQ (rounds 6-9) / RecIntL (rounds 10-14, 19-21 & reselection) / RecIntQL (rounds 15-18). The product was agarose size-purified, A$_{260}$nm quantified and added to construct synthesis PCR in 3× molar amount of the anticipated recovered RNA (judged by test extensions) that yielded it.

This final PCR for construct transcription in the subsequent selection round used 1 μM of the indicated construct synthesis primer, plus 1 μM RTri (or HDVrt for the reselection), in GoTaq HotStart reactions or (where indicated in source data) the GeneMorph II kit for mutagenesis (Agilent).
References and Notes


**Acknowledgments:** We thank S. Thomsen, M. McKie and M. Sharrock for important discussions, and B.T. Porebski and S.L. James for assistance with sequencing analysis. This work was supported by a Homerton College junior research fellowship (J.A.), a Harvard-Cambridge summer fellowship (A.R.) and by the Medical Research Council (J.A., E.G., A.M., P.H.; programme no. MC_U105178804).

Conditions for selections are included as source data 1 for Figures 1, 2 and 4. Numerical data for Figures 1c and 3b are included in Figure 1 - source data 2 and Figure 3 - source data 1 respectively. Numerical data and calculations for Figure 8 and its supplements 1 and 2 are supplied as Figure 8 - source data 1. Numerical data and calculations for Figure 8 - figure supplement 3 are supplied as Figure 8 - source data 2. A more extensive selection of substrate concentration-dependent error rates is supplied in Figure 9 - source data 1. Sequences of ribozymes, triplet synthesis templates, and oligonucleotides used in this study are supplied in Supplementary files 1, 2 and 3 respectively.
Figure legends

Figure 1. Monomer polymerisation and triplet polymerisation.

(A) Scheme outlining initial derivation of a triplet polymerase activity from a mononucleotide polymerase ribozyme via directed evolution. Z RPR truncation effects are shown in Figure 1 - figure supplement 1, the selection cycle is outlined in Figure 1 - figure supplement 2, and the selection conditions of rounds 1-7 are listed in Figure 1 - source data 1. Below, modes of action and secondary structures of the mononucleotide polymerase ribozyme (Z RPR) and a triplet polymerase ribozyme (0core), both depicted surrounding primer (tan) / template (grey) duplexes with a mononucleoside triphosphate (NTP) or trinucleotide triphosphate (triplet) substrate present (red). Here the templates are hybridised to the ribozyme upstream of the primer binding site, flexibly tethered to enhance local concentration and activity (via L repeats of an AACA sequence, e.g. L = 5 in templates SR1-4 below). Z RPR residues comprising its catalytic core (Zcore) are black; mutations in 0core arising from directed evolution of Zcore are in teal.

(B) Primer extension by the Z RPR using monomers (1 mM NTPs) or by 0core using triplets (5 μM $\text{ppp}AUA$ & $\text{ppp}CGC$) on a series of 6-nucleotide repeat templates (SR1-4, examples below) with escalating secondary structure potential that quenches Z RPR activity beyond the shortest template SR1 ($-7^\circ C$ ice 17 d, 0.5 μM/RNA). Extension by the triplet polymerase ribozyme 0core can overcome these structure tendencies up to the longest template SR4.

(C) Triplet concentration dependence of extension using templates SR1 (grey circles) and SR3 (black triangles) by 0core ($\text{ppp}AUA$ & $\text{ppp}CGC$, 0.1 μM of primer A10, template and ribozyme, $-7^\circ C$ supercooled 15 d, ± s.d., n = 3); shown below is a model of cooperative triplet-mediated unfolding of template SR3 structure to explain the sigmoidal triplet concentration dependence (red curve) of extension upon it. Numerical values are supplied in Figure 1 - source data 2.

Figure 1 - source data 1. Selection conditions of rounds 1-7.

Figure 1 - source data 2. Triplet concentration-dependent extension values.

Figure 2. Emergence of cooperativity during in vitro evolution.

(A) Composition of the round 21 selection pool as a % of total pool sequences; selection conditions of rounds 8-21 are listed in Figure 2 - source data 1. Secondary structures of ribozyme
type 1-6 archetypes are shown in Figure 2 - figure supplement 1, and comparison of their activities to that of the polyclonal selection pool is shown in Figure 2 - figure supplement 2. (B) Primer extension with triplets by these emergent triplet polymerase ribozyme types 2-6 (‘Rz’, alongside the starting Zcore ribozyme with ‘cap+’ sequence from selection, see Figure 3a), alone or with added truncated type 1 (+1, see Figure 3a) which boosted their triplet polymerase activities (0.5 μM Rzs/A10 primer/SR3 template, 5 μM pppAUA & pppCGC, −7˚C ice 16 h).

**Figure 2 - source data 1.** Selection conditions of rounds 8-21.

**Figure 3. Heterodimer formation and behaviour.**

(A) Secondary structures of the most common type 1 and type 5 clones from the selection, with *in vitro*-selected 3’ domain and core mutations coloured orange (type 1) or blue (type 5). 5’ truncation of type 1 (faded, including its putative primer/template interacting region), yielding the minimal type 1 variant ‘1’, maintained its enhancement activity (see Figure 3 - figure supplement 1). The effects of transplanting the indicated 5’ hairpin ‘cap+’ element from type 5 to other ribozymes are shown in Figure 3 - figure supplement 2. The inset shows type 5 ‘cap+’ hairpin element alteration to ‘cap−’ (yielding type 5<sup>cap−</sup>).

(B) Gel mobility shift characteristic of complex formation resulting from mixing of type 5 ribozyme with type 1 RNA (equimolar, or with the indicated equivalents). Type 5<sup>cap−</sup> loses this shift and its susceptibility to type 1 activity enhancement (Figure 3 - figure supplement 1). Below, shifted band intensities with increasing type 1 addition are plotted (quantified relative to the indicated type 5<sup>+</sup> lane intensities, n = 4 ± s.d.), signifying 1:1 heterodimer formation; numerical values are supplied in Figure 3 - source data 1.

(C) Type 1 enhancement allows type 5 variants to synthesise long RNAs using triphosphorylated oligonucleotide (11 nt) or short triplet (3 nt) substrates (Sub, 3.6 μM or 5 μM each, substrate sequences in grey beside lanes; 0.4 μM primer A11/template I-8, 2 μM each Rz, −7˚C ice for 16 d). This activity is independent of template tethering ([Wochner et al., 2011](#)), as comparable synthesis is achieved by versions of type 5 whose 5’ regions allow or avoid hybridisation to the template (type 5<sup>cis</sup> or type 5<sup>trans</sup> respectively, schematic above, sequences in Supplementary file 1). The average extent of ligation at the end of the reaction amongst all junctions in a lane is shown beneath each lane.

**Figure 3 - source data 1.** Relative intensities of shifted bands when varying type 1 equivalents.
**Figure 4. A trans-acting heterodimeric triplet polymerase.**

Top, scheme outlining derivation of the final t5\(^{+1}\) triplet polymerase archetype from the type 1 and type 5 RNAs (shown in Figure 3a) by reselection using the conditions in Figure 4 - source data 1. Below, the secondary structure of this ribozyme heterodimer, 135 nt (1) & 153 nt (t5) long, is depicted operating in trans on a non-tethered primer/template duplex. Type 5 3’ domain bases that re-emerged after randomisation during reselection are coloured black in the t5 3’ domain. Ribozyme development is summarized in Figure 4 - figure supplement 1; all ribozyme sequences are listed in Supplementary file 1.

**Figure 4 - source data 1.** Selection conditions of rounds 1-5 of the reselection.

**Figure 5. Triplet-mediated structured and functional template copying.**

(A) Extension on three structured hairpin templates (4S, 6S, 8S) with increasing stability. Top, the mfold-predicted (Markham & Zuker, 2005) structure and T\(_M\) of the most stable 8S template; below, primer extensions on these templates by t5\(^{+1}\) triplet polymerase (with 5 µM each encoded triplet) or type 1-enhanced Z polymerase ribozyme (Z RPR\(^{\text{cap}^+{}^{+1}}\), with 1 mM each NTP) (2 µM ribozyme, 0.5 µM 4S, 6S or 8S template and primer A9, –7°C ice 25 d). The self-complementary region in each template is indicated between each pair of lanes (shaded by triplet), with the encoded triplet substrate sequences at the left (in grey, with 5’ template overhangs in brackets). Syntheses using different substrate compositions and concentrations are shown in Figure 5 - figure supplement 1 and Figure 5 - figure supplement 2. While all hairpin templates are robustly copied by t5\(^{+1}\), synthesis by the Z RPR is completely arrested by the 6S and 8S hairpins.

(B) Synthesis of the broccoli aptamer. The native secondary structure is shown above (Tan: bases from ‘+’ strand synthesis primer. Green: bases from triplets. Outlined green: primer binding site for the ‘–’ strand synthesis). Below left, t5\(^{+1}\)-catalysed synthesis of fluorescent broccoli aptamer, run alongside standard (Broc+, synthesized by in vitro transcription), and stained for RNA with SYBR Gold (magenta) or folded with DFHBI-1T ligand (green fluorescence) (2 µM t5\(^{+1}\), 1 µM BBrc10/TBrc, 5 µM each triplet (in grey), −7°C ice 22 d). Below right, ‘–’ strand synthesis on Broc+ standard (0.5 µM without ligand in ribozyme extension buffer, 0.5 µM FBrcb6 primer, 2 µM t5\(^{+1}\), 5 µM each triplet (in grey), −7°C ice 38 d). t5\(^{+1}\) is able to synthesize both full-length functional (fluorescent) Broccoli ‘+’ and encoding ‘–’ strands.
Figure 6. Ribozyme self-synthesis and assembly of its own catalytic domain.

(A) t5^{+1}-catalysed syntheses of the five catalytic domain ‘+’ and ‘−’ segments via triplet extension of primers (grey) in −7°C ice. Triplets are coloured by segment and shown alongside the lanes; longer oligonucleotide substrates (faded) were provided for α^{+} and ε^{+} syntheses to combat ribozyme-template pairing as shown in Figure 6 - figure supplement 2 and Figure 6 - figure supplement 3. The triplets were supplied at 5 μM (α^{+} to ε^{+}), 10 μM (β^{−} to ε^{−}), or 20 μM (α^{−}) each, with 0.5 μM primer/template (P/T; 1 μM for Fβ6/Tβ) and oligonucleotides equimolar to template sites. Use of substrates of more heterogenous compositions and lengths are shown in Figure 6 - figure supplement 4 and Figure 6 - figure supplement 5 respectively. Densitometry gave yields of full-length products (boxed, by % of total primer), and geometric mean of the final extents of ligation (78%) across all 70 junctions in this self-synthesis context. These segment sequences derive from t5^{b}, a t5 variant with a neutral signature mutation (Supplementary file 1).

(B) Secondary structure representation of a t5 catalytic domain (αβ^{+}/γδε^{+}, t5^{b} sequence), formed via non-covalent assembly of t5^{+1}-synthesised ‘+’ strand fragments in Figure 6 - figure supplement 1, coloured by segment and synthesis substrate as in (A).

(C) Activity of ribozyme-synthesised αβ^{−}/γδε^{+} (B), compared to protein-synthesised αβ^{+}/γδε^{+} and full-length t5^{b} equivalents. These were assayed for synthesis of a δ− strand segment on a ribozyme-synthesised δ^{−} template, with added in vitro transcribed type 1 (2 μM each Rz, 5 μM triplets, 0.5 μM P/T, −7°C 0.25× ice 10 d). The ribozyme-synthesized and assembled αβ^{+}/γδε^{+} ribozyme is as active as in vitro transcribed equivalents, and can efficiently utilize ribozyme-synthesized RNA (δ^{−}) as a template.

Figure 7. Triplet-initiated template sequence copying.

(A) Extension by t5^{+1} of fluorescein-labelled primers bound to either the 3’ (A10) or 5’ (pppba) ends of a template (T8GAA, 10 μM pppGAA, 0.5 μM/RNA, −7°C ice 69 h), demonstrating extension in either 5’-3’ (A10) or 3’-5’ (pppba) directions.

(B) Synthesis of β^{+} on Tβ template via t5^{+1}-catalysed polymerisation of the substrates indicated on the right (2 μM t5^{+1}, 5 μM each triplet, 0.5 μM template (lane 2: 0.5 μM hexanucleotide), −7°C ice 9 d). Extension products in lanes 2-5 were eluted from template, PAGE-separated and SYBR-
Gold stained alongside in vitro transcribed full-length segment control (lane 1). Lane 3 shows full-length synthesis of β+ segment from triplets alone.

(C) Triplet based replication of the γ segment. Top left, synthesis of γ+ from the indicated substrates (1 μM t5+1, 5 μM each triplet, 2 μM TyHP template (lane 2: 2 μM Bioγ7 primer), −7°C ice 7 d). Biotinylated extension products in lanes 2-4 were isolated from template, PAGE-separated and SYBR-Gold stained alongside in vitro transcribed staining marker (Mγ m1, lane 1). This indicated 10% yield (per template) of full-length synthesis of γ+ segment from triplets alone (the final band in lane 4), which was purified for use as a template in γ− segment synthesis (bottom right, 1 μM t5+1, 5 μM each triplet, 0.05 μM template, with 0.05% Tween-20, −7°C ice 27 d). Extension products in lanes 5 & 6 were eluted from template, PAGE-separated and SYBR-Gold stained alongside in vitro transcribed full-length segment control (Mγ − m1, lane 7). This indicated 6% yield (per template) of full-length synthesis of γ− segment from triplets alone (the final band in lane 6).

Figure 8. Fidelity of ribozyme-catalysed triplet polymerisation.

(A) To estimate its fidelity, each ribozyme was provided with an equimolar mix of all 64 triplet substrates (pppNNN at 5 μM each) for primer extension using templates containing twelve representative trinucleotide sequences (N′N′N′). Deep sequencing of extension products identified the triplets added opposite each template trinucleotide, yielding position-specific error tendencies; the overall fidelity was calculated as a geometric mean of positional errors at each triplet position (n and s.d. of this value shown for ribozymes assayed multiple times, see Figure 8 - source data 1 for analysis of collated errors). The triplet polymerases exhibit diverse fidelity profiles; fidelity profiles of other type 5 variants are shown in Figure 8 - figure supplement 1.

(B) Collation of error rates by base type and position for type 5 with (t5+1) and without (αβγδ+1) the ε fidelity domain. Positional and overall fidelities are calculated as geometric means (see Figure 8 - source data 1); individual positional fidelities are plotted in Figure 8 - figure supplement 2.

(C) Schematic summary of effects of triplet minor groove modification upon the fidelity phenotype. In the depicted trinucleotide RNA duplex segment, spheres represent minor groove groups potentially available for hydrogen bonding in a sequence-general manner. For three of these groups (highlighted in black), we assayed whether their modification in substrates (2F = 2′
fluoro, 2S = 2-thio) affected the fidelity domain’s mismatch discrimination capabilities (detailed in Figure 8 - figure supplement 3, with data and calculations in Figure 8 - source data 2). These groups are labelled with the fraction of fidelity phenotype retained ($\phi$) when discriminating between the indicated modified substrates. Colour reflects the impact of that group’s modification upon the fidelity phenotype, with red denoting a strong disruptive effect, and yellow weak or negligible effects.

**Figure 8 - source data 1.** Analysis of collated errors by ribozymes in the fidelity assay.

**Figure 8 - source data 2.** Calculation of residual fidelity phenotypes in Figure 8c.

**Figure 9. Substrate pool interactions improve triplet fidelity.**

Applying the fidelity assay (Figure 8a, using t5+1) to single templates with only an encoded triplet and a mismatching one as substrates (at 5 μM each), we observed that relative mispair incorporation was proportionally reduced (by 61% (left) and 73% (right)) upon introduction of complementary triplets. Using all 64 triplets ($\text{pppNNN}$) has an analogous effect upon these pair/mispair comparisons with fidelity progressively improved upon increasing overall $\text{pppNNN}$ concentrations, with example effects on other triplets and overall fidelity presented in Figure 9 - figure supplement 1, and comprehensive error rates and ratios in Figure 9 - source data 1.

**Figure 9 - source data 1.** Collated error rates and ratios at different substrate concentrations.
Figure supplement legends

Figure 1 - figure supplement 1. Templated ligase activity from a mononucleotide polymerase.

Extension of primer A11 on template HTI by the Z RPR (Wochner et al., 2011) and a 3’ truncation of it (Zcore, Figure 1a) at −7°C in ice for 8 d (0.5 μM of each RNA), adding as substrates either a long oligonucleotide (‘11 nt’, pppGCGAAGCGUGU at 0.5 μM), a triplet (‘3 nt’, pppGCG at 5 μM), or NTPs (‘1 nt’, at 1 mM each). Gel densitometry was used to calculate the percentage of primer extended (shown below each lane). Z RPR has only a very limited template ligation activity (slightly exceeding background nonenzymatic ligation (Rohatgi et al., 1996)). In contrast, 3’ truncation (which abolished mononucleotide polymerase activity) accommodated and enhanced templated ligation.

Figure 1 - figure supplement 2. Selection scheme for in vitro evolution of triplet polymerase activity.

Illustration of the in vitro evolution strategy to enrich iterative templated ligase activity from libraries of selection constructs. Centre, an example selection construct is shown comprising Zcore with a 3’ 30 nt random domain followed by a fixed hairpin-forming reverse transcription site (1). This construct is flexibly tethered at its 5’ in cis (Tagami et al., 2017; Attwater et al., 2010) (via L repeats (3-8) of a flexible AACA linker) to a template region (grey) bound to 5’-dual biotin modified primer (tan). This construct requires successive polymerisation of two pppUCG triplet substrates (red) in −7°C ice followed by self-ligation to its triphosphorylated 5’ terminus to attach itself to the biotinylated primer (2). Active triplet polymerases are then identified and size-purified by denaturing PAGE to deplete unreacted ribozymes and primer (3), then captured on streptavidin beads and subjected to a denaturing wash to ensure covalent linkage to primer. Remaining dual-purified bead-linked ribozymes were then reverse transcribed (4). cDNAs were eluted with NaOH and neutralised, before acting as templates for ‘+’ strand DNA synthesis from 5’ biotinylated ‘rescue’ oligonucleotides that only prime correctly on the complete synthesised sequence (5), ensuring that constructs that deviated from the specified templated triplet polymerization fail to prime on cDNA preventing their ‘+’ strand synthesis. ‘+’ strands were then captured on fresh streptavidin beads, stripped of cDNAs, reamplified and diversified to generate selection construct for a new round of selection (6). See Materials and Methods for protocol
details and Figure 1 - source data 1, Figure 2 - source data 1, Figure 4 - source data 1 and Supplementary file 3 for oligonucleotides used in selections.

**Figure 2 - figure supplement 1. Secondary structures of type 1-6 ribozymes.**

The archetypes shown are active truncations of the most common/enriched variants of each type. Types 2-6 are depicted hybridised to a flexibly tethered primer (tan) / template (grey) / substrate (red) duplex analogous to the selection construct (Figure 1 - figure supplement 2), supporting extension on the SR1-4 templates (Figure 2b). Coloured bases deviate from Zcore or derive from the random sequence region. Type 1, the most common type in the output, displayed a high degree of sporadic mutation in the catalytic core domain, consistent with a loss of catalytic function. Shown here are the four most common core mutations, which were each present in >45% of type 1 sequences.

**Figure 2 - figure supplement 2. Clonal versus polyclonal activity.**

Type 1 & 2 RNAs, and the indicated polyclonal ribozyme pools late in the selection, were transcribed as selection constructs generated with Tri8AUAM (then annealed to primer A10). Their abilities to incorporate eight $pppAUA$ triplets then self-ligate in this selection construct context were compared (left, 0.1 μM each RNA, 2 μM $pppAUA$, in −7°C ice for 5 d; right, 0.2 μM each RNA, 2 μM $pppAUA$ and $pppAUG^{3'd}$, in −7°C ice for 9 d). Left: Despite being the most active type (Figure 2b), type 2 activity fell far short of the round 19 pool polyclonal activity. Right: The incorporation of wobble pairing 3′-deoxy ‘terminator’ triplet $pppAUG^{3'd}$ (see results section on fidelity, yielding a faster-migrating band than incorporation of the cognate $pppAUA$ triplet) was decreased sharply after round 19, as calculated by densitometry and averaging amongst the ligation junctions in each lane, indicative of enrichment of higher fidelity triplet polymerases in the selection pool.

**Figure 3 - figure supplement 1. Parameters of type 1 activity enhancement.**

(A) PAGE of primer extension reactions comprising combinations of type 5 and type 1 variants (0.2 μM primer A10, template SR3, and type 5 variant, here annealed together and combined with
0.2 μM of separately-annealed type 1 variant after buffer addition; −7°C in ice for 63 h, 2 μM each pppAUA, pppAUG₃ᵈ, pppCGC). Full-length type 1 enhanced the activity of type 5 as did 5’ truncated type 1 (‘1’, Figure 3a) without its duplex-interacting A-minor motif (Shechner et al., 2009) and its ‘cap–’ sequence (originally present on some type 1 sequences in the selection). Replacing the type 5 ribozyme’s selected 5’ ‘cap+’ sequence with ‘cap–’ (type 5cap–) abolished its enhancement by type 1 RNA.

(B) Stoichiometry of type 1 enhancement of type 3 activity (0.5 μM primer A10, template SR3, 2 μM each pppAUA, pppGUA₃ᵈ, pppCGC in −7°C ice for 20 h); quantification of average extension per primer in each lane is consistent with the 1:1 stoichiometry implied by type 1’s ~50% abundance in the round 21 selection pool.

Figure 3 - figure supplement 2. Type 1 enhancement of parental ribozymes.

Left, transplanting the conserved 5’ ‘cap+’ sequence from types 2-6 allows the Zcore but not the 0core ribozyme triplet polymerase activity to be enhanced by type 1; shown are PAGE of extensions of primer A10 on template CCCMisAUG by the indicated ribozyme cores (5 μM of pppAUG and pppCCC, 0.5 μM of each RNA with type 1 annealed and added separately, −7°C in ice for 22 h). Right, RNA polymerase activity of Z RPR using NTPs is enhanced by type 1 addition, but only when modified with the 5’ ‘cap+’ sequence (Z RPRcap+), extending primer A10 on tethered template HTI (0.5 μM of each RNA, 4 mM of each NTP, at 4°C for 68 h).

Figure 4 - figure supplement 1. Summary of ribozyme development in this work.

Left, map of triplet polymerase selection and engineering. Ribozyme sequences are listed in Supplementary file 1. Right, secondary structures of the central ribozymes in this work.

Figure 5 - figure supplement 1. Substrates for structured template copying.

Primer extension reactions are shown using sets of defined triplet substrates (above the lanes), or random pools of all 64 triplets (pppNNN), each at the indicated concentrations on three templates with progressively longer, more stable secondary structure elements (right, Tₘs estimated using (Zuker, 2003)) (0.5 μM primer A9, 0.5 μM 4S, 6S, 8S templates, 2 μM t₅₊₁, −7°C ice for 25 d). Structure invasion and primer extension across the structured template region (beginning at the
dotted red line on the image) occurred in all cases (with increased efficiency at higher substrate concentrations for the 8S template (5µM)), including extensions where template secondary structures were allowed to form in the absence of triplets (*, triplets added last after 30 min supercooled at -7˚C).

**Figure 5 - figure supplement 2. Substrate concentration dependence of structured template copying.**

Extension on the 8S structured template with a defined set of complementary triplets (left) or dinucleotides (right), at the indicated substrate concentrations (0.5 µM primer A9, template 8S, 2 µM t5⁺, -7˚C ice for 25 d). Left, quantification of the extent of extension into the structured template region below each lane (measured by the fraction of third triplet addition to primers already extended by two) recapitulates a sigmoidal relationship with triplet concentration. Right, dinucleotides (like mononucleotides) exhibit negligible extension into the template hairpin structure.

**Figure 6 - figure supplement 1. Ribozyme catalytic domain self-synthesis and assembly.**

Top, conditions and yields for self-synthesis and assembly of the catalytic t5 domain (as αβ⁺ and γδε⁺ fragments, Figure 6b) by t5⁺. To obtain maximal amounts of fully synthesized and assembled t5⁺ ribozyme for activity testing (Figure 6c), we implemented t5 ‘+’ strand synthesis using sequence-specified triplet substrates for segment syntheses (as shown). The synthesis scale corresponds to the limiting component present. *

*: Unlike other segment ligation steps, the δε ligation was carried out using freeze-thaw cycling (Mutschler et al., 2015) (2.5 h at -30˚C, 21 h at -7˚C, then 0.5 h at 37˚C), which modestly improved yield.

Below, scheme of t5⁺-catalysed t5 synthesis and assembly reactions.

PAGE separations of syntheses (xt) alongside purified reference segments and fragments (in bold) were stained with SYBR Gold to quantify boxed full-length products. These were excised for use in subsequent assembly steps (illustrated by dashed black arrows). Synthetic schemes of colour-coded segments are shown beside corresponding lanes, denoting ‘+’ strand synthesis (bold primer and dashes) on ‘−’ strand template. Fully synthesized segments are shown as bold lines. Assembly reactions involve use of fully synthesized segments as the final substrate in a primer extension.
reaction ($\alpha + \beta, \gamma + \delta \epsilon$) or direct templated ligation of synthesized segments ($\delta + \epsilon$). The $\alpha \beta^+$ and $\gamma \delta \epsilon^+$ fragments associate spontaneously to form an active triplet polymerase ribozyme (Figure 6b), tested for activity in Figure 6c. The synthesised segment sequences were derived from the $t5^b$ variant of $t5$, which differs by one neutral signature mutation from $t5$ in the $\alpha$ segment, and by another neutral signature mutation from $t5^a$ in the $\epsilon$ segment (mutations highlighted in Supplementary file 1); these neutral mutations, not present in the $t5$ and $t5^a$ used to synthesise the segments (above), allowed verification of the synthetic origin of the product fragments by sequencing, which revealed the correct signature mutations in each fragment ruling out contamination of the synthesized ‘$+$’ strand products by the synthesizing (‘$-$’ strand) triplet polymerase ribozyme.

**Figure 6 - figure supplement 2. Substrate competition attenuates inhibitory $\epsilon^+ / \epsilon^-$ pairing during self-synthesis.**

(A) Longer oligonucleotide substrates are required to maximise synthesis of full-length $\epsilon^+$ segment (0.5 $\mu$M $Fe9$ primer/$Te$ template/oligonucleotide substrate and 5 $\mu$M each triplet (left, omitted where equivalent oligonucleotide substrate was present), 13 d in $-7^\circ C$ ice). Replacing two triplets with a preformed hexanucleotide substrate ($pppUGAAUG$) boosts full-length $\epsilon^+$ product synthesis by $t5^{+1}$ (left panel), but not by the type 6$^{+1}$ ribozyme - with a different accessory domain (see Figure 2 - figure supplement 1, sequence in Supplementary file 1) - which synthesizes $\epsilon^+$ independent of this hexanucleotide (right panel, allowing some full-length $\epsilon^+$ synthesis with only triplets). This likely reflects unfavourable competition between the $t5$ ribozyme’s own $\epsilon^+$ domain and triplet substrates for complementary pairing to the $\epsilon^-$ RNA used as template. Consistent with this hypothesis, increasing $t5^{+1}$ concentrations with this template is counterproductive (middle panel). For both ribozymes, using a preformed nonanucleotide substrate $pppUUUUUCAUG$ at the end of the template boosted full-length product over use of a $pppUUCAUG$ hexanucleotide + $pppUUU$ triplet, or just the three constituent triplets (S=9 vs. 6/3 vs. 3/3/3); this is not due to the individual triplet sequences as $t5^{+1}$ can efficiently incorporate $pppUUU$ and $pppUUC$ (see Figure 5b, Figure 6a).

(B) Doubling triplet concentrations to 10 $\mu$M can attenuate $t5^{+1}$’s requirement for the $pppUGAAUG$ hexanucleotide substrate during $\epsilon^+$ synthesis. This suggests that at higher triplet concentrations they successfully compete with this part of $t5$ for $\epsilon^-$ template hybridisation (0.5 $\mu$M $Fe9$ primer/$Te$ template/longer oligonucleotide substrates, 13 d in $-7^\circ C$ ice). With a fourfold
excess of primer/template duplex and substrates (right lane), the amount of full-length segment generated exceeds the ribozyme added, evidence of multiple turnover of full-length ε⁺ product. 

**Figure 6 - figure supplement 3. Ribozyme stabilisation attenuates inhibitory δ⁺/δ⁻ pairing during self-synthesis.**

Left, δ⁺ synthesis using 5 μM of each complementary triplet substrate (–hex) or 1 μM of pppUGACAU (+hex) replacing the final two triplets (0.25 μM primer Fδ7/template HTδ, 0.5 μM each ribozyme, 6 d in –7°C ice). Type 5 with the initially isolated ε domain (type 5⁺) needs hexamer substrate for δ⁺ synthesis, but with the reselected ε domain (t5⁻) does not. See Supplementary file 1 for ribozyme sequences.

Right, model of substrate/ribozyme competition for template binding. The region at the end of the δ segment exhibits irregular pairing in the initially isolated ε domain (above), and appears vulnerable to δ⁻ template pairing during δ synthesis with triplet substrates. Reselection strengthened base-pairing in the corresponding ε domain stem (below), replacing a G-U wobble with a cognate G-C pair and a U.U mispair with cognate U-A pair, stabilizing the δ⁻/ε domain junction and potentially reducing competition with triplets during δ segment synthesis.

**Figure 6 - figure supplement 4. Ribozyme segment synthesis with random substrate pools.**

For maximum self-synthesis yield, we had used specific triplet substrate sets. Here we compare synthesis by 0.5 μM t5⁺⁺ of the five t5 ‘+’ segments (30 d in –7°C ice) using specific triplets vs. random triplet pools or reduced G-content random triplet pools. Reactions included 5 μM of each triplet in specific triplet sets (‘tri’, as in Figure 6a ‘+’ syntheses), random pppNNN, or a low-G pppNNN. The low-G pppNNN had the same overall triplet concentration as pppNNN, but individual triplets were five-fold less common for each constituent G (~10/2/0.4/0.08 μM for 0/1/2/3 Gs per triplet; some primer/template (P/T) concentrations were reduced here to ensure excess substrate over template). Synthesis is compared with and without some longer oligonucleotide substrates (indicated below, equimolar to template sites, replacing the corresponding triplets in ‘tri’ substrate mixes). Importantly, for all ‘+’ strand segments, full-length products are generated using pppNNN with yields approaching those obtained when using specific triplets (calculated by densitometry, above the gel); intriguingly the low-G pppNNN pools often gave superior yields, possibly helped by the higher concentrations of weaker-binding AU-rich triplets therein.
Figure 6 - figure supplement 5. Ribozyme segment synthesis with mixed length substrate pools.

δ⁺ (left) or γ⁺ (right) syntheses were performed (using 0.5 μM each of t5⁺1 and the primer/templates above, 15 d in −7°C ice) with their constitutive triplet sets alone (*, see Figure 6a) or the random-sequence substrate mixes indicated. Introduction of dinucleotide and mononucleotide substrates (pppNN, pppN i.e. NTPs) decreases full-length product band intensity but not ligations performed (here quantified whilst assuming each extension product was formed from the fewest substrates possible), with increasing numbers of extension products deviating from the starting triplet register. Dinucleotides and mononucleotides appear poor substrates in the absence of triplets.

Figure 8 - figure supplement 1. Fidelity of type 5 variants.

Positional error rates and fidelities (determined as in Figure 8a) for type 5 ribozyme variants; the overall fidelity was calculated as a geometric mean of positional errors at each triplet position (n and s.d. of this value shown for ribozymes assayed multiple times, see Figure 8 - source data 1). Fidelity of the initial type 5 isolate (tethered to the assay templates) is modestly improved in the absence of type 1. However, with type 1, fidelity is improved when operating fully in trans (type5⁺1). Reselection and stabilisation of the ε domain yields a further fidelity improvement (t5⁺1) but ε truncation (αβγδ⁺1) reverts the pattern of error tendencies along the triplet towards that of the starting core. See Supplementary file 1 for ribozyme sequences.

Figure 8 - figure supplement 2. Fidelity domain influence upon template- and position-specific error rates.

At each position in each of the twelve triplet templates (top) tested in the fidelity assay (Figure 8a), three different errors are possible after analysis of sequence composition of added triplets (see Figure 8 - source data 1). Errors are represented by discs, coloured by mutation (e.g. A to G : green; A to C : blue; G to C : blue). Disc size is proportional to error frequency without the fidelity domain (its occurrence using αβγδ⁺1). The effect of the ε fidelity domain upon these error tendencies is plotted on the y-axis (by comparison of error rates between t5⁺1 and αβγδ⁺1). Many of the most significant errors (using αβγδ⁺1) exhibit the greatest proportional reductions.
with the fidelity domain, including substantial (>10-fold) reductions in several misincorporations at the third position. Improvements in fidelity are variable but focused on the 2\textsuperscript{nd} and 3\textsuperscript{rd} triplet positions, while the 1\textsuperscript{st} triplet position shows an increase specifically in A to G mutations, indicative of a tolerance for G-U wobble pairing.

**Figure 8 - figure supplement 3. Determination of residual fidelity phenotype when using minor groove-modified substrates.**

Top: framework for estimation of residual fidelity phenotype (\(\phi\)). Shown are model logistic curves for incorporation of a triplet e.g. GAU (at a fixed concentration, opposite 3’-AUC-5’) versus increasing concentrations of a mismatching triplet e.g. GGU. A ribozyme ‘+’ with a fidelity function (e.g. t5\(^a\), blue curve) will incorporate equal amounts of the matched and mismatched triplets at a different concentration of the mismatched triplet compared to a ribozyme ‘−’ without a fidelity function (e.g. \(\alpha\beta\gamma\delta\), red curve). \(x_+\) and \(x_-\) represent the lns of these concentrations for + and −; their separation (\(x_+ - x_-\)) is a proxy for the strength of the ‘+’ fidelity phenotype. We measured the relative incorporation (W) of triplet pairs by the two triplet polymerases at test concentrations by the two triplet polymerases at test concentrations (t) of mismatched triplet, chosen to maximise the difference in the ribozymes’ resulting fractional incorporations (marked by a blue dot vs. a red dot on respective curves). If a triplet modification (*, right) interferes with mismatch discrimination by ‘+’, it would shift the ‘+’ relative incorporation curve (blue) towards that of ‘−’ (red), reducing the difference in fractional incorporation. Assuming curve steepness (k) remains constant, the residual phenotype (\(\phi\)) for that modification is described by the new separation (\(x_+* - x_-*\)) as a proportion of the original separation (\(x_+ - x_-\)); numerical values and calculations from the measurements described below are supplied in Figure 8 - source data 2.

Middle: measurements of ratios of incorporation vs. misincorporation (W) for unmodified and modified triplet pairs (at the indicated test concentrations) by the t5\(^a+1\) triplet polymerase compared to the \(\varepsilon\) fidelity domain-truncated \(\alpha\beta\gamma\delta+1\) (at 0.5 \(\mu\)M each). The expected triplet additions are in grey along the left of each gel; the average W from n independent experiments, calculated via densitometry of products containing slower-migrating G misincorporations, is shown below each lane along with the average \(\phi\). Primers (P), templates (T) and additional triplets were at 0.5/0.5/5 \(\mu\)M, incubated for three days in \(-7^\circ\)C ice. Left: presence of a 2-thiouracil (2SU) at the 3\textsuperscript{rd} position of the triplet abolishes the fidelity domain’s ability to discriminate against a 2\textsuperscript{nd} position wobble pair (P: A10, T: CCCMisGAU, +5 \(\mu\)M pppCCC). Centre-left: the same
modification (2SU) also abolished the fidelity domain’s preference for G misincorporation at the 1st position (P: Fγ7, T: TγAGU, +5 μM HO-CUG). Centre-right: in contrast, a 2SU at the 1st position of the triplet exerted no clear influence upon 3rd position wobble discrimination (P: A10, T: CCCMisUGA, +5 μM pppCCC). Right: replacement of a 2’ hydroxyl group with a 2’ fluoro (2’F) at the 1st triplet position likewise had no effect upon 3rd position mismatch discrimination (P: Fγ7, T: TγAGU, +5 μM HO-CUG).

Below: Measurement of fidelity phenotype in the presence or absence of a downstream triplet. Residual phenotypes indicate attenuation or abolishment of 1st position (left panel, P Fγ7, T TγAGU), 2nd position (middle panel, P Fγ7, T TγGAU), and 3rd position (right panel, P Fγ7, T TγAGU) fidelity effects from downstream triplet absence. No effects are seen (upon 3rd position discrimination) from the presence or absence of a 5’ triphosphate on the downstream triplet (ϕppp).

In the absence of a triplet bound downstream on the template, the fidelity phenotype is severely compromised, suggesting this adjacent triplet:template duplex plays a critical role in positioning the fidelity domain relative to the incoming triplet.

**Figure 9 - figure supplement 1. Influence of random triplet pool concentrations upon triplet misincorporation tendencies.**

Left: 0.2 μM t5⁺⁺ was tested in the fidelity assay at a series of pppNNN concentrations. The observed overall error rate (black circles) in these reactions halved as pppNNN concentrations were increased to 7 μM each (after which extension activity began to reduce and error rates increase). Diverse patterns were observed in the concentration dependence of individual triplet positional errors, of which a selection are displayed here (×) coloured according to whether they raise (red-yellow) or lower (blue) the GC content of the expected triplet, and coloured more intensely for higher GC content triplets.

Right: For a broader set of errors with 0.5 μM t5⁺⁺ (those with >0.3% incidence at 5 μM pppNNN, x-axes) the changes in their incidences from raising concentrations of each pppNNN from 2 μM to 10 μM substrate are plotted on the y-axes. The frequencies of several mutations that increase triplet GC content in GC-rich triplets are reduced at 10 μM pppNNN concentrations (top panel). Conversely, the frequencies of some mutations that reduce GC content are increased at 10 μM pppNNN concentrations (bottom panel); these suggest relative depletion of available GC rich triplets as overall triplet concentrations rise. Such pool behaviour can explain the reduction of
overall error rates in ~5 µM $^{PPN}$NNN triplet pools, as misincorporations that increase GC content (e.g. G opposite U) are responsible for more errors than those that decrease it (bottom left).

Values and ratios for a wider range of errors are supplied in Figure 9 - source data 1.
Table 1. Sequencing of ribozyme-synthesised $\beta^+$ segment.

<table>
<thead>
<tr>
<th>Internal triplets'</th>
<th>10 µM each tri:</th>
<th>10 µM each PPPNNN</th>
<th>10 µM average, low-G PPPNNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>average (%)</td>
<td>93.3</td>
<td>92.8</td>
<td>97.3</td>
</tr>
<tr>
<td>$A$</td>
<td>99.4</td>
<td>97.0</td>
<td>98.2</td>
</tr>
<tr>
<td>$C$</td>
<td>99.5</td>
<td>98.8</td>
<td>99.5</td>
</tr>
<tr>
<td>$G$</td>
<td>97.8</td>
<td>99.2</td>
<td>97.3</td>
</tr>
<tr>
<td>$U$</td>
<td>99.0</td>
<td>99.5</td>
<td>99.0</td>
</tr>
<tr>
<td>$A$</td>
<td>99.6</td>
<td>99.3</td>
<td>99.4</td>
</tr>
<tr>
<td>$G$</td>
<td>99.1</td>
<td>98.6</td>
<td>97.8</td>
</tr>
<tr>
<td>$U$</td>
<td>99.1</td>
<td>99.4</td>
<td>98.0</td>
</tr>
<tr>
<td>$G$</td>
<td>95.5</td>
<td>99.1</td>
<td>99.3</td>
</tr>
<tr>
<td>$A$</td>
<td>99.5</td>
<td>98.6</td>
<td>99.0</td>
</tr>
<tr>
<td>$C$</td>
<td>99.1</td>
<td>99.1</td>
<td>99.3</td>
</tr>
<tr>
<td>$A$</td>
<td>99.8</td>
<td>98.7</td>
<td>98.9</td>
</tr>
<tr>
<td>$G$</td>
<td>98.9</td>
<td>94.3</td>
<td>98.8</td>
</tr>
<tr>
<td>$U$</td>
<td>98.9</td>
<td>81.4</td>
<td>98.8</td>
</tr>
<tr>
<td>$A$</td>
<td>99.5</td>
<td>96.6</td>
<td>98.8</td>
</tr>
<tr>
<td>$G$</td>
<td>99.4</td>
<td>97.4</td>
<td>97.6</td>
</tr>
<tr>
<td>$C$</td>
<td>99.3</td>
<td>99.4</td>
<td>97.3</td>
</tr>
<tr>
<td>$U$</td>
<td>97.8</td>
<td>59.8</td>
<td>97.4</td>
</tr>
<tr>
<td>$U$</td>
<td>98.8</td>
<td>42.3</td>
<td>97.4</td>
</tr>
<tr>
<td>$U$</td>
<td>96.3</td>
<td>96.5</td>
<td>93.2</td>
</tr>
</tbody>
</table>

Shown are the individual base fidelities (%) along the $\beta^+$ sequences (top) synthesised by t5$^{+1}$, using the six specific triplets (tri), or random (PPPNNN) or compositionally-biased random (low-G PPPNNN, see Figure 6 - figure supplement 4) substrate pools, from F$\beta$6 primer (the first six positions at the left) with template T$\beta$ (1 µM each RNA, 13 d –7°C ice). For their sequencing, extension products were eluted from templates, and full-length products were gel-purified, ligated to adaptor, reverse-transcribed and PCR amplified. For compositional analysis, a small percentage of unrelated amplified products were excluded (those with >9 mutations vs. the expected $\beta^+$ sequence; similar levels were excluded if a >6 mutation threshold was applied, 0.2%/0.2% & 3.7%/4.2% & 3.7%/3.8% for tri & PPPNNN & low-G PPPNNN). These sequences mostly appeared to derive from off-target priming and extension of F$\beta$6 on the ribozyme in the presence of PPPNNN.

The sequencing of products generated from specific triplets provides an estimate of background error arising from amplification and sequencing. The final triplet constitutes an error hot-spot - likely to mutate to a more mutationally stable triplet during self-replication - exacerbated in PPPNNN samples by the inability of the fidelity domain to operate in the absence of a downstream triplet (Figure 8 - figure supplement 3). The geometric average of internal triplet position fidelities is used to gauge overall t5$^{+1}$ fidelity during RNA synthesis. While overall fidelity drops from defined to random triplets (98.8% to 96.7%), much of this loss in fidelity can be recovered by adjusting the triplet composition to a low-G random pool, where reductions in G-U wobble pairing more than compensate for increases in rarer misincorporations opposite template C.

Supplementary file 1. Ribozyme sequences.

All sequences are written in a 5’-to-3’ direction, generated by GMP transcription of the corresponding PCR-generated dsDNA of the sequence downstream of 5T7 sequence duplex. All transcripts were PAGE-purified.
HDV ribozyme sequences (blue) were transcribed in series with reselected type 5 ribozymes and cleave themselves off during transcription (Schurer et al., 2002) to yield precise 3’ ends with 2’, 3’-cyclic phosphates; this presence of group did not affect type 5 activity.

Sequences corresponding to the 5’ ‘cap+’ or ‘cap−’ regions from the selection, presenting a target for type 1 interaction, are coloured light green, with alternative arbitrary inert 5’ hairpin-forming sequences in dark green. Single stranded sequences capable of hybridization with sites at the 3’ (or 5’, type 5\textsuperscript{cis}) ends of certain templates to endow flexible ribozyme-template duplex tethering (Wochner et al., 2011; Attwater et al., 2010) are coloured yellow. Accessory domains 3’ of the catalytic core are in bold.

Supplementary file 2. Transcription of triplets.
Each target was transcribed from oligonucleotides 5T7 (Supplementary file 3) and 5’-(var)-TATAGTGAGTCGTATTAATTTCGCGGGCGAGATCGATC’, where the (var) overhang encodes (is the DNA reverse complement of) the sequence indicated below. Guide yields: >15 nmol = ***, 10-15 nmol = **, 5-10 nmol = *, <5 nmol = ~. †: These transcriptions yield one main product suitable for use as markers to identify comigrating triplets with a similar G/AU/C content.

Supplementary file 3. Oligonucleotide sequences.
All sequences are written in a 5’-to-3’ direction. DNA sequences are coloured grey. RNA sequences are coloured black, with the exception of hammerhead ribozyme (brown) and HDV ribozyme (blue) sequences transcribed in series that cleave themselves off during transcription (Schurer et al., 2002) to yield precise 5’ and 3’ ends respectively. All RNAs were denaturing PAGE-purified, and DNAs were not, unless otherwise noted (’(non)-GP’). Competing oligonucleotides, listed by templates, were purified using RNeasy columns (Qiagen). Primer and oligonucleotide binding sites on templates are underlined.
**A**

- **Lane 1:** $\text{pppGCCAUCAAAGCUUGAGAGCAUCUU}$ alone ($\beta^+$ sequence)
- **Lane 2:**
  - $\text{pppGCCAUC ppAAP ppGGG UGA ppGAG ppCCA ppCUU}$
- **Lane 3:**
  - $\text{pppGCC ppAUC ppAAP ppGGG UGA ppGAG ppCCA ppCUU}$
- **Lane 4:**
  - $\text{pppAUC ppAAP ppGGG UGA ppGAG ppCCA ppCUU}$
- **Lane 5:** No Rz,
  - $\text{pppAUC ppAAP ppGGG UGA ppGAG ppCCA ppCUU}$

**B**

**β** synthesis

- **Lane 1:** $\text{pppGCCAUCAAAGCUUGAGAGCAUCUU}$ alone ($\beta^+$ sequence)

**C**

**γ** synthesis

- **Lane 1:**
  - $\text{pppGAUCGAGAGCGCGAGCCUUCGUGGCCC^p}$ alone ($\gamma^*$ sequence)
- **Lane 2:**
  - Bio-$\text{GGAUGCA ppGGG GGG GCA GGC UUC ppGGU ppGCG}$
- **Lane 3:**
  - Bio-$\text{GAPGAGAGCGCGAGCCUUCGUGGCCC}$
- **Lane 4:**
  - Bio-$\text{GAPGAGAGCGCGAGCCUUCGUGGCCC}$
- **Lane 5:**
  - Bio-$\text{GAPGAGAGCGCGAGCCUUCGUGGCCC}$
- **Lane 6:**
  - Bio-$\text{GAPGAGAGCGCGAGCCUUCGUGGCCC}$
- **Lane 7:**
  - Bio-$\text{GAPGAGAGCGCGAGCCUUCGUGGCCC}$

**Lane: 1 2 3 4 5 6 7**
### Average ribozyme error rates (%) per triplet position:

<table>
<thead>
<tr>
<th>Triplet position</th>
<th>Zcore</th>
<th>type 2'F</th>
<th>type 3'</th>
<th>type 4'</th>
<th>type 5'</th>
<th>type 6'</th>
<th>t5'</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>3.3</td>
<td>3.7</td>
<td>8.3</td>
<td>7.5</td>
<td>5.2</td>
<td>9.9</td>
<td>4.4</td>
</tr>
<tr>
<td>2nd</td>
<td>6.7</td>
<td>9.6</td>
<td>7.1</td>
<td>6.7</td>
<td>9.7</td>
<td>3.4</td>
<td>3.5</td>
</tr>
<tr>
<td>3rd</td>
<td>15.7</td>
<td>14.7</td>
<td>6.6</td>
<td>6.8</td>
<td>15.7</td>
<td>4.3</td>
<td>5.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Overall fidelity (%)</th>
<th>s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>91.3</td>
<td>±0.73(2)</td>
</tr>
<tr>
<td>90.5</td>
<td>±0.61(2)</td>
</tr>
<tr>
<td>92.7</td>
<td>±0.73(2)</td>
</tr>
<tr>
<td>93.0</td>
<td>±0.61(2)</td>
</tr>
<tr>
<td>89.7</td>
<td>±0.73(2)</td>
</tr>
<tr>
<td>95.3</td>
<td>±0.61(2)</td>
</tr>
<tr>
<td>93.8</td>
<td>±0.73(2)</td>
</tr>
<tr>
<td>97.4</td>
<td>±0.61(2)</td>
</tr>
</tbody>
</table>

#### Base encoded: A G C U

<table>
<thead>
<tr>
<th>1st position fidelity (%)</th>
<th>% incorporated</th>
</tr>
</thead>
<tbody>
<tr>
<td>92.5</td>
<td>0.1 0.1 0.2</td>
</tr>
<tr>
<td>96.6</td>
<td>0.7 0.1 0.4</td>
</tr>
<tr>
<td>90.2</td>
<td>0.0 0.0 0.9 99.1</td>
</tr>
<tr>
<td>93.2</td>
<td>0.3 0.2 0.5 98.6</td>
</tr>
<tr>
<td>99.6</td>
<td>0.0 0.2 3.2 98.0</td>
</tr>
<tr>
<td>97.5</td>
<td>2.2 0.2 0.0 98.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2nd position fidelity (%)</th>
<th>% incorporated</th>
</tr>
</thead>
<tbody>
<tr>
<td>98.5</td>
<td>0.3 0.3 0.1</td>
</tr>
<tr>
<td>93.3</td>
<td>0.2 0.5 98.1 0.3</td>
</tr>
<tr>
<td>96.2</td>
<td>0.2 0.5 98.0 0.6</td>
</tr>
<tr>
<td>98.0</td>
<td>0.2 0.2 0.0 99.1</td>
</tr>
<tr>
<td>96.8</td>
<td>0.0 0.2 1.0 99.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3rd position fidelity (%)</th>
<th>% incorporated</th>
</tr>
</thead>
<tbody>
<tr>
<td>93.7</td>
<td>2.2 0.2 0.0 98.0</td>
</tr>
<tr>
<td>98.7</td>
<td>0.0 0.2 0.0 99.1</td>
</tr>
<tr>
<td>99.7</td>
<td>0.0 0.2 0.0 99.1</td>
</tr>
<tr>
<td>99.1</td>
<td>2.2 0.2 0.0 98.0</td>
</tr>
</tbody>
</table>

### φ values:

- \( \phi = 1.0 \), \( A^{2F}GU \) vs. \( A^{2F}GG \)
- \( \phi = 0.8 \), \( U^{2S}GA \) vs. \( U^{2S}GG \)
- \( \phi = -0.2 \), \( GAU^{(2S)} \) vs. \( GGU^{(2S)} \)
- \( \phi = 0.0 \), \( AGU^{(2S)} \) vs. \( GGU^{(2S)} \)
<table>
<thead>
<tr>
<th>Incorporation of mispair vs. encoded triplet</th>
<th>Encoded:</th>
<th>Mispair:</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 μM $pppNNN$</td>
<td>$pppACC$</td>
<td>$pppGCC$</td>
</tr>
<tr>
<td>27%</td>
<td>$pppGAC$, $pppGCC$</td>
<td>31%</td>
</tr>
<tr>
<td>5 μM $pppNNN$</td>
<td>$pppACC$, $pppGCC$, $pppGGU$, $pppGGC$</td>
<td>12%</td>
</tr>
<tr>
<td>5 μM $pppNNN$</td>
<td>$pppGAC$, $pppGGC$, $pppGUC$, $pppGCC$</td>
<td>0.71%</td>
</tr>
<tr>
<td>10 μM $pppNNN$</td>
<td>$pppGCC$</td>
<td>8%</td>
</tr>
<tr>
<td>10 μM $pppNNN$</td>
<td>$pppGCC$</td>
<td>0.69%</td>
</tr>
<tr>
<td>10 μM $pppNNN$</td>
<td>$pppGCC$</td>
<td>0.69%</td>
</tr>
<tr>
<td>10 μM $pppNNN$</td>
<td>$pppGCC$</td>
<td>0.69%</td>
</tr>
</tbody>
</table>
Substrate (nt): 1 3 11 1 3 11 1 3
Ribozyme: none
Z RPR
Zcore

Extension (%): .01 .07 .07 35 .58 .22 .62 2.2 2.9
Pri
Polyclonal output: R19 59
R20 89
R21 90
no AUG 3'dp
AUA incorp. (%):
Full length self-ligation %: .07 .35 .30 .71
AUG 3'd
AUA

Pri
AUG 3'd

Rz: 1 type 2 R19 polyclonal

Polyclonal output: pppAUA incorp. (%): R19 59
R20 89
R21 90
no AUG 3'd
**Type 5**

- **Pri**
- **type 1**
- **+1**

**Type 5**

- **cap**
- **+1**

**Type 3** (μM): 0.25

**Type 1** (μM): 0, 0.25, 0.05, 0.5

**Triplets added per primer:** 0.005, 0.136, 0.044, 0.115
### Table: Substrates and Structured Template

<table>
<thead>
<tr>
<th>Substrates</th>
<th>4S</th>
<th>6S</th>
<th>8S</th>
</tr>
</thead>
<tbody>
<tr>
<td>(μM): 5*</td>
<td>5</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>5*</td>
<td>5</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

**Structured template:**

- **Template 4S**
  - 5′-CUGC<sub>5′</sub>CAAC<sub>3′</sub><sup>PPP</sup>GGU<sup>5′</sup>-<br>
  - T<sub>M</sub> ~ 93°C

- **Template 6S**
  - 5′-CUGC<sub>5′</sub>CAAC<sub>3′</sub><sup>PPP</sup>GGU<sup>5′</sup>-<br>
  - T<sub>M</sub> ~ 89°C

- **Template 8S**
  - 5′-CUGC<sub>5′</sub>CAAC<sub>3′</sub><sup>PPP</sup>GGU<sup>5′</sup>-<br>
  - T<sub>M</sub> ~ 82°C
Substrates:

$\text{PPP GGU PPP GUG PPP GAG}$

$\text{PPP UGU PPP GCG PPP GAC}$

$\text{PPP ACA PPP CUC PPP CUU}$

Substrates:

$\text{PPP GGU PPP UGU PPP GCG PPP GAC}$

$\text{PPP ACA PPP CUC}$

(length $\mu$M):

0 0.5 1 2 3 5 10

(extension $\geq 3$):

0 0.03 0.07 0.22 0.43 0.6 0.7

(extension $\geq 2$):

0 2 3 5 10 20 40 80
<table>
<thead>
<tr>
<th>Synthesis</th>
<th>Scale (pmol), Volume (μl)</th>
<th>Ribozyme</th>
<th>Primer</th>
<th>Template</th>
<th>Oligomers</th>
<th>Triplet</th>
<th>Incubation time (d)</th>
<th>Full-length product (%)</th>
<th>Recovered product (pmol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>β</td>
<td>2000, 1000</td>
<td>2 μM t5</td>
<td>pppGCCAUc</td>
<td>2 μM Tβ</td>
<td>none</td>
<td>10 μM pppAAA, pppGCU, pppUGA, pppGAG, pppCAU, pppCUU</td>
<td>14</td>
<td>31</td>
<td>183 β⁺</td>
</tr>
<tr>
<td>α+β</td>
<td>170, 340</td>
<td>2 μM t5</td>
<td>1 μM Bioα9</td>
<td>1 μM TαL</td>
<td>0.5 μM β⁺</td>
<td>10 μM pppCGA, pppUCU, pppAAA</td>
<td>24</td>
<td>14</td>
<td>19 αβ⁺</td>
</tr>
<tr>
<td>ε</td>
<td>3000, 6000</td>
<td>0.5 μM t5</td>
<td>pppGCAAAA</td>
<td>0.5 μM Tε</td>
<td>0.5 μM pppUGAAUG</td>
<td>5 μM pppCGC, pppGUG, pppCUU, pppCGU, pppGAG</td>
<td>22</td>
<td>21</td>
<td>521 ε⁺</td>
</tr>
<tr>
<td>δ</td>
<td>4000, 2000</td>
<td>2 μM t5</td>
<td>pppGCAUAG</td>
<td>2 μM Tδ</td>
<td>2 μM pppUGACAU</td>
<td>5 μM pppCGC, pppCAA, pppCGU, pppUCU, pppCCA</td>
<td>20</td>
<td>15</td>
<td>432 δ⁺</td>
</tr>
<tr>
<td>δε</td>
<td>400, 800</td>
<td>0.5 μM t5</td>
<td>pppGCAUAG</td>
<td>0.5 μM Tδ</td>
<td>0.5 μM pppGCAU</td>
<td>none</td>
<td>6*</td>
<td>36</td>
<td>115 δε⁺</td>
</tr>
<tr>
<td>γ+δε</td>
<td>115, 460</td>
<td>0.5 μM t5</td>
<td>pppGCAUAG</td>
<td>0.5 μM TγL</td>
<td>0.25 μM δε⁺</td>
<td>5 μM pppGAG, pppGCG, pppGCA, pppGCC, pppUUC, pppGGU, pppGGC</td>
<td>11</td>
<td>8</td>
<td>5.2 γδε⁺</td>
</tr>
</tbody>
</table>

**Diagram:**
- αβ⁺ xt
- β⁺ xt
- γδε⁺ xt
- δ⁺ xt
- ε⁺ xt
- γδε⁺
Ribozyme: 3' terminal substrates

<table>
<thead>
<tr>
<th>3' terminal substrates</th>
<th>0.5 µM t5⁺¹</th>
<th>2 µM t5⁺¹</th>
<th>2 µM type 6⁺⁺¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>pppAUG</td>
<td>3 6 9</td>
<td>3 6 9</td>
<td>3 6 9</td>
</tr>
<tr>
<td>pppUUC</td>
<td>3 6 9</td>
<td>3 6 9</td>
<td>3 6 9</td>
</tr>
<tr>
<td>pppUUU</td>
<td>3 6 9</td>
<td>3 6 9</td>
<td>3 6 9</td>
</tr>
<tr>
<td>pppGAG</td>
<td>3 6 9</td>
<td>3 6 9</td>
<td>3 6 9</td>
</tr>
<tr>
<td>pppAUG</td>
<td>3 6 9</td>
<td>3 6 9</td>
<td>3 6 9</td>
</tr>
<tr>
<td>pppUGA</td>
<td>3 6 9</td>
<td>3 6 9</td>
<td>3 6 9</td>
</tr>
<tr>
<td>pppCGU</td>
<td>3 6 9</td>
<td>3 6 9</td>
<td>3 6 9</td>
</tr>
<tr>
<td>pppCUU</td>
<td>3 6 9</td>
<td>3 6 9</td>
<td>3 6 9</td>
</tr>
<tr>
<td>pppGUG</td>
<td>3 6 9</td>
<td>3 6 9</td>
<td>3 6 9</td>
</tr>
</tbody>
</table>

[tri] (µM): 5 10 5 10 5

full-length product (%): 29 34 13 30 33
Substrate:

|            | NNN | NNN | NNN | NNN | NNN | NNN | NNN | NNN | NNN | NNN | NNN | NNN | NNN | NNN | NNN | NNN | NNN | NNN | NNN | NNN | NNN | NNN |
|------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|            | dG(c) | dG(c) | dG(c) | dG(c) | dG(c) | dG(c) | dG(c) | dG(c) | dG(c) | dG(c) | dG(c) | dG(c) | dG(c) | dG(c) | dG(c) | dG(c) | dG(c) | dG(c) | dG(c) | dG(c) | dG(c) | dG(c) |

Oligomers:

- **pppGACAAA**
- **pppCAUCUU**
- **pppUGACAU**
- **pppUGAAUG**
- **pppUUUUUCAUG**

P/T:

- Fe9/9HTαHP
- Fβ6/β
- Fγ7/γHP
- Fδ7/δ
- Fe6/ε

[P/T] (μM):

- .5 .5 .5 .35 .5 .5 .5 .35 1 1 .35 .5 .5 .5 .5 .5 .5 .5 .5 .5 .5 .5 .5 .5 .5 .5 .5 .5 .35 .5 .5 .5 .5 .35
<table>
<thead>
<tr>
<th>[each pppNNN] (μM):</th>
<th>*5</th>
<th>5</th>
<th>5</th>
<th>5</th>
<th>5</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>[each pppNN] (μM):</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>20</td>
<td>40</td>
<td>40</td>
<td>0</td>
</tr>
<tr>
<td>[each pppN] (μM):</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>80</td>
<td>320</td>
<td>320</td>
<td>0</td>
</tr>
</tbody>
</table>

- Average ligations per primer: 1.54, 0.98, 0.95, 0.91, 0.91, 0.03, 0
- Extension product triplet register (%): 100, 100, 86, 72, 58, 16, 0

- Average ligations per primer: 2.16, 2.81, 2.83, 2.86, 2.94, 1.02, 0
- Extension product triplet register (%): 100, 100, 96, 84, 70, 2, 0
## Average error rates (%) per triplet position:

<table>
<thead>
<tr>
<th>Rz:</th>
<th>type 5$^{+1}$</th>
<th>type 5</th>
<th>type 5$^{s+1}$</th>
<th>t5$^{+1}$</th>
<th>αβγδ$^{+1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>6.3</td>
<td>5.5</td>
<td>5.9</td>
<td>4.4</td>
<td>2.5</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>3.4</td>
<td>2.1</td>
<td>2.7</td>
<td>1.8</td>
<td>4.7</td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
<td>4.3</td>
<td>3.7</td>
<td>3.1</td>
<td>1.7</td>
<td>9.1</td>
</tr>
</tbody>
</table>

### Overall fidelities (%):

|          | 95.3           | 96.2    | 96.1           | 97.4       | 94.5        |

s.d.$^{(n)}$: ±0.73$^{(2)}$ ±1.1$^{(3)}$ ±0.61$^{(3)}$ ±0.31$^{(4)}$
Encoded triplet: $t_{5+1} = \alpha \beta \gamma = A \beta \gamma$

$\frac{\alpha \beta \gamma}{\alpha \beta \gamma}$

decreases error

e $\epsilon$ increases error

e $\epsilon$ increases error

e $\epsilon$ increases error

e $\epsilon$ increases error

$\frac{\alpha \beta \gamma}{\alpha \beta \gamma}$

decreases error

$\frac{\alpha \beta \gamma}{\alpha \beta \gamma}$

decreases error

e $\epsilon$ increases error

e $\epsilon$ increases error

e $\epsilon$ increases error

e $\epsilon$ increases error

e $\epsilon$ increases error
Unmodified triplets

![Graph showing Unmodified triplets]

Modified triplets (*)

![Graph showing Modified triplets (*)]

\[
\begin{align*}
\ln \left( \frac{W}{W^*} \right) &= k(x_+ - x_) \\
\ln \left( \frac{W^*}{W} \right) &= k(x_+^* - x_-^*) \\
\phi &= \frac{(x_+^* - x_-^*)}{(x_+ - x_-)} = \frac{\ln(W^*/W^*)}{\ln(W/W^*)}
\end{align*}
\]

\[\text{If } \phi < 1, \text{ the modification enhances the fidelity phenotype (for GAU vs. GGU)}\]

\[\phi = -0.2, \pm 0.23 \text{ (s.d.)} \]

\[\phi = 0.0, \pm 0.18 \text{ (s.d.)} \]

\[\phi = 0.8, \pm 0.52 \text{ (s.d.)} \]

\[\phi = 1.0, \pm 0.19 \text{ (s.d.)} \]

\[\phi = 0.5, \pm 0.18 \text{ (s.d.)} \]

\[\phi = 0.0, \pm 0.54 \text{ (s.d.)} \]

\[\phi = 0.3, \pm 0.04 \text{ (s.d.)} \]

\[\phi = 0.9, \pm 0.09 \text{ (s.d.)} \]
Mutation incidence with 5 μM \textsuperscript{ppp}NNN (%)

Mutation incidence with 10 μM \textsuperscript{ppp}NNN

Mutation incidence with 2 μM \textsuperscript{ppp}NNN

Average error contribution (%) of classes of substitutions by effect on GC content:

<table>
<thead>
<tr>
<th>pppNNN concentration (μM of each triplet)</th>
<th>Error occurrence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5 μM</td>
<td>2.78 0.28 0.59</td>
</tr>
<tr>
<td>5 μM</td>
<td>1.42 0.14 0.44</td>
</tr>
</tbody>
</table>

Mutation effect upon triplet GC content:

<table>
<thead>
<tr>
<th>Mutation incidence with 5 μM \textsuperscript{ppp}NNN (%)</th>
<th>Mutation incidence with 10 μM \textsuperscript{ppp}NNN</th>
<th>Mutation incidence with 2 μM \textsuperscript{ppp}NNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>3</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>30</td>
</tr>
<tr>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
</tbody>
</table>

Error occurrence (%) 10

Mutation effect upon triplet GC content:

<table>
<thead>
<tr>
<th>Mutation incidence with 5 μM \textsuperscript{ppp}NNN (%)</th>
<th>Mutation incidence with 10 μM \textsuperscript{ppp}NNN</th>
<th>Mutation incidence with 2 μM \textsuperscript{ppp}NNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>3</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>30</td>
</tr>
<tr>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
</tbody>
</table>

Mutation effect upon triplet GC content:

<table>
<thead>
<tr>
<th>Mutation incidence with 5 μM \textsuperscript{ppp}NNN (%)</th>
<th>Mutation incidence with 10 μM \textsuperscript{ppp}NNN</th>
<th>Mutation incidence with 2 μM \textsuperscript{ppp}NNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>3</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>30</td>
</tr>
<tr>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
</tbody>
</table>

Overall error rate 0.5 μM \textsuperscript{ppp}NNN 2.78 0.28 0.59

5 μM \textsuperscript{ppp}NNN