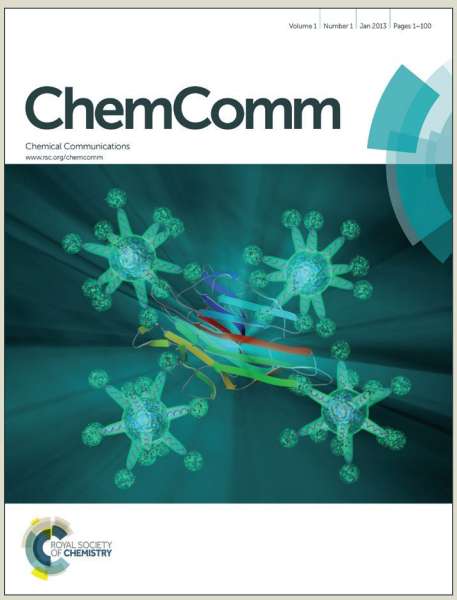
ChemComm

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemcomm

Journal Name

COMMUNICATION

Thiolated uridine substrates and templates improve the rate and fidelity of ribozyme-catalyzed RNA copying

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Noam Prywes,[a,b†] Yale S. Michaels,[a,c†] Ayan Pal,[a,d] Seung Soo Oh[a,d] and Jack W. Szostak[a,b,d]*

Ribozyme-catalyzed RNA polymerization is inefficient and error prone. Here we demonstrate that two alternative bases, 2-thio-uridine (s^2U) and 2-thio-ribo-thymidine (s^2T), improve the rate and fidelity of ribozyme catalyzed nucleotide addition as NTP substrates and as template bases. We also demonstrate the functionality of s^2U and s^2T-containing ribozymes.

The RNA world hypothesis posits that early life forms utilized RNA molecules, both as genetic polymers and as chemical catalysts.[1] Central to this theory is the emergence of an RNA enzyme (ribozyme) capable of catalyzing RNA polymerization.[2] The plausibility of such a ribozyme polymerase is supported by the existence of numerous and critical catalytic RNAs in contemporary organisms such as self-splicing introns,[3] the peptidyl transferase core of the ribosome[4] and a multitude of self-cleaving RNAs.[5]

Inspired by biological ribozymes, researchers have successfully evolved ribozymes in the laboratory capable of RNA-catalyzed RNA polymerization starting from pools of random oligonucleotides.[6,7] *In vitro* evolution and subsequent engineering have yielded, from such random pools, a class of ribozymes capable of polymerizing RNA in a template directed manner,[8] culminating recently with the development of the tC19Z ribozyme which is able to polymerize a strand of RNA longer than itself.[9] Despite this progress, even state of the art ribozyme polymerases achieve very low yields of full-length products on templates longer than 10 nucleotides and operate with poor fidelity, with UTP addition across from a G template being the most frequent error due to their propensity to form a stable G-U wobble base-pair.[9] Consequently, a ribozyme with the capacity for complete and efficient self-replication has yet to be isolated.

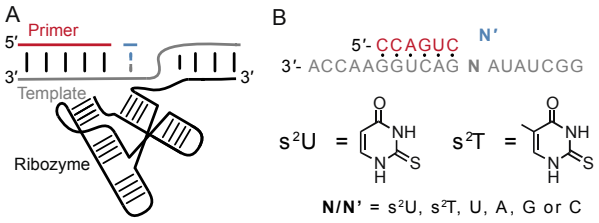


Fig. 1 b1-233t ribozyme schematic. (a) Schematic of the b1-233t ribozyme. The primer (red) is extended by a single nucleotide (blue) directed by a template (grey). The template is base paired to both the primer and the ribozyme (black). (b) Sequence of the 3′ portion of the primer (red) and the full sequence of the template (grey). N indicates the position in the template that is varied in this study. N′ indicates the NTP. Structures of s^2U and s^2T are shown.

One compelling method for potentially improving the fidelity and efficiency of ribozyme polymerases, and ultimately achieving ribozymes capable of self-replication, is to explore the use of non-canonical genetic polymers. Nucleic acids with alternative backbone linkages, sugar structures or nucleobases have been shown to fold into functional enzymes and aptamers with various advantages compared with their natural counterparts.[10,11,12,13] Ribozymes with backbones modified to include a mixture of 2′-5′ and 3′-5′ phosphate linkages have also been shown to be active.[14,15] 2′-5′ linkages are a natural byproduct of non-enzymatic polymerization under prebiotic conditions and are more amenable to duplex dissociation following replication, a necessity for subsequent copying of daughter strands. Aptamers evolved to contain one alternative nucleobase, 7-(2-thienyl)imidazo[4,5-b]pyridine, displayed tighter ligand binding than their wildtype counterparts.[11] In a

[a] *Howard Hughes Medical Institute, Department of Molecular Biology and Center for Computational and Integrative Biology, Massachusetts General Hospital, 185 Cambridge Street, Boston, Massachusetts 02114, USA.*
[b] *Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford St., Cambridge, Massachusetts 02138, USA.*
[c] *Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford OX3 9DS, UK*
[d] *Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA.*
*Email: szostak@molbio.mgh.harvard.edu
† These authors contributed equally to this work.
Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

separate study, when two alternative nucleobases, 2-amino-8-(1′-β-D-2-deoxyribofuranosyl)-imidazo[1,2-a]-1,3,5-triazin-4(8H)one and 6-amino-5-nitro-3-(1′-β-D-2′-deoxyribofuranosyl)-2(1H)-pyridone, were used, a tight binding (low nM affinity) aptamer emerged from an earlier round of *in vitro* evolution than is typical.[13]

As the high error rates observed in ribozyme-catalyzed RNA replication are largely a consequence of G-U wobble base pairing,[7,9] we chose to explore 2-thio-uridine (s^2U) and 2-thio-ribo-thymidine (s^2T) as potential alternatives to uridine in the context of increasing ribozyme efficiency and fidelity because these uridine analogues destabilize wobble pairing with guanine and form more stable pairs with adenine.[16,17,18,19,20] We have previously demonstrated the utility of s^2U and s^2T for improving fidelity and efficiency of non-enzymatic RNA polymerization.[21] Furthermore, thiolated uracil-analogues are abundant in biology; notably, s^2U in the tRNA anticodon loop plays a context-dependant role in prohibiting or tolerating wobble pairing.[22,23,24] s^2T has been found to increase the thermostability of tRNA in archaea[25] and thermophilic bacteria.[26] More recently, Attwater *et al.* showed that s^2UTP may serve as a more facile alternative to UTP as a substrate for ribozyme polymerization; however in the context of the highly evolved tC19Z ribozyme, incorporation of s^2U leads to a block in transcription.[10]

We addressed the question of ribozyme efficiency and fidelity in the simplified context of single-nucleotide addition. The b1-233t ribozyme catalyzes the addition of one to three ribonucleotides to an external RNA primer in a template directed fashion.[7] Prior characterisation of b1-233t revealed that addition of UTP across from a G (erroneous, wobble-pairing template) proceeds more efficiently than UTP addition across from an A (correct template).[7] We sought to improve the efficiency and fidelity of this ribozyme by using s^2U and s^2T, two U analogues with increased affinity for A and reduced affinity for G.[18,20,21] Following the procedure of Ekland and Bartel,[7] we combined the b1-233t ribozyme with a fluorescently labeled primer paired to a template with a single, variable nucleotide—the first unpaired template base 3′ of the primer-template duplex (Figure 1, supplemental table 1). UTP, s^2UTP or s^2TTP were added to a reaction mixture containing the b1-233t ribozyme (2.5 μM), primer (1 μM), template (2 μM) and buffer (30 mM Tris pH = 8, 60 mM $MgCl_2$, 200 mM KCl and 600 μM EDTA). We then quantified primer extension rates across a range of NTP concentrations for templates containing either an A or a G base at the coding site and determined the Michaelis-Menten parameters K_M and k_{cat}.

Consistent with previous reports, UTP is added at comparable rates across from an A-containing template (k_{cat} = 4.81x10^{-2} hr^{-1}, K_m = 7.34 mM, k_{cat}/K_M = 6.56 x 10^{-3} mM^{-1} hr^{-1}) or a G-containing template (k_{cat} = 5.68x10^{-2} hr^{-1}, K_M = 9.31 mM, k_{cat}/K_m = 6.10 x 10^{-3} mM^{-1} hr^{-1}) (Figure 2a, Figure S2, Supplemental Table 2). Interestingly, the thiolated UTP analogue s^2UTP exhibits more stable binding across from an A-containing template than UTP (K_M UTP = 7.34 mM, K_M s^2UTP = 2.84 mM, Figure S2, Supplemental Table 2). Furthermore s^2UTP is polymerized with diminished efficiency across from a G-containing template compared to UTP (k_{cat}/K_m UTP = 6.1 x10^{-3} mM^{-1} hr^{-1}, k_{cat}/K_m s^2UTP = 1.6x10^{-3} mM^{-1} hr^{-1}), confirming the hypothesis that substituting s^2UTP for UTP in single nucleotide primer extension reactions inhibits the formation of G-templated wobble base-pairs (Figure 2a). We also observed the same trend for s^2TTP: addition of s^2TTP across from an A-containing template increased k_{cat} and lowered K_M compared to UTP while addition of s^2TTP across from a G-containing template was impaired both in addition rate and binding strength compared to UTP (Supplemental Table 2). We found that both s^2UTP and s^2TTP exhibited at least a six-fold better selectivity for Watson-Crick template primer extension than UTP (Figure 2c). As NTP substrates, the thiolated uracil analogues that we tested each substantially improved both the efficiency and selectivity of ribozyme-catalyzed primer extension. Despite the utility of s^2UTP and s^2TTP as substrates for single-nucleotide addition, thiolated uracil analogues have been shown, in the context of the tC19Z polymerase ribozyme, to impair the copying of long templates.[10] We reason that these nucleobases may interfere with tertiary structural interactions between the growing primer strand and the ribozyme, which was evolved to use the canonical nucleobases. Further evolution with thiolated nucleobase substrates might overcome this hurdle.

In addition to assessing s^2UTP and s^2TTP as NTP substrates, we also evaluated s^2U and s^2T as template coding bases for ribozyme-catalyzed primer extension. Using identical reaction conditions as those described above, we determined kinetic parameters for b1-233t ribozyme-catalyzed ATP and GTP addition to a primer across from a template containing U, s^2U or s^2T in the coding position (Figure 2b, Supplemental table 2). When U was replaced by s^2U as the template coding base for ribozyme-catalyzed primer extension and ATP is used as a

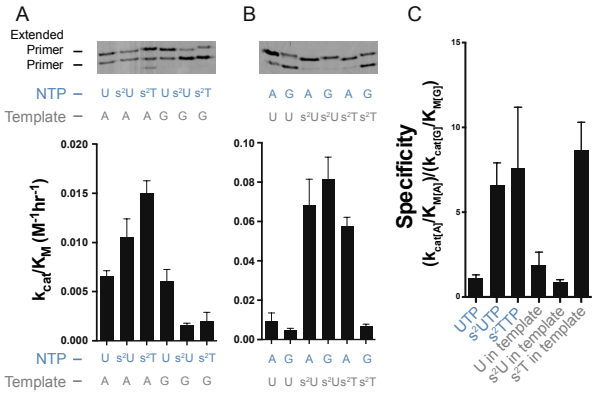


Fig. 2 Ribozyme kinetics and specificity. (a) Gel image showing the extension of the primer (bottom band) by the indicated nucleotide triphosphate (blue) when paired with the indicated template uridine analog in the template (grey) after 3 days. Reaction conditions were as specified by Bartel and colleagues (Ekland 1996). In the bar graph below, k_{cat}/K_M for each of the pairings is plotted. (b) Same as in (a) but with uridine analogs as NTPs and with A and G in the template. (c) The specificity of the reaction was calculated as the k_{cat}/K_M value for the U:A pair divided by that of the U:G pair.

substrate, the K_M decreased by 30% and k_{cat} increased five-fold (Figure S2, Supplemental Table 2). Though replacing U with s^2U in the template coding position increases the overall efficiency of ATP primer extension, it also serves as a more efficient template for GTP, diminishing the selectivity of addition by 55% as compared to U (Figure 2b–c). This is in accord with previous studies that have shown that s^2U can, under some circumstances, pair more tightly to G than U does.[22,27] Replacing U with s^2T in the template coding position, however, improves both the efficiency and selectivity of ribozyme-catalyzed primer extension. As a template for ATP addition, s^2T reduces the K_M two-fold and increases the k_{cat} threefold as compared to a U-containing template (Figure S2, Supplemental Table 2). As a template for GTP addition, a s^2T-containing template does not significantly affect the efficiency of primer extension compared to a U-containing template ($p = 0.33$, unpaired t test). Consequently, a s^2T-containing template improves both the efficiency and the selectivity of accurate primer extension by six-fold and five-fold, respectively, when compared to a canonical U-containing template (Figure 2b,c). Taken together, these data suggest that thiolated uracil analogues generally improve multiple parameters of ribozyme-catalyzed monomer addition both as templates and as substrates.

Given the efficiency of thiolated uracil analogues as templates and substrates of nucleotide addition, we asked whether ribozymes fully substituted with s^2U or s^2T were capable of catalyzing RNA primer extension (Figure 3a). Consequently, we sought to synthesize modified versions of the b1-233t ribozyme containing s^2U or s^2T in place of U. We first transcribed a 26-mer DNA (Supplemental table 1) oligonucleotide template containing 8 As with T7 RNA Polymerase (T7-RNAP) using a reaction mixture comprised of GTP, CTP, ATP and s^2UTP (75mM each) and assayed the resulting transcript by liquid chromatography mass spectrometry (LCMS) (Figure S3). The major product (m/z = 8989.8828) agreed with the predicted molecular weight for the s^2U-containing transcript (m/z = 8989.8827), confirming the previously established[28] capacity for T7-RNAP to polymerize RNA oligonucleotides using thiolated nucleotide triphosphates as substrates.

Next, we transcribed a DNA template for the b1-233t ribozyme *in vitro* using T7-RNAP, substituting s^2UTP or s^2TTP for UTP in the reaction mix. T7-RNAP was able to synthesize full-length transcripts with both s^2UTP and s^2TTP (Figure 3a). We term these new, s^2U or s^2T-containing versions of the ribozyme b1-233t-s^2U and b1-233t-s^2T, respectively. We assessed the capacity of these modified ribozymes to catalyze single-nucleotide addition reactions. Remarkably, despite analogue substitution at all uracil positions, both b1-233t-s^2U and b1-233t-s^2T were able to extend an RNA primer, albeit at much slower rates than the wildtype b1-233t ribozyme, with b1-233t-s^2T requiring three days to achieve appreciable primer extension at the detection limit of our assay (Figure 3b). We reasoned that the class 1 ligase ribozyme, the evolutionary precursor to b1-233t might also be amenable to substitution of s^2U and s^2T substitution. We generated two variants of the class 1 ligase by replacing UTP with s^2UTP or s^2TTP in the *in vitro* transcription reaction and assessed the capacity of each ribozyme to ligate itself to an external primer.[6] Again, both

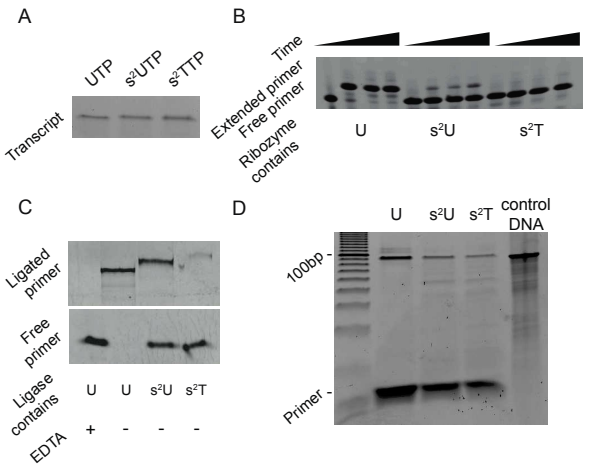


Fig. 3 Transcription, function and reverse transcription using modified rNTPs. (a) PAGE gel showing T7 transcription products after gel purification for transcription reactions containing the indicated triphosphate. (b) Gel showing ribozyme polymerase activity for WT and thiolated ribozyme varients. Time points are 0, 1, 2 and 3 days. (c) Gel showing class 1 ligase activity for wildtype, s^2U and s^2T containing varients. The reaction in lane one is a negative control run at an inhibitory EDTA concentration (200 mM). (d) Gel showing reverse transcription of the ribozymes from (b). These reactions were treated with RNaseH and RNaseA to remove residual RNA so that only cDNA would remain.

thiolated ligase variants were functionally active albeit at rates markedly lower than the wildtype class 1 ligase (Figure 3c). Though functional ribozymes have been evolved to contain non-canonical nucleotides,[29,30] to our knowledge, there is only one previous example[31] wherein a previously isolated ribozyme can be completely retrofitted with a new nucleobase and maintain functionality. In a diverse prebiotic chemical space, tolerance for chemical substitutions may have conferred a fitness advantage on catalytic RNAs.

In vitro evolution would likely yield more efficient thiolated-ribozyme polymerases. The ability to reverse-transcribe RNA containing s^2U and s^2T would be critical for such experiments.[32] We tested whether SuperScript™ reverse transcriptase could reverse transcribe b1-233t-s^2U and b1-233t-s^2T (Figure 3d, Supplemental table 1), and confirmed that RNA containing s^2U and s^2T can indeed be reverse-transcribed by conventional methods; thus, all of the tools required to evolve a thiolated ribozyme are readily accessible.

Here we have comprehensively assessed s^2U and s^2T as template bases and as NTP substrates for the b1-233t ribozyme.[7] We show that, in the context of this particular ribozyme, s^2U and s^2T have several advantages over U in terms of both the rate and the fidelity of primer extension. We have also demonstrated that both of these thiolated Uracil analogues are compatible substrates for T7 RNA polymerase, thus allowing us to synthesize several ribozymes containing s^2U or s^2T in place of U. Remarkably, s^2U or s^2T substitutions

are tolerated in the b1-233t ribozyme at all U residues simultaneously. The capacity for primer extension is maintained in both variants, albeit at substantially reduced efficiency. Not only do thiolated uracil analogues improve several dimensions of RNA-catalyzed nucleotide addition, ribozymes containing s^2U and s^2T are functionally active and can be generated by protein polymerases. These thiolated ribozyme variants can be reverse transcribed paving the way for future *in vitro* evolution of thio-nucleotide substituted ribozymes.

We thank Dr. Aaron Engelhart and Mr. Tony Z. Jia for their thoughtful advice in preparing this manuscript. We also thank Dr. Todd M. Lowe (UCSC) for useful discussions and Dr. Jonathan C. Blain and Dr. Victor Lelyveld for their technical assistance.

Notes and references

1 W. Gilbert, *Nature*, 1986, **319**.
2 M. P. Robertson and G. F. Joyce, *Cold Spring Harbor Perspectives in Biology*, 2012, **4**.
3 D. Herschlag and T. R. Cech, *Biochemistry*, 1990, **29**, 10159–10171.
4 H. F. Noller, V. Hoffarth and L. Zimniak, *Science*, 1992, **256**, 1416–1419.
5 A. Serganov and D. J. Patel, *Nat Rev Genet*, 2007, **8**, 776–790.
6 D. P. Bartel and J. W. Szostak, Science, 1993, **261**, 1411–1418.
7 E. H. Ekland and D. P. Bartel, *Nature*, 1996, **382**, 373–376.
8 W. K. Johnston, P. J. Unrau, M. S. Lawrence, M. E. Glasner and D. P. Bartel, *Science*, 2001, **292**, 1319–1325.
9 A. Wochner, J. Attwater, A. Coulson and P. Holliger, *Science*, 2011, **332**, 209–212.
10 J. Attwater, S. Tagami, M. Kimoto, K. Butler, E. T. Kool, J. Wengel, P. Herdewijn, I. Hirao and P. Holliger, *Chem. Sci.*, 2013, **4**, 2804.
11 M. Kimoto, R. Yamashige, K.-I. Matsunaga, S. Yokoyama and I. Hirao, *Nat Biotechnol*, 2013, 1–6.
12 V. B. Pinheiro, A. I. Taylor, C. Cozens, M. Abramov, M. Renders, S. Zhang, J. C. Chaput, J. Wengel, S.-Y. Peak-Chew, S. H. McLaughlin, P. Herdewijn and P. Holliger, *Science*, 2012, **336**, 341–344.
13 K. Sefah, Z. Yang, K. M. Bradley, S. Hoshika, E. Jiménez, L. Zhang, G. Zhu, S. Shanker, F. Yu, D. Turek, W. Tan and S. A. Benner, *Proc Natl Acad Sci USA*, 2013, 1–6.
14 A. E. Engelhart, M. W. Powner and J. W. Szostak, *Nature Chem*, 2013, **5**, 390–394.
15 C. Cozens, H. Mutschler, G. M. Nelson, G. Houlihan, A. I. Taylor and P. Holliger, *Angew Chem Int Ed Engl*, 2015, **54**, 15570–15573.
16 R. K. Kumar and D. R. Davis, *Nucleic Acids Res*, 1997, **25**, 1272–1280.
17 J. Sheng, A. Larsen, B. D. Heuberger, J. C. Blain and J. W. Szostak, *J Am Chem Soc*, 2014, **136**, 13916–13924.
18 N. A. Siegfried, R. Kierzek and P. C. Bevilacqua, *J Am Chem Soc*, 2010, **132**, 5342–5344.
19 H. O. Sintim and E. T. Kool, *J Am Chem Soc*, 2006, **128**, 396–397.
20 S. M. Testa, M. D. Disney, D. H. Turner and R. Kierzek, *Biochemistry*, 1999, **38**, 16655–16662.
21 B. D. Heuberger, A. Pal, F. Del Frate, V. V. Topkar and J. W. Szostak, *J Am Chem Soc*, 2015, **137**, 2769–2775.
22 A. Rodriguez-Hernandez, J. L. Spears, K. W. Gaston, P. A. Limbach, H. Gamper, Y.-M. Hou, R. Kaiser, P. F. Agris and J. J. Perona, *J Mol Biol*, 2013, **425**, 3888–3906.
23 R. Vormbrock, R. Morawietz and H. G. Gassen, *Biochim. Biophys. Acta*, 1974, **340**, 348–358.
24 M. Yoshida, K. Takeishi and T. Ukita, *Biochim. Biophys. Acta*, 1971, **228**, 153–166.
25 N. Horie, M. Hara-Yokoyama, S. Yokoyama, K. Watanabe, Y. Kuchino, S. Nishimura and T. Miyazawa, *Biochemistry*, 1985, **24**, 5711–5715.
26 K. Watanabe, T. Oshima, M. Saneyoshi and S. Nishimura, *FEBS Lett.*, 1974, **43**, 59–63.
27 M. J. O. Johansson, A. Esberg, B. Huang, G. R. Björk and A. S. Byström, *Mol Cell Biol*, 2008, **28**, 3301–3312.
28 C. Nakayama and M. Saneyoshi, *Journal of Biochemistry*, 1984, **96**, 1501–1509.
29 M. Hollenstein, C. J. Hipolito, C. H. Lam and D. M. Perrin, *Chembiochem*, 2009, **10**, 1988–1992.
30 T. M. Tarasow, S. L. Tarasow and B. E. Eaton, *Nature*, 1997, **389**, 54–57.
31 H. Sun, S. Jiang, J. Caton-Williams, H. Liu and Z. Huang, *RNA*, 2013, **19**, 1309–1314.
32 A. D. Ellington and J. W. Szostak, *Nature*, 1990, **346**, 818–822.