

# Structural interpretation of the effects of threo-nucleotides on nonenzymatic template-directed polymerization

Wen Zhang<sup>1,2,3</sup>, Seohyun Chris Kim<sup>1,4</sup>, Chun Pong Tam<sup>1,4</sup>, Victor S. Lelyveld<sup>1,2</sup>, Saikat Bala<sup>5</sup>, John C. Chaput<sup>5</sup> and Jack W. Szostak<sup>1,2,4,\*</sup>

<sup>1</sup>Howard Hughes Medical Institute and Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, MA 02114, USA, <sup>2</sup>Department of Genetics, Harvard Medical School, Boston, MA 02114, USA, <sup>3</sup>Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202, USA, <sup>4</sup>Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA and <sup>5</sup>Department of Chemistry and of Pharmaceutical Sciences, University of California, Irvine, CA 92697, USA

Received October 28, 2020; Revised November 25, 2020; Editorial Decision November 27, 2020; Accepted December 17, 2020

## ABSTRACT

The prebiotic synthesis of ribonucleotides is likely to have been accompanied by the synthesis of non-canonical nucleotides including the threo-nucleotide building blocks of TNA. Here, we examine the ability of activated threo-nucleotides to participate in nonenzymatic template-directed polymerization. We find that primer extension by multiple sequential threo-nucleotide monomers is strongly disfavored relative to ribo-nucleotides. Kinetic, NMR and crystallographic studies suggest that this is due in part to the slow formation of the imidazolium-bridged TNA dinucleotide intermediate in primer extension, and in part because of the greater distance between the attacking RNA primer 3'-hydroxyl and the phosphate of the incoming threo-nucleotide intermediate. Even a single activated threo-nucleotide in the presence of an activated downstream RNA oligonucleotide is added to the primer 10-fold more slowly than an activated ribonucleotide. In contrast, a single activated threo-nucleotide at the end of an RNA primer or in an RNA template results in only a modest decrease in the rate of primer extension, consistent with the minor and local structural distortions revealed by crystal structures. Our results are consistent with a model in which heterogeneous primordial oligonucleotides would, through cycles of replication, have given rise to increasingly homogeneous RNA strands.

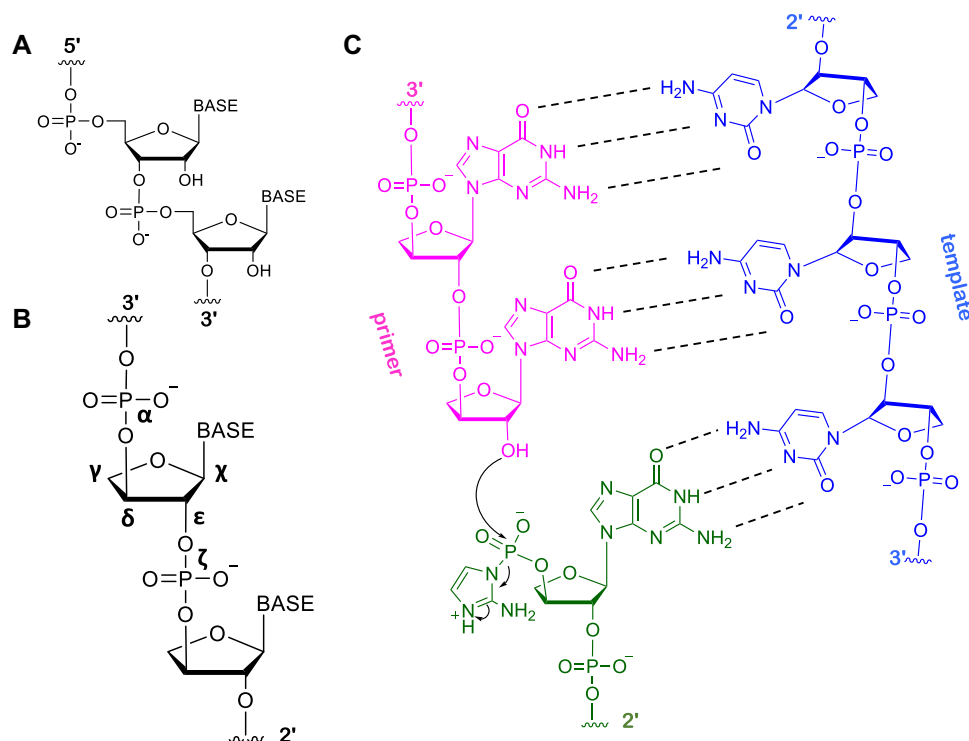
## INTRODUCTION

Studies of the path from the prebiotic world to the RNA World have long contended with the chemical complexity of RNA. Eschenmoser proposed a comprehensive synthetic approach, suggesting that the origin of RNA might be elucidated by a systematic exploration of the properties of its chemical analogues and structural relatives (1,2). Such a constructive approach addresses several seemingly intractable questions in abiogenesis. In particular, are ribonucleic acids chemically unique in their informational and functional properties? Is it possible that a chemically simpler genetic polymer might have arisen prior to RNA? Or, could RNA have been the winner in a competition with alternative polymers? It was quickly recognized that, in order to have been a true ancestor of RNA, any such a polymer would require the ability to transmit its information to RNA (3). This broad pursuit might be further constrained by recognizing that, to be a viable genetic polymer preceding or competing with RNA, such a species must also be capable of being replicated in a template-directed manner by a nonenzymatic process.

Eschenmoser and colleagues proposed and synthesized a promising candidate in this search: threose nucleic acid (TNA). The TNA backbone is composed of phosphodiester-linked 3'→2' α-L-threofuranose sugars and, therefore, has one fewer atom in its repeating unit than RNA (4). The possible prebiotic relevance of TNA has been bolstered by the observation that threose is one of the higher yield products of Kiliani-Fischer sugar synthesis (5). Threose is chemically simpler than ribose not only because of its shorter carbon chain length but also its reduced stereochemical complexity (Figure 1A, B). Prebiotically plausi-

\*To whom correspondence should be addressed. Tel: +1 617 726 5102; Fax: +1 617 643 3328; Email: szostak@molbio.mgh.harvard.edu  
Present addresses:

Chun Pong Tam, Beam Therapeutics, 26 Landsdowne St., Cambridge, MA 02139, USA.  
Saikat Bala, Polypeptide Laboratories Inc., Torrance, CA 90503, USA.



**Figure 1.** Comparison of RNA and threose nucleic acid (TNA) structures. (A) Native RNA chemical structure. (B) TNA chemical structure. Backbone and glycosidic torsion angles are labelled. (C) Chemical representation of hypothetical TNA template (blue) directed ligation of a TNA primer (pink) and a TNA ligator (green), of which the 3'-phosphate is activated as a 2-aminoimidazolidine.

ble pathways for producing threo-nucleotides have recently been proposed (6). TNA synthesis is simpler than RNA synthesis, because the threose sugar component of TNA can be generated from two molecules of glycolaldehyde, whereas the ribose component of RNA requires the sequential reaction of glycolaldehyde followed by glyceraldehyde. Critically, TNA can form stable Watson–Crick duplexes with itself, DNA and RNA, such that the exchange of genetic information between these polymers appears feasible (Figure 1C) (4). Structural studies demonstrate that TNA binds RNA to form right-handed double-helical duplexes (7–9) despite the fact that the contracted chemical structure of the TNA backbone decreases the distance between sequential phosphorus atoms.

To address the question of the replicative capacity of TNA, we have previously reported the enzymatic copying of DNA to TNA and vice versa, using both wild type and engineered polymerases (10–12). However, it remains unclear whether TNA can support efficient template-dependent copying by nonenzymatic chemistry. Heuberger and Switzer have demonstrated nonenzymatic synthesis of guanosine oligo-ribonucleotides on a TNA template (13). We subsequently reported nonenzymatic template-directed primer extension using 2-methylimidazole-activated 2'-amino-threose nucleotides as substrates, using DNA, RNA and TNA templates (14). In both cases, however, the TNA-templated nonenzymatic copying reactions were much slower than the corresponding RNA-templated reactions, irrespective of the precise identity of the activated mononucleotides used in primer extension reactions. More recently

the Krishnamurthy group showed that oligomers of amino-terminated TNA or RNA oligonucleotides could be ligated together with similar efficiency on all-RNA templates, but that ligation on chimeric TNA-RNA templates was much less efficient (15). The consistent observation of slow kinetics for TNA-directed nonenzymatic copying provides a biophysical argument for the emergence of RNA as opposed to TNA based life.

Here, we examine the behaviour of TNA in nonenzymatic template-directed primer extension reactions from both a kinetic and a structural perspective. We find that activated TNA monomers alone are not effective substrates for multi-step nonenzymatic primer extension reactions, and even single activated threo-nucleotides in an otherwise all-RNA context are incorporated with poor efficiency. Through a series of kinetic and co-crystallization studies, we analyse the origins of these experimental observations. Our results suggest that the reactive intermediate in primer extension, an imidazolium-bridged dinucleotide composed of 3'-3' linked threo-nucleotides, is both generated very slowly and once formed, binds to the template with suboptimal geometry for primer 3'-OH attack. In contrast we find that a single TNA nucleotide at the primer terminus or at the +1 template site only modestly slows the rate of primer extension, consistent with crystallographic observations that such substitutions have relatively minor structural consequences. Taken together, we propose that structural constraints explain the diminished ability of TNA to engage in nonenzymatic template-directed primer extension, and we discuss the consequences of these observations for the plau-

sibility of the involvement of TNA in the origin of primitive cells.

## MATERIALS AND METHODS

### RNA and TNA oligonucleotides preparation

Native, LNA-containing, and TNA-containing RNA oligonucleotides used for primer extension and crystallographic studies were custom-synthesized by Exiqon Inc. (Woburn, MA) or IDT Inc. (San Jose, CA), or were synthesized in-house on an Expedite 8900 DNA/RNA synthesizer. TNA phosphoramidites were synthesized as previously reported (16,17). Oligonucleotides synthesized in-house were deprotected using AMA (1:1 v/v aqueous mixture of 30% w/v ammonium hydroxide and 40% w/v methylamine) for 20 min at 65°C, followed by desilylation with Et<sub>3</sub>N•3HF. Oligonucleotides were HPLC purified on an Agilent ZORBAX Eclipse-XDB C18 column using 25 mM triethylammonium bicarbonate in H<sub>2</sub>O (pH 7.5) with gradient elution from 0% to 30% acetonitrile over 40 mins. The oligonucleotides were collected, lyophilized, desalted, and concentrated as appropriate for primer extension and crystallization experiments. Oligonucleotides were characterized by LC-MS (see SI for details).

### Synthesis of $\alpha$ -L-threofuranosyl nucleoside-3'-phosphoro-(2-aminoimidazolidine) (2AIptG and 2AIptC)

The activated monomers 2AIptG and 2AIptC were synthesized essentially as previously reported for the corresponding activated ribonucleotides (18). Briefly, to a solution of 30 mM  $\alpha$ -L-threofuranosyl-3'-monophosphate (tGMP and tCMP) (17,19) and 2-aminoimidazolium chloride (5 eq.) in DMSO was added triphenylphosphine (PPh<sub>3</sub>, 10 eq.), triethylamine (TEA, 5 eq.) and 2,2'-dipyridyldisulfide (DPDS, 9 eq.) under stirring at room temperature. The reaction was allowed to proceed for 4 h, followed by precipitation of the crude product, supernatant removal and product purification by reverse phase flash chromatography. For detailed experimental procedures, see Supporting Information 1.4.

### Primer extension assays

Primer (1  $\mu$ M) and template (1.5  $\mu$ M) were pre-mixed in 15  $\mu$ l of a solution containing 200 mM HEPES pH 8.0 and 200 mM MgCl<sub>2</sub>. The primer extension reaction was initiated by adding the indicated activated monomers (2AIptG, 2AIptC, 2AIpC or 2AIpG) from 200 mM stock solutions to a final concentration of 40 mM in the reaction. At the indicated time points, 1  $\mu$ l aliquots were removed and quenched with 29  $\mu$ l of 90% (v/v) aqueous formamide containing 15 mM EDTA. Samples were heated at 95°C for 2 min, cooled to room temperature, and analysed on 18% (19:1) denaturing PAGE gels containing 7 M urea. Gels were imaged using a Typhoon 9410 scanner. The resulting gel images were analysed using ImageJ software.

### Crystallization

RNA samples (0.5 mM duplex or hairpin in water) were heated to 80°C for 5 min, slowly cooled to room temper-

ature (20°C), and incubated at 4°C overnight prior to crystallization. In the self-complementary RNA duplexes, the overhanging 5'-termini were composed of locked 5-Me-C residues, in order to rigidify the template in the A-form conformation and facilitate crystallization. Nucleic Acid Mini Screen Kits, Natrix (Hampton Research) and Nuc-Pro-HTS (Jena Bioscience) were used to screen crystallization conditions at 18°C using the sitting-drop method. An NT8 robotic system and Rock Imager (Formulatrix, Waltham, MA) were used for crystallization screening and monitoring the crystallization process. Optimal crystallization conditions are listed in the Supporting Information (Supplementary Table S1).

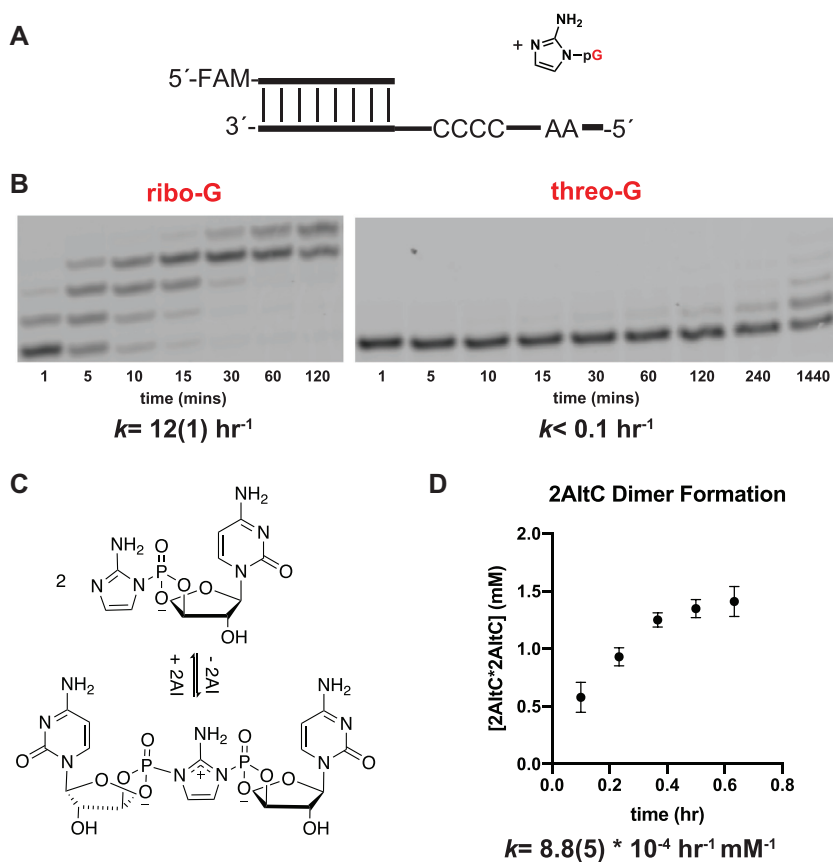
### Data collection, structure determination and refinement

Crystals were mounted and soaked for 2 min in a drop of mother liquor solution which contained 30% glycerol as a cryo-protectant, then transferred to liquid nitrogen. Diffraction data were collected at a wavelength of 1 Å under a liquid nitrogen stream at -174°C on Beam line 821 or 822 at the Advanced Light Source in the Lawrence Berkeley National Laboratory (USA). The crystals were exposed for 1 s per image with a 1° oscillation angle. The distances between detector and the crystal were set to 200–300 mm. Raw diffraction data was processed using HKL2000 (20). Structures were solved by molecular replacement (21) using our previously determined structures (PDB codes 5VCI, 5HBX or 5HBW) as search models. All structures were refined using Refmac (22) in the CCP4 suite. The refinement protocol included simulated annealing, positional refinement, restrained B-factor refinement, and bulk solvent correction. The topologies and parameters for locked nucleic acids, threose mononucleotide (tGMP), threose nucleotide residues (tA and tG) and 3'-3'-imidazolium-bridged di-threose guanosine intermediate (tGp-AI-ptG) were created using ProDrg in the CCP4 suite (23) and applied during refinement. After several cycles of refinement, a number of ordered waters were added. Data collection, phasing, and refinement statistics of the determined structures are listed in the Supporting Information (Supplementary Tables S2 and S3).

## RESULTS

### Activated threo-nucleotides are poor substrates for nonenzymatic primer extension

*Primer extension using 2AIptG as substrate.* We first sought to determine whether an RNA primer could be extended with multiple activated threo-nucleotide monomers. To generate the monomer, we activated  $\alpha$ -L-threofuranosyl guanosine 3'-monophosphate (tGMP) to its corresponding 3'-phosphoro-2-aminoimidazolidine (2AIptG) following the standard procedure for the preparation of activated ribonucleotides (24) (details in Supporting Information). For our initial test of primer extension using activated tG, we used an all-RNA primer and an RNA template oligonucleotide with the templating sequence 3'-CCCCAA-5' (Figure 2) (16). Although primer extension was observable in the presence of 40 mM activated threo-nucleotide monomer 2AIptG, it was extremely slow (est. <0.1 h<sup>-1</sup>). In contrast,



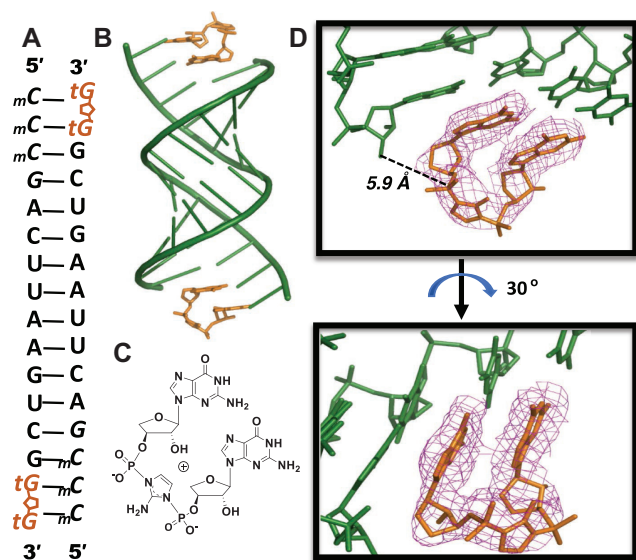
**Figure 2.** Comparison of nonenzymatic primer extension using either an activated ribonucleotide or an activated threo-nucleotide. (A) Schematic representation of a primer extension reaction with either an activated ribo-nucleotide (left) or an activated threo-nucleotide (right). (B) Gel electrophoresis images and rates of primer extension for either an activated ribo-nucleotide (left) or an activated threo-nucleotide (right). All reactions were performed at pH 8.0, in 200 mM HEPES, 200 mM  $\text{Mg}^{2+}$ , 40 mM 2AIpG. Values are the mean rate constant  $\pm$  SD in parentheses, with the last digit reported being the last significant figure and the one in which error arises from triplicate experiments. (C) Schematic representation of the dimerization equilibrium of 2-aminoimidazole activated threo-C. (D) Accumulation of imidazolium-bridged threo-C at 60 mM 2AIpC over the first 38 min. The observed initial rate of the formation of imidazolium-bridged threo-C is  $8.8 \times 10^{-4} \text{ h}^{-1} \text{ mM}^{-1}$ . See Supplementary Data for kinetics analysis of primer extension, dimer formation and hydrolysis (Supplementary Data, Supplementary Figure S4 and S5, 11–14).

primer extension reactions containing the activated ribonucleotide monomer 2AIpG yielded >95% extended product within one hour, and the observed rate constant for disappearance of unreacted primer was  $12 \text{ h}^{-1}$  (Figure 2B). Thus, primer extension with 2AIpG alone is at least 100-fold slower than with 2AIpG, in the context of an RNA primer and an RNA template. We considered the possibility that the primer template context might affect primer extension with threo-nucleotides, e.g. geometrical constraints might lead to a requirement for a TNA primer and/or a TNA template. However, all tested combinations including a primer ending in a TNA residue and a TNA template led to even slower primer extension with 2AIpG as the monomer (Supplementary Figure S1).

We have previously shown that nonenzymatic RNA primer extension is mediated by a highly reactive imidazolium-bridged dinucleotide intermediate, which has  $\sim 100$  fold greater affinity for the template than a mononucleotide, and has a preorganized conformation that is highly favorable for nucleophilic attack by the 3'-hydroxyl of the primer (25). We therefore attempted to

observe the formation of the corresponding TNA dinucleotide intermediate under primer extension conditions (Figure 2C). As previous reports suggest that nucleobase identity has little impact on the rate of formation of the imidazolium-bridged dinucleotide intermediate (25,26), we utilized 2AIpC as a model substrate, which allowed us to investigate intermediate formation at a much higher concentration. When we incubated purified 2AIpC at a concentration of 60 mM at pH 8 and followed the reaction by  $^{31}\text{P}$ -NMR, we found that the imidazolium-bridged TNA dinucleotide is formed, but the observed rate of  $8.8 \times 10^{-4} \text{ h}^{-1} \text{ mM}^{-1}$  is approximately 4–5-fold slower than the corresponding rates for the activated ribonucleotides 2AI-rA and 2AI-rC ( $3.2$  and  $4.5 \times 10^{-3} \text{ h}^{-1} \text{ mM}^{-1}$ ) (26,27) (Figure 2D). Nonenzymatic copying rates are dependent on multiple factors, including the rate of formation of imidazolium-bridged dinucleotide intermediates and their reactivities. Our observations suggest that the slow rate of primer extension with activated threo-nucleotide monomers is due at least in part to the slower formation of the necessary imidazolium-bridged intermediate.





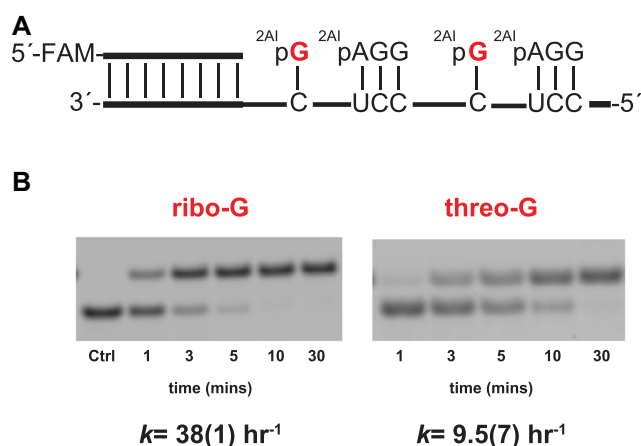
**Figure 3.** Structure of the imidazolium-bridged tG-dinucleotide intermediate bound to an RNA primer/template complex. (A) Diagram and designed secondary structure of RNA primer/template complex. The imidazolium-bridged tG-dinucleotide intermediate in primer extension is shown bound to the template overhang. The tGp-AI-ptG dinucleotides (orange) are bound at both ends of the duplex. (B) Overall crystal structure of TNA-D1 complex. (C) Chemical structure of the imidazolium-bridged threo-guanosine dinucleotide intermediate. (D) Local views of the tG-intermediate bound to the template by two Watson-Crick base-pairs. Pink mesh indicates the corresponding  $2F_o-F_c$  maps contoured at  $1.5\sigma$ . The distance between the 3'-OH group of the primer and the P atom of the adjacent tG of the bound intermediate is 5.9 Å, and the O-P-N angle of attack is  $122^\circ$ .

*2AIptG forms an imidazolium-bridged intermediate that binds to the template.* Nonenzymatic template-directed primer extension with activated ribonucleotides proceeds through a covalent imidazolium-bridged dinucleotide intermediate (25). As shown above, activated tG monomers do form an equivalent intermediate, albeit more slowly. In order to visualize the structure of the TNA intermediate, and to gain insight into the reasons for its slow formation and possibly slow reactivity with the primer/template complex, we undertook a crystallographic analysis. In prior work, we obtained direct structural evidence for the imidazolium-bridged RNA intermediate (Gp-AI-pG) by first co-crystallizing an RNA-dGMP complex and then exchanging the unactivated nucleotides with an activated guanosine-5'-phosphoro-(2-aminoimidazole) by soaking the crystals in a solution of activated monomer (28). Here, we carried out a similar procedure by soaking the same RNA-dGMP co-crystals with a solution of the activated threo-nucleotide monomer 2AIptG. Using a self-complementary LNA-modified RNA sequence, 5'-mCmCmCGACUUAAGUCG-3' (Figure 3A), we sought to observe formation of the dinucleotide intermediate on the 5'-mCmC binding site on the template. We screened a range of soak times to estimate the appropriate timescale for exchange and intermediate formation. The optimal structure was determined after soaking the RNA-dGMP crystal in a 20 mM 2AIptG solution for 24 h. At each end of the A-helical duplex, we observe density consistent with

imidazolium-bridged TNA guanosine dinucleotides (tGp-AI-ptG) bound to the template (TNA-D1, Figure 3B and C). It is noteworthy that formation of the tGp-AI-ptG intermediate in the crystal was much slower than formation of the corresponding RNA intermediate Gp-AI-pG (24 h versus  $\sim 1$  h) (28), consistent with our solution phase NMR observations. Both guanine nucleobases in the imidazolium-bridged TNA intermediate tGp-AI-ptG are well ordered and are bound to the template through Watson-Crick base pairing. Critically, the electron density between the two nucleotides is consistent with the formation of an imidazolium bridge. The threose sugars are only moderately well-ordered in the complex, making it difficult to draw conclusions about the sugar conformation in tGp-AI-ptG. Nevertheless, the phospho-imidazolium-phospho bridge is sufficiently well ordered to enable an estimate of the attack geometry. The distance between the 3'-OH of the primer and the incoming P atom of tGp-AI-ptG molecule is  $\sim 5.9$  Å in the refined model. This distance is significantly longer than that observed in the RNA complex containing native Gp-AI-pG (4.6 Å) (28). Furthermore the O-P-N angle of attack is  $122^\circ$ , which is less than the previously observed angles of  $132^\circ$  and  $170^\circ$  for the all-RNA complex. Both the longer distance and smaller angle could contribute to the slower primer extension observed with activated threo-nucleotides. At the other end of the duplex, the tGp-AI-ptG dinucleotide is significantly disordered, which makes it impossible to determine a reliable 3'-OH to P distance.

*Primer extension with 2AIptG and an activated downstream helper oligoribonucleotide.* Although primer extension with multiple sequential TNA monomers is extremely slow, we considered the possibility that an isolated TNA monomer adjacent to an activated RNA oligonucleotide might show improved reactivity. For example, the formation of an imidazolium-bridged TNA-RNA intermediate might be faster if the imidazole moiety of 2AIptG could attack the phosphate of an adjacent 2AIpG more readily due to the reduced steric bulk around the ribonucleotide phosphate, which is held away from the sugar by C5'. We examined the kinetics of primer extension in this format using an RNA primer/template duplex, with binding sites on the template for 2AIptG and an adjacent 2AI activated RNA trimer (Figure 4). In this context, the rate of primer extension with the activated threo-nucleotide 2AIptG was only four-fold slower than with the corresponding ribonucleotide 2AIpG (Figure 4B). This simplest explanation for this result is that mixed TNA-RNA intermediates both form and react more rapidly than TNA-TNA intermediates, permitting threo-nucleotide incorporation with only a modest kinetic defect.

To further investigate possible reasons for poor primer extension with threo-nucleotide monomers, we obtained two additional structures of RNA oligonucleotides with bound threo-nucleotide monomers. We first co-crystallized the unactivated  $\alpha$ -L-threofuranosylguanine-3'-monophosphate (tGMP) monomer with a short partially self-complementary RNA 7-mer (TNA-M1, sequence: 5'-mCmUGUACA-3', italic: locked nucleotides, mC: locked 5-methylcytidine, mU: locked 5-methyluridine, Supplementary Figure S2). This oligonucleotide self-anneals to form a central six base pair duplex, flanked by single overhanging



**Figure 4.** Evaluation of threose nucleotides in nonenzymatic primer extension. (A) Schematic representation of primer extension using an RNA template. 2AIpG represents 2-aminoimidazole activated guanosine monomer and 2AIpAGG represents 2-aminoimidazole activated RNA trimer helper. (B) Gel electrophoresis images and rates of primer extension for 2-aminoimidazole activated nucleotides. All reactions were performed at pH 8.0, 200 mM HEPES, 200 mM  $Mg^{2+}$ , 20 mM 2AIpG, 1 mM of activated RNA trimer. Values are reported as the mean rate constant  $\pm$  SD in parentheses, with the last digit reported being the last significant figure and the one in which error arises from triplicate experiments. See Supplementary Data for kinetics analysis of primer extension (Supplementary Data, Supplementary Figures S6 and S7).

5-Me-C residues, to which the complementary G or tG monomer can bind. When incubated with tGMP, crystals grew in the hexagonal  $P6_3$  space group, and we determined the structure to a resolution of 2.36 Å. Similar to our previously reported RNA-GMP complex structure (29), the helical duplexes are end-to-end stacked, and a tGMP nucleotide is bound to the templating 5-methylcytidine residue at each end in a Watson-Crick base pair (Supplementary Figure S2B). The sugar is in the  $C4'$ -*exo* conformation in both tGMP ligands (Supplementary Figure S2C). The well-defined electron density of the 3'-phosphate of tGMP shows that the distance between the 3'-OH of the RNA primer and the phosphorus atom of tGMP is  $\sim 6.7$  Å, slightly longer than the distance observed for the ribo-nucleotide monomer ( $\sim 6.3$  Å) (29).

To examine whether a structured downstream oligonucleotide might help to preorganize the noncovalently bound tGMP, an effect we have previously observed with RNA monomers, we co-crystallized a tGMP monomer sandwiched within a gapped RNA hairpin structure (30,31). The resulting structure (TNA-M2, Supplementary Figure S2D and E) reveals that the sandwiched tGMP nucleotide is in a well-ordered Watson-Crick pair, and the tGMP sugar is in the  $C4'$ -*exo* conformation (Supplementary Figure S2F). Based on the well-defined electron density of the tGMP 3'-phosphate, the distance between the 3'-OH of the primer and the incoming phosphorus atom is  $\sim 4.6$  Å, similar to but slightly longer than the corresponding distance observed for a bound ribonucleotide ( $\sim 4.4$  Å) (30). In both structures, the longer distance between the primer 3'-OH and the bound tGMP monomer appears to be influenced by the structure of the threose-nucleotide, in which the phosphate

group is directly attached to the threose-furanosyl ring instead of being held away from the sugar by the 5'-methylene of ribose.

One noteworthy conformational difference between complexes TNA-M1 and TNA-M2 is that the bound monomers have  $\chi$  torsion angles that differ by  $>40^\circ$ . Without any helper to enhance the binding, the  $\chi$  angle of tGMP in TNA-M1 is  $-177^\circ$ , which is closer to the characteristic  $\chi$  torsion angle in an A-form helix. When the threose-nucleotide is sandwiched between two well-structured RNA duplexes, the N-glycosidic bond of the TNA monomer is rotated, resulting in a  $\chi$  angle of  $-134^\circ$ . This latter angle is closer to those observed for the covalently incorporated TNA residues in the TNA-P1 and TNA-T1 (see below) stem-loop structures (Table 1), suggesting that the conformational constraints imposed by the sandwiched RNA complex induce a threose-nucleotide geometry that is distinct from that of the free species. Moreover, the threose-nucleotide monomers in TNA-M1 and TNA-M2 have different  $\gamma$  torsion angles from the backbone-linked threose-nucleotides in TNA-P1 and TNA-T1 (Table 1). In the absence of covalent linkage to the RNA backbone at either the 3' or 2' terminus, the tGMP monomers in both TNA-M1 and TNA-M2 complexes may be in a more relaxed conformation, to differing extents depending on the presence of a proximal downstream oligonucleotide.

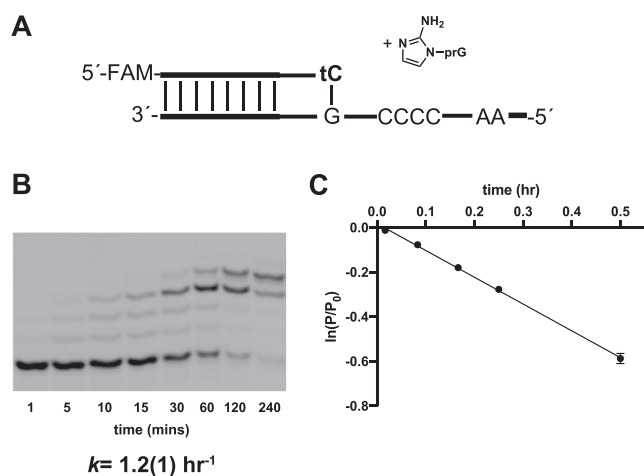
#### Kinetic and structural consequences of TNA residues in the primer and the template

*Effect of a threose-nucleotide at the primer terminus.* We then asked whether nonenzymatic extension beyond a terminal threose-nucleotide at the end of an RNA primer was observable. Our previous results with 2'- $NH_2$  modified threose-nucleotides suggested that primer extension from a threose-nucleotide should be possible (14). However, because the amino sugar modification is known to affect nucleotide conformation (32) and because the amino group is more nucleophilic than a hydroxyl (33), it was unclear how well primer extension would proceed from an unmodified threose-nucleotide. We therefore synthesized an RNA primer containing a single terminal threose-cytidine (tC) residue at its 3' end (2' end when tC is in the terminal position, 5'-(FAM)-AGUGAGUAACGtC-2', Figure 5A). We then compared rates of RNA-templated primer extension using activated 2AIpG monomer as a substrate for the extension of this tC-terminal primer versus an all-RNA control primer. In the control RNA primer/template system, primer extension proceeded rapidly, with an observed rate of  $12\text{ h}^{-1}$ . With the tC-terminal primer, however, the observed rate was reduced to  $1.2\text{ h}^{-1}$ , 10-fold slower than the RNA extension reaction. After incubation for 24 h, the tC-terminated primer was primarily converted a +4 nucleotide product (Figure 5B). Thus, once the primer is extended with a ribonucleotide, continued primer extension can proceed to the end of the template.

*Structural consequences of a single threose-nucleotide at the primer terminus.* To gain insight into the structure of a reaction center containing a priming threose-nucleotide, we crystallized an RNA hairpin-monomer complex (TNA-P1,

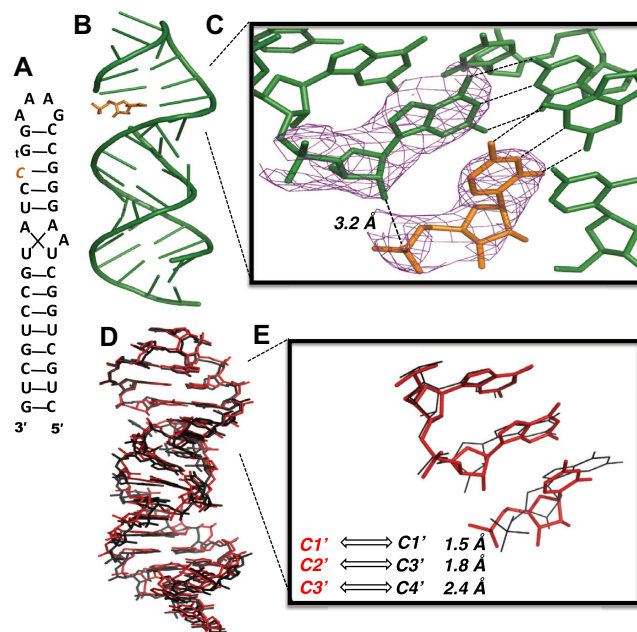
**Table 1.** Torsion angles in degrees of TNA residues. (na = not available)

Complex	Status of Residue	$\alpha$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\chi$
TNA-P1	Primer	-76.5	-177.4	163.6	na	na	-121.3
TNA-T1	Template	-75.4	-177.1	140.1	-96.2	-111.4	-111.4
TNA-M1	Paired monomer	na	-137.1	172.0	na	na	-177.2
TNA-M2	Paired monomer	na	-138.2	171.4	na	na	-134.7
2'-NH <sub>2</sub> tG (13)	Free monomer	na	na	163.1	na	na	-138.2



**Figure 5.** Primer extension with an RNA primer ending in a threonucleotide. (A) Schematic representation of a primer extension reaction. Primer extension reactions were carried out in triplicate using 40 mM 2AlpG, pH 8.0, in 200 mM HEPES, 200 mM Mg<sup>2+</sup>. (B) Representative PAGE analysis of result. Values are the mean rate constant  $\pm$  SD in parentheses, with the last digit reported being the last significant figure and the one in which error arises from triplicate experiments. (C) Plot of  $\ln(P/P_0)$  as a function of time. The rate of extension was determined from linear least-square fits of the data from three independent experiments.

Figure 6A), in which the RNA stem contained a 2' terminal threoguanosine (tG) residue followed by a single-nucleotide gap. One unactivated cytidine monophosphate (CMP) mononucleotide was sandwiched noncovalently between the tG primer and the downstream helper oligonucleotide, to mimic the bound monomer in the ground state prior to primer extension. As in our previously determined RNA-monomer hairpin structure (34), the entire RNA-monomer hairpin crystallized in the R3 space group, with one complex per asymmetric unit (Figure 6B). We determined the structure to a resolution of 2.7 Å. The tG-containing RNA, downstream RNA helper and bound CMP monomer assembled to form the desired hairpin, with the tetraloop of one hairpin docked with the receptor motif of a separate hairpin complex. The terminal tG residue at the 2'-end of the primer was well ordered, with the guanine nucleobase Watson-Crick base-paired with the templating cytidine. The threose sugar is in a C4'-*exo* conformation with the 2' and 3' hydroxyls pseudo-diaxial (Figure 6C), consistent with a prior report from Egli et al. in which a TNA residue was present in the middle of an oligonucleotide (8). The 2'-OH group of the tG residue is oriented toward the phosphate of the incoming CMP monomer, which is Watson-Crick paired with the template. The distance between the P atom of the bound CMP and the 2'-



**Figure 6.** Three-nucleotide substitution at the end of an RNA primer has minor structural effects. (A) Diagram and designed secondary structure of RNA hairpin-monomer complex TNA-P1. The primer and template are linked together to form a hairpin structure. A CMP monomer (orange) is sandwiched between the TNA guanosine (tG) at the end of the primer and the flanking downstream oligonucleotide. (B) Overall crystal structure of TNA-P1 complex. Green: primer-template hairpin and downstream RNA oligonucleotides. Orange: CMP monomer. (C) Local view of the terminal tG residue and the adjacent CMP monomer. Pink mesh indicates 2F<sub>o</sub> - F<sub>c</sub> maps contoured at 2.5 σ. (D) Side view of the superimposed structures of the previously reported all-RNA hairpin structure (5VCI) and the TNA-P1 complex. Red: TNA-P1 structure; black: 5VCI structure. (E) Local view of the superimposed structures. Distances between corresponding atoms in the threose sugar and ribose sugar versions are labelled (middle nucleotide of overlaid structures).

OH of the tG primer is  $\sim 3.2$  Å, which appears favourable for attack, however as the monomer is unactivated CMP, the structure is not informative about the angle of attack with respect to the leaving group.

The overall structure of the TNA-P1 complex is highly similar to that of the RNA hairpin-monomer complex presented in our recent report (PDB: 5VCI) (34). The two structures are fully superimposable (RMSD = 0.96 Å), including the RNA backbone and nucleobase positions (Figure 6D). Thus, a single threonucleotide at the end of an RNA duplex is accommodated with minimal structural perturbation. The largest deviations between the two structures are at the primer terminus, where the tG residue is shifted and twisted compared to the native RNA primer, with distances in the range of 1.5–2.4 Å between the corresponding atoms



of the threose and ribose sugar rings (Figure 6E). The tG residue has a glycosidic linkage torsion  $\chi$  angle of  $-121^\circ$  (Table 1), compared to a typical  $\chi$  angle of  $-168^\circ$  for the ribonucleotides of the primer (26).

**Primer extension on templates containing TNA nucleotides.** We next sought to investigate the effect on primer extension of single or multiple TNA residues in the template. Previous work from Switzer et al. showed that primer extension across an all TNA template region was possible but slow (12). We therefore began by comparing primer extension on a template region consisting of four ribo- or threocytidine residues, using an all-RNA primer and activated ribo-guanosine (2AIPG) as the monomer substrate (Supplementary Figure S3). Consistent with the previous results of Switzer et al. (13), primer extension on the (tC)<sub>4</sub> template ( $1.3 \text{ h}^{-1}$ ) was roughly ten-fold slower than on a (rC)<sub>4</sub> template ( $12 \text{ h}^{-1}$ ). (Supplementary Figure S3B and S10)

In a prebiotic situation in which both ribo- and threonucleotides were present, templates in which isolated threo-nucleotides were interspersed in an otherwise RNA oligonucleotide might be common. However, the effect of an isolated threo-nucleotide in an otherwise all-RNA template has not been previously investigated. We therefore synthesized two new template oligonucleotides (5'-AAGGG(t)GCCAGUCAGUCUACGC-3'), one all-RNA and one containing a single threo-guanosine nucleotide (tG) at the +1 position (Figure 7). When the tG residue was present in the template strand at the +1 position the pseudo-first order reaction rate for the first addition was  $5.6 \text{ h}^{-1}$ , compared to  $11 \text{ h}^{-1}$  for the all-RNA template (Figure 7B). However, extension of the +1 product to longer products was much slower (estimated rate constant for conversion of +1 to +2/3/4 products  $0.71 \text{ h}^{-1}$ ), a decrease comparable to that seen for a TNA residue at the end of the primer. Thus, a TNA:RNA base-pair seems to significantly slow primer extension, irrespective of whether the TNA residue is in the primer or the template. However, subsequent extension to the +3 and +4 products was not observably inhibited, and almost complete conversion of the primer to fully extended +4 products was observed after 24 hours.

**Structural consequences of a single TNA residue in the template strand.** To gain structural insight into the conformation of the reaction center in TNA-templated nonenzymatic primer extension, we co-crystallized and determined the structure of an RNA hairpin complex, similar to that used above, but this time containing a single TNA residue on the template strand of a gapped stem. The RNA hairpin-monomer complex (TNA-T1) contains a single tG residue at the appropriate position to serve as the template to bind a free CMP monomer (Figure 8A). Broadly, TNA-T1 and TNA-P1 shared similar structural features, including an RNA A-form stem, tetraloop/receptor motif, and sandwiched monomers (Figure 8B). The TNA-T1 complex also crystallized in the R3 space group, and we determined the structure to a resolution of 2.95 Å. The tG residue on the template strand is well-ordered, forming three hydrogen bonds with the bound CMP monomer through a Watson-Crick base-pairing interaction. The tG threose sugar ring is in the C4'-*exo* conformation, and the glycosidic torsion

angle  $\chi$  of tG is  $-111^\circ$ , similar to the tA residue seen at the end of the primer. However, the noncovalently bound CMP is poorly ordered, suggesting a significant degree of conformational heterogeneity. The electron density of the monomer's sugar and phosphate is poorly resolved, yielding only a rough estimate of 4 Å for the distance between the 3'-OH of primer and the P atom of the CMP (Figure 8C).

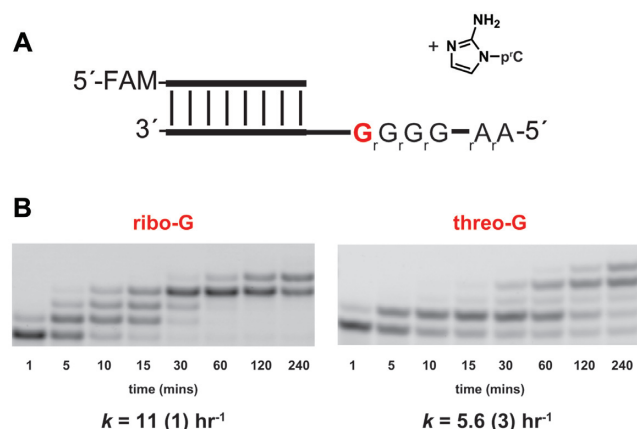
Superimposing the TNA-T1 structure over the all-RNA complex 5VCI, we find the two complexes are very similar in overall conformation (RMSD = 0.64 Å, Figure 8D). The tG threose sugar, constrained by stacked upstream and downstream RNA nucleotides, is shifted by 1.1–1.7 Å compared to the native RNA ribose sugar ring (Figure 8E). The torsion angles of the tG backbone are presented in Table 1. As expected, the tG residues in the TNA-P1 and TNA-T1 complexes share similar  $\alpha$  and  $\gamma$  torsion angles, likely due to similar structural constraints on the threo-nucleotide residue in the RNA context of both hairpin structures. Overall, a single TNA nucleotide substitution on the template does not dramatically disrupt the RNA structure.

## DISCUSSION

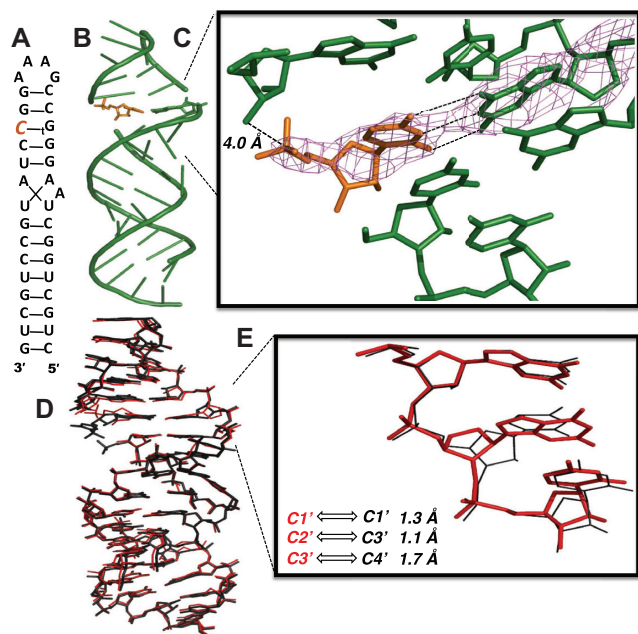
For many years, the RNA World hypothesis has been a guiding force for studying the transition from simple chemistry to complex cellular life, bolstered by recent advances in prebiotically plausible pathways for the generation of RNA building blocks from simple feedstocks (35,36) and in the nonenzymatic copying of RNA templates in prebiotically plausible compartments (37). At the same time, we and others have explored alternative genetic materials with the potential to store and replicate genetic information (1). Among these alternatives, TNA is a striking candidate that, because of its simpler chemistry, might present a lower hurdle for abiogenesis (4). However, TNA appears to have poor reactivity in nonenzymatic copying reactions, regardless of whether the threo-nucleotide 2'-position is the native -OH group, as shown in our current work, or the highly nucleophilic -NH<sub>2</sub> group (14).

Our structural studies provide reasonable explanations for the moderately reduced rates of primer extension from a primer ending in a threo-nucleotide, and across from a threo-nucleotide in the template strand. In the corresponding interrupted RNA hairpin complexes TNA-P1 and TNA-T1, which have single threo-nucleotide substitutions at the end of the primer or in the template strand, the overall conformations are generally comparable to those of the all-RNA complexes. In this stem-loop RNA context, TNA residues adopt an extended conformation so as to be accommodated within the RNA A-form duplex with minimal global distortion, and with the formation of well-ordered base pairs with complementary RNA nucleotides on the opposite strand. These structural findings are in agreement with the observation that TNA strands form thermodynamically stable heteroduplexes with RNA (4,8). The modeled 2'-O to 5'-P distance for nucleophilic attack in the case of the threo-nucleotide terminated primer is also comparable to that seen in fully RNA complexes. Given that duplex structure is unperturbed and that the local structure is only minimally distorted, it is reasonable that only moderately





**Figure 7.** Evaluation of nonenzymatic primer extension with 2-aminimidazole activated ribo-cytidine (2AIpC) across from a single templating threoguanosine (tG) residue. (A) Schematic representation of a primer extension reaction of 2AIpC with template containing either ribo-G (rG) or threo-G (tG). (B) Gel electrophoresis images and rates of primer extension with 2-aminimidazole activated cytidine. All reactions were performed at pH 8.0, 200 mM HEPES, 200 mM  $\text{Mg}^{2+}$ , 40 mM 2AIpC. Values are reported as the mean rate constant  $\pm$  SD in parentheses, with the last digit reported being the last significant figure and the one in which error arises from triplicate experiments. See Supplementary Data for kinetics analysis of primer extension (Supplementary Data, Supplementary Figures S8 and S9).



**Figure 8.** A threo-nucleotide substitution in an RNA template results in only minor and local structural changes. (A) Diagram and designed secondary structure of RNA hairpin-monomer complex. A CMP monomer (orange) is sandwiched between the 3'-end of the primer and a downstream oligonucleotide. The TNA substitution in the template is denoted by  $tG$  (B): Overall crystal structure of TNA-T1 complex. (C) Local view of the  $tG$  template residue, showing Watson-Crick base-pairing to the bound CMP monomer. Pink mesh indicates the corresponding  $2F_o - F_c$  maps contoured at  $2.5 \sigma$ . (D) Side view of the superimposed structures of the previously reported RNA hairpin structure (5VCI) and the TNA-T1 complex. Red: TNA-T1 structure; black: 5VCI structure. (E) Local view of the superimposed structures. The distances between the positions of the corresponding atoms in the threose sugar and the corresponding ribose sugar are labeled.

reduced reactivity is observed for ribonucleotide addition to a threo-nucleotide terminated primer or across from a threo-nucleotide in the template.

Remarkably, the extremely slow rate of nonenzymatic primer extension with 2AI-activated threo-nucleotide monomers alone appears to result largely from the failure of activated threo-nucleotides to form significant levels of the imidazolium-bridged dinucleotide intermediate in solution, with a possible further contribution from slower reaction of intermediate that does form with the primer. The reaction of two 2AI-activated threo-nucleotides with each other to form the imidazolium-bridged dinucleotide intermediate involves the attack of the unprotonated N3 of the 2AI moiety of one threo-nucleotide on the phosphate of a second activated monomer, with its protonated 2AI being the leaving group. The low level of imidazolium-bridged dinucleotide seen in solution was at first surprising, because the reacting groups are distant from the threose sugar that distinguishes TNA from RNA. However, the phosphate of an activated threo-nucleotide is more sterically hindered than that of a ribo-nucleotide, because the TNA phosphate is directly linked to the threose furanosyl ring such that it is flanked by the nucleobase and the sugar as well as the covalently attached 2AI moiety. We speculate that this steric hindrance is responsible for the low levels of imidazolium-bridged TNA intermediate in solution, and hence the slow rate of primer extension with 2AI-activated threo-nucleotide monomers. Nevertheless our observations show that the imidazolium-bridged TNA intermediate can form *in crystallo*, similarly to activated RNA monomers (28), albeit much more slowly. The formation of the imidazolium-bridged TNA dinucleotide in the crystal may be facilitated when two activated threo-nucleotide monomers are held in close proximity for extended periods of time as a result of being bound next to each other on the template. Additional experiments will be required to identify all of the factors that control the formation of the imidazolium-bridged TNA intermediate both in solution and *in crystallo*.

Our crystallographic findings show that both TNA monomers and the TNA imidazolium-bridged dinucleotide do bind to the template by Watson-Crick base pair-

ing, and the threose sugar is consistently in the C4'-*exo* conformation. The unactivated 3'-phosphate in single threo-nucleotides is shifted towards the helix major groove, increasing the distance from the attacking 3'-OH of the primer. Compared to single TNA residues in the primer or template, the  $\gamma$  torsion angles of free monomers adopt distinct values that influence the orientation of the 3'-phosphate group. The noncovalently bound TNA monomer, unconstrained by linkage to the RNA backbone, appears to relax to a conformation that is unfavorable to primer extension. In the case of the imidazolium-bridged dinucleotide, the threose sugars are disordered, but the appearance of electron density between the two threo-nucleotides clearly indicates the formation of the critically important imidazolium-bridge. The estimated distance between the primer 3'-OH and the P atom of the adjacent TNA nucleotide is significantly longer than for a ribonucleotide (5.9 Å vs 4.6 Å), suggesting that even if 2AI-activated TNA monomers do form the intermediate and bind the template, primer extension would be very slow.

Our results cast doubt upon the viability of TNA as a self-replicating polymer, at least in the context of nonenzymatic template-directed primer extension with phosphorimidazolide-activated threo-nucleotide monomers. The copying of a C4 template region with 2AI-activated tG monomers was extremely inefficient no matter whether the primer ended in a ribo- or a threo-nucleotide, and no matter whether the template region was composed of a ribo- or a threo-nucleotides. Since even copying in an essentially all TNA context was ineffective, the inability of 2AIptG monomers to mediate efficient multi-step primer extension cannot simply reflect the local structural context of the reaction. Instead, the more likely explanation follows from our observation that formation of an imidazolium-bridged tG-p2AIp-tG dinucleotide is inefficient, such that the low levels of this key intermediate lead to poor primer extension irrespective of the local structural context.

We have previously considered a scenario in which ribonucleotides were the predominant but not the only prebiotic synthetic product, such that heterogeneous oligonucleotides might have assembled by non-templated chemistry. Just as in our previous consideration of arabinonucleotides, we now find that isolated threo-nucleotides in the template can be copied over by RNA primer extension, but more slowly than an all-RNA template. Similarly, a primer ending in a single threo-nucleotide can be extended with ribonucleotide monomers, but with a ten-fold kinetic defect compared with an all-RNA primer. Interestingly, the kinetic defect associated with an arabino-terminated primer is more severe (~100-fold). In contrast, the kinetic defect associated with primer extension with threo-nucleotide monomers is context dependent. The most likely explanation for the slow polymerization of the arabino-substrate is the suboptimal local conformation of the arabino-nucleotide backbone, because the rate of formation of the arabino-intermediate in solution is comparable to that of the RNA counterpart. However, the slow nonenzymatic copying rates observed with TNA appear to be due to both slow formation of the dinucleotide intermediate and the unfavorable local conformation. Our results

suggest that the major route for the incorporation of threo-nucleotides into a growing chain would be through reaction of the primer with a mixed TNA-RNA imidazolium-bridged intermediate, for which the kinetic defect is only about 4-fold relative to ribonucleotide incorporation. Nevertheless this suggests that in the presence of, for example, an equimolar mixture of activated threo- and ribonucleotides, RNA synthesis would predominate. RNA has an exceptional double helical structure with the ribose sugars in the 3'-*endo* conformation; as a consequence the RNA primer/template/intermediate complex is preorganized so as to favour rapid primer extension. Our experiments show that in this context, RNA copying has a significant kinetic advantage over alternative nucleic acids including not only TNA, but also ANA and DNA (26). Our experimental results are therefore consistent with a model in which relatively homogeneous RNA oligonucleotides emerged from a much more heterogeneous mixture of activated nucleotides through the self-purifying capacity of nonenzymatic RNA replication.

## DATA AVAILABILITY

Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under accession numbers 6U7Y, 6U7Z, 6U89, 6U8F, 6U8U.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank the Szostak lab for helpful discussions, insightful commentary and careful revision of the manuscript. We thank Dr L. Zhou for assistance with LC-MS analysis. X-ray diffraction data were collected at the Advanced Light Source (ALS) SIBYLS beamlines 821 and 822, a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the Department of Energy, Office of Basic Energy Sciences, through the Integrated Diffraction Analysis Technologies (IDAT) program, supported by DOE Office of Biological and Environmental Research. The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health, National Institute of General Medical Sciences, and the Howard Hughes Medical Institute.

## FUNDING

J.W.S. is an Investigator of the Howard Hughes Medical Institute; National Science Foundation [CHE-1607034, in part]; Simons Foundation [290363 to J.W.S.]; National Science Foundation [MCB: 1946312]; W.M. Keck Foundation (to J.C.C.). Funding for open access charge: Howard Hughes Medical Institute and the Simons Collaboration on the Origins of Life.

*Conflict of interest statement.* None declared.

## REFERENCES

- Eschenmoser, A. (1999) Chemical etiology of nucleic acid structure. *Science*, **284**, 2118–2124.
- Eschenmoser, A. (1997) Towards a chemical etiology of nucleic acid structure. *Orig. Life Evol. Biosphys.*, **27**, 535–553.
- Orgel, L. (2000) A simpler nucleic acid. *Science*, **290**, 1306–1307.
- Schöning, K.-U., Scholz, P., Guntha, S., Wu, X., Krishnamurthy, R. and Eschenmoser, A. (2000) Chemical etiology of nucleic acid structure: the  $\alpha$ -threofuranosyl-(3'→2') oligonucleotide system. *Science*, **290**, 1347–1351.
- Serianni, A.S., Clark, E.L. and Barker, R. (1979) Carbon-13-enriched carbohydrates. Preparation of erythrose, threose, glyceraldehyde, and glycolaldehyde with  $^{13}\text{C}$ -enrichment in various carbon atoms. *Carbohydr. Res.*, **72**, 79–91.
- Kim, H.-J. and Benner, S.A. (2017) Prebiotic stereoselective synthesis of purine and noncanonical pyrimidine nucleotide from nucleobases and phosphorylated carbohydrates. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 11315–11320.
- Ebert, M.-O., Mang, C., Krishnamurthy, R., Eschenmoser, A. and Jaun, B. (2008) The Structure of a TNA–TNA complex in solution: NMR study of the octamer duplex derived from  $\alpha$ -(L)-threofuranosyl-(3'-2')-CGAATTCG. *J. Am. Chem. Soc.*, **130**, 15105–15115.
- Pallan, P.S., Wilds, C.J., Wawrzak, Z., Krishnamurthy, R., Eschenmoser, A. and Egli, M. (2003) Why does TNA cross-pair more strongly with RNA than with DNA? An answer from X-ray analysis. *Angew Chem. Int. Ed. Engl.*, **42**, 5893–5895.
- Anosova, I., Kowal, E.A., Sisco, N.J., Sau, S., Liao, J.Y., Bala, S., Rozners, E., Egli, M., Chaput, J.C. and Van Horn, W.D. (2016) Structural insights into conformation differences between DNA/TNA and RNA/TNA chimeric duplexes. *ChemBioChem.*, **17**, 1705–1708.
- Horhota, A., Zou, K., Ichida, J.K., Yu, B., McLaughlin, L.W., Szostak, J.W. and Chaput, J.C. (2005) Kinetic analysis of an efficient DNA-dependent TNA polymerase. *J. Am. Chem. Soc.*, **127**, 7427–7434.
- Chaput, J.C. and Szostak, J.W. (2003) TNA synthesis by DNA polymerases. *J. Am. Chem. Soc.*, **125**, 9274–9275.
- Chaput, J.C., Ichida, J.K. and Szostak, J.W. (2002) DNA polymerase-mediated DNA synthesis on a TNA template. *J. Am. Chem. Soc.*, **125**, 856–857.
- Heuberger, B.D. and Switzer, C. (2006) Nonenzymatic oligomerization of RNA by TNA templates. *Org. Lett.*, **8**, 5809–5811.
- Blain, J.C., Ricardo, A. and Szostak, J.W. (2014) Synthesis and nonenzymatic template-directed polymerization of 2'-amino-2'-deoxythreose nucleotides. *J. Am. Chem. Soc.*, **136**, 2033–2039.
- Bhowmik, S. and Krishnamurthy, R. (2019) The role of sugar-backbone heterogeneity and chimeras in the simultaneous emergence of RNA and DNA. *Nat. Chem.*, **11**, 1009–1018.
- Zhang, S. and Chaput, J.C. (2012) Synthesis of threose nucleic acid (TNA) phosphoramidite monomers and oligonucleotide polymers. *Curr. Protoc. Nucleic Acid Chem.*, **50**, 4.51.1–4.51.26.
- Sau, S.P., Fahmi, N.E., Liao, J.-Y., Bala, S. and Chaput, J.C. (2016) A scalable synthesis of  $\alpha$ -L-threose nucleic acid monomers. *J. Org. Chem.*, **81**, 2302–2307.
- Li, L., Prywes, N., Tam, C.P., O'Flaherty, D.K., Lelyveld, V.S., Izgu, E.C., Pal, A. and Szostak, J.W. (2017) Enhanced nonenzymatic RNA copying with 2-aminoimidazole activated nucleotides. *J. Am. Chem. Soc.*, **139**, 1810–1813.
- Bala, S., Liao, J.-Y., Mei, H. and Chaput, J.C. (2017) Synthesis of  $\alpha$ -L-threofuranosyl nucleoside 3'-monophosphates, 3'-phosphoro (2-methyl) imidazolides, and 3'-triphosphates. *J. Org. Chem.*, **82**, 5910–5916.
- Otwinowski, Z. and Minor, W. (1997) Processing X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
- Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D. Biol. Crystallogr.*, **53**, 240–255.
- Schüttelkopf, A.W. and Van Aalten, D.M. (2004) PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. *Acta Crystallogr. D. Biol. Crystallogr.*, **60**, 1355–1363.
- Joyce, G., Inoue, T. and Orgel, L. (1984) Non-enzymatic template-directed synthesis on RNA random copolymers: Poly (C, U) templates. *J. Mol. Biol.*, **176**, 279–306.
- Walton, T. and Szostak, J.W. (2016) A highly reactive imidazolium-bridged dinucleotide intermediate in nonenzymatic RNA primer extension. *J. Am. Chem. Soc.*, **138**, 11996–12002.
- Kim, S.C., Zhou, L., Zhang, W., O'Flaherty, D.K., Rondo-Brovetto, V. and Szostak, J.W. (2020) A model for the emergence of RNA from a prebiotically plausible mixture of ribonucleotides, arabinonucleotides and 2'-deoxynucleotides. *J. Am. Chem. Soc.*, **142**, 2317–2326.
- Walton, T. and Szostak, J.W. (2017) A kinetic model of nonenzymatic RNA polymerization by cytidine-5'-phosphoro-2-aminoimidazole. *Biochemistry*, **56**, 43, 5739–5747.
- Zhang, W., Walton, T., Li, L. and Szostak, J.W. (2018) Crystallographic observation of nonenzymatic RNA primer extension. *Elife*, **7**, e36422.
- Zhang, W., Tam, C.P., Wang, J. and Szostak, J.W. (2016) Unusual base-pairing interactions in monomer–template complexes. *ACS Cent. Sci.*, **2**, 916–926.
- Zhang, W., Tam, C.P., Zhou, L., Oh, S.S., Wang, J. and Szostak, J.W. (2018) Structural rationale for the enhanced catalysis of nonenzymatic RNA primer extension by a downstream oligonucleotide. *J. Am. Chem. Soc.*, **140**, 2829–2840.
- Prywes, N., Blain, J.C., Del Frate, F. and Szostak, J.W. (2016) Nonenzymatic copying of RNA templates containing all four letters is catalyzed by activated oligonucleotides. *Elife*, **5**, e17756.
- Tereshko, V., Gryaznov, S. and Egli, M. (1998) Consequences of replacing the DNA 3'-oxygen by an amino group: high-resolution crystal structure of a fully modified N3'→P5' phosphoramidate DNA dodecamer duplex. *J. Am. Chem. Soc.*, **120**, 269–283.
- O'Flaherty, D.K., Zhou, L. and Szostak, J.W. (2019) Nonenzymatic template-directed synthesis of mixed-sequence 3'-NP-DNA up to 25 nucleotides long inside model protocells. *J. Am. Chem. Soc.*, **141**, 10481–10488.
- Zhang, W., Tam, C.P., Walton, T., Fahrenbach, A.C., Birrane, G. and Szostak, J.W. (2017) Insight into the mechanism of nonenzymatic RNA primer extension from the structure of an RNA-GpppG complex. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 7659–7664.
- Powner, M.W., Gerland, B. and Sutherland, J.D. (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature*, **459**, 239.
- Becker, S., Feldmann, J., Wiedemann, S., Okamura, H., Schneider, C., Iwan, K., Crisp, A., Rossa, M., Amatov, T. and Carell, T. (2019) Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides. *Science*, **366**, 76–82.
- Adamala, K. and Szostak, J.W. (2013) Nonenzymatic template-directed RNA synthesis inside model protocells. *Science*, **342**, 1098–1100.