

DNA Polymerase-Mediated Synthesis of Unbiased Threose Nucleic Acid (TNA) Polymers Requires 7-Deazaguanine To Suppress G:G Mismatching during TNA Transcription

Matthew R. Dunn,^{†,§} Andrew C. Larsen,[§] Walter J. Zahurancik,^{||} Nour Eddine Fahmi,[§] Madeline Meyers,^{‡,§} Zucui Suo,^{||} and John C. Chaput^{*,‡,§}

[†]School of Life Sciences, [‡]Department of Chemistry and Biochemistry, and [§]The Biodesign Institute, Arizona State University, Tempe, Arizona 85287-5301, United States

^{||}Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, United States

Supporting Information

ABSTRACT: Threose nucleic acid (TNA) is an unnatural genetic polymer capable of undergoing Darwinian evolution to generate folded molecules with ligand-binding activity. This property, coupled with a nuclease-resistant backbone, makes TNA an attractive candidate for future applications in biotechnology. Previously, we have shown that an engineered form of the *Archaeon* replicative DNA polymerase 9^oN, known commercially as Terminator DNA polymerase, can copy a three-letter genetic alphabet (A,T,C) from DNA into TNA. However, our ability to transcribe four-nucleotide libraries has been limited by chain termination events that prevent the synthesis of full-length TNA products. Here, we show that chain termination is caused by tG:dG mismatching in the enzyme active site. We demonstrate that the unnatural base analogue 7-deazaguanine (7dG) will suppress tGTP misincorporation by inhibiting the formation of Hoogsteen tG:dG base pairs. DNA templates that contain 7dG in place of natural dG residues replicate with high efficiency and >99% overall fidelity. Pre-steady-state kinetic measurements indicate that the rate of tCTP incorporation is 5-fold higher opposite 7dG than dG and only slightly lower than dCTP incorporation opposite either 7dG or dG. These results provide a chemical solution to the problem of how to synthesize large, unbiased pools of TNA molecules by polymerase-mediated synthesis.

We have developed a polymerase-mediated replication system that makes it possible to copy genetic information back and forth between deoxyribose nucleic acid (DNA) and threose nucleic acid (TNA) (Figure 1).¹ Under appropriate conditions, an engineered form of the *Archaeon* replicative DNA polymerase 9^oN, known commercially as Terminator DNA polymerase (New England BioLabs, Inc.), can copy sequence-defined DNA templates into TNA through the sequential addition of TNA nucleotide triphosphates.^{2,3} TNA templates purified by denaturing polyacrylamide gel electrophoresis (PAGE) can be reverse-transcribed back into cDNA using Superscript II (SSII), which is a highly engineered form of a reverse transcriptase isolated from the Maloney murine leukemia virus.⁴ Amplification of the cDNA using the

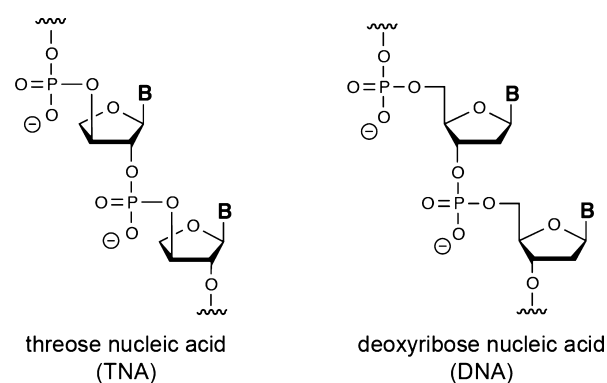


Figure 1. Molecular structure of TNA and DNA. Constitutional structure for the linearized backbone of α -L-threofuranosyl-(3'-2') nucleic acid, TNA (left), and DNA (right). TNA is an unnatural genetic polymer composed of repeating α -L-threose sugars that are vicinally connected by 2',3'-phosphodiester linkages.

polymerase chain reaction (PCR) followed by strand separation allows the process to be performed iteratively.

When transcription and reverse transcription are performed on DNA templates of a known sequence, the overall fidelity of TNA replication can be determined by sequencing the cDNA product. DNA libraries composed of a three-letter genetic alphabet (A,T,C) have been shown to transcribe with high efficiency, and individual sequences replicate with >99% fidelity.¹ This level of fidelity is remarkable, considering the structural differences between DNA and TNA. However, attempts to synthesize TNA libraries that contain a four-letter genetic alphabet (A,C,T,G) have failed due to chain termination events that inhibit TNA synthesis on unbiased DNA templates.² Investigation into this problem revealed that DNA templates with low numbers of isolated dG residues could be transcribed into full-length TNA, but the cDNA obtained after reverse transcription contained a high level of G→C mutations.¹ The low fidelity of TNA replication on four-nucleotide templates can be overcome by replacing tCTP in the transcription mixture with dCTP; however, this change leads to a mixed-backbone TNA–DNA heteropolymer that may not be

Received: November 7, 2014

Published: March 18, 2015

suitable for some applications.¹ This problem has created a need for new replication strategies that can be used to generate large, unbiased pools of TNA for *in vitro* selection.

Recognizing the limitations of G-rich templates, we decided to evaluate the ability of Terminator DNA polymerase to read through G-repeats in a DNA template. We used a primer extension assay in which the DNA template was engineered to contain a central region of 0–3 consecutive dG residues that are flanked by an arbitrary sequence composed of A, C, and T residues (Figure 2A). Terminator was challenged to extend a

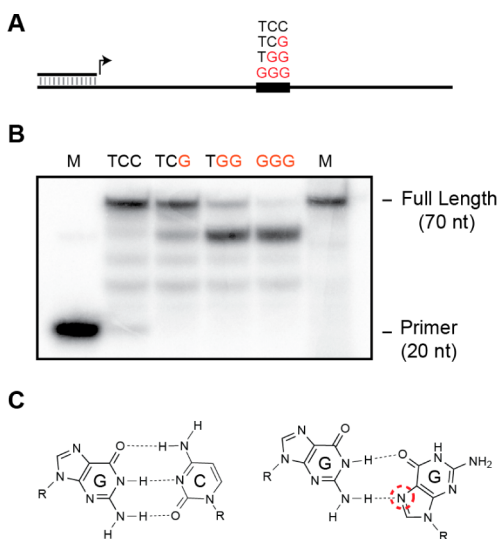


Figure 2. Sequence specificity of Terminator DNA polymerase. (A) Primer extension assay used to evaluate TNA synthesis on dG-containing DNA templates. (B) Primer extension results for DNA templates that contain consecutive G-residues. (C) Molecular structure of a normal G:C Watson–Crick base pair and a G:G Hoogsteen base pair. The analogue of 7-deazaguanine replaces the N-7 nitrogen with a methylene group (red circle).

DNA primer annealed to a DNA template using TNA triphosphates (tNTPs) that were obtained by chemical synthesis.⁵ If the polymerase is able to extend the primer to the end of the DNA template, it will produce an elongated product that is extended by 50 TNA residues. However, if the polymerase is unable to read through the G residues, it will produce a truncated product that is easily detected by denaturing PAGE.

Analysis of the resulting primer extension assay products indicates that Terminator is able to generate full-length TNA product when 0 or 1 dG residue is present in the DNA template (Figure 2B). However, the polymerase stalls on templates that contain 2 or more consecutive dG residues, indicating that G-repeats inhibit the transcription of TNA polymers on DNA templates. This result, in conjunction with our previous observation that TNA replication on dG-containing templates leads to an abundance of G→C mutations in the cDNA product,¹ indicates that Terminator misincorporates tGTP monomers opposite dG. While this could be caused by deformation of the duplex in the active site or partial stacking of G in the absence of hydrogen bonds, we postulate that tGTP forms a tG:dG Hoogsteen base pair in the enzyme active site. Hoogsteen base pairing between guanine residues occurs when a guanine base in one strand adopts a *syn* conformation relative to the sugar moiety, which allows the second guanine residue to form two hydrogen bonds on the

major groove face of the first guanine base (Figure 2C).⁶ Crystallographic analysis of a Watson–Crick DNA duplex containing an internal G:G base pair reveals that a G:G Hoogsteen base pair is structurally similar to a G:C base pair.⁷ This structural similarity is one reason why many polymerases (natural and engineered) struggle to read through templates that are rich in G nucleotides.⁸

To test the hypothesis that G:G Hoogsteen base pairing was responsible for the chain termination events that occur when TNA polymers are synthesized on unbiased DNA libraries (L3), we constructed two pools of single-stranded DNA molecules. The first pool contained an unbiased mixture of all four natural nucleobases (A,C,T,G). The second pool was identical in length and composition to the first pool but contained the unnatural base analogue 7dG in place of natural dG. The pools were constructed by PCR using appropriate sets of nucleotide triphosphates and a PEG-modified PCR primer to enable separation of the strands by denaturing PAGE.

Preliminary analysis of the DNA products produced by PCR suggested that the 7dG reaction generated very little product when compared to the all-natural dNTPs reaction (Figure S1). Although we initially interpreted this result to mean that 7dG was a poor substrate for deep vent (exo-) DNA polymerase, further analysis revealed that 7dG quenches the fluorescence of UV-excitable dyes. In fact, only when the DNA was labeled with a fluorescent primer or stained with SYTO 60, a dye that excites in the IR range, were the DNA products from both reactions observed in equal abundance. This unanticipated phenomenon, though not entirely unprecedented, is a reminder of how the chemical and physical properties of modified DNA can differ from those of natural genetic polymers.⁹

Using the dG- and 7dG-containing templates, we performed a primer extension assay for 6 h at 55 °C to compare the efficiency of TNA synthesis on the two pools of DNA molecules. As illustrated in Figure 3, Terminator DNA

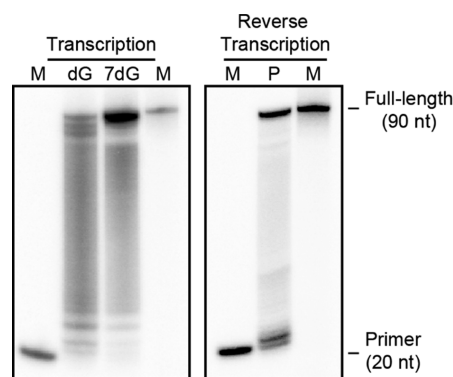


Figure 3. Transcription and reverse transcription of a four-nucleotide TNA library. PAGE analysis reveals that TNA transcription by Terminator DNA polymerase (left panel) and reverse transcription by Superscript II (right panel) proceed with high efficiency on DNA templates that contain 7-deazaguanine in place of the natural guanine base.

polymerase was unable to copy natural DNA templates into TNA, as evidenced by the absence of any significant full-length product and the high degree of banding on the gel. However, efficient synthesis was observed for the primer extension reaction performed on the pool of 7dG-containing DNA templates, which were designed to promote TNA synthesis by inhibiting the formation of G:G mispairs. In this case, the

polymerase was able to extend the DNA primer by 70 sequential residues to produce a significant amount of full-length product. This result provides the first example of an unbiased TNA library constructed by polymerase-mediated primer extension. It also confirms the hypothesis that G:G mispairing in the enzyme active site was responsible for polymerase stalling on natural DNA templates.

Because a complete replication cycle requires copying the pool of TNA strands back into DNA, we tested the ability for SSII to reverse-transcribe the TNA back into DNA. Full-length TNA strands isolated from the transcription of 7dG-containing templates were purified by denaturing PAGE and incubated with SSII for 2 h at 55 °C. Consistent with our previous analysis of a three-nucleotide library,¹ primer extension analysis revealed that the four-nucleotide library was an efficient template for reverse transcription (Figure 3).

To ensure that Therminator and SSII produced an accurate copy of each template, we measured the fidelity of TNA replication by sequencing the cDNA product. This assay measures the overall fidelity of replication, which includes the transcription of DNA into TNA, followed by the reverse transcription of TNA back into DNA (Figure 4). Several controls were used to eliminate possible contamination by the starting DNA template. First, we showed that PAGE-purified

TNA could not be amplified by PCR unless it was reverse-transcribed into cDNA (Figure S2). Second, the primer–template complex used for TNA synthesis was designed to contain a T:T mismatch that would produce a T:A watermark in the cDNA product. Only those sequences that were transcribed into TNA and reverse-transcribed back into cDNA would carry the watermark. With these controls in place, we sequenced 850 nucleotide positions from cDNA isolated after TNA transcription and reverse transcription. Analysis of the DNA sequences yielded an overall fidelity of replication of >99% (Figure S3). This level of fidelity is similar to what we observed for the replication of three-nucleotide templates and supports the robust transcription of TNA polymers on 7dG-containing DNA templates.¹

To determine if 7dG overcomes polymerase stalling by increasing the rate of tCTP incorporation, we used pre-steady-state kinetics to measure the rate constants for incorporation of dCTP or tCTP opposite a templating dG or 7dG residue. Briefly, a pre-incubated solution of Therminator DNA polymerase (100 nM) and a 5'-radiolabeled DNA primer–template substrate (20 nM) in a reaction buffer at 37 °C was rapidly mixed with dCTP or tCTP (100 μM) for various times before being quenched with EDTA. The data were fit to a single-exponential equation, $[\text{product}] = A[1 - \exp(-k_{\text{obs}}t)]$, to yield k_{obs} , the observed rate constant of nucleotide incorporation (Table 1). Notably, the rate constant of correct

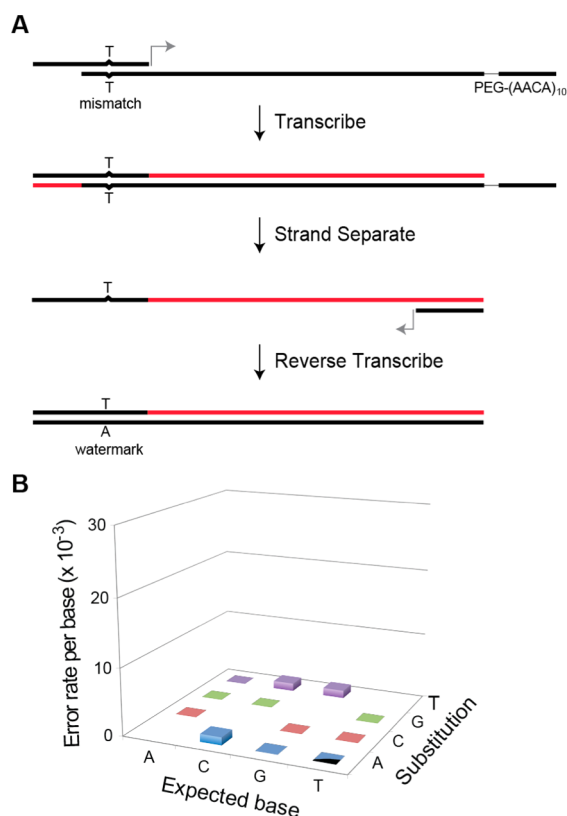


Figure 4. Fidelity of TNA replication using 7-deazaguanine. (A) Schematic representation of the transcription and reverse transcription process used to evaluate the fidelity of TNA replication. DNA is shown in black, and TNA is shown in red. The primer–template complex contains a T:T mismatch, which produces a T-to-A transversion in the cDNA strand. The transversion represents a watermark to ensure that the sequenced DNA was produced by TNA transcription and reverse transcription. (B) The mutation profile reveals that TNA transcription and reverse transcription proceeded with an error rate of 3.8×10^{-3} (>99% fidelity).

Table 1. Pre-Steady-State Kinetic Parameters for Therminator-Mediated Single-Nucleotide Extension

5'-d-CGCAGCCGTCCAACCAACTCA 3'-d-GCGTCGGCAGGTTGGTTGAGT-X-GCAGCTAGGTTACGGCAGG		
NTP	X	k_{obs} (s^{-1})
tCTP	dG	0.9 ± 0.2
tCTP	7dG	3.8 ± 0.5
tGTP	7dG	0.62 ± 0.09
tGTP	dG	0.68 ± 0.06
dCTP	dG	4.2 ± 0.9
dCTP	7dG	5.0 ± 0.7
dGTP	dG	0.39 ± 0.01
dGTP	7dG	0.40 ± 0.06

dCTP or incorrect dGTP incorporation was not significantly affected by substituting the templating dG with 7dG. This indicates that the unnatural templating nucleotide 7dG did not significantly alter the interactions between the polymerase–DNA complex and incoming nucleotide at the enzyme active site. In contrast, the rate constant of tCTP incorporation increased by 4.2-fold opposite the templating 7dG relative to dG and is comparable to the rate constant for normal dCTP incorporation opposite dG (Table 1). Thus, tCTP is able to pair with 7dG nearly as well as a natural dCTP pairs with dG. Furthermore, opposite the templating 7dG, the rate constant for tGTP misincorporation is 6.1-fold lower than that for correct tCTP incorporation, which is comparable the 10.8-fold difference between dCTP:dG and dGTP:dG incorporations. Thus, the switching of dG to 7dG in the DNA template not only solves the polymerase stalling problem but also maintains relatively high DNA→TNA transcription fidelity.

The ability to code and decode sequence-defined genetic polymers, like TNA, provides access to many non-biological applications that will benefit materials science, nanotechnology, and molecular medicine.¹⁰ For example, TNA is highly resistant

to nuclease degradation, making it a stable scaffold for future diagnostic and therapeutic applications. In addition, because TNA has the ability to evolve in response to imposed selection constraints, it could also be used to enhance our understanding of why nature chose RNA as the molecular basis of life's genetic material.¹¹ Previous work in this area has been limited by the absence of polymerases that could be used to study alternative chemistries of life. As this paradigm is now changing, we may soon discover that many different types of genetic polymers exhibit the characteristic signatures of heredity and evolution—two important hallmarks of life.^{12,2a}

In summary, we have developed a method for replicating unbiased pools of TNA polymers that uses 7-deazaguanine in the DNA template to suppress G:G mispairing during TNA transcription with Terminator DNA polymerase. Characterization of the replication cycle reveals that the TNA synthesis proceeds with high efficiency and high overall sequence fidelity. We suggest that this approach could be used to explore the functional properties of TNA by *in vitro* selection.

■ ASSOCIATED CONTENT

📄 Supporting Information

Additional information and experimental methods. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*john.chaput@asu.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank other members of the Chaput laboratory for helpful comments and suggestions. This work was supported by National Science Foundation grants 1304583 to J.C.C. and MCB-0960961 to Z.S. W.J.Z. is supported by graduate training grant T32 GM008512 from National Institutes of Health.

■ REFERENCES

- (1) Yu, H.; Zhang, S.; Dunn, M. R.; Chaput, J. C. *J. Am. Chem. Soc.* **2013**, *135*, 3583.
- (2) For examples of polymerase-mediated TNA synthesis, see:
 - (a) Yu, H.; Zhang, S.; Chaput, J. C. *Nat. Chem.* **2012**, *4*, 183.
 - (b) Horhota, A.; Zou, K.; Ichida, J. K.; Yu, B.; McLaughlin, L. W.; Szostak, J. W.; Chaput, J. C. *J. Am. Chem. Soc.* **2005**, *127*, 7427.
 - (c) Chaput, J. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 9274.
 - (d) Kempeneers, V.; Vastmans, K.; Rozenski, J.; Herdewijn, P. *Nucleic Acids Res.* **2003**, *31*, 6221.
 - (e) Ichida, J. K.; Zou, K.; Horhota, A.; Yu, B.; McLaughlin, L. W.; Szostak, J. W. *J. Am. Chem. Soc.* **2005**, *127*, 2802.
 - (f) Ichida, J. K.; Horhota, A.; Zou, K.; McLaughlin, L. W.; Szostak, J. W. *Nucleic Acids Res.* **2005**, *33*, 5219.
 - (g) Zhang, S.; Chaput, J. C. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 1447.
- (3) For papers on Terminator DNA polymerase, see:
 - (a) Gardner, A. F.; Jack, W. E. *Nucleic Acids Res.* **2002**, *30*, 605.
 - (b) Gardner, A. F.; Jack, W. E. *Nucleic Acids Res.* **1999**, *27*, 2545.
- (4) Arezi, B.; Hogrefe, H. *Nucleic Acids Res.* **2009**, *37*, 473.
- (5) (a) Zhang, S.; Chaput, J. C. *Curr. Protein Nucleic Acid Chem.* **2012**, *4*, No. 4.51.1. (b) Zhang, S.; Yu, H.; Chaput, J. C. *Curr. Protein Nucleic Acid Chem.* **2013**, *4*, No. 4.54.1.
- (6) Cantor, C. R. *Principles of Nucleic Acid Chemistry*; Springer-Verlag: New York, 1984; p 120.
- (7) Skelly, J. V.; Edwards, K. J.; Jenkins, T. C.; Neidle, S. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 804.
- (8) Staiger, N.; Marx, A. *ChemBioChem* **2010**, *11*, 1963.
- (9) (a) Kelley, S. O.; Barton, J. K. *Chem. Biol.* **1998**, *5*, 413. (b) Innis, M. A.; Myambo, K. B.; Gelfand, D. H.; Brow, M. A. D. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 9436.
- (10) For reviews, see:
 - (a) Lutz, J.-F.; Ouchi, M.; Liu, D. R.; Sawamoto, M. *Science* **2013**, *341*, 628.
 - (b) Chaput, J. C.; Yu, H.; Zhang, S. *Chem. Biol.* **2012**, *19*, 1360.
- (11) (a) Eschenmoser, A. E. *Science* **1999**, *284*, 2118. (b) Joyce, G. F. *Nature* **2002**, *418*, 214.
- (12) Pinheiro, V. B.; Taylor, A. I.; Cozens, C.; Abramov, M.; Renders, M.; Zhang, S.; Chaput, J. C.; Wengel, J.; Peak-Chew, S.-Y.; McLaughlin, S. H.; Herewihjn, P.; Holliger, P.; et al. *Science* **2012**, *336*, 341.