Developing Engineered Polymerases for Practical Applications in Synthetic Biology

by

Matthew Ryan Dunn

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved September 2015 by the
Graduate Supervisory Committee:

John C. Chaput, Chair
Joshua LaBaer
Douglas Lake
Marco Mangone

ARIZONA STATE UNIVERSITY

December 2015

ABSTRACT

Advances in chemical synthesis have enabled new lines of research with unnatural genetic polymers whose modified bases or sugar-phosphate backbones have potential therapeutic and biotechnological applications. Maximizing the potential of these synthetic genetic systems requires inventing new molecular biology tools that can both generate and faithfully replicate unnatural polymers of significant length. Threose nucleic acid (TNA) has received significant attention as a complete replication system has been developed by engineering natural polymerases to broaden their substrate specificity. The system, however, suffers from a high mutational load reducing its utility. This thesis will cover the development of two new polymerases capable of transcribing and reverse transcribing TNA polymers with high efficiency and fidelity. The polymerases are identified using a new strategy wherein gain-of-function mutations are sampled in homologous protein architectures leading to subtle optimization of protein function. The new replication system has a fidelity that supports the propagation of genetic information enabling in vitro selection of functional TNA molecules. TNA aptamers to human alpha-thrombin are identified and demonstrated to have superior stability compared to DNA and RNA in biologically relevant conditions. This is the first demonstration that functional TNA molecules have potential in biotechnology and molecular medicine.

DEDICATION

This work is dedicated to my mom who instilled in me a desire to learn and to always better myself and to my wife for her boundless support through my degree.

ACKNOWLEDGMENTS

I would like to thank my supervisor, Dr. John Chaput, for providing a research environment that has allowed me to produce this work, his helpful insight when progress seemed lost, and training in scientific writing. I also owe significant gratitude to the Chaput lab members, especially Dr. Andrew Larsen, who were always present to discuss and troubleshoot problems, assist with lab work, and discuss life. I'd also like to thank my undergraduate mentees for their help and for their patience and feedback as I developed into a better mentor.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

EMERGENCE OF SYNTHETIC GENETICS

Synthetic Genetics

DNA and RNA are nucleic acid polymers that comprise the molecular basis of life through their unique ability to store genetic information. Once thought to be limited to storage, advances over the last 30 years have expanded their functional repertoire.[1,2] These findings revolutionized how researchers thought about nucleic acids inciting a movement to not only better understand nucleic acid properties and discover the chemical and physical basis of molecular information encoding and folding, but to also expand these chemical properties by systematically varying nucleobase, sugar, and backbone moieties.[3–22] This advance gave rise to a collection of unnatural genetic polymers (**Figure 1.1**), often called xenonucleic acids (XNA), and the field of synthetic genetics. At the heart of synthetic genetics is to understand, exploit, and redesign genetic systems for new applications in analogy to what synthetic biology does on cellular and organismal systems.[23–26]

Each XNA polymer has unique chemical, physical, and biological properties that can be harnessed to derive novel applications. For instance, some XNA systems have shown to be unrecognizable by natural enzymes.[4,27–40] While this phenomenon greatly increases their potential in molecular medicine and pharmaceutical development, this property challenges the merger of XNA chemistry into biological systems. It has become increasingly evident when attempting to replicate unnatural genetic polymers in the lab as the extent of chemical dissimilarity between XNA and natural systems often bars the use of natural polymerases for propagating genetic information stored in XNA polymers.[41–44]

New advances in polymerase engineering and screening techniques are beginning to overcome this hurdle and enable the replication of non-natural genetic polymers making it possible for the first time to explore the sequence space of synthetic genetic polymers.

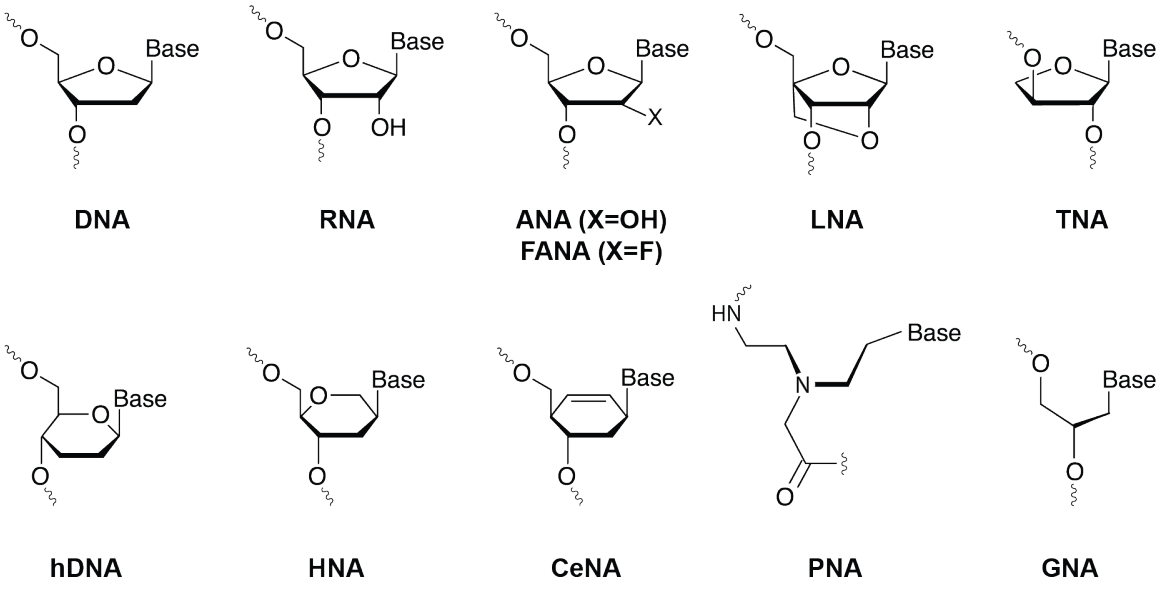

Figure 1.1. Chemical Structure of Sugar-Modified XNA Molecules. Synthetic genetics includes a range of alternative genetic systems each of which have unique chemical and biological properties. This list is limited to genetic systems with sugar-modified backbones.

Threose Nucleic Acid

Developed by Eschenmoser and colleagues in a screen to identify unnatural genetic systems, α-L-threofuranosyl-(3'→2') nucleic acid (TNA) has emerged as one of the first genetic polymers with broad application in synthetic genetics (**Figure 1.2**).[45,46] TNA, like RNA, is composed of a repeating sugar phosphate backbone where the natural five-carbon ribose sugar is replaced with an unnatural four-carbon threose sugar. TNA has attracted significant attention as an RNA analog for biomedicine because TNA polymers maintain stable Watson-crick duplexes opposite complementary strands of

TNA as well as DNA and RNA (**Figure 1.3B and 1.4**), are resistant to nuclease degradation (**Figure 1.3A**), and have been shown to fold into functional structures capable of ligand binding.[27,42,45,47]

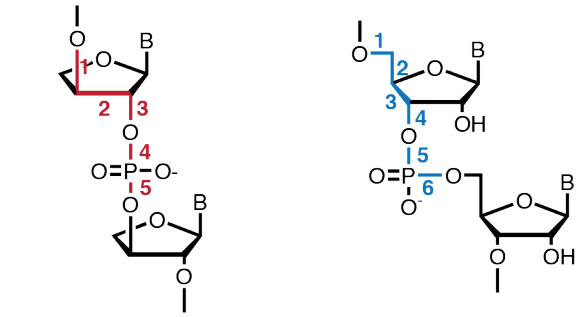

Figure 1.2. Constitutional Structure for the Linearized Backbone of Threose Nucleic Acid (TNA) and Ribose Nucleic Acid (RNA). TNA is an unnatural genetic polymer composed of repeating α-L-threose sugars vicinally connected by 3'→2'-phosphodiester linkages. The backbone bonds are numbered to demonstrate the shorter backbone repeat unit.

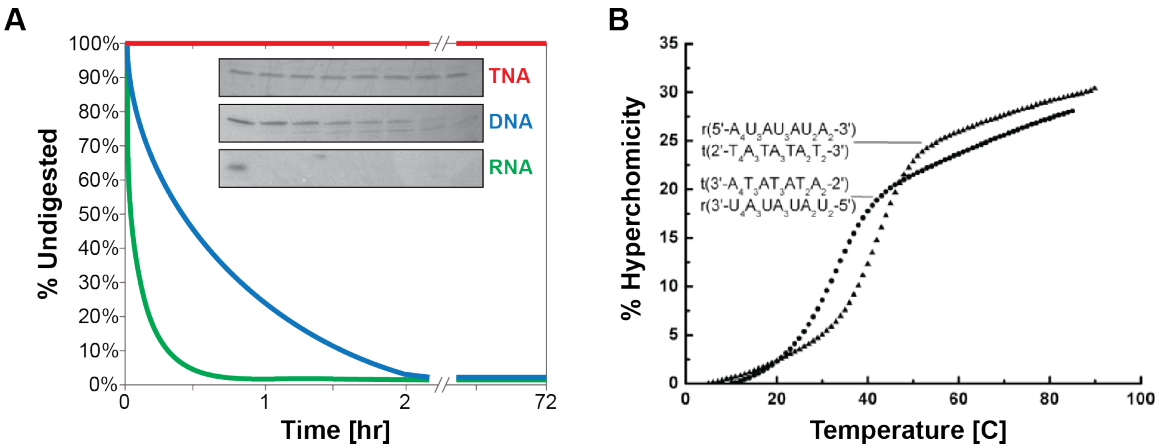

Figure 1.3. Nuclease Stability and Base Pairing of TNA. (A) A time course of the stability of TNA, DNA, and RNA polymers in the presence of both DNase and RNase demonstrates TNA's nuclease resistance.[27] RNA digests were sampled over 3 minutes in 30 second increments, DNA digests over 1 hour in 5 minute increments, and TNA digests over 3 days in 12 hour increments. (B) The melting curves for RNA-TNA duplexes demonstrates a cooperative melting profile suggesting the duplexes for stable Watson-crick duplexes.[48,49]

The solution structure of a self-complementary TNA octamer reveals that TNA-TNA duplexes adopt an A-form helical structure that closely approximates the helical geometry of A-form RNA (**Figure 1.4**).[50] This observation is surprising given that TNA polymers have a backbone repeat unit that is one atom shorter than that found in DNA and RNA (**Figure 1.2**). Shedding insight into the mechanism of pairing, the crystal structure of a DNA duplex containing a single TNA demonstrates that TNA stretches its 3'→2' vicinal linkage to match the natural 5'→3' trans-diaxial positioning of DNA and RNA.[51]

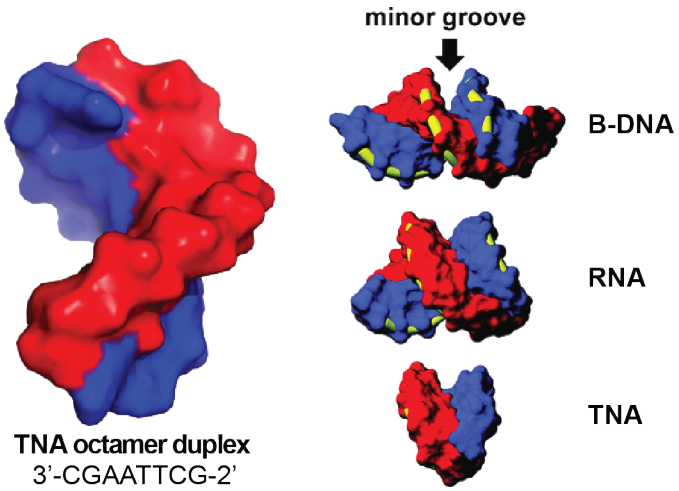

Figure 1.4. Structural Comparison of DNA, RNA, and TNA Duplexes. NMR structure of TNA compared to the structures of DNA and RNA. TNA adopts a confirmation similar to A-form RNA.[50]

In addition to possible applications in biomedicine, TNA has been the focus of numerous studies assessing its feasibility as an ancestral genetic polymer capable of propagating and evolving early genetic information. The TNA world, which arose through a series of discrete chemical steps in a pool of other primordial informational polymers, would eventually give rise to modern RNA-based systems based on a DNA genome and protein enzymes.[52,53] TNA is an attractive pre-RNA candidate due to its chemical simplicity under prebiotic conditions and its ability to exchange genetic information with DNA and RNA.[54] However, TNA molecules have yet to demonstrate catalytic activity, which is a hallmark of pre-RNA polymers.

Further research to develop biologically relevant molecules for both biomedicine and origins of life will require the use of in vitro selection to identify functional TNA molecules from diverse pools.

Overview of In Vitro Selection

For more than two decades, in vitro selection experiments have been performed on DNA and RNA. In vitro selection, also known as SELEX (systematic evolution of ligands by exponential enrichment) is a powerful molecular biology technique commonly used to evolve nucleic acid molecules that can bind to a desired target or catalyze a specific chemical reaction (**Figure 1.5**).[1,55] Starting from a large library of nucleic acid molecules (typically 10^{15} different sequences), functional molecules are isolated though iterative rounds of enrichment and amplification.[2,56] In a manner analogous to natural selection, molecules that survive the selection increase in abundance, while molecules that lack the desired function go extinct. There have been a number of methods for separating functional from non-functional molecules including affinity chromatography, membrane partitioning, gel electrophoresis, capillary electrophoresis, cell and viral-based strategies, and fluorescence-based sorting. By increasing the selective pressure, for example by reducing the incubation time or substrate concentration, it is possible to identify individual molecules with superior functional properties if present in the original population. In vitro selection has been used to study basic questions in molecular evolution as well as develop molecules with practical applications in molecular medicine. In addition to Macugen, which is an RNA aptamer approved by the FDA for the treatment of macular degeneration, many aptamers are now in clinical trials in the U.S. and Europe.[57,58] Moving beyond natural genetic polymers requires polymerases that can transcribe, reverse transcribe, and amplify the nucleic acid sequences encoded in unnatural genetic polymers.

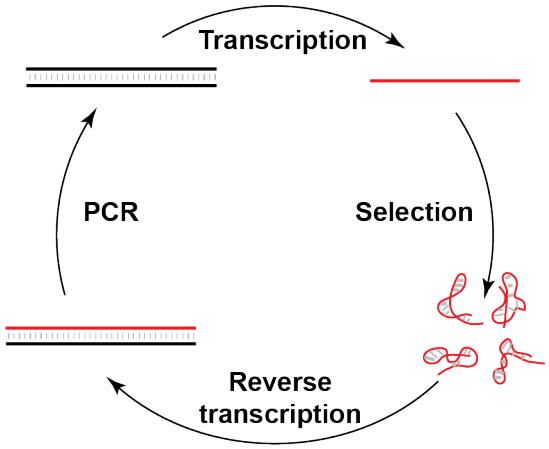

Figure 1.5. Overview of In Vitro Selection for functional RNA molecules. A diverse library of DNA molecules (black) is converted to RNA (red). The RNA pool is then separated based on functional activity. Surviving molecules are reverse transcribed back into DNA and amplified by PCR generating an enriched population of molecules ready for the next round of selection.

Identification of a Polymerase Capable of Generating TNA Polymers

In order to enable in vitro selection of functional TNA molecules, a polymerase capable of generating long TNA polymers was needed. In a screen to identify DNA-dependent TNA polymerases, an engineered form of the family B DNA polymerase from *Thermococcus* sp. 9°N was discovered. This enzyme, bearing the exonuclease silencing mutations (D141A and E143A) and the A485L mutation, known commercially as Therminator DNA polymerase, was shown to extend a significant length of TNA under mutagenic conditions (**Figure 1.6**).[59–62] Therminator DNA polymerase can copy three-nucleotide DNA libraries into TNA with modest efficiency and fidelity; however, the polymerase stalls when it encounters templating dG residues.[42]

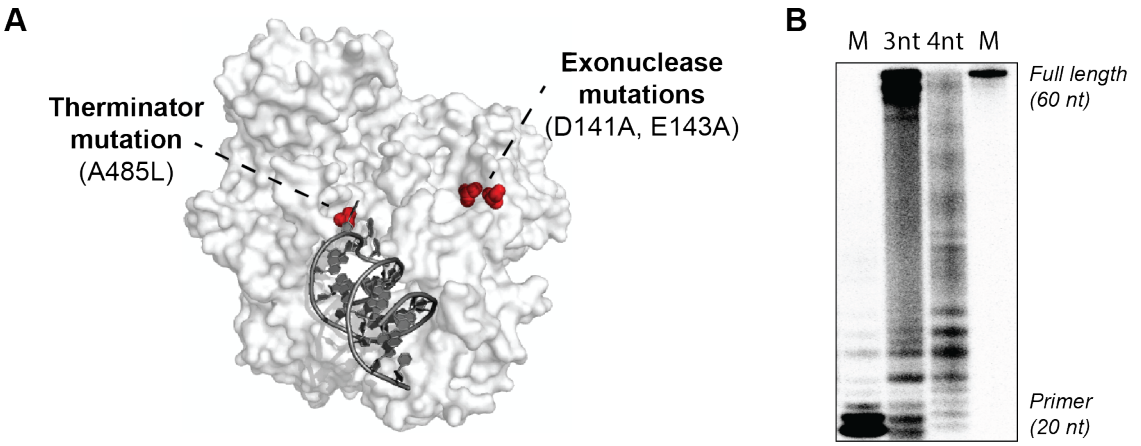

Figure 1.6. Structure and TNA Polymerization of Therminator DNA Polymerase. (A) 3D structure of *Thermococcus* sp. 9N DNA polymerase highlighting the exonuclease silencing and A485L mutations (PDB: 4K8X).[63] (B) Therminator-mediated TNA transcription efficiency on three nucleotide and four nucleotide libraries. Radioactive products were analyzed by phopshorimaging.

In Vitro Selection of a TNA Aptamer

Noticing this problem and motivated to determine if TNA can support functional activity, an in vitro selection strategy was developed to select for structures with discrete ligand-binding properties using a library devoid of G nucleobases.[64,65] Since a TNA-dependent DNA polymerases was not available to reverse transcribe TNA polymers back into DNA for amplification using polymerase chain reaction (PCR), a one-enzyme selection strategy was devised that establishes a genotype-phenotype link by displaying individual TNA molecules (phenotype) on their encoding double-stranded DNA sequence (genotype).[65] In this strategy, termed DNA display (**Figure 1.7A**), a self-priming library is extended with TNA. The TNA is then liberated from its template by strand displacement primer extension generating a single-stranded TNA pool fused on its 3' end to its encoding double-stranded DNA template. After selection, functional TNA

molecules can be propagated by amplifying the attached DNA sequence by PCR. Using the target human α-thrombin, TNA aptamers were identified with high specificity and affinities as high as 200 nM (**Figure 1.7B**).[42] This was the first demonstration that function is not limited to natural polymers.

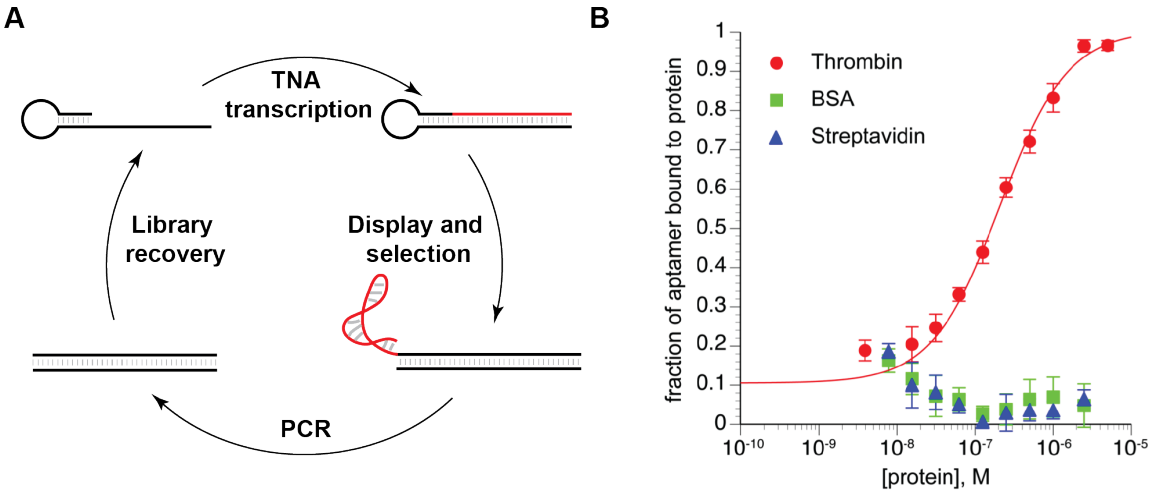

Figure 1.7. Strategy for in vitro TNA evolution. (A) A diverse library of DNA molecules is ligated to a hairpin primer. The hairpin primer is then extended with TNA triphosphates generating a chimeric DNA-TNA duplex. A second primer is then annealed to the hairpin and a strand displacing polymerase generates a double stranded DNA liberating the bound TNA molecule. Functional TNA molecules are then separated and amplified by PCR to generate an enriched pool of DNA molecules. The selection cycle is then repeated. (B) Binding isotherms for an anti-thrombin TNA aptamer with thrombin, streptavidin, and BSA substrates.[42]

Identification of TNA Reverse Transcriptases

DNA display, while capable of enabling structures, is intrinsically limited by the utilization of the transcription primer to physically link the TNA to the encoding DNA. In higher-level selection strategies, chemically functionalized primers are needed to introduce different moieties. Therefore, a two-enzyme selection strategy, where one enzyme generates TNA from a DNA template and the other regenerates a DNA pool from TNA, frees the primer for other functionalities.

To identify a polymerase capable of reverse transcribing TNA, a screen was performed that identified an engineered version of MMLV reverse transcriptase known commercially as Superscript II that could reliably reverse transcribe TNA under mutagenic conditions (**Figure 1.8**).[66] Superscript II is capable of reverse transcribing a significant portion of a three nucleotide TNA library into complementary DNA.

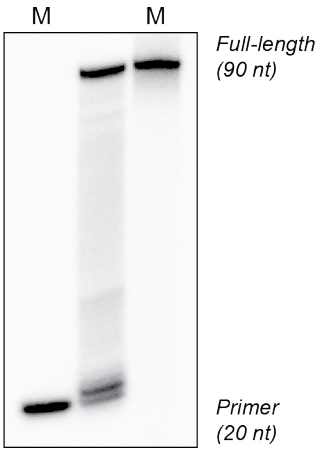

Figure 1.8. TNA Reverse Transcription by Superscript II. Superscript II is capable of reverse transcribing a significant portion of a three nucleotide TNA library into complementary DNA.[27] M denotes primer and full-length markers. Radioactive products were visualized by phosphorimaging.

The development of a complete replication system (DNA to TNA to DNA) allowed for the first time the measurement of the aggregate fidelity of TNA replication.[27] DNA sequences composed of a three-letter genetic alphabet (A, T, C) are converted to and from TNA with high efficiency, and individual sequences replicate with >99% fidelity generating fewer than 3 errors per 1000 bases incorporated (**Figure 2.1**). However, additional biases are observed for DNA sequences composed of a four-letter genetic alphabet (A, T, C, G). The addition of templating dG results in a mutation rate of greater than 80 misincorporations per 100 bases.[27]

Summary

Considering the structural differences between DNA and TNA it is outstanding that only minor modifications to natural polymerases enable a high level of processivity and fidelity. However, this level of fidelity is significantly lower than that seen in natural polymerases and poses a potential problem for many applications of TNA including in vitro selection. Since in vitro selection requires an enrichment of functional molecules from a highly diverse pool, high levels of mutation would effectively neutralize the enrichment of genetic information harboring functional phenotypes. The sole use of three-nucleotide systems is possible but the phenotypic potential of four-nucleotide systems is greater and therefore desired.

CHAPTER 2

DNA POLYMERASE-MEDIATED SYNTHESIS OF UNBIASED THREOSE NUCLEIC ACID (TNA) POLYMERS REQUIRES 7-DEAZAGUANINE TO SUPPRESS G:G MISPAIRING DURING TNA TRANSCRIPTION

Publication Note

This research was originally published in the *Journal of the American Chemical Society*. Dunn, M.R., Larsen, A.C., Zahurancik, W.J., Fahmi, N.E., Meyers, M.M., Suo, Z., Chaput, J.C. (2015) DNA Polymerase-Mediated Synthesis of Unbiased Threose Nucleic Acid (TNA) Polymers Requires 7-Deazaguanine To Suppress G:G Mispairing during TNA Transcription. *J. Am. Chem. Soc.*, 2015, 137(12), pp 4014–4017.

Contribution Statement

M.R.D. and M.M. performed all TNA work. A.C.L. improved 7dG utilization. W.J.F. synthesized TNA triphosphates. W.Z. and Z.S. performed kinetic analyses. M.R.D. and J.C.C. designed the study and wrote the paper. All authors discussed results and commented on the manuscript.

Introduction

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 Archaean replicative DNA polymerase 9°N, known commercially as Therminator 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 mispairing in the enzyme active site during TNA transcription. 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 1.2**).[27] Under appropriate conditions, an engineered form of the Archaean replicative DNA polymerase 9°N, known commercially as Therminator DNA polymerase (New England BioLabs, Inc.), can copy sequence-defined DNA templates into TNA through the sequential addition of TNA nucleotide triphosphates.[42,59–62,65,67] 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.[68] Amplification of the cDNA using the 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 (**Figure 2.1A**).[27] 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.[42,59–62,65,67,69,70] 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.[27] The low fidelity of TNA replication on four-nucleotide templates (**Figure 2.1B**) can be overcome by replacing tCTP in the transcription mixture with dCTP (**Figure 2.1C**); however, this change leads to a mixed-backbone TNA–DNA heteropolymer that may not be suitable for some applications.[27] 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.

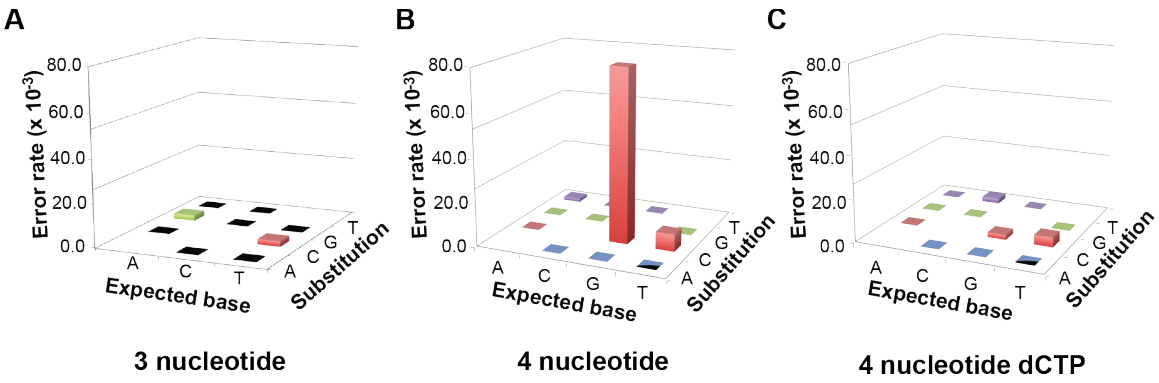

Figure 2.1. TNA Transcription Fidelity of Therminator DNA Polymerase on Templates of Varying Complexity. TNA transcription fidelity on (A) three nucleotide templates (A, T, C), (B) four nucleotide templates (A, T, C, G), and (C) four nucleotide templates with tCTP replaced by natural dCTP.

Results

Recognizing the limitations of G-rich templates, we decided to evaluate the ability of Therminator 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 2.2A**). Therminator was challenged to extend a DNA primer annealed to a DNA template using TNA triphosphates (tNTPs) that were obtained by chemical synthesis.[71,72] 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.

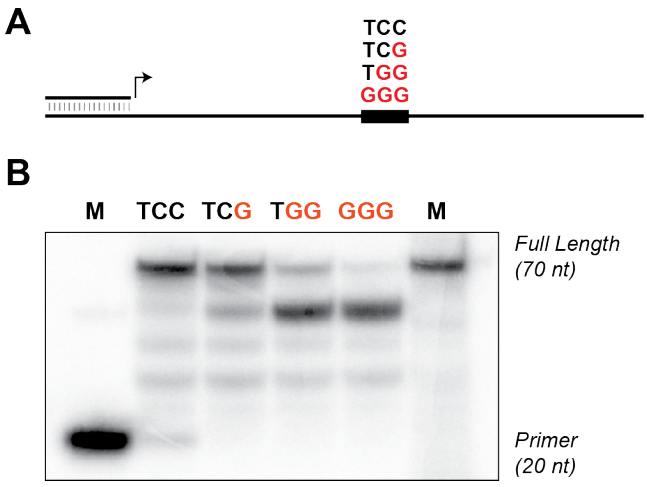

Figure 2.2. Sequence Specificity of Therminator DNA Polymerase. (A) Schematic of the primer extension assay used to evaluate TNA synthesis on dG-containing DNA templates. (B) Primer extension results for DNA templates that contain consecutive G-residues. Radioactive products were visualized by phosphorimaging.

Analysis of the resulting primer extension assay products indicates that Therminator is able to generate full-length TNA product when 0 or 1 dG residue is present in the DNA template (**Figure 2.2B**). 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, indicates that Therminator misincorporates tGTP monomers opposite dG.[27] 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 2.3**).[73] 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.[74] This structural similarity is one reason why many polymerases (natural and engineered) struggle to read through templates that are rich in G nucleotides.[75]

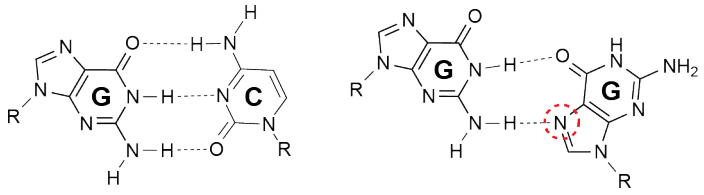

Figure 2.3. Proposed Hoogsteen G:G Mispairing Mode. Molecular structure of a normal G:C Watson−Crick base pair and a G:G Hoogsteen base pair. The analogue of 7-deazaguanine replaces the N-7 nitrogen with a methylene group (red circle).

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 2.4**). 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.[76,77]

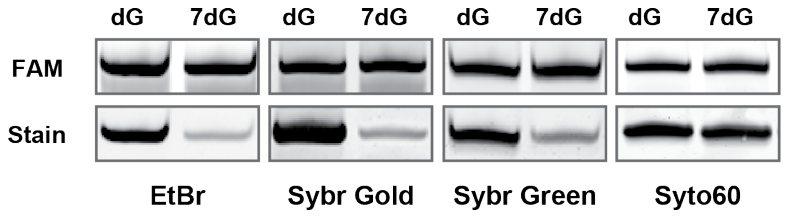

Figure 2.4. Evidence of Fluorescence Quenching by 7dG-Containing DNA. PCR amplified DNA carrying a fluorescein-label was analyzed on four different denaturing PAGE gels. Each gel was imaged by fluorescence and separately stained with ethidium bromide (EtBr), Sybr Gold, Sybr Green, or Syto60. Fluorescence imaging (top) demonstrates that an equivalent amount of dG- and 7dG-containing DNA is present in each lane. When the same gels are stained with ethidium bromide, Sybr Gold, Sybr Green, and Syto60 (bottom) only the Syto60 stain shows equal amounts of PCR product for the dG and 7dG amplicons. Reduced band intensity for the EtBr, Sybr Gold, and Sybr Green stains is due to quenching of the fluorescent dyes by 7dG residues in the DNA. The Syto60 stain does not lead to fluorescence quenching because it emits in the infrared region.

Using the dG- and 7dG-containing templates, we performed a primer extension assay for 6 hours at 55°C to compare the efficiency of TNA synthesis on the two pools of DNA molecules. As illustrated in **Figure 2.5**, Therminator DNA 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.

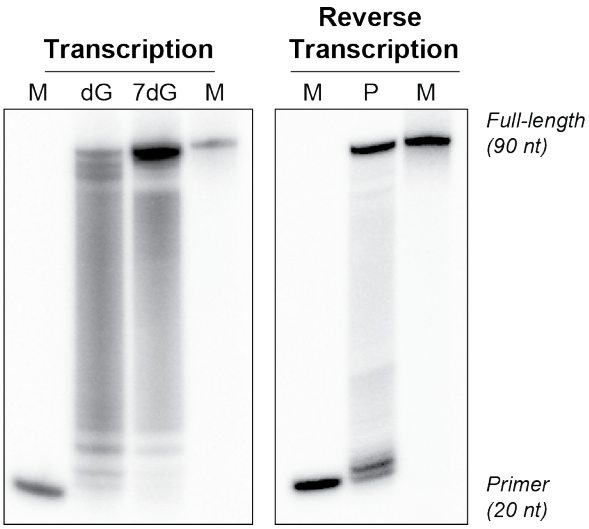

Figure 2.5. Transcription and Reverse Transcription of a Four-Nucleotide TNA Library. PAGE analysis reveals that TNA transcription by Therminator DNA polymerase (left panel) and reverse transcription by Superscript II (right panel) proceed with high efficiency on DNA templates that contain 7-deazaguanine in place of the natural guanine base. Radioactive products were visualized by phosphorimaging.

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 hours at 55°C. Consistent with our previous analysis of a three-nucleotide library, primer extension analysis revealed that the four-nucleotide library was an efficient template for reverse transcription (**Figure 2.5**).[27]

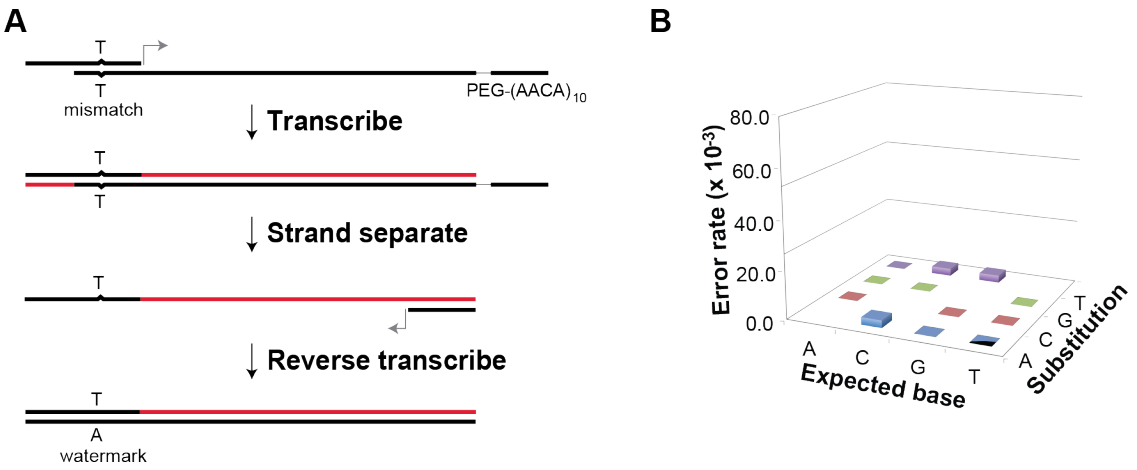

Figure 2.6. Fidelity of TNA Replication Using 7-Deazaguanine. (A) Schematic representation of the transcription and reverse transcription process used to evaluate the fidelity of TNA replication. DNA is shown in black, TNA is shown in red. The primer-template complex contains a T-T mismatch, which produces 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 proceeds with an error rate of 3.8 x 10^{-3} (>99% fidelity).

To ensure that Therminator and Superscript II 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 2.6A**). 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 2.7**). 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 2.6B and 2.8**). 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 (**Table 2.1**).[27]

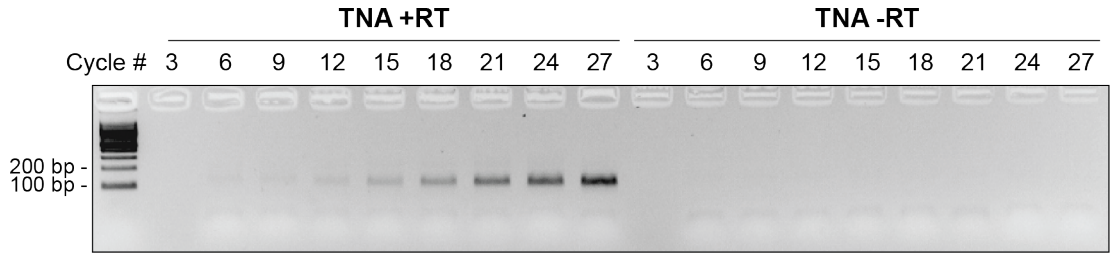

Figure 2.7. PCR Assay for DNA Contaminants. PCR analysis confirms that the TNA sample is devoid of DNA contaminants that could interfere with the TNA fidelity assay. DNA amplicons are only observed when the TNA sample is reverse transcribed into cDNA. Products were visualized by staining with ethidium bromide.

Table 2.1. Comparison of the TNA Replication Fidelity on Different Templates Using Therminator DNA Polymerase and Superscript II Reverse Transcriptase.

Template	Bases Read	Total Error
3NT	1050	3.8×10^{-3}
4NT	1650	35.8×10^{-3}
4NT-7	850	3.5×10^{-3}

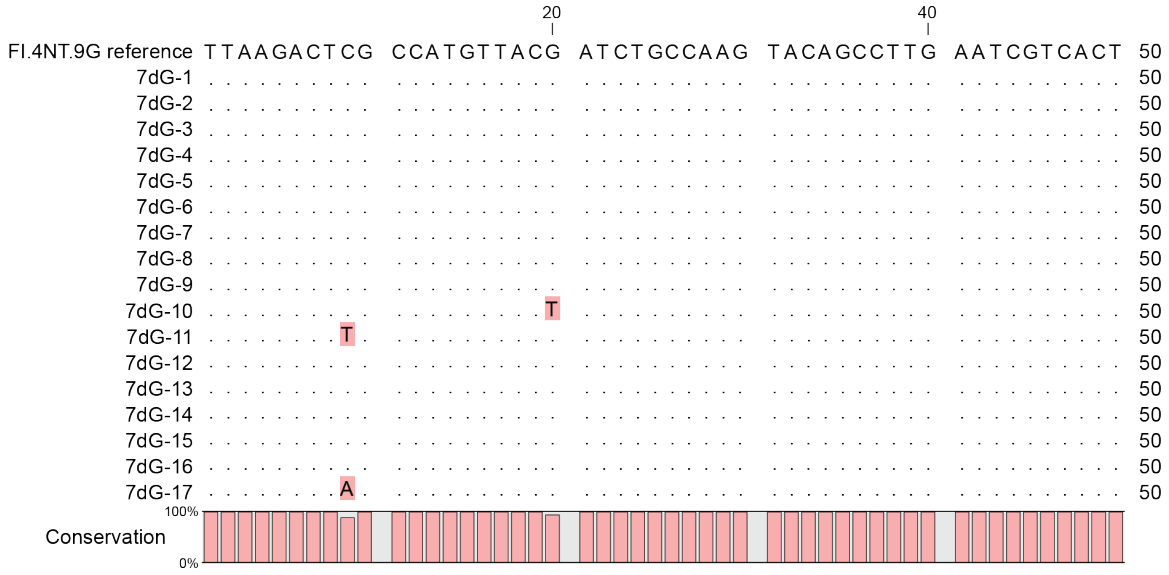

Figure 2.8. DNA Sequence Alignment Used to Assess TNA Replication Fidelity. An alignment of the sequence reads to the FI.4NT.9G reference sequence indicates that 3 mutations were observed in a total of 850-nt positions. This yields an overall fidelity of 99.6% for TNA replication using 7dG containing templates. Interestingly, no G to C transversions were observed in the data, demonstrating that 7dG inhibits G:G mispairing in the enzyme active site.

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, [product] = A [1 – exp(−k_{obs}t)], to yield k_{obs}, the observed rate constant of nucleotide incorporation (**Table 2.2**). Notably, the rate constant of correct 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 2.2**). 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.

Table 2.2. 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
dCTP	dG	4.2 ± 0.9
dCTP	7dG	5.0 ± 0.7
dGTP	dG	0.39 ± 0.01
dGTP	7dG	0.40 ± 0.06

Discussion

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.[46,78] 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.[48,79] 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.[41,42]

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 Therminator 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.

Experimental

General Information

DNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), purified by denaturing polyacrylamide gel electrophoresis, electroeluted, concentrated by ethanol precipitation, and quantified by UV absorbance. TNA triphosphates (tNTPs) were obtained through a multistep chemical synthesis as previously reported.[71,72] 7-deaza-dGTP was purchased from Roche (Mannheim, Germany). dNTPs were purchased from Sigma (St. Louis, MO). [γ-^{32}P]ATP was purchased from Perkin Elmer (Walktham, MA). 7-deaza-dG-CE Phosphoramidite was purchased from Glen Research (Sterling, VA). The enzymes T4 polynucleotide kinase, Therminator DNA polymerase, and exonuclease-deficient Deep Vent DNA polymerase (DV exo-) were purchased from New England Biolabs (Ipswich, MA). Optikinase was purchased from Affymetrix (Santa Clara, CA). SuperScript II reverse transcriptase, Sybr Gold, and Sybr Safe were purchased from Invitrogen (Grand Island, NY). CloneJET PCR cloning kit was purchased from Fermentas (Waltham, MA). Syto60 dye was purchased from LI-COR (Lincoln, NE).

DNA Sequences

The sequences below are written from 5'→3'. Primer and template sequences were synthesized by IDT, except for L3 and the 7dG-modified 41-mer template, which were synthesized in house. N = A:T:C:G = 1:1:1:0.5, X = dG or 7dG.

Primers	
PBS2	GACACTCGTATGCAGTAGCC
PBS2 FAM	/56-FAM/GACACTCGTATGCAGTAGCC
PBS2.mismatch	CTTTTAAGAACCGGACGAACGACACTCGTTTGCAGTAGC
PBS1	TGTCTACACGCAAGCTTACA
PBS1.PEG	AACAAACAAACAAACAAACAAACAAACAAACAA ACAAACA/iSp9/TGTCTACACGCAAGCTTACA
extra.primer	CTTTTAAGAACCGGACGAAC
21-mer primer	CGCAGCCGTCCAACCAACTCA
Templates	
4NT.0g	ACTATTCAACTTACAATCCTATCAACCTTATAATC CACTTGGCTACTGCATACGAGTGTC
4NT.1g	ACTATTCAACTTACAATCGTATCAACCTTATAATC CACTTGG CTACTGCATACGAGTGTC
4NT.2g	ACTATTCAACTTACAATGGTATCAACCTTATAATC CACTTGGCTACTGCATACGAGTGTC
4NT.3g	ACTATTCAACTTACAATGGGATCAACCTTATAATC CACTTGGCTACTGCATACGAGTGTC
FI.4NT.9G	TGTCTACACGCAAGCTTACATTAAGACTCGCCATG TTACGATCTGCCAAGTACAGCCTTGAATCGTCACTGGCT ACTGCATACGAGTGTC
L3 library:	TGTCTACACGCAAGCTTACA - N50 – GGCTACTGCATACGAGTGTC
41mer template:	GGACGGCATTGGATCGACGXTGAGTTGGTTGGACGGCTG CG

Imaging DNA Strands Containing 7dG Nucleotides in Place of dG

DNA strands that contain 7dG are known to quench the fluorescence of UV excitable dyes like ethidium bromide and Sybr Gold.[76,77,80,81] During imaging, this can lead to weaker than expected band intensity on a gel that can be misconstrued as a low yielding PCR reaction. Because this phenomenon is not widely known, we demonstrated this problem using various imaging techniques to observe PCR amplified DNA on a denaturing PAGE. PCR reactions were performed in 100 µL volumes that contained 100 pmol of DNA primer (FAM PBS2 and PBS1) and 5 pmol of DNA template in 1x Thermopol buffer supplemented with 400 nM dATP, dTTP, dCTP, either dGTP or 7-deaza-dGTP, and 0.04 U/µL DV exo-. The solution was thermocycled as follows: 2 min. at 95°C followed by 15 cycles of 15 sec at 95°C, 15 sec at 58°C, and 45 sec for dGTP or 2.25 min. for 7dGTP at 72°C. PCR reactions were separately analyzed on four different PAGE gels. The fluorescein signal was imaged using a Typhoon TRIO scanner. The gels were then individually stained with either ethidium bromide, Sybr Gold, Sybr Green, or Syto60. Ethidium bromide, Sybr Gold, and Sybr Green stained gels were imaged on a Biorad GelDoc, and the Syto60 stained gel was imaged on a LI-COR Odyssey CLx imager.

Synthesis of DNA Templates Containing 7-Deaza-Deoxyguanosine (7dG)

Single-stranded DNA templates containing 7dG in place of dG were generated by PCR. Each PCR reaction (1,000 µL total volume) contained 1,000 pmol of each DNA primer (PBS2 and PBS1.PEG.long) and 50 pmol of DNA template or single-stranded library. Reactions were performed in 1x Thermopol buffer [20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1% Triton X-100, pH 8.8] supplemented with

400 nM dATP, dTTP, dCTP, 7-deaza-dGTP, and 0.04 U/μL DV exo-. The solution was divided into 50 μL reactions and thermocycled as follows: 2 min. at 95°C followed by 15 cycles of 15 sec at 95°C, 15 sec at 58°C, and 2.25 min. at 72°C. The reactions were combined, concentrated by lyophilization, and supplemented with 5 mM EDTA and 50% (w/V) urea. The solution was heat denatured for 5 min. at 95°C, and the PEG-modified strand was purified by 10% denaturing urea PAGE. The corresponding band was excised, electroeluted, and concentrated by ethanol precipitation. The precipitated pellet was re-suspended in 20 μL of nuclease free water and quantified by UV absorbance. A reaction schematic and example PAGE of the PCR are shown in **Figure 2.9**.

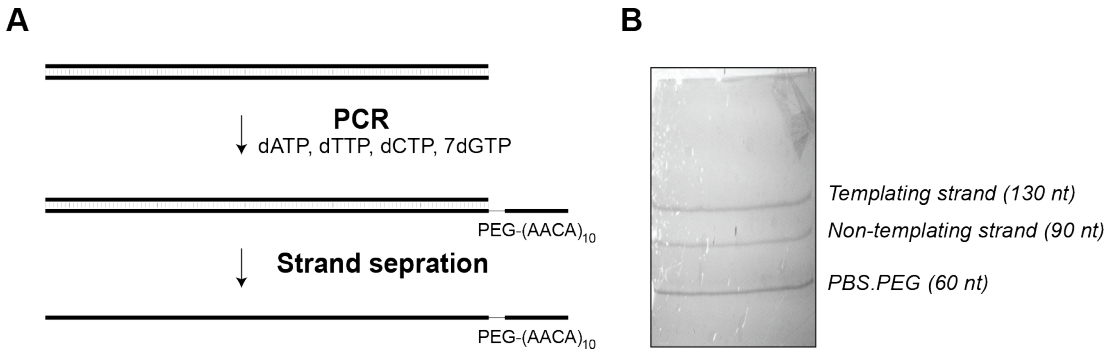

Figure 2.9. Amplification and Gel Mobility of 7dG DNA Template. (A) Double-stranded DNA library L3 was amplified by PCR with 7dGTP in place of natural dGTP. One of the PCR primers (PBS1.PEG) is modified with a PEG spacer and a 40 nucleobase overhang for denaturing urea PAGE separation. (B) The 40 nucleobase overhang provides sufficient mobility change on PAGE. Products were visualized by UV shadowing.

Small-Scale TNA Synthesis on dG- and 7dG-Containing DNA Templates

PBS2 primer (50 pmol) was radiolabeled using T4 polynucleotide kinase and [γ-^{32}P]ATP. TNA transcriptions reactions were performed in a 10 μL volume containing 10 pmol of synthetic template or 7dG template generated by PCR and 5 pmol of radiolabeled

PBS2 primer. The primer and template were annealed in 1x Thermopol buffer by heating for 5 min. at 95°C and cooling for 10 min. at 4°C. 0.1 U/all Therminator DNA polymerase was premixed with 1.25 mM $MnCl_2$. The premixed Therminator/$MnCl_2$ solution and 100 μM tNTPs were added to the primer template complex and the solution was incubated for 3 hours at 55°C for 4NT.0g, 4NT.1g, 4NT.2g, and 4NT.3g templates. A 6-hour incubation was used for the L3 library. The reaction products were analyzed by 20% denaturing PAGE and visualized by phosphor-imaging with a Storm scanner.

Large-Scale TNA Synthesis of 7dG-Modified DNA Template for the Fidelity Assay

Large-scale TNA synthesis reactions were performed in a 250 μL volume reaction containing 250 pmol of extra.PBS2.mismatch DNA primer and an equal amount of FI.4NT.9G DNA template generated by PCR with 7-deaza-dGTP. The primer and template were annealed in 1x Thermopol buffer by heating for 5 min. at 95°C and cooling for 10 min. at 4°C. 0.1 U/all Therminator DNA polymerase was premixed with 1.25 mM $MnCl_2$. The premixed Therminator/$MnCl_2$ solution and 100 μM tNTPs were added to the primer template complex and the solution was incubated for 6 hours at 55°C. TNA was purified from the PEG-modified DNA template by 10% denaturing urea PAGE, electroeluted, and concentrated by ethanol precipitation. The precipitated pellet was re-suspended in 20 μL of nuclease free water and quantified by Nanodrop. This strategy produces a TNA strand with a single T-T mismatch that can be identified as a watermark (point mutation) only after reverse transcription, PCR amplification, and DNA sequencing. A reaction schematic and example PAGE of the transcription are shown in **Figure 2.10**.

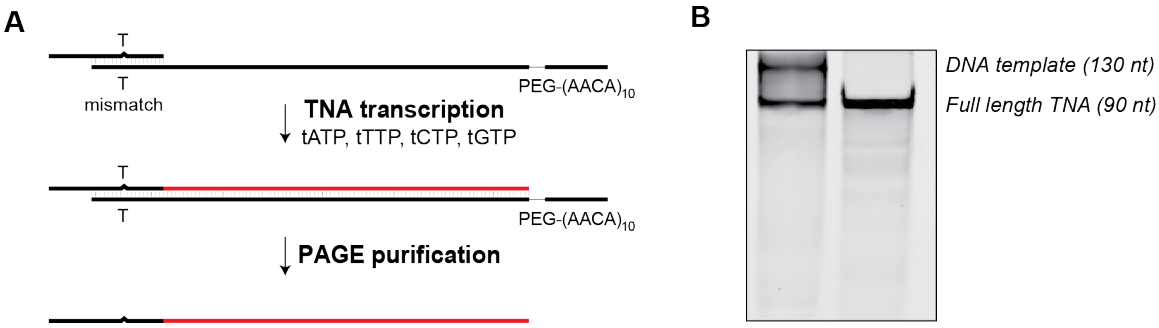

Figure 2.10. Transcription and gel Mobility of Template and TNA Product. (A) Single-stranded template containing either dG or 7dG is transcribed into TNA. The resulting TNA-DNA duplex is purified by denaturing urea PAGE. (B) The 40 nuclease overhand provides sufficient mobility change on PAGE. A small amount of TNA transcription was run and the bands were visualized using ethidium bromide to demonstrate band mobility. The large-scale TNA transcription gel is not stained with ethidium and rather visualized by UV-shadowing on a TLC plate.

TNA Reverse Transcription

PBS1 (50 pmol) primer was radiolabeled using T4 polynucleotide kinase and [γ-^{32}P]ATP. TNA was reverse transcribed in a final volume of 10 µL. 5 pmol radiolabeled DNA primer PBS1 was annealed to 10 pmol TNA template in 1x First Strand Buffer [50 mM Tris-HCl, 75 mM KCl, 3 mM $MgCl_2$, pH 8.3] by heating for 5 min. at 95°C and cooling for 10 min. at 4°C. 500 µM dNTPs and 10 mM DTT was added to the primer template complex and incubated for 2 min. at 42°C. Finally, 3 mM $MgCl_2$, 1.5 mM $MnCl_2$, and 10 U/µL SuperScript II reverse transcriptase was added to the reaction and incubated for 1 hour at 42°C. The reaction products were analyzed by 20% denaturing PAGE and visualized by a storage phosphor screen with a Storm scanner.

PCR Analysis of the TNA Sample Before and After Reverse Transcription into cDNA

PCR was used to assay the TNA sample for DNA contaminants that could interfere with the fidelity study. Two 100 μL PCR reactions were performed on 1% volume of the TNA product and TNA RT solutions in 1x Thermopol buffer, supplemented with 400 nM dNTPs, 100 pmol of DNA primer (extra.primer and PBS1), and 0.04 U/μL DV exo-. The reactions were cycled as follows: 2 min. at 95°C followed by 27 cycles of 15 sec at 95°C, 15 sec at 58°C, and 45 sec at 72°C. To monitor DNA amplification, aliquots (8 μL) were removed every 3 cycles and stored on ice. Once the PCR reaction was complete, the samples were analyzed on a 3% agarose gel stained with ethidium bromide.

Fidelity Analysis by DNA Sequencing

PCR amplified DNA (1 pmol) was ligated into a pJET vector following manufacturer's protocol. The ligated product was transformed into XL1-blue E. coli, cloned, and sequenced at the ASU sequencing facility. Sequencing results were analyzed using CLC Main Workbench.

Pre-Steady-State Kinetic Analysis

The 21-mer primer strand was radiolabeled using Optikinase and [γ-^{32}P]ATP for 3 hours at 37°C following manufacturer's protocol. The 5′-radiolabeled primer was then annealed to each 41-mer template by incubating the primer and template in a 1:1.15 ratio for 5 min. at 95°C before cooling slowly to room temperature over several hours.

All pre-steady-state assays were performed at 37°C in reaction buffer R (20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 0.1 mg/mL BSA, 0.1 mM EDTA, 5 mM DTT, 10% glycerol, pH 8.8), and were carried out using a rapid chemical quench-flow apparatus. A pre-incubated solution of 100 nM Therminator and 20 nM 5'-radiolabeled DNA substrate in buffer R containing 1.25 mM $MnCl_2$ was rapidly mixed with 100 μM dNTP or tNTP and 2 mM Mg^{2+} in reaction buffer R for various times before quenching with 0.37 M EDTA. All reported concentrations are final after mixing.

The reaction products were separated by 17% denaturing urea PAGE. The gel image was exposed to a storage phosphor screen, imaged using a Typhoon TRIO, and analyzed by ImageQuant. The plots of product concentration versus time were prepared using Kaleidagraph. Data were fit to a single-exponential equation ([product] = A[1 – exp($-k_{obs}t$)]), where k_{obs} is the observed reaction rate and A is the product amplitude. All reactions were performed in duplicate.

CHAPTER 3

INTRODUCTION TO POLYMERASE ENGINEERING

Introduction

The discovery that 7dG improves the fidelity of four nucleotide TNA transcriptions enabled, for the first time, the generation of highly complex libraries for in vitro selection. This solution, however, encountered fundamental challenges in execution due to the difficulty of amplifying 7dG-containing molecules and the cost of 7dG substrates. Surprisingly, 7dG nucleotide triphosphates are almost 10 times as expensive as their natural counterpart, which quickly becomes excessive when working on the large scales required for in vitro selection. Therefore, rather than trying to modify the substrates themselves to improve fidelity, developing a new generation of polymerases with higher fidelity and processivity for TNA that do not require 7dG would be desirable.

Polymerases serve a central biological function of maintaining both the storage and flow of genetic information. They perform these tasks on long nucleic acid stretches with a precisely defined catalytic efficiency, processivity, and fidelity that facilitate both life and adaptive evolution. Because of these properties, polymerases have obtained a paramount role in biotechnology and have generated a surge of interest in both improving the physical properties (fidelity, processivity, thermostability, salt chemical stability, etc.) of existing polymerases and identifying new variants with unique properties.

For this same reason, engineered polymerases have emerged as powerful tools in the synthesis of unnatural genetic polymers.[41,59,75] While current XNA polymerases are capable of converting short lengths (<90 nt) of DNA templates into XNAs, they often function with reduced activity compared to their wild-type counterparts. This is

presumably because polymerases have evolved over billions of years to maintain cellular integrity by excluding damaged and non-cognate substrates from the genome[82]. This high level of specificity provides a formidable barrier to the design of engineered polymerases with expanded substrate specificities. Magnifying this barrier is the large size of polymerases leading to an enormous potential sequence space to explore for new activity.

Recent developments in the polymerase engineering field have made it possible to screen large portions of sequence space for new activity. However, these methods currently rely on diversity introduced through random mutation in a predetermined polymerase. While, this strategy has led to some fruitful results, unguided strategies lead to large fractions of the total mutational load present in degenerate sites that have no effect on polymerase function. This dilutes the search efficiency. Better understanding of the sequence-function relationship of polymerases, more specifically the identification of critical residues in the polymerase scaffold that alter substrate specificity, would allow for the design of focused libraries that hone searches to regions of sequence space most likely to harbor to functional members.

Polymerase Structure and Function

Polymerases are highly specialized enzymes that catalyze the synthesis of nucleic acid polymers by directing the addition of nucleotide 5'-triphosphates onto the 3' end of a growing strand. Although all polymerases accomplish this using a two-metal ion, catalytic incorporation of nucleotides onto a template, free 3' hydroxyl, there exists a great deal of variation in sequence, structure, and function.[83]

Polymerases are currently classified into seven families (A, B, C, D, X, Y, RT) based on their substrates and products. The most studied polymerases belong to the

replicative A and B Families, which are found in all clades of life and the viruses that infect those organisms[84]. Family C polymerases are solely found in prokaryotes and function to replicate bacterial chromosomes, and family-D polymerases are from Archaea.[84,85] Family X and Y polymerases have function in DNA repair mechanisms and are usually more error-prone, and the reverse transcriptase family consists of polymerases generally found in viruses that catalyze the conversion of RNA back into DNA, but eukaryotic constituents also exist and include telomerase. [86,87]

Family A polymerases are the most studied due to their broad applications in molecular biology. DNA pol I and its exonuclease-deficient counterpart Klenow were the first polymerases to be characterized enzymatically and structurally.[88–90] The structure provided insight into the mechanism and functionality of polymerases. The polymerase resembles a right hand whose finger, palm, and thumb embrace a templating DNA duplex and incoming substrate (**Figure 3.1A**). Interestingly, most polymerases share the same three core domain, with slight modifications in domain topology, suggesting a long term evolutionary relationship between all of the polymerases families.[91,92] Additional 3'→5' and 5'→3' exonuclease domains are present in some polymerases (**Figure 3.1B**). The finger and thumb domains are primarily responsible for binding the template and the incoming nucleotide triphosphate. Catalyzing the nucleotides incorporation into the growing strand is magnesium ion coordinated by two arginine residues found in the palm domain.[83,93,94] Additional structural categories of polymerases exist suggesting alternate evolutionary histories, but all polymerases contain two catalytic arginine residues.[95]

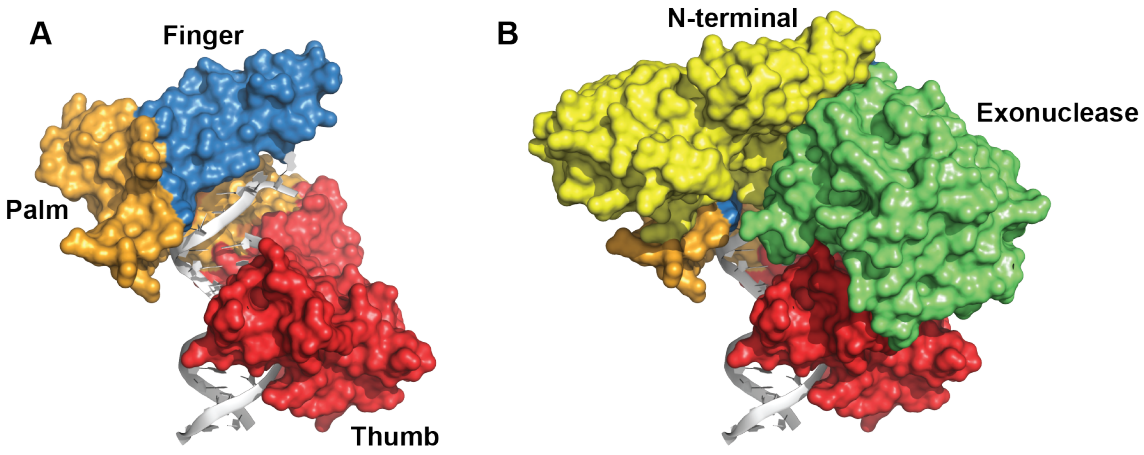

Figure 3.1. Polymerase Structural Domains. (A) Surface rendering of the finger (blue), palm (orange), and thumb (red) domains of the DNA polymerase from Thermococcus sp. 9°N (PDB 4K8X).[96] The three domains wrap around their DNA substrate white in analogy to a right hand. (B) Accessory N-terminal (yellow) and exonuclease (green) domains are present in this polymerase; however, not all polymerases contain these domains.

Family B polymerases, isolated mainly from hyperthermophilic Archaea but also found less frequently in all clades of life, are also widely used in molecular biology and biotechnology. Structurally similar to family A polymerases, family B polymerases have gained interest through their high thermostability, fidelity, and affinity for their primer-template complexes.[97,98]

Polymerase Engineering Efforts

Some of the first efforts in polymerase engineering were performed on A family polymerases.[99–108] These often included the deletion of entire domains from the polymerase structure. Two important examples are the Klenow and Stoffel fragments where the 5'→3' exonuclease domains from DNA polymerase I from *E. coli* and

Thermococcus aquaticus respectively were removed to greatly increase the processivity of these polymerases in vitro.[109,110]

Improvements in genetic engineering and bioinformatic tools led to more systematic investigations of polymerase function and the identification of new proteins with unique functions. The polymerase Phusion available from Life Technologies was generated through the fusion of the *Pyrococcus furiosis* family B polymerase with the thermostable, non-specific DNA binding domain from *Sulfolobus solfataricus*. The addition of the DNA-binding domain to Pfu DNA polymerase yielded a 9-fold increase in enzyme processivity without affecting the enzyme's other physical properties making it one of the best polymerases on the market for amplifying long DNA fragments.[111] Additionally, Phusion is recorded to have error rates as low at 4.4×10^{-7}, which is approximately 50-fod lower than Taq. Other commercially available examples of fusion DNA polymerases include Finnzyme's Phusion, Bio-Rad's iProof, Stratagene's PfuUltra II Fusion HS DNA polymerases.

A large polymerase engineering effort has been directed towards developing enzymes better suited for polymerase chain reaction (PCR).[112,113] PCR remains one of the most important applications of DNA polymerases in biotechnology due to its ability to rapidly amplify small amounts of DNA sample and has achieved wide-spread use from molecular cloning to diagnostics. PCR relies on thermal cycling, consisting of cycles of repeated heating and cooling of the reaction for DNA melting, primer binding, and enzymatic replication of the DNA. Engineering efforts have been directed to improving thermostability, fidelity, and resistance to inhibitors or DNA damage.[103,111,114–118]

Polymerases with both low and high fidelity for error-prone PCR[107] or faithful gene amplification have been identified.

Examples of Engineered Substrate Specificity

Replicative DNA polymerases are highly effective in distinguishing cognate base pairs from non-cognate or damaged substrates. Despite ribonucleotide concentration more than 10-fold higher than deoxynucleotides in a bacterial cell, *E. coli* DNA polymerase I displays $10^4 \sim 10^5$-fold selectivity against NTPs.[119,120] Polymerases achieve remarkably high fidelity by discriminating cognate base pairs from non-cognate substrates through electrostatic interactions primarily in the active site but also throughout the polymerase. Nucleotide incorporation is generally highly sensitive to aberrations in the geometric shape and change of the interaction between the incoming and templating bases. Surprisingly, this interaction is weakly dependent on Watson-Crick base pairing (with the exception of polY polymerases, which appear to rely on hydrogen bonding to a much greater extent).[8,15,121–127] Further, extension of the incorporated base depends on the presence of cognate minor groove hydrogen bond donor and acceptor groups and aberration in duplex geometry by misincorporation can impair extension activating polymerase stalling and editing mechanisms.[128–131]

The development of Sanger sequencing required the generation of polymerases accepting of dideoxy and base modified substrates.[69,70,132–135] For example, a single amino acid substation (F667Y) in Taq DNA polymerases increased its utilization of dideoxy substrates by over a thousand-fold.[77,136] In family B polymerases, the A488L mutation in Vent (or A485L mutation in *Thermococcus* sp. 9°N known commercially as Therminator DNA polymerase) improve ddNTP incorporation by over 10-fold in combination with the D141A and E143A exonuclease silencing mutations.[69,70,137] NEB identified a variant of 9°N DNA polymerases containing three additional mutations (L408S, Y409A, and P410V) known as Therminator III capable of incorporation 3'-amino- and 3'-azido-blocked nucleotides. Mutations in the region, known as the A-motif and C-motif has led to a series of polymerases variants capable of incorporating ribonucleotide and Cy-dye among others.[75,100,105,120,138–140]

Interestingly, family A and B polymerases were shown to have different intrinsic biases for incorporating unnatural substrates[69]. Vent A488L and Therminator DNA polymerase have demonstrated remarkable ability to incorporate a broad range of unnatural nucleotide substrates including TNA.[31,41,59,62,141–143]

A new generation of family B polymerases capable of limited synthesis of broad range of XNA substrates was developed using a technique called compartmentalized self-replication.[41] In this technique polymerase variants generated by random mutagenesis then expressed in *E. coli* are encapsulated in water/oil droplet emulsions to create a library of artificial cells in which the DNA and polymerase of each enzyme variant are compartmentalized in a single microreactor. Polymerases capable of extend a biotin-modified, DNA primer annealed to a region of their encoding DNA with xNTPs create a

stable DNA-XNA hybrid duplex that can be recovered by affinity chromatography. The recovered genes are then reintroduced into new bacteria and the process is repeated. These polymerases demonstrated modest fidelity and ability to replicate libraries of XNA polymers. Using the polymerases, HNA aptamers were generated against hen-egg lysozyme and HIV Tar-1 by in vitro selection.

Summary

Polymerases are an ancient class of enzymes that have evolved to exclude non-cognate substrates for the viability of their host. Engineering attempts have managed to overcome this intrinsic barrier and are beginning to generate XNA polymerases. These few examples have generated a foundation for the development of a new generation of polymerases with new or improved activities over their predecessors. To do so, identification of the determinants of substrate specificity and new technologies for the identification of active variants from large pools are required.

CHAPTER 4

ACHIEVING IMPROVED POLYMERASE ACTIVITY WITH UNNATURAL SUBSTRATES BY SAMPLING HOMOLOGOUS PROTEIN ARCHITECTURES

Publication Note

This research has been submitted to *Nature Chemistry*. Dunn, M.R., Otto, C., Fenton, K., Chaput, J.C. Achieving Improved Polymerase Activity with Unnatural Substrates by Sampling Homologous Protein Architectures. *Nat. Chem.* (Submitted).

Contribution Statement

M.R.D. performed conservation analysis. M.R.D. and C.O. generated polymerase variants and performed activity assays. K.E.F. expressed and purified the polymerases. M.R.D. and J.C.C. designed the study and wrote the paper. All authors discussed results and commented on the manuscript.

Introduction

The ability to synthesize and propagate genetic information encoded in the framework of xeno-nucleic acid (XNA) polymers would inform a wide range of topics from the origins of life to synthetic biology. While directed evolution has produced examples of engineered polymerases that accept XNA substrates, these enzymes function with reduced activity relative to their natural counterparts. Here, we describe a biochemical strategy that enables the discovery of optimal polymerase architectures for a given unnatural polymerase function. We show that it is possible to identify homologs with properties that are not present in the donor scaffold by transferring gain-of-function mutations between orthologous protein structures. This strategy was validated in separate examples that use engineered polymerases to synthesize RNA and TNA. The ability to

combine phenotypes from different donor and recipient scaffolds provides a new paradigm in polymerase engineering where natural structural diversity can be used to create enzymes with distinct and desirable properties.

The ability to encode and decode genetic information in XNA polymers is a major goal of synthetic biology and an effort that would improve our understanding of why nature chose DNA as the molecular basis of life's genetic material.[46] In principle, information transfer could be accomplished by non-enzymatic template-directed polymerization in which XNA building blocks assembled on a DNA template are coupled together using purely chemical methods. Over the past few decades, this approach has been used to 'transcribe' natural DNA or RNA templates into a wide range of biopolymer analogues, including systems with modified bases and backbones, as well as other systems with diverse chemical functionality.[144–150] In a dramatic recent advance, Liu and coworkers extended this concept to include the synthesis, selection, and amplification of peptide nucleic acids (PNA) from a library of 108 different PNA molecules.[64] This proof-of-principle demonstration, which showed how PNA could be made to evolve in vitro, was later expanded to include sequence-defined polymers with chemical compositions that are structurally unrelated to DNA.[151] However, despite significant progress, enzyme-free polymerization methods remain a challenging strategy for XNA synthesis due to the reduced coupling efficiency observed for monomeric and short polymeric units.

Enzyme-mediated strategies that mimic certain biosynthetic pathways found in nature provide an alternative approach to generating synthetic polymers. DNA replication and RNA transcription, for example, use polymerases to synthesize information-carrying polymers that maintain the flow of genetic information in biological systems.

Unfortunately, modified nucleotides are often poor substrates for natural polymerases. This is presumably due to the fact that polymerases are an ancient class of enzymes that have evolved over billions of years to maintain cellular integrity by excluding damaged and non-cognate substrates from the genome.[82] This high level of specificity provides a formidable barrier to the design of engineered polymerases that can store and retrieve genetic information in xeno-nucleic acid polymers with diverse backbone architectures.

Notwithstanding the challenges of polymerase substrate selectivity, new advances in polymerase screening techniques have uncovered a number of engineered polymerases that can accept chemically modified substrates.[152,153] Several laboratories have identified variants of natural DNA polymerases that are able to expand the genetic alphabet by incorporating unnatural bases along side the naturally occurring bases of A, C, G, and T (U).[14,121,154] Previously, our laboratory identified a polymerase set that will convert sequence-specific information back and fourth between threose nucleic acid (TNA) and DNA.[27] In a recent advance, Holliger and colleagues have developed several polymerases that can 'transcribe' and 'reverse transcribe' artificial genetic polymers composed of TNA, hexitol nucleic acid (HNA), locked nucleic acid (LNA), cyclohexyl nucleic acid (CeNA), arabinonucleic acid (ANA), 2'-fluoro-arabino nucleic acid (FANA)[41]. Some of these enzymes have been used to evolve functional XNA molecules with ligand binding and catalytic activity.[41,42,155–157]

Although enzyme-mediated examples provide the foundation for storing and accessing information in sequence-defined polymers, all of the XNA polymerases developed to date function with reduced activity relative to their natural counterparts. This constraint limits the efficiency of in vitro selection experiments that aim to explore

the structural and functional landscape of artificial genetic polymers and could lead to a build up of errors that will eventually scramble the intended molecular message. Here, we report the development and implementation of a protein design strategy that was used to refine the activity of engineered polymerases generated by molecular evolution or library screening. This strategy uses the natural structural diversity of orthologous proteins to identify the optimal polymerase architecture for a given unnatural polymerase function. By transferring beneficial mutations between homologous scaffolds, we show that it is possible to generate new polymerases that function with superior activity.

Results

Previous observations of the low fidelity of threose nucleic acid (TNA) synthesis by an engineered DNA polymerase stimulated us to explore the structural factors that govern polymerase activity.[27,59,61] Recognizing that compelling structural and phylogenetic evidence exists to support an adaptive pathway in the evolution of DNA and RNA polymerases, we hypothesized that substrate specificity could be defined by a minimum set of specificity determining residues (SDRs).[158] This hypothesis suggests that engineered polymerases could be developed with greater efficiency by targeting the subset of amino acid positions that control substrate recognition, which in turn would help advance the development of XNA polymerases by reducing the synthetic burden imposed by unnatural substrates.

We further postulated that orthologs—genes from different species that evolved from a common ancestor and share the same function—provided an interesting new strategy for expanding the functional properties of engineered polymerases. Because of subtle differences in protein folding and dynamics, we predicted that gain-of-function

mutations discovered in one polymerase scaffold would exhibit different phenotypic properties when transferred to a homologous protein scaffold. We call this approach scaffold sampling as it allows newly discovered mutations to sample slightly different regions of protein fold space. Although it is known that mutations can be transferred between orthologs, efforts to harness the natural structural diversity of known protein folds remains limited.[159] Early work by Stemmer et al. examined DNA shuffling as a method for protein recombination, but this strategy requires iterative rounds of directed evolution.[160] By contrast, scaffold sampling is an optimization strategy that functions independent of Darwinian evolution.

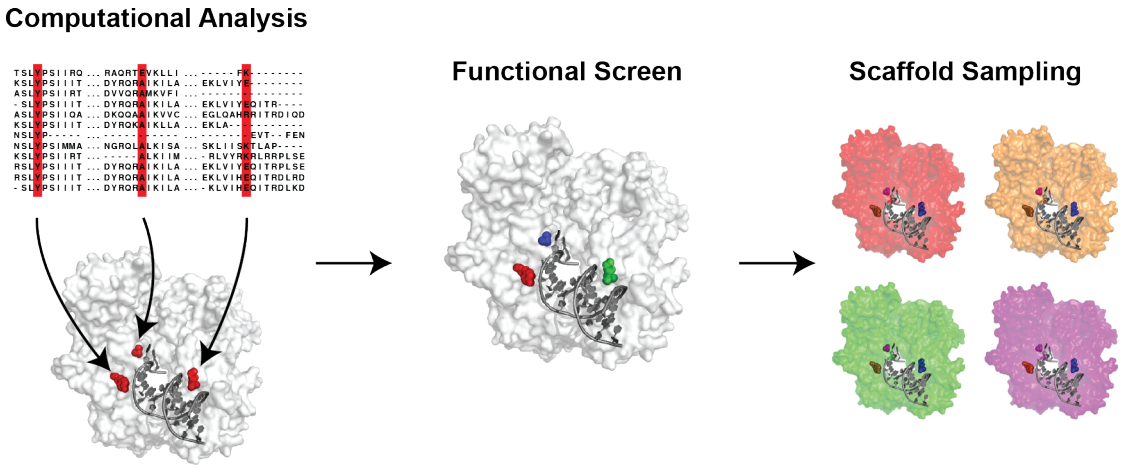

Figure 4.1. Polymerase Engineering Strategy. Overview of the process of identifying polymerases with enhanced activity for incorporating non-natural nucleotides. The first step is to identify specificity determining residues (SDRs) within a set of related polymerases. Briefly, a computational analysis is used to identify amino acid positions that are conserved in sequence and structure, and located within 10 Å of the DNA primer-template complex. Positions that have been described previously in the literature are given higher priority in functional assays. SDRs are then tested for non-cognate substrate activity in a model polymerase scaffold. Mutations with the strongest gain-of-function

activity are then examined in homologs for properties that are not present in the donor scaffold.

The combination of both suppositions led us to derive a biochemical strategy that could be used to optimize polymerase functions (**Figure 4.1**). For this study, we chose family B DNA polymerases as a model system because (i) a number of engineered variants have been developed to function with enhanced activity for unnatural substrates, including TNA triphosphates (tNTPs) (**Table 4.1**); (ii) family B polymerases are found throughout the tree of life, including Archaea, Eubacteria, and Eukarya, and in the viruses that infect these organisms, indicating that substantial sequence diversity exists between related members; and (iii) the three-dimensional coordinates of several high resolution crystal structures are available in the protein databank (PDB), which makes it possible to perform structural comparisons between polymerases from different clades.[41,61,63,70,75,161]

We used a computational procedure to identify putative SDRs in a representative set of family B polymerases.[162] Comparative phylogenetic and structural data allowed us to identify 144 amino acid positions that are conserved in both sequence and structure, 50 of which lie within 10 Å of the primer-template complex (**Figure 4.2**). We considered these 50 sites to be putative SDRs due to their close proximity to the enzyme active site.[163] A thorough review of the literature revealed that 8 of these sites overlapped with positions known to impact the recognition of unnatural substrates, and were therefore given a higher priority in functional assays (**Figure 4.3A and Table 4.1**). Three of the 8 sites, residues 409, 491, and 664, form primary contacts to the primer-template complex, while other positions are more distal and could form critical second-shell contacts that help define the shape and electrostatic potential of the enzyme active site (**Figure 4.3B**).

Relative to most other positions in the polymerase family, these sites are highly conserved at the amino acid level (>83% vs. 55% sequence identity) (**Figure 4.3C and 4.4**).

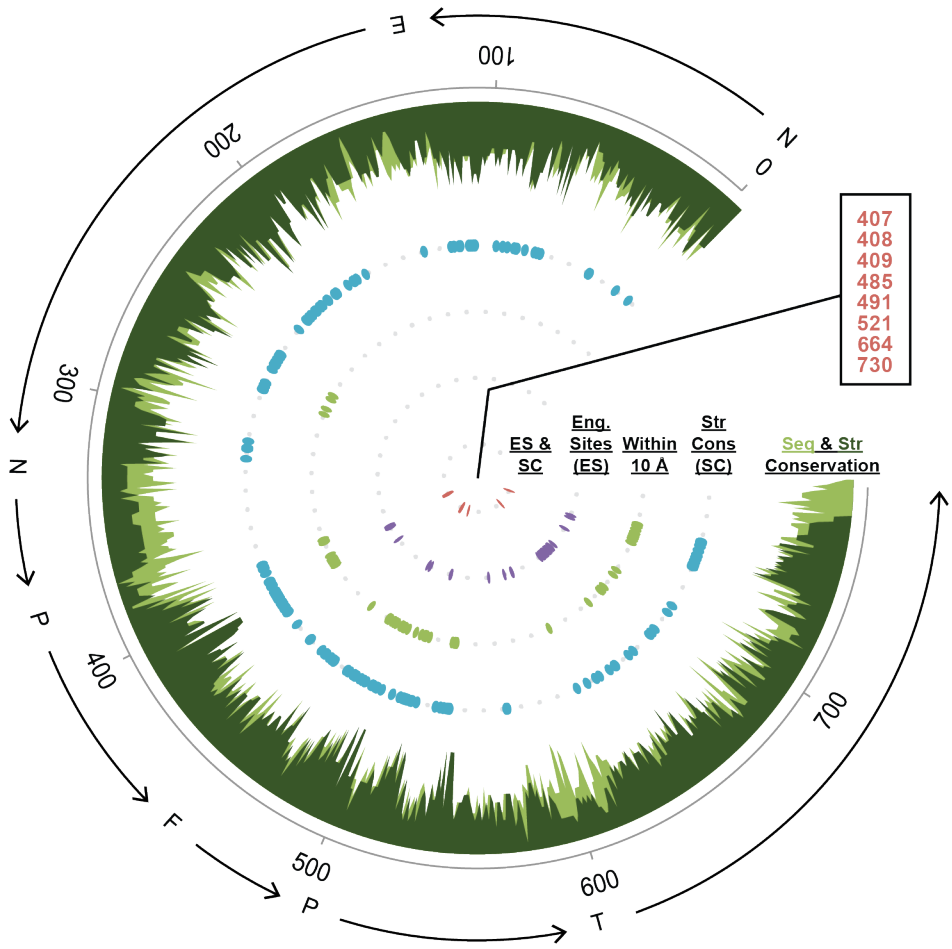

Figure 4.2. Specificity Determining Residues in Family B DNA Polymerases. Circular diagram showing the sequence (light green) and structural (dark green) conservation of representative family B polymerases. Inner rings depict amino acid positions that are conserved in both sequence and structure (blue dots), the subset of conserved sites that lie within 10 Å of the DNA primer-template complex (green dots), and amino acid positions with reported substrate modifying activity (purple dots). Sites where computational and empirical data overlap (red dots) represent the top 8 specificity determining residues (SDRs). Domain abbreviations: N, N-terminus; E, exonuclease; P, palm; F, finger; and T, thumb.

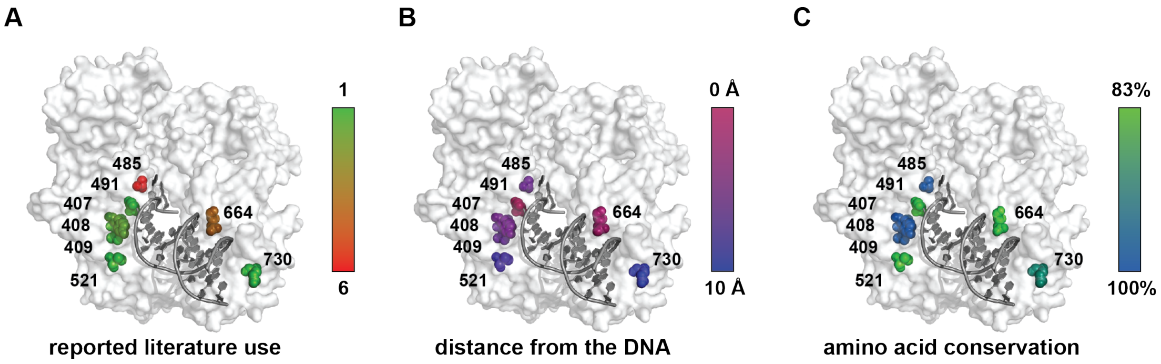

Figure 4.3. Analysis of Top 8 Predicted Substrate Determining Residues. Heat maps showing the 8 predicted SDRs mapped onto the surface of 9n DNA polymerase (PDB: 4K8X).[63] Residues are colored by their frequency of reported literature use (A); distance from the primer-template complex (B); and evolutionary conservation (C).

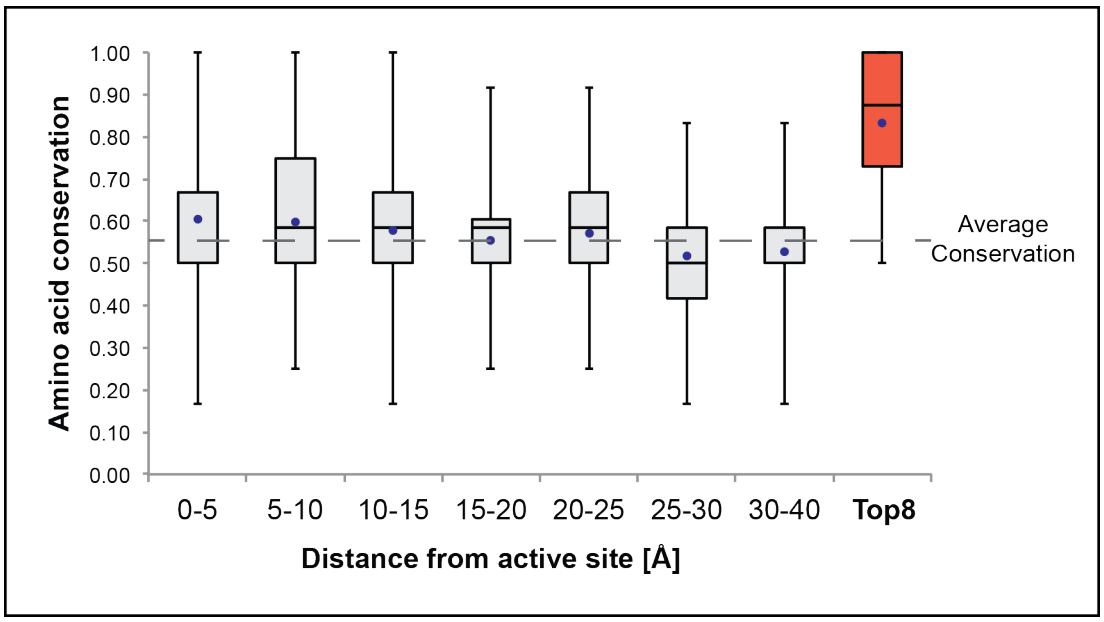

Figure 4.4. Global Sequence Conservation of Family B DNA Polymerases. Sequence conservation at the amino acid level was measured for 12 representative family B DNA polymerases isolated from Archaea, Eubacteria and Eukarya, and the viruses that infect these organisms. Overall sequence conservation (dashed line) is similar to sequence conservation measurements sampled at increasing distances from the enzyme active site.

Table 4.1. Gain-of-Function Mutations Identified in Family B DNA Polymerases.

Wild-Type Polymerase	Mutation	Conferred Activity	Reference
9n	A485L	3’-deoxyapionucleotides	[36]
9n	N/A	FANA	[142,143]
9n	A485L	RNA	[161]
9n	L408Q	RNA	[75]
9n	G77S, E276D, A281V, E300K, N568K, Y653C	RNA	[75]
9n	I171V, K476E, I488F, I618F	RNA	[75]
9n	A485L	TNA	[61,62]
Pfu, Tgo	V93Q	2’-deoxyuridine	[164–166]
Phi29	Y390F (9n: 491)	8-oxo-2'-deoxyguanosine	[167]
Tgo	A485L, E654Q, E658Q, K659Q, V661A, E664Q, Q665P, D669A, K671Q, T676K, R709K	ceNA, LNA, ANA, FANA	[41]
Tgo	L403P, A485L, P657T, E658Q, K659H, Y663H, E664K, D669A, K671N, T676I	ceNA, LNA, ANA, FANA	[41]
Tgo	A485L, V589A, E609K, I610M, K659Q, E664Q, Q665P, R668K, D669Q, K671H, K674R, T676R, A681S, L704P, E730G	ceNA, HNA	[41]
Tgo	409G, 485L, 664L	RNA, 2-N_3–substituted RNAs, 5meC and Ψ NTPs	[168]
Vent	Y410V/A486L (9n: 409/485)	RNA	[70]

To test the hypothesis that scaffold sampling could be used to identify functional differences between homologous protein scaffolds, we searched the literature for examples of engineered polymerases that contained one or more of the top 8 SDRs predicted by our computational analysis. We focused our search on DNA polymerases that have been developed to synthesize RNA, which has long been a model system for polymerase engineering. The enzyme that best satisfied our criteria is a mutant version of an Archaean DNA polymerase isolated from the species *Thermococcus gorgonarius* (Tgo).[168] This polymerase, referred to as Tgo-QGLK, contains mutations at 3 of the 8 predicted SDRs (Y409G, A485L, and E664K) as well as additional mutations at the 3'→5'-exonuclease (exo-) silencing (D141A and E143A) and uracil-stalling (V93Q) positions (**Figure 4.5**).

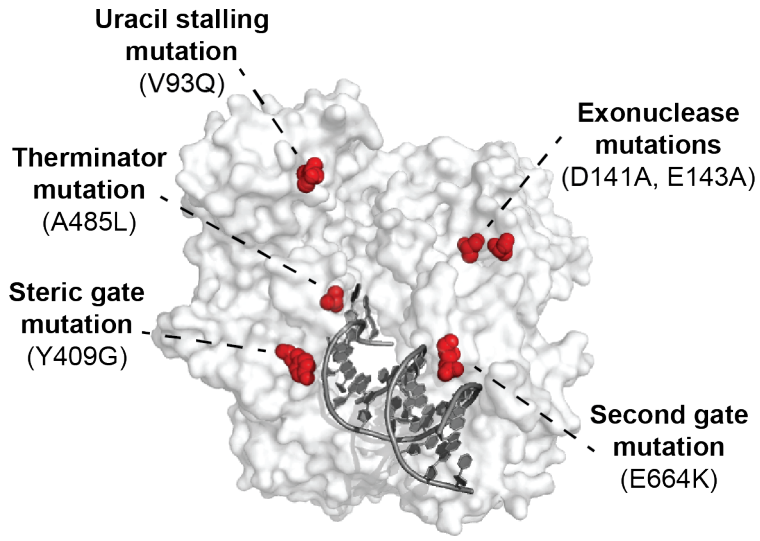

Figure 4.5. RNA Gain-of-Function Mutations. Mutations required for RNA polymerase activity in Tgo DNA polymerase were mapped onto the structure of 9n DNA polymerase (PDB: 4K8X). Abbreviations: Q: glutamine, G: glycine, L: leucine, K: lysine, and A: alanine.

We constructed a series of engineered polymerase variants by sequentially introducing the QGLK mutations into a related Archaean DNA polymerase isolated from *Thermococcus* sp. 9°N exo- (9n). Following expression and purification from *E. coli*, we challenged 9n-QGLK and its variants to extend a DNA primer-template complex with ribonucleotide triphosphates (NTPs). We found that the QGLK mutations endow 9n with strong RNA synthesis activity, as evidenced by the ability of 9n-QGLK to efficiently extend a DNA primer with 40 sequential ribonucleotides (**Figure 4.6A**). With the exception of 9n-GLK, the remaining polymerases each stalled after 10-15 nucleotide incorporations. The observation that the GLK mutations function with high activity is consistent with the model that these sites occupy SDR positions in the polymerase scaffold. Notably, mutation V93Q, which was not identified in our analysis of SDR positions, exhibits no gain-of-function activity over the GLK mutations in the 9n scaffold (**Figure 4.6A**).

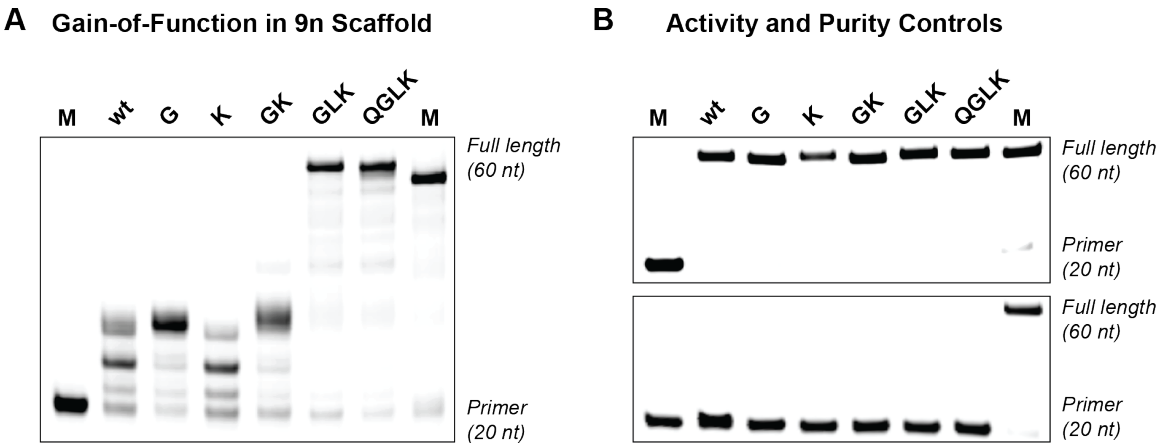

Figure 4.6. Introduction of RNA Gain-of-Function Mutations into 9n DNA Polymerase. Cumulative RNA polymerase activity of the QGLK mutations in the 9n DNA polymerase scaffold (A). The full-length marker was generated using dNTPs. Polymerase activity assays were performed in the presence (B) and absence (C) of dNTPs to confirm that each recombinant polymerase was functional and free of dNTP or NTP contamination. Products were visualized by IR fluorescence.

Control experiments were performed in the presence and absence of dNTP substrates to ensure that the enzymes were properly folded, functional, and free of dNTP contamination that could lead to a false positive in the absence of additional DNA or RNA nucleotide triphosphates (**Figure 4.6B**). In all cases, full-length DNA product was observed when dNTPs were present in the polymerase activity assay, while the primer remained unextended in the absence of dNTPs.

To test the generality of transferring functions between polymerases, we performed the same primer-extension assay on a set of related Archaean DNA polymerases isolated from the species Tgo, *Thermococcus kodakarensis* (Kod), and *Pyrococcus* sp. deep vent (DV). Because these enzymes were not readily available, DNA constructs encoding exo- versions of each polymerase along with the QGLK mutations (~2.4 kb) were obtained from DNA2.0. The gene products were cloned into an expression

vector, sequence verified, and expressed and purified from *E. coli*. Testing of the recombinant enzymes in a polymerase activity assay revealed that the QGLK variants have strong RNA synthesis activity relative to their wild-type counterparts (**Figure 4.7A**), indicating that the selected mutations endow all four wild-type scaffolds with the ability to synthesize RNA. RNase digestion of the extension products confirmed NTPs were the incorporated nucleotide (**Figure 4.7B**). While this assay demonstrated that mutations and their encoding phenotypes could be transferred between homologous scaffolds, it did not address other issues of enzyme performance such as template bias.

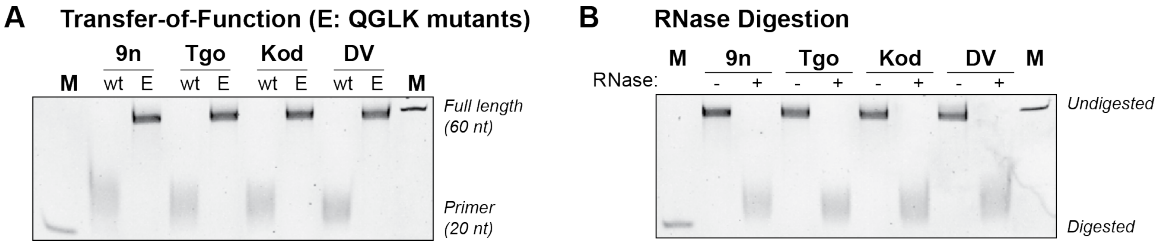

Figure 4.7. Scaffold Sampling RNA Gain-of-Function Mutations. RNA polymerase activity for wild-type and engineered 9n, Tgo, Kod, and DV polymerases bearing the QGLK mutations (A). RNase digestion of primer-extended RNA products generated by DNA-dependent RNA synthesis (B). Products were visualized by IR fluorescence.

It is well known that many polymerases have intrinsic sequence biases that can limit their activity in functional assays.[169] This problem is overlooked when polymerases are tested on individual templates that by design are unstructured and devoid of nucleotide repeats. In an effort to determine the extent to which a polymerase scaffold can impact the phenotypic expression of beneficial mutations, we evaluated the QGLK polymerases under passive and stringent conditions. In this assay, each polymerase was challenged to synthesize RNA by extending a DNA primer that was annealed to either a single sequence (passive) or a large population of random sequences containing a range of challenging template contexts (stringent). Strikingly, the four QGLK polymerases exhibit markedly different levels of efficiency when challenged with a library of sequences as compared to a single sequence (**Figure 4.8**). Comparing the four scaffolds activity on the stringent template over time demonstrated that both the Tgo and DV scaffolds were able to covert 40% of the starting primer into full length RNA within four hours while the 9n and Kod scaffolds had negligible activity (**Figure 4.9**). The Tgo scaffold was slightly faster than the DV scaffold as seen by its generation of 50% full-length product in 1.2 hours as compared to the 2 hours DV required. Both the 9n and Kod scaffolds produced small amounts of product very slowly over the 8-hour period, but were never able to reach a significant amount of full-length product. In this particular case, strong RNA synthesis activity was observed for Tgo and DV, while 9n and Kod generated little or no full-length product. We then measured the aggregate fidelity of RNA replication for the four QGLK polymerases using an assay that analyzes the DNA product isolated when a DNA template is copied into RNA, purified, and reverse copied back into DNA (**Figure 2.6A**). The fidelity data revealed a more refined look into the

activity of the polymerases by demonstrating that DV is 5-fold more faithful than Tgo at RNA synthesis, which is interested given that both polymerases have equally low sequence bias when compared to 9n and Kod (**Table 4.2**). The fact that strong phenotypic differences are observed between homologous protein architectures engineered with the same mutations, provided the first clear evidence that scaffold sampling could be used to identify optimal structures for a given polymerase function.

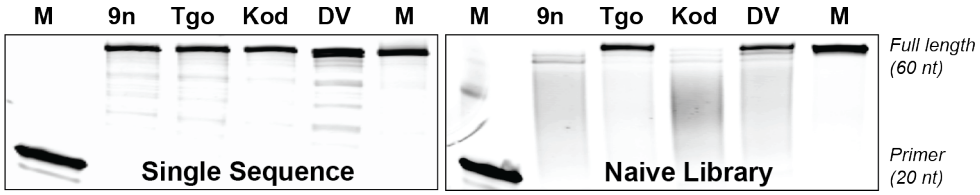

Figure 4.8. Polymerase Efficiency Assay. RNA polymerase activity assay performed on a single sequence and a naïve DNA library using 9n, Tgo, Kod, and DV polymerases engineered with QGLK mutations. Products were visualized by IR fluorescence.

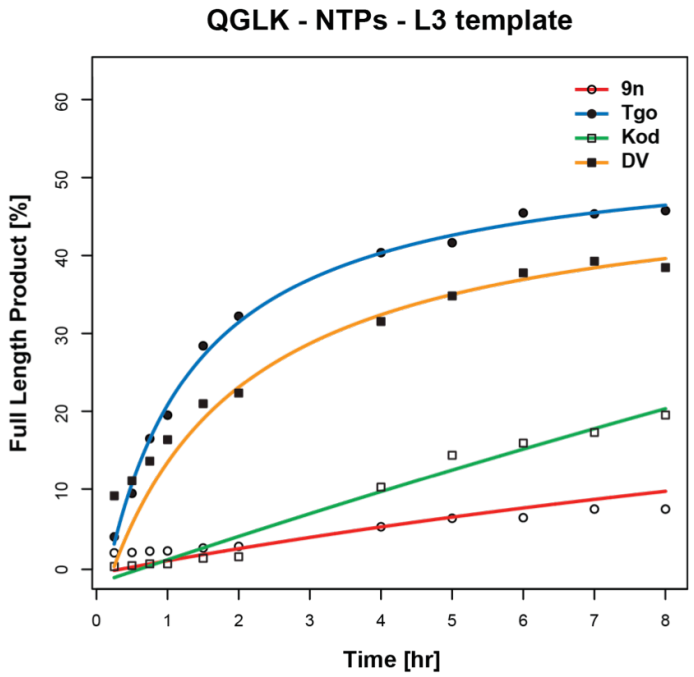

Figure 4.9. Transcription Efficiency Time Course of QGLK Mutants. Polymerase activity of the four scaffolds bearing the QGLK mutations were monitored for their ability to extend NTP substrates on the stringent L3 template over an 8 hour period. Full-length product generated at each time interval was plotted with respect to time.

Table 4.2. Fidelity for RNA Polymerase Variants.

Polymerase Variant	Backbone	Bases Read	Mutation Rate
9n-QGLK	RNA	1100	2.0×10^{-3}
DV-QGLK	RNA	1400	1.0×10^{-3}
Kod-QGLK	RNA	1100	4.0×10^{-3}
Tgo-QGLK	RNA	600	5.0×10^{-3}

Next, we sought to expand the concept of scaffold sampling to include the development of an engineered polymerase with XNA activity. In particular, we wanted to generate a polymerase that could synthesize TNA—an artificial genetic polymer

composed of repeating α-L-threose sugars that are vicinally linked by 2'→3'-phosphodiester bonds (**Figure 1.2**).[45] In addition to serving as a DNA analogue, TNA is an attractive candidate for therapeutic and diagnostic applications due to its stability against nuclease degradation.[27] Unfortunately, the current TNA polymerase (9n-L: sold commercially as Therminator DNA polymerase) suffers from a strong sequence bias that precludes the synthesis of many TNA polymers.[62] This problem is due to a propensity for G-G mispairing in the enzyme active site, which leads to a rise in G to C transversions when TNA is reverse-transcribed back into DNA.[27] While the fidelity of TNA synthesis can be improved with 7-deazaguanine (7dG), a base analogue that suppresses G-G mispairing, 7dG is a costly solution due to the scale at which most TNA transcription reactions are performed.[170]

In an attempt to identify a TNA polymerase variant that functions with improved fidelity, we targeted SDRs at positions 485 and 664, which are known to be strong gatekeepers of substrate specificity.[161,168] Cassette mutagenesis was used to construct variants of 9n that contained random mutations at positions 485 and 664. Polymerase variants were isolated and tested in a polymerase activity assay that involved extending a DNA primer-template complex with chemically synthesized tNTPs.[71,72] This analysis revealed a preference for arginine (R) at position 485 and isoleucine (I), glutamine (Q), histamine (H), or lysine (K) at position 664 (**Figure 4.10**). Subsequent screening of co-variants led to the identification of A485R and E664I, termed 9n-RI, as an efficient DNA-dependent TNA polymerase (**Figure 4.11**). Surprisingly, we found that 9n-RI has a diminished capacity for DNA synthesis, indicating that the RI mutations alter the wild-type activity of the enzyme (**Figure 4.12**).

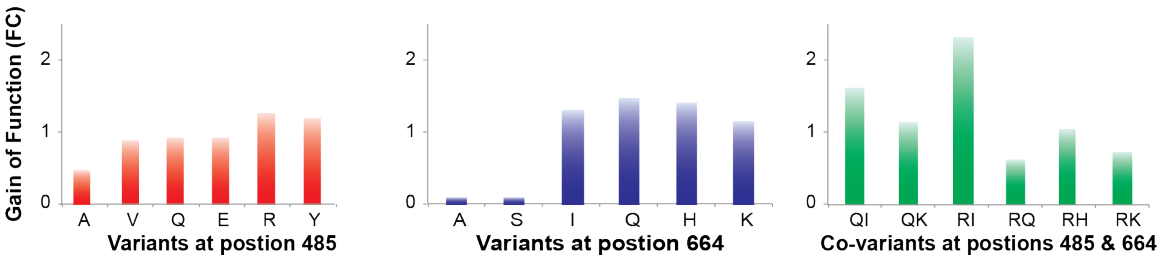

Figure 4.10. Focused Screening for Improved TNA Polymerases. Variant and covariant screening at positions 485 and 664 demonstrate the RI mutations additive effect for improving TNA transcription efficiency.

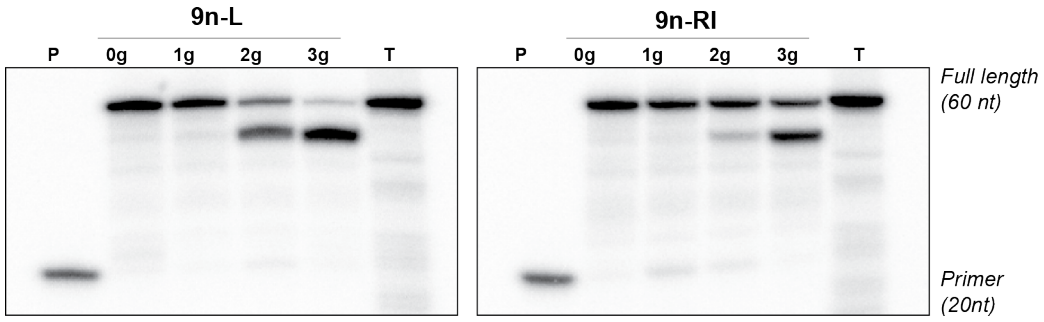

Figure 4.11. Reduced Sequence Bias of 9n-RI TNA Polymerase. Primer extension results for DNA templates that contain consecutive G-residues. 9n-RI polymerase can transcribe a higher fraction of full-length product through challenging templates. Refer to **Figure 2.2A** for template schematic. Products were visualized by IR fluorescence.

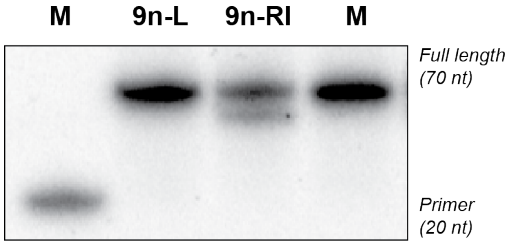

Figure 4.12. 9n-RI Polymerase has Reduced Activity for DNA Synthesis. The engineered polymerase 9n-RI is less efficient at DNA synthesis than 9n-L (sold commercially as Therminator DNA polymerase) when tested at the enzyme concentrations typically used for TNA synthesis. This result suggests that 9n-RI is beginning to change its substrate preference from dNTPs to tNTPs rather than broaden its substrate specificity, which is a more common phenomenon among engineered polymerases. Products were visualized by IR fluorescence.

Next, we examined whether the strong TNA synthesis activity produced by the RI mutations could be recapitulated in other polymerase scaffolds. For this experiment, the RI mutations were inserted into Tgo, Kod, and DV. Following expression, purification from *E. coli*, and functional and nucleotide purity validation (**Figure 4.13**), the RI-mutant polymerases were tested in side-by-side assays against an identical set of polymerases that carried only the Therminator (A485L) mutation. The resulting primer-extension assay revealed that each of the four RI-polymerases could efficiently extend the DNA primer into full-length TNA product (**Figure 4.13**), which shows that the selected mutations function equally well in four different polymerase scaffolds. By contrast, the A485L mutant polymerases exhibit noticeably less activity when tested on the same DNA template. The fact that the RI-mutant polymerases are able to copy a difficult template into TNA demonstrates that the RI-polymerases function with reduced sequence bias relative to their Therminator A485L counterparts.

Encouraged by the enhanced activity of the RI-mutant polymerases, we decided to investigate the impact of scaffold sampling on the fidelity and efficiency of TNA synthesis using a more rigorous test of enzyme performance than simple primer extension on a single well-defined template. We began by measuring the aggregate fidelity of TNA replication using an assay that analyzes the DNA product isolated when a DNA template is copied into TNA, purified, and reverse copied back into DNA (**Figure 2.6A**). For the reverse-transcription step, we used an engineered version of MMLV reverse transcriptase that was previously identified as an efficient TNA-dependent DNA polymerase.[27] By comparing the fidelity profiles produced by the four different polymerase architectures, we found that Kod-RI functions with markedly higher replication fidelity (~8-fold) than the other three RI-polymerases or the intermediate polymerases 9n-L and 9n-R (**Figure 4.15 and Table 4.3**). Furthermore, the enhanced fidelity of Kod-RI was accompanied by a meaningful gain in template synthesis efficiency as evidenced by the ability of Kod-RI to synthesize a full-length TNA library in 3 hours as compared to the 15 hours required for 9n-RI (**Figure 4.16**). Taken together, the dramatic improvements in TNA synthesis fidelity and efficiency provide strong evidence that scaffold sampling can be used to discover new phenotypic properties that are not present in the starting donor scaffold.

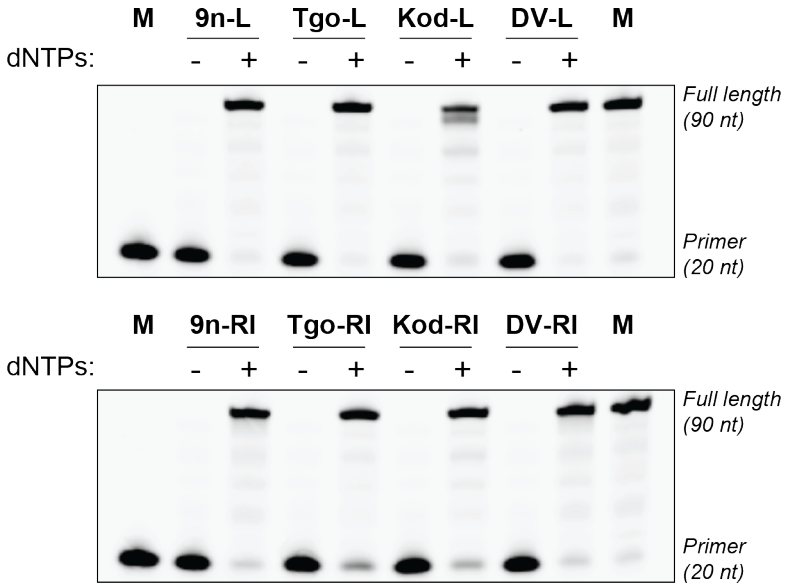

Figure 4.13. Control Experiments for Engineered TNA Polymerases. Polymerase activity assays were performed in the presence and absence of dNTPs to confirm that each recombinant polymerase was functional and free of dNTP or NTP contamination. Products were visualized by IR fluorescence.

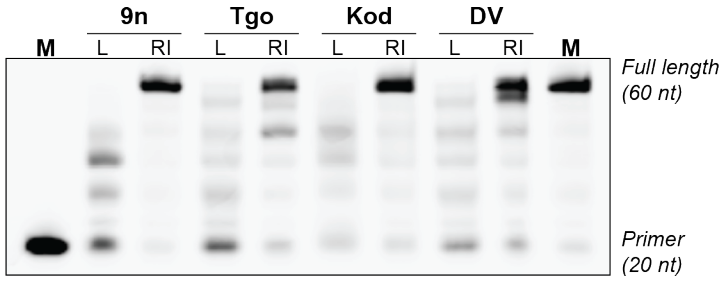

Figure 4.14. Scaffold Sampling TNA Gain-of-Function Mutations. TNA polymerase activity for engineered 9n, Tgo, Kod, and DV polymerases bearing the L (A485L) and RI (A485R and E664I) mutations. Positions 485 and 664 were identified by mutational analysis of predicted SDRs. Products were visualized by IR fluorescence.

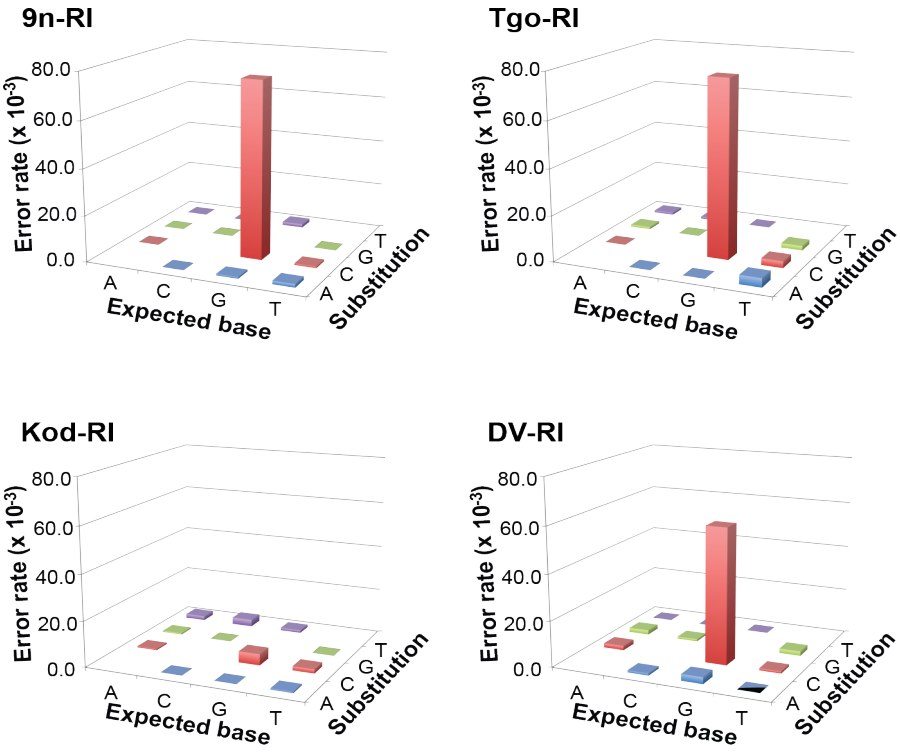

Figure 4.15. RI Polymerase Fidelity Profiles. Fidelity of TNA replication using 9n-RI, Tgo-RI, Kod-RI, and DV-RI. Tgo TNA polymerase has a five-fold reduced error rate. See **Table 4.3** for values.

Table 4.3. Fidelity for TNA Polymerase Variants.

Polymerase Variant	Backbone	Bases Read	Mutation Rate
9n-L	TNA	1300	8.0×10^{-2}
9n-R	TNA	1550	6.8×10^{-2}
9n-RI	TNA	1300	8.1×10^{-2}
DV-RI	TNA	1000	7.1×10^{-2}
Kod-RI	TNA	1600	1.3×10^{-2}
Tgo-RI	TNA	1000	8.9×10^{-2}

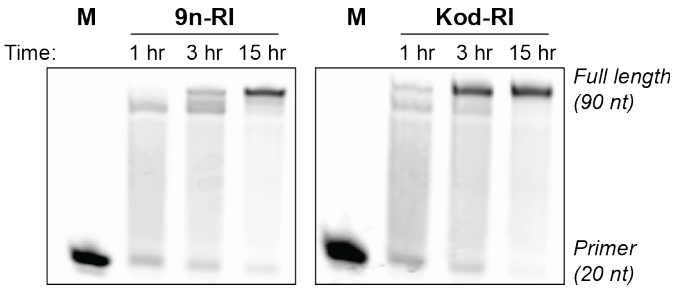

Figure 4.16. Polymerase Efficiency Comparison of Kod-RI and 9n-RI TNA Polymerases. TNA polymerase activity assay performed on a naïve DNA library using 9n-RI and Kod-RI. Products were visualized by IR fluorescence.

Rationalizing the superior activity of Kod-RI requires understanding the individual roles of the RI mutations. Some insights can be gained from the crystal structure of wild-type Kod, which has been solved in both the apo and binary form with DNA (PDB: 1WNS and 4K8Z, respectively).[63,171] Comparison of the two structures indicates that a pronounced structural movement occurs upon DNA binding wherein the thumb domain rotates inward to bind the DNA in a groove formed by the thumb and palm domains. The DNA itself adopts a standard B-form helix with the sugar residues favoring the 2' endo conformation. Although Kod forms numerous direct contacts to the phosphate backbone, very few interactions are observed to the nucleobases or sugar oxygen atoms. The A485 residue is located on the backside of the finger domain, suggesting that the increased bulk of the arginine residue favors rotation of the finger domain toward the DNA helix, possibly altering the geometry of the enzyme active site. E664 contacts the DNA by interacting with coordinated water molecules in the minor groove of the DNA helix. Substitution of this residue for a hydrophobic side chain could increase TNA synthesis efficiency by weakening contacts to primer-template complex; however, further experiments are need to support this claim.

Discussion

We have developed a protein design strategy that was used to refine the activity of two classes of engineered polymerases. This strategy is based on the concept that homologous protein structures provide access to slightly different regions of protein fold space where critical gain-of-function mutations can endow the recipient scaffold with new phenotypes that are not present in the starting donor scaffold. By surveying known and newly discovered mutations in related protein architectures, we show that it is possible to identify the optimal structure for a given unnatural polymerase function. During the course of our study, we elucidated several key properties that shed new light on the determinants of polymerase selectivity and the effect that scaffold sampling can have on the enzymatic properties of engineered polymerases. Together, these developments lay the foundation for future efforts that aim to create robust enzymatic systems that allow artificial genetic polymers to replicate in vitro and eventually inside living cells.[3]

DNA and RNA polymerases are highly selective enzymes that have very little tolerance for modified substrates. This stringent level of specificity makes it difficult to identify enzymes that will recognize unnatural substrates, either as nucleoside triphosphates or as templates. Although directed evolution and high throughput library screening methods have been used to broaden the substrate specificity profile of natural polymerases, such approaches can be difficult to implement and are overly expensive when using XNA substrates that are not commercially available.[152,153] Our method of identifying specificity determining residues contrasts sharply with these earlier strategies

by reducing the search through protein sequence space to a subset of amino acid positions that has an increased likelihood of functioning as determinants of substrate specificity.

As a demonstration, we used a computational method to identify 50 amino acid positions that were predicted to control the substrate specificity of family B polymerases. Biochemical and mutational analysis showed that 3 of predicted sites play a key role in the RNA synthesis activity of a known engineered polymerase, which indicates that our computational analysis converged on some of the same amino acid positions previously discovered by directed evolution.[168] In a subsequent example, we developed a highly efficient TNA polymerase by surveying less than 20 single and double point mutations at two computationally predicted sites. This later case is an interesting demonstration of library efficiency where a successful polymerase was generated by sampling only a small fraction of the 256 million co-variants that can be assembled for a typical family B polymerase. Taken together, these examples provide the motivation to explore other SDR positions in future engineering projects.

Perhaps the most surprising result to come from our study was the impact that scaffold sampling can have on the enzymatic properties of engineered polymerases. One might expect a priori that transferring mutations between homologous protein scaffolds would have a neutral or even deleterious effect on polymerase function due to the potential for disrupting important contacts in the native tertiary structure. However, contrary to this view, our results suggest that scaffold sampling makes it possible to create superior enzymes by combining phenotypes from different donor and recipient scaffolds. This observation, which to our knowledge has not been described previously, provides an interesting new paradigm in protein engineering where structural homology

can be used to improve the functional properties of gain-of-function mutations. Although our study focused on the development of XNA polymerases, one could imagine extending this concept to other proteins generated by computational design or directed evolution.

Although the mechanism of scaffold sampling is not fully understood, it is clear from our study that different combinations of mutations prefer different structural architectures. For example, the GLK mutations required for RNA synthesis favor Tgo and deep vent, while the RI mutations necessary for TNA synthesis prefer the Kod scaffold. Given that Archaean family B polymerases have high structural homology (α-carbon RMSD ~1.5 Å), the functional differences observed are likely due to dynamic changes in protein motion rather than structural changes in protein folding. Thus, it would seem that homologous scaffolds, being similar in structure, place the enzymes within close functional space to the desired activity, and scaffold sampling provides a local search around this space. However, in analogy to Monte-Carlo simulations, one would expect the best results to surface only when you are close to an optimal solution.

In summary, we have developed a polymerase engineering strategy that can be used to improve the activity of newly discovered mutations. We suggest that it many be possible to apply scaffold sampling to other protein classifications where existing enzymes suffer from problems related to low catalytic activity or poor solubility. This line of research provides an exciting opportunity to develop new enzymes that impact broad areas of synthetic biology, molecular medicine, and biotechnology.

Experimental

General Information

DNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), purified by denaturing polyacrylamide gel electrophoresis, electroeluted, concentrated by ethanol precipitation, and quantified by UV absorbance. NTPs and dNTPs were purchased from Sigma (St. Louis, MO). TNA triphosphates (tNTPs) were obtained by chemical synthesis as previously described.[59,61] AccuPrime DNA polymerase and SuperScript II reverse transcriptase were purchased from Invitrogen (Grand Island, NY). CloneJET PCR cloning kit was purchased from Fermentas (Waltham, MA). RNase A and hen egg lysozyme were purchased from Sigma. The 9n-L gene was kindly provided by Andreas Marx in a pGDR11 expression vector. DNA encoding the genes for DV exo-, Kod exo-, and Tgo exo- were purchased from DNA2.0 (Menlo Park, CA). DpnI restriction enzyme was purchased from NEB (Ipswich, MA). Ultracell YM-30 concentrators were purchased from Millipore (Billerica, MA). Costar Spin-X protein concentrating columns were purchased from Sigma (St. Louis, MO).

DNA Sequences

The sequences below are written from 5'→3' using IDT abbreviations for modified positions.

<u>Extension and Fidelity Primers</u>

PBS2	gacactcgtatgcagtagcc
PBS2 IR800	/5IRD800//iSp18/gacactcgtatgcagtagcc
PBS2.mismatch	cttttaagaaccggacgaacgacactcgtTtgcagtagcc
PBS1	tgtctacacgcaagcttaca
extra.primer	cttttaagaaccggacgaac

Site-Directed Mutagenesis Primers

9n E664A for	ggaaaaactggtgattcatgcgcaaattacccgtgatctg
9n E664A rev	cagatcacgggtaatttgcgcatgaatcaccagtttttcc
9n E664H for	ggaaaaactggtgattcatcatcaaattacccgtgatctg
9n E664H rev	cagatcacgggtaatttgatgatgaatcaccagtttttcc
9n E664I for	ggaaaaactggtgattcatattcaaattacccgtgatctg
9n E664I rev	cagatcacgggtaatttgaatatgaatcaccagtttttcc
9n E664K for	ggaaaaactggtgattcataaacaaattacccgtgatctg
9n E664K rev	cagatcacgggtaatttgtttatgaatcaccagtttttcc
9n E664Q for	ggaaaaactggtgattcatcagcaaattacccgtgatctg
9n E664Q rev	cagatcacgggtaatttgctgatgaatcaccagtttttcc
9n E664S for	ggaaaaactggtgattcatagccaaattacccgtgatctg
9n E664S rev	cagatcacgggtaatttggctatgaatcaccagtttttcc
9n L485A for	gctggattatcgtcagcgcgcgattaaaattctggccaac
9n L485A rev	gttggccagaattttaatcgcgcgctgacgataatccagc
9n L485E for	gctggattatcgtcagcgcgaaattaaaattctggccaac
9n L485E rev	gttggccagaattttaatttcgcgctgacgataatccagc
9n L485Q for	gctggattatcgtcagcgccagattaaaattctggccaac
9n L485Q rev	gttggccagaattttaatctggcgctgacgataatccagc
9n L485R for	gctggattatcgtcagcgccgtattaaaattctggccaac
9n L485R rev	gttggccagaattttaatacggcgctgacgataatccagc
9n L485V for	gctggattatcgtcagcgcgtgattaaaattctggccaac
9n L485V rev	gttggccagaattttaatcacgcgctgacgataatccagc
9n L485Y for	gctggattatcgtcagcgctatattaaaattctggccaac
9n L485Y rev	gttggccagaattttaatatagcgctgacgataatccagc
DV A485L for	gattacaggcaacggctgatcaaaatcctggcg
DV A485L rev	cgccaggattttgatcagccgttgcctgtaatc
DV E664I for	aagctagttatttacatccagatcacgaggccc
DV E664I rev	gggcctcgtgatctggatgtaaataactagctt
DV L485A for	gattacaggcaacgggcaatcaaaatcctggcg
DV L485A rev	cgccaggattttgattgcccgttgcctgtaatc

DV L485R for	gattacaggcaacggcgtatcaaaatcctggcg
DV L485R rev	cgccaggattttgatacgccgttgcctgtaatc
Kod A485L for	gattacaggcagaggctgatcaagatcctggca
Kod A485L rev	tgccaggatcttgatcagcctctgcctgtaatc
Kod E664I for	aagctggtgatccacatccagataacgagggat
Kod E664I rev	atccctcgttatctggatgtggatcaccagctt
Kod L485A for	gattacaggcagagggcaatcaagatcctggca
Kod L485A rev	tgccaggatcttgattgccctctgcctgtaatc
Kod L485R for	gattacaggcagaggcgtatcaagatcctggca
Kod L485R rev	tgccaggatcttgatacgcctctgcctgtaatc
Tgo A485L for	gattacaggcaacgactgatcaaaatccttgct
Tgo A485L rev	agcaaggattttgatcagtcgttgcctgtaatc
Tgo E664I for	aagctggtgatccacatccagataacgagggat
Tgo E664I rev	atccctcgttatctggatgtggatcaccagctt
Tgo L485A for	gattacaggcaacgagcaatcaaaatccttgct
Tgo L485A rev	agcaaggattttgattgctcgttgcctgtaatc
Tgo L485R for	gattacaggcaacgacgtatcaaaatccttgct
Tgo L485R rev	agcaaggattttgatacgtcgttgcctgtaatc

Templates for Extension and Fidelity Assays

4nt.1g	actattcaacttacaattgcatcaaccttataatccacttggctactgcatacgagtgtc
4nt.1g-2g-3g	ttaaactcgggtacaccttggatctccaagaatctactggctactgcatacgagtgtc
4nt.3ga	tgtctacacgcaagcttacaccattctttaacagatcatcactcgacatttatatgacaa cattaacctcggctactgcatacgagtgtc
fidelity.temp	tgtctacacgcaagcttacattaagactcgccatgttacgatctgccaagtacagccttga atcgtcactggctactgcatacgagtgtcaaaaaaaaaa

Sequence and Structural Analysis of Family B DNA polymerases

A list of family B DNA polymerases with structural coordinates available in the protein databank (PDB) was compiled by searching the PDB for family B DNA polymerases. In cases where multiple structures were present for the same enzyme, the polymerase with the most complete 3D dataset was chosen as the representative member. In total, we identified 12 unique DNA polymerases that derive from diverse evolutionary clades including Archaea, Eubacteria, and Eukarya, and the viruses that infect these organisms. The selected polymerases are: *Desulfurococcus* sp. Tok (1QQC), *Escherichia coli* DNA Polymerase II (3MAQ), Herpes Simplex virus type 1 (2GV9), Phi 29 (2PZS), *Pyrococcus abyssi* (4FM2), *Pyrococcus furiosus* (2JGU), Enterobacteria Phage RB69 (1CLQ), *Sulfolobus solfataricus* (1S5J), *Thermococcus gorgonarius* (1TGO), *Thermococcus kodakarensis* (4K8Z), *Thermococcus* sp. 9°N-7 (4K8X), Yeast pol delta (3IAY).[63,172–181] For these polymerases, we calculated the sequence and structural conservation at each amino acid position. Sequence conservation was calculated using MUSCLE. Structural conservation was calculated with PyMol by separating the proteins into their individual domains (N-terminal, exonuclease, palm, finger, and thumb) and then performing a multiple structure alignment on the individual domains. The sequence and structural analyses produced two separate histograms that were overlaid with the sequence of our model DNA polymerase (*Thermococcus* 9°N). The list of conserved sites was then narrowed to the subset of amino acid positions that are located within 10 Å of the DNA primer-template complex.

Site-Directed Mutagenesis

The polymerase genes were cloned into the pGDR11 vector and amino acid substitutions were sequentially introduced by site-directed mutagenesis.[162] Briefly, 30 ng of plasmid was combined with 5 pmol of primer and 0.1 units/μL AccuPrime DNA polymerase in 1x AccuPrime buffer. The solution was thermocycled as follows: 2 min. at 95°C followed by 16 cycles of 30 sec at 95°C, 60 sec at 55°C, and 7.5 min. at 72°C. After cycling, DpnI (5 units) was added and the solution was incubated for 1 hour at 37°C to digest the starting DNA plasmid. The undigested DNA was transformed into XL1-blue chemically competent *E. coli*, cloned, and the correct mutations were verified by sequencing the entire gene (ASU Core Facility).

Polymerase Expression and Purification

An *E. coli* strain engineered with the polymerase of interest was streaked onto solid media and a starter culture was established by inoculating LB-ampicillin liquid media with a single colony. The starter culture was used to inoculate an expression culture at OD_{600} of 0.1 in LB-ampicillin. The culture was grown at 37°C with shaking until reaching an OD_{600} of 0.8 at which time protein expression was induced with 1 mM IPTG. After an additional 4 hours of growth at 37°C, the cells were pelleted, re-suspended in nickel purification buffer [50 mM phosphate, 250 mM sodium chloride, 10% glycerol, pH 8], sonicated with 0.1 mg/mL hen egg lysozyme, and heat treated for 15 min. at 75°C. The lysate was clarified by centrifugation for 15 min. at 4,000 rpm. Polymerases were purified from the clarified lysate by binding to a nickel affinity resin. The column was washed with nickel purification buffer and eluted with nickel purification buffer supplemented with 75 mM imidazole. Eluted protein was exchanged

into storage buffer [10 mM Tris-HCl, 100 mM KCl, 1 mM DTT, 0.1 mM EDTA, pH 7.4] and concentrated using a protein concentrating spin column and stored at 4°C. The polymerase concentrations were analyzed by denaturing PAGE and normalized.

Polymerase Activity Assay

Polymerase activity assays were performed in a 10 μL reaction volume containing 5 pmol of DNA primer-template complex (see table below for the templates and reaction times used in these experiments and the materials section for the primer-template sequence). The complex was labeled at the 5' end of the DNA primer with IR800 dye. The primer-template complex was annealed in 1x ThermoPol buffer [20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1% Triton X-100, pH 8.8] by heating for 5 min. at 95°C and cooling for 10 min. at 4°C. For assays performed with tNTPs, the polymerase (1 μL) was pretreated with 1 mM $MnCl_2$ then added to the reaction mixture. For reactions with dNTPs, NTPs, or no added triphosphates, the polymerase was added directly to the reaction mixture. Last, the nucleotide triphosphates (100 μM) were added to the reaction and the solution was incubated at 55°C. The reactions were quenched in stop buffer [1x Tris-boric acid buffer, 20 mM EDTA, 7 M urea, pH 8] and denatured by incubating for 5 min. at 90°C. Reaction products were analyzed by denaturing PAGE and visualized using a LI-COR Oddysey CLx imager.

The following template and extension times were used:

xNTP	Polymerase mutant	Template	Extension time
dNTPs	varies	varies	30 min
NTPs	QGLK	4nt.1g	3 hours
tNTPs	L	4nt.3ga	1 hour
tNTPs	RI	4nt.1g-2g-3g	30 min

For PAA time courses, the reaction volume was increased to 25 μL. After the reaction commenced, 1 ug of reaction was aliquotted into 30 uL of stop buffer. Following separation by denaturing PAGE, the products were visualized using a LI-COR Oddysey CLx imager and the amount of full length and truncated products were quantified using the ImageQuant Software. Percent full-length product was calculated by diving the amount of full-length product by the sum of full-length and truncated. The data was plotted and a non-linear regression was fit to $V_2 \sim (FL_{min} + FL_{max} * t / (t + FL_{50})$ using R where FL_{min} is the minimum amount of full length product, FL_{max} is the maximum amount of full-length product, FL_{50} is half full length product, and t is time.

Polymerase Controls for Activity and Contamination

To ensure that the recombinant polymerases were functional and free of dNTP or NTP contamination, the enzymes were challenged to extend a DNA primer-template complex in the presence and absence of dNTPs using conditions described in the polymerase activity assay. Polymerase activity assays were analyzed by denaturing PAGE and imaged using a LI-COR Odyssey CLx.

RNase Digestion

To verify that the QGLK polymerases generated full-length RNA products, the extended material was digested with 2.5 μg of RNase A for 1 hour at 37°C. The reactions were quenched in stop buffer (see above) and denatured by incubating for 5 min. at 70°C before cooling to room temperature. Reaction products were analyzed by denaturing PAGE and imaged using a LI-COR Odyssey CLx.

Fidelity Analysis

Fidelity reactions were performed by extending a DNA primer-template complex in a 100 μL reaction volume containing 100 pmol of fidelity.temp and 100 pmol of PBS2.mismatch primer. The primer and template were annealed in 1x ThermoPol buffer by heating for 5 min. at 95°C and cooling for 10 min. at 4°C. The polymerase (10 uL) was added to the reaction mixture. For TNA extensions, the polymerase was pretreated with 1 mM $MnCl_2$. The reactions were initiated by addition of the nucleotide triphosphates (100 μM). Following a 4 hour incubation at 55°C, the reactions were quenched in stop buffer and denatured at 90°C for 5 min. Elongated primers were purified by denaturing PAGE, electroeluted, and concentrated using an YM-30 concentrator device.

The purified transcripts (TNA or RNA) were reverse transcribed in a final volume of 100 μL. PBS1 primer (100 pmol) was annealed to the template in 1x First Strand Buffer [50 mM Tris-HCl, 75 mM KCl, 3 mM $MgCl_2$, pH 8.3] by heating for 5 min. at 90°C and cooling for 10 min. at 4°C. Next, 500 μM dNTPs and 10 mM DTT was added and the reaction was allowed to incubate for 2 min. at 42°C. Finally, 3 mM $MgCl_2$ and 10 U/μL SuperScript II reverse transcriptase were added and the reaction was allowed to

incubate for 1 hour at 42°C. For reactions performed on TNA templates, 1.5 mM $MnCl_2$ was included in the reaction mixture.

After reverse transcription, the PCR amplified DNA (1 pmol) was ligated into a pJET vector following manufacturer's protocol. The ligated product was transformed into XL1-blue *E. coli*, cloned, and sequenced (ASU Core Facility). Sequencing results were analyzed using CLC Main Workbench. Sequences lacking the T to A watermark were discarded as they were generated from the starting DNA template rather than replicated material. The error rate for each of the nine possible substitution (for example, T → C, T → G, or T → A) was determined as follows: $\mu_{exp.\rightarrow obs.}$ = (#observed/#expected) * 1000. The total error rate was determined by summing the error rate for each substitution.

CHAPTER 5

A TWO-ENZYME TNA REPLICATION SYSTEM ENABLING IDENTIFICAITON OF NUCLEASE RESISTANT TNA APTAMERS

Introduction

Recent advances in polymerase engineering have enabled the replication of synthetic genetic polymers following a mechanism in which DNA is 'transcribed' into xenonucleic acids (XNA) and 'reverse transcribed' back into DNA. By introducing a selective amplification step into the replication cycle, functional XNA molecules have been evolved that can bind to a specific target or catalyze a chemical reaction. However, despite these advances, the process of coding and decoding genetic information in XNA is limited by the fidelity and catalytic efficiency of XNA polymerases. Here we identify the large fragment of *Geobacillus stearothermophilus* (Bst) DNA polymerase I as a threose nucleic acid (TNA) reverse transcriptase. Bst generates twice as much RT product while maintaining a three-fold lower mutation rate than Superscript II, the previous choice for TNA reverse transcription. Using the improved replication system, we identify specific TNA aptamers to human α-thrombin. When challenged to retain binding activity in the presence of nucleases TNA aptamers retain function, while RNA aptamers are degraded and lose activity. The superior efficiency and fidelity of Bst polymerase enables the identification of functional TNA molecules with biotechnological potential.

Advances in nucleic acid chemistry have enabled the development of a wide range of artificial genetic polymers with broad therapeutic and biotechnological applications. Maximizing the potential of these synthetic genetic systems requires

developing new molecular biology tools that can both generate and faithfully replicate unnatural polymers. Threose nucleic acid (TNA) (**Figure 1.2**) has received significant attention, due to the development of complete replication system by engineering natural polymerases to broaden their substrate specificity.[27,48,59–62,66,71,170] Replication proceeds by transcribing a DNA template into TNA, reverse transcribing the TNA into cDNA, and then amplifying the cDNA by PCR. However, current TNA replication systems have yet to achieve a fidelity that enables in vitro selection of functional TNA molecules.[42]

Recently an improved polymerase capable of transcribing TNA polymerase has been developed. Kod-RI has a five-fold improved fidelity, higher activity, and is capable of transcribing through complex sequence contexts as compared to its predecessor Therminator DNA polymerase. Development of a more efficient TNA polymerase prompted investigation of new reverse transcriptases capable of higher efficiency reverse transcription. Previously, an engineered version of Moloney Murine Leukemia Virus reverse transcriptase called Superscript II (SSII) was identified to reverse transcribe short stretches of threose nucleic acid. Following careful optimization of the reaction conditions, SSII emerged as a suitable TNA reverse transcriptase capable of generating significant amounts of full length cDNA from a complex library of TNA.[27] The combined fidelity of Kod-RI TNA transcription and SSII reverse transcription provides noteworthy fidelity, generating only a single misincorporations per 50 nucleobases. While a promising advance, this error rate can have detrimental effects when selecting for functional TNA molecules of lengths greater than 50 nucleobases. Identifying enzymes with higher fidelity is imperative to enhance the in vitro selection process by reducing the loss of functional genetic information through mutation.

In a previously reported screen for TNA reverse transcription activity, the large fragment from Geobacillus stearothermophilus DNA polymerase I (Bst) was found to harbor modest TNA reverse transcription activity.[66] While initially deemed less functional than SSII, Bst was later demonstrated to additionally be capable of both transcribing and reverse transcribing limited stretches of glycerol nucleic acid polymers.[141,182] This finding reinvigorated efforts to examine the functional properties of Bst on TNA templates.

Results

To better evaluate the reverse transcription activity of Bst, we directly compared Bst and SSII activity. Reverse transcription is most often examined by quantitative PCR following the generation of cDNA. While a useful strategy, this indirect readout makes it challenging to accurately quantify the fraction of template converted to cDNA. Therefore, we chose to monitor the reverse transcription directly by gel mobility analysis (**Figure 5.1A**). First, we transcribed a DNA library composed of two primer binding sites flanking a fifty nucleotide random region to generate a complex pool of TNA molecules. The TNA was recovered from its templating DNA following strand separation by denaturing gel electrophoresis. Both Bst and SSII were challenged to reverse transcribe the TNA into cDNA by extending a fluorescently labeled primer hybridized to the TNA in the presence and absence of Mn^{2+} (**Figure 5.1B**). In agreement with prior observations, SSII was capable of generating full-length cDNA only when Mn^{2+} was present. In contrast, Bst was able to generate a similar fraction of full-length product both in the presence and absence of Mn^{2+}. Bst polymerase generated twice the amount of full-length product as SSII, converting more than half of the TNA population into full-length cDNA.

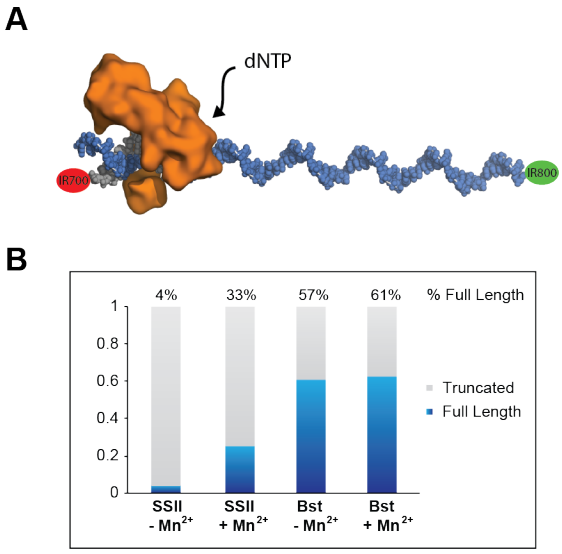

Figure 5.1. Comparison of Bst and SSII TNA Reverse Transcription. (A) Schematic of TNA reverse transcription. Utilizing a IR680 labeled DNA primer (gray), a reverse transcriptase (orange) extends the DNA primer with dNTPs utilizing a TNA template (blue). (B) Comparison of Bst and SSII TNA reverse transcriptase activity in the presence and absence of the mutagenic Mn^{2+} ion.

We hypothesized that Bst's reverse transcriptase activity could be further improved by optimizing buffer conditions and enzyme concentration. Starting from the standard 3 mM Mg^{2+} present in Thermopol buffer, we gradually increased the Mg^{2+} concentration to 13 mM. Activity correlated positively with the Mg^{2+} concentration, eventually peaking at 6 mM. The additional 3 mM Mg^{2+} doubled the amount of full-length product (**Figure 5.2A**). Independently, the unit concentration of Bst also demonstrated a positive correlation with activity, revealing an optimal concentration of 1.6 U/µL (**Figure 5.2B**). The optimized conditions combined for a 3-fold improvement in the amount of full-length cDNA product.

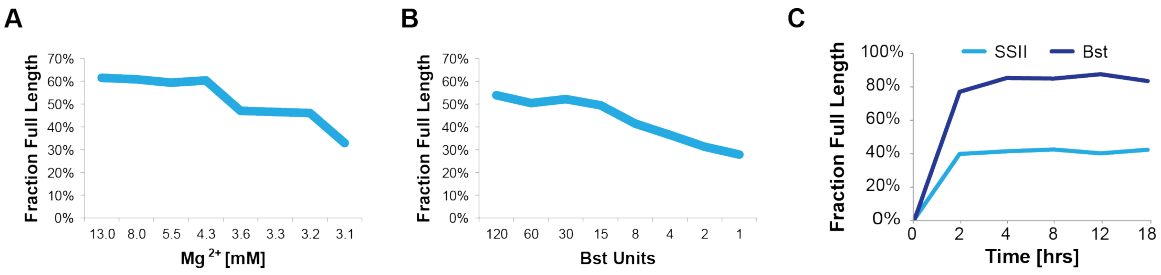

Figure 5.2. Bst Reaction Condition Optimization. Both increased Mg^{2+} (A) and enzyme (B) concentration in the reverse transcription reaction have a beneficial effect on the amount of full-length cDNA generated. (C) Comparison of full-length cDNA production over time in both polymerases' optimal conditions.

Having optimized the conditions for Bst RT activity, we compared its efficiency to SSII. To do so, we analyzed the amount of full-length cDNA both Bst and SSII generated over an 18-hour period. The amount of full-length product for both enzymes plateaued by two hours (**Figure 5.2C**). However, Bst polymerase was able to generate more than twice the amount of full-length product than SSII. Close examination revealed that SSII produced a noticeably higher amount of truncated product as indicated by a smear beneath the full-length product. This is most likely due to polymerase stalling on challenging sequence contexts and suggests that Bst has reduced sequence bias compared to SSII.

Having established that Bst is capable of generating more full-length cDNA than SSII, we analyzed the fidelity of TNA replication using Bst. We measured the aggregate fidelity of the entire replication process (DNA→TNA→DNA) (**Figure 2.6A**) by transcribing TNA from a single DNA template sequence. The TNA was recovered from its templating DNA following strand separation by denaturing gel electrophoresis and checked for purify by PCR. Since TNA is not capable of being amplified by standard

polymerases, only remaining DNA template would generate a PCR product (**Figure 5.3**). Bst was then used to reverse transcribe TNA into cDNA in the presence and absence of manganese ions. As a control, SSII was subjected to the same assay. After reverse transcription and PCR amplification, the DNA product was cloned into pJET and transformed into chemically competent E. coli. Plasmid from individual colonies was collected, sequenced, and analyzed for mutation (**Table 5.1**). Including Mn^{2+} in the RT reaction had a drastic effect on the fidelity of Bst, generating one error for every five bases incorporated. Removing Mn^{2+} from the reaction produced a 50-fold improvement in the fidelity resulting in a single error for every 250 bases incorporated, a mutation rate three-fold lower than SSII.

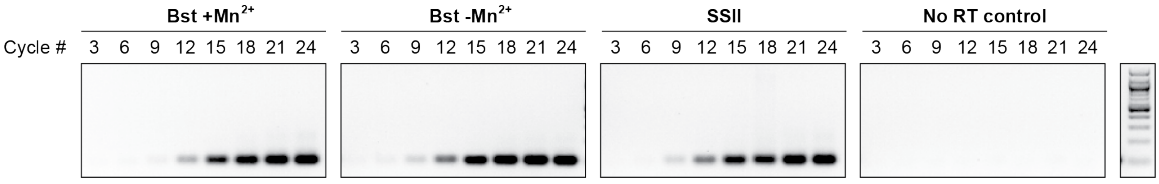

Figure 5.3. PCR Amplification of Reverse Transcribed TNA. cDNA from TNA reverse transcriptions were amplified by PCR and analyzed at multiple cycle numbers to identify the relative amounts of full length product in each reaction. Amplification of a portion of the TNA sample prior to reverse transcription (No RT) demonstrated that there was no contaminating DNA in the TNA samples. Gels were visualized by ethidium bromide staining.

Table 5.1. Fidelity of TNA Reverse Transcription with Bst DNA Polymerase and SSII Reverse Transcriptase.

Enzyme	Bst	Bst	SSII
Mn^{2+}	+	-	+
# Mutations	87	3	24
# Bases Read	400	650	1600
Mutation Rate ($x10^{-3}$)	217	4.6	15

The strong TNA RT activity and fidelity of Bst has the potential to greatly improve the enrichment during in vitro selection experiments. In vitro selection libraries used to identify nucleic acid aptamers or catalysts often contain randomized regions of 40-50 nucleobases. Since the mutation rate associated with TNA replication using SSII is approximately one mutation in every 50 bases, a typical library would receive an average of one mutation in each member of the population during each round of replication. This level of mutation will effectively reduce the enrichment for each round of in vitro selection and make it significantly more difficult to identify highly functional molecules. Bst's improved fidelity helps overcome this limitation, as one mutation would only be expected for every five members of a library with 50 nts, enabling the propagation of functional genetic information.

We then challenged the replication system with Bst to identify TNA aptamers that bind to human α-thrombin. After generating a library of 10^{10} unique TNA molecules, we subjected the library to three rounds of capillary electrophoresis SELEX (**Figure 5.4**).[183–188] After the third round, the library was cloned into pJET, transformed into chemically competent E. coli, and isolates were sequenced. Ten unique TNA sequences were chosen randomly as no two sequences were identical, generated by transcription with a

fluorescently labeled primer, purified, and tested for solution binding using a dot blot membrane partitioning strategy. In this strategy, a nitrocellulose membrane selectively captures protein and protein-aptamer complexes, while unbound aptamer passes through and is captured on the underlying nylon membrane. The binding affinities for the 10 TNA aptamers ranged from 200 to 1000 nM (**Table 5.2**). TNA aptamer T1 demonstrated the best affinity to human alpha-thrombin with a Kd of 225 nM (**Figure 5.5A**). Additionally, T1 had a high level of specificity, showing negligible binding to BSA, hemoglobin, and RNase A (**Figure 5.5B**). Deletion analysis of the constant flanking regions demonstrated that the 5' primer had little effect on activity. while removal of the 3' primer reduced affinity two-fold (**Figure 5.6**). This suggests that the 3' constant region is important either for proper folding of the aptamer or for recognition of human α-thrombin.

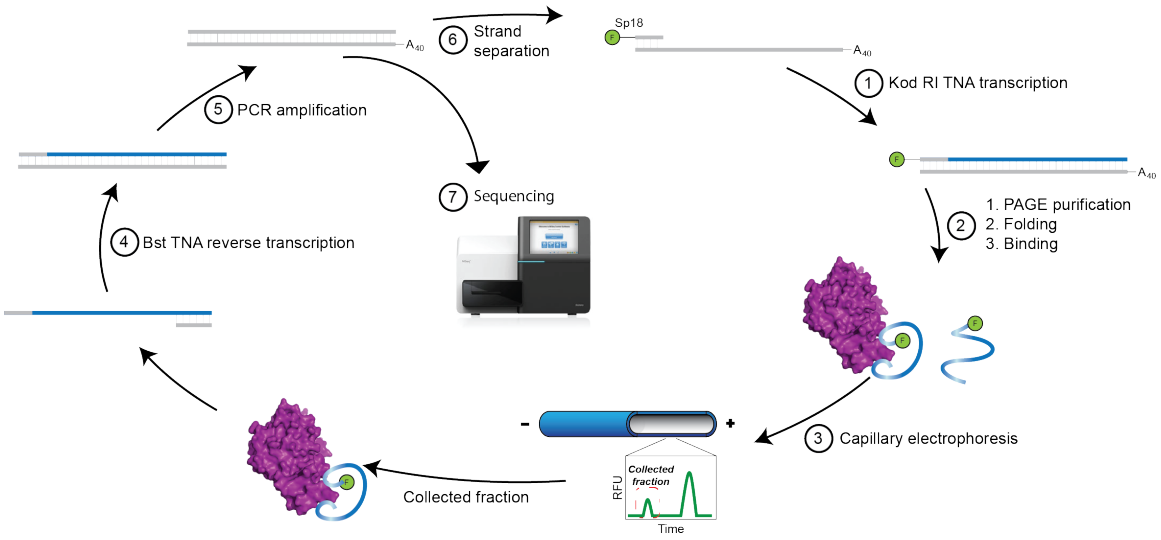

Figure 5.4. Capillary Electrophoresis SELEX Strategy. To begin, (1) Kod-RI extends a FAM-labeled DNA primer bound to a DNA template containing a 5'-PEG overhang. (2) TNA is purified by denaturing PAGE, resuspended in binding buffer, folded, and then incubated with target protein. (3) Protein-TNA complexes are separated from unbound TNA by capillary electrophoresis. (4) The collected fraction is then reverse transcribed back into DNA with Bst. (5) cDNA is then PCR amplified with a single 5'-PEG overhang primer generating two asymmetrically sized strands. (6) The longer template strand is then purified from the non-template strand by denaturing PAGE. After 3 rounds of CE-SELEX, (7) the pool is then sequenced and individual library members are tested for activity.

Table 5.2: TNA Aptamers Characterized. Constant primer binding sites are colored: PBS8 (blue), PBS2 (green), and PBS7 (purple).

Name	Kd	Sequence
T1	911	GTCCCCTTGGGGATACCACCTGAGTATAGTTAAGTAGTTTGTGAGTAATTGTTATTGTAGAATATTGATG
T2	624	GTCCCCTTGGGGATACCACCTGTGGAGGTTGTTAGGTGGGTAGAGGGAAGAGATAGGTAGTGTGATGGGG
T3	889	GTCCCCTTGGGGATACCACCGTGTAGGAGTGGTTGGTGAGTTGGTGGGGTTATGTTGAAGGTGGTGAGG
T4	671	GTCCCCTTGGGGATACCACCAGTGAATAGTAGTGGGTAGTTGGTGAGTGTGAGTGAGGGAGGGAAGGGTG
T5	783	GTCCCCTTGGGGATACCACCTTGGAGGAGGAGCGATAGGGGGAGTGTGATATGGGTTTGTTGGAGTTGGG
T6	1043	GTCCCCTTGGGGATACCACCTTGGTAGGAAGGGGTGTTGATGAGTAAGGGGTGGTAGTAAGGTGTGG
T7	1183	GTCCCCTTGGGGATACCACCCAGGGTTGGAAGAGTGGGTGGTAGTTGGTTGTGGTGTTGTTGTGGTGGG
T8	799	GTCCCCTTGGGGATACCACCGAGGAGGGGTGAGGGGTGAAATGGGAGTGTAAGGGTTTTGAGTGTTGTGG
T9	843	GTCCCCTTGGGGATACCACCTCGATTGGCAGTTGCGTCGTTATTGCGGTTTGATTATATGAGTGGCGATT
T1-87	225	GTCCCCTTGGGGATACCACCTGAGTATAGTTAAGTAGTTTGTGAGTAATTGTTATTGTAGAATATTGATGTCTCAATGCACTGACGATCC
T1-2	513	GACACTCGTATGCAGTAGCCTGAGTATAGTTAAGTAGTTTGTGAGTAATTGTTATTGTAGAATATTGATG
T1-27	2824	GACACTCGTATGCAGTAGCCTGAGTATAGTTAAGTAGTTTGTGAGTAATTGTTATTGTAGAATATTGATGTCTCAATGCACTGACGATCC

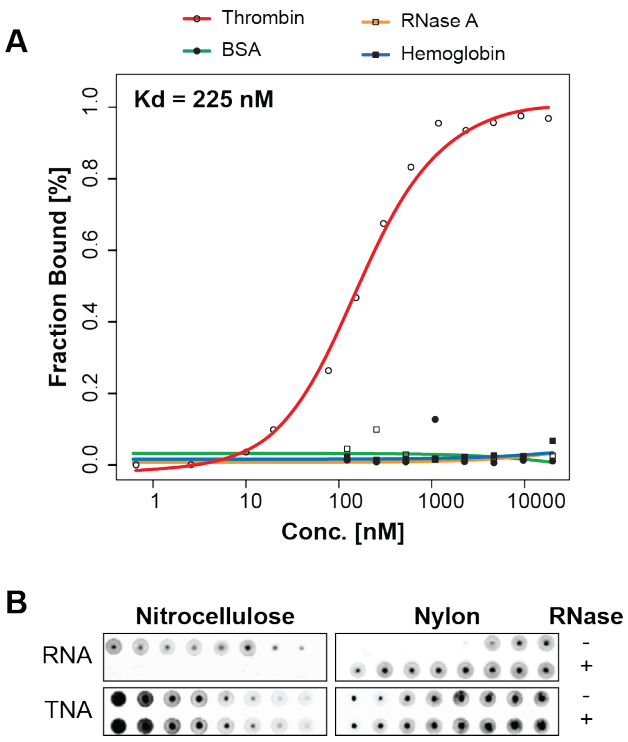

Figure 5.5. Affinity, Specificity, and Nuclease Stability of a Thrombin TNA Aptamer. (A) Solution binding isotherms of aptamer T1 against human α-thrombin, BSA, hemoglobin, and RNase A. (B) Comparison of TNA and RNA aptamer function in the presence and absence of RNase A. Signal on the nitrocellulose membrane represents aptamer bound to thrombin, while signal on the nylon membrane represents free aptamer. Membranes were visualized by IR fluorescence.

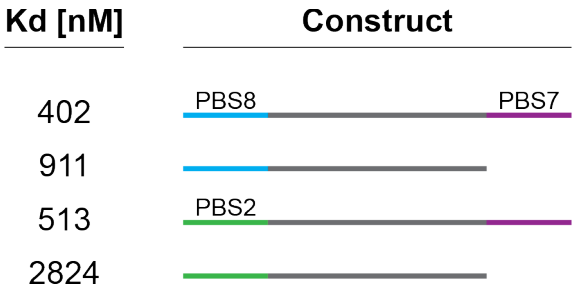

Figure 5.6. T1 Truncation Analysis. Modified T1 aptamers missing the PBS7 constant region or replacing PBS8 constant region with a different sequence were tested for solution binding affinities.

To analyze the nuclease resistance of TNA aptamer T1, we sought to have a control aptamer comprised of an RNA backbone. We generated RNA aptamers to thrombin using the same strategy. After 3 rounds of CE-SELEX, RNA aptamer R1 was found to have a comparable affinity of 200 nM to T1 (**Figure 5.7 and Table 5.3**). We then subjected both the T1 and the R1 aptamers to nuclease treatment with RNase A and analyzed their binding capabilities before and after treatment. As expected, T1 demonstrated binding in both the presence and absence of RNase, while R1 binding was completely abolished after RNase treatment (**Figure 5.5B**).

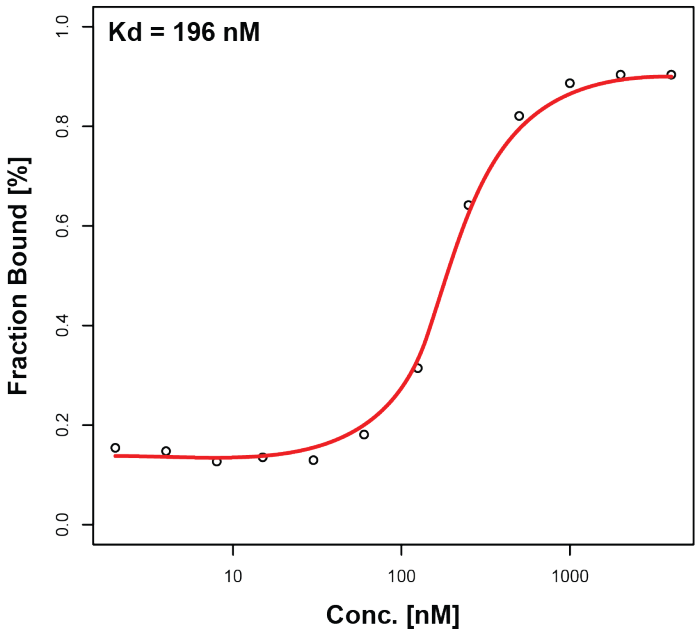

Figure 5.7. R1 Aptamer Solution Binding Affinity and Specificity. Solution binding isotherms of RNA aptamer R1 against human α-thrombin.

Table 5.3: RNA Aptamers Characterized. PBS8 constant region is colored blue.

Name	Kd	Sequence
R1	188	GTCCCCTTGGGGATACCACCGGTGAGGGATAAATTGATGTGATG AGGTGGTCGGGGGTGTACCAGTATC
R2	190	GTCCCCTTGGGGATACCACCTAGCCCAGGGTACAATTGGAGCAT GGGGGGTTTGTCGAGAGATGCGTATT
R3	184	GTCCCCTTGGGGATACCACCTTTGGGACGGGCGGCTGCCCTTGT ACCAGGTGGAGTTATCAGCAAGATC

Discussion

In summary, we demonstrated that Bst DNA polymerase is an efficient TNA reverse transcriptase. The combined TNA replication system of Kod-RI transcription and Bst reverse transcription has enabled the identification of TNA aptamers that contain all four natural bases. The selected aptamers bind to human α-thrombin with modest affinities ranging from 200 to 1000 nM. The TNA aptamer T1 binds human α-thrombin selectively and retains full function in the presence of nucleases. This work demonstrates the potential therapeutic and biotechnological applications of functional TNA molecules. Additionally, the replication system has potential in many non-biological applications like information storage and nanotechnology.[27,46,78,189] Finally, Bst provides an improved starting point for polymerase engineering experiments aimed to further improve the reverse transcription capabilities of unnatural genetic polymers.

Experimental

General Information

DNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), purified by denaturing polyacrylamide gel electrophoresis, electroeluted, concentrated by ethanol precipitation, and quantified by UV absorbance. dNTPs were purchased from Sigma (St. Louis, MO). TNA triphosphates (tNTPs) were obtained by chemical synthesis as previously described. SuperScript II reverse transcriptase and Proteinase K were purchased from Invitrogen (Grand Island, NY). CloneJET PCR cloning kit was purchased from Fermentas (Waltham, MA). Deep Vent (exo-) and Bst DNA polymerase large fragment were obtained from NEB (Ipswich, MA). Ultracell YM-10 concentrators were purchased from Millipore (Billerica, MA).

DNA Sequences

The sequences below are written from 5' → 3' using IDT abbreviations for modified positions. N is standard IUPAC nomenclature denoting a random position composed of an equal number of all 4 nucleobases.

Reverse Transcription Oligos

PBS2	gacactcgtatgcagtagcc
PBS1	tgtctacacgcaagcttaca
PBS2 IR800	/5IRD800//iSp18/gacactcgtatgcagtagcc
PBS1 IR700	/5IRD700//iSp18/tgtctacacgcaagcttaca
PBS2.mismatch	/56-FAM/cttttaagaaccggacgaacgacactcgtttgcagtagcc
extra.primer	cttttaagaaccggacgaac
fidelity.temp	tgtctacacgcaagcttacattaagactcgccatgttacgatctgccaagtacagccttgaatcgt cactggctactgcatacgagtgtc/3InvdT/
L3	tgtctacacgcaagcttaca-N50-ggctactgcatacgagtgtc

Selection Oligos

PBS7	ggatcgtcagtgcattgaga
PBS7 PEG	(aaca)10/iSp18/ggatcgtcagtgcattgaga
PBS8	gtccccttggggataccacc
PBS8 FAM	/56-FAM/gtccccttggggataccacc
L3 78	ggatcgtcagtgcattgaga -N50- ggtggtatccccaaggggac

Generation of DNA Template for TNA Transcription

Single-stranded DNA templates were generated by PCR. Each PCR reaction (1,000 µL total volume) contained 1,000 pmol of each DNA primer (PBS7 PEG Long and PBS8) and 50 pmol of double stranded DNA template. Reactions were performed in 1x Thermopol buffer [20 mM Tris-HCl, 10 mM $(NH4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1% Triton X-100, pH 8.8] supplemented with 400 nM dNTPs, and 0.04 U/µL DV exo-. The solution was divided into 50 µL reactions and thermocycled as follows: 2 min at 95°C followed by 15 cycles of 15 sec at 95°C, 15 sec at 58°C, and 2.25 min at 72°C. The reactions were combined, concentrated by lyophilization, and supplemented with 5 mM EDTA and 50% (w/V) urea. The solution was heat denatured for 5 min at 95°C, and the PEG-modified strand was purified by 10% denaturing urea PAGE. The corresponding band was excised, electroeluted, and concentrated by ethanol precipitation. The precipitated pellet was re-suspended in 20 µL of nuclease free water and quantified by UV absorbance.

TNA Transcription

TNA template was generated in a 100 μL reaction volume containing 50 pmol of DNA primer-template complex. The complex was labeled at the 5' end of the DNA primer with IR800 dye. The primer-template complex was annealed in 1x ThermoPol buffer [20 mM Tris-HCl, 10 mM $(NH_4)2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1% Triton X-100, pH 8.8] by heating for 5 min at 95°C and cooling for 10 min at 4°C. Kod-RI TNA polymerase (10 μL) was pretreated with 1 mM $MnCl_2$ then added to the reaction mixture. The nucleotide triphosphates (100 μM) were then added to the reaction and the solution was incubated for 4 hours at 55°C. The reactions were treated with 10 μL of Proteinase K for 1 hour at 55°C. The reactions were quenched by adding 50% (w/V) urea and a final concentration of 10 mM EDTA and denatured by incubating for 5 min. at 95°C. TNA was purified by 20% denaturing PAGE, electroeluted, and concentrated on a YM-10 microcentrifuge column. TNA concentration was quantified by UV.

TNA Reverse Transcription

For SSII reactions, 5 pmol PBS1 IR700 primer was annealed to 10 pmol TNA template in a final volume of 10 μL in 1× First Strand Buffer [50 mM Tris-HCl, 75 mM KCl, 3 mM $MgCl_2$, pH 8.3] by heating for 5 min at 95°C and cooling for 10 min at 4°C. 500 μM dNTPs and 10 mM DTT was added to the primer template complex and incubated for 2 min at 42°C. Finally, 3 mM $MgCl_2$, 1.5 mM $MnCl_2$ (in Mn^{2+} positive reactions), and 10 U/μL SuperScript II reverse transcriptase was added to the reaction and incubated for up to 8 h at 42°C.

For Bst reactions, 5 pmol PBS1 IR700 primer was annealed to 10 pmol TNA template in a final volume of 10 μL in 1× Thermopol buffer by heating for 5 min at 95°C and cooling for 10 min at 4°C. 500 μM dNTPs and 3 mM $MgCl_2$ was added to the primer template complex. Finally, 1.5 mM $MnCl_2$ (in Mn^{2+} positive reactions), and 1.6 U/μL Bst DNA polymerase was added to the reaction and incubated for up to 8 h at 55°C. Reaction products were analyzed by 10% denaturing PAGE and imaged using a LI-COR Odyssey CLx. The amount of full-length product was quantified in the LI-COR software.

Fidelity Analysis

cDNA from reverse transcribed TNA was PCR amplified (1 pmol) and ligated into a pJET vector following manufacturer's protocol. The ligated product was transformed into chemically competent XL1-blue *E. coli*, cloned, and sequenced (ASU Core Facility). Sequencing results were analyzed using CLC Main Workbench. Sequences lacking the T to A watermark were discarded as they were generated from the starting DNA template rather than replicated material. The error rate was determined by dividing the number of substitutions from the expected sequence over the total number of bases sequenced.

Capillary Electrophoresis SELEX

For each round of selection, DNA template and FAM-labeled TNA were generated as described above. FAM-labeled TNA was suspended in 20 μL of 1x Binding Buffer [10 mM Tris, 100 mM NaCl, 1 mM $MgCl_2$, pH 8.3] and folded by heating for 5 min at 95°C then cooling for 10 min at room temperature. CE was performed on a Beckman PA800 plus with an eCAP Neutral Capillary (50 μm ID, 60 cm long). The mobility of the TNA sample was determined in the absence of protein to determine the

collection window. CE proceeded by first rinsing the capillary with 100 mM NaOH, then distilled water, followed by 1x TGK buffer [177 mM glycine, 25 mM tris base, 5 mM KH_2PO_4, pH 8.3] for 2 min each at 20 psi. TNA same was injected for 4 s at 1 psi and then separated for 20 min at 30 kV constant. Afterwards, 500 nM thrombin was added to the TNA sample and incubated for 30 min at room temperature. CE-SELEX proceeded with 5 injections collecting the bound protein complex from the unbound peak into 50 μL of 1x Thermopol buffer. The collected TNA was then reverse transcribed for 2 hours by adding the remaining reagents for Bst TNA reverse transcription as described above in a 100 μL final volume.

After 3 rounds of selection, isolates were identified as described in the fidelity analysis section.

RNA Primer Extension and Reverse Transcriptions

For RNA primer extension, the protocol is identical as in the TNA transcription section with the substitution of Tgo-QGLK enzyme and NTP substrates for Kod-RI and tNTPs. Reverse transcription was performed with Superscript II following manufacturer's recommended protocol.

Dot Blot Solution Binding Analysis

IR800-labeled TNA molecules were synthesized by primer extension and purified by denaturing polyacrylamide gel electrophoresis as described above Folded structures were incubated with thrombin poised at concentrations spanning the expected Kd (typically 1 nM–10 μM) at room temperature. After 1 hour, the protein-bound TNA molecules were partitioned away from the unbound fractions using vacuum to pass the solution through a layer of nitrocellulose and nylon membranes. Both membranes were

quantified by imaging on the LICOR, and dissociation constants were calculated with R using nonlinear least-squares regression to the equation $F = F_{min} + F_{max} * C/(C + Kd)$ where F is the fraction bound, F_{min} and F_{max} are the minimum and maximum bound fractions, Kd is the dissociation constant, and C is the protein concentration.

CHAPTER 6

CONCLUSIONS AND FUTURE OUTLOOKS

Starting from a two-enzyme system composed of Therminator DNA polymerase and Superscript II reverse transcriptase that was capable of replicating 3 nucleotide systems, this thesis expanded TNA replication to four nucleotides by identifying new polymerase variants with improved activity and fidelity. To do so, we began by identifying gain-of-function mutations in the *Thermococcus* sp. 9°N scaffold that greatly improved polymerase activity. We then optimized polymerase function using a new strategy called scaffold sampling in which gain-of-function mutations are tested in homologous protein architectures. By doing so, subtle differences between tertiary structures position the gain-of-function mutations in beneficial environments effectively fine tuning polymerase function. The new TNA polymerase, Kod-RI (A485R and E664I) has high activity and fidelity generating only 15 errors per 1000 nucleobase incorporations. We then identify the large fragment from *Bacillus stearothermophilus* DNA polymerase I to be an effective TNA reverse transcriptase. The combination of the two polymerases, Kod-RI transcription and Bst reverse transcription, are an efficient and faithful replication system generating fewer than 4 errors per 1000 nucleobase incorporations. Overall, the Kod-RI and Bst replication system is 16 times more faithful and has much higher activity that its predecessor.

While the Kod-RI and Bst replication system is the most advanced TNA replication system to date, both its fidelity and activity pale in comparison to natural polymerases. The error rate of Taq DNA polymerase is approximately 1 misincorporation per 10 million bases, which is 40,000 times more faithful than the TNA replication

system. While the current TNA replication system is incapable of supporting life through genomic replication, it is useful in other applications. In vitro selection for functional nucleic acid molecules, for example, is typically performed with libraries containing randomized regions of 40-50 nucleobases. Since the mutation rate associated with TNA replication is one error in every 250 incorporations, only 1 of every 5 library members would receive a mutation per round of replication effectively enabling the propagation of functional genetic information.

Using this replication system, we developed TNA aptamers to human alpha-thrombin with modest affinity and high specificity. The TNA aptamers have a high intrinsic nuclease resistance providing them potential in biotechnological applications and molecular medicine and have the potential to replace natural nucleic acid polymers in aptamer development.

Looking into the future, XNA polymers have potential in a wide range of applications from biopharmaceuticals and molecular medicine to nanotechnology and materials engineering. This is due to the chemical and physical diversity seen among different XNA chemistries. These applications range from the development of aptamers that can bind to target proteins or small molecules, to enzymes and small oligos capable of regulating gene expression inside living cells, to generating expanded conformational geometries and chemical properties in nucleic acid nanotechnology that are currently inaccessible using solely natural polymers. To date there have been four XNA aptamers to biologically relevant targets.[41,42] Additionally there are a number of XNA enzymes capable of hydrolyzing RNA backbones.[155] PNA and LNA in particular have been used extensively in combination with natural polymers to regulate gene expression via a

siRNA mechanism. Finally, GNA polymers have been shown to specifically fold into nucleic acid origami structures with mirror image symmetry.[190] This demonstrates the potential conformational space that XNAs can access as compared to natural polymers. Further, unique combinations of different XNAs and natural polymers can further increase the accessible conformational space of XNA nanomaterial.

While the polymerases presented in this work provide a replication system capable of supporting in vitro evolution, continued development of polymerases with higher activity, processivity, fidelity, and specificity for their XNA substrates and templates will be needed. These XNA replication systems have additional potential in synthetic biology within living cells where engineered expression networks can be made completely orthogonal by coupling to XNA genetic information.

These applications along with those that have yet to be developed would be greatly aided by better understanding of the structural properties of XNA polymers. The understanding of the conformational space and constraints of XNA polymers would allow for better development of XNA nanotechnology as well as the generation of computer models for the rational design of both biological and non-biological XNA technologies. Structural data on the polymerases that replicate XNA polymers would enable further improvements to XNA replications systems towards highly efficient and specific activities for their XNA targets.

References

1. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science (80-.).* **249,** 505–10 (1990).

2. Joyce, G. F. In vitro evolution of nucleic acids. *Curr. Opin. Struct. Biol.* **4,** 331–6 (1994).

3. Herdewijn, P. & Marliere, P. Toward Safe Genetically Modified Organisms through the Chemical Diversification of Nucleic Acids. *Chem. Biodivers.* **6,** 791 (2011).

4. Appella, D. H. Non-natural nucleic acids for synthetic biology. *Curr. Opin. Chem. Biol.* **13,** 687–96 (2009).

5. Brudno, Y. & Liu, D. R. Recent Progress Toward the Templated Synthesis and Directed Evolution of Sequence-Defined Synthetic Polymers. *Chem. Biol.* **16,** 265–276 (2009).

6. Kool, E. T. Active site tightness and substrate fit in DNA replication. *Annu. Rev. Biochem.* **71,** 191–219 (2002).

7. Zhang, S., Switzer, C. & Chaput, J. C. The resurgence of acyclic nucleic acids. *Chem. Biodivers.* **7,** 245–58 (2010).

8. Henry, A. A. & Romesberg, F. E. Beyond A, C, G and T: Augmenting nature's alphabet. *Curr. Opin. Chem. Biol.* **7,** 727–733 (2003).

9. Kool, E. T. Replacing the nucleobases in DNA with designer molecules. *Acc. Chem. Res.* **35,** 936–43 (2002).

10. Piccirilli, J. A., Krauch, T., Moroney, S. E. & Benner, S. A. Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. *Nature* **343,** 33–37 (1990).

11. Krueger, A. T. & Kool, E. T. Model systems for understanding DNA base pairing. *Curr. Opin. Chem. Biol.* **11,** 588–594 (2007).

12. Lavergne, T., Malyshev, D. A. & Romesberg, F. E. Major groove substituents and polymerase recognition of a class of predominantly hydrophobic unnatural base pairs. *Chemistry (Easton).* **18,** 1231–1239 (2012).

13. Mitsui, T. *et al.* An unnatural hydrophobic base pair with shape complementarity between pyrrole-2-carbaldehyde and 9-methylimidazo[(4,5)-b]pyridine. *J. Am. Chem. Soc.* **125,** 5298–5307 (2003).

14. Moran, S., Ren, R. X.-F. F., Rumney, S. & Kool, E. T. Difluorotoluene, a nonpolar isostere for thymine, codes specifically and efficiently for adenine in DNA replication. *J. Am. Chem. Soc.* **119,** 2056–2057 (1997).

15. Malyshev, D. A., Young, J. S., Ordoukhanian, P. & Romesberg, F. E. PCR with an expanded genetic alphabet. *J. Am. Chem. Soc.* **131,** 14620–14621 (2009).

16. Yang, Z., Chen, F., Chamberlin, S. G. & Benner, S. A. Expanded genetic alphabets in the polymerase chain reaction. *Angew. Chemie* **49,** 177–180 (2010).

17. Hollenstein, M., Hipolito, C. J., Lam, C. H. & Perrin, D. M. A self-cleaving DNA enzyme modified with amines, guanidines and imidazoles operates independently of divalent metal cations (M2+). *Nucleic Acids Res.* **37,** 1638–1649 (2009).

18. Hollenstein, M., Hipolito, C. J., Lam, C. H. & Perrin, D. M. A DNAzyme with three protein-like functional groups: Enhancing catalytic efficiency of M2+-independent RNA cleavage. *ChemBioChem* **10,** 1988–1992 (2009).

19. Vaish, N. K., Fraley, A. W., Szostak, J. W. & McLaughlin, L. W. Expanding the structural and functional diversity of RNA: analog uridine triphosphates as candidates for in vitro selection of nucleic acids. *Nucleic Acids Res.* **28,** 3316–3322 (2000).

20. Vaught, J. D. *et al.* Expanding the chemistry of DNA for in vitro selection. *J. Am. Chem. Soc.* **132,** 4141–4151 (2010).

21. Delaney, J. C. *et al.* Efficient replication bypass of size-expanded DNA base pairs in bacterial cells. *Angew. Chemie* **48,** 4524–7 (2009).

22. Krueger, A. T., Peterson, L. W., Chelliserry, J., Kleinbaum, D. J. & Kool, E. T. Encoding phenotype in bacteria with an alternative genetic set. *J. Am. Chem. Soc.* **133,** 18447–51 (2011).

23. Endy, D. Foundations for engineering biology. *Nature* **438,** 449–453 (2005).

24. Reardon, S. Visions of synthetic biology. *Science (80-.).* **333,** 1242–3 (2011).

25. Pennisi, E. Two steps forward for synthetic biology. *Science (80-.).* **325,** 928–9 (2009).

26. Benner, S. A. & Sismour, A. M. Synthetic biology. *Nat. Rev. Genet.* **6,** 533–43 (2005).

27. Yu, H., Zhang, S., Dunn, M. R. & Chaput, J. C. An Efficient and Faithful in Vitro Replication System for Threose Nucleic Acid. *J. Am. Chem. Soc.* **135,** 3583–91 (2013).

28. Keefe, A. D. & Cload, S. T. SELEX with modified nucleotides. *Curr. Opin. Chem. Biol.* **12,** 448–56 (2008).

29. Watts, J. K., Deleavey, G. F. & Damha, M. J. Chemically modified siRNA: tools and applications. *Drug Discov. Today* **13,** 842–55 (2008).

30. Watts, J. K. *et al.* Differential stability of 2′F-ANA•RNA and ANA•RNA hybrid duplexes: Roles of structure, pseudohydrogen bonding, hydration, ion uptake and flexibility. *Nucleic Acids Res.* **38,** 2498–511 (2010).

31. Damha, M. J. *et al.* Hybrids of RNA and arabinonucleic acids (ANA and 2'F-ANA) are substrates of ribonuclease H. *J. Am. Chem. Soc.* **120,** 12976–7 (1998).

32. Noronha, A. M. *et al.* Synthesis and biophysical properties of arabinonucleic acids (ANA): Circular dichroic spectra, melting temperatures, and ribonuclease H susceptibility of ANA-RNA hybrid duplexes. *Biochemistry* **39,** 7050–62 (2000).

33. Kasahara, Y., Irisawa, Y., Ozaki, H., Obika, S. & Kuwahara, M. 2',4'-BNA/LNA aptamers: CE-SELEX using a DNA-based library of full-length 2'-O,4'-C-methylene-bridged/linked bicyclic ribonucleotides. *Bioorg. Med. Chem. Lett.* **23,** 1288–1292 (2013).

34. Maier, T., Przylas, I., Strater, N., Herdewijn, P. & Saenger, W. Reinforced HNA backbone hydration in the crystal structure of a decameric HNA/RNA hybrid. *J. Am. Chem. Soc.* **127,** 2937–43 (2005).

35. Martín-Pintado, N. *et al.* The solution structure of double helical arabino nucleic acids (ANA and 2'F-ANA): effect of arabinoses in duplex-hairpin interconversion. *Nucleic Acids Res.* **40,** 9329–39 (2012).

36. Kataoka, M., Kouda, Y., Sato, K., Minakawa, N. & Matsuda, A. Highly efficient enzymatic synthesis of 3'-deoxyapionucleic acid (apioNA) having the four natural nucleobases. *Chem. Commun.* **47,** 8700–2 (2011).

37. Cummins, L. L. *et al.* Characterization of fully 2'-modified oligoribonucleotide hetero- and homoduplex hybridization and nuclease sensitivity. *Nucleic Acids Res.* **23,** 2019–24 (1995).

38. Kurreck, J., Wyszko, E., Gillen, C. & Erdmann, V. A. Design of antisense oligonucleotides stabilized by locked nucleic acids. *Nucleic Acids Res.* **30,** 1911–8 (2002).

39. Deleavey, G. F. & Damha, M. J. Designing chemically modified oligonucleotides for targeted gene silencing. *Chem. Biol.* **19,** 937–954 (2012).

40. Ferrari, N., Bergeron, D. & Tedeschi, A. Characterization of Antisense Oligonucleotides Comprising 2'-Deoxy-2'-Fluoro-β-D-Arabinonucleic Acid (FANA) Specificity, Potency, and Duration of Activity. *Ann. N. Y. Acad. Sci.* **102,** 91–102 (2006).

41. Pinheiro, V. B. *et al.* Synthetic genetic polymers capable of heredity and evolution. *Science (80-.).* **336,** 341–4 (2012).

42. Yu, H., Zhang, S. & Chaput, J. C. Darwinian evolution of an alternative genetic system provides support for TNA as an RNA progenitor. *Nat. Chem.* **4,** 183–7 (2012).

43. Joyce, G. F. Directed evolution nucleic acid enzymes. *Annu. Rev. Biochem.* **73,** 791– 836 (2004).

44. Szostak, J. W. In vitro genetics. *Trends Biochem. Sci.* **17,** 89–93 (1992).

45. Schöning, K. *et al.* Chemical etiology of nucleic acid structure: the alpha-threofuranosyl-(3'->2') oligonucleotide system. *Science (80-.).* **290,** 1347–1351 (2000).

46. Chaput, J. C., Yu, H. & Zhang, S. The emerging world of synthetic genetics. *Chem. Biol.* **19,** 1360–71 (2012).

47. Wu, X., Guntha, S., Ferencic, M., Krishnamurthy, R. & Eschenmoser, A. Base-pairing systems related to TNA: α-threofuranosyl oligonucleotides containing phosphoramidate linkages. *Org. Lett.* **4,** 1279–82 (2002).

48. Eschenmoser, A. Chemical etiology of nucleic acid structure. *Science (80-.).* **284,** 2118–24 (1999).

49. Eschenmoser, A. The TNA-family of nucleic acid systems: properties and prospects. *Orig. Life Evol. Biosph.* **34,** 277–306 (2004).

50. Ebert, M.-O., Mang, C., Krishnamurthy, R., Eschenmoser, A. & Jaun, B. The structure of a TNA-TNA complex in solution: NMR study of the octamer duplex derived from α-(L)-Threofuranosyl-(3′-2′)-CGAATTCG. *J. Am. Chem. Soc.* **130,** 15105–15 (2008).

51. Wilds, C. J., Wawrzak, Z., Krishnamurthy, R., Eschenmoser, A. & Egli, M. Crystal structure of a B-form DNA duplex containing (L)-α-threofuranosyl (3'→2') nucleosides: A four-carbon sugar is easily accommodated into the backbone of DNA. *J. Am. Chem. Soc.* **9,** 1347–51 (2002).

52. Engelhart, A. E. & Hud, N. V. Primitive genetic polymers. *Cold Spring Harb. Perspect. Biol.* **2,** a002196 (2010).

53. Joyce, G. F. Bit by bit: The darwinian basis of life. *PLoS Biol.* **10,** e1001323 (2012).

54. Orgel, L. E. A Simpler Nucleic Acid. *Science (80-.).* **290,** 1306–7 (2000).

55. Wilson, D. S. & Szostak, J. W. In vitro selection of functional nucleic acids. *Annu. Rev. Biochem.* **68,** 611–47 (1999).

56. Gold, L., Polisky, B., Uhlenbeck, O. & Yarus, M. Diversity of oligonucleotide functions. *Annu. Rev. Biochem.* **64,** 763–97 (1995).

57. Ni, X., Castanares, M., Mukherjee, A. & Lupold, S. E. Nucleic acid aptamers: clinical applications and promising new horizons. *Curr. Med. Chem.* **18,** 4206–14 (2011).

58. Keefe, A. D., Pai, S. & Ellington, A. Aptamers as therapeutics. *Nat. Rev. Drug Discov.* **9,** 537–50 (2010).

59. Chaput, J. C. & Szostak, J. W. TNA synthesis by DNA polymerases. *J. Am. Chem. Soc.* **125,** 9274–5 (2003).

60. Kempeneers, V., Vastmans, K., Rozenski, J. & Herdewijn, P. Recognition of threosyl nucleotides by DNA and RNA polymerases. *Nucleic Acids Res.* **31,** 6221–6 (2003).

61. Horhota, A. *et al.* Kinetic analysis of an efficient DNA-dependent TNA polymerase. *J. Am. Chem. Soc.* **127,** 7427–34 (2005).

62. Ichida, J. K., Horhota, A., Zou, K., McLaughlin, L. W. & Szostak, J. W. High fidelity TNA synthesis by Therminator polymerase. *Nucleic Acids Res.* **33,** 5219–25 (2005).

63. Bergen, K., Betz, K., Welte, W., Diederichs, K. & Marx, A. Structures of KOD and 9°N DNA polymerases complexed with primer template duplex. *Chembiochem* **14,** 1058–62 (2013).

64. Brudno, Y., Birnbaum, M. E., Kleiner, R. E. & Liu, D. R. An in vitro translation, selection and amplification system for peptide nucleic acids. *Nat. Chem. Biol.* **6,** 148–55 (2010).

65. Ichida, J. K. *et al.* An in vitro selection system for TNA. *J. Am. Chem. Soc.* **127,** 2802–3 (2005).

66. Chaput, J. C., Ichida, J. K. & Szostak, J. W. DNA polymerase-mediated DNA synthesis on a TNA template. *J. Am. Chem. Soc.* **125,** 856–7 (2003).

67. Zhang, S. & Chaput, J. C. Synthesis and enzymatic incorporation of α-L-threofuranosyl adenine triphosphate (tATP). *Bioorg. Med. Chem. Lett.* **23,** 1447–1449 (2013).

68. Arezi, B. & Hogrefe, H. Novel mutations in Moloney Murine Leukemia Virus reverse transcriptase increase thermostability through tighter binding to template-primer. *Nucleic Acids Res.* **37,** 473–81 (2009).

69. Gardner, A. F. & Jack, W. E. Acyclic and dideoxy terminator preferences denote divergent sugar recognition by archaeon and Taq DNA polymerases. *Nucleic Acids Res.* **30,** 605–13 (2002).

70. Gardner, A. F. & Jack, W. E. Determinants of nucleotide sugar recognition in an archaeon DNA polymerase. *Nucleic Acids Res.* **27,** 2545–53 (1999).

71. Zhang, S. & Chaput, J. C. Synthesis of threose nucleic acid (TNA) triphosphates and oligonucleotides by polymerase-mediated primer extension. *Curr. Protoc. Nucleic Acid Chem.* 44.54.1 (2013).

72. Zhang, S. & Chaput, J. C. Synthesis of threose Nucleic Acid (TNA) phosphoramidite monomers and oligonucleotide polymers. *Curr. Protoc. Nucleic Acid Chem.* 44.51.1 (2012). doi:10.1002/0471142700.nc0451s50

73. Cantor, C. R. *Principles of Nucleic Acid Chemistry*. (Springer-Verlag, 1984).

74. Skelly, J. V, Edwards, K. J., Jenkins, T. C. & Neidle, S. Crystal structure of an oligonucleotide duplex containing G.G base pairs: influence of mispairing on DNA backbone conformation. *Proc. Natl. Acad. Sci.* **90,** 804 (1993).

75. Staiger, N. & Marx, A. A DNA polymerase with increased reactivity for ribonucleotides and C5-modified deoxyribonucleotides. *Chembiochem* **11,** 1963–6 (2010).

76. Kelley, S. O. & Barton, J. K. DNA-mediated electron transfer from a modified base to ethidium: pi-stacking as modulator of reactivity. *Chem. Biol.* **5,** 413–25 (1998).

77. Innis, M. A., Myambo, K. B., Gelfand, D. H. & Brow, M. A. DNA sequencing with Thermus aquaticus DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Proc. Natl. Acad. Sci.* **85,** 9436 (1988).

78. Lutz, J.-F., Ouchi, M., Liu, D. R., Sawamoto, M. & Pinheiro, V. B. Sequence-Controlled Polymers. *Science (80-.).* **336,** 341–4 (2012).

79. Joyce, G. F. The antiquity of RNA-based evolution. *Nature* **418,** 214–21 (2002).

80. Latimer, L. J. P. & Lee, J. S. Ethidium bromide does not fluoresce when intercalated adjacent to 7-deazaguanine in duplex DNA. *J. Biol. Chem.* **266,** 13849–51 (1991).

81. Li, H. & Seela, F. Parallel and antiparallel DNA: fuorescence quenching of ethidium bromide by 7-deazapurines. *Nucleosides. Nucleotides Nucleic Acids* **24,** 865–8 (2005).

82. Albà, M. M. Replicative DNA polymerases. *Genome Biol.* **2,** 3 (2001).

83. Steitz, T. A. DNA Polymerases: Structural Diversity and Common Mechanisms. *J. Biol. Chem.* **274,** 17395–17398 (1999).

84. Braithwaite, D. K. & Ito, J. Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res.* **21,** 787–802 (1993).

85. Cann, I. K. O. & Ishino, Y. Archaeal DNA replication: Identifying the pieces to solve a puzzle. *Genetics* **152,** 1249–1267 (1999).

86. Moon, A. F. *et al.* The X family portrait: Structural insights into biological functions of X family polymerases. *DNA Repair (Amst).* **6,** 1709–1725 (2007).

87. Sale, J. E., Lehmann, A. R. & Woodgate, R. Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat. Publ. Gr.* **13,** 141–152 (2012).

88. Kornberg, A. The biologic synthesis of deoxyribonucleic acid. 665–680 (1959).

89. Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. & Steitz, T. A. Structure of large fragment of Escherichia coli DNA polymerase I complexed with dTMP. *Nature* **313,** 762–766 (1985).

90. Beese, L. S., Derbyshire, V. & Steitz, T. a. Structure of DNA polymerase I Klenow fragment bound to duplex DNA. *Science (80-.).* **260,** 352–355 (1993).

91. Mönttinen, H. A. M., Ravantti, J. J., Stuart, D. I. & Poranen, M. M. Automated structural comparisons clarify the phylogeny of the right-hand-shape polymerases. *Mol. Biol. Evol.* **31,** (2014).

92. Beard, W. A. & Wilson, S. H. Structural design of a eukaryotic DNA repair polymerase: DNA polymerase beta. *Mutat. Res.* **460,** 231–244 (2000).

93. Castro, C. *et al.* Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nat. Struct. Mol. Biol.* **16,** 212–218 (2009).

94. Castro, C. *et al.* Two proton transfers in the transition state for nucleotidyl transfer catalyzed by RNA- and DNA-dependent RNA and DNA polymerases. *Proc. Natl. Acad. Sci.* **104,** 4267–4272 (2007).

95. Werner, F. & Grohmann, D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.* **9,** 85–98 (2011).

96. Rodriguez, A. C., Park, H.-W., Mao, C. & Beese, L. S. Crystal structure of a pol alpha family DNA polymerase from the hyperthermophilic archaeon Thermococcus sp. 9 degrees N-7. *J. Mol. Biol.* **299,** 447–62 (2000).

97. Kong, H., Kucera, B. & Jack, W. E. Characterization of a DNA polymerase from the Hyperthermophile archaea Thermococcus litoralis. *J. Biol. Chem.* **268,** 1965–75 (1993).

98. Lundberg, K. S. *et al.* High-fidelity amplification using a thermostable DNA polymerase isolated from Pyrococcus furiosus. *Gene* **108,** 1–6 (1991).

99. Fa, M., Radeghieri, A., Henry, A. A. & Romesberg, F. E. Expanding the substrate repertoire of a DNA polymerase by directed evolution. *J. Am. Chem. Soc.* **126,** 1748–54 (2004).

100. Xia, G. *et al.* Directed evolution of novel polymerase activities: mutation of a DNA polymerase into an efficient RNA polymerase. *Proc. Natl. Acad. Sci.* **99,** 6597–602 (2002).

101. Glick, E., Vigna, K. L. & Loeb, L. A. Mutations in human DNA polymerase eta motif II alter bypass of DNA lesions. *EMBO J.* **20,** 7303–7312 (2001).

102. Vichier-Guerre, S., Ferris, S., Auberger, N., Mahiddine, K. & Jestin, J. L. A population of thermostable reverse transcriptases evolved from Thermus aquaticus DNA polymerase I by phage display. *Angew. Chemie* **45,** 6133–7 (2006).

103. Ghadessy, F. J., Ong, J. L. & Holliger, P. Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl. Acad. Sci.* **98,** 4552–7 (2001).

104. Ghadessy, F. J. *et al.* Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution. *Nat. Biotechnol.* **22,** 755–59 (2004).

105. Ramsay, N. *et al.* CyDNA: Synthesis and replication of highly Cy-Dye substituted DNA by an evolved polymerase. *J. Am. Chem. Soc.* **132,** 5096–104 (2010).

106. Camps, M., Naukkarinen, J., Johnson, B. P. & Loeb, L. A. Targeted gene evolution in Escherichia coli using a highly error-prone DNA polymerase I. *Proc. Natl. Acad. Sci.* **100,** 9727–9732 (2003).

107. Patel, P. H., Kawate, H., Adman, E., Ashbach, M. & Loeb, L. A. A Single Highly Mutable Catalytic Site Amino Acid Is Critical for DNA Polymerase Fidelity. *J. Biol. Chem.* **276,** 5044–51 (2001).

108. Chelliserrykattil, J., Cai, G. & Ellington, A. D. A combined in vitro/in vivo selection for polymerases with novel promoter specificities. *BMC Biotechnol.* **1,** 13 (2001).

109. Klenow, H. & Henningsen, I. Selective elimination of the exonuclease activity of the deoxyribonucleic acid polymerase from Escherichia coli B by limited proteolysis. *Proc. Natl. Acad. Sci.* **65,** 168–175 (1970).

110. Lawyer, F. C., Stoffel, S., Saiki, R. K. & Chang, S. K. High level expression, purification and enzymatic characterization of full length Thermus aquaticus DNA polymerase and truncated from deficient in 5 to 3 exonuclease activity. *J. Biol. Chem.* **2,** 275–287 (1993).

111. Wang, Y. *et al.* A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. *Nucleic Acids Res.* **32,** 1197–207 (2004).

112. Mullis, K. *et al.* Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* **51,** 263–73 (1986).

113. Motz, M. *et al.* Elucidation of an archaeal replication protein network to generate enhanced PCR enzymes. *J. Biol. Chem.* **277,** 16179–16188 (2002).

114. Pavlov, A. R., Belova, G. I., Kozyavkin, S. A. & Slesarev, A. I. Helix-hairpin-helix motifs confer salt resistance and processivity on chimeric DNA polymerases. *Proc. Natl. Acad. Sci.* **99,** 13510–13515 (2002).

115. Kermekchiev, M. B., Kirilova, L. I., Vail, E. E. & Barnes, W. M. Mutants of Taq DNA polymerase resistant to PCR inhibitors allow DNA amplification from whole blood and crude soil samples. *Nucleic Acids Res.* **37,** 1–14 (2009).

116. Baar, C. *et al.* Molecular breeding of polymerases for resistance to environmental inhibitors. *Nucleic Acids Res.* **39,** 1–12 (2011).

117. D'Abbadie, M. *et al.* Molecular breeding of polymerases for amplification of ancient DNA. *Nat. Biotechnol.* **25,** 939–943 (2007).

118. Gloeckner, C., Sauter, K. B. M. & Marx, A. Evolving a thermostable DNA polymerase that amplifies from highly damaged templates. *Angew. Chemie* **46,** 3115–7 (2007).

119. Kornberg, A. & Baker, T. *DNA replication*. (W H Freeman & Company, 1992).

120. Astatke, M., Ng, K., Grindley, N. D. & Joyce, C. M. A single side chain prevents Escherichia coli DNA polymerase I (Klenow fragment) from incorporating ribonucleotides. *Proc. Natl. Acad. Sci.* **95,** 3402–3407 (1998).

121. Malyshev, D. A. *et al.* Efficient and sequence-independent replication of DNA containing a third base pair establishes a functional six-letter genetic alphabet. *Proc. Natl. Acad. Sci.* **109,** 12005–12010 (2012).

122. Gao, J., Liu, H. & Kool, E. T. Expanded-size bases in naturally sized DNA: Evaluation of steric effects in Watson-Crick pairing. *J. Am. Chem. Soc.* **126,** 11826–31 (2004).

123. Delaney, J. C. *et al.* High-fidelity in vivo replication of DNA base shape mimics without Watson-Crick hydrogen bonds. *Proc. Natl. Acad. Sci.* **100,** 4469–73 (2003).

124. Liu, H. *et al.* A four-base paired genetic helix with expanded size. *Science (80-.).* **302,** 868–71 (2003).

125. Tyler, R. C. *et al.* Comparison of cell-based and cell-free protocols for producing target proteins from the Arabidopsis thaliana genome for structural studies. *Proteins* **59,** 633–43 (2005).

126. Mizukami, S., Kim, T. W., Helquist, S. A. & Kool, E. T. Varying DNA base-pair size in subangstrom increments: Evidence for a loose, not large, active site in low-fidelity Dpo4 polymerase. *Biochemistry* **45,** 2772–2778 (2006).

127. Kim, T. W., Delaney, J. C., Essigmann, J. M. & Kool, E. T. Probing the active site tightness of DNA polymerase in subangstrom increments. *Proc. Natl. Acad. Sci.* **102,** 15803–15808 (2005).

128. Johnson, S. J. & Beese, L. S. Structures of mismatch replication errors observed in a DNA polymerase. *Cell* **116,** 803–16 (2004).

129. Huang, M. M., Arnheim, N. & Goodman, M. F. Extension of base mispairs by Taq DNA polymerase: implications for single nucleotide discrimination in PCR. *Nucleic Acids Res.* **20,** 4567–4573 (1992).

130. Miller, H. & Grollman, A. P. Kinetics of DNA polymerase I (Klenow fragment exo-) activity on damaged dna templates: Effect of proximal and distal template damage on DNA synthesis. *Biochemistry* **36,** 15336–15342 (1997).

131. Carver, T. E., Hochstrasser, R. A. & Millar, D. P. Proofreading DNA: recognition of aberrant DNA termini by the Klenow fragment of DNA polymerase I. *Proc. Natl. Acad. Sci.* **91,** 10670–10674 (1994).

132. Tabor, S. & Richardson, C. C. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci.* **84,** 4767–4771 (1987).

133. Tabor, S. & Richardson, C. C. Selective inactivation of the exonuclease activity of bacteriophage T7 DNA polymerase by in vitro mutagenesis. *J. Biol. Chem.* **264,** 6447–6458 (1989).

134. Arezi, B., Hansen, C. J. & Hogrefe, H. H. Efficient and high fidelity incorporation of dye-terminators by a novel archaeal DNA polymerase mutant. *J. Mol. Biol.* **322,** 719–729 (2002).

135. Evans, S. J. *et al.* Improving dideoxynucleotide-triphosphate utilisation by the hyper-thermophilic DNA polymerase from the archaeon Pyrococcus furiosus. *Nucleic Acids Res.* **28,** 1059–1066 (2000).

136. Tabor, S. & Richardson, C. C. A single residue in DNA polymerases of the Escherichia coli DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. *Proc. Natl. Acad. Sci.* **92,** 6339–6343 (1995).

137. Brown, J. A. & Suo, Z. Unlocking the sugar 'steric gate' of DNA polymerases. *Biochemistry* **50,** 1135–42 (2011).

138. Patel, P. H. & Loeb, L. A. DNA polymerase active site is highly mutable: evolutionary consequences. *Proc. Natl. Acad. Sci.* **97,** 5095–100 (2000).

139. Ong, J. L., Loakes, D., Jaroslawski, S., Too, K. & Holliger, P. Directed evolution of DNA polymerase, RNA polymerase and reverse transcriptase activity in a single polypeptide. *J. Mol. Biol.* **361,** 537–50 (2006).

140. Wanninger-Weiß, C., Di Pasquale, F., Ehrenschwender, T., Marx, A. & Wagenknecht, H.-A. Nucleotide insertion and bypass synthesis of pyrene- and BODIPY-modified oligonucleotides by DNA polymerases. *Chem. Commun.* 1443–5 (2008). doi:10.1039/b718002k

141. Tsai, C.-H., Chen, J. & Szostak, J. W. Enzymatic synthesis of DNA on glycerol nucleic acid templates without stable duplex formation between product and template. *Proc. Natl. Acad. Sci.* **104,** 14598–603 (2007).

142. Peng, C. G. & Damha, M. J. Probing DNA polymerase activity with stereoisomeric 2′-fluoro-β-D-arabinose (2′F-araNTPs) and 2′-fluoro-β-D-ribose (2′F-rNTPs) nucleoside 5′-triphosphates. *Can. J. Chem.* **86,** 881–91 (2008).

143. Peng, C. G. & Damha, M. J. Polymerase-directed synthesis of 2'-deoxy-2'-fluoro-β-d-arabinonucleic acids. *J. Am. Chem. Soc.* **129,** 5310–1 (2007).

144. Leitzel, J. C. & Lynn, D. G. Template-directed ligation: from DNA towards different versatile templates. *Chem. Rec.* **1,** 53–62 (2001).

145. Chaput, J. C. & Switzer, C. Non-enzymatic transcription of an isoG-isoC base pair. *J. Am. Chem. Soc.* **122,** 12866–12867 (2000).

146. Chaput, J. C., Sinha, S. & Switzer, C. 5-Propynyluracil·diaminopurine: an efficient base-pair for non-enzymatic transcription of DNA. *Chem. Commun.* 1568–9 (2002). doi:10.1039/b204535d

147. Rosenbaum, D. M. & Liu, D. R. Efficient and Sequence-Specific DNA-Templated Polymerization of Peptide Nucleic Acid Aldehydes. *J. Am. Chem. Soc.* **125,** 13924–13925 (2003).

148. Heuberger, B. O. & Switzer, C. Nonenzymatic oligomerization of RNA by TNA templates. *Org. Lett.* **8,** 5809–11 (2006).

149. Koppitz, M., Nielsen, P. E. & Orgel, L. E. Formation of oligonucleotide-PNA-chimeras by template-directed ligation. *J. Am. Chem. Soc.* **120,** 4563–9 (1998).

150. Ura, Y., Beierle, J. M., Leman, L. J., Orgel, L. E. & Ghadiri, M. R. Self-assembling sequence-adaptive peptide nucleic acids. *Science (80-.).* **325,** 73–7 (2009).

151. Niu, J., Hili, R. & Liu, D. R. Enzyme-free translation of DNA into sequence-defined synthetic polymers structurally unrelated to nucleic acids. *Nat. Chem.* **5,** 282–92 (2013).

152. Loakes, D. & Holliger, P. Polymerase engineering: towards the encoded synthesis of unnatural biopolymers. *Chem. Commun.* 4619–31 (2009). doi:10.1039/b903307f

153. Chen, T. & Romesberg, F. E. Directed polymerase evolution. *FEBS Lett.* **588,** 219–29 (2014).

154. Yang, Z., Chen, F., Alvarado, J. B. & Benner, S. A. Amplification, mutation, and sequencing of a six-letter synthetic genetic system. *J. Am. Chem. Soc.* **133,** 15105–15112 (2011).

155. Taylor, A. I. *et al.* Catalysts from synthetic genetic polymers. *Nature* (2014). doi:10.1038/nature13982

156. Kimoto, M., Yamashige, R., Matsunaga, K., Yokoyama, S. & Hirao, I. Generation of high-affinity DNA aptamers using an expanded genetic alphabet. *Nat. Biotechnol.* **31,** 453–7 (2013).

157. Sefah, K. *et al.* In vitro selection with artificial expanded genetic information systems. *Proc. Natl. Acad. Sci.* **111,** 1449–1454 (2014).

158. Sousa, R. Structural and mechanistic relationships between nucleic acid polymerases. *Trends Biochem. Sci.* **21,** 186–190 (1996).

159. Chen, R. & Jeong, S.-S. Functional prediction: identification of protein orthologs and paralogs. *Protein Sci.* **9,** 2344–53 (2000).

160. Stemmer, W. P. C. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370,** 389–91 (1994).

161. McCullum, E. O. & Chaput, J. C. Transcription of an RNA aptamer by a DNA polymerase. *Chem. Commun.* **20,** 2938–40 (2009).

162. Li, L., Shakhnovich, E. I. & Mirny, L. A. Amino acids determining enzyme-substrate specificity in prokaryotic and eukaryotic protein kinases. *Proc. Natl. Acad. Sci.* **100,** 4463–8 (2003).

163. Morley, K. L. & Kazlauskas, R. J. Improving enzyme properties: when are closer mutations better? *Trends Biotechnol.* **23,** 231–7 (2005).

164. Connolly, B. A. Recognition of deaminated bases by archaeal family-B DNA polymerases. *Biochem. Soc. Trans.* **37,** 65–8 (2009).

165. Fogg, M. J., Pearl, L. H. & Connolly, B. a. Structural basis for uracil recognition by archaeal family B DNA polymerases. *Nat. Struct. Biol.* **9,** 922–7 (2002).

166. Wardle, J. *et al.* Uracil recognition by replicative DNA polymerases is limited to the archaea, not occurring with bacteria and eukarya. *Nucleic Acids Res.* **36,** 705–11 (2008).

167. De Vega, M. & Salas, M. A highly conserved Tyrosine residue of family B DNA polymerases contributes to dictate translesion synthesis past 8-oxo-7,8-dihydro-2'-deoxyguanosine. *Nucleic Acids Res.* **35,** 5096–107 (2007).

168. Cozens, C., Pinheiro, V. B., Vaisman, A., Woodgate, R. & Holliger, P. A short adaptive path from DNA to RNA polymerases. *Proc. Natl. Acad. Sci.* **109,** 8067–72 (2012).

169. Ross, M. G. *et al.* Characterizing and measuring bias in sequence data. *Genome Biol.* **14,** R51 (2013).

170. Dunn, M. R. *et al.* DNA Polymerase-Mediated Synthesis of Unbiased Threose Nucleic Acid (TNA) Polymers Requires 7-Deazaguanine To Suppress G:G Mispairing during TNA Transcription. *J. Am. Chem. Soc.* **137,** 4014–7 (2015).

171. Hashimoto, H. *et al.* Crystal structure of DNA polymerase from hyperthermophilic archaeon Pyrococcus kodakaraensis KOD1. *J. Mol. Biol.* **306,** 469–77 (2001).

172. Zhao, Y. *et al.* Crystal structure of an archaebacterial DNA polymerase. *Structure* **7,** 1189–99 (1999).

173. Wang, F. & Yang, W. Structural insight into translesion synthesis by DNA Pol II. *Cell* **139,** 1279–89 (2009).

174. Liu, S. *et al.* Crystal structure of the herpes simplex virus 1 DNA polymerase. *J. Biol. Chem.* **281,** 18193–200 (2006).

175. Berman, A. J. *et al.* Structures of phi29 DNA polymerase complexed with substrate: the mechanism of translocation in B-family polymerases. *EMBO J.* **26,** 3494–505 (2007).

176. Gouge, J., Ralec, C., Henneke, G. & Delarue, M. Molecular recognition of canonical and deaminated bases by P. abyssi family B DNA polymerase. *J. Mol. Biol.* **423,** 315–36 (2012).

177. Kim, S. W., Kim, D.-U. U., Kim, J. K., Kang, L.-W. W. & Cho, H.-S. S. Crystal structure of Pfu, the high fidelity DNA polymerase from Pyrococcus furiosus. *Int. J. Biol. Macromol.* **42,** 356–61 (2008).

178. Shamoo, Y. & Steitz, T. A. Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. *Cell* **99,** 155–66 (1999).

179. Savino, C. *et al.* Insights into DNA replication: the crystal structure of DNA polymerase B1 from the archaeon Sulfolobus solfataricus. *Structure* **12,** 2001–8 (2004).

180. Hopfner, K.-P. *et al.* Crystal structure of a thermostable type B DNA polymerase from Thermococcus gorgonarius. *Proc. Natl. Acad. Sci.* **96,** 3600–5 (1999).

181. Swan, M. K., Johnson, R. E., Prakash, L., Prakash, S. & Aggarwal, A. K. Structural basis of high-fidelity DNA synthesis by yeast DNA polymerase delta. *Nat. Struct. Mol. Biol.* **16,** 979–86 (2009).

182. Chen, J. J. *et al.* Enzymatic primer-extension with glycerol-nucleoside triphosphates on DNA templates. *PLoS One* **4,** e4949 (2009).

183. Drabovich, A. P., Berezovski, M., Okhonin, V. & Krylov, S. N. Selection of smart aptamers by methods of kinetic capillary electrophoresis. *Anal. Chem.* **78,** 3171–3178 (2006).

184. Mendonsa, S. D. & Bowser, M. T. In vitro selection of aptamers with affinity for neuropeptide Y using capillary electrophoresis. *J. Am. Chem. Soc.* **127,** 9382–9383 (2005).

185. Mendonsa, S. D. & Bowser, M. T. In vitro evolution of functional DNA using capillary electrophoresis. *J. Am. Chem. Soc.* **126,** 20–21 (2004).

186. Mendonsa, S. D. & Bowser, M. T. In Vitro Selection of High-Affinity DNA Ligands for Human IgE Using Capillary Electrophoresis. *Anal. Chem.* **76,** 5387–5392 (2004).

187. Ruff, P., Pai, R. B. & Storici, F. Real-Time PCR-Coupled CE-SELEX for DNA Aptamer Selection. *ISRN Mol. Biol.* **2012,** 1–9 (2012).

188. Tok, J., Lai, J., Leung, T. & Li, S. F. Y. Selection of aptamers for signal transduction proteins by capillary electrophoresis. *Electrophoresis* **31,** 2055–2062 (2010).

189. Pinheiro, V. B. & Holliger, P. Towards XNA nanotechnology: New materials from synthetic genetic polymers. *Trends Biotechnol.* **32,** 321–328 (2014).

190. Zhang, R. S., McCullum, E. O. & Chaput, J. C. Synthesis of two mirror image 4-helix junctions derived from glycerol nucleic acid. *J. Am. Chem. Soc.* **130,** 5846–7 (2008).