UNIVERSITY OF CALIFORNIA,
IRVINE

Evolution and Structural Elucidation of TNA Polymerases

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Pharmacological Sciences

by

Victoria Maola Gross

Dissertation Committee:
Professor John Chaput, Chair
Professor Brian Paegel
Professor Celia Goulding

2026

DEDICATION

To

my parents - my role models for hard work

to my husband- for standing by me with patience and love through it all

to my daughter - for giving everything new meaning

TABLE OF CONTENTS

LIST OF FIGURES

Page

ACKNOWLEDGMENTS

I am especially grateful to my advisor, Professor John Chaput. You are a role model for hard work, curiosity, creativity, and perseverance. Thank you for challenging me to tackle hard questions, and for being the first person to clap when we found the answers. Your unwavering encouragement, guidance, and belief in me has brought me to where I am today. I am forever thankful for your support, willingness to talk science or life, and for the many years of great memories. It has been both an honor and a privilege to study under you and be part of your lab family.

I have been supported by many past and current members of the Chaput group along my graduate career. Dr. Eric Yik gave me many of my first tutorials in the lab and worked in parallel with me on the droplet-based evolution work—thank you for making the long days fun, for patiently answering my many questions, and for always being willing to grab a coffee. To Mohammad, Erica, Alex and Joy: I lucked out with the most kindhearted lab mates. Thank you for always being willing to lend a helping hand, crack a joke, and for always cheering each other on. Dr. Nicholas Chim has been with me from my first day to my last day, always showing up with kindness, patience, and a welcoming heart. I am deeply grateful to you for providing me with such a strong scientific foundation, and I will forever miss harvesting crystals together. I would also like to thank Arlene Ngor for her dedication to organizing and maintaining supplies, and for helping troubleshoot equipment in the direst of moments.

I would like to thank my committee members, Dr. Brian Paegel and Dr. Celia Goulding, for their invaluable guidance and thoughtful feedback throughout my doctoral journey. Thank you for challenging me, for your time, patience, and the care you have shown in helping me grow both intellectually and professionally.

I would also like to thank the staff at the Advanced Light Source (ALS), Stanford Synchrotron Radiation Light Source (SSRL), and the National Synchrotron Light Source II (NSLS) for technical assistance.

I am deeply grateful for the financial support that made this work possible, including funding from the Henry Wood Elliot Foundation, the Vertex Pharmaceuticals Fellowship, and the UROP and UROP/CALIT2 Interdisciplinary Research Team Fellowships. I would also like to thank sources of financial support awarded to the Chaput lab: National Science Foundation Grants CHM: 2001434, MCB: 1946312, and 2433788, Division of Chemistry of Life's Processes (CLP) and Genetic Mechanisms (GM), and the University of California.

Figures 1 and 3 of the introduction are a reproduction of the material from Nikoomanzar, A., Chim, N., Yik, E. J. & Chaput, J. C. Engineering polymerases for applications in synthetic biology. *Q. Rev. Biophys.* 53, e8 (2020), used with permission under CC-BY.

Chapter 1 of this dissertation is a reprint of the material as it appears in Maola, V.A†., Yik, E.J†., Hajjar, M. Lee, J., Holguin, M., Quijano, R., Nguyen, K., Ho, K., Medina, J., Chim, N., Chaput, JC*. Directed evolution of a highly efficient TNA polymerase achieved by homologous recombination. *Nature Catalysis* 7, 1173–1185 (2024)., used with permission from Springer

Nature. The co-authors listed in this publication are Dr. Eric Yik, Mohammad Hajjar, Manuel Holguin, Riley Quijano, Kalvin Nguyen, Katherine Ho, Jenny Medina, Dr. Nicholas Chim, and Professor John C. Chaput.

Chapter 2 of this dissertation is a reprint of the material as it appears in Yik, E. J[†]., Maola, V. A[†]. & Chaput, J. C. Engineering TNA polymerases through iterative cycles of directed evolution. *Methods Enzymol.* 691, 29-59 (2023). https://doi.org/10.1016/bs.mie.2023.04.014, used with permission from Elsevier. The co-authors listed in this publication are Dr. Eric Yik and Professor John C. Chaput.

Chapter 3 of this dissertation is a reprint of the material as it appears in Hajjar, M.[†], Maola, V.A.[†], Lee, J., Holguin, M.J., Quijano, R.N., Nguyen, K.K., Ho, K.L., Medina, J.V., Botello-Cornejo, E., Barpuzary, B., Chim, N., Chaput, J.C. Directed evolution of a TNA polymerase identifies independent paths to fidelity and catalysis. *Nature Communications*, 2025. doi.org/10.1038/s41467-025-67652-1, used with permission from Springer Nature Limited. The co-authors listed in this publication are Mohammad Hajjar, Joy Lee, Manuel Holguin, Riley Quijano, Kalvin Nguyen, Katherine Ho, Jenny Medina, Elionel Botello-Cornejo, Bhawna Barpuzary, Dr. Nicholas Chim, and Professor John C. Chaput.

VITA

Victoria Maola Gross

2026 Ph.D. in Pharmacological Sciences, University of California, Irvine
2024 M.S in Pharmacological Sciences, University of California, Irvine
2020 B.A in Anthropology, University of California, Irvine

FIELD OF STUDY
Molecular biology & biochemistry; protein engineering; directed evolution; x-ray crystallography

PUBLICATIONS
1. Medina EL†, Maola VA†, Hajjar M, Ko GK, Ho EJ, Horton AR, Chim N, and Chaput JC. Rapid evolution of a highly efficient RNA polymerase by homologous recombination. ***Nature Chemical Biology****, 2026*. doi.org/10.1038/s41589-025-02124-7
2. Hajjar M†, Maola VA†, Lee J†, Holguin MJ, Quijano RN, Nguyen KK, Ho KL, Medina JV, Botello-Cornejo E, Barpuzary B, Chim N, Chaput JC. Directed evolution of a TNA polymerase identifies independent paths to fidelity and catalysis. ***Nature Communications***, 2025. doi.org/10.1038/s41467-025-67652-1
3. Maola VA†, Yik EJ†, Hajjar M, Lee JJ, Holguin MJ, Quijano RN, Nguyen KK, Ho KL, Medina JV, Chim N, and Chaput JC. Directed evolution of a highly efficient TNA polymerase achieved by homologous recombination**.** ***Nature Catalysis****,* 2024. doi.org/10.1038/s41929-024-01233-1
4. Yik EJ†, Maola VA†, Chaput JC. Engineering TNA polymerases through iterative cycles of directed evolution. ***Methods in Enzymology***, 2023. doi.org/10.1016/bs.mie.2023.04.014
5. Li Q, Maola VA, Chim N, Hussain J, Lozoya-Colinas A, and Chaput JC. Synthesis and Polymerase Recognition of Threose Nucleic Acid Triphosphates Equipped with Diverse Chemical Functionalities. ***J. Am. Chem. Soc.***, 2021. **doi.org/**10.1021/jacs.1c08649
6. Mikami A, Datta D, Kundu J, Kumar V, Erande N, Medina E, Maola, VA, Lee EM, Matsuda S, Harp J, Egli M, Chaput JC, Manoharan M. Chiral-4′-methyl-α-(L)-threofuranosyl nucleic acids (MTNA): Synthesis, structure, binding affinity, nuclease stability, and polymerase recognition (Manuscript in preparation).
7. Hajjar M, Horton A, Maola VA, Chim N, Chaput JC. How Custom Polymerases are Driving Innovation in Synthetic Biology. **Curr. Opin. Chem. Biol.** (Manuscript in preparation).

PATENT APPLICATIONS
1. U.S. Patent Application No.: PCT/US2024/011595 filed on Jan 16, 2024. Engineered Polymerase for Threose Nucleic Acid Synthesis. Inventors: Chaput JC, Maola VA, Yik EJ.
2. U.S. Patent Application Serial No.: 63/864,960 filed on August 15, 2025. Engineered RNA Polymerase. Inventors: Chaput JC, Medina EL, Maola VA.

AWARDS AND FUNDING
- Henry Wood Elliot Award, Spring 2025
- UROP/CALIT2 Interdisciplinary Research Team Fellow, Winter-Spring 2024
- Vertex Scholar, Spring 2023
- UROP Fellow, Chaput Lab Team Research, Summer 2022

ABSTRACT OF THE DISSERTATION

Evolution and Structural Elucidation of TNA Polymerases
by
Victoria Maola Gross
Doctor of Philosophy in Pharmacological Sciences
University of California, Irvine, 2026
Professor John Chaput, Chair

DNA polymerases are central to biology and biotechnology due to their ability to faithfully propagate genetic information. Expanding the substrate scope of these enzymes beyond natural nucleic acids represents a major challenge in enzyme engineering and is essential to the practical use of new classes of genetic polymers. This thesis describes the directed evolution, technological development, and structural analysis of polymerases capable of synthesizing α-L-threofuranosyl nucleic acid (TNA), a synthetic genetic polymer with a noncanonical sugar backbone that is resistant to nucleases and acid-mediated degradation.

To overcome the inherent incompatibility between natural polymerases and TNA substrates, an integrated evolutionary strategy was employed. Starting from a homologous recombination library, iterative rounds of selection and random mutagenesis were used to traverse the polymerase fitness landscape and identify mutation combinations that collectively enable efficient TNA synthesis. This campaign yielded polymerase variant 10–92, which catalyzes TNA synthesis with near-natural performance, achieving catalytic rates of ~1 nt/s and fidelity > 99%. Structural analysis of the closed ternary complex revealed widespread structural remodeling to accommodate the altered geometry of TNA, highlighting the role of distributed mutations in enabling new enzymatic functions.

Central to this effort was the development of a recombination-based diversification strategy, designed to interface with an established droplet-based microfluidic screening platform. This workflow enabled the generation of a highly diverse polymerase library and reliable recovery of active variants, supporting iterative rounds of evolution. These advances streamlined the enrichment and regeneration of functional variants, and established a general framework that can be adapted to a variety of DNA-modifying enzymes.

Finally, by capturing and characterizing multiple evolutionary intermediates along the evolutionary trajectory of the 10-92 TNA polymerase, this thesis provides mechanistic insight into one example of how new enzymatic activities emerge. Structural and biochemical analyses reveal that improvements in catalytic efficiency and replication fidelity arise from distinct molecular bases, challenging the longstanding view that accuracy and catalysis are coupled events. Together, these findings establish recombination-guided evolution, high-throughput screening, and structural interrogation as powerful and complementary tools for expanding the enzymatic recognition of synthetic genetic polymers, with implications for synthetic biology, biotechnology, and the study of enzyme evolution.

INTRODUCTION

DNA Synthesis by Natural Polymerases

DNA is the cornerstone of life. Its ability to encode a vast array of genetic information from simple organisms to complex life makes it an exciting and mysterious technology. Central to the significance of DNA is its ability to be faithfully replicated from parent to daughter strands during cell division, ensuring the inheritance of traits across generations. In the 1950s, Arthur Kornberg discovered the first DNA polymerase from *Escherichia coli (E. coli)*, establishing DNA polymerases as the enzymes responsible for genome replication during cell division and marking them among the most essential molecular machines in biology[1].

DNA polymerases follow a primer extension mechanism where a single strand of parental DNA serves as a template for the synthesis of a complementary daughter strand[2]. In this process, the growing daughter strand functions as a primer and is extended in a 5′-3′ direction by the addition of deoxyribonucleoside triphosphates (dNTPs) to the terminal 3′ hydroxyl group (Fig. 1). This reaction requires divalent metal ions, typically Mg^{2+}, to coordinate the incoming dNTP and activate the 3′ hydroxyl group for phosphodiester bond formation[3]. Since their initial discovery, many other DNA polymerases have been identified and subsequently classified into seven families - A, B, C, D, X, Y, and RT -categorized based on function and structure.

Speed and fidelity are critical parameters for DNA synthesis in rapidly dividing cells. For each nucleotide incorporation, a polymerase must distinguish the correct nucleoside triphosphate from an excess of incorrect and non-cognate (NTP) substrates. Due to their functional roles, the rate and fidelity of DNA synthesis can vary widely between different DNA polymerases.

Replicative DNA polymerases found in A- and B-families have rates that can exceed 100 nt s^{-1} and intrinsic fidelities in the range of one error in 10^5–10^6 incorporation events[5,6]. In addition, many polymerases have 3′-5′ exonuclease proofreading activity that exists as a separate domain or a tightly bound subunit, which can remove non-complementary nucleotides after phosphodiester bond formation (Fig. 1)[6,7]. These domains increase the fidelity of DNA synthesis by 10-fold (10^6–10^7 nt s^{-1}) relative to polymerases lacking a proofreading domain[6,8]. By comparison, repair polymerases, such as pol β (X-family), are much slower at DNA synthesis and less faithful than replicative DNA polymerases, often functioning with rates in the range of ~10 nt s^{-1} and fidelities on the order of 1 error in 10^2–10^4 incorporation events[6,9]. However, the reduced activity of repair polymerases is expected given their functional role in repairing damaged sites in genomic DNA by various cellular repair mechanisms. Furthermore, DNA polymerases harbor strict substrate specificity, reflecting their evolutionarily optimized recognition of the natural deoxyribonucleotide substrates. Together, fidelity, processivity, and substrate specificity define polymerase performance.

Structural studies by x-ray crystallography reveal that almost all polymerases adopt a catalytic domain that is described by analogy to a right hand, with distinct palm, fingers and thumb subdomains[10]. The palm subdomain is the reaction core of the enzyme, where a β-sheet scaffold supports the active site and homes the catalytic aspartic residues that are essential for phosphodiester bond formation. Surrounding this core are the finger and thumb subdomains, which are predominantly α-helical, and play complementary roles in processivity. In the absence of the substrate, the α-helical finger subdomain adopts an open conformation. The binding of the incoming nucleoside triphosphate induces the fingers to close, facilitating catalysis. The thumb

subdomain stabilizes the DNA duplex and facilitates translocation. Together, polymerases orchestrate highly dynamic and coordinated molecular interactions that govern DNA synthesis.

Propagation of an Artificial Genetic System

Over the past several decades, synthetic biology has sought to expand the chemical diversity of genetic polymers[11]. This effort has led to the development of xeno-nucleic acids (XNAs, Fig. 2), a broad class of artificial genetic polymers in which the canonical ribose or deoxyribose sugar is replaced with alternative chemical structure[12]. These modifications dramatically alter the physicochemical properties of the resulting polymers, with some XNAs exhibiting strong resistance to nuclease degradation and increased acid stability. As a result, XNAs have emerged as promising candidates for applications such as therapeutic oligonucleotides[13], molecular diagnostics[14], and allele specific gene-silencing reagents[15].

While many examples of XNA exist, there are six artificial genetic polymers that successfully replicate in a Darwinian evolution system to date: 1,5-anhydrohexitol nucleic acid (HNA), arabino nucleic acid (ANA), 2′-fluoro-arabino nucleic acid (FANA), cyclohexenyl nucleic acid (CeNA), locked nucleic acids (LNA) and α -L-threose nucleic acid (TNA)[16,17,18,19]. Among the various XNAs that have been explored, TNA is of particular interest in the Chaput group. Initially identified as part of a large systematic study designed to address the question of why nature chose ribose rather than a different aldose sugar as the backbone unit for DNA and RNA, the creation of TNA contributed to the foundation of synthetic genetics[20]. TNA is composed of a four-carbon sugar backbone rather than the five-carbon ribose or deoxyribose found in RNA and DNA. This results in a structurally simpler, but more rigid polymer that extends in a 2′-3′ direction, differing from the 3′-5′ phosphodiester linkages in DNA and RNA.

Despite these structural changes, TNA can still store information in the sequence of the canonical nucleobases adenine, guanine, cytosine, and thymine. Thus, TNA can form stable Watson–Crick base pairs with complementary strands of DNA, RNA, and itself[21]. As an aside, this cross-pairing ability has fueled interest in TNA as a potential ancestral genetic polymer, lending support to hypotheses that early genetic systems may have relied on simpler backbone chemistries prior to the emergence of RNA[18]. Beyond origins-of-life research, TNA exhibits properties that are attractive for biotechnology and medicine, including enhanced resistance to acid-mediated degradation and nucleases[22].

Substantial progress has already been made toward enabling synthetic genetic polymers, including the development of scalable chemical routes for the preparation of monomers and triphosphates[23,24,25]**.** Despite these advances, the widespread adoption of XNAs like TNA remains limited by challenges associated with their synthesis. For example, some XNAs can be prepared by solid-phase chemical synthesis, this approach is costly, environmentally toxic, and poorly suited for generating long oligos or libraries. In contrast, enzymatic synthesis offers a scalable and efficient route for XNA synthesis. However, the active sites of natural polymerases are generally incompatible with the altered sugar chemistry, having evolved to precisely accommodate the geometry and properties of their canonical substrates (i.e., DNA and RNA). As a result, achieving efficient and reliable enzymatic synthesis of XNAs represent a fundamental challenge in the field.

Despite this incompatibility, several encouraging examples of polymerase promiscuity have been reported, revealing that natural DNA polymerases can catalyze DNA dependent XNA synthesis[20,21]. These critical studies established that existing enzymes can serve as practical

starting points for the reprogramming of polymerases. Early generations of engineered polymerases have already demonstrated the feasibility of this approach[26,27,28]. However, it is recognized that further engineering is required to achieve robust activity.

Strategies for Gene Diversification

Directed evolution in the laboratory attempts to mimic that of biological evolution, but on a highly accelerated and controlled setting. Successful directed evolution relies on strategies that generate genetic diversity and effective identification of variants with enhanced properties. The most popular methods for gene diversification include random mutagenesis, focused mutagenesis, and recombination[30]. Additionally, computational tools can be used to guide gene diversification strategies by identifying residues or regions most likely to contribute to enhanced function.

Random mutagenesis can be further categorized into non-chemical (*in vivo)* and chemical (*in vitro)* methods. Non-chemical *in vivo* strategies increase mutation rates by increasing the error rate within DNA replication, but this approach is unfavorable because it also induces mutations to the host genome. Orthogonal replication systems overcome this limitation by restricting mutagenesis to the target DNA, enabling higher mutation efficiencies without compromising host viability[31,32]. Common *in vitro* chemical and physical strategies include use of alkylation-based reagents, deaminating agents, substitution with base analogues such as 2-aminopurine, and UV irradiation[33,34,35,36]. While all these methods introduce mutations throughout the target sequence, chemical mutagenesis is less commonly used in evolutionary campaigns due to biased rates of transversion and transition mutations. Of all the methods, *in vitro* error-prone polymerase chain reaction (epPCR) is the preferred mutagenesis technique. epPCR requires a low fidelity DNA polymerase subjected to increased magnesium

concentrations, supplementation with manganese, mutagenic dNTP analogues, or biased ratios of dNTPs[37,38,39]. epPCR is favorable because it requires no structure-function information and is easy to use.

Site-saturated mutagenesis (SSM) is a type of focused mutagenesis strategy in which a user-selected amino acid residue or multiple residues can be systematically substituted to all other 19 alternative amino acid residues[40,41,42]. This allows researchers to test every possible substitution at defined positions. All SSM methods rely on synthetic oligonucleotides that contain degenerate codons, or codons made with a mixed population of nucleotides, that are incorporated during PCR or oligo annealing. Each degenerate codon encodes multiple amino acids for a given position. The three ways to build SSM libraries are by cassette mutagenesis, whole plasmid amplification, or PCR-based mutagenesis. In cassette mutagenesis, a small DNA segment around the target residue is excised from the plasmid, and a synthetic mutagenized DNA "cassette" is inserted in its place. This method is practical for single sites or multiple sites that are in close proximity[42]. With whole plasmid amplification, the entire plasmid is amplified using primers that contain the degenerate codons[43]. The original parent plasmid is removed with DPN1, and the resulting mutated plasmids are transformed directly into *E. coli*. This method is simple with no subcloning required. PCR-based mutagenesis is the most widely used strategy, where PCR is used to amplify fragments containing degenerate codons[44]. The fragments are then assembled using cloning or other DNA assembly methods. This method is highly flexible, as users can introduce mutations to both nearby and distant residues simultaneously. The caveat to SSM is that it requires detailed structure or biochemical information to identify key residues worth investigating.

Unlike epPCR or focused mutagenesis, recombination is a strategy that mimics that of natural evolution. In homologous recombination, beneficial mutations are combined from variants of the same gene or homologous genes from different organisms. This rearrangement of existing mutations often produces large functional gain with fewer deleterious mutations. The original DNA shuffling method was pioneered by Stemmer, where five β-lactamase genes were randomly digested with DNase and allowed to freely reassemble in a PCR reaction without additional primers[45]. Following three rounds of recombination, the winning variants were back crossed with the parent genes to remove any neutral mutations. Aside from fragmentation recombination methods, Arnold and colleagues developed a staggered extension process (StEP) in which a typical PCR protocol is modified to prematurely abort the elongation step by jumping to the heat denaturation step in the next PCR cycle[46]. Subsequent annealing allows the previously incomplete extension products to switch templates, effectively recombining multiple templates into one amplicon. However, this might be challenging if parent genes have too low of sequence homology. The caveats to recombination are potential parent contamination, low diversity with chimeric libraries, and challenges on long genes with low sequence similarity.

A variety of computational gene diversification strategies have been used for directed evolution, including sequence-based or structure-based methods, molecular dynamics, and machine learning[47,48,49]. These computational approaches help identify residues, regions, or patterns that can be targeted for diversification. Additionally, machine learning, for example, can aid the workflow by building sequence-to-function maps based on quantified data from screening the variant pool. While there is no single algorithm that suits all protein engineering applications, the most important consideration is choosing an algorithm with abundant, high-quality datasets.

Selection and Screening

Advances in mutagenesis, recombination, and computational design have enabled the generation of increasingly large and diverse genetic libraries. However, library construction represents only one component of directed evolution; an equally critical challenge is the development of general, high-throughput workflows for identifying variants that exhibit desired phenotypes. Library evaluation can be categorized into selection-based or screening-based methods. Selection-based methods enable the evaluation of large protein libraries ($> 10^9$) by coupling functional performance or survival to enrichment. This directly eliminates undesired variants while allowing positive candidates to be subjected to further rounds of directed evolution. With screening-based methods, each variant is tested individually using biochemical or biophysical assays.

Popular selection methods include display-based selection and compartmentalization[2]. With display-based selection, each protein variant is physically connected to the DNA that encodes it. This genotype and phenotype linkage can either be direct, such as in mRNA display, or linked indirectly via a carrier, such as in phage display. As the earliest display method, phage display is one of the most studied and successful methods for evolving peptides or proteins to bind to a specific ligand[50,51,52]. The platform has since been modified for the evolution of polymerases, where a bacteriophage displays a protein variant in close proximity to a tethered DNA primer-template duplex, enabling selection of active enzymes through incorporation of biotin-labeled nucleotides (Fig 3a). Phage particles that acquire biotin through successful DNA synthesis are selectively captured using streptavidin, while the inactive variants are washed away. The genetic material of the active population is recovered and amplified through reinfection of *E. coli*. Successful examples of engineered polymerases include Stoffel fragment

(SF) of Taq DNA polymerase for ribonucleoside triphosphate[52], 2′-methoxy nucleoside triphosphate[53], and the unnatural PICS:PICS self pair[54].

With compartmentalization, the protein and gene are isolated together in a "container," such as a living cell, water-in-oil, or water-in-oil-in-water emulsions. Some of the most notable examples of compartmentalization for the evolution of polymerases include compartmentalized self-replication (CSR), compartmentalized self-tagging (CST), and droplet-based optical polymerase sorting (DrOPS).

CSR (Fig. 3b) is a selection-based technology developed by the Holliger group that leverages the native replicative activity of DNA polymerases[55]. A library of polymerase genes expressed in *E. coli* is encapsulated in bulk emulsions. Inside the droplet, active polymerases self-replicate by amplifying their encoding gene by PCR. The quality of amplification is directly proportional to the activity of the polymerase. Through iterative rounds, the most active variants will have more self-replication, resulting in an accumulation of the gene over rounds of selection, while outcompeting inactive variants. Examples of polymerase evolution campaigns utilizing CSR include engineering for recognition of α-phosphorthioate dNTP[55] and recognition of dNTPs with hydrophobic base analogs[56].

Like CSR, CST is a selection technique in which activity enables survival (Fig. 3c)[16]. A library of variants is transformed and expressed in *E. coli*. The cells are co-encapsulated with a biotinylated primer and a magnetic bead tethered to streptavidin. Following cell lysis, active polymerase variants extend the biotinylated primer, which is annealed to the encoding plasmid template. The emulsions are disrupted, and successfully replicated biotinylated genomic DNA is captured by the streptavidin bead and carried forward to the next round of selection. CST is well known for engineering XNA polymerases and advancing the field of synthetic genetics[16].

DrOPS was developed by the Chaput group[57,58] and is used in the work of this dissertation (Fig. 3d). Despite being a relatively new technique, DrOPS is a high-throughput activity-based screening platform that directly measures polymerase activity on a single-cell level. Like the previously mentioned methods, a library of polymerase variants is transformed and expressed in *E. coli.* The cells are then encapsulated with buffer, a self-priming hairpin with a 5′ fluorophore, a short 3′ quencher, and TNA triphosphates (tNTPs). Following cell lysis, polymerase variants are challenged to incorporate tNTPs across the DNA template to disrupt the donor-quencher pair. Successful synthesis enables the fluorescent signal. Following an incubation, droplets are optically sorted on a custom fluorescent activated droplet sorter[57]. Active variants are collected, and their encoding DNA is recovered and amplified for further rounds of evolution. Previously work in the Chaput group used the DrOPS platform to identify Kod-RSGA, a DNA dependent TNA polymerase engineered by deep mutational scanning and the systematic identification of epistatic mutations[27]. For the first time in our group, this thesis implemented iterative rounds of directed evolution using the DrOPS platform.

In this thesis, I describe the directed evolution of a DNA polymerase capable of efficiently and faithfully synthesizing TNA. By combining homologous recombination with targeted mutagenesis, this work enabled the systematic exploration of sequence space and the identification of combinations of mutations that collectively confer near-natural TNA polymerase activity. This integrated evolutionary strategy overcame key limitations of earlier approaches and resulted in a polymerase with dramatically improved catalytic efficiency and fidelity relative to previously reported variants.

This thesis is organized as follows. Chapter 1 details the design and execution of the evolutionary campaign that yielded an efficient TNA polymerase. Chapter 2 elaborates the experimental methodology behind our homologous recombination technique, as well as iterative cycles of directed evolution using DrOPS. Chapter 3 presents the structural and biochemical characterization of key evolutionary intermediates and the final evolved enzyme, elucidating the structural basis for near-natural rates of TNA synthesis. Finally, Chapter 4 discusses the broader implications of this work and outlines future directions for expanding the enzymatic accessibility of synthetic genetic polymers.

Figures for the Introduction

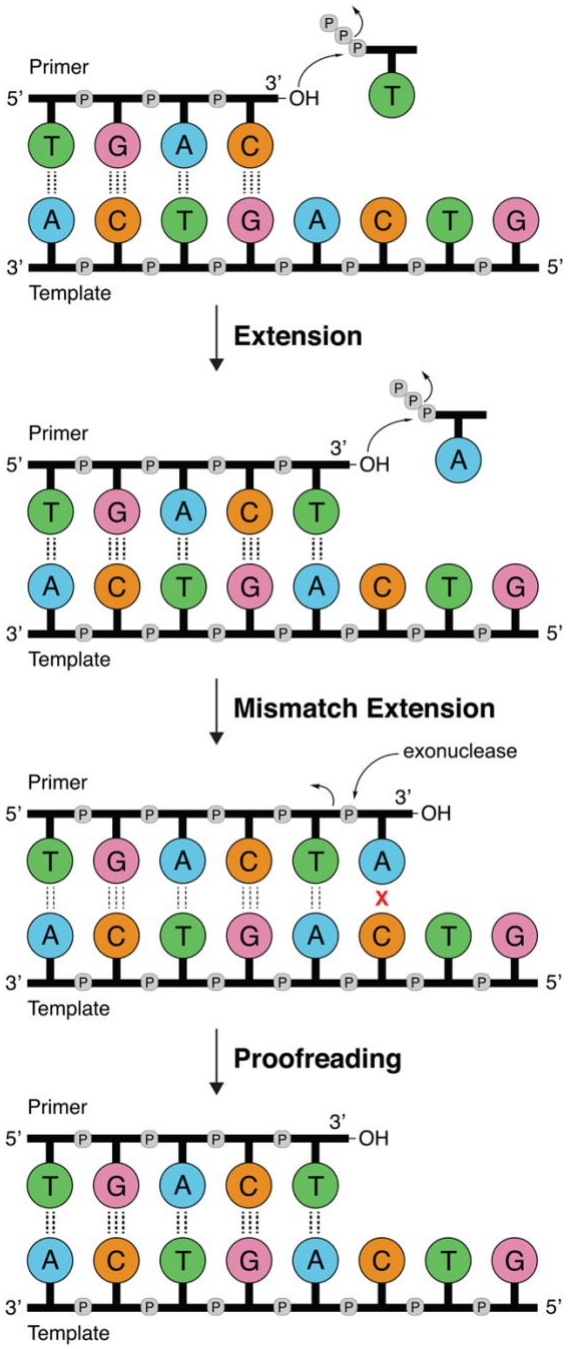


Figure 0.1 DNA synthesis mechanism.

Natural DNA polymerases extend a DNA primer in the 5'-3' direction using the template to determine the sequence of the growing strand. Polymerases with 3'-5' exonuclease activity have the ability to correct mistakes by removing terminal nucleotides that are incorrectly paired with the template. Adapted from Nikoomanzar, A. et al. *(2020), licensed under CC-BY 4.0.*

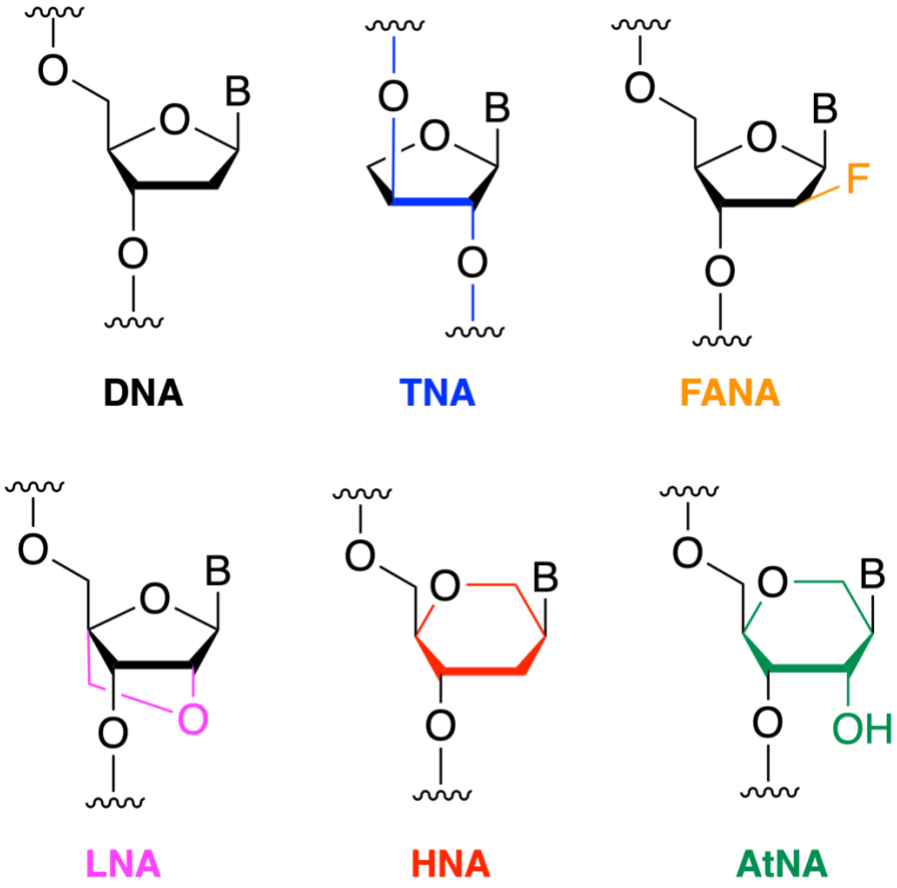


Figure 0.2 DNA and examples of xeno-nucleic acids (XNAs) and their chemical structures.

Chemical structures of DNA compared with representative xeno-nucleic acids (XNAs), highlighting differences in sugar architecture that distinguish natural and synthetic genetic polymers.

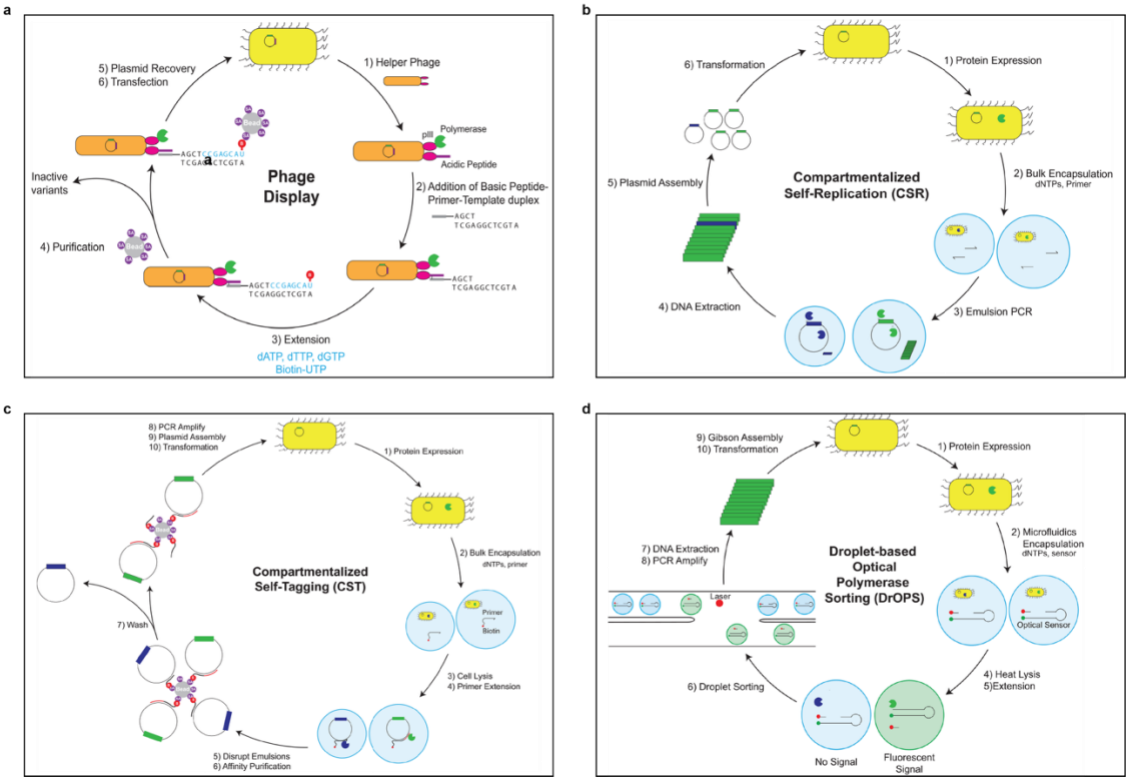


Figure 0.3 Examples of Screening Platforms.

(a). Phage display links polymerase variants to their DNA primer–template duplex on the surface of bacteriophage particles, enabling activity-based selection of enzymes that incorporate a biotin-tagged substrate for capture on streptavidin beads, followed by recovery and amplification of functional variants in E. coli. (b). *Compartmentalized self-replication (CSR) encapsulates individual polymerase-expressing* E. coli *cells in emulsion droplets, where active enzymes amplify their own genes by PCR—linking catalytic activity to gene replication—so that functional variants progressively enrich through iterative rounds of selective amplification. (c). Compartmentalized self-tagging (CST) encapsulates* E. coli *cells expressing polymerase variants in emulsions, where active enzymes extend a biotinylated primer annealed to their encoding plasmid, stabilizing the primer–plasmid complex for streptavidin-based capture and selective recovery of functional genes for subsequent amplification and evolution. (d). Droplet-based optical polymerase sorting (DrOPS) encapsulates individual* E. coli *cells expressing polymerase variants into microfluidic water-in-oil droplets, where active enzymes generate full-length extension products that activate a fluorescence sensor, enabling fluorescent droplets to be isolated by FADS and the enriched genes to be recovered, amplified, and subjected to subsequent rounds of evolution. Adapted from Nikoomanzar, A. et al. (2020), licensed under CC-BY 4.0*

References for the Introduction

1. Garcia-Diaz, M. & Bebenek, K. Multiple functions of DNA polymerases. *Crit. Rev. Plant Sci.* 26, 105–122 (2007).

2. Cooper, G. M. DNA replication. In *The Cell: A Molecular Approach* 2nd edn (Sinauer Associates, 2000). Steitz, T. A. & Steitz, J. A.

3. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl Acad. Sci. USA* 90, 6498–6502 (1993).

4. Ito, J. & Braithwaite, D. K. Compilation and alignment of DNA polymerase sequences. *Nucleic Acids Res.* 19, 4045–4057 (1991).

5. Kunkel, T. A. DNA replication fidelity. *J. Biol. Chem.* 279, 16895–16898 (2004).

6. Nikoomanzar, A., Chim, N., Yik, E. J. & Chaput, J. C. Engineering polymerases for applications in synthetic biology. *Q. Rev. Biophys.* 53, e8 (2020).

7. Kunkel, T. A. & Bebenek, K. DNA replication fidelity. *Annu. Rev. Biochem.* 69, 497–529 (2000).

8. Loeb, L. A. & Monnat, R. J. Jr. DNA polymerases and human disease. *Nat. Rev. Genet.* 9, 594–604 (2008).

9. Wu, W.-J., Yang, W. & Tsai, M.-D. How DNA polymerases catalyse replication and repair with contrasting fidelity. *Nat. Rev. Chem.* 1, 0068 (2017).

10. Brautigam, C. A. & Steitz, T. A. Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.* 8, 54–63 (1998).

11. Chaput, J. C., Egli, M. & Herdewijn, P. The XNA alphabet. Nucleic Acids Res. 53, gkaf635 (2025).

12. Chaput, J. C. & Herdewijn, P. What is XNA? *Angew. Chem. Int. Ed.* 58, 11570–11572 (2019)

13. Lozoya-Colinas, A., Yu, Y. & Chaput, J. C. Functionally enhanced XNA aptamers discovered by parallelized library screening. *J. Am. Chem. Soc.* 145, 25789–25796 (2023).

14. Yang, K. & Chaput, J. C. REVEALR: a multicomponent XNAzyme-based nucleic acid detection system for SARS-CoV-2. *J. Am. Chem. Soc.* 143, 8957–8961 (2021).

15. Wang, Y., Wang, Y., Song, D., Sun, X., Li, Z., Chen, J. Y. & Yu, H. An RNA-cleaving threose nucleic acid enzyme capable of single point mutation discrimination. *Nat. Chem.* 14, 350–359 (2022)

16. Pinheiro, V. B. *et al.* Synthetic genetic polymers capable of heredity and evolution. *Science* 336, 341–344 (2012).

17. Alves Ferreira-Bravo, I. *et al.* Selection of 2′-deoxy-2′-fluoroarabinonucleotide (FANA) aptamers that bind HIV-1 reverse transcriptase with picomolar affinity. *Nucleic Acids Res.* 43, 9587–9599 (2015).

18. 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, 183–187 (2012).

19. Taylor, A. I. *et al.* Catalysts from synthetic genetic polymers. *Nature* 518, 427–430 (2015).

20. Eschenmoser, A. Chemical etiology of nucleic acid structure. *Science* 284, 2118–2124 (1999).

21. Schöning, K.-U., Scholz, P., Wu, X., Guntha, S., Delgado, G., Krishnamurthy, R. & Eschenmoser, A. The α-L-threofuranosyl-(3′→2′)-oligonucleotide system ('TNA'): synthesis and pairing properties. *Helv. Chim. Acta* 85, 4111–4153 (2002).

22. Culbertson, M. C., Temburnikar, K., Sau, S. P., Liao, J. Y., Bala, S. & Chaput, J. C. Evaluating TNA stability under simulated physiological conditions. Bioorg. Med. Chem. Lett. 26, 2418–2421 (2016).

23. Hollenstein, M. Nucleoside triphosphates—building blocks for the modification of nucleic acids. Molecules 17, 13569–13591 (2012).

24. Sau, S. & Chaput, J. C. A gram-scale HPLC-free synthesis of TNA triphosphates using an iterative phosphorylation strategy. Org. Lett. 19, 4379–4382 (2017).

25. Sau, S. P. & Chaput, J. C. A one-pot synthesis of α-L-threofuranosyl nucleoside triphosphates (tNTPs). Bioorg. Med. Chem. Lett. 26, 3271–3273 (2016).

26. Houlihan, G., Arangundy-Franklin, S. & Holliger, P. Engineering and application of polymerases for synthetic genetics. Curr. Opin. Biotechnol. 48, 168–179 (2017).

27. Nikoomanzar, A., Vallejo, D., Yik, E. J. & Chaput, J. C. Programmed allelic mutagenesis of a DNA polymerase with single amino acid resolution. *ACS Synth. Biol.* 9, 1873–1881 (2020).

28. Medina, E., Yik, E. J., Herdewijn, P. & Chaput, J. C. Functional comparison of laboratory-evolved XNA polymerases for synthetic biology. *ACS Synth. Biol.* 10, 1429–1437 (2021).

29. Wang, Y., Ngor, A. K., Nikoomanzar, A. & Chaput, J. C. Evolution of a general RNA-cleaving FANA enzyme. *Nat. Commun.* 9, 5067 (2018).

30. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* 16, 379–394 (2015).

31. Rix, G., Watkins-Dulaney, E.J., Almhjell, P.J. *et al.* Scalable continuous evolution for the generation of diverse enzyme variants encompassing promiscuous activities. *Nat Commun* 11, 5644 (2020). https://doi.org/10.1038/s41467-020-19539-6

32. Diercks, C. S., Sondermann, P., Rong, C., Gillis, T. G., Ban, Y., Wang, C., Dik, D. A. & Schultz, P. G. An orthogonal T7 replisome for continuous hypermutation and accelerated evolution in *E. coli*. *Science* 389, 618–622 (2025).

33. Lai, Y. P., Huang, J., Wang, L., Li, J. & Wu, Z. R. A new approach to random mutagenesis *in vitro*. *Biotechnol. Bioeng.*86, 622–627 (2004).

34. Myers, R. M., Lerman, L. S. & Maniatis, T. A general method for saturation mutagenesis of cloned DNA fragments. *Science* 229, 242–247 (1985).

35. Freese, E. Specific mutagenic effect of base analogues on phage T4. *J. Mol. Biol.* 1, 87–105 (1959).

36. Bridges, B. A. & Woodgate, R. Mutagenic repair in *Escherichia coli*: products of the *recA* gene and of the *umuD* and *umuC* genes act at different steps in UV-induced mutagenesis. *Proc. Natl Acad. Sci. USA* 82, 4193–4197 (1985).

37. Leung, D. W., Chen, E. & Goeddel, D. V. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* 1, 11–15 (1989).

38. Zaccolo, M., Williams, D. M., Brown, D. M. & Gherardi, E. An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *J. Mol. Biol.* 255, 589–603 (1996).

39. Cadwell, R. C. & Joyce, G. F. Mutagenic PCR. *PCR Methods Appl.* 3, S136–S140 (1994).

40. Matteucci, M. D. & Heyneker, H. L. Targeted random mutagenesis: the use of ambiguously synthesized oligonucleotides to mutagenize sequences immediately 5′ of an ATG initiation codon. *Nucleic Acids Res.* 11, 3113–3121 (1983).

41. Reidhaar-Olson, J. F. & Sauer, R. T. Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. *Science* 241, 53–57 (1988).

42. Wells, J. A., Vasser, M. & Powers, D. B. Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites. *Gene* 34, 315–323 (1985).

43. Byrappa, S., Gavin, D. K. & Gupta, K. C. A highly efficient procedure for site-specific mutagenesis of full-length plasmids using Vent DNA polymerase. *Genome Res.* 5, 404–407 (1995).

44. Quan, J. & Tian, J. Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS One* 4, e6441 (2009).

45. Stemmer, W. P. C. DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proc. Natl Acad. Sci. USA* 91, 10747–10751 (1994).

46. Zhao, H., Giver, L., Shao, Z. *et al.* Molecular evolution by staggered extension process (StEP) *in vitro* recombination.*Nat. Biotechnol.* 16, 258–261 (1998).

47. Ng, P. C. & Henikoff, S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31, 3812–3814 (2003).

48. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 7, e46688 (2012).

49. Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y. & Liang, J. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.*34, W116–W118 (2006).

50. Smith, G. P. & Petrenko, V. A. Phage display. *Chem. Rev.* 97, 391–410 (1997).

51. Jestin, J.-L., Kristensen, P. & Winter, G. A method for the selection of catalytic activity using phage display and proximity coupling. *Angew. Chem. Int. Ed.* 38, 1124–1127 (1999).

52. Xia, G., Chen, L., Sera, T., Fa, M., Schultz, P. G. & Romesberg, F. E. Directed evolution of novel polymerase activities: mutation of a DNA polymerase into an efficient RNA polymerase. *Proc. Natl Acad. Sci. USA* 99, 6597–6602 (2002).

53. 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, 1748–1754 (2004).

54. Leconte, A. M., Chen, L. & Romesberg, F. E. Polymerase evolution: efforts toward expansion of the genetic code. *J. Am. Chem. Soc.* 127, 12470–12471 (2005).

55. Ghadessy, F. J., Ong, J. L. & Holliger, P. Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl Acad. Sci. USA* 98, 4552–4557 (2001).

56. Loakes, D., Gallego, J., Pinheiro, V. B., Kool, E. T. & Holliger, P. Evolving a polymerase for hydrophobic base analogues. *J. Am. Chem. Soc.* 131, 14827–14837 (2009).

57. Larsen, A. C., Dunn, M. R., Hatch, A., Sau, S. P., Youngbull, C. & Chaput, J. C. A general strategy for expanding polymerase function by droplet microfluidics. *Nat. Commun.* 7, 11235 (2016).

58. Vallejo, D., Nikoomanzar, A., Paegel, B. M. & Chaput, J. C. Fluorescence-activated droplet sorting for single-cell directed evolution. *ACS Synth. Biol.* **8**, 1430–1440 (2019)

CHAPTER 1: Directed Evolution of a Highly Efficient TNA Polymerase Achieved by Homologous Recombination

Victoria A. Maola[1†], Eric J. Yik[1†], Mohammad Hajjar[1], Joy J. Lee[1], Manuel J. Holguin[1], Riley N. Quijano[1], Kalvin K. Nguyen[1], Katherine L. Ho[1], Jenny V. Medina[1], Nicholas Chim[1], and John C. Chaput[1-4]*

† These authors contributed equally

Affiliations

[1]Department of Pharmaceutical Sciences, University of California, Irvine, CA 92697-3958 USA

[2]Department of Chemistry, University of California, Irvine, CA 92697-3958 USA

[3]Department of Molecular Biology and Biochemistry, University of California, CA 92697-3958 USA

[4]Department of Chemical and Biomolecular Engineering, University of California, Irvine, CA 92697-3958 USA

This chapter is a reprint of the material as it appears in Nature Catalysis: *Nat Catal* **7**, 1173–1185 (2024). doi: 10.1038/s41929-024-01233-1 used with permission from Springer Nature Limited.

1.1 Abstract

Reprogramming DNA polymerases to synthesize xeno-nucleic acids (XNA) is an important challenge that tests current enzyme engineering tools. Here, we describe an evolutionary campaign aimed at generating an XNA polymerase that can efficiently make α-L-threofuranosyl nucleic acid (TNA) – an artificial genetic polymer that is recalcitrant to nucleases and resistant to acid-mediated degradation. Starting from a homologous recombination library, iterative cycles of selection were performed to traverse the fitness landscape in search of neutral mutations with increased evolutionary potential. Subsequent directed evolution of focused mutagenic libraries yielded 10-92, a newly engineered TNA polymerase that functions with a catalytic rate of ~1 nt/s and >99% fidelity. A crystal structure of the closed ternary complex reveals the degree of structural change required to remodel the active site pocket for improved TNA synthesis activity. Together, these data demonstrate the importance of recombination as a strategy for evolving XNA polymerases with considerable practical value for biotechnology and medicine.

1.2 Introduction

Engineering DNA polymerases to synthesize artificial genetic polymers with novel backbone architectures (also known as xeno-nucleic acids or XNAs) is an important challenge that tests current enzyme engineering strategies and offers a route to valuable reagents that can advance future applications in biotechnology and medicine[1-3]. Although computational methods have solved many problems in protein design[4,5], attempts to create XNA polymerases by in silico prediction have thus far proven unsuccessful, which is presumably due to the complicated reaction pathway[6] and dynamic motion[7] that replicative DNA polymerases follow for each cycle of nucleotide addition. We and others have explored this problem using rational design and high-

throughput screening to identify mutations that broaden the substrate specificity of DNA polymerases toward XNA triphosphates[8-14]. However, these efforts, which examined an assortment of sugar chemistries have proven more challenging than related studies aimed at generating polymerase variants that recognize chemical modifications made to the nucleobase[15,16], 2' sugar position[17,18], and phosphodiester linkage[19,20] of DNA. This observation suggests that the amino acid changes required to reprogram DNA polymerases with heightened levels of XNA synthesis activity remain mostly unknown. Uncovering the identity of these residues, which are likely to be unique for each XNA system, is a challenging combinatorial problem that will require deploying ultrahigh-throughput screening tools to search new regions of the fitness landscape for structural variants that are more adept at XNA synthesis.

Since previous attempts at generating α-L-threofuranosyl nucleic acid (TNA) polymerases have resulted in enzymes with catalytic activities that are inferior to natural DNA polymerases[8-10], we considered the possibility that major structural changes would be required to improve the geometry of the TNA triphosphate (tNTP) in the active site pocket. Insight into this problem can be found in the crystal structure of Kod-RI, a first-generation TNA polymerase, which shows poor recognition of the templating base and a suboptimal distance between the 2'-hydroxyl on the TNA primer and the α-phosphate of the tNTP substrate[21]. As these studies were performed using polymerase libraries generated by saturation mutagenesis and deep mutational scanning[8-10], we felt that homologous recombination, a technique based on the recombination of protein fragments from related species or synthetic libraries, offered an opportunity to explore new regions of the fitness landscape[22,23]. Classical studies have shown that recombination libraries produce a higher frequency of functional proteins than error-prone methods because most of the resulting mutations derive from natural orthologs, which are less likely to introduce unwanted stop codons into the

sequence[24]. Previously, recombination was used to engineer DNA polymerases for the ability to bypass lesions, function in the presence of environmental inhibitors, and incorporate fluorescent tags for next-generation sequencing[25-27].

TNA was chosen as the target for polymerase activity to allow comparison with previous efforts[8-10] and to provide a stringent test of XNA substrate recognition, as the backbone repeat unit of TNA is one atom shorter than that of DNA and RNA (Fig. 1.1a)[28]. Unlike other potentially natural derivatives of RNA[29], TNA is capable of forming stable antiparallel duplex structures with itself and with complementary strands of DNA and RNA[28]. It's unusual 2',3'-linked phosphodiester backbone is recalcitrant to nucleolytic enzymes[30] and stable against acid-mediated degradation[31], making it an attractive genetic system for therapeutic applications[32]. In vitro selection experiments have shown that TNA is capable of folding into shapes with ligand binding and catalytic activity[33-38], and its recognition by Bst DNA polymerase, a TNA reverse transcriptase, allows digital information stored in TNA to be recovered and read by DNA sequencing[39]. Finally, TNA has been evaluated as a building block for improving the biostability and activity of gene silencing applications involving DNAzymes[40], antisense oligonucleotides[41], and RNA interference pathways[42].

Here we show how recombination can be used to reprogram natural DNA polymerases for the ability synthesize XNA. Following a process of in vitro selection and directed evolution, we isolated a TNA polymerase, termed 10-92, that functions with a catalytic rate of ~1 nt/s and >99% fidelity. A crystal structure of the closed ternary complex bound to the primer-template duplex and an incoming tNTP reveals the degree of structural change required to remodel the active site pocket for improved TNA synthesis activity. Together, these data demonstrate the importance of

recombination as a strategy for evolving highly efficient and faithful XNA polymerases that can be used as tools to synthesize artificial genetic polymers.

1.3 Results

Library Generation by Homologous Recombination

Because natural polymerases are constrained by a long history of evolutionary contingency, sequences with strong XNA synthesis activity are expected to be quite rare in sequence space[43]. We, therefore, chose to design our starting DNA library based on four hyperthermophilic archaeal B-family DNA polymerases, which are more accepting of modified nucleotides than other polymerase classes[44]. Four 2.4 kilobase (kb) genes encoding polymerases isolated from the archaeal species *Thermococcus kodakarensis* (Kod), *Thermococcus gorgonarius* (Tgo), *Pyrococcus* Deep Vent (DV), and *Thermococcus* 9°N (9°N) were prepared as synthetic double-stranded (ds) DNA by high fidelity gBlock assembly (Appendix A: Supplementary Fig. 1, Supplementary Tables 1-2). The Tgo, DV, and 9°N genes are 89-96% identical to Kod DNA polymerase. Each gene carried the exonuclease silencing mutations (D141A and E143A) and four beneficial mutations (A485R, N491S, R606G, and T723A) that derive from our previous best TNA polymerase, Kod-RSGA (Fig. 1b)[10], and were expected to confer modest TNA synthesis activity on the synthetic orthologs[45]. Additionally, the library contained 10 conserved regions (Fig. 1.1c) that were required for synthetic recombination using the polymerase chain reaction (PCR). Homologous recombination (Fig. 1.1d) was achieved using a DNA shuffling approach in which 44 individual PCR reactions were performed to fragment each gene into 11 pieces that were reassembled in a single PCR reaction (Appendix A: Supplementary Fig. 2) to create a pool of >4 million polymerase variants[46]. DNA sequencing of individual clones isolated from the pool

revealed a library composed of the desired recombinants, recombinants with novel crossovers caused by template switching, and random single nucleotide point mutations (SNPs). In contrast to our previous TNA polymerase libraries[8-10], this population of sequences was expected to explore a larger range of sequence space due to the combinatorial nature of the assembly process[47].

Droplet-Based Polymerase Evolution

Iterative rounds of directed evolution were accomplished using the massively parallel screening technique of droplet-based optical polymerase sorting (DrOPS, Appendix A: Supplementary Table 3)[48]. DrOPS is a two-step microfluidic approach in which *E. coli* expressing different polymerase variants are individually encapsulated in uniform water-in-oil (w/o) droplets along with the reagents necessary to complete a polymerase activity assay (Fig. 1.1e), including chemically synthesized tNTPs[49]. The droplet population (~10^7) is collected, thermally lysed off-chip by prolonged heating at 95°C, and incubated at 55°C to promote TNA synthesis on a self-priming DNA template. Polymerases that successfully copy the template into full-length TNA generate a fluorescent signal by disrupting a fluorophore-quencher pair located at the 5' end of the DNA template. Fluorescent droplets are then recovered by passing the droplet population through a fluorescence-activated droplet sorting (FADS) device, which partitions droplets based on user-defined fluorescence values. Genes encoding active polymerases are recovered, amplified by PCR, reinserted into an expression vector, and transformed back into *E. coli* for another round of selection (Fig. 1.1f). After 5 rounds of selection for TNA synthesis activity, a screen of 500 members identified clone 5-270 as a promising starting point for subsequent optimization by directed evolution (Appendix A: Supplementary Figs. 3-5). The sequence of clone 5-270 differs from its closest natural homolog, Kod DNA polymerase, by 40 amino acid mutations and 1 Leu insertion located on α-helix 12 of the palm subdomain (Fig. 1.2a).

To identify variants with improved TNA synthesis activity, we performed a directed evolution campaign that began by intentionally mutating the thumb subdomain of clone 5-270 and carrying the library through two additional rounds of droplet sorting. Screening 100 clones from round 7 identified clone 7-47 as an improved variant with 4 mutations relative to 5-270 (Fig. 1.2a). Mutagenesis of the thumb subdomain of clone 7-47, followed by one round of selection, uncovered clone 8-64 as the top performing variant isolated from round 8. Clone 8-64 differs from 7-74 by a single point mutation (A741P) located in α-helix 21 in the thumb subdomain (Fig. 1.2a). Finally, we mutagenized the finger and palm subdomains of clone 8-64 and carried the new library through two more rounds of selection. Screening 100 clones from round 10 identified clone 10-92 as a highly efficient TNA polymerase with 5 additional mutations, two of which occur in the finger subdomain responsible for tNTP recognition (Fig. 1.2a,b). Comparative sequence analysis shows that 10-92 is a highly chimeric enzyme comprised of pieces from all four parent polymerases (Fig. 1.2c, Appendix A: Supplementary Fig. 6), differing from natural Kod polymerase by 1 insertion and 50 amino acid mutations.

The enzymatic and biochemical properties of the 10-92 TNA polymerase were evaluated using standard polymerase assays. The results demonstrate that 10-92 is strikingly more efficient at TNA synthesis in standard magnesium buffer than any of its evolutionary progenitors (Fig. 1.3a,b, Appendix A: Supplementary Figs. 7-8), capable of extending a DNA primer to full-length product with 40 nucleotide additions occurring in 45 seconds (Fig. 1.3c). TNA synthesis activity appears to be slightly more efficient from a TNA primer than a DNA primer (Appendix A: Supplementary Fig. 9). The enhanced catalytic activity of 10-92 allows the enzyme to support TNA synthesis with reduced tNTP concentrations (4-fold) relative to Kod-RSGA, making it a more efficient enzyme for TNA synthesis. It also allows the enzyme to discriminate against 2'-

deoxynucleoside triphosphates (dNTPs) to a higher degree (4-fold) than Kod-RSGA (Fig. 1.3d,e), suggesting that 10-92 is on the path to becoming a more specialized enzyme for TNA synthesis. Consistent with its improved substrate recognition ability, the polymerase was found to be more accepting of base-modified tNTPs (Fig. 1.3f, Appendix A: Supplementary Fig. 10), which are valuable substrates for functionally enhanced TNA aptamers, known as threomers[35]. The fidelity of TNA synthesis was assessed by sequencing the product of a complete cycle of TNA replication (DNA→TNA→DNA). This assay, which measures the aggregate fidelity of replication using controls to eliminate PCR artifacts, reveals that 10-92 functions with >99% template copying fidelity, yielding only 3 mistakes out of 1000 nucleotide incorporation events (Appendix A: Supplementary Figs. 11-12). This is a notable improvement over previous generations, which suffered from G-G mismatches during TNA synthesis[50]. Last, we confirmed that 10-92 maintains the extreme thermal stability of its natural homologs by showing that the TNA synthesis activity of 10-92 is undiminished after 6 hours of heating at 90°C (Fig. 1.3g). We attribute the extreme thermal stability of 10-92 to the cell lysis step of the DrOPS protocol that requires the polymerase to survive a thermal challenge of 5 minutes at 95°C. Together, the combined properties of high catalytic activity, high fidelity, and high thermal stability make the 10-92 TNA polymerase an important tool in synthetic biology.

Intrigued by the catalytic efficiency of 10-92, we decided to measure the rate of TNA synthesis using the technique of polymerase kinetic profiling (PKPro)[51]. In contrast to single-nucleotide primer-extension assays, which measure the efficiency at which a single nucleotide is added to the 3' end of a DNA primer[52], PKPro provides a more holistic view of XNA synthesis by measuring the average rate of nucleotide incorporation across a template[51]. As such, researchers are able to evaluate the entire catalytic cycle, which includes substrate recognition, nucleotide incorporation

onto the growing strand, and continued extension from the modified position. For this assay, an unlabeled DNA primer-template duplex was extended with TNA by incubating the substrate with 10-92 and tNTPs at 55°C. The primer-extension reactions were stopped at designated times by plunging the reaction vessels into powdered dry ice. Once all the time points were collected, the reactions were thawed on a cold plate in the presence of EvaGreen, previously identified as an optimal intercalating dye for TNA-DNA duplexes[51], and rates were evaluated by measuring the fluorescence signal of each sample. Background subtracted plots yield a rate of 58.08 ± 1.15 nt/min for 10-92 (Fig. 4). This rate of TNA synthesis is noticeably faster (Fig. 1.4) than the rates observed for our first, second, and third generation TNA polymerases of Kod-RI (0.08 ± 0.00 nt/min), Kod-RS (1.86 ± 0.17 nt/min), and Kod-RSGA (5.74 ± 0.16 nt/min), respectively. Similar trends were observed in time course studies analyzed by denaturing PAGE (Appendix A: Supplementary Fig. 8).

Structure Activity Analysis

We examined the contribution of the 10 adaptive mutations acquired in the directed evolution of 10-92 by assessing the TNA synthesis activity of single point reversions in which each mutation was individually reverted back to the amino acid residue observed in 5-270. Each single point reversion of the 10-92 TNA polymerase was prepared by site-directed mutagenesis, cloned, and purified from *E. coli* lysate (Appendix A: Supplementary Fig. 13). TNA synthesis activity evaluated at 45 and 60 seconds indicates that the G615D, P741A, and P548T reversions have the most dramatic effect on the polymerase by slowing the rate of TNA synthesis (Appendix A: Supplementary Fig. 14). Interestingly, these mutations each involve the loss of either a glycine or proline residue in the 10-92 enzyme, indicating their potential importance in enhancing or

constraining the flexibility of the folded structure. Additionally, we observe a noticeable loss of activity when the 10 directed evolution mutations were transferred to Kod-RSGA (Appendix A: Supplementary Fig. 15), highlighting the epistatic nature of the acquired mutations.

Synthesis of Base-Modified TNA Aptamers

To illustrate the potential for 10-92 to function as an important tool for synthetic biology, we compared the synthesis of three TNA aptamers using the TNA polymerases 10-92 and Kod-RSGA, which was the chemical basis of our starting library[10]. The aptamer sequences chosen for this study were selected from in vitro evolution experiments that were performed previously against the targets of HIV reverse-transcriptase (HIV-RT)[53], the receptor binding domain (RBD) of the S1 protein from SARS-CoV-2[54], and the full-length S1 protein from SARS-CoV-2[35]. These aptamers comprise sequences of increasing synthetic difficulty as they transition from standard base chemistry for the HIV-RT aptamer to base-modified sequences carrying amide-linked phenylalanine and tryptophan side chains at the C-5 position of uracil bases for the RBD and S1 aptamers, respectively. Side-by-side primer extension studies reveal that 10-92 is strikingly more effective at synthesizing standard and base-modified TNA aptamers than Kod-RSGA by extending a 5' labeled DNA primer annealed to the complementary DNA template (Fig. 1.5). Even the tryptophan chemistry, which historically has been a difficult chemotype to prepare by primer extension, was generated as a distinct full-length product with minor truncation bands as viewed by denaturing polyacrylamide gel electrophoresis (PAGE). The binding activity of the enzymatically generated and PAGE purified TNA aptamers was confirmed by biolayer interferometry (BLI) using TNA sequences carrying a 5' biotin-labeled version of the DNA primer. Curve fitting of the resulting BLI sensorgrams reveals that the data conforms to a 1:1 binding

model with calculated dissociation constants (K_D) in the low nanomolar range. As expected, TNA aptamers carrying the aromatic side chains function with slower off-rates than the standard base sequence (Fig. 1.5), making base-modified threomers possible reagents for diagnostic and therapeutic applications.

Structural Analysis

The robust catalytic activity of the 10-92 TNA polymerase warrants a better understanding of the structural alterations needed to reprogram a natural DNA polymerase for strong XNA synthesis activity. Insights into this problem were obtained by solving the crystal structure of the closed ternary complex of 10-92 with the primer-template (P-T) duplex and incoming tNTP bound in the enzyme active site. Protein crystals grown in the presence of a chemically synthesized chain-terminating 2′-deoxy-α-L-threofuranosyl thymidine triphosphate ($tTTP^d$)[55] and excess tATP trapped the enzyme in a catalytically active conformation. Diffraction-quality crystals were identified from a screen of ~900 conditions performed in a hanging-drop format using full-length and C-terminal truncated versions of the protein. The best crystal carried a 15 amino acid C-terminal deletion and resolved to a resolution limit of 2.7Å. This structure was solved by molecular replacement using a previous TNA polymerase structure as the search model (Appendix A: Supplementary Table 4)[21].

The 10-92 TNA polymerase adopts a disk-shaped architecture (Fig. 1.6) that is consistent with all known archaeal B-family DNA polymerases[56]. Superposition of the structure against the natural Kod DNA polymerase[57] reveals major conformational changes to the catalytic domain as well as smaller changes to the N-terminal (NTD) and exonuclease (Exo) domains (Fig. 1.6, Appendix A: Supplementary Fig. 16). These structural rearrangements, caused by the accumulation of neutral

and beneficial mutations throughout the protein scaffold, primarily impact coordination of the enzyme to the P-T duplex (Appendix A: Supplementary Table 5). Regions of the NTD domain shift by as much as 2Å to interact with the unpaired region of the template strand, which is well resolved as it extends into the NTD (Fig. 1.6a). In the thumb subdomain, large positional rearrangements of the α-helical secondary structures (up to 7Å) bring the thumb closer to the paired region of the DNA duplex (Fig. 1.6b), presumably improving the translocation efficiency of the enzyme. Another dramatic structural change occurs in the palm subdomain, which contains the catalytic aspartate residues required for chemical bond formation. Here, the Leu381 insertion, acquired by recombination with *Pyrococcus* DV, causes the α12-helix to shift by 3Å relative to its position in Kod DNA polymerase (Fig 1.6c). This change to the protein sequence, along with neighboring mutations, is responsible for remodeling the enzyme active site and improving coordination to the DNA template.

Close inspection of the crystal structure reveals the presence of a snuggly fit tATP substrate in the active site pocket formed by the palm and finger subdomains (Fig. 1.7a, Appendix A: Supplementary Fig. 17). Most of the mutations in this region of the polymerase reside in the O-helix of the finger subdomain, which is responsible for recognizing the incoming tNTP substrate. These mutations account for a smaller active site pocket as compared to a previously solved structure of an ancestral variant known as Kod-RI (Appendix A: Supplementary Fig. 18)[21]. The tATP substrate is stabilized by a series of Van der Waals interactions with the base and sugar moieties and electrostatic contacts to the phosphate tail, including the coordination of a tightly bound Mg^{2+} ion (Fig 1.7b). Consistent with fast enzyme kinetics, the templating base (dT8) forms a favorable coplanar base pair with the tATP substrate (Fig 1.7c, Appendix A: Supplementary Table 5), thus correcting a suboptimal base pair geometry previously observed in the ancestral

Kod-RI TNA polymerase structure[21]. The improved base pairing ability of the incoming tATP substrate leads to a bond distance of 3.9Å from the 2′ carbon atom of the threose sugar on the chain terminating primer to the α-phosphate of the incoming tATP substrate, which closely approximates the distance (3.7Å) observed in the structure of natural Kod DNA polymerase[57]. The remodeled active site represents an improvement over the ancestral Kod-RI polymerase, which exhibited a bond distance of 4.6Å between equivalent atoms.

Comparison of the 10-92 and wild-type Kod DNA polymerase structures provided an opportunity to interpret the functional role of the most prominent directed evolution mutations identified in our reversion study. As noted above, reversion of the directed evolution mutations Pro548, Gly615, and Pro741 to their wild-type residue led to reduced levels of TNA synthesis activity (Appendix A: Supplementary Fig. 14), indicating that these mutations play an important role in TNA synthesis. In the palm subdomain, Pro548 along with several mutations acquired by homologous recombination and one directed evolution mutation (His550), cause noticeable rearrangements to the active site pocket (Appendix A: Supplementary Figs. 18-19). This includes the truncation of β-strands 18, 21, and 22, which reside in close proximity to the catalytic residues Asp405, Asp541, and Asp543. Collectively, this cluster of mutations is responsible for improving the geometry of the reaction for in-line attack on the tNTP substrate.

The mutations of Gly615 and Pro741 by directed evolution play a more subtle role in TNA synthesis by improving contacts to the P-T duplex. In the wild-type structure, Arg612, Asp614, Trp615 form a structured hydrogen bonding triad that allows Arg612 and Trp615 to contact the minor groove of the P-T duplex. The Gly615 mutation disrupts this triad, leading to a loss of the $3_{10}1$ helix (Appendix A: Supplementary Fig. 1) and greater flexibility of the Arg613 side chain. We speculate that this change may improve the translocation efficiency of TNA synthesis by

allowing Arg613 to contact the newly synthesized TNA strand as the primer transitions from DNA into TNA. At a more distant region of the duplex, Arg743 forms an ion pair with Glu609 in the wild-type structure, positioning Arg743 to contact the negatively charged phosphate groups along the template strand. The Pro741 mutation in the 10-92 polymerase shifts the coordination of Arg744 from Glu610 to prolines 740 and 741, which may allow for better coordination of the TNA-DNA heteroduplex.

1.4 Discussion

Replicative DNA polymerases are responsible for duplicating the genomes of organisms during cell division. These enzymes follow a template-dependent primer extension mechanism in which the growing daughter strand is extended in the 5'-3' direction through the sequential addition of correctly paired nucleotides. This process of oligonucleotide synthesis has been widely regarded as a viable path for synthesizing artificial genetic polymers ranging in composition from close structural analogs of DNA to XNAs with unusual sugar-phosphate backbones[1-3]. As expected, the activity of natural polymerases toward unnatural substrates tends to vary according to the types of modifications found in the substrate. As an example, 2'-fluoroarabino nucleic acid (FANA), a genetic material differing from DNA by a single hydrogen to fluorine substitution at the 2' carbon position of the furanose ring is recognized by several naturally occurring DNA polymerases[58]. By comparison, more chemically divergent substrates, like TNA and hexose nucleic acid (HNA), cause natural polymerases to chain terminate at the primer or following a small number of nucleotide incorporations[59,60]. This problem of poor substrate recognition is presumably due to the strict geometric control that polymerases impose on their cognate DNA substrates to ensure proper

nucleotide selection, chemical bond formation, and continued primer extension following nucleotide incorporation[61].

Reprogramming DNA polymerases to efficiently synthesize a wide range of chemically diverse XNAs remains an important goal in synthetic biology. Our approach to this problem focused on TNA as a model system for understanding how DNA polymerases could be engineered to recognize different classes of XNA substrates. However, none of the library diversification strategies attempted previously were able to produce an enzyme with robust TNA synthesis activity[8-10], suggesting that major structural changes would be required to properly align the TNA substrate in the active site pocket. In hindsight, this result is not entirely unexpected as the sugar plays an important role in the mechanism of DNA synthesis and requires careful alignment to allow in-line attack of the primer on the nucleoside triphosphate[62]. In an effort to explore new regions of the fitness landscape, we designed and constructed a homologous recombination library that derived from four archaeal DNA polymerases, each carrying the RSGA mutations known to enhance TNA substrate recognition[10]. This polymerase evolution strategy was inspired by the success of previous molecular breeding approaches in which the genes of DNA polymerase orthologs were shuffled and subjected to selection for improved DNA synthesis activity[25-27]. Following five rounds of enrichment and five additional rounds of directed evolution with intentional mutagenesis, we identified clone 10-92 as a highly divergent enzyme capable of synthesizing TNA at a rate of ~1 nt/s. This level of polymerase activity is within 5-200-fold of the rate of DNA synthesis observed for the natural archaeal DNA polymerases used as the basis of our library[63]. Interestingly, our strategy of selecting for improve catalytic activity also improved the fidelity of the enzyme, which is consistent with the notion that high catalytic activity requires faithful nucleotide selection.

The crystal structure of the ternary complex offered valuable insight into the catalytic activity achieved by the 10-92 TNA polymerase. Comparison of the 10-92 structure against its closest natural homolog, Kod DNA polymerase[57], reveals the large degree of structural alteration imparted by the homologous recombination and directed evolution process. Regions of the catalytic domain, and to lesser extent the N-terminal and exonuclease domains, underwent major structural refinement with some secondary structural elements shifting by >7Å relative to their position in Kod DNA polymerase. This finding suggests that the evolutionary distance required to transition natural DNA polymerases into highly specialized XNA enzymes may be larger than previously thought, as structural refinement of the active site pocket required major topological changes in the protein architecture. Exploring this problem with other genetic systems will test the robustness of our selection strategy and offer a possible route to valuable reagents that can advance future biotechnology applications.

In summary, we describe the evolution and structure of a highly efficient TNA polymerase that can be used to expand the field of synthetic genetics. These results offer a possible path for evolving new examples of XNA polymerases that can recognize distinct classes of genetic materials, including those for which such enzymes do not exist.

1.5 Methods

Design and construction of recombination library

The custom homologous recombination library was designed based on the sequence alignment of four homologous hyperthermophilic family-B DNA polymerases isolated from archaeal species *Thermococcus kodakaraensis* (Kod, GenBank: KP682508.1), *Thermococcus gorgonarius* (Tgo, GenBank: KP682507.1), *Pyrococcus* species GB-D Deep Vent (DV,

GenBank: KP682509.1), and Thermococcus sp. 9°N (9°N, GenBank: KP682506.1). The genes, ordered as gBlocks (Appendix A: Supplementary Table 1), were disassembled and reassembled by synthetic DNA shuffling[46] using a collection of primers (Appendix A: Supplementary Table 2). For extensive details, see Appendix A: Supplementary Methods.

Droplet-based optical polymerase sorting, recovery of sorted DNA, and library regeneration

Cells expressing the recombinant polymerase library were prepared by transforming 300 ng of the plasmid library into 20 µL of XL1-blue *E. coli*. Recovered cells were then used to inoculate 50 mL of LB-carbenicillin (50 µg/mL) liquid medium in a 250 mL baffled flask. The liquid culture was grown to confluency overnight at 37°C with shaking at 225 rpm. Overnight (500 µL, 1:100 v/v) was used to inoculate 50 mL of LB-carbenicillin (50 µg/mL) liquid medium in a 250 mL baffled flask and grown at 37°C with shaking at 225 rpm. At an OD600 of ~0.6 au, the culture was cooled to 25°C, induced with 1 mM IPTG, and incubated overnight at 25°C with shaking at 225 rpm. After expression, 1 mL of cell culture was collected and centrifuged for 5 min at 3,220 x g, and the supernatant was discarded. The cells were washed with 1 mL of commercial 1x ThermoPol buffer (20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1% TritonX-100, pH 8.8) for a total of four washes (4 mL). The bacterial pellet was suspended in 2 mL of 1x ThermoPol buffer and the absorbance was measured at 600 nm. Cells were diluted to an OD of 0.05 to enable encapsulation at occupancies of 0.1 cells per droplet according to a Poisson distribution. Just prior to emulsification, the cells were mixed with the reagents for the polymerase activity assay (PAA). The polymerase activity assay consists of 1 µM of a self-priming hairpin template labeled with Cy3 at the 5' end (Appendix A:

Supplementary Table 2, 30merHP.V2), 2 μM of a 3' end labeled Iowa Black quencher sequence (Appendix A: Supplementary Table 2, QP08.Iowa), and 100 μM of TNA triphosphates (tNTPs) in 1x ThermoPol buffer. tNTPs were obtained by chemical synthesis as described previously[49,64]. Single emulsions, formed utilizing a flow focusing geometry, were collected under a layer of mineral oil in 1.5 mL plastic micro-centrifuge tubes and incubated for 5 mins at 95°C to lyse the cells, followed by incubation at 55°C for various times. Following incubation, droplets were injected into a FADS device capable of droplet sorting. Droplet formation and sorting were performed using custom PDMS chips[48,65] and the custom LabVIEW v17.0 software was utilized to control user-defined parameters within the microfluidics experiments.

Plasmid DNA was recovered from the population of positively sorted droplets present in the collection tube by extraction with Pico-Break following the manufacturer's recommended protocol. The polymerase variant genes are PCR amplified and Gibson assembly is performed to insert the amplified genes back into plasmids. For extensive details, see Appendix A: Supplementary Methods.

Construction of mutagenic PCR libraries for directed evolution

Mutagenic PCR was performed using Kod-WT (exo-) DNA polymerase, which harbors two exonuclease silencing mutations (D141A, E143A) and is preferred for blunt ended amplification. The polymerase was expressed and purified from *E. coli*[66]. The mutagenic PCR was performed with biased dNTP ratios, and supplemented $MnCl_2$ and $MgCl_2$ as described previously[67,68]. The PCR reaction was performed with the following final concentration: 1x ThermoPol, 1 μM forward primer,1 μM reverse primer, 1 mM TTP, 1 mM dCTP, 0.2 mM dGTP, 0.2 mM dATP,

5.5 mM $MgCl_2$, 0.013 μM Kod-WT (exo-) polymerase, 100 ng of 5-270 plasmid and 2 fold serial dilution of $MnCl_2$ from 500-31.25 μM. PCR cycles were performed as followed: denature at 95°C for 2 min followed by 30 cycles of denature at 95°C for 1 min, anneal at 60°C for 1 min, extend at 72°C for 2.5 min, and followed by agarose gel analysis (Image Lab 6.1, Bio-Rad). For extensive details, see Appendix A: Supplementary Methods.

Colony picking and polymerase activity screen

Gibson assembly material from post regeneration were transformed into XL1 Blue *E. coli*, recovered and plated onto agar plates containing 50 ng/μL carbenicillin. Single colonies were picked and placed into 4 mL of LB media containing 50 ng/μL carbenicillin and grown in a shaking incubator at 225 rpm, overnight at 37°C. Overnight culture (40 μL) was inoculated into fresh 4 mL LB media containing 50 ng/μL carbenicillin and grown to an OD_{600} ~0.6. Cells were induced with 1 mM IPTG and expressed at 25°C, 225 rpm for 18-20 h. Cells were harvested, centrifuged (1,500 x g, 25°C, 10 min) and suspended in 10 mM Tris-HCl, pH 8.0, 500 mM NaCl and 10% glycerol and transferred to 1.5 mL centrifuge tubes. Tubes were placed into 70°C Eppendorff ThermoMixer with heated lid for 1 h, cooled for 1 h on ice, and centrifuged (16,300 x g, 4°C, 1 h). Clarified supernatant was assessed with Bradford dye against 10, 5, and 0 μM of purified parent polymerase.

Polymerase activity assay was performed with the following final concentrations: 1x ThermoPol, 0.5 μM of template 4, 0.5 μM of 5′-IR800-primer PBS8.20mer, 100 μM tNTPs, and 1 or 2 μL of lysate in a final volume of 20 μL. Primer-template were annealed in 1x ThermoPol at 90°C for 5 min and immediately placed on ice for 5 min, followed by the addition of tNTPs

and lysate on ice. Reactions were placed into thermal cycler at 55°C. At designated time points 1 µL of reaction was removed, mixed with 39 µL of quenching buffer (95% formamide, 25 mM EDTA) and denatured at 95°C for 10 min. Denatured samples were loaded (10 µL) into a 15% urea denaturing PAGE and run at 10 W for 1.5 h and visualized on Li-Cor Odyssey CLx Imager, Image Studio Lite v5.2.

Recombinant polymerase expression and purification of TNA polymerases

Protein expression and purification was performed using standard methods[66]. For extensive details, see Appendix A: Supplementary Methods.

Thermal stability challenge

Purified polymerase was quantified to 10 µM and aliquoted (100 µL) into 6, 1.5 mL Eppendorf tubes and placed into an Eppendorff ThermoMixer set 90°C with a thermal lid to prevent evaporation. At the designated time (1-6 h) a tube was removed and placed on ice until activity evaluation. The collected tubes were spun at 16,300 x g for 5 min at 4°C and UV quantified using a NanoDrop instrument. The polymerase was assayed for activity of TNA synthesis across a DNA template. Polymerase activity screen was performed with the following final concentrations: 1 µM primer-template duplex, 100 µM tNTPs, 1 µM heat-treated Polymerase in a 20 µL reaction. After 30 minutes of incubation, a 1 µL aliquot of the reaction was quenched by the addition of 39 µL of quenching buffer and visualized by denaturing polyacrylamide gel electrophoresis with fluorescent imaging on a Li-Cor Odyssey CLx Imager, Image Studio Lite v5.2.

Substrate recognition

TNA polymerases were evaluated for the ability to recognize and incorporated C5′-modified uracil TNA triphosphates (Trp and Phe) previously generated by chemical synthesis[69]. Polymerase activity screen was performed as described above with the exception of 2 equivalents of polymerase and combination of the C-5 modified tUTP substrate with standard tCTP, tGTP, and tATP substrates for a final concentration of 100 μM.

BLI validation of TNA aptamers

DNA templates (1.2 μM) encoding known TNA aptamers were annealed to a biotinylated-PBS8 primer (1 μM) in 1x ThermoPol at 90°C for 5 min and snap cooled, followed by the addition of 100 μM of respective TNA base mixes and 2 μM of 10-92 for a total volume of 750 μL. Primer extension was performed at 55°C. The reaction was quenched by freezing the sample, vacuum concentrating to a volume of <100 μL and adding 250 μL of quenching buffer. The quenched sample was heated to 90°C for 5 min and loaded onto a 15% urea-denaturing polyacrylamide gel and ran at 15W for 2.5 hrs. The product band was excised, electroeluted, buffer exchanged into water and UV quantified. For BLI analysis, the aptamer was diluted to 100 nM final concentration (1 mL) with BLI binding buffer (10mM HEPES pH 7.0, 150 mM NaCl, 3 mM EDTA, 0.05% Tween-20) heated to 90°C for 5 minutes and snap cooled. The protein target was prepared by performing 2-fold serial dilutions starting from 100 nM to obtain the following concentrations: 25, 12.5, 6.25, and 3.125 nM.

BLI studies were performed with four different concentrations of target and one buffer only sensor to determine the background. Prior to testing, sensors were equilibrated in BLI

binding buffer for at least 30 min. After aptamer folding and immobilization onto the biosensor tip, the BLI run was performed with the following steps: a buffer only baseline 60 s to equilibrate sensors, loading the aptamer for 200 s, a second buffer only baseline for 200 s, an association phase with the target protein for 600 s, and a dissociation phase for 600 s. The plate and reagents were incubated at 30°C for the duration of the experiment and for 10 min prior to each run. Data was acquired and analyzed using the Octet Data Acquisition software v12.0 and Analysis HT software v12.0.1.55, respectively. For full kinetic measurements, the buffer only baseline sample was used to subtract the background from all samples before applying Savitzky-Golay filtering and fitting both the association and dissociation curves together and applying a global fit to determine K_D and other metrics.

TNA polymerase fidelity

TNA polymerase fidelity was determined by sequencing the product of a complete replication cycle[9] and analyzing the Sanger sequences using CLC Main Workbench 23.0.5. For extensive details, see Appendix A: Supplementary Methods.

TNA versus DNA primed TNA synthesis

The DNA template (EM619, 1 μM) was annealed to either a TNA or DNA PBS8 primer (PBS8, 0.5 μM) in 1x ThermoPol at 95°C for 5 min and immediately snap cooled, followed by the addition of 10-92 polymerase (0.5 μM). The reaction was initiated with the addition of 100 μM of tNTPs for a total volume of 20 μL. Primer extension was performed at 55°C and quenched at designated timepoints with the addition of quenching buffer (95% Formamide, 25 mM EDTA). The quenched samples were heated to 95°C for 5 min and loaded onto a 15% urea-denaturing

polyacrylamide gel and ran at 15W for 1.5 hrs.

10-92 for reversion analysis

Ten single point reversions were introduced into the backbone of 10-92, creating ten new constructs. Forward primers were designed to contain the desired reversion mutation (Appendix A: Table 2) and reverse primers were ordered for whole plasmid PCR. The resulting PCR reaction was seeded into a KLD treatment and transformed into NEB 5alpha *E. coli* for colony picking. The resulting constructs were validated by Sanger sequencing, expressed in *E. coli*, purified as described above, validated by SDS-PAGE, and challenged in a primer extension.

Cloning Kod-RSGA with directed evolution mutations from rounds 6-10

The 10 mutations acquired by directed evolution were cloned into the backbone of Kod-RSGA, creating the construct Kod-RSGA+DE mutations. A gBlock was ordered to contain the linear dsDNA sequence encompassing the 10 mutations (Appendix A: Supplementary Table 1, Kod-RSGA+DE mutations). Corresponding primers (Appendix A: Supplementary Table 2, Vector Fwd KodRSGA+ DE mutations and Vector Rvs KodRSGA+ DE mutations) were ordered to facilitate the amplification of the remaining gene and expression vector from the original pGDR11-*kod*-RSGA construct. The resulting vector portion was used after a DNA clean up, DpnI treatment, and agarose gel purification. The insert and vector were ligated through Gibson assembly and transformed into NEB 5alpha *E. coli* for colony picking. The resulting construct was validated by Sanger sequencing, expressed in *E. coli*, and purified as described above, and challenged in a primer extension.

Polymerase kinetics

TNA polymerase kinetic measurements were performed by monitoring the reaction progress over time as the DNA primer-template (P-T) duplex (Appendix A: Supplementary Table 2, PBS8 & EM619) was extended with tNTPs. For each time point, 3 individual reaction replicates (15 μL) from a single master mix poised under single-turnover conditions with equimolar concentrations of polymerase (0.2 μM) and pre-annealed P-T duplex (0.2 μM each, heated at 90°C for 5 min, cooled on ice) in ThermoPol buffer (NEB, 2X) were pre-equilibrated for 5 minutes at 55°C. The reactions were then initiated by the addition of a preheated tNTPs mixture (200 μM of each tNTP). The reactions were stopped at designated time points by plunging the reaction vessel into powdered dry ice. Once all the time points were collected, the frozen reactions were thawed on a cold plate by adding 15 μL (6 μM) EvaGreen® dye, previously identified as the optimal intercalating dye to monitor synthesis of TNA on a DNA template[51]. Each reaction (25 μL) was transferred to a clear V-bottom 96-well plate and fluorescence intensity was measured (ex: 487/20 nm, em: 528/20 nm) using a Synergy Neo2 plate reader with the Gen5 v3.14.03 software (BioTek).

Baseline fluorescence was measured using Kod-WT (exo-), the parent polymerase with limited TNA synthesis ability[70], which was used to collect a 0 second time point by freezing the reaction immediately after the addition of the tNTP solution. Additionally, the maximum fluorescence value was obtained from a 2 hour reaction performed with each polymerase. Because Kod-RI does not reach full-length product, the maximum fluorescence value of Kod-RS was also used to for Kod-RI. The raw fluorescence data from the plate reader was normalized by subtracting the baseline and then dividing by the difference of the maximum (2 hour reaction) and minimum (baseline) fluorescence values. The fluorescence data was then converted to

nucleotides per polymerase as previously described [51] by multiplying data with the conversion factor F = [Template] * L / [Polymerase], where L is the length of extension in bases, then plotted over time. Rates in nucleotides per minute are extracted as the slope of the linear range of each curve. Kinetics data were processed using Microsoft Excel v16.66.

Cloning of truncated polymerases for crystallography

The 10-92 gene was PCR amplified from pGDR11-10-92 using the 10-92-Fwd and 10-92-Rvs primers (Appendix A: Supplementary Table 2) containing *NdeI* and *NotI* restriction enzyme sites, respectively. Two 10-92 C-terminal truncations, 10-92_760 and 10-92_750, were PCR amplified by substituting the 10-92-Rvs primer with either 10-92-Rvs_760 or 10-92-Rvs_750 primer (Appendix A: Supplementary Table 2), both containing the *NotI* restriction enzyme site. For extensive details, see Appendix A: Supplementary Methods.

Protein crystallization and structure determination

The primer-template (P/T) duplex was prepared by combining equal parts of the primer and template strands in Kod buffer (50 mM Tris-HCl pH 8.5, 200 mM NaCl, 0.1 mM EDTA, 1 mM DTT) and supplemented with 20 mM $MgCl_2$, heating at 95°C for 5 min and slow cooling to 10°C over 10 min. An initial binary complex was prepared by incubating 10-92_760 (6 mg mL^{-1}) with 1.2 M equivalents of the annealed P/T duplex at 37°C for 30 min. 3 M excess of dtTTP was added and incubated at 37°C for 30 min. The resulting n+1 binary complex was further incubated with 10 M excess tATP at 37°C for another 30 min to yield the final ternary complex. The10-92 ternary complex was screened against ~900 conditions in a hanging-drop format using a Mosquito crystallization robot (SPT LabTech). For extensive details about

crystallization conditions, see Appendix A: Supplementary Methods.

A 2.73Å dataset was collected at the Advanced Light Source (Lawrence Berkeley National Laboratory, Berkeley, CA) from a single crystal. Images were indexed, integrated, and merged using XDS[71,72]. The initial model was determined by molecular replacement using Phaser[73] with PDB structure 5VU8 as the search model, and the final model was determined using iterative rounds of manual building through Coot[74] and refinement with phenix[75]. The stereochemistry and geometry of all structures were validated with Molprobity[76] with the final refinement parameters summarized in Appendix A, Supplementary Table 4. Final coordinates and structure factors have been deposited in the Protein Data Bank. All molecular graphics were prepared with PyMOL[77].

1.6 Data availability

The crystal structure of 10-92 was deposited in the protein database (8T3X). The repository furthermore contains PDB structures of all ensemble refinements presented in this work. Other data are available in the main text, Supplementary Information, or from the authors upon reasonable request. Source data are provided with the original publication.

1.7 Acknowledgments

This work was supported by a grant to JC from the NSF (MCB 1946312).

1.8 Author Contributions

JC, VM, and EY conceived the project and designed the experiments. EY and VM evolved the enzyme, EY, VM, MH performed the biochemical characterization. VM, JL, MH, MJH, RQ,

KN, KH, JM, and NC contributed to protein structure determination. JC wrote the manuscript. All authors reviewed and commented on the manuscript.

1.9 Competing Interests

JC, VM, EY, and the University of California-Irvine have filed a patent application (PCT/US24/11595) on the composition and activity of the 10-92 TNA polymerase. No financial competing interests are declared.

1.10 Figures

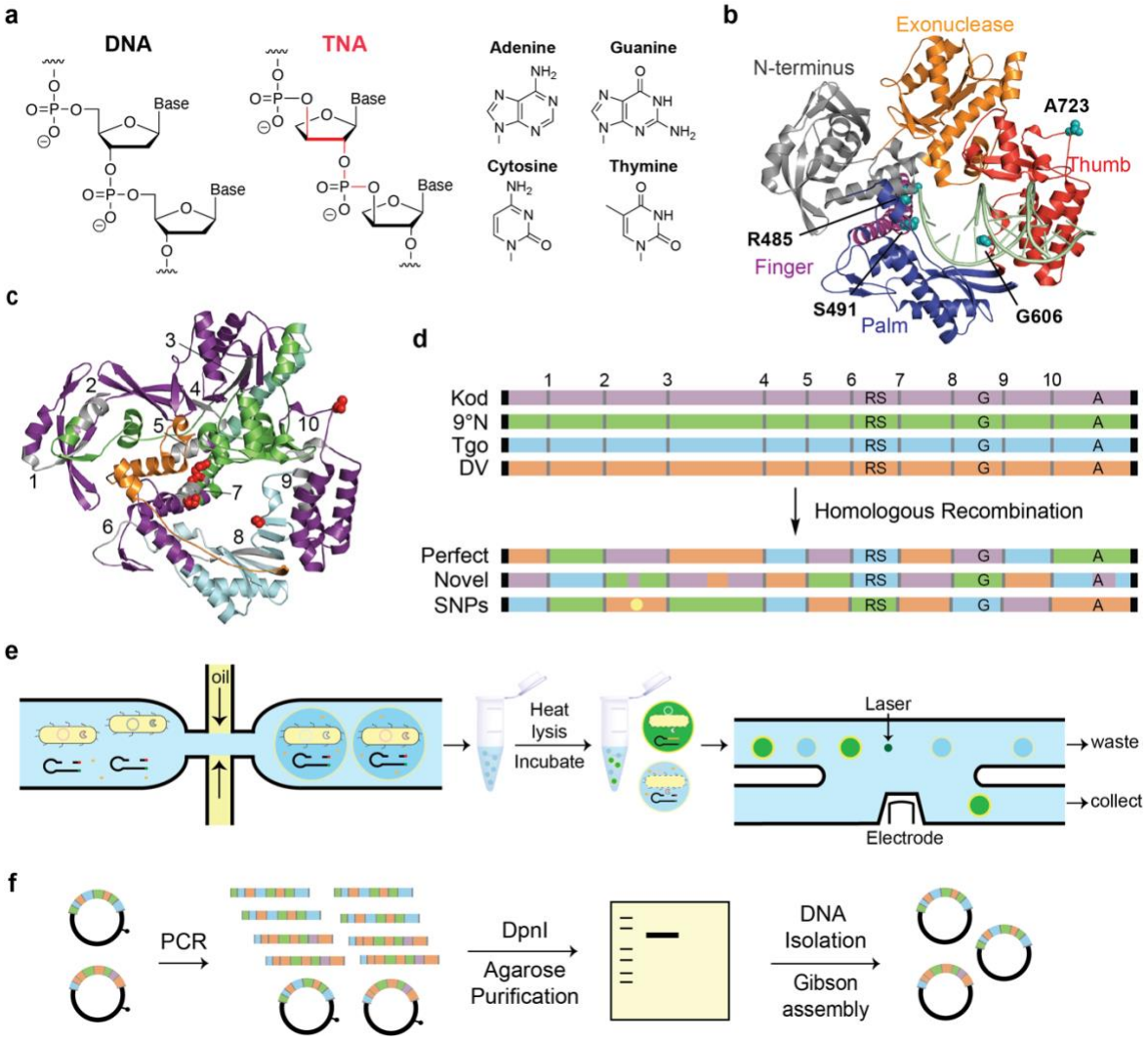


Figure 1.1 Homologous recombination as a path to engineered XNA polymerases.

(a) Chemical structure of the linearized backbone for natural DNA and 3',2'-α-L-threofuranosyl nucleic acid (TNA) with accompanied bases. (b) Crystal structure of Kod-RSGA TNA polymerase colored by domain with the RSGA mutations shown as space-filling models (PDB: 7RSU). (c) Structural representation of an example recombination product generated by DNA shuffling. Recombinant segments are colored according to their parent gene, with conserved recombination sites numbered and shown in grey. (d) Schematic illustration of a DNA shuffling approach applied to archaeal DNA polymerases. (e) Overview of the microfluidic polymerase engineering process known as droplet-based optical polymerase sorting (DrOPS). Microfluidic devices are used for droplet generation and sorting. Intermediate steps of E. coli lysis and primer extension occur off-chip. (f) Library assembly following iterative rounds of directed evolution involves Gibson assembly of PCR amplified genes with the backbone expression vector.

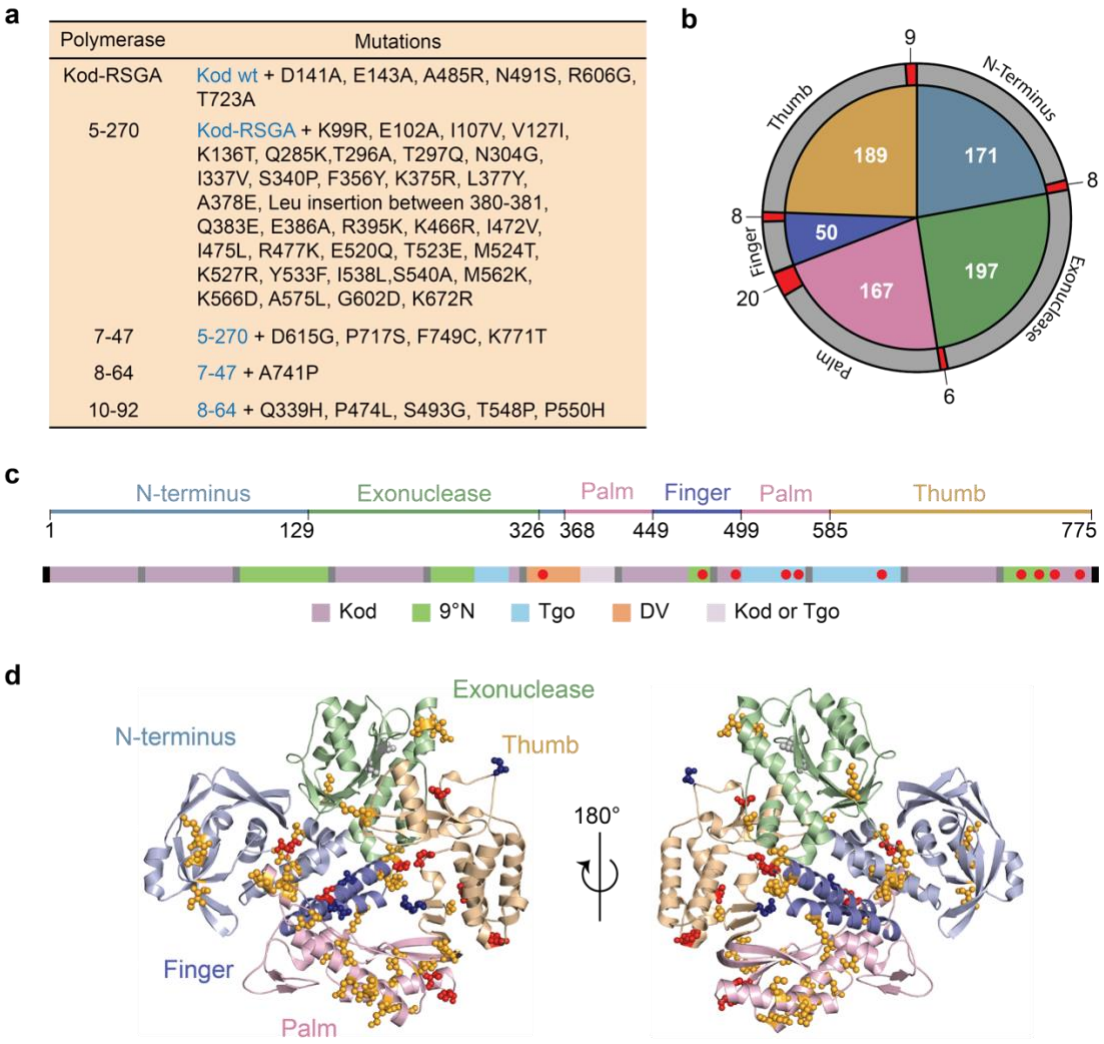


Figure 1.2 Directed evolution of a highly efficient TNA polymerase.

(a) Table summarizing the mutations that arose during the directed evolution of the 10-92 TNA polymerase. 5-270 is the product of DNA recombination and five rounds of enrichment for TNA synthesis activity, while 7-47, 8-64, and 10-92 are the products of directed evolution with intentional mutagenesis of clones 5-270, 7-47, and 8-64. (b) Pie chart denoting the distribution of amino acid mutations found in 10-92. Total residues per domain (inner values) and mutations observed for each domain (outer values). (c) Genomic map showing the recombination segments and mutations (red dots) that arose in the directed evolution of 10-92 from 5-270. (d) Ball-and-stick representation of the amino acid differences observed between wild-type Kod DNA polymerase and 10-92, mapped onto the crystal structure of wild-type Kod (PDB: 5OMF). Color scheme: exonuclease silencing mutations (gray), RSGA mutations from best previous TNA polymerase (blue), mutations acquired by recombination (orange), mutations acquired by directed evolution (red).

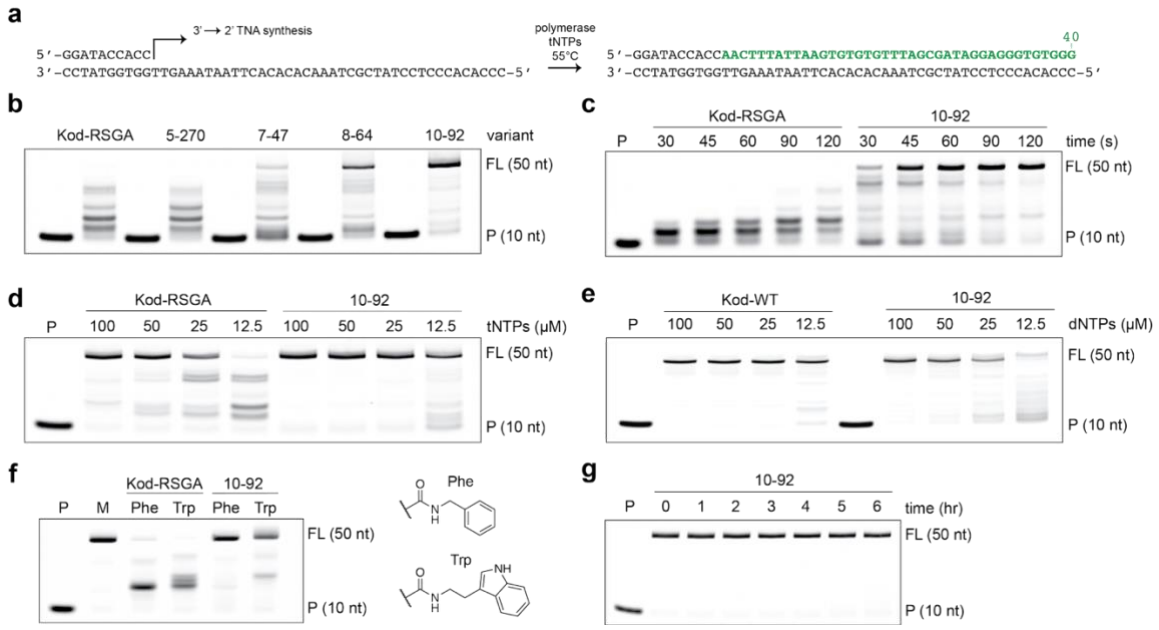


Figure 1.3 Functional analysis of TNA polymerases.

(a) Primer-extension assay. Sequence of the DNA primer-template duplex (black) before and after extension with TNA (green). (b) Activity screen of TNA polymerases. Key enzymes identified during the evolution process were challenged to extend the DNA primer with 40 TNA residues in a reaction time of 2 minutes. (c) Time course comparing the TNA synthesis activity of Kod-RSGA to 10-92. (d) tNTP substrate dependency. Product formation was evaluated after 30 minutes of incubation with decreasing substrate concentrations. (e) dNTP substrate specificity. Product formation was evaluated after 1 minute of incubation with decreasing substrate concentrations. (f) Synthesis of functionally enhanced TNA. Product formation was compared after 30 minutes of incubation for tNTP mixtures containing C5-modified tUTP monomers bearing either phenylalanine (Phe) or tryptophan (Trp) side chains with amide linkages. (g) Thermal stability. TNA synthesis activity was assessed after 1-6 hours of polymerase incubation at 90°C. Unless otherwise noted, reactions were performed by incubating a 5′ IR680-labeled DNA primer-template duplex (1 mM) with 1 mM polymerase and 100 mM tNTPs in 1X ThermoPol buffer at 55°C. C5-modified tNTP mixtures contain 2 mM polymerase. Primer-extension reactions were analyzed by denaturing polyacrylamide gel electrophoresis with fluorescent imaging on a LI-COR imager. P, primer; F, full-length product; M, marker. Representative gels of two independent experiments are shown.

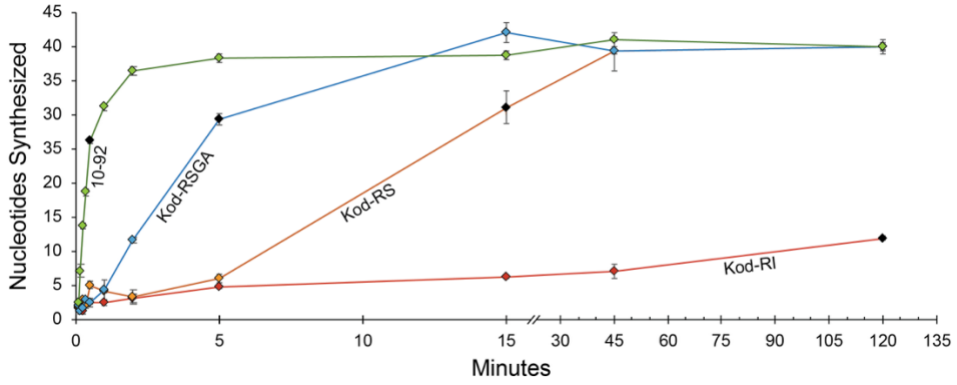


Figure 1.4 Kinetic analysis of TNA synthesis.

Kinetic curves are shown for primer extension reactions catalyzed by Kod-RI (red), Kod-RS (orange), Kod-RSGA (blue), and 10-92 (green). The end of the linear range for each polymerase is marked by a black diamond. The slope of the linear range corresponds to the rate of TNA synthesis in nucleotides per minute. Error bars represent the standard error of the mean (S.E.M) of 3 independent reaction replicates.

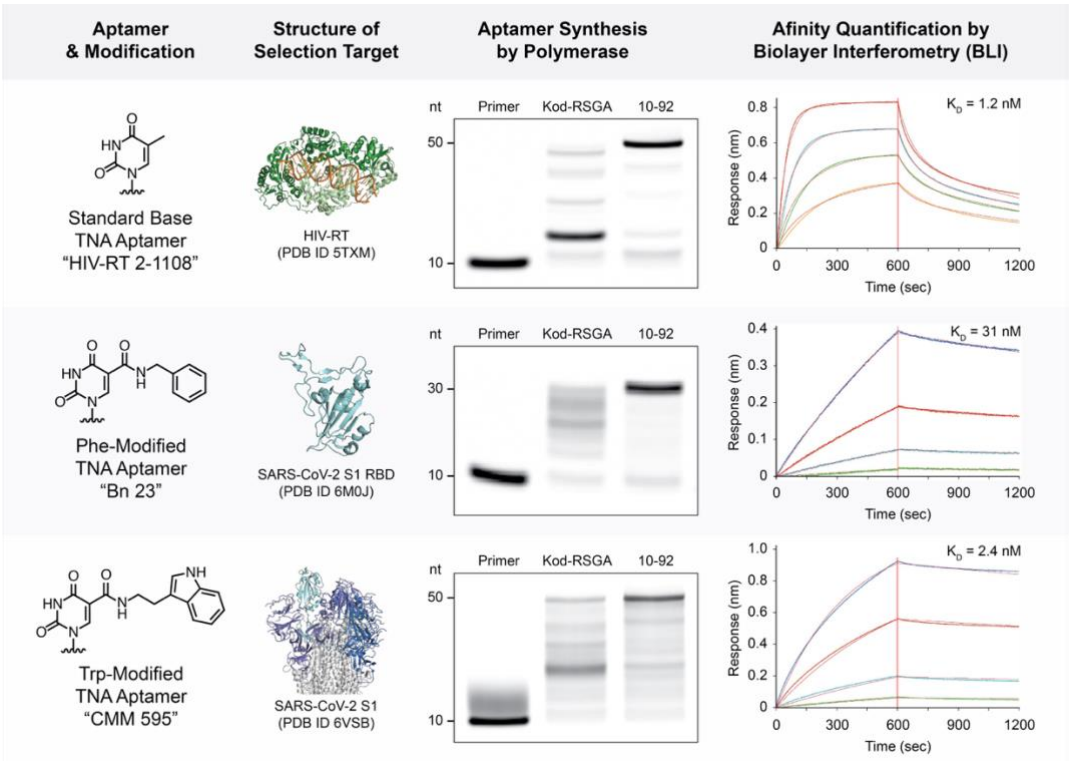


Figure 1.5 Enhanced functional activity.

The 10-92 TNA polymerase synthesizes standard and base modified TNA aptamers with higher efficiency than Kod-RSGA. Representative gels of two independent experiments are shown. BLI sensorgrams confirm that the enzymatically synthesized aptamers bind to their cognate protein targets.

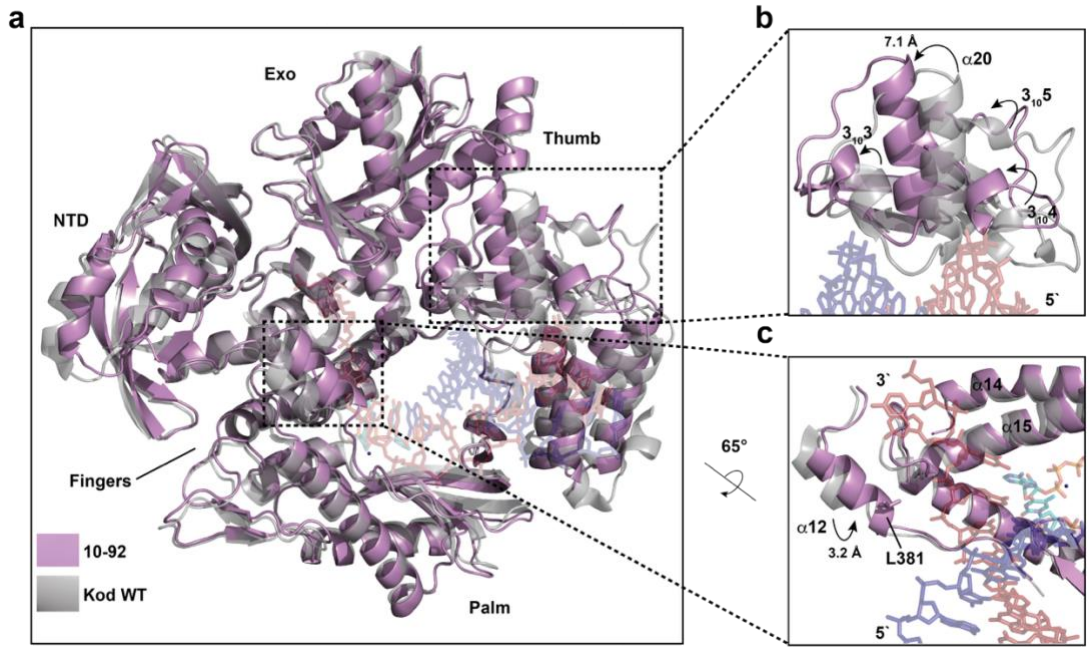


Figure 1.6 Structural topology of the 10-92 TNA polymerase.

(a) Structural overlay of the closed ternary conformations of the 10-92 TNA polymerase (purple, PDB: 8T3X) and Kod DNA polymerase (transparent gray, PDB: 5OMF) reveal distinct conformational rearrangements (RMSD 2Å) in the NTD and thumb regions of the enzyme. Secondary structural transitions observed in the (b) thumb and (c) palm impact duplex recognition. The palm subdomain contains the Leu381 insertion obtained by homologous recombination. Primer, template, and TNA substrates are colored blue, red, and teal, respectively.

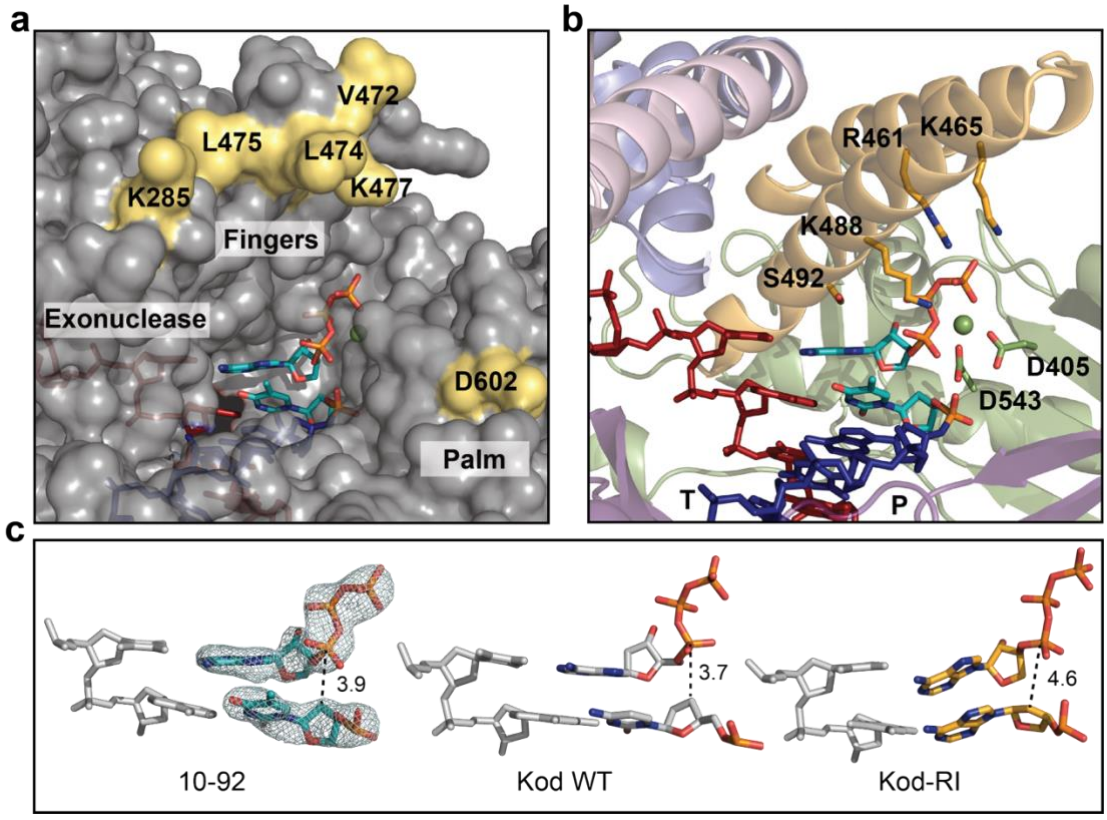


Figure 1.7 Active site of the 10-92 TNA polymerase.

(a) Surface rendering of the active site region of the 10-92 TNA polymerase trapped in a closed ternary conformation with a tATP substrate (teal) snugly fit in an active site cavity. Mutations observed in the active site pocket are shown in yellow. (b) The tATP substrate is stabilized by direct contacts with residues in the finger subdomain (orange stick) and a magnesium ion (green spheres), which is further coordinated by catalytic residues, D405 and D543 (green sticks) in the palm subdomain. Primer and template are shown in blue and red, respectively. (c) Comparison of the nascent base pair between the 10-92 and Kod-RI TNA polymerases. TNA and DNA nucleotides are shown as stick models, with a polder map of the tT^d and tATP substrates for 10-92 shown in gray and contoured at 4.0σ.

1.11 References

1. Nikoomanzar, A., Chim, N., Yik, E. J. & Chaput, J. C. Engineering polymerases for applications in synthetic biology. *Q. Rev. Biophys.* 53, e8 (2020).

2. Houlihan, G., Arangundy-Franklin, S. & Holliger, P. Engineering and application of polymerases for synthetic genetics. *Curr. Opin. Biotechnol.* 48, 168-179 (2017).

3. Aschenbrenner, J. & Marx, A. DNA polymerases and biotechnological applications. *Curr. Opin. Biotechnol.* 48, 187-195 (2017).

4. Tunyasuvunakool, K. *et al.* Highly accurate protein structure prediction for the human proteome. *Nature* 596, 590-596 (2021).

5. Wang, J. *et al.* Scaffolding protein functional sites using deep learning. *Science* 377, 387-394 (2022).

6. Chim, N., Meza, R. A., Trinh, A. M., Yang, K. & Chaput, J. C. Following replicative DNA synthesis by time-resolved X-ray crystallography. *Nat. Commun.* 12, 2641 (2021).

7. Doublie, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. *Nature* 391, 251-258 (1998).

8. Larsen, A. C. *et al.* A general strategy for expanding polymerase function by droplet microfluidics. *Nat. Commun.* 7, 11235 (2016).

9. Nikoomanzar, A., Vallejo, D. & Chaput, J. C. Elucidating the Determinants of Polymerase Specificity by Microfluidic-Based Deep Mutational Scanning. *ACS Synth. Biol.* 8, 1421-1429 (2019).

10. 1Nikoomanzar, A., Vallejo, D., Yik, E. J. & Chaput, J. C. Programmed allelic mutagenesis of a DNA polymerase with single amino acid resolution. *ACS Synth. Biol.* 9, 1873-1881 (2020).

11. Pinheiro, V. B. *et al.* Synthetic genetic polymers capable of heredity and evolution. *Science* 336, 341-344 (2012).

12. Hoshino, H., Kasahara, Y., Kuwahara, M. & Obika, S. DNA Polymerase Variants with High Processivity and Accuracy for Encoding and Decoding Locked Nucleic Acid Sequences. *J. Am. Chem. Soc.* 142, 21530-21537 (2020).

13. Liu, C. *et al.* Phosphonomethyl Oligonucleotides as Backbone-Modified Artificial Genetic Polymers. *J. Am. Chem. Soc.* 140, 6690-6699 (2018).

14. Kuwahara, M. *et al.* Systematic analysis of enzymatic DNA polymerization using oligo-DNA templates and triphosphate analogs involving 2',4'-bridged nucleosides. *Nucleic Acids Res.* 36, 4257-4265 (2008).

15. Loakes, D., Gallego, J., Pinheiro, V. B., Kool, E. T. & Holliger, P. Evolving a polymerase for hydrophobic base analogues. *J. Am. Chem. Soc.* 131, 14827-14837 (2009).

16. Ramsay, N. *et al.* CyDNA: synthesis and replication of highly Cy-dye substituted DNA by an evolved polymerase. *J. Am. Chem. Soc.* 132, 5096-5104 (2010).

17. Chen, T. *et al.* Evolution of thermophilic DNA polymreases for the recognition and amplification of C2'-modified DNA. *Nat. Chem.* 8, 556-562 (2016).

18. Freund, N. *et al.* A two-residue nascent-strand steric gate controls synthesis of 2'-O-methyl- and 2'-O-(2-methoxyethyl)-RNA. *Nat. Chem.* 15, 91-100 (2023).

19. Lelyveld, V. S., Fang, Z. & Szostak, J. W. Trivalent rare earth metal cofactors confer rapid NP-DNA polymerase activity. *Science* 382, 423-429 (2023).

20. Arangundy-Franklin, S. *et al.* A synthetic genetic polymer with an uncharged backbone chemistry based on alkyl phosphonate nucleic acids. *Nat. Chem.* 11, 533-542 (2019).

21. Chim, N., Shi, C., Sau, S. P., Nikoomanzar, A. & Chaput, J. C. Structural basis for TNA synthesis by an engineered TNA polymerase. *Nat. Commun.* 8, 1810 (2017).

22. Crameri, A., Raillard, S.-A., Bermudez, E. & Stemmer, W. P. C. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* 391, 288-291 (1998).

23. Ness, J. E. *et al.* Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. *Nat. Biotechnol.* 20, 1251-1255 (2002).

24. Drummond, D. A., Silberg, J. J., Meyer, M. M., Wilke, C. O. & Arnold, F. H. On the conservative nature of intragenic recombination. *Proc. Natl. Acad. Sci. U.S.A.* 102, 5380-5385 (2005).

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

26. Baar, C. *et al.* Molecular breeding of polymerases for resistance to environmental inhibitors. *Nucleic Acids Res.* 39, e51 (2011).

27. Leconte, A. M. *et al.* Directed evolution of DNA polymerases for next-generation sequencing. *Angew. Chem. Int. Ed. Engl.* 49, 5921-5924 (2010).

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

29. Eschenmoser, A. Chemical etiology of nucleic acid structure. *Science* 284, 2118-2124 (1999).

30. Culbertson, M. C. *et al.* Evaluating TNA stability under simulated physiological conditions. *Bioorg. Med. Chem. Lett.* 26, 2418-2421 (2016).

31. Lee, E. M., Setterholm, N. A., Hajjar, M., Barpuzary, B. & Chaput, J. C. Stability and mechanism of threose nucleic acid toward acid-mediated degradation. *Nucleic Acids Res.* 51, 9542-9551 (2023).

32. Dunn, M. R., Jimenez, R. M. & Chaput, J. C. Analysis of aptamer discovery and technology. *Nat. Rev. Chem.* 1, 0076 (2017).

33. 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, 183-187 (2012).

34. Dunn, M. R., McCloskey, C. M., Buckley, P., Rhea, K. & Chaput, J. C. Generating biologically stable TNA aptamers that function with high affinity and thermal stability. *J. Am. Chem. Soc.* 142, 7721-7724 (2020).

35. McCloskey, C. M. *et al.* Evolution of Functionally Enhanced alpha-l-Threofuranosyl Nucleic Acid Aptamers. *ACS Synth. Biol.* 10, 3190-3199 (2021).

36. Li, X., Li, Z. & Yu, H. Selection of threose nucleic acid aptamers to block PD-1/PD-L1 interaction for cancer immunotherapy. *Chem. Commun.* 56, 14653-14656 (2020).

37. Wang, Y. *et al.* An RNA-cleaving threose nucleic acid enzyme capable of single point mutation discrimination. *Nat. Chem.* 14, 350-359 (2022).

38. Wang, Y. *et al.* A Threose Nucleic Acid Enzyme with RNA Ligase Activity. *J. Am. Chem. Soc.* 143, 8154-8163 (2021).

39. Yang, K., McCloskey, C. M. & Chaput, J. C. Reading and Writing Digital Information in TNA. *ACS Synth. Biol.* 9, 2936-2942 (2020).

40. Wang, Y., Nguyen, K., Spitale, R. C. & Chaput, J. C. A biologically stable DNAzyme that efficiently silences gene expression in cells. *Nat. Chem.* 13, 319-326 (2021).

41. Wang, F. *et al.* Synthetic alpha-l-Threose Nucleic Acids Targeting BcL-2 Show Gene Silencing and in Vivo Antitumor Activity for Cancer Therapy. *ACS Appl. Mater. Interfaces* 11, 38510-38518 (2019).

42. Shigeo, M. e. a. Shorter is better: the a-(L)-threofuranosyl nucleic acid modification improves stability, potency, safety, and Ago2 binding and mitigates off-target effects of small interfering RNAs. *J. Am. Chem. Soc.* 145, in press (2023).

43. Gould, S. J. *Wonderful life : the Burgess Shale and the nature of history*. 1st edn, (W.W. Norton, 1989).

44. Hottin, A. & Marx, A. Structural Insights into the Processing of Nucleobase-Modified Nucleotides by DNA Polymerases. *Acc. Chem. Res.* 49, 418-427 (2016).

45. Dunn, M. R., Otto, C., Fenton, K. E. & Chaput, J. C. Improving Polymerase Activity with Unnatural Substrates by Sampling Mutations in Homologous Protein Architectures. *ACS Chem. Biol.* 11, 1210-1219 (2016).

46. Yik, E. J., Maola, V. A. & Chaput, J. C. Engineering TNA polymerases through iterative cycles of directed evolution. *Methods Enzymol.* 691, 29-59 (2023).

47. Minshull, J. & Stemmer, W. P. Protein evolution by molecular breeding. *Curr. Opin. Chem. Biol.* 3, 284-290 (1999).

48. Vallejo, D., Nikoomanzar, A., Paegel, B. M. & Chaput, J. C. Fluorescence-Activated Droplet Sorting for Single-Cell Directed Evolution. *ACS Synth. Biol.* 8, 1430-1440 (2019).

49. Liao, J.-Y., Bala, S., Ngor, A. K., Yik, E. J. & Chaput, J. C. P(V) Reagents for the Scalable Synthesis of Natural and Modified Nucleoside Triphosphates. *J. Am. Chem. Soc.* 141, 13286-13289 (2019).

50. Mei, H. *et al.* Synthesis and Evolution of a Threose Nucleic Acid Aptamer Bearing 7-Deaza-7-Substituted Guanosine Residues. *J. Am. Chem. Soc.* 140, 5706-5713 (2018).

51. Nikoomanzar, A., Dunn, M. R. & Chaput, J. C. Evaluating the rate and substrate specificity of laboratory evolved XNA polymerases. *Anal. Chem.* 89, 12622-12625 (2017).

52. Goodman, M. F., Creighton, S., Bloom, L. B. & Petruska, J. Biochemical basis of DNA replication fidelity. *Crit. Rev. Biochem. Mol. Biol.* 28, 83-126 (1993).

53. Kundu, N., McCloskey, C. M., Hajjar, M. & Chaput, J. C. Parameterizing the Binding Properties of XNA Aptamers Isolated from a Low Stringency Selection. *Biochemistry* 62, 3245-3254 (2023).

54. Lozoya-Colinas, A., Yu, Y. & Chaput, J. C. Functionally Enhanced XNA Aptamers Discovered by Parallelized Library Screening. *J. Am. Chem. Soc.*, in press (2023).

55. Bala, S. *et al.* Synthesis of 2'-Deoxy-alpha-l-threofuranosyl Nucleoside Triphosphates. *J. Org. Chem.* 83, 8840-8850 (2018).

56. Rodriguez, A. C., Park, H.-W., Mao, C. & Beese, L. S. Crystal structure of a Pol A family DNA polymerase from the hyperthermophilic Archaeon *Thermococcus sp. 9°N-7 J. Mol. Biol.* 299, 447-462 (2000).

57. Kropp, H. M., Betz, K., Wirth, J., Diederichs, K. & Marx, A. Crystal structures of ternary complexes of archaeal B-family DNA polymerases. *PLoS One* 12, e0188005 (2017).

58. Wang, Y., Ngor, A. K., Nikoomanzar, A. & Chaput, J. C. Evolution of a general RNA-cleaving FANA enzyme. *Nat. Commun.* 9, 5067 (2018).

59. Chaput, J. C. & Szostak, J. W. TNA synthesis by DNA polymerases. *J. Am. Chem. Soc.* 125, 9274-9275 (2003).

60. Vastmans, K. *et al.* Enzymatic incorporation in DNA of 1,5-anhydrohexitol nucleotides. *Biochemistry* 39, 12757-12765 (2000).

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

62. Steitz, T. A., Smerdon, S. J., Jager, J. & Joyce, C. M. A unified polymerase mechanism for nonhomologous DNA and RNA polymerases. *Science* 266, 2022-2025 (1994).

63. Hogrefe, H. H., Cline, J., Lovejoy, A. E. & Nielson, K. B. DNA polymerases from hyperthermophiles. *Methods Enzymol.* 334, 91-116 (2001).

64. Sau, S. P., Fahmi, N. E., Liao, J.-Y., Bala, S. & Chaput, J. C. A scalable synthesis of α-L-threose nucleic acid monomers. *J. Org. Chem.* 81, 2302-2307 (2016).

65. Vallejo, D., Nikoomanzar, A. & Chaput, J. C. Directed evolution of custom polymerases using droplet microfluidics. *Methods Enzymol.* 644, 227-253 (2020).

66. Nikoomanzar, A., Dunn, M. R. & Chaput, J. C. Engineered polymerases with altered substrate specificity: expression and purification. *Curr. Protoc. Nucleic Acid Chem.* 69, 4.75 (2017).

67. Cadwell, R. C. & Joyce, G. F. Randomization of genes by PCR mutagenesis. *PCR Methods Appl.* 2, 28-33 (1992).

68. Chaput, J. C. & Szostak, J. W. Evolutionary optimization of a nonbiological ATP binding protein for improved folding stability. *Chem. Biol.* 11, 865-874 (2004).

69. Li, Q. *et al.* Synthesis and Polymerase Recognition of Threose Nucleic Acid Triphosphates Equipped with Diverse Chemical Functionalities. *J. Am. Chem. Soc.* 143, 17761-17768 (2021).

70. Medina, E., Yik, E. J., Herdewijn, P. & Chaput, J. C. Functional Comparison of Laboratory-Evolved XNA Polymerases for Synthetic Biology. *ACS Synth. Biol.* 10, 1429-1437 (2021).

71. Kabsch, W. Xds. *Acta Crystallogr. D Biol. Crystallogr.* 66, 125-132 (2010).

72. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* 66, 133-144 (2010).

73. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658-674 (2007).

74. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486-501 (2010).

75. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* 68, 352-367 (2012).

76. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 66, 12-21 (2010).

77. The PYMOL molecular graphics program (DeLano Scientific, San Carlos, CA, USA, 2002).

CHAPTER 2: Engineering TNA Polymerases through Iterative Cycles of Directed Evolution

Eric J. Yik[1†], Victoria A. Maola[1†], and John C. Chaput[1-4]*

Affiliations

[1]Department of Pharmaceutical Sciences, University of California, Irvine, CA 92697-3958 USA

[2]Department of Chemistry, University of California, Irvine, CA 92697-3958 USA

[3]Department of Molecular Biology and Biochemistry, University of California, CA 92697-3958 USA

[4]Department of Chemical and Biomolecular Engineering, University of California, Irvine, CA 92697-3958 USA

[†] These authors contributed equally

This chapter is a reprint of the material as it appears in Methods in Enzymology: *Methods Enzymol.* 691, 29-59 (2023). https://doi.org/10.1016/bs.mie.2023.04.014, used with permission from Elsevier.

2.1 Abstract

DNA polymerases are important tools for biotechnology, synthetic biology, and chemical biology as they are routinely used to amplify and edit genetic information. However, natural polymerases do not recognize artificial genetic polymers (also known as xeno-nucleic acids or XNAs) with unique sugar-phosphate backbone structures. Directed evolution offers a possible solution to this problem by facilitating the discovery of engineered versions of natural polymerases that can copy genetic information back and forth between DNA and XNA. Here we report a directed evolution strategy for discovering polymerases that can synthesize threose nucleic acid (TNA) on DNA templates. The workflow involves library generation and expression in *E. coli*, high-throughput microfluidics-based screening of uniform water-in-oil droplets, plasmid recovery, secondary screening, and library regeneration. This technique is sufficiently general that it could be applied to a wide range of problems involving DNA modifying enzymes.

2.2 Introduction

Directed evolution is a protein engineering technique that is commonly used to change the substrate specificity of an enzyme or to introduce new physical or chemical properties into a protein scaffold[1]. The process mirrors natural selection by allowing a population of sequences to change over time as it adapts to increasing levels of selective pressure. During a directed evolution experiment, proteins with desired phenotypes increase in abundance, while inactive variants go extinct. As an example, efforts to increase the thermal stability of a protein often involve forcing a population of sequences to survive a thermal denaturation step that is designed to separate stably folded proteins from the more weakly folded population. Although approaches to directed evolution have changed over time due to improvements in molecular biology

techniques, the process itself continues to play an important role in biotechnology where specialized enzymes are needed to catalyze chemical reactions that differ from those found in nature.

Library generation is a critical step in the directed evolution process as it defines the region of sequence space that will be surveyed in hopes of identifying new variants with desired functional properties. The most common techniques used for library construction include rational design by site-directed mutagenesis, site-saturation mutagenesis, deep mutational scanning, error-prone PCR (epPCR), and homologous recombination[2]. Most of these techniques rely on the introduction of random or controlled sequence diversity into a gene that leads to the production of nonsynonymous mutations in translated proteins. Homologous recombination, also known as DNA shuffling, offers a different route to library construction by allowing sequence diversity to accumulate through a process that recombines useful mutations found in a population of related sequences. In its original form, DNA shuffling involved the random fragmentation of a gene followed by its reassembly using the polymerase chain reaction (PCR) to produce a pool of chimeric sequences[3]. In recent years, this procedure has been refined using synthetic DNA fragments that allow researchers to precisely define the recombination sites[4].

In vitro selection technologies that link the phenotype of a protein to its genetic sequence are vital to the success of a directed evolution experiment[5]. Library size and selection format are important parameters to consider when choosing a selection technology. Since many excellent reviews have been dedicated to this subject, we will briefly highlight a few examples. Colony screening is the simplest approach to protein engineering. It requires the least amount of technical expertise and instrumentation but is inherently low throughput due to the need to evaluate library members on an individual basis. mRNA display and phage display allow very large library

diversities (10^9-10^{12}) to be queried *en masse* but are better suited for binding than catalysis[6]. Compartmentalization strategies, such as the one described in this protocol, offer attractive routes to enzyme engineering because they provide environments that support multiple turnover activity. Compartmentalized self-replication (CSR)[7], compartmentalized self-tagging (CST)[8], and droplet-based optical polymerase sorting (DrOPS) are examples of in vitro selection technologies that are performed in a water-in-oil (w/o) droplet[9]. These approaches have been particularly successful in the evolution of DNA polymerases for novel activities[10].

This chapter focuses on methods that were developed to change the substrate specificity of natural DNA polymerases (Fig. 2.1). We were interested in engineering DNA polymerases for the ability to synthesize threose nucleic acid (TNA) on DNA templates. Although this example is highly specific, the steps taken were sufficiently general that they could be applied to many other problems in protein engineering. We examine the intricacies of library generation by homologous recombination using synthetic DNA fragments that are randomly assembled by PCR. This portion of the method is extensive, as a detailed description is required to generate a library by homologous recombination. We then describe how the library is carried through iterative cycles of selection and amplification with a special emphasis on library recovery and regeneration. The intricate details of our microfluidic-based approach of droplet-based optical polymerase sorting (DrOPS) are only lightly discussed as this technology was the subject of a previous chapter[11]. We suggest that the robustness of this method should enable others to broadly expand the substrate specificity of natural DNA polymerases.

2.3 Library Design

2A. Designing a library for homologous recombination

In this section, we describe the *in silico* process of designing genes that will be used for homologous recombination. We recommend using molecular biology cloning software, such as CLC Main, when designing and annotating genes for library assembly. These programs allow users to establish electronic workflows that simplify vector design, increase transparency, and minimize errors. Genes are typically designed at the DNA level within the framework of a vector map. Annotated vector maps should contain all the necessary information for a host cell to propagate and express the gene, including the origin of replication, antibiotic resistance, cloning sites, promoter region, and protein affinity tags. Although the choice of which expression vector is best for a given protein is beyond the scope of this protocol, users should consider such variables as desired protein expression level and the expression of endogenous proteins that may impact the selection strategy. For example, in our system, the newly expressed polymerases are challenged to synthesize TNA on a self-priming DNA hairpin, which is susceptible to nuclease degradation. To overcome this problem, we chose the vector pGDR11, which contains a T5 promoter that allows for protein expression in nuclease deficient cell lines such as XL1-Blue.

Library assembly by homologous recombination requires the user to identify homologs of the parent enzyme. This is typically done by BLAST analysis using a parent sequence downloaded from a protein sequence database (e.g., Genbank or Uniprot). Once the genes chosen for homologous recombination are identified, they can be modified with known beneficial mutations that a user may wish to include in the library, and each sequence is codon optimized to ensure efficient expression in *E. coli*. Homologous recombination sites are then chosen by identifying regions of the gene that exhibit high sequence identity between the various homologs. We recommend a minimum of 21 consecutive nucleotides to obtain T_m values in the range of 55-70°C. As needed, these regions are modified by codon swapping to ensure 100% sequence identity at the

recombination sites. Additionally, the parent gene should be modified at the 5' and 3' ends with short segments (20-40 nt) matching the vector insertion site. These overlapping regions are needed to insert the gene into the expression vector via Gibson assembly[12] (Fig. 2.2). Finally, the computationally designed genes are ordered as linear, dsDNA gBlocks, which allows for sequence lengths up to 3 kilobases (kb). Once received, gBlocks are used directly without the need for PCR amplification or purification.

The example described in this chapter involves the directed evolution of an engineered DNA polymerase with improved DNA-dependent TNA polymerase activity. TNA is an artificial genetic polymer comprised of repeating units of a-L-threofuranosyl nucleotides that are connected by 2',3'-phosphodiester linkages[13]. TNA is considered a difficult polymer for polymerase recognition as its backbone repeat unit is one atom shorter than that of DNA and RNA. Previous work focused on an archaeal B-family DNA polymerase isolated from the hyperthermophilic species *Thermococcus kodakaraensis* (Kod). High throughput screening assays led to the discovery of Kod-RSGA, an engineered form of Kod DNA polymerase carrying the exonuclease silencing mutations D141A and E143A as well as the TNA recognition mutations A485R, N491S, R606G, and T723A[14]. In effort to design a library for homologous recombination, BLAST analysis of Kod DNA polymerase was performed to identify other archaeal B-family DNA polymerases, including *Thermococcus gorgonarius* (Tgo), *Pyrococcus* species GB-D Deep Vent (DV), and *Thermococcus* species 9°N. The four genes were aligned, codon optimized, and modified in silico to contain the six mutations found in Kod-RSGA. Sequence alignment identified 10 recombination sites that were further modified by codon swapping to ensure 100% sequence identity for homologous recombination. The four genes (Kod-RSGA, Tgo-RSGA, DV-RSGA, and 9°N-RSGA) were reviewed for accuracy and ordered as gBlocks from IDT. Additionally, the gBlocks

contained overlapping segments of 24 and 21 nt at the 5′ and 3′ ends, respectively, to facilitate their insertion into the pGDR11 vector via Gibson assembly[12].

Materials

National Center for Biotechnology Information: BLAST and GenBank

CLC Main Benchmark

Custom DNA gBlock (IDT, Coralville, IA)

1. Identify the enzyme or protein to be optimized by directed evolution.
2. Perform a BLAST analysis to identify related genes that will be used for homologous recombination.
3. Confirm the sequence of each enzyme or protein by downloading the sequence from GenBank. In the current study, the following hyperthermophilic archaeal B-family DNA polymerases were used: *Thermococcus kodakaraensis* (Kod, KP682508.1), *Thermococcus gorgonarius* (Tgo, KP682507.1), *Pyrococcus* species GB-D Deep Vent (DV, KP682509.1), *Thermococcus* sp. 9°N (9°N, KP682506.1).
4. Perform a protein sequence alignment to compare the length and sequence composition of each enzyme at the DNA and protein levels.
5. As needed, modify the wild-type sequence at the DNA level to include beneficial mutations known to improve the desired activity. In our case, each polymerase sequence was modified to contain the exonuclease silencing mutations D141A and E143A and the gain-of-function mutations A485R, N491S, R606G, and T723A.

6. From the DNA sequence alignment, identify regions of high sequence identity that can be used for homologous recombination. If needed, codon swap residues in conserved regions to achieve 100% sequence identity. Recombination sites are typically ≥21 nt to obtain T_m values in the range of 55-70°C.
7. Digitally insert the optimized gene sequences into the vector. In our case, the genes were inserted between a 5′ sequence encoding a His-tag and a SalIrestriction site in a pGDR11 vector.
8. Identify flanking regions at the 5′ and 3′ ends of gene that can be used to insert the genes into a protein expression vector by Gibson assembly.

   Suggested primer length is at least 20-40 nt.
9. Review the sequences for accuracy before ordering as gBlocks from IDT.

2B. Amplification of the expression vector for Gibson assembly

Here, we discuss primer design and provide technical recommendations for successful amplification of the vector backbone required for Gibson assembly[12]. As mentioned above, the parent gene is purchased as a gBlock with short segments (20-40 nt) matching the vector insertion site. These overlapping regions are needed to insert the gene into the expression vector via Gibson assembly (Fig. 2.2). Considerations for primer design include nucleotide length, composition, and salt concentration, as these factors affect the melting temperature of primers in a PCR reaction. The primer sequences should be designed in the 5′ to 3′ direction with a recommended T_m range of 55-70°C, which can be accomplished with a 15-40 nt sequence. In addition, the T_m range of the forward and reverse primer pairs should vary by ≤3°C. For example, the forward primer (5′-TGA GTC GAC CTG CAG CCA AGC TTA ATT AGC TGA GC-3′) is 35 nt in length with a T_m of

66.3°C, while the reverse primer (5′-GCT GCC GCG CGG CAC CAG-3′) is 18 base pair in length with a T_m of 68.3°C for our vector backbone generation. We found that *in silico* mapping of primers in CLC Main, in conjunction with IDT OligoAnalyzer™ Tool, will increase the chances of proper primer alignment and favorable PCR amplification.

When generating the vector backbone by PCR, we recommend using a high-fidelity polymerase, such as Q5 DNA polymerase, which produces blunt ended DNA for Gibson assembly (Gibson et al., 2009). In a standard Q5 PCR reaction, the following concentrations were used: 0.5 μM of each forward and reverse primer, 0.4 mM dNTPs, 5 ng of parent plasmid, and 0.5 μL of Q5 polymerase (0.02 U/μL) in a 50 μL reaction. For a standard PCR, we used the following program: 95°C for 2.5 min followed by 30 cycles: (1) denature at 95°C for 30 sec, (2) anneal at 55-72°C for 45 sec, (3) extend at 72°C, 30 sec per kb, and a final polishing step at 72°C for 1 min. The PCR amplicon lengths and purity are verified by agarose gel electrophoresis. In cases of PCR failure, we recommend troubleshooting the reaction by considering the primer sequence, length, and orientation; annealing temperature of the PCR reaction; integrity of the PCR reagents; and plasmid concentration.

Materials

Custom DNA oligonucleotides

Nuclease free water

Ice bucket

Q5 DNA Polymerase

Q5 polymerase buffer

Deoxyribose nucleotides

Thermal cycler

Gel Loading Dye, Blue 6x

General Recommendations for DNA Electrophoresis Protocol (Thermo Scientific, Waltham, MA)

Agarose powder

Gel Doc XR+ System

DNA Clean and Concentrator-5

NanoDrop

1. Design primers in the 5′ to 3′ direction on the *in silico* vector map generated in section 2.1. Primer lengths should vary between 15-40 nt with a T_m range of 55-70°C.

   Design the vector forward primer at the 3′ end of the gene. Design the vector reverse primer at the 5′ end of the gene and copy sequence of the complementary strand.
2. Prepare vector forward and reverse primers as 10 µM working stocks for PCR.
3. Setup a negative control PCR reaction (50 µL) on ice with the following final concentrations: 1x Q5 buffer, 0.5 µM vector forward primer, 0.5 µM vector reverse primer, 0.4 mM dNTPs, and 0.5 µL Q5 DNA polymerase (0.02 U/µL).
4. Setup a PCR reaction (50 µL) on ice with the following final concentrations: 1x Q5 buffer, 0.5 µM vector forward primer, 0.5 µM vector reverse primer, 0.4 mM dNTPs, 0.5 µL Q5 DNA polymerase (0.02 U/µL), and 5 ng of parent plasmid from a miniprep.
5. Thermal cycle the vector amplification with the following suggestions: initial denaturation at 95°C for 2.5 min; followed by 30 thermal cycles: (1) denature at 95°C for

30 sec (2) anneal at 68°C for 45 sec, and (3) extend at 72°C for 2.5 min; polishing step at 72°C for 1 min.

PCR reaction optimization should include the anneal temperature, primer concentration, extension times, and plasmid concentration.

6. Mix 2 µL of PCR reaction with 1 µL of Gel Loading Dye, Blue (6x) and 3 µL of nuclease free water. Load mixture into 1% ethidium-bromide agarose gel with the appropriate DNA ladder and run at 120V for 45 min.
7. Visualize agarose gel under UV in Gel Doc XR+ for proper amplicon size.

2C. PCR cleanup and DpnI treatment

Following successful amplification of the vector backbone required for Gibson assembly, the amplicon product should be gel purified to reduce the potential for parental plasmid contamination in subsequent steps. We recommend performing a PCR cleanup prior to DpnI treatment to remove unwanted salts, excess primers, dNTPs, and polymerase. This ensures an optimal buffering system for the DpnI mediated degradation of the methylated or hemi-methylated parent plasmids. We also recommend incubating the PCR products with DpnI at 37°C for 4-5 h, compared to the suggested time of 1h, followed by heat inactivation of the enzyme by incubating at 80°C for 20 min.

Materials

Ice bucket

1.5 mL centrifuge tube

DNA Clean and Concentrator-5

DNA wash buffer (10 mM Tris-HCl, pH 7.4, 80% ethanol)

DpnI

rCutSmart™ (NEB)

1. Transfer PCR reactions into a 1.5 mL centrifuge tube and mix with Zymo DNA binding buffer from DNA Cleanup and Concentrator in a ratio of 300 µL of DNA binding buffer to 100 µL of PCR reaction.
2. Load mixture into Zymo IC column and centrifuge at 14,000 r.c.f for 30 sec. Dispose of flow through and reuse collection tube.
3. Wash Zymo IC column with 2x500 µL DNA wash buffer and centrifuge at 14,000 r.c.f for 30 sec. Dispose of flow through and reuse collection tube between steps.
4. Dry Zymo IC column by centrifugation at 14,000 r.c.f for 3 min.
5. Place Zymo IC column into a clean 1.5 centrifuge tube and elute DNA with 2x20 µL warmed (37°C) nuclease free water. Centrifuge the Zymo IC column at 14,000 r.c.f for 1 min.

   Prewarmed H_2O may help with DNA extraction.
6. Thaw the DpnI reagents (rCutSmart™ buffer and DpnI enzyme) on ice.
7. Transfer the eluted DNA into a 200 µL PCR tube with 5 µL of rCutSmart™ buffer and 5 µL of DpnI enzyme.
8. Incubate at 37°C for 4-5 h in an Eppendorf ThermoMixer or Thermal cycler, followed by an inactivation step at 80°C for 20 min. The product can be held at 24°C following the incubation until agarose gel purification.

2D. Agarose Gel Purification

Agarose gel purification is an essential step for removing parent plasmids and unwanted amplicons from the reaction mixture. We suggest following the recommendations for DNA electrophoresis (Thermo Scientific) to determine the desired gel percentage. Once the agarose gel percentage is determined, a wide lane comb can be generated by taping several wells together, best applied for loading volumes of ≥50 µL. Agarose gel thickness can affect the migration and visualization of the amplicon. For example, when using a Bio-Rad 7x7 cm UVTP tray, we suggest using 50 mL of agarose gel solution for casting and allowing the gel to solidify for 1 h prior to use. We suggest running the agarose gel at a low voltage of 60-70V until the loading dye migrates through ¾ of the length of the agarose gel. It is important to track the migration of the desired oligonucleotide in reference to the gel loading dye migration. Since agarose gel percentage affects the migration of the oligonucleotide ladder and loading dye, one should empirically determine the migration pattern of the loading dye prior to running a purification gel.

When slicing the agarose gel for DNA extraction, we recommend visualizing the band with a blue light (495 nm) which is less damaging to DNA than a UV light. All surfaces, razor blades, and forceps should be sterilized with 70% ethanol prior to cutting the gel. Isolation of the DNA amplicon can be accomplished using a commercial agarose purification kit. We recommend the Zymo Agarose Dissolving Buffer (ADB) and Zymo IC columns for DNA isolation. We recommend thoroughly cleaning the interior and exterior of the Zymo IC column after loading the DNA to remove any excess salt or agarose gel matrix. Elute the purified amplicon with nuclease free water and quantify the DNA concentration by UV absorbance.

Materials

Bio-Rad comb for GT systems; 1.5 mm thick, 5.5 cm long

Agarose powder

Gel Loading Dye, Blue 6x

Agarose Dissolving Buffer

Zymo IC column

DNA wash buffer (10 mM Tris-HCl, pH 7.4, 80% Ethanol)

Eppendorf ThermoMixer

70% Ethanol for sterilization

1. Pour a 2% (w/v) ethidium bromide agarose gel with 2 cm wells and solidify to room temperature.

   Comb size can be modified depending on scale of reactions.

2. Add enough Blue Loading Dye Blue 6x to the amplified DNA for a final concentration of 1x and mix by pipetting.
3. Load the DNA and appropriate DNA ladder into the 2% agarose gel. Run at 70V for 2 h.
4. Visualize the agarose gel under blue light (495 nm) and excise the appropriate band.

   Ensure the cutting tools are sterile to avoid contamination.

5. Transfer the excised gel piece into a 1.5 mL low binding centrifuge tube.
6. Add 300 µL of Agarose Dissolving Buffer (ADB) for every 100 mg of agarose gel piece, up to 400 mg.
7. Incubate at 37°C, 400 rpm, for at least 1 h in an Eppendorf ThermoMixer.
8. Once fully dissolved, load aqueous solution into a Zymo IC DNA column and centrifuge at 14,000 r.c.f. for 30 sec.

9. Add 500 µL DNA wash buffer into the column. Mix by inverting the column 10 times and aspirate the buffer into waste. Repeat once more.
10. Rinse the exterior of the column with 4x1 mL of fresh DNA wash buffer. Place column into a clean 1.5 mL low binding centrifuge tube.
11. Wash the column with 4x200 µL of DNA wash buffer and centrifuge at 14,000 r.c.f. for 30 sec. Discard the flow through after each spin and reuse the collection tube.
12. Remove excess ethanol by drying the column, centrifuging at 14,000 r.c.f. for 3 min.
13. Elute the DNA with 2x10 µL of pre-warmed (37°C) nuclease-free H_2O into a 1.5 mL centrifuge tube, for a total of 20 µL. Let stand for 1 minute then centrifuge at 14,000 r.c.f. for 1 min.

2E. Gibson Assembly of the parent genes and linear vector and transformation

Gibson assembly is an advanced cloning technique commonly used for multi-fragment gene assembly in a single isothermal reaction[12]. In this technique, the gene, prepared as a gBlock, is inserted into a linearized protein expression vector in a reaction that includes a 5′ exonuclease, a DNA polymerase, and a DNA ligase. The exonuclease removes nucleotides on the 5′ ends of the dsDNA and linear vector, exposing 3′ overhangs that allow for hybridization of the insert and vector. The DNA polymerase fills in the gaps with DNA and the ligase mediates phosphodiester bond formation sealing the nicks between the insert and vector. The newly synthesized constructs are transformed into a cloning strain, such as DH5α competent *E. coli*, for repairs and propagation. We recommend sequencing individual clones to ensure property assembly of the gene. If mutations are present, it is best to correct these positions by site-directed mutagenesis rather than ordering a new gBlock.

We recommend the following steps for successful Gibson assembly. First, we recommend resuspending the gBlock in nuclease free water at 37°C for at least 10 min. Importantly, users should resist the temptation to vortex the sample, which can lead to DNA damage. Second, our preferred reaction conditions for Gibson assembly are: 100 ng of vector, 1 µL of gBlock, and 10 µL of 2x Gibson assembly master mix in a final 20 µL reaction. In other examples, such as gene amplification following selection, we recommend using a ratio of 100 ng of vector backbone to 100 ng of insert. Third, we prefer a reaction time of 60 min at 50°C from the suggested 15 to 60 min incubation[12]. Fourth, we recommend column purifying the Gibson product, then transforming 1 µL of the product into 50 µL aliquot of DH5α competent *E. coli* with the addition of BME (1 µL, 28 µM) prior to the heat shock step to increase the transformation efficiency by degrading carbohydrates on the cell surface[15].

Materials

gBlock

vector backbone

Ice bucket

2x Gibson Assembly Master Mix

Thermal cycler

DH5α competent E. *coli* cells

BME, b-mercaptoethanol

Super Optimal broth with Catabolite repression (SOC)

New Brunswick Innova 44 Incubator

50 µg/mL carbenicillin Luria-Bertani (LB) agar plates

Nuclease free water

10x ThermoPol

100 µM gene flanking primers (fragment 1 forward and 11 reverse)

Taq DNA polymerase

Deoxyribose nucleotides

Gel Loading Dye, Blue 6x

Agarose powder

10 mg/mL Ethidium Bromide (EtBr)

Gel Doc XR+ System

1x Phosphate Saline Buffer (PBS) and 50% glycerol

50 µg/mL carbenicillin LB media

Eppendorf Centrifuge 5804

Plasmid Miniprep Kit

1. Suspend gBlocks in nuclease free water according to the manufacturer recommended guidelines.

   Do not vortex the DNA. Place at 37°C to aid reconstitution.

2. Mix 100 ng of purified vector with 1 µL of gBlock and add nuclease free water to a final volume of 10 µL in a PCR tube.
3. Add 10 µL of 2x Gibson master mix and mix by pipette for a total of 20 µL.
4. Place PCR tube into thermal cycler at 50°C for 1 h.
5. Add 300 µL of Zymo DNA binding buffer to PCR tube and transfer mixture into Zymo IC column.

6. Spin DNA through Zymo IC column at 14,000 r.c.f for 30 sec.
7. Wash Zymo IC column with 2x500 µL DNA wash buffer and centrifuge at 14,000 r.c.f for 30 sec.
8. Dry Zymo IC column by centrifugation at 14,000 r.c.f for 3 min.
9. Place Zymo IC column into a clean 1.5 centrifuge tube and elute DNA with 2x10 µL warmed (37°C) nuclease free water. Centrifuge the Zymo IC column at 14,000 r.c.f for 1 min.
10. Remove a 50 µL aliquot of DH5α competent cells and place on ice for 5 min to thaw cells.
11. Add 1 µL of purified Gibson assembly product to the competent cells. Mix by flicking the tube 10 times and incubate on ice for 30 min.
12. Supplement the transformation with 1 µL of 1.4 mM BME for a final concentration of 28 µM. Mix and immediately place into 42°C water bath for 30 sec.

    BME has been shown to increase transformation efficiency (Janjua et al., 2014).
13. Place cells on ice for 2 minutes immediately following heat shock.
14. Add 1 mL of warmed SOC to cells and place into New Brunswick Innova 44 at 37°C, 225 rpm for 1 h.
15. Plate 250 µL of recovered cells onto 50 µg/mL carbenicillin agar plates and place into 37°C standing incubator overnight.
16. For colony PCR, include a negative control PCR reaction (10 µL) with the following final concentration: 1x ThermoPol buffer, 0.5 µM of each gene flanking primer (fragment 1 forward primer and fragment 11 reverse primer), 0.4 mM dNTPs, and 0.2 µL Taq DNA polymerase (100 U/mL).

17. Setup colony PCR reaction (20x10 µL) with the following final concentration: 1x ThermoPol buffer, 0.5 µM of gene flanking primer set (fragment 1 forward and 11 reverse) , 0.4 mM dNTPs, 0.2 µL Taq DNA polymerase (100 U/mL). Pick colony directly off agar plates.
18. For each individual reaction tube, swirl the single colony in the reaction mixture, remove 1 µL of PCR reaction and mix into a separate PCR tube containing 25 µL of 1x PBS, 50% glycerol (glycerol stocks).
19. Thermal cycle the colony PCR with the following suggestions: initial denaturation 95°C for 2 min; followed by 30 thermal cycles: (1) denature at 95°C for 15 sec, (2) anneal at 58°C for 15 sec, and (3) extend at 72°C for 2.5 min; polishing step at 72°C for 1 min.
20. Mix 9 µL of PCR reaction with 2 µL of Gel Loading Dye, Blue (6x) and load mixture into 1% ethidium-bromide agarose gel with appropriate DNA ladder and run at 120V for 45 min.
21. Visualize agarose gel for proper amplicon size under UV with Gel Doc XR+.
22. Transfer the glycerol stocks that amplified (26 µL containing 1 µL of PCR reaction with 25 µL of 1x PBS, 50% glycerol) into 4 mL of 50 µg/mL carbenicillin LB media. Place media into New Brunswick Innova 44 at 37°C, at 225 rpm overnight.
23. Harvest the cells after 16-22 h of growth by centrifugation at 4000 rpm, 5 min in Eppendorf Centrifuge 5804.
24. Follow manufacturer instructions for Plasmid Miniprep II to prepare plasmids for sequencing.
25. Sequence the plasmid to verify gene assembly. If mutations are observed during sequencing, perform site-directed mutagenesis to repair genes before moving forward.

2F. Fragment Generation and Homologous Recombination

In this section, we discuss the generation of a homologous recombination library using a 3-step process (Fig. 2.3). First, the parent gene is fragmented into individual pieces by PCR using forward and reverse primers specific to each fragment. Second, the gene fragments are randomly reassembled by overlap PCR. Last, PCR is performed to amplify the full-length material from step 2.

As mentioned in section 2.3a, recombination sites were identified as short segments that are shared between the parent homologs. These sites flank the individual gene fragments and serve as primer binding sites for PCR amplification of the gene fragments. Refer to the suggestions of primer design in section 2.2. In our study, the parent genes contained 11 fragmentation pieces, which required 44 individual PCR reactions to generate the 44 gene fragments as individual amplicons (Fig. 2.3A). PCR reactions contained final concentrations: 0.5 µM of each primer set, 0.4 mM dNTPs, 5 ng of parent plasmid, and 0.5 µL of Q5 polymerase (0.02 U/µL) in a 50 µL reaction. For a standard PCR cycling, we recommend the following program: denature at 95°C for 2.5 min followed by 30 cycles: (1) denature at 95°C for 30 sec, (2) anneal at 55-72°C for 45 sec, (3) extend at 72°C for 30 sec per kilobase (kb) and a final polishing step at 72° C for 1 min. Amplicon length and purity are analyzed by agarose gel electrophoresis. Upon verification, a PCR cleanup was performed, and the amplicons were eluted in nuclease free water and UV quantified.

The second step is a fragment reassembly process that uses overlap PCR to randomly generate all possible sequence combinations defined by the fragmentation library (Fig. 2.3A). We recommend using 100 ng of total DNA per every 1000 bases. Since the length of our gene is 2,631 bp, we used a total DNA of 263.1 ng for the overlap PCR reaction. We made 10 ng/µL working stocks by combining 25 ng of each homologous parent fragment 1 for a total of 100 ng in a final volume of 10 µL. This process was repeated for fragments 2-11 to obtain 11 individual working

stocks. As an example, fragment 1 (179 bp) composes 6.8% of out of the total gene size (2,631 bp). Therefore, 6.8% of the 263.1 ng of DNA required for the overlap PCR is 17.9 ng. From a 10 ng/μL working stock, we added 1.79 μL of the fragment 1 in addition to the required amount for subsequent fragments 2 to 11, along with 1x Q5 buffer, 0.4 mM dNTPs, and 0.5 μL of Q5 polymerase (0.02 U/μL) in a 50 μL reaction. For the overlap PCR, the recommended cycling is as follows: initial denaturation 94°C for 30 sec, followed by 9 thermal cycles of (1) denature at 94°C for 15 sec, (2) extend at 72°C for 30 sec per kb min with a decrease of 0.5°C per cycle. Then, an additional 5 thermal cycles of (3) denature at 94°C for15 sec, (4) anneal at 67.5° for 15 sec with a decrease of 0.5°C per cycle, (5) extend at 72°C for 30 sec per kb; with a polishing step of 72°C for 2 min.

The final step is amplification of the shuffled gene library (Fig. 2.3B). The overlap PCR reaction (4 μL) from step 2, is used for a PCR with 0.5 μM of the gene flanking primers (fragment 1 forward and fragment 11 reverse), 0.4 mM dNTPs, and 0.5 μL of Q5 polymerase (0.02 U/μL) in a 50 μL reaction. We recommend 20-40 cycles of PCR with the following conditions: initial denaturation 94°C for 30 sec; followed by 17 thermal cycles (1) denature at 94°C for 15 sec, (2) anneal at 67°C for 15 sec with a decrease of 0.5°C per cycle, (3) extend at 72°C for 30 sec per kb. Additional 23 thermal cycles of (4) denature at 94°C for 15 sec, (5) anneal at 59°C for 15 sec, and (6) extend at 72°C for 30 sec per kb; with a polishing step of 72°C, 2 min. It is essential to verify assembly of the full gene by agarose gel electrophoresis. Successful PCR amplicons should undergo DNA cleanup and Dpn1 treatment (section 2C), followed by agarose gel purification (section 2D) in preparation for Gibson assembly (section 2E and 2G).

Materials

Custom DNA oligonucleotides

Ice Bucket

Q5 polymerase buffer

Q5 DNA polymerase

Deoxyribose nucleotides (dNTPs)

Thermal cycler

Deoxyribose nucleotides

Gel Loading Dye, Blue 6x

Agarose powder

Gel Doc XR+ System

1. Design custom primers based on recombination sites between all polymerases in section 2A.

   Primers should range in length from 15 to 25 base pairs for optimal PCR.

2. Order primers. Prepare as 100 µM stocks with nuclease free water and dilute to 10 µM working stock.
3. Setup a negative control PCR reaction (50 µL) with the following final concentrations: 1x Q5 buffer, 0.5 µM fragment 1 forward primer, 0.5 µM fragment 11 reverse primer, 0.4 mM dNTPs, and 0.5 µL Q5 DNA polymerase.
4. Setup individual PCR reactions (50 µL) with the following final concentrations: 1x Q5 buffer, 0.5 µM of each fragment primer set (fragment 1 forward and fragment 1 reverse primers), 0.4 mM dNTPs, 0.5 µL Q5 DNA polymerase, and 5 ng of parent plasmid.

Repeat for each fragment from each of the 4 parents using the designated primer sets (fragment 2 forward and reverse, etc.) for a total of 44 individual PCR reactions.

5. Thermal cycle the fragment amplification with the following suggestions: initial denaturation 95°C for 2 min; 30 thermal cycles: (1) denature at 95°C for 30 sec, (2) anneal at 68.6°C for 30 sec, and (3) extend at 72°C for 2.5 min; polishing step at 72°C for 1 min.
6. Mix 2 µL of PCR reaction, 1 µL Gel Loading Dye, Blue (6x), and 3 µL nuclease free water. Load samples into a 2% ethidium-bromide agarose gel, with appropriate DNA ladder, and run at 120V for 45 min. Visualize agarose gel under UV and image.
7. Each PCR reaction was mixed with 300 µL of Zymo DNA binding buffer and transfer to Zymo IC column. Centrifuge columns at 14,000 r.c.f for 30 sec and discard flow through.
8. Wash Zymo IC column with 2x500 µL Zymo DNA wash buffer, centrifuge at 14,000 r.c.f for 30 sec and discard flow through.
9. Dry Zymo IC column by centrifugation at 14,000 r.c.f for 3 min.
10. Place Zymo IC column into a clean 1.5 centrifuge tube and elute DNA with 2x25 µL warmed (37°C) nuclease free water. Centrifuge the Zymo IC column at 14,000 r.c.f for 1 min.
11. Quantify the DNA concentration by UV absorbance.
12. Mix 25 ng of each individual fragment cassette (Kod, Tgo, DV, 9°N) for fragment 1 for a total of 100 ng in 10 µL (10 ng/ µL stock concentration) for assembly. Repeat process for all subsequent fragments 2-11.
13. Use mass ratio between fragment sizes and the entire gene to determine the amount of DNA to add for the overlap PCR.

14. We suggest a concentration of 100 ng of DNA per 1 kilobase. For this study, the gene size is 2,631 base pair, therefore a total of 263.1 ng was used for the overlap PCR.

    Example: Fragment 1 cassette is 179 base pairs out of 2,631 total gene, which is 6.8% of the gene. Therefore, 17.9 ng of fragment 1 is required. Add 1.79 µL of DNA from the 10 ng/µL stock concentration.

15. Setup a 50 µL PCR reaction with the following final concentrations: 1x Q5 buffer, 0.4 mM dNTPs, 0.5 µL Q5 DNA polymerase, and a total of 263.1 ng of DNA.
16. For overlap PCR, thermal cycle with the following conditions: initial denaturation 94°C, 30 sec, followed by 9 thermal cycles of (1) denature at 94°C for 15 sec, (2) anneal at 72°C for 1.5 min with a decrease of 0.5°C per cycle. Additional 5 thermal cycles of (3) denature at 94°C for 15 sec, (4) anneal at 67.5°C for 15 sec with a decrease of 0.5°C per cycle, and (5) extend at 72°C for 1.5 min; with a polishing step of 72°C for 2 min.
17. Setup a negative PCR reaction (50 µL) with the following final concentrations: 1x Q5 buffer, 0.4 mM dNTPs, 0.5 µL of Q5 DNA polymerase, nuclease free water in the presence of 0.5 µM fragment 1 forward, and 0.5 µM fragment 11 reverse gene flanking primers.
18. Setup a PCR reaction (50 µL) with the following final concentrations: 1x Q5 buffer, 0.4 mM dNTPs, 0.5 µL of Q5 DNA polymerase, nuclease free water in the presence of 0.5 µM gene flanking fragment 1 forward and fragment 11 reverse primers, and 4 µL of PCR product from step 16.
19. Perform amplification of PCR product with the following conditions: initial denaturation 94°C for 30 sec; followed by 17 thermal cycles of (1) denature at 94°C for 15 sec, (2) anneal at 67°C for 15 sec with a decrease of 0.5°C per cycle, and (3) extend at 72°C for

1.5 min. Additional 23 thermal cycles of (4) denature at 94°C for 15 sec, (5) anneal at 59°C for 15 sec, and (6) extend at 72°C for 1.5 min; with a polishing step of 72°C for 2 min.

20. Mix 2 µL of PCR reaction with 1 µL of Gel Loading Dye, Blue (6x) and 3 µL of nuclease free water and load into 1% ethidium-bromide agarose gel, with appropriate DNA ladder, and run at 120V for 45 min.
21. Refer to section 2C for PCR cleanup and DpnI treatment of the PCR amplicon
22. Refer to section 2D for Agarose gel purification

2G. Gibson assembly of the recombination library and linear vector

Materials

Refer to materials in section 2E

1. Mix 100 ng of vector with 100 ng of gel purified gene and add nuclease free water to a final volume of 10 µL in a PCR tube.
2. Refer to section 2E and follow steps 2 to 25.

2H. Plasmid scale-up and purification

Plasmid scale-up is necessary to ensure that sufficient coverage of the library is available following Gibson assembly. We recommend scaling the library through five independent transformations with DH5α competent cells. The cells are transformed with manufacturer recommendations, plated, and incubated over night at 37°C. Cells are scrapped off the agar plates and midi prepped to recover the plasmid.

Material

Ice bucket

DH5α competent cells

Beta-mercaptoethanol (BME)

SOC media

New Brunswick Innova 44 Incubator

50 µg/mL carbenicillin LB agar plates

50 µg/mL carbenicillin LB media

L-shape Cell Spreader

50 mL conical tube

ZymoPURE Express Plasmid Midiprep Kit

1. Remove 4x50 µL aliquot of DH5α competent cells and place on ice for 5 min to thaw cells.
2. Add 1 µL of purified Gibson assembly product to the competent cells. Mix by flicking the tube 10 times and incubate on ice for 30 min.
3. Supplement the transformation with 1 µL of 1.4 mM BME for a final concentration of 28 µM. Mix and immediately place into 42°C water bath for 30 sec.
4. Place cells on ice for 2 min immediately following heat shock.
5. Add 1 mL of warmed SOC to cells and place into New Brunswick Innova 44 at 37°C, 225 rpm for 1 h.
6. Plate 250 µL of recovered cells onto 50 µg/mL carbenicillin LB agar plates and place into 37°C standing incubator overnight.

7. Scrape the colonies by adding 1 mL of 50 μg/mL carbenicillin LB media to the agar plate. Use an L-shape cell spreader to gently scrape *E. coli* cells off.
8. Pull all cells into a 50 mL conical tube and add 50 μg/mL carbenicillin LB media to a final volume of 25 mL.
9. Perform ZymoPURE Express Plasmid Midiprep Kit following the manufacturer's recommended instructions.
10. Extract DNA from column with 2x300 μL warmed (37°C) nuclease free water and quantify the DNA concentration by UV absorbance.

2.4 Cell growth, sample preparation for Droplet-based Optical Polymerase Sorting (DrOPS) and plasmid recovery

In this section, we describe the process of expressing the polymerase library in *E. coli*, and recovery of the DNA from the droplet following droplet-base optical polymerase sorting (DrOPS) (Fig. 2.4 & 2.5). Two independent microfluidic devices are utilized to generate and sort water/oil emulsions. A complete description of the microfluidic devices and instrumentation setup was described in a previous chapter[11]. The cell line (XL1-Blue Supercompetent Cells) used for our polymerase expression is a nuclease deficient *E. coli* strain that prevents the degradation of our fluorescently labeled DNA hairpin. We utilize the lac operon repressor system to attenuate expression level. A key step to ensure successful generation of microdroplets is buffer exchanging and cleaning of *E. coli* cells into the reaction buffer. Once the cells are clean and quantified, an activity reaction mixture is made with the following components: 1x ThermoPol, 1 μM Cy3-fluorescently labeled DNA hairpin, 2 μM Iowa Black complementary sequence, and 100 μM of tNTPs. Excess amount of Iowa Black complementary oligonucleotide is added to ensure complete

quenching. Post emulsification, droplets are subjected to a heat lysis at 95°C for 5 min and incubated at 55°C for various length of time depending on the desired selection stringency. A microfluidic device is utilized to sort these single layer emulsions on a custom fluorescently activated droplet sorter (FADS). A key step in recovering droplets efficiently is the washing of the collection line with excess amount of buffer used to generate the emulsions. Emulsions are disrupted by the addition of fluorocarbon oils containing surfactants and a vortex. A DNA cleanup step is performed to remove excess fluorocarbon oils prior to amplification of the gene.

Materials

Ice bucket

XL1-Blue Supercompetent Cells

SOC media

50 μg/mL carbenicillin LB media

New Brunswick Innova 44 Incubator

Isopropyl-1-thio-β-D-galactopyranoside (IPTG)

14 mL round-bottom Falcon tube

Eppendorf Centrifuge 5804

10x ThermoPol

1.5 mL centrifuge tube

Eppendorf 5415D centrifuge in 4°C deli

5′-Cy3 labeled self-priming hairpin template

3′-Iowa Black FQ quencher

Threose nucleic acid (TNA)

Custom DNA oligonucleotides

Q5 polymerase buffer

Q5 DNA polymerase

Deoxyribose nucleotides

Thermal cycler

Gel Loading Dye, Blue 6x

Agarose gel

Gel Doc XR+ System

1. Mix 300 ng of library plasmid with 20 µL of XL1-Blue Supercompetent Cells and place on ice for 30 min.
2. Heat shock cells at 42°C for 30 sec and immediately place on ice for 2 min.
3. Add 1 mL of warmed SOC to cells and place into New Brunswick Innova 44 at 37°C, 225 rpm for 1 h.
4. Transfer the recovered cells into 500 mL baffled flask containing 50 mL of 50 µg/mL carbenicillin LB media and place into New Brunswick Innova 44 at 37°C, 225 rpm, overnight.
5. Transfer 500 µL of overnight into a fresh 50 mL of 50 µg/mL carbenicillin LB media and place into New Brunswick Innova 44 at 37°C, 225 rpm, until OD_{600} reaches ~0.600 au.
6. Induce protein expression by adding IPTG to a final concentration of 1 mM. Place the cells back into the New Brunswick Innova 44 at 25°C, 225 rpm, 20 h.
7. Transfer 1 mL of cells into a 14 mL round-bottom falcon tube and centrifuge cells at 1,811 r.c.f for 5 min in Eppendorf centrifuge.

8. Discard supernatant, wash cells with 1 mL of 1x ThermoPol, and centrifuge cells at 1,811 r.c.f for 5 min in Eppendorf centrifuge. Repeat four times.
9. After the fourth wash, suspend cells with 2 mL of 1x ThermoPol and measure OD_{600}.
10. Transfer the appropriate volume of the suspended cell to a 1.5 mL centrifuge tube and centrifuge at 1,811 r.c.f for 5 min in Eppendorf 5415D centrifuge.
11. Carefully discard supernatant as to not disturb the pellet.
12. Prepare 250 µL reaction mixture for the fluorescence-based polymerase activity assay containing the following final concentrations: 1x ThermoPol buffer, 1 µM of a 5′-Cy3 labeled self-priming hairpin template, 2 µM of a 3′-Iowa Black FQ quencher complementary to the 5′ end of the template, and 100 µM of tNTPs.
13. Follow (Vallejo et al., 2020) for microfluidic device fabrication and instrumentation for droplet generation.
14. Following droplet generation, droplets are incubated at 95°C for 5 min, cooled to 24°C for 5 min, and incubated at 55°C for variable time depending on desired selection stringency (18 - 2 h).
15. Follow (Vallejo et al., 2020) for microfluidic device fabrication, droplet sorting, and recovery of DNA.
16. Setup a negative control PCR reaction (50 µL) with the following final concentrations: 1x Q5 buffer, 0.5 µM fragment 1 forward primer, 0.5 µM fragment 11 reverse primer, 0.4 mM dNTPs, and 0.5 µL Q5 DNA polymerase.
17. Setup PCR reaction (50 µL) for each extraction with the following final concentrations: 1x Q5 buffer, 0.5 µM fragment 1 forward primer, 0.5 µM fragment 11 reverse primer, 0.4 mM dNTPs, 0.5 µL Q5 DNA polymerase, and 4 µL of the purified recovered DNA.

18. Thermal cycle the gene amplification with the following suggestions: initial denaturation 95°C for 2.5 min; followed by 30 thermal cycles: (1) denature at 95°C for 30 sec, (2) anneal at 70°C for 30 sec, and (3) extend at 72°C for 2.5 min; polishing step at 72°C for 1 min.
19. Mix 2 μL of PCR reaction with 1 μL of Gel Loading Dye, Blue (6x) and 3 μL of nuclease free water. Load mixture into 1% ethidium-bromide agarose gel, with the appropriate DNA ladder, and run at 120V for 45 min.
20. Visualize agarose gel under UV in Gel Doc XR+ for proper amplicon size.
21. Refer to section 2C for PCR cleanup and DpnI treatment
22. Refer to section 2D for agarose gel purification
23. Refer to section 2E and 2G for Gibson assembly

2.5 Colony picking and activity screen

In this section, we discuss a simple approach for screening enriched variants for polymerase activity. Library members are transformed into *E. coli* and plated onto an antibiotic containing agar plate. Cells are diluted in a ratio of 1:20 before plating to allow for isolated colony to grow overnight at 37°C. Single colonies are picked and placed in 4 mL of LB media with antibiotics for overnight growth at 37°C with shaking at 225 rpm. A small amount of overnight culture (40 μL) is inoculated into 4 mL of fresh media, grown to the appropriate cell density, and induced by the addition of Isopropyl β-d-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. Following induction, cells are incubated overnight at 25°C with shaking at 225 rpm, although temperatures may vary depending on the expression vector. The following day, cells are harvested and suspended in polymerase lysis buffer. The suspended cells are subjected to a heat

lysis at 70°C for 1 h, followed by cooling on ice for 1 h, and centrifuge to pellet cell debris and denatured proteins. The clarified lysate is estimated with 1x Bradford dye before the primer extension assay in which polymerases are challenged to extend tNTPs on a labeled DNA primer-template duplex. The reactions are quenched with stop buffer and run on a denaturing polyacrylamide gel. Gels are imaged, and analysis is performed on the banding pattern of the variant polymerase as compared to the parent polymerase.

Materials

Ice bucket

XL1-Blue Supercompetent Cells

SOC media

50 μg/mL carbenicillin LB agar plates

50 μg/mL carbenicillin LB media

New Brunswick Innova 44 Incubator

Isopropyl-1-thio-β-D-galactopyranoside (IPTG)

14 mL round-bottom Falcon tube

Eppendorf Centrifuge 5804

Polymerase lysis buffer (10 mM Tris-HCl, pH 8, 500 mM NaCl)

1.5 mL centrifuge tube

Eppendorf 5415D centrifuge in 4°C deli

200 μL PCR tubes

10x ThermoPol

5′-IR800-DNA primer

DNA template

Threose nucleic acid (TNA)

Quenching buffer (25 mM EDTA, 95% formamide)

SequaGel UreaGel 29:1 Denaturing Gel System

Oddessy Li-Cor Imager

1x Bradford Dye

1. Transfer 300 ng of plasmid purified from Gibson assembly of directed evolution output and transform with 20 µL XL1 Blue cells and incubate for 30 min on ice.
2. Heat shock cells at 42°C for 30 sec and immediately place back on ice for 2 min.
3. Add 1 mL of warmed SOC to cells and place into New Brunswick Innova 44 at 37°C, 225 rpm for 1 h.
4. Plate 250 µL of recovered cells onto 50 µg/mL carbenicillin LB agar plates and place into 37°C standing incubator overnight.
5. Transfer 4 mL of 50 µg/mL carbenicillin LB media into a 14 mL round bottom Falcon tube. Pick single colonies for each tube (100-500 in batches of 20) and place into New Brunswick Innova 44 at 37°C, 225 rpm, overnight.
6. Transfer 40 µL of overnight into 4 mL of fresh 50 µg/mL carbenicillin LB media in a 14 mL round bottom Falcon tube and place into New Brunswick Innova 44 at 37°C, 225 rpm, until OD_{600} reaches ~0.600 au.
7. Induce cells with a final concentration of 1 mM IPTG and place into New Brunswick Innova 44 at 25°C, 225 rpm, for 20 h.

8. Centrifuge cells at 4000 rpm for 5 min in Eppendorf Centrifuge 5804 and discard media supernatant.
9. Suspend cells with 100 µL of polymerase buffer (10 mM Tris-HCl, pH 8, 500 mM NaCl, 10% glycerol) and transfer into a 1.5 mL centrifuge tube.
10. Incubate the 1.5 mL centrifuge tube at 70°C for 1 h.
11. Place the 1.5 mL centrifuge tube on ice for 1 h.
12. Centrifuge the 1.5 mL centrifuge tube at 4°C, 14,000 r.c.f for 1 h to pellet cell debris.
13. Mix 4 µL of the supernatant with 200 µL of Bradford dye to verify the presence of protein.
14. Prepare the activity assay mixture (20 µL) with the following final concentrations: 1x ThermoPol, 1 µM 5′-IR800-label DNA primer, 1 µM DNA template in the absence of 100 µM tNTPs, and 2 µL clarified lysate.
15. Incubate the mixture at 90°C for 5 min and immediately place in ice for 5 min.
16. To the reaction mixture, add 100 µM tNTPs and 2 µL clarified lysate on ice.
17. Incubate the reaction at 55°C at various time points.
18. Remove 1 µL of the reaction mixture and add with 39 µL of quenching buffer (25 mM EDTA and 95% formamide).
19. Incubate quenched reaction at 95°C for 10 min to denature DNA and TNA heteroduplex.
20. Prepare denaturing PAGE under manufacturer's instruction for SequaGel UreaGel 29:1 Denaturing Gel System.
21. Load 10 µL of denatured sample into the denaturing PAGE and run at 10 Watts for 1.5 h.
22. Image gel on Li-Cor Odyssey to observe extension of tNTPs onto the DNA primer.

2.6 Mutagenic PCR

Mutagenic PCR is a molecular biology technique that is commonly used to introduce random mutations into a gene. In a typical mutagenic PCR reaction, a low fidelity polymerase, such as Taq, is used. However, Taq has been shown to incorporate nucleotides resulting in an overhanging at the 3' termini. This is problematic for Gibson assembly, as the overhang will cause an undesired frame shift. In lieu of multiple rounds of PCR between a low fidelity and high-fidelity polymerase, we recommend utilizing a recombinantly expressed high-fidelity polymerase with the exonuclease activity silenced. The polymerase concentration will need to be empirically determined. We recommend biasing the concentration of triphosphates in favor of pyrimidines to reduce G·T wobble mismatches and G →A transitions[16]. It is helpful to perform a concentration gradient of manganese to determine acceptable amounts to favor the generation of proper size amplicons and prevent uncontrolled amplification.

Materials

10x ThermoPol

Manganese Chloride • 6 H_2O ($MnCl_2$)

Custom oligonucleotide primers

1.5 mL centrifuge tube

200 µM PCR tubes

Deoxyribose nucleotides (dNTPs)

1. Prepare a fresh 50 mM stock of $MnCl_2$ and dilute to 1 mM.

2. Setup a negative control PCR reaction (50 µL) without plasmid, with the following final concentrations: 1x ThermoPol buffer, 1 µM each forward and reverse primers, 1 mM TTP, 1 mM dCTP, 0.2 mM dGTP, 0.2 mM dATP, 5.5 mM $MgCl_2$, 0.013 µM Kod-WT (exo -) polymerase.
3. Setup PCR reactions (4x50 µL) with the following final concentrations: 1x ThermoPol buffer, 1 µM each fragment 1 forward and fragment 11 reverse primer, 1 mM TTP, 1 mM dCTP, 0.2 mM dGTP, 0.2 mM dATP, 5.5 mM $MgCl_2$, 0.013 µM Kod-WT (exo -) polymerase, and 100 ng plasmid. Supplement each reaction with either 250, 125, 62.5, 31.25 or 0 µM $MnCl_2$.
4. Thermal cycle the PCR according to the following suggestions: initial denaturation 95°C for 2 min; followed by 30 cycles thermal cycling of (1) denature at 95°C for 1 min, (2) anneal at 60°C for 1 min, and (3) extend at 68°C for 2.5 min; hold at 4°C.

   To determine the optimal annealing temperature, a gradient PCR can be performed. Further optimizations can also be performed on extension time and temperature.
5. Analyze 2 µL aliquots of PCR reactions on a 1% (w/v) ethidium bromide agarose gel with the appropriate DNA ladder to validate amplicon size. Gel band intensities should be robust and clean with successful PCR.

   Troubleshooting parameters include optimization of primer length or concentration, polymerase concentration, $MnCl_2$ concentration, or extension time.
6. Refer to section 2C for PCR clean up and DpnI treatment for Gibson assembly.
7. Refer to section 2D for Agarose Gel Purification.
8. Refer to section 2E and 2G for Gibson assembly.
9. Refer to section 2H for Plasmid scale-up and purification.

10. Refer to section 2.4 for Cell growth, sample preparation for Droplet-based Optical Polymerase Sorting (DrOPS) and plasmid recovery.
11. Refer to section 2C for PCR clean up and DpnI treatment.
12. Refer to section 2D for Agarose Gel Purification.
13. Refer to section 2E and 2G for Gibson assembly.
14. Refer to section 2H for Plasmid scale-up and purification.
15. Refer to section 2.5 for Colony picking and activity screen.

2.7 Notes

- Problems with PCR amplification of the library can usually be fixed by gradient PCR, primer design, or cycle optimization.
- Utilize Thermo Scientific general recommendations for DNA electrophoresis to determine the appropriate agarose gel percentage for the expected amplicon size.
- We highly recommend DpnI treatment and agarose gel purification of the PCR amplified expression vector before Gibson assembly.
- Increased time and enzyme concentration may improve the efficiency of DpnI treatment.
- For mutagenic PCR, one may need to empirically determine the amount of polymerase and $MnCl_2$ to optimize the PCR.

2.8. Summary and conclusions

This chapter describes a directed evolution strategy used to engineer archaeal B-family DNA polymerases to recognize TNA, an unnatural nucleic acid analogue (Schöning et al., 2000). We provide a special emphasis on library design and generation by homologous recombination

using synthetic gene fragments that are randomly reassembled at engineered recombination sites. We first describe how parent scaffolds are identified, acquired, and cloned into a preferred expression vector by Gibson assembly. Next, we describe the 3-step process of homologous recombination beginning with fragment generation, where the encoding DNA for each parent scaffold is fragmented into cassettes by PCR using primers that are specific for each fragment. Second, we describe how the cassettes are randomly assembled by overlap PCR, and third, a final PCR is performed to amplify full-length material from step 2. We provide a detailed methodology for expression of the polymerase library in *E. coli* and lightly describe the preparation of single cell encapsulation into uniform water-in-oil droplets, and recovery of the DNA from the droplet following droplet-based optical polymerase sorting (DrOPS). The genes from the active library are PCR amplified, purified, and ligated to the expression vector for colony picking and activity screening. The activity screen reveals variants with desired functional activities, whose genetic information can inform the basis of the next round of directed evolution. To further diversify the sequence landscape, a protocol for mutagenic PCR is also included.

This contribution provides a systematic protocol to generate a library of mutant enzymes with various fitness levels across a diverse sequence landscape. This procedure allows users to screen extremely diverse libraries without relying on NGS sequencing. These methodologies are general and can be of value to other enzyme engineering studies on DNA modifying enzymes.

2.9 Acknowledgments

We wish to thank members of the Chaput lab for their helpful comments on the manuscript. This work was supported by grants from the National Science Foundation (CHM: 2001434 and MCB: 1946312).

2.10 References

1. Zeymer, C. & Hilvert, D. Directed evolution of protein catalysts. *Annu. Rev. Biochem.* **87**, 131–157 (2018).

2. Arnold, F. H. & Georgiou, G. *Directed Evolution Library Creation: Methods and Protocols* (Humana Press, 2003).

3. Stemmer, W. P. C. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389–391 (1994).

4. Ness, J. E. *et al.* Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. *Nat. Biotechnol.* **20**, 1251–1255 (2002).

5. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16**, 379–394 (2015).

6. Levin, A. M. & Weiss, G. A. Optimizing the affinity and specificity of proteins with molecular display. *Mol. BioSyst.* **2**, 49–57 (2006).

7. Ghadessy, F. J., Ong, J. L. & Holliger, P. Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl Acad. Sci. USA* **98**, 4552–4557 (2001).

8. Pinheiro, V. B. *et al.* Synthetic genetic polymers capable of heredity and evolution. *Science* **336**, 341–344 (2012).

9. Larsen, A. C., Dunn, M. R., Hatch, A., Sau, S. P., Youngbull, C. & Chaput, J. C. A general strategy for expanding polymerase function by droplet microfluidics. *Nat. Commun.* **7**, 11235 (2016).

10. Nikoomanzar, A., Chim, N., Yik, E. J. & Chaput, J. C. Engineering polymerases for applications in synthetic biology. *Q. Rev. Biophys.* **53**, e8 (2020).

11. Vallejo, D., Nikoomanzar, A. & Chaput, J. C. Directed evolution of custom polymerases using droplet microfluidics. *Methods Enzymol.* **644**, 227–253 (2020).

12. Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A. III & Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

13. Schöning, K. U., Scholz, P., Guntha, S., Wu, X., Krishnamurthy, R. & Eschenmoser, A. Chemical etiology of nucleic acid structure: the α-threofuranosyl-(3′→2′) oligonucleotide system. *Science* **290**, 1347–1351 (2000).

14. Nikoomanzar, A., Vallejo, D., Yik, E. J. & Chaput, J. C. Programmed allelic mutagenesis of a DNA polymerase with single amino acid resolution. *ACS Synth. Biol.* **9**, 1873–1881 (2020).

15. Janjua, S., Younis, S., Deeba, F. & Naqvi, S. M. S. High efficiency DNA transformation protocol for *Escherichia coli*using combination of physico-chemical methods. *Int. J. Agric. Biol.* **16**, 132–138 (2014).

16. McCullum, E. O., Williams, B. A. R., Zhang, J. & Chaput, J. C. Random mutagenesis by error-prone PCR. *Methods Mol. Biol.* **634**, 103–109 (2009).

2.11 Figures

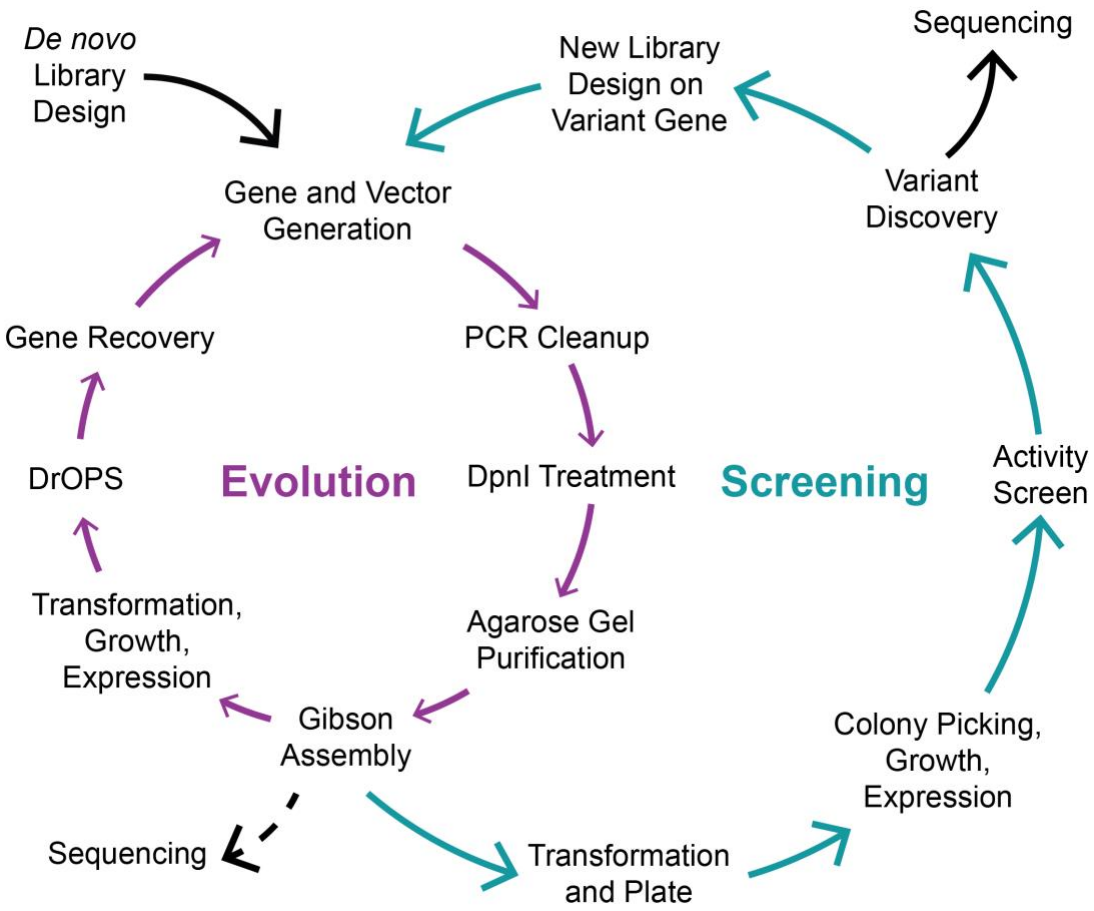


Figure 2.1 From In silico design to plasmid generation

A de novo library is designed through in silico modeling to verify sequence composition, primer binding sites, and proper translation. Genes are used directly from gBlocks or generated by PCR amplification with a high-fidelity polymerase, along with PCR generation of the vector backbone. The PCR amplicons are DpnI treated, agarose gel purified, and ligated through Gibson assembly. The vector library is transformed into E. coli, grown, expressed, and encapsulated into microdroplets with a fluorescent based assay. Active droplets are sorted, recovered, gene amplified, and prepared for iterative rounds of selection or activity screening. Key variants discovered from activity screens are sequenced by Sanger sequencing and become the backbone for the next library design.

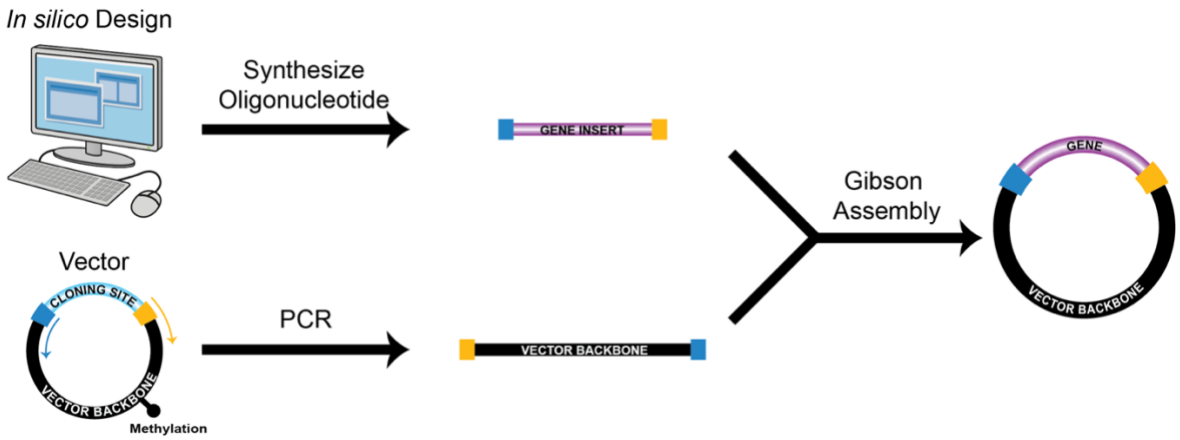


Figure 2.2 From In silico design to plasmid generation

Plasmid vector maps are prepared in silico with molecular biology cloning software to ensure that each gene used for homologous recombination is properly designed and assembled. Critical to this process is the design of synthetic recombination sites that are 100% identical across the homologs. The genes are ordered as gBlocks, which prepares the material as linear dsDNA. Separately, the vector backbone is generated by PCR and prepared for gene ligation by Gibson assembly. The gene inserts and vector backbone contain overlapping regions required for vector assembly.

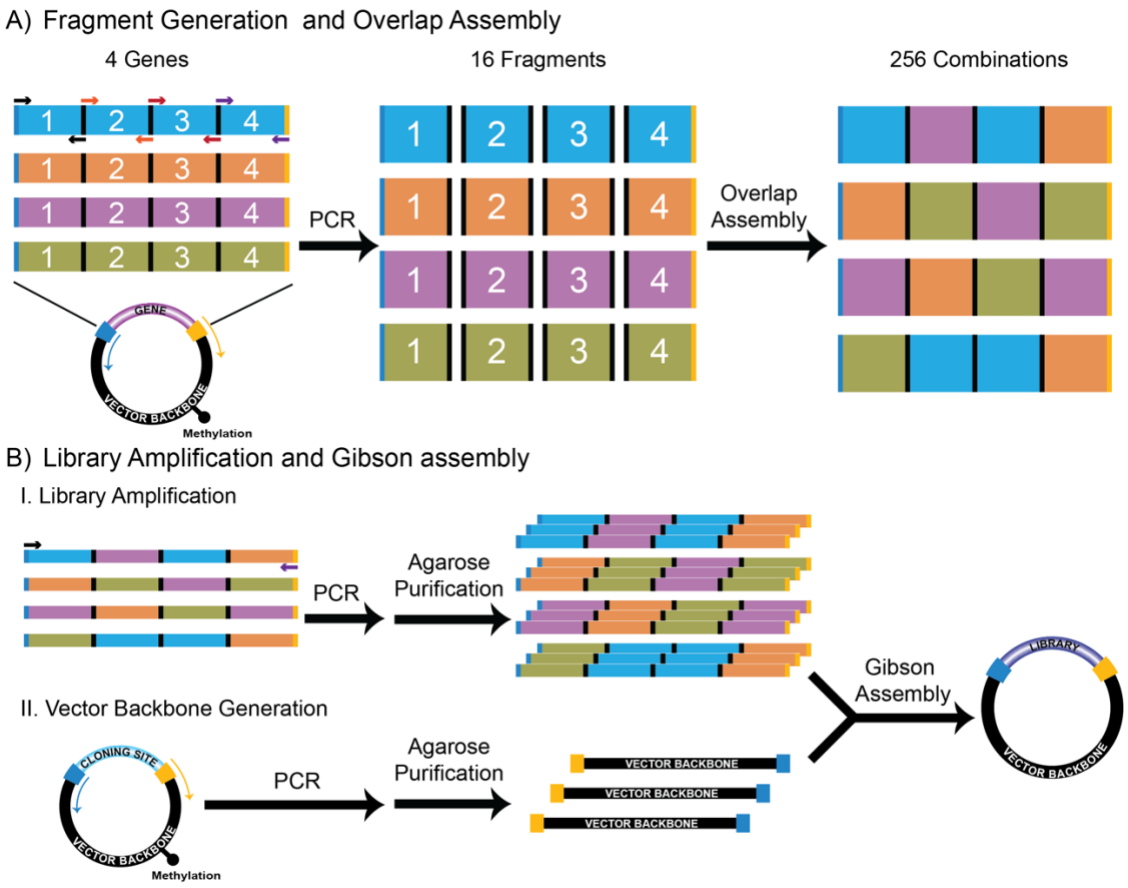


Figure 2.3 Generating a homologous recombination library by fragmentation and reassembly

A) Fragmentation and reassembly. Parent genes designed at the DNA level contain conserved regions across homologs that define the boundaries between segments of the genes that will be individually PCR amplified to create a library of DNA fragments. Fragments can vary in length and number depending on how the library was designed. The total number of gene fragments will equal the number of gene fragment regions times the number of homologs. Once the gene fragments are synthesized, the library is prepared by synthetic homologous recombination in a PCR reaction that contains all the gene fragments plus forward and reverse primer sites defining the vector overlap regions that extend past the 5' and 3' ends of the gene. B) Insertion of the library into a protein expression vector. The DNA library is amplified by PCR and inserted into a protein expression vector by Gibson assembly. It is important to agarose purify both the gene insert and vector backbone to isolate the expected amplicon.

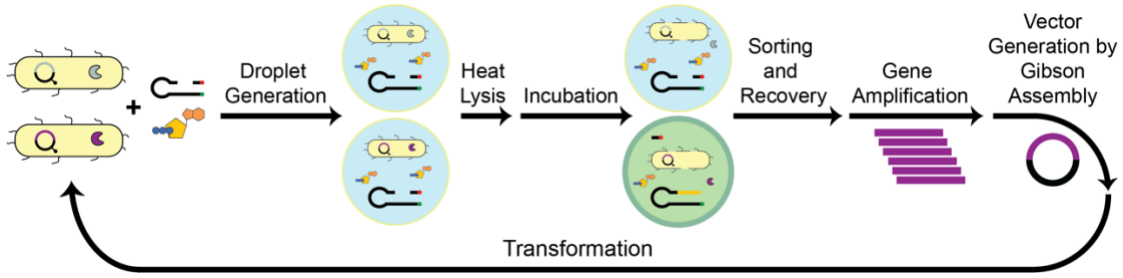


Figure 2.4 Directed evolution strategy for evolving TNA polymerases by droplet-based optical polymerase sorting (DrOPS).

A library of polymerase variants generated by homologous recombination or error-prone PCR is inserted into a protein expression vector by Gibson assembly. The plasmid library is transformed into E. coli, grown and induced to express the polymerase library. Post-induction, E. coli cells are encapsulated into microdroplets with the reagents necessary for performing a fluorescence-based polymerase activity assay. Droplets are collected, subjected to thermal lysis, and incubated at 55°C. Functionally active polymerases capable of extending a labelled DNA hairpin template with TNA generate a fluorescent signal by blocking a DNA quencher strand from annealing with the template. Fluorescent droplets are sorted by a custom fluorescently activated droplet sorter (FADS) mounted on a microfluidic instrument. Fluorescent droplets are collected, and the encoding DNA is amplified for another round of directed evolution.

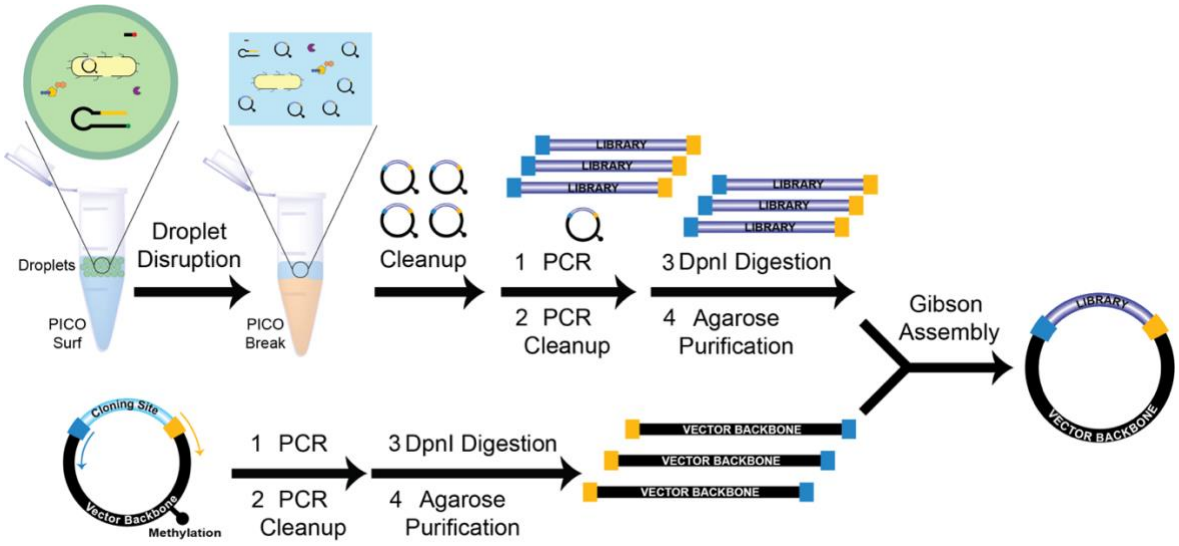


Figure 2.5 Plasmid extraction, gene regeneration, and plasmid assembly

Emulsion of positively collected droplets are broken with the addition of PICO Break in a 1:1 ratio of droplet and PICO Surf, which creates a biphasic mixture. The aqueous phase (top) containing the plasmids of interest are extracted and purified. The genes from the active library are PCR amplified, purified, treated with DpnI, and agarose gel purified prior to Gibson assembly with the newly generated expression vector backbone.

CHAPTER 3: Directed Evolution of a TNA Polymerase Identifies Independent Paths to Fidelity and Catalysis

Mohammad Hajjar[1†], Victoria A. Maola[1†], Joy J. Lee[1†], Manuel J. Holguin[1], Riley N. Quijano[1], Kalvin K. Nguyen[1], Katherine L. Ho[1], Jenny V. Medina[1], Elionel Botello-Cornejo[1], Bhawna Barpuzary[1], Nicholas Chim[1]*, and John C. Chaput[1-4]*

[†]These authors contributed equally

Affiliations

[1]Department of Pharmaceutical Sciences, University of California, Irvine, CA 92697, USA

[2]Department of Chemistry, University of California, Irvine, CA 92697, USA

[3]Department of Molecular Biology and Biochemistry, University of California, Irvine, CA 92697, USA

[4]Department of Chemical and Biomolecular Engineering, University of California, Irvine, CA 92697, USA

This chapter is a reprint of the material as it appears in *Nat Commun* 17, 925 (2026). https://doi.org/10.1038/s41467-025-67652-1, used with permission from Springer Nature Limited.

3.1 Abstract

Directed evolution facilitates functional adaptations through stepwise changes in sequence that alter protein structure. While most campaigns yield solutions that maintain the framework of a rigid protein architecture, a few have produced enzymes with more notable structural differences. One example is a polymerase that was evolved to synthesize threose nucleic acid (TNA) with near-natural activity. Understanding how this enzyme arose provides a model for studying pathways that guide enzymes toward more productive regions of the fitness landscape. Here, we trace the evolutionary trajectory of an unnatural polymerase by solving crystal structures of key intermediates along the pathway and evaluating their biochemical activity. Contrary to the view that fidelity is a product of increased catalytic efficiency, we find that accuracy and catalysis are decoupled activities guided by separate ground-state and transition-state discrimination events. Together, these results offer a glimpse into the forces responsible for shaping the emergence of new enzyme functions.

3.2 Introduction

Directed evolution is a powerful approach for generating enzymes that are needed to drive future applications in biotechnology and medicine[1-3]. Through recursive cycles of diversification and selection, populations of molecules adapt to novel biochemical functions that lie outside the sphere of activities found in nature[4]. Though effective, most protein evolution studies are unable to match the catalytic efficiency of natural enzymes[5]. The central challenge lies in understanding how to reshape the active site pocket to accommodate noncognate substrates while preserving the catalytic activity of the native enzyme fold. Although aspects of this problem have been studied by examining the products of directed evolution campaigns[6-9], understanding the molecular basis of functional adaptation requires structural information about key intermediates along the pathway, each representing a critical step toward the enzyme's final, functional form[10]. By studying the evolutionary journey from a low activity variant to a highly functional enzyme, it may be possible to learn basic design principles that could accelerate future selections or expand the predictive power of AI algorithms to include dynamic enzymes with complicated reaction mechanisms.

In recent work, we demonstrated the directed evolution of a polymerase, called 10-92, that can synthesize α-L-threofuranosyl nucleic acid (TNA) with near-natural activity (Fig. 3.1)[11]. Biochemical studies reveal that 10-92 achieves a catalytic rate of ~1 nt s^{-1} and >99% fidelity by sequentially adding TNA triphosphates (tNTPs) to the 2'-hydroxyl group of the growing primer, yielding an unnatural 3'→2' linked TNA product (Fig. 3.1a). The enzyme was discovered using the massively parallel screening technique of droplet-based optical polymerase sorting (DrOPS)[12,13] to query a large combinatorial library prepared by recombining the sequences of four hyperthermophilic archaeal B-family DNA polymerases (DNAP) (*Thermococcus kodakarensis*, Kod; *Thermococcus gorgonarius*, Tgo; *Pyrococcus Deep Vent*, DV; and *Thermococcus 9°N*, 9°N)

(Fig. 3.1b, 3.1c). Each polymerase carried the exonuclease silencing mutations (D141A and D143A) and four beneficial mutations (A485R, N491S, R606G, and T723A) found in Kod-RSGA, our previous best TNA polymerase (TNAP)[14].

A crystal structure of the 10-92 TNAP uncovered large movements in the protein architecture that improved substrate binding, including positional rearrangements of α-helical secondary structures (up to 7 Å) that bring the thumb subdomain closer to the paired region of the DNA duplex and other large movements that realign the catalytic triad in the palm subdomain (Appendix B: Supplementary Fig. 1,2)[11]. Together, these structural changes enabled the transition from an enzyme capable of highly efficient DNA synthesis to one that is now specialized for TNA synthesis (Appendix B: Supplementary Fig. 3). Whether these structural changes arose early or late in the evolutionary process remains to be established, as do the underlying mechanisms responsible for achieving enhanced catalytic activity.

Here we study the evolutionary trajectory of the 10-92 TNAP by solving crystal structures of key intermediates along the pathway and evaluating their biochemical activity. Crystal structures were solved as binary and closed ternary complexes, and biochemical data were collected for fidelity, catalytic efficiency, thermal stability, and substrate specificity. Among the more notable trends is the observation that fidelity and catalysis appear to occur as decoupled activities, possibly guided by separate ground- and transition-state discrimination events. Molecular insights into these processes were obtained from crystal structures that were used to map the process of active site refinement across successive generations of the enzyme. Together, these findings provide a detailed mechanistic framework for understanding how directed evolution reshapes enzyme function to achieve accurate and efficient catalysis.

3.3 Results

Biochemical analysis of the evolutionary intermediates

To follow the biochemical changes that occurred along the evolutionary path, we evaluated the polymerases that most influenced the evolution of the 10-92 TNAP for fidelity, activity, thermal stability, and substrate specificity. This analysis included Kod-RSGA as well as intermediates 5-270, 7-47, and 8-64 (Fig. 3.1c and Appendix B: Supplementary Fig. 4), the most active variants identified in rounds 5, 7, and 8, respectively, and the parent sequences used for further diversification[11].

Our analysis of fidelity and catalysis reveals a general trend in which steep gains in fidelity occur in earlier generations of the enzyme, whereas analogous improvements in activity are delayed until later stages of the evolutionary trajectory (Fig. 3.1d, Appendix B: Supplementary Figs. 5-7 and Supplementary Table 1). However, it should be noted that we do observe a concomitant improvement in fidelity as catalysis improves, presumably due to fine-tuning of the active site near the end of the evolutionary trajectory. Nevertheless, the overall trend of early improvements in fidelity and later improvements in catalysis caught our attention as it appears to counter the prevailing view that a selection for improved catalytic activity through the use of progressively shorter incubation times, as was the case for 10-92[11], should indirectly enrich for enzymes with higher fidelity[15,16]. This is because less faithful polymerases are more likely to stall at misincorporation sites, thereby failing to complete the extension within the allotted time[17-22]. Thus, variants that maintain rapid synthesis under stringent incubation times are expected to exhibit the combined properties of enhanced catalytic efficiency with increased fidelity. While this interpretation is consistent with the selection results overall, our biochemical analysis of the evolutionary intermediates indicates that variants can emerge that exhibit enhanced fidelity

without a proportional gain in catalytic activity. This observation implies that fidelity and catalysis may have determinants that can be separately tuned during the early stages of evolution.

We also found that all of the polymerases tested retain their extreme thermostability, showing no detectable loss in TNA synthesis activity after prolonged exposure to 90°C for 6 hours (Fig. 3.1d and Appendix B: Supplementary Fig. 8). This result highlights the exceptional structural integrity of the variants and challenges the view that directed evolution inevitably compromises protein folding stability[5]. Such trade-offs are particularly relevant in enzyme engineering campaigns, where enhanced activity often comes at the expense of reduced thermostability[23]. Remarkably, the 10-92 TNAP, with its 51 amino acid mutations relative to its closest natural homolog, Kod DNAP (Fig. 3.1c), exhibits no measurable loss of thermostability (Appendix B: Supplementary Fig. 8). This observation supports the view that protein stability promotes evolvability[24], which may be a result of the thermal lysis step of the DrOPS selection protocol[25].

Finally, the 10-92 TNAP also exhibits a striking shift in substrate specificity, defined here as the ability for the polymerase to distinguish TNA versus DNA substrates based on their catalytic rates of incorporation. Kinetic analysis across generations of the polymerase reveals a consistent reluctance of the enzymes to relinquish their ancestral DNA synthesis activity until the final stages of evolution, at which point an inversion (13-fold) in substrate specificity occurs, favoring TNA synthesis over DNA synthesis (Appendix B: Supplementary Fig. 7). This inversion of substrate preference marks the transition of the enzyme from a broadly acting generalist to one that is increasingly specialized for TNA synthesis on DNA templates. Such a shift is a hallmark of functional divergence, arising from the accumulation of mutations that reconfigure the active site toward a new function.

A binary complex capturing an unexpected preorganized state

As an initial step toward solving the puzzle of how 10-92 emerged as a highly active enzyme from a diverse pool of synthetic homologs, crystal structures of major intermediates identified along the evolutionary pathway were first solved as binary structures of the post-catalytic complex with a bound DNA primer-template (P/T) duplex that was enzymatically extended with two TNA residues (3'-tTtA-2') at the 3' terminus. Four structures (5-270, 7-47, 8-64, and 10-92) spanning a resolution limit of 2.17 – 2.56 Å were determined by molecular replacement (Appendix B: Supplementary Table 2). Consistent with all known archaeal B-family DNAPs, the enzymes adopt a disk-shaped architecture (Appendix B: Supplementary Figs. 9-10) defined by an N-terminal domain (NTD), an exonuclease domain, and a catalytic domain comprised of the finger, palm, and thumb subdomains[26]. In all cases, the duplex is well resolved and tightly bound between the palm and thumb subdomains (Appendix B: Supplementary Fig. 9).

Surprisingly, comparative structural analysis reveals an unexpected, closed conformation for the finger subdomain in all four of the TNAP structures. In this conformation, helices α14 and α15 are shifted by ~22° relative to their position in the open binary structure of wild-type Kod DNAP (Fig. 3.2a-e)[27]. This finding contrasts with all known binary structures of replicative DNAPs, which typically exhibit an open finger conformation ready to accept the incoming triphosphate[28]. Until now, the closed state has only been observed in structures of ternary complexes trapped in a catalytically competent state of the reaction cycle (Fig. 3.2b,c)[29,30].

While it is possible that one could consider this observation a crystallization artifact, several lines of evidence argue against this option. First, the TNAP variants (5-270, 7-47, 8-64, and 10-92) crystallized under two distinct precipitation conditions, suggesting that the conformation reflects a *bona fide* structural feature rather than a condition-specific artifact.

Second, the closed binary conformation is absent in previous binary structures of Kod-RI[31] and Kod-RSGA[32] (Appendix B: Supplementary Fig. 11), further strengthening the view that the conformational difference arises from sequence changes rather than crystallization chemistry. Finally, the closed binary conformation emerged in variants that also exhibit changes in their biochemical activity, providing a functional correlation that supports its structural relevance.

To understand the molecular basis of this unique conformation, we analyzed the interdomain interactions stabilizing the closed state. In the binary form of natural Kod DNAP, only limited contacts exist between the exonuclease domain and finger subdomain (Fig. 3.2f)[27], whereas noticeably more contacts are observed as Kod transitions to a closed ternary conformation, supporting a catalytically competent state with the dNTP poised for chemical bond formation (Fig. 3.2g)[26]. Contrasting the natural system, the closed binary conformation of the TNAP intermediate observed in 5-270, 7-47, 8-64, and 10-92, exhibits a robust network of electrostatic and hydrophobic interactions that preorganize the enzyme in a catalytically active conformation despite the absence of a bound tNTP (Fig. 3.2h).

The structural similarity between the binary TNAP complexes and the closed ternary structure of 10-92 previously solved by X-ray crystallography is striking, particularly in the geometry of the active site, which differs only by the absence of substrate-specific contacts (Appendix B: Supplementary Fig. 12)[11]. Because the closed binary motif first appears in the 5-270 intermediate and is absent in the binary structure of Kod-RSGA[32] (Appendix B: Supplementary Fig. 11), we infer that this conformational switch arose from recombination events that facilitated TNA synthesis by stabilizing the enzyme in a catalytically-productive state. In this scenario, the finger subdomain exists in a dynamic equilibrium between the open and closed conformations, reminiscent of natural DNAPs[20]. However, upon correct substrate binding and the ensuing

conformational change[19], the preorganized state would provide a selective advantage by biasing the enzyme toward the closed conformation, thus offering more time for catalysis to occur prior to substrate release.

The catalytically active state of the enzyme

To visualize how catalytic activity evolved within the TNAP lineage, we determined the crystal structures of key intermediates in a catalytically competent state with a closed finger conformation and a P/T duplex and tATP substrate bound in the enzyme active site. A key aspect of this effort was the chemical synthesis of a chain-terminating 2'-deoxy-α-L-threofuranosyl thymidine-3'-triphosphate (dtTTP) analog, prepared in 8 synthetic steps from a 3'-protected α-L-threofuranosyl thymidine (tT) precursor using a Barton deoxygenation strategy to remove the 2' hydroxyl group (Appendix B: Supplementary Figs. 13-16)[33,34]. This substrate enabled us to capture the reaction state immediately following tNTP binding, but before chemical bond formation. Using this approach, we successfully obtained high resolution crystal structures for 5-270 and 8-84, resolving to 3.03 Å and 2.38 Å, respectively (Appendix B: Supplementary Table 3).

By combining this data with previously solved structures for Kod-RI[31] and 10-92[11], the least and most optimized versions of the enzyme, respectively, a structural framework was established to study the biochemical changes that occurred along the evolutionary path. Superposition of the four TNAP structures (Kod-RI, 5-270, 8-64, and 10-92) reveals that a global divergence in structure occurred following recombination (Fig. 3.3a,e and Appendix B: Supplementary Figure 17). The backbone Cα root mean square deviation (RMSD) relative to wild-type Kod DNAP increases from ~0.5 Å for Kod-RI to ≥1.0 Å for 5-270, 8-64, and 10-92, with movements of secondary structural elements observed in the thumb, palm, and N-terminal domains. These changes correspond to a progressive remodeling of the active site pocket (Fig. 3b-f), which aligns with observed

biochemical gains in TNA synthesis activity (Fig. 3.1d). Surprisingly, most of the 51 acquired mutations found in 10-92 are located >20 Å from the catalytic core (Fig. 3.3a, and Appendix B: Supplementary Fig. 18), emphasizing the importance of long-range structural interactions in fine-tuning the active site geometry[35].

Importantly, deep learning models failed to predict these transitions. Structure predictions by AlphaFold3 yielded only canonical B-family DNAP structures, with open binary conformations and no evidence of the experimentally observed rearrangements (Appendix B: Supplementary Figs. 19-23). This observation underscores a key limitation of current AI-based modeling, which is the inability to accurately capture long-range mutations and evolution-driven conformational shifts that underlie new protein functions. Future improvements in predictive accuracy will require experimental benchmarks and better mechanistic understanding of conformational plasticity during enzyme evolution.

Reshaping the Active Site Pocket

Our previous crystal structure of the 10-92 TNAP trapped in a closed ternary complex revealed how recombination and molecular evolution enabled substantial active site remodeling to support efficient TNA synthesis[11]. The selected mutations produced local backbone shifts, including high-energy rearrangements, that restructured the active site pocket to promote optimal hydrogen bonding, electrostatic stabilization, and van der Waals complementarity with the bound tNTP. In this conformation, the catalytic aspartate triad (D405, D541, D543) is precisely arranged to coordinate metal ions and facilitate robust nucleotide transfer chemistry.

By integrating new structural data from intermediates along the evolutionary path, we now provide a high-resolution, atomic-level view of how these changes unfolded. A central theme to emerge is the stepwise refinement of the active site geometry, beginning with a disruptive

expansion of the active site volume followed by precise, substrate-specific remodeling. For example, the active site volume increases ~250% from wild-type Kod DNAP to Kod-RI, before contracting through 5-270 and 8-64 to a more compact volume of 692 $Å^3$ observed in 10-92, which is approximately 10% smaller than the starting enzyme (Fig. 3.3c–d, Appendix B: Supplementary Fig. 24). This process of active site refinement allows the enzyme to accommodate the smaller size of the tNTP substrate, noting that TNA derives from a four-carbon threose sugar, as compared to the five-carbon sugar found in DNA[36].

At the local level, analysis of the catalytic triad reveals a clear stepwise refinement across the lineage. Overlaying the active site from successive TNAP generations shows that the catalytic residues undergo a series of positional and rotational movements, supported by Polder maps, that ultimately align the aspartate residues into a catalytic configuration closely mimicking the arrangement found in wild-type Kod DNAP (Fig. 3.3e,f and Appendix B: Supplementary Figs. 25, 26). These refinements enable TNA substrate recognition while preserving the structural features essential for metal ion coordination and catalysis.

Our structural analysis reveals that base pair recognition and reaction center geometry emerged as two distinct evolutionary pressures shaping the evolutionary trajectory. The former is reflected in the conformation of the nascent Watson-Crick base pair formed between the incoming tATP and the templating thymine base. Following the active site distortion observed in Kod-RI, base pair geometry was largely restored in 5-270, as measured for the parameters of opening, propeller, and buckle (Fig. 3.4, Appendix B: Supplementary Table 4). By comparison, the geometry of the reaction center improved more gradually across generations of the enzyme, with the distance from the C2' group of the primer to the α-phosphate of the bound tATP steadily decreasing from 5.8-6.7 Å in 5-270 (Appendix B: Supplementary Fig. 27) to eventually 3.9 Å in

10-92. This final value closely approximates the ideal distance of 3.7 Å observed in wild-type Kod DNAP (Fig. 3.4b). Together, these findings underscore the evolutionary mechanism underpinning TNAP optimization; namely, early gains in fidelity correspond to improved base pair recognition, while later gains in catalytic activity emerge from a fine-tuning of the active site geometry. This structural and functional convergence culminates in a highly specialized enzyme whose performance reflects both macro-scale architectural adaptation and micro-scale atomic-level precision.

Mutational Landscape Underlying the Activity of 10-92.

The mutational landscape underlying the evolution of the 10-92 TNAPs reveals a clear shift from broadly distributed, largely conserved substitutions found in 5-270 to a set of later-arising, non-conserved mutations that cluster in the thumb subdomain and strongly impact TNA synthesis activity (Fig. 3.5 and Appendix B: Supplementary Table 5). This distribution is unsurprising as the mutations found in 5-270 are the result of homologous recombination, while those in 10-92 derive from focused mutagenesis[11]. Structural analyses show that mutations occur across both surface exposed and buried positions, with most mutations residing far from the catalytic center (Fig. 3.5), again underscoring the importance of long-range interactions in shaping new protein functions[35]. Computational free-energy predictions indicate that while early substitutions are generally neutral with respect to stability, most of the function-driving mutations obtained from directed evolution are destabilizing (Fig. 3.5), highlighting the trade-off between structural integrity and catalytic enhancement[23]. Limited reversion analysis of later-stage mutations demonstrates that at least one function-driving mutation is present in each of the key intermediates (Fig 3.5)[11]. Together, these data suggest that polymerase specialization for TNA synthesis emerged through a process involving the accumulation of large numbers of mostly

neutral mutations via homologous recombination that change the protein architecture, followed by a smaller set of adaptive mutations that refine catalytic activity at the cost of local stability.

Identifying the precise set of mutations responsible for improved fidelity is challenging because 5-270 carries 35 mutations, all derived from natural diversity. These substitutions collectively remodel the enzyme architecture, produce major structural alterations that reshape the active site, improve the geometry of the nascent Watson-Crick base pair of the incoming tNTP, and facilitate formation of the preorganized conformation observed in the binary structures. While this extensive remodeling likely underlies the fidelity gains observed in 5-270, the large number of changes makes it difficult to determine which mutations are adaptive drivers versus neutral passengers that simply accumulate through the process of directed evolution. Nevertheless, some mutations do appear to contribute disproportionately to these global changes, including the 381L insertion and substitutions within the thumb subdomain (G602D and K672R). Future studies aimed at disentangling the functional role of these and other mutations will be critical for elucidating the mechanistic basis of the fidelity improvement observed early in the evolutionary path.

Interpreting the mutations responsible for improved catalytic activity is aided by a single-point reversion analysis performed on the ten amino acid mutations observed between 5-270 and 10-92[11]. This analysis revealed that the top four most important mutations in terms of their contribution to catalytic improvement (D615G, A741P, T548P, and S493G) involve the gain of either a glycine or proline residue, while the fifth most important mutation (P550H) transitioned a proline to histidine (Figs. 3.5, 3.6). Together, these changes highlight the importance of flexibility and rigidity as a mechanism for fine-tuning the activity of the catalytic center. While these mutations are located ~20 Å (except S493G which is ~13 Å) from the active site, defined as the 2' terminus of the primer, they have a profound impact on enzyme function. The D615G and A741P

mutations, first observed in 7-47 and 8-64, respectively, directly contact the primer and template strands, respectively (Fig. 3.6 and Appendix B: Supplementary Figs. 28, 29). These interactions balance the need for the enzyme to recognize the duplex tight enough to allow for catalysis yet weak enough to facilitate efficient translocation. The T548P and P550H mutations, both observed in 10-92, induce structural changes in a loop that optimizes the position of the catalytic triad for efficient coordination to divalent metals (Fig 3.6 and Appendix B: Supplementary Figs. 28-30), allowing for enhanced catalytic efficiency. Finally, the S493G mutation observed in 10-92, is located in the active site pocket (Fig. 3.6 and Appendix B: Supplementary Figs. 28, 29). This mutation appears to reduce suboptimal interactions between the enzyme and templating base, resulting in better substrate binding and catalysis.

3.4 Discussion

This study provides a high-resolution structural and biochemical view of how directed evolution shapes the structure and function of an enzyme to enable the synthesis of an artificial TNA polymer. By mapping the evolutionary trajectory of a highly efficient TNA polymerase isolated from a molecular evolution study involving recombination and directed evolution[11], we uncover a series of coordinated molecular events that separate the emergence of fidelity from catalytic efficiency, two properties traditionally thought to be coupled activities in enzyme evolution[15-17].

Contrary to prevailing assumptions, we observe that fidelity and catalysis diverge early in the evolutionary process and are refined through distinct structural pathways. Fidelity improves rapidly across early variants, corresponding to restoration of Watson–Crick base pair geometry and nascent base pair stability. In contrast, gains in catalytic activity and altered substrate

specificity appear only in later generations and correlate with progressive remodeling of the active site architecture and reaction center geometry. These findings suggest that ground-state and transition-state discrimination can evolve independently, offering a modular mechanism by which enzymes acquire new function. A possible reason for this distinction could be that transition-state optimization is more difficult to realize than ground-state optimization, as the free-energy of transition-state complexes are much higher than those of ground-state complexes[16].

A defining feature of the evolved polymerase lineage is the adoption of a preorganized, closed finger conformation in the absence of a bound nucleotide. This catalytically competent architecture, which first appears in intermediate 5-270, the most active variant identified after recombination and five rounds of positive selection for TNA synthesis activity, is stabilized by a network of interactions that shift the conformational equilibrium away from the open conformation commonly associated with the binary complex of replicative DNA polymerases. This structural transition may represent a critical inflection point in the evolutionary process, enabling the enzyme to escape the kinetic penalties associated with accommodating the unnatural TNA substrate. Notably, this conformational switch was not predicted by deep learning-based structure prediction models (AlphaFold3), underscoring current limitations in modeling dynamic, evolution-driven enzyme conformations.

Thermal profiling further reveals that even extensive mutational remodeling, 51 mutations in the case of 10-92, does not compromise the structural integrity of the enzyme fold. Instead, the evolved polymerase retains full activity following prolonged exposure at 90 °C. We attribute this remarkable thermostability to the stringent selection environment imposed by the DrOPS screening workflow, which includes a high-temperature lysis step. These findings challenge the view that functional gains necessarily incur stability trade-offs and highlight the value of

integrating environmental constraints into the selection process. Additionally, this observation supports the view that intrinsic protein folding stability is an important criterion for promoting protein evolvability[24]

The evolution of substrate preference provides deeper insight into the adaptive landscape navigated by the polymerase. Across the evolutionary trajectory, successive variants exhibited progressively higher TNA synthesis rates, while DNA synthesis activity remained predominant. A marked inversion in substrate specificity emerged only in variant 10-92 (Appendix B: Supplementary Figure 7). The preference of 10-92 for TNA likely stems from its remodeled active site, whose reduced volume reflects structural adaptation to the smaller size of the TNA triphosphate. This inversion in specificity represents a clear case of functional divergence and illustrates how exploration of sequence space (the ensemble of all possible protein sequences of a given length) can yield enzymes that are not only structurally robust, but also precisely tuned for novel chemical tasks.

Taken together, our findings illustrate one example of how directed evolution can orchestrate long-range structural rearrangements and local active site refinement to generate an enzyme with tailor-made activity. The modular separation of fidelity and catalysis observed here offers a new framework for understanding enzyme adaptation and informs future efforts to design polymerases for applications in synthetic biology and medicine. More broadly, this work underscores the importance of combining structural biology with evolutionary trajectories to illuminate mechanisms of molecular innovation that remain opaque to current predictive models.

3.5 Methods

Reagents

DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT, Coralville, Iowa). TNA triphosphates were obtained by chemical synthesis previously[34,37]. ThermoPol buffer, Q5 DNA polymerase, Xl1 Blue competent *E. coli*, BL21(DE3) competent *E. coli*, Taq DNA polymerase, *NdeI* and *NotI* restriction enzymes were purchased from New England Biolabs (Ipswich, MA). Qiaprep plasmid minikit was purchased from Qiagen (San Diego, CA). Amicon centrifugal filter, hen egg white lysozyme, and chemical reagents including dNTPs, sodium chloride, manganese chloride, magnesium chloride, Tris-HCl, isopropyl β-D-thiogalactoside (IPTG), 1,4-dithiothreitol (DTT), ampicillin, and ammonium persulfate (APS) were purchased from Sigma Aldrich (St. Louis, Missouri). TOPO TA cloning kit, ethylenediaminetetraacetic acid (EDTA), and Dynabeads M270 were purchased from Thermofisher Scientific (Waltham, Massachusetts). SequalGel UreaGel 29:1 Denaturing Gel System was purchased from National Diagnostics (Atlanta, GA). Tetramethyl-ethylenediamine (TEMED) was purchased from Bio-Rad (Hercules, California). Chromatography columns were purchased from Cytiva (Little Chalfont, United Kingdom). EvaGreen® dye was purchased from Biotium (Fremont, CA). Clear V-bottom 96-well plates were purchased from Greiner (Monroe, NC). Plastic 1.5 mL micro-centrifuge tubes were purchased from Sigma-Aldrich (St. Louis, MO). Crystallization screens and individual crystallization reagents were purchased from Hampton Research (Aliso Viejo, CA), NeXtal Biosciences (Holland, OH), and Molecular Dimensions (Holland, OH). The Mosquito crystallization robot was purchased from SPT LabTech (Covina, CA). pET-21b(+) plasmid was purchased from Novagen Technology (Glendale, CA). Whole plasmid and Sanger sequencing were performed using Plasmidsaurus (Los Angeles, CA) and Genewiz, respectively (San Diego, CA).

TNAP expression and purification for biochemical assays

Plasmids for TNAP expression and purification were generated previously[38]. Briefly, a pGDR11 vector containing the TNAP of interest was transformed into XL1 Blue *E. coli*, grown, induced for expression, and harvested. Cells were resuspended and sonicated in TNAP buffer (10 mM Tris-HCl, pH 8.0, 500 mM NaCl, 10% glycerol). The cell lysate was heat treated at 70 °C for 1 h, cooled on ice for 1 h, and centrifuged (23,708 x g, 4°C, 20 min). The clarified supernatant was treated with a final concentration of 0.5% (v/v) PEI for 15 min and centrifuged (23,708 x g, 4°C, 20 min) to remove nucleic acids. Ammonium sulfate precipitation was performed (final concentration: 60% (w/v)), centrifuged (23,708 x g, 4°C, 20 min), and the protein pellet was resuspended in equilibration buffer (10 mM Tris-HCl, pH 8.0, 50 mM NaCl, 10% glycerol). The TNAP was purified by hand on a 5 mL heparin affinity column and eluted by stepwise addition of buffers containing: 10 mM Tris-HCl, pH 8.0, 10% glycerol and increasing concentrations of NaCl (50, 100, 250, 500, 750 mM). Eluted TNAP at 500 mM NaCl were pooled and concentrated to 100 μM for biochemical assays.

TNAP kinetics

TNAP kinetic measurements were performed by monitoring the reaction progress as the DNA primer-template duplex (PBS8: GTCCCCTTGGGGATACCACC & EM619: CCCACACCCTCCTATCGCTAAACACACACTTAATAAAGTTGGTGGTATCCCCAAGGG GAC, Appendix B: Supplementary Table 1) was extended with dNTPs or tNTPs to measure DNA or TNA synthesis, respectively, over time[39]. For each time point, 2 individual reaction replicates (15 μL) from a single master mix poised under single-turnover conditions with equimolar concentrations of TNAP (0.2 μM) and pre-annealed duplex (0.2 μM each, heated at 90°C for 5

min, cooled on ice) in 1x ThermoPol buffer [20 mM Tris-HCl, 10 mM $(NH_4)_2S)_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1% Triton X-100, (pH 8.8)] were pre-equilibrated for 5 minutes at 55°C. The reactions were then initiated by adding a preheated triphosphates mixture (200 μM of each). The reactions were stopped at designated time points by plunging the reaction vessel into powdered dry ice. Once all the time points were collected, the frozen reactions were thawed on a cold plate by adding 15 μL (6 μM) EvaGreen® dye, previously identified as the optimal intercalating dye to monitor the synthesis on a DNA template[39]. Each reaction (25 μL) was transferred to a clear V-bottom 96-well plate, and fluorescence intensity was measured (ex: 487/20 nm, em: 528/20 nm) using a Synergy Neo2 plate reader with the Gen5 v3.14.03 software (BioTek).

For both DNA and TNA synthesis, baseline fluorescence was measured using Kod (exo-), the parent polymerase with negligible TNA synthesis ability, which was used to collect a 0-second time point by freezing the reaction immediately after adding tNTPs. The maximum fluorescence value was obtained from a 15-minute reaction with each polymerase. The raw fluorescence data from the plate reader was normalized by subtracting the baseline and dividing by the difference between the maximum (15-min reaction) and minimum (baseline) fluorescence values. The fluorescence data was then converted to nucleotides per polymerase[39] by multiplying data with the conversion factor F = [Template] * L / [Polymerase], where L is the length of extension in bases, then plotted over time. Rates in nucleotides per minute are extracted as the slope of the linear range of each curve. Data were processed using Microsoft Excel v16.66.

TNAP fidelity

Fidelity analysis was performed in hydrogel particles[40]. In brief, an aliquot of ~12 million hydrogel particles displaying the acrydite PBS8 primer (GTCCCCTTGGGGATACCACC, Appendix B: Supplementary Table 1) was placed into a 1.5 mL and washed twice with 100 µL 1x ThermoPol buffer. The hydrogels were then resuspended in 100 µL 1x ThermoPol buffer containing 2 µM 4NT.9G fidelity template (Supplementary Table 1) and 1% KF-6012 and heated at 95°C for 3 min. The reaction mixture was then snap-cooled on ice for 5 min to promote primer-template annealing. To initiate TNA synthesis, 2 µL of tNTPs (5 mM stock) and 20 µL of TNAP (10 µM stock) was added, mixed, and then incubated at 55°C on a ThermoMixer C for 1 hour. After the reaction is complete, hydrogels were washed with 100 µL breaking buffer (10 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM EDTA, 1% (v/v) SDS, 1% (v/v) Triton X-100) twice, incubated with 100 µL 100 mM NaOH (37°C, 3 min) to strip the template, and neutralized with 100 µL breaking buffer. The hydrogels were resuspended in 100 µL 1x ThermoPol buffer containing 2 µM PBS7.PBS9 overhang primer (Appendix B: Supplementary Table 1) with a double nucleotide mismatch (TT-TT), 0.01% KF-6012, 3 mM magnesium sulfate and then heated at 95°C for 3 min. The reaction mixture was then snap-cooled on ice for 5 min to promote primer-template annealing. To initiate the reverse transcription reaction, 10 uL of dNTPs (5 mM stock) and 20 µL of Bst-LF (10 µM stock) was added, mixed, and then incubated at 50°C on a ThermoMixer C for 4 hours. The cDNA was PCR amplified with Taq DNA polymerase using PBS8 and PBS9: CTTTTAAGAACCGGACGAAC as primers (Appendix B: Supplementary Table 1). PCR material was ligated into a TOPO vector using a TOPO-TA kit and cloned into DH5α competent *E. coli* cells. Individual clones were picked, grown in liquid LB media, and sequenced. DNA sequences were aligned and analyzed with the template using CLC Main software. Error rates were

determined as $\mu_{exp \rightarrow obs}$ = (#observed/#expected) × 1000. The total error rate was determined by summing the error rate for each substitution.

TNAP thermal challenge

5 x 1.5 mL Eppendorf tubes, each containing 100 µL Kod or TNAP (10 µM), were placed into an Eppendorff ThermoMixer set at 90°C with a thermal lid to prevent evaporation. After 6 hours, the tubes were spun at 16,3000 x g for 5 min at 4°C and UV quantified using a NanoDrop instrument. TNA synthesis was assayed with a primer-template duplex (PBS8_short: /5IRD680/GTCCCCTTGG & EM619: CCCACACCCTCCTATCGCTAAACACACACTTAATAAAGTTGGTGGTATCCCCAAGGG GAC, Appendix B: Supplementary Table 1). Polymerase activity screen was performed with the following final concentrations: 1 µM primer-template duplex, 100 µM tNTPs, 1 µM heat-treated polymerase in a 20 µL reaction. After 30 minutes of incubation, a 1 µL aliquot of the reaction was quenched by the addition of 39 µL of quenching buffer and visualized by denaturing polyacrylamide gel electrophoresis with fluorescent imaging on a Li-Cor Odyssey CLx Imager, Image Studio Lite v5.2.

TNAP cloning, expression, and purification for crystallography

Full-length *tnap* genes in pGDR11 were PCR amplified using the TNAP-Fwd (ATCCATATGATCCTCGACACTGACTAC) and TNAP-Rvs primers (ACGCATG*CGGCCGC*TCAAGTTCCTTTCGGCGTCAG), Appendix B: Supplementary Table 1, containing *NdeI* and *NotI* restriction enzyme sites, respectively. C-terminal truncation constructs for all TNAPs, except for 8-64, were PCR amplified by substituting the TNAP-Rvs primer with

TNAP-Rvs_760 (ACGCATG*CGGCCGC*TCACTTCTGGTAGCGCAGGTC, Appendix B: Supplementary Table 1); 8-64-Rvs_760: ACGCATG*CGGCCGC*TCAAGTTCCTTTCGGCGTCAG was used to amplify 8-64, which harbored acquired mutations at the C-terminal. All reverse primers contain the *NotI* restriction enzyme site. Purified PCR product of each gene construct and pET-21b(+) were digested with *NdeI* and *NotI* restriction enzymes and ligated, resulting in pET21-*5-270*, pET21-*5-270_760* pET21-*7-47*, pET21-*7-47_760*, pET21-*8-64*, pET21-*8-64_760*, and pET21-*10-92*. Post verification by whole plasmid sequencing, all TNAP constructs were expressed and purified[31]. Briefly, BL21(DE3) cells harboring pET21-*tnap* were grown aerobically at 37 °C in LB medium containing 100 μg mL^{-1} ampicillin. At an OD_{600} of 0.8, expression of a tagless TNAP was induced with 0.8 mM isopropyl β-D-thiogalactoside at 18 °C for 20 hr. Cells were harvested by centrifugation for 20 min at 3315 × *g* at 4 °C and lysed in 40 mL lysis buffer (10 mM Tris-HCl pH 7/6.5 - full-length/truncated TNAP, 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 10% glycerol, 5 mg hen egg white lysozyme) by sonication. The cell lysate was centrifuged at 23,708 × *g* for 30 min and the clarified supernatant was heat treated for 20 min at 70 °C and centrifuged again at 23,708 × *g* for 30 min. The supernatant was loaded onto 5 mL HiTrap Q HP and heparin HP columns assembled in series with the efflux of the Q column loaded in front of the heparin column. After washing with lysis buffer, the Q column was removed and TNAP was eluted from the heparin column with a high salt buffer (10 mM Tris-HCl pH 7/6.5 - full-length/truncated TNAP, 1 M NaCl, 0.1 mM EDTA, 1 mM DTT, 10% glycerol) using a linear gradient. Eluted fractions containing TNAP were visualized by SDS-PAGE, pooled, and concentrated using a 30 kDa cutoff Amicon centrifugal filter. Further purification was achieved by size exclusion chromatography (Superdex 200 HiLoad 16/600) pre-equilibrated with TNAP buffer (50 mM Tris-HCl pH 8.5, 200 mM NaCl,

0.1 mM EDTA, 1 mM DTT). Purified TNAP was concentrated to 20 mg mL^{-1} for crystallization trials.

TNAP crystallography

The binary DNA template, T1: /5Cy5/AAACGTACGCAGTTCGCG, was HPLC purified sample while the remaining oligonucleotides, P: CGCGAACTGCG and T2: TATGCACGTACGCAGTTCGCG, were ordered with standard desalting (Appendix B: Supplementary Table 1). The primer-template duplexes, P/T1 and P/T2 for the binary and ternary complexes, respectively, were prepared by combining equal parts of the respective primer and template strands in TNAP buffer and supplemented with 20 mM MgCl2, heating at 95°C for 5 min and slow cooling to 10°C over 10 min.

All constructs were screened against ~900 conditions in a hanging-drop format using a Mosquito crystallization robot (SPT LabTech). Positive crystal hits were further optimized in 24-well hanging drop trays, with each drop consisting of 1 µL of sample mixed with 1 µL of mother liquor over 500 µL mother liquor in every well. The specific crystallization conditions for each TNAP complex are described below. Six diffraction datasets were collected at synchrotron sources (Advanced Light Source, Stanford Synchrotron Radiation Lightsource, and National Synchrotron Light Source II) from single crystals. Unless specified, images were indexed, integrated, and scaled using XDS[41]. Data collection statistics are summarized in Tables S2 and S3. Initial models were determined by molecular replacement using Phaser[42]; for the binary complexes, the Kod-RSGA binary complex (7RSU) was initially used as the search. However, the 10-92 closed ternary complex (8T3X) was subsequently used due to the closed conformation of the binary structures.

8T3x was used as the search model for the ternary complexes. All final models were determined using iterative rounds of manual building through Coot[43] and refinement with phenix[44]. The stereochemistry and geometry of all structures were validated with Molprobity[45] with the final refinement parameters summarized in Tables S2 and S3. Final coordinates and structure factors have been deposited in the Protein Data Bank.

Binary complexes were prepared by incubating full-length TNAPs (6 mg mL^{-1}) with 1.2 molar equivalents of the annealed P/T1 at 37°C for 30 min. 5 M excess of tTTP and tATP were added and incubated at 37°C for 30 min separately. The resulting 5-270 binary complex crystallized in 0.1 M calcium acetate, 0.1 M 2-(N-morpholino)ethanesulfonic acid pH 6, and 15 % polyethylene glycol 400 while the remaining binary complexes crystallized in 0.06 M magnesium chloride hexahydrate, 0.06 M calcium chloride dihydrate, 0.1 M imidazole, 0.1 M 2-(N-morpholino) ethanesulfonic acid pH 6, 17.5–19 % glycerol, and 7.5–9 % polyethylene glycol 4000. The images for the 7-47 binary complex were indexed, integrated, and scaled using iMOSFLM.

Initial binary complexes were prepared by incubating truncated TNAPs (6 mg mL^{-1}) with 1.2 M equivalents of the annealed P/T2 duplex at 37°C for 30 min. 3 M excess of dtTTP was added and incubated at 37°C for 30 min. The resulting n+1 binary complex was further incubated with 10 M excess tATP at 37°C for another 30 min to yield final ternary complexes. 5-270 ternary complex crystals grew in 0.2 M sodium iodide, 20 % polyethylene glycol 3350, and 0.1 M calcium chloride dihydrate and 8-84 ternary complex crystals grew in 0.2 M sodium iodide, 20 % polyethylene glycol 3350, and 0.1 M taurine.

TNAP structural analysis

Molecular visualizations, RMSD, and distance and torsion angle measurements were performed using licensed PyMOL (version 2.5.2; https://www.pymol.org)[46] and open-source PyMOL (version 3.1.0; https://github.com/schrodinger/pymol-open-source). Structural alignments for visualization and analysis in Figures 3 and 4 were carried out using the 'align' function in PyMOL, applied to the entire protein–nucleic acid complex. Polder omit maps were generated in Phenix using phenix.polder and visualized in PyMOL[47]. Local base-pair and base-pair step parameters are computed using Web 3DNA [48]. 2D interaction maps were assembled with data from Mapiya and DNAproDB. Pocket volumes were calculated using PyVOL.

Integrated Mutational Analysis

Sequence alignments used to generate the fitness landscape (Appendix B: Supplementary Fig. 3) were performed using the EMBOSS Needle algorithm[49] on the EMBL–EBI platform[50]. Conservation was quantified with BLOSUM62 substitution scores[51].

Functional effects of reversions to 5-270 were quantified from the 45-sec primer extension assay as previously reported[11]. Band intensities were measured by densitometry in ImageJ[52] as the fraction of full-length product (top band) relative to the total pixel density of each lane. Activities were normalized to the performance of 10-92.

Mutational effects on folding stability were estimated using FoldX5[53] with the BuildModel command applied to a repaired Kod^{exo-} closed ternary structure (PDB ID: 5OMF). Reported $\Delta\Delta G$ values are in kcal mol^{-1}.

Distances between each mutated alpha carbon atom and the active site, defined by the sugar moiety of the dideoxy primer terminus (C2' in TNA or C3' in DNA) were calculated in PyMOL on ternary complexes of Kod, 5-270, and 10-92.

Relative solvent-accessible surface area (SASA) was computed on ternary structures using FreeSASA[54] with the default parameters. Values were reported as relative exposure (%) in ternary complexes of Kod, 5-270, and 10-92.

AlphaFold3 predictions

AlphaFold3[55] structural predictions were performed on the online server provided by Google Deep Mind (https://alphafoldserver.com/) on 2024-08-16, before the initial public release of the 10-92 ternary structure (PDB ID: 8T3X) on 2024-08-28. Two AlphaFold3 runs (denoted AF3-1 and AF3-2) were performed for every evaluated structure, producing 5 models per run (models 0 – 4, denoted by subscripts). The binary structural predictions were obtained by inputting the sequence of residues 1-776 in the presence of a DNA primer (CGCGAACTGC), DNA template (AAACGTACGCAGTTCGCG) and 2 Mg^{2+} ions. Ternary structural predictions were obtained by inputting the sequence of residues 1-760 in the presence of a DNA primer (CGCGAACTGCGT), DNA template (TATGCACGTACGCAGTTCGCG), an ATP, and 2 Mg^{2+} ions.

2′-deoxy-α-L-threofuranosyl thymidine-3′-triphosphate (dtTTP) synthesis

Synthetic procedures: All moisture sensitive reactions were performed in anhydrous solvents under an argon or nitrogen atmosphere. All commercial reagents, solvents and anhydrous solvents were used without further purification. Reaction progresses were monitored by thin layer chromatography using glass-backed analytical Silica Plate with UV active F254 indicator. Flash column chromatography was performed with Silica Flash® P60 silica gel (40-63 μm particle size) for most of the crude reaction mixture. Yields are reported as isolated yields of pure compounds. ^{1}H, ^{13}C, and ^{31}P NMR spectra were analysed on 400 MHz NMR spectrometer (Bruker Avance NEO). ^{1}H values are reported in parts per million relative to Me_4Si or corresponding deuterium solvents used as an internal standard. ^{13}C values are reported in parts per million relative to corresponding deuterium solvents used as an internal standards. ^{31}P NMR values are reported in parts per million relative to an external standard of 85% H3PO4. Splitting patterns are designated as follows: s, singlet; d, doublet; dd, doublet of doublets; t, triplet; m, multiplet. High resolution mass spectrometry (HRMS) data were acquired using the electrospray ionization time-of-flight (ESI-TOF) method at the University of California, Irvine Mass Spectrometry Core Facility.

Chemical synthesis, Step 1: To a mixture containing the activated nucleoside monophosphate **1**[33] (240 mg, 0.6087 mmol, 1 equiv) and 1.2 equiv of 1-(2-(pyrenesulfonyl)ethyl)pyrophosphate (343 mg, 0.7305 mmol) was added along with 8 equiv of a premade solution of $ZnCl_2$ (664 mg, 4.8670 mmol) in 3.2 mL anhydrous DMF under a nitrogen atmosphere. The mixture was stirred at room temperature for 5 h and the reaction progress was monitored by HPLC (MeCN/0.05 M TEAB buffer, from 0% to 70% over 42 min). After consumption of the starting material, the product was precipitated by adding the reaction mixture dropwise to stirred diethyl ether (170 mL). The precipitate was collected by centrifuging at 20,000 x g for 5 min at room temperature. The pellet

was resuspended in 30 mL 20% H_2O in MeCN containing 2% *N*,*N*-diisopropylethylamine (DIPEA) and centrifuged at 20,000 x g for 5 min and the supernatant was collected. The above process was repeated two times, and the combined supernatants were evaporated under diminished pressure. The crude was loaded on a silica gel column (packed with 2% H_2O/isopropanol containing 1% DIPEA) by liquid loading (2 mL of 2% H_2O/isopropanol containing 1% DIPEA) with eluents 5% to 10% H_2O/isopropanol containing 1% DIPEA and then 5% H_2O in (isopropanol/MeCN 1:1) containing 1% DIPEA. The fractions containing the product were collected and evaporated under diminished pressure at 30-40°C to afford pure 240 mg (0.2758 mmol) fully protected nucleoside triphosphate **2**. ^{31}P NMR (162 MHz, D_2O) δ -11.93, -12.85 (d, J = 15.6 Hz), -23.10.

Chemical synthesis, Step 2: A solution of fully protected nucleoside triphosphate 240 mg (0.2758 mmol) in 30-33% aqueous NH_4OH was stirred for 18 h at 37°C in a sealed tube. After the reaction, the solvent was evaporated under diminished pressure. The solid was resuspended with MilliQ water (5 mL) and the aqueous solution was washed with DCM (10 mL) and EtOAc (10 mL). The organic portion was discarded, and the aqueous extract was collected and evaporated under diminished pressure. The crude solid was resuspended with 2 mL of RNAse free water, filtrated by 0.22 μm syringe filter and dropwise added to the forty times volume of acetone (80 mL) at room temperature containing $NaClO_4$ (390 mg, 3.1815 mmol, 15 equiv). The resulting suspension was centrifuged at 20,000 x g for 5 min at room temperature. The supernatant was discarded, and the pellet was washed with organic solution (acetone/DCM, 10:1, 2 x 30 mL) to afford the **1a** (85 mg, 0.1577 mmol, 26% two-step yield). ^{31}P NMR (162 MHz, D_2O) δ -8.09, -12.42 (d, J = 19.5 Hz), -22.52 (t, J = 19.5 Hz, 1H).

3.6 Data availability.

The crystallographic data generated in this study have been deposited in the Protein Data Bank under accession codes 9OAT [https://doi.org/10.2210/pdb9OAT/pdb], 9OAU [https://doi.org/10.2210/pdb9OAU/pdb], 9OAV [https://doi.org/10.2210/pdb9OAV/pdb], 9OAW [https://doi.org/10.2210/pdb9OAW/pdb], 9OAX [https://doi.org/10.2210/pdb9OAX/pdb], and 9OAY [https://doi.org/10.2210/pdb9OAY/pdb]. The Protein Data Bank furthermore contains additional structures referenced in this work: 4K8Z [https://doi.org/10.2210/pdb4k8z/pdb], 5OMF [https://doi.org/10.2210/pdb5omf/pdb], 7OMB [https://doi.org/10.2210/pdb7omb/pdb], 8T3X [https://doi.org/10.2210/pdb8t3x/pdb]. The kinetics, fidelity, and thermostability data generated in this study are provided in Appendix B: Supplementary Information. Source data are provided as Source Data files with the original publication.

3.7 Acknowledgements.

We wish to thank members of the Chaput lab for helpful comments and suggestions, and the staff at the Advanced Light Source (ALS), Stanford Synchrotron Radiation Lightsource (SSRL), and the National Synchrotron Light Source II (NSLS) for technical assistance. This work was supported in part by the National Science Foundation (2433788) Division of Chemistry of Life's Processes (CLP) and Genetic Mechanisms (GM), and the University of California.

3.8 Author Contributions.

JC and NC conceived the study and designed the research plan. MH, VM, JL, MJH, RQ, KN, KH, JM, EB, and BB performed the experiments and analyzed the data. JC and NC wrote the manuscript with input from all authors.

3.9 Competing interests.

JC, VM, and the University of California-Irvine have filed a patent application (PCT/US24/11595) on the composition and activity of the 10-92 TNA polymerase. No financial competing interests are declared. The remaining authors declare no competing interests.

3.10 References

1. Turner, N. J. Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol.* 5, 567-573 (2009).

2. Buller, R. *et al.* From nature to industry: Harnessing enzymes for biocatalysis. *Science* 382, eadh8615 (2023).

3. Davis, A. M., Plowright, A. T. & Valeur, E. Directing evolution: the next revolution in drug discovery? *Nat. Rev. Drug Discov.* 16, 681-698 (2017).

4. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat Rev Mol Cell Biol* 10, 866-876 (2009).

5. Tokuriki, N. & Tawfik, D. S. Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* 19, 596-604 (2009).

6. Lo Surdo, P., Walsh, M. A. & Sollazzo, M. A novel ADP- and zinc-binding fold from function-directed in vitro evolution. *Nat Struct Mol Biol* 11, 382-383 (2004).

7. Chao, F. A. *et al.* Structure and dynamics of a primordial catalytic fold generated by in vitro evolution. *Nat. Chem. Biol.* 9, 81-83 (2013).

8. Blomberg, R. *et al.* Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* 503, 418-421 (2013).

9. Suzuki, T. *et al.* Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. *Nat. Chem. Biol.* 13, 1261-1266 (2017).

10. Yang, G., Miton, C. M. & Tokuriki, N. A mechanistic view of enzyme evolution. *Protein Sci.* 29, 1724-1747 (2020).

11. Maola, V. A. *et al.* Directed evolution of a highly efficient TNA polymerase achieved by homologous recombination. *Nature Catalysis* 7, 1173-1185 (2024).

12. Larsen, A. C. *et al.* A general strategy for expanding polymerase function by droplet microfluidics. *Nat. Commun.* 7, 11235 (2016).

13. Vallejo, D., Nikoomanzar, A., Paegel, B. M. & Chaput, J. C. Fluorescence-Activated Droplet Sorting for Single-Cell Directed Evolution. *ACS Synth. Biol.* 8, 1430-1440 (2019).

14. Nikoomanzar, A., Vallejo, D., Yik, E. J. & Chaput, J. C. Programmed allelic mutagenesis of a DNA polymerase with single amino acid resolution. *ACS Synth. Biol.* 9, 1873-1881 (2020).

15. Ravasio, R. *et al.* A minimal scenario for the origin of non-equilibrium order. *ArXiv* (2025).

16. Tawfik, D. S. Accuracy-rate tradeoffs: how do enzymes meet demands of selectivity and catalytic efficiency? *Curr. Opin. Chem. Biol.* 21, 73-80 (2014).

17. Beard, W. A., Shock, D. D., Vande Berg, B. J. & Wilson, S. H. Efficiency of correct nucleotide insertion governs DNA polymerase fidelity. *J. Biol. Chem.* 277, 47393-47398 (2002).

18. Johnson, S. J. & Beese, L. S. Structures of mismatch replication errors observed in a DNA polymerase. *Cell* 116, 803-816 (2004).

19. Johnson, K. A. Role of induced fit in enzyme specificity: a molecular forward/reverse switch. *J. Biol. Chem.* 283, 26297-26301 (2008).

20. Berezhna, S. Y., Gill, J. P., Lamichhane, R. & Millar, D. P. Single-molecule Forster resonance energy transfer reveals an innate fidelity checkpoint in DNA polymerase I. *J. Am. Chem. Soc.* 134, 11261-11268 (2012).

21. Wu, E. Y. & Beese, L. S. The structure of a high fidelity DNA polymerase bound to a mismatched nucleotide reveals an "ajar" intermediate conformation in the nucleotide selection mechanism. *J. Biol. Chem.* 286, 19758-19767 (2011).

22. 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, 183-187 (2012).

23. Tokuriki, N. *et al.* Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nat. Commun.* 3, 1257 (2012).

24. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. U.S.A.* 103, 5869-5874 (2006).

25. Vallejo, D., Nikoomanzar, A. & Chaput, J. C. Directed evolution of custom polymerases using droplet microfluidics. *Methods Enzymol.* 644, 227-253 (2020).

26. Kropp, H. M., Betz, K., Wirth, J., Diederichs, K. & Marx, A. Crystal structures of ternary complexes of archaeal B-family DNA polymerases. *PLoS One* 12, e0188005 (2017).

27. 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, 1058-1062 (2013).

28. Steitz, T. A. DNA polymerases: structural diversity and common mechanisms. *J. Biol. Chem.* 274, 17395-17398 (1999).

29. Li, Y., Korolev, S. & Waksman, G. Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of Thermus aquaticus DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* 17, 7514-7525 (1998).

30. Doublie, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. *Nature* 391, 251-258 (1998).

31. Chim, N., Shi, C., Sau, S. P., Nikoomanzar, A. & Chaput, J. C. Structural basis for TNA synthesis by an engineered TNA polymerase. *Nat. Commun.* 8, 1810 (2017).

32. Li, Q. *et al.* Synthesis and Polymerase Recognition of Threose Nucleic Acid Triphosphates Equipped with Diverse Chemical Functionalities. *J. Am. Chem. Soc.* 143, 17761-17768 (2021).

33. Bala, S. *et al.* Synthesis of 2'-Deoxy-alpha-l-threofuranosyl Nucleoside Triphosphates. *J. Org. Chem.* 83, 8840-8850 (2018).

34. Liao, J.-Y., Bala, S., Ngor, A. K., Yik, E. J. & Chaput, J. C. P(V) Reagents for the Scalable Synthesis of Natural and Modified Nucleoside Triphosphates. *J. Am. Chem. Soc.* 141, 13286-13289 (2019).

35. Gromiha, M. M. & Selvaraj, S. Importance of long-range interactions in protein folding. *Biophys Chem* 77, 49-68 (1999).

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

37. Sau, S. P., Fahmi, N. E., Liao, J.-Y., Bala, S. & Chaput, J. C. A scalable synthesis of α-L-threose nucleic acid monomers. *J. Org. Chem.* 81, 2302-2307 (2016).

38. Nikoomanzar, A., Dunn, M. R. & Chaput, J. C. Engineered polymerases with altered substrate specificity: expression and purification. *Curr. Protoc. Nucleic Acid Chem.* 69, 4.75 (2017).

39. Nikoomanzar, A., Dunn, M. R. & Chaput, J. C. Evaluating the rate and substrate specificity of laboratory evolved XNA polymerases. *Anal. Chem.* 89, 12622-12625 (2017).

40. Medina, E. L. & Chaput, J. C. Measuring XNA polymerase fidelity in a hydrogel particle format. *Nucleic Acids Res.* 53 (2025).

41. Kabsch, W. Xds. *Acta Crystallogr. D Biol. Crystallogr.* 66, 125-132 (2010).

42. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658-674 (2007).

43. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486-501 (2010).

44. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* 75, 861-877 (2019).

45. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* 27, 293-315 (2018).

46. The PYMOL molecular graphics program (DeLano Scientific, San Carlos, CA, USA , 2002).

47. Liebschner, D. *et al.* Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr D Struct Biol* 73, 148-157 (2017).

48. Li, S., Olson, W. K. & Lu, X. J. Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res.* 47, W26-W34 (2019).

49. Needleman, S. B. & Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* 48, 443-453 (1970).

50. Madeira, F. *et al.* The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Res.* 52, W521-W525 (2024).

51. Henikoff, S. & Henikoff, J. G. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U.S.A.* 89, 10915-10919 (1992).

52. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671-675 (2012).

53. Schymkowitz, J. *et al.* The FoldX web server: an online force field. *Nucleic Acids Res.* 33, W382-388 (2005).

54. Mitternacht, S. FreeSASA: An open source C library for solvent accessible surface area calculations. *F1000Res* 5, 189 (2016).

55. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630, 493-500 (2024).

3.11 Figures

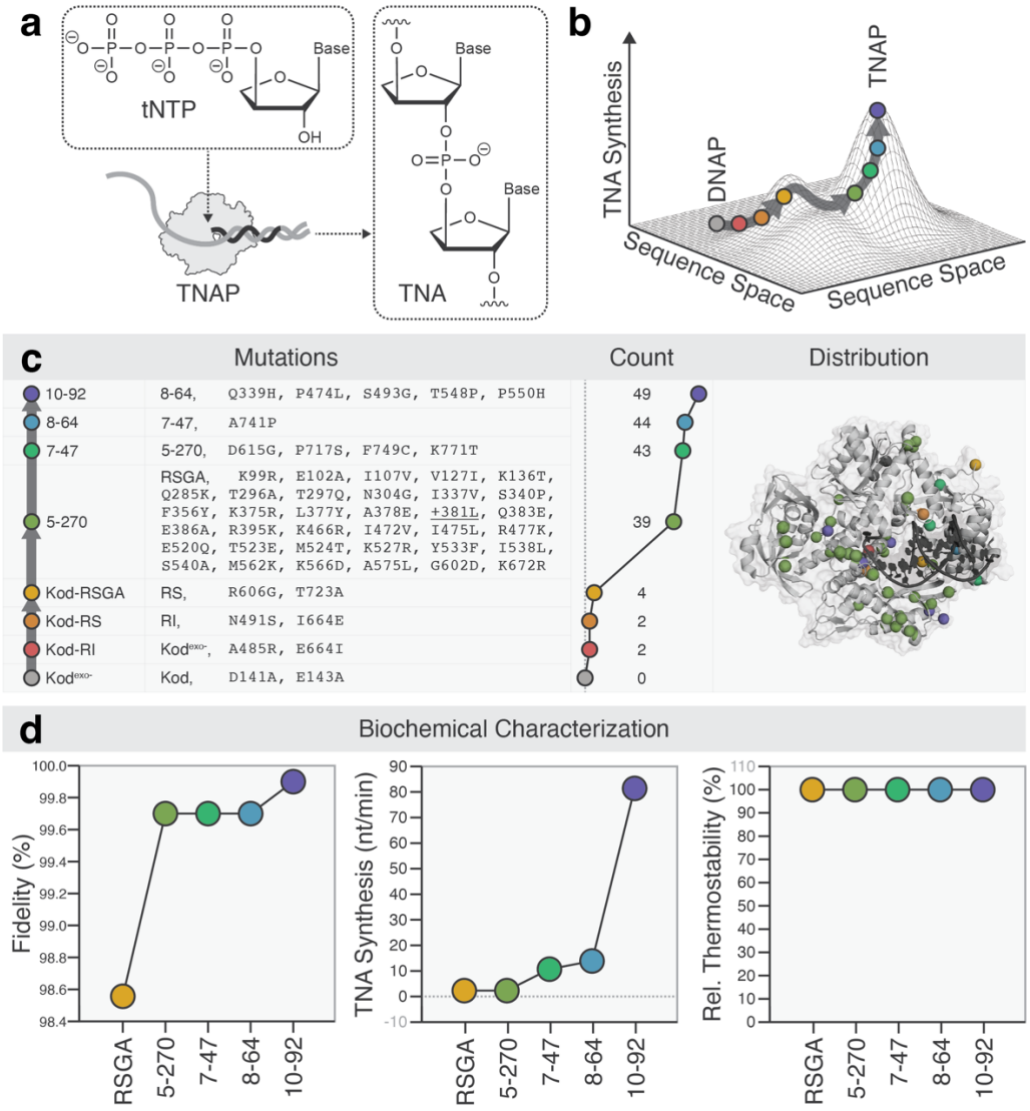


Figure 3.1 Evolution of the 10-92 TNA polymerase.

***a**, Schematic representation of a TNA polymerase (TNAP) copying DNA into TNA by adding tNTPs to a growing primer. **b**, A hypothetical fitness landscape conceptualizing the path taken from a family of natural DNAPs to a highly evolved TNAP. **c**, TNAP generations are color-coded relative to Kod DNAP. Mutations are counted and spatially mapped onto the structure of Kod DNAP (PDB: 5OMF) with the DNA duplex shown in black. **d**, Biochemical analysis of successive generations of TNAPs from Kod-RSGA to 10-92. Quantitative scatter plots display the values observed for catalysis, fidelity, and thermal stability. Source data are provided in Supplementary Figure 6 (fidelity) and additional Source Data files (kinetics and thermostability).*

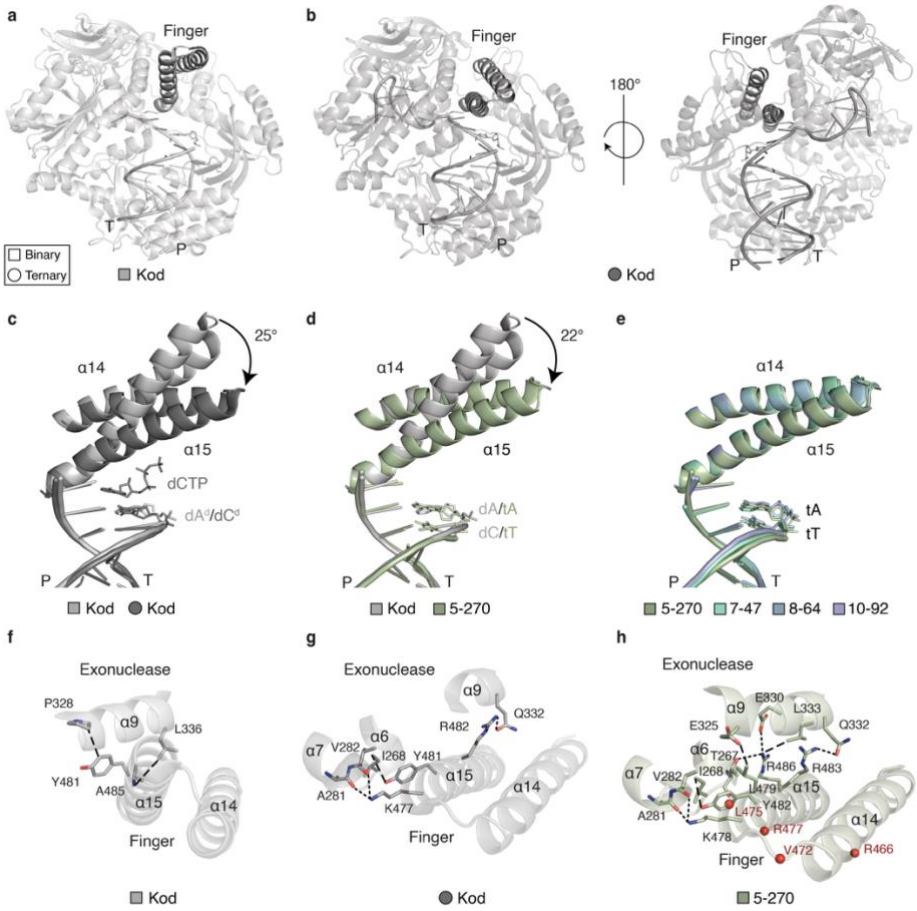


Figure 3.2 The finger subdomain of TNA polymerases

***a-b**, Global architecture of Kod DNAP captured in an open binary complex with dsDNA (A, PDB: 4K8Z) and closed ternary structure in complex with dsDNA and dATP (B, PDB: 7OMB). The finger subdomain is shown in black. **c**, Overlay of the open (light gray) and closed (dark gray) conformations of the finger subdomain found in Kod DNAP. **d**, Overlay of the finger subdomain of Kod DNAP with 5-270, both crystallized as a binary complex. **e**, Overlay of the finger subdomains of 5-270, 7-47, 8-64, and 10-92, each crystallized as a binary complex. **f-h**, Molecular interactions observed in the finger subdomain of Kod binary (f), Kod ternary closed conformation (g), and 5-270 binary (h). Residue labels: natural residues found in Kod DNAP (black); acquired mutations (red spheres, labels). Short dashed lines denote electrostatic interactions, long dashed lines show hydrophobic interactions. TNAP generations are color-matched to Figure 1.*

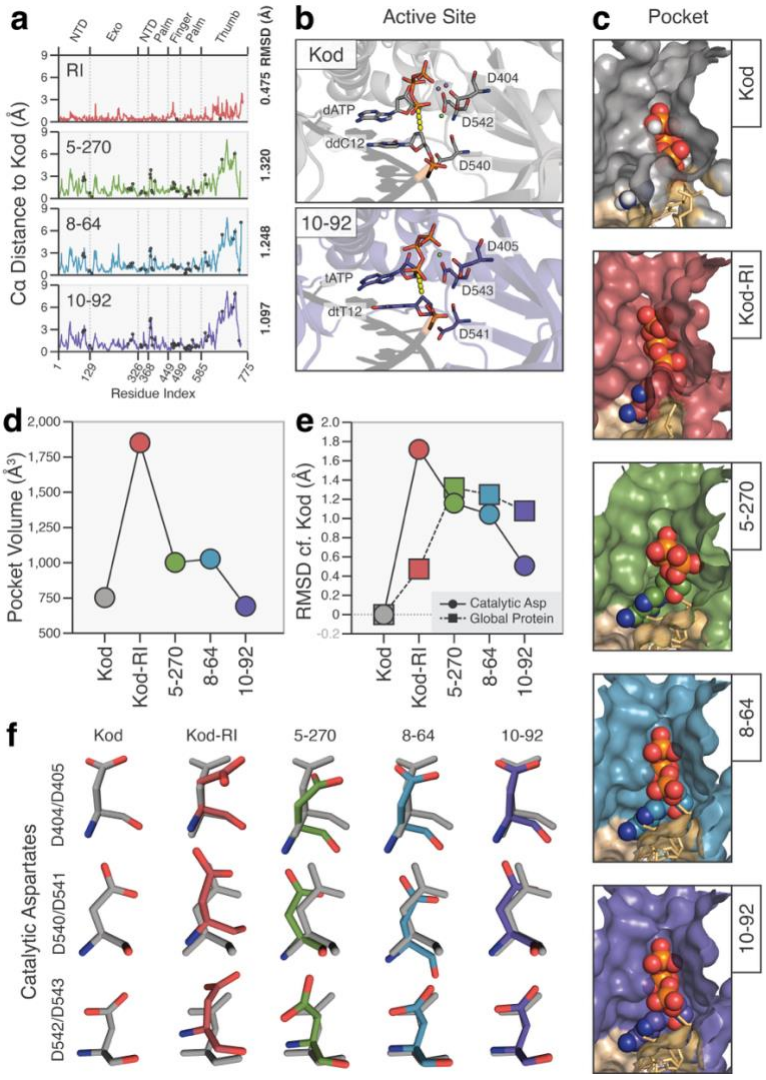


Figure 3.3 Stepwise changes in global and local structural conformation.

***a**, The distance in angstroms from all resolved C-a positions in the ternary structures of TNAP generations relative to their position in Kod DNAP. Black dots indicate the position of the acquired mutations. **b**, Catalytic aspartate triads observed in the active site of Kod (top) and 10-92 (bottom) ternary structures. Dashed yellow line indicates the distance from the primer to the alpha phosphate of the incoming nucleotide. Spheres indicate the position of divalent metal ions. **c**, Shape-filling representation of the active site pocket of progressive generations of TNAPs. Color code: primer (yellow), template (wheat), and triphosphates are colored by element. **d**, Comparison the active site pocket volume (Å^3) for Kod, Kod-RI, 5-270, 8-64, and 10-92. **e**, Root mean square deviation (RMSD) of C-a positions calculated for global (squares) and local (circles) regions of TNAP generations relative to Kod DNAP. Local RMSD values reflect all-atom calculations for the catalytic triad as shown in panel B. **f**, Stepwise optimization of the catalytic aspartate triad across TNAP generations. Overlays show the residue position in the TNAP relative to Kod (gray) via global structural alignment. TNAP generations are color-matched to Figure 1.*

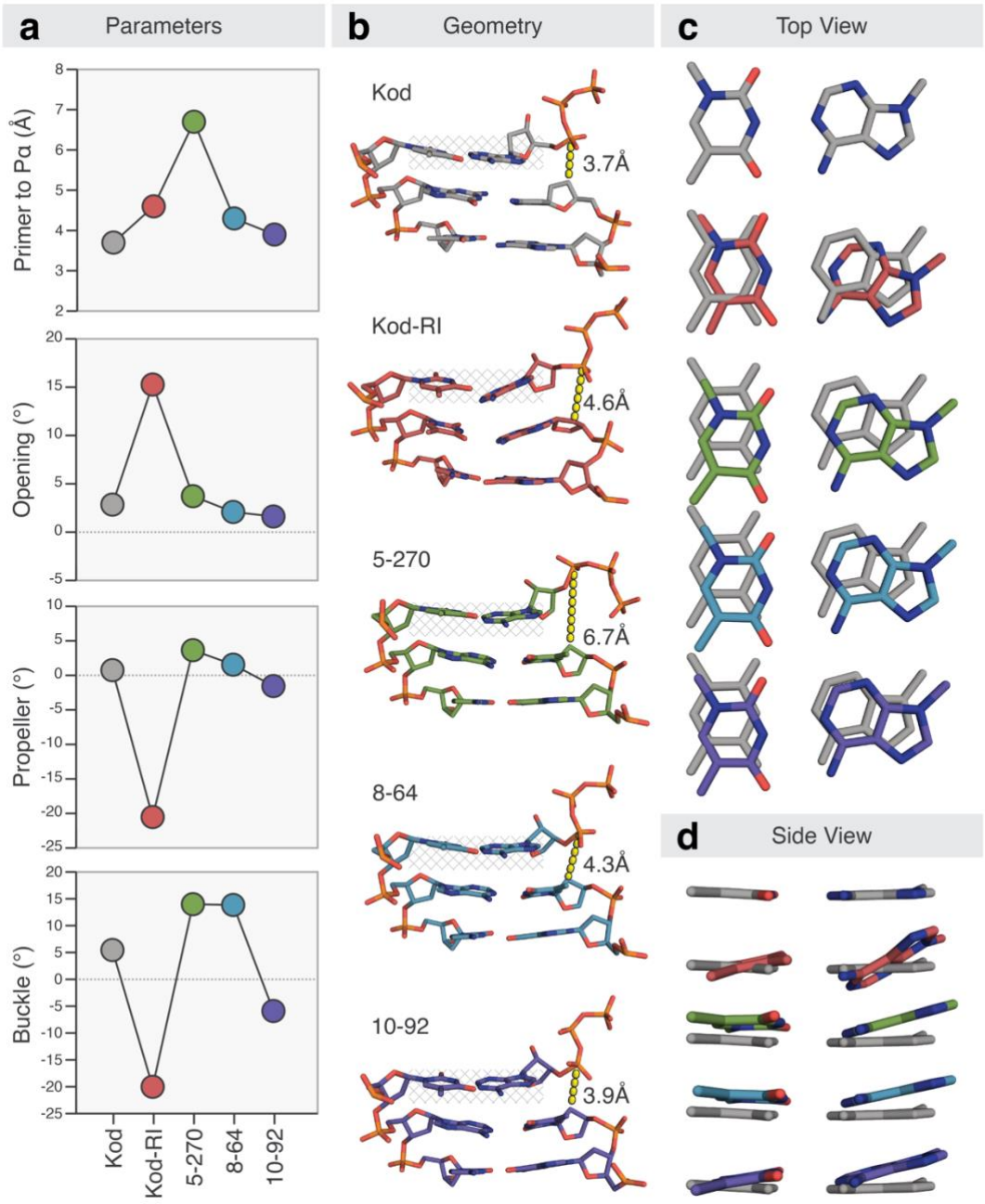


Figure 3.4 Structural optimization of the nascent base pair and reaction center

***a**, Nucleic acid structural parameters plotted across generations of TNAP. Measurements reflect changes in the distance from the primer to the a-phosphate and key conformational rotations observed for the nascent base pair formed between the templating base (dT) and incoming triphosphate (dATP or tATP). **b**, Structural comparisons of the nascent base pair and reaction center observed for Kod DNAP and generations of TNAPs. DNA and TNA are shown as stick models with the distance from the primer to the a-phosphate denoted with a dashed yellow line and numerical value. **c-d**, Top (c) and side (d) views of the geometry of the nascent base pair across generations of TNAPs via global structural alignment to Kod DNAP. TNAP generations are color-matched to Figure 1.*

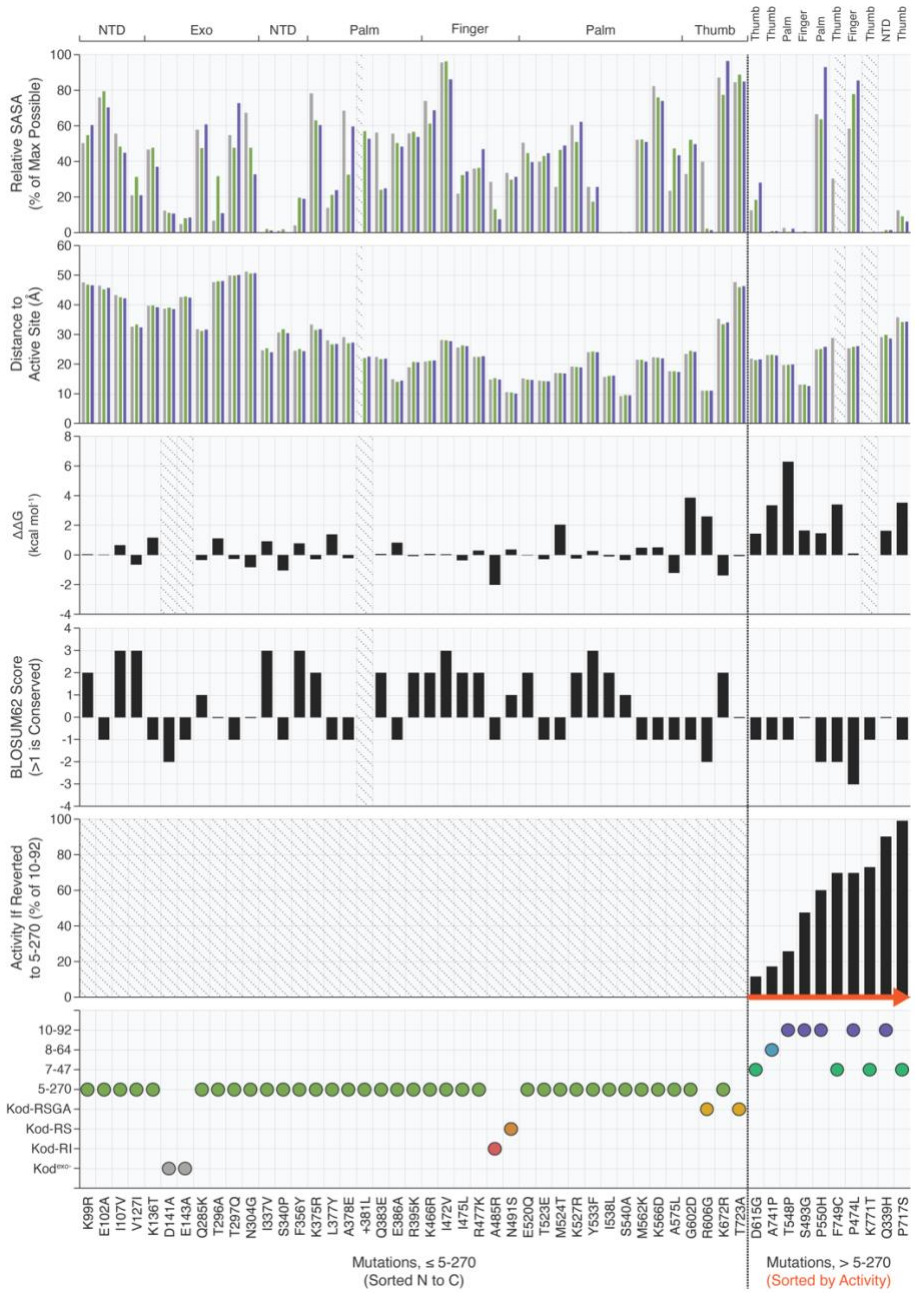


Figure 3.5 Integrated analysis of 51 mutations acquired during the evolution of 10-92

Mutations are ordered along the x-axis, with those observed in 5-270 arranged N→C and those acquired after 5-270 separated by a dashed line and ranked by functional importance. For each site, we report the generation of appearance, the effect of reversion on activity (% of 10-92, extracted from prior data[11]), BLOSUM62 score, predicted stability change (ΔΔG, kcal·mol⁻¹), distance to the active site, and relative solvent accessibility in ternary complexes (Kod^{exo-}, gray; 5-270, green; 10-92, purple). Boxes without data or where metrics are not applicable are marked with a gray diagonal dotted pattern.

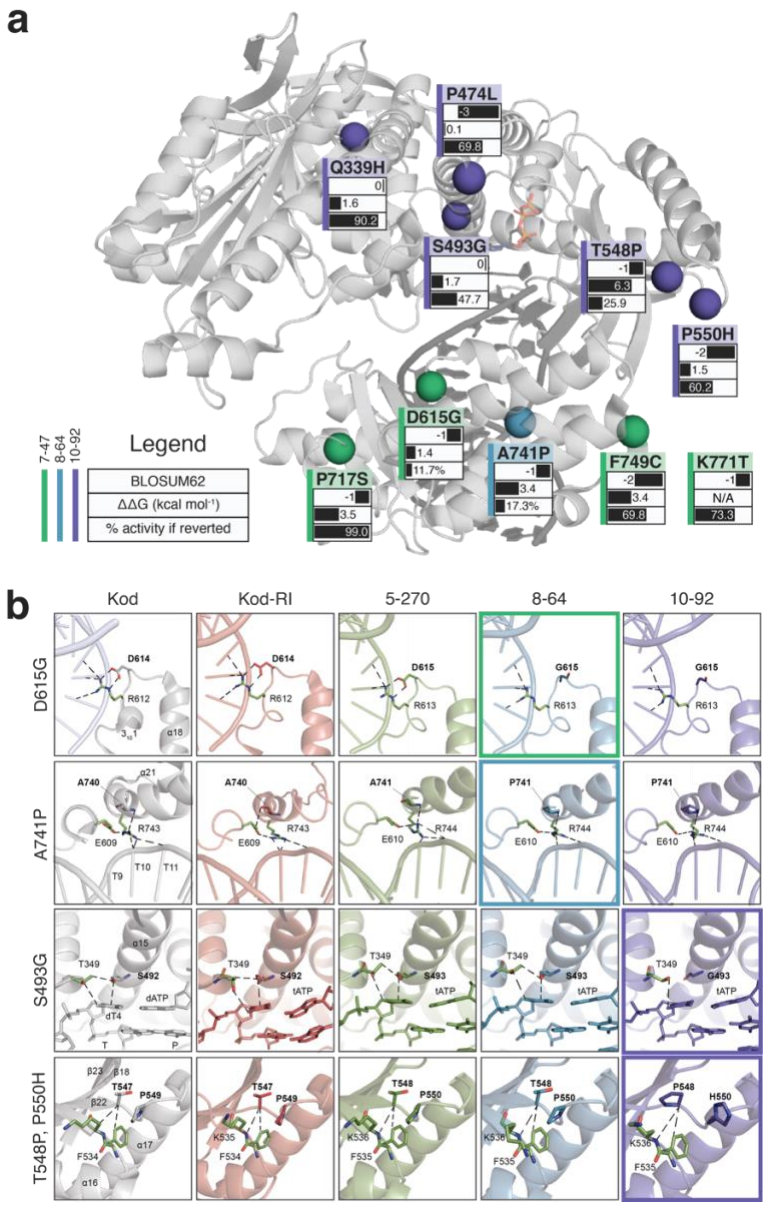


Figure 3.6 Structural analysis of key mutations driving improved catalysis

***a**, Ten mutations acquired during error prone mutagenesis between 5-270 and 10-92 are mapped onto the Kod^{exo-} scaffold (PDB ID 5OMF) by showing the Ca atoms of each mutated residue as spheres. The K771T mutation near the C-terminus was not resolved in the crystal structure. Labels indicate the mutation, its BLOSUM62 score, predicted ΔΔG (kcal mol^{-1}), and the remaining percentage of activity when the mutation is individually reverted. Color coding reflects the generation in which the mutation first appeared (mint green, 7-47; blue, 8-64; purple,10-92). **b**, Panels depict key structural changes throughout the evolutionary trajectory by tracking residues impacting activity across crystallized closed-ternary structures. Dashed lines indicate inter- and intramolecular interactions with other residues in the protein (green) and the primer-template duplex. Boxes with thicker borders denote the first appearance of the mutations.*

CHAPTER 4: Summary and Conclusions

4.1 Summary of Results

Like most XNAs, the enzymatic synthesis of TNA has been a challenge. The unnatural backbone of a TNA polymer disrupts the precise molecular interactions required for nucleotide recognition and incorporation by DNA polymerases. Early attempts to synthesize TNA by engineered polymerases were low-efficiency reactions with sub-optimal fidelity, limiting their practical utility. The aim of this work was to evolve and solve the structure of a highly efficient TNA polymerase that can be used to expand the field of synthetic genetics.

In chapter 1, (published in Maola, V.A., Yik, E.J., Hajjar, M. *et al*, 2024) we identified an optimized TNA polymerase by designing and constructing a homologous recombination library. The starting library derived from four archaeal DNA polymerases, each carrying the RSGA mutations known to enhance TNA substrate recognition[1]. After five rounds of enrichment and five additional rounds of directed evolution with targeted mutagenesis, we identified clone 10-92 as a highly divergent enzyme capable of synthesizing TNA at a rate of ~1 nt/s. To gain structural insight into this enhanced activity, we also solved the crystal structure of the ternary complex. Comparison of the 10-92 structure against its closest natural homolog, Kod DNA polymerase revealed structural remodeling imparted by the evolution process. Major conformational changes were observed within the catalytic domain, and to a lesser extent the N-terminal and exonuclease domains. These data suggest that the evolutionary distance required to convert natural DNA polymerases into highly specialized XNA enzymes may be greater than previously appreciated, as optimization of the active site pocket required substantial topological changes in the global protein architecture. Extending this approach to other genetic systems will help assess the generality of

this strategy and may provide access to valuable of resources that can advance future biotechnology applications.

In chapter 2, (published in Yik, E. J., Maola, V. A., & Chaput, J. C., 2023), we describe in detail the directed evolution strategy used to engineer the 10-92 polymerase presented in chapter 1. We emphasize library design and construction through homologous recombination. The workflow details how parent scaffolds are identified, acquired, and cloned into a preferred expression vector, then we describe the 3-step process of homologous recombination. We also provide a detailed methodology for expression of the polymerase library in *E. coli* and lightly describe the preparation of single cell encapsulation into uniform water-in-oil droplets, and recovery of the DNA from the droplet following droplet-based optical polymerase sorting (DrOPS). The genes from the active library are PCR amplified, purified, and ligated to the expression vector for colony picking and activity screening. The activity screen reveals variants with desired functional activities, whose genetic information can inform the basis of the next round of directed evolution. To further diversify the sequence landscape, a protocol for mutagenic PCR is also included. Together, these methods establish a general and scalable platform for exploring polymerase sequence-function relationships and enable the screening of highly diverse enzyme libraries, without relying on next-generation sequencing. This protocol is sufficiently general such that it can be applied to a wide range of DNA modifying enzymes.

In chapter 3, (published in Hajjar, M., et al, 2025), we present the structural and biochemical analysis of how directed evolution reshapes enzyme architecture to enable efficient synthesis of TNA. By mapping the evolutionary trajectory of a highly efficient TNA polymerase isolated from a molecular evolution study involving recombination and directed evolution[2], this work reveals that fidelity and catalytic efficiency, properties often assumed to be intrinsically

linked, emerge through distinct molecular pathways[3-5]. We observe that fidelity improves early in the evolutionary process through restoration of Watson–Crick base pair geometry and stabilization of the nascent base pair, whereas substantial gains in catalytic activity and altered substrate specificity arise only in later generations, coinciding with the extensive remodeling of the active site. These findings suggest that ground-state and transition-state discrimination can evolve independently, offering a modular mechanism by which enzymes acquire new function. A defining feature of the evolved polymerase lineage is the adoption of a preorganized, closed finger conformation in the absence of a bound nucleotide. This catalytically competent architecture is stabilized by a network of interactions that shift the conformational equilibrium away from the open conformation commonly associated with the binary complex of replicative DNA polymerases. This structural transition may represent a critical inflection point in the evolutionary process, enabling the enzyme to escape the kinetic consequences associated with accommodating the unnatural TNA substrate. Across the evolutionary trajectory, successive variants exhibited progressively higher TNA synthesis rates, while DNA synthesis activity remained predominant. A marked inversion in substrate specificity emerged only in variant 10-92, which likely stems from its remodeled active site, whose reduced volume reflects structural adaptation to the smaller size of the TNA triphosphate. This inversion in specificity represents a turning point of functional divergence and illustrates how exploration of sequence space (the ensemble of all possible protein sequences of a given length) can yield enzymes that are not only structurally robust, but also precisely tuned for novel chemical tasks.

4.2 Future Perspectives and Ending Notes

Taken together, our findings illustrate an example of how directed evolution can orchestrate long-range structural rearrangements and local active site refinement to generate an enzyme with

tailor-made activity. The separation of fidelity and catalysis observed in this work offers a potential new framework for understanding enzyme adaptation and informs future efforts to design polymerases for applications in synthetic biology and medicine. More broadly, this work underscores the importance of combining structural biology with evolutionary trajectories to illuminate mechanisms of molecular innovation that remain opaque to current predictive models.

While the work in this dissertation demonstrates a meaningful improvement for the synthesis of XNA polymers, there remains a gap for an XNA replication system that functions independently of DNA, since current systems rely on DNA as an informational intermediate (i.e., XNA→DNA→XNA). Achieving a fully autonomous XNA replication system (i.e., XNA→XNA) is a critical conceptual and technical milestone that would greatly enhance our understanding of polymerase activity and produce a necessary tool for the field of synthetic genetics. The selection platform (DrOPS) created by the Chaput group has immense potential to accomplish not only this, but also to generate a suite of XNA enzymes for the synthetic genetic toolbox, including enzymes that can perform ligation and restriction digestion.

An additional opportunity for further development lies in streamlining lysate screening to more efficiently identify interesting variants. Currently, this labor-intensive workflow relies on hand-picking colonies and expressing variants individually in 4-mL culture tubes. Each variant is then assayed by primer extension and analyzed through end-point detection by denaturing polyacrylamide electrophoresis (PAGE). The current workflow is resource-intensive, requiring abundant access to consumables, as well as extended processing time due to many necessary gel runs. This workflow could be much improved by the transition to high-density expression system such as in deep-well plates. Additionally, implementing a pre-screening step prior to urea-PAGE analysis would allow for expressed variants to be evaluated in a microplate-format primer

extension assay. This rapid, fluorescence-based readout would allow users to identify variants meeting predefined activity thresholds, that then could then be subject to detailed resolution by denaturing PAGE to assess details like extension patterns. This two-tiered strategy may increase screening throughput by reserving gel-based analysis for most promising candidates.

4.3 References

1. Nikoomanzar, A., Vallejo, D., Yik, E. J. & Chaput, J. C. Programmed allelic mutagenesis of a DNA polymerase with single amino acid resolution. *ACS Synthic Biology* 9, 1873-1881, (2020).

2. Maola, V. A. *et al.* Directed evolution of a highly efficient TNA polymerase achieved by homologous recombination. *Nature Catalysis* 7, 1173-1185 (2024).

3. Ravasio, R. *et al.* A minimal scenario for the origin of non-equilibrium order. *ArXiv* (2025).

4. Tawfik, D. S. Accuracy-rate tradeoffs: how do enzymes meet demands of selectivity and catalytic efficiency? *Curr. Opin. Chem. Biol.* 21, 73-80 (2014).

5. Beard, W. A., Shock, D. D., Vande Berg, B. J. & Wilson, S. H. Efficiency of correct nucleotide insertion governs DNA polymerase fidelity. *J. Biol. Chem.* 277, 47393-47398 (2002).

APPENDIX A: Supplementary Data of Chapter 1

Supplementary Methods

Reagents

DNA oligonucleotides and gBlocks were purchased from Integrated DNA Technologies (IDT, Coralville, Iowa). TNA triphosphates were obtained by chemical synthesis as described previously [1,2]. ThermoPol buffer, Q5 DNA polymerase, NEB 5-alpha competent *E. coli*, Taq DNA polymerase, *NdeI* and *NotI* restriction enzymes, DpnI, and Gibson assembly were purchased from New England Biolabs (Ipswich, MA). XL1-Blue competent cells were purchased from Agilent (Santa Clara, CA). DNA clean up kit and ZymoPURE plasmid midiprep kit were purchased from Zymo (Irvine, CA). Express DNA Miniprep kit and Plasmid midiprep kit were purchased from Biomiga (San Diego, CA). Chemical reagents including dNTPs, manganese chloride and ammonium persulfate (APS) were purchased from Sigma Aldrich (St. Louis, Missouri). TOPO TA cloning kit, ethylenediaminetetraacetic acid (EDTA) were purchased from Thermofisher Scientific (Waltham, Massachusetts). SequalGel UreaGel 29:1 Denaturing Gel System was purchased from National Diagnostics (Atlanta, GA). Tetramethyl-ethylenediamine (TEMED) was purchased from Bio-Rad (Hercules, California). Heparin affinity columns were purchased from GE Healthcare (Little Chalfont, United Kingdom). Polydimethylsiloxane (PDMS) base and curing agent were purchased from Dow Corning (Midland, MI). SU-8 2025 photoresist was purchased from Fisher Scientific (Hampton, NH). Fluorinated oil HFE-7500 was purchased from 3M Novec (St Paul, MN), and Pico-Surf™ 1 surfactant, Pico-Glide™ 1, and Pico-Break™ 1 were all purchased from Dolomite Microfluidics (UK). EvaGreen® dye was purchased from Biotium (Fremont, CA). Clear V-bottom 96-well plates were purchased from Greiner (Monroe, NC). Plastic 1.5 mL micro-centrifuge tubes were

purchased from Sigma-Aldrich (St. Louis, MO). Tygon tubing (EW-06419-01) was purchased from Cole-Parmer (Vernon Hills, IL). The SMC ITV0011-2UMS digital pressure regulator was purchased from Automation Distribution (Hatfield, PA). Microfluidic reagents HFE 7500, and Pico-Surf were purchased from 3M Novec (North Cordova, IL) and Sphere Fluidics, (Cambridge, UK), respectably. A 552 nm laser was purchased from Coherent OBIS LS (Santa Clara, CA). An apochromatic objective was purchased from Motic (Kowloon Bay, Hong Kong). A field-gated programmable array (FPGA, USB-7856R) was purchased from National Instruments (Austin, TX). The high-voltage amplifier was purchased from Trek Inc 2210 (Lockport, NY). Crystallization screens were purchased from Hampton Research (Aliso Viejo, CA), NeXtal Biosciences (Holland, OH), and Molecular Dimensions (Holland, OH). The Mosquito crystallization robot was purchased from SPT LabTech (Covina, CA). pET21 plasmid was purchased from Novagen Technology (Glendale, CA). The final crystallography reagents include HEPES, purchased from Sigma Aldrich (St. Louis, MO), polyethylene glycol 6000 and lithium chloride both purchased from Hampton research (Aliso Viejo, CA). All Sanger sequencing was performed by Genewiz (San Diego, CA)

Design and construction of recombination library

A custom homologous recombination library was designed based on the sequence alignment of four homologous hyperthermophilic family-B DNA polymerases isolated from archaeal species *Thermococcus kodakaraensis* (Kod, GenBank: KP682508.1), *Thermococcus gorgonarius* (Tgo, GenBank: KP682507.1), *Pyrococcus* species GB-D Deep Vent (DV, GenBank: KP682509.1), and Thermococcus sp. 9°N (9°N, GenBank: KP682506.1). Each gene was modified to contain the following amino acid mutations: E141A, D143A, A485R, N491S, R606G, and T723A. The

gene encoding each enzyme was translated *in silico*, aligned, and sequence verified. Following in silico validation, 11 regions were identified as segments (termed cassettes) for subsequent homologous recombination by DNA shuffling. The boundaries between each cassette (7-10 amino acids) were modified by codon swapping to ensure perfect homology between each of the four parent genes. This was required for homologous recombination by overlap PCR. Finally, the genes were modified to include overlap regions with the backbone vector, as required for Gibson assembly. The genes were ordered as gBlocks (Supplementary Table 1).

The vector backbone required for Gibson assembly was generated with 1x Q5 Buffer, 0.5 μM Vector-Fwd (Supplementary Table 2), 0.5 μM Vector-Rvs, 5 ng of pGDR11 9°N-WT, 0.4 mM dNTPs, 0.5 μL Q5 DNA polymerase (0.02 U/μL). PCR was performed as followed: denature at 95°C for 2.5 min followed by 30 cycles: denature at 95°C for 30 sec, anneal at 68°C for 45 sec, extended at 72°C for 2.5 min and final step at 72°C for 1 min. PCR reaction (2 μL) was visualized on a 1% ethidium-bromide agarose gel and run at 120 V for 45 min to confirm correct amplicon size. Zymo PCR cleanup was performed as followed: 100 μL of PCR reaction was mixed with 300 μL of DNA binding buffer and loaded onto Zymo IC column. The column was spun at 14,000 rcf for 30 sec and washed with 500 μL x2 of Zymo wash buffer at 14,000 rcf for 30 sec. Column was further dried at 14,000 rcf for 3 min. DNA was eluted from column with 25 μL x2 of nuclease free water and UV quantified.

gBlocks were reconstituted at the recommended volume with nuclease free water, mixed with 100 ng (1 μL) of the vector backbone and assembled through NEB Gibson assembly at 55°C for 1 h. Transformation of Gibson assembly was performed as followed: 1 μL of Gibson reaction

was transformed into 50 µL NEB 5-alpha competent *E. coli* and incubated on ice for 30 min. Prior to heat shock, 28 µM of BME was added and mixed with cells. Immediately, cells were heat shocked at 42°C for 30 sec and immediately placed back on ice for 2 min. Cells were recovered with 1 mL of Super Optimal broth (SOC) and placed into a shaking incubator at 225 rpm for 1 h at 37°C. Recovered cells (250 µL) were plated onto 50 ng/mL carbenicillin agar plates. Plates were incubated overnight at 37°C. Individual colonies were picked and colony PCR was performed as followed: denature at 95°C for 2 min followed by 30 cycles: denature at 95°C for 15 sec, anneal at 58°C for 15 sec, extend at 72°C for 2.5 min final step at 72°C for 1 min with Fragment1-Fwd (Supplementary Table 2) and Fragment11-Rvs. PCR reactions were analyzed by 1% ethidium bromide agarose gels to verify the correct amplicon size. Sanger sequencing was performed to verify the sequence composition of each gene. Site directed mutagenesis was performed with Q5 to repair improper nucleotide mutations. Post-repair PCR reactions were analyzed by 1% ethidium bromide agarose gel to verify full vector amplicon. PCR reaction (1 µL) was treated with NEB KLD, transformed, and plated on 50 ng/mL carbenicillin agar plates. Colony PCR was performed, and gene-containing colonies were sequenced by Sanger sequencing.

A total of 44 fragment cassettes were individually generated by PCR using primers (Supplementary Table 2) that were specific to internal regions of each gene. PCR reactions contained a final concentration of 1x Q5 buffer, 0.5 µM Fwd primer, 0.5 µM Rvs primer, 0.4 mM dNTPs, and 0.5 µL Q5 DNA polymerase (0.02U/µL) in a final volume of 50 µL. PCR cycles were performed as followed: denature at 95°C for 2 min followed by 30 cycles: denature at 95°C for 30 sec, anneal at 68.6°C for 45 sec, extend at 72°C for 2.5 min and a final step at

72°C for 1 min. Each fragment was mixed with Zymo DNA binding buffer and loaded onto Zymo IC column for purification. DNA was washed 4x 500 µL DNA wash buffer (10 mM Tris-HCl, pH 7.4, 80% ethanol). DNA was eluted from the column with 2x 25 µL of nuclease free water and UV quantified.

DNA shuffling of the 44 gene fragments proceeded by PCR. First, each version of fragment 1 (Kod, Tgo, 9N, DV) were mixed (25 ng each) to produce 100 ng of DNA in a 10 µL volume, generating a 10 ng/µL working stock for assembly. This process was repeated for each gene fragment region. Mass ratios between the fragment sizes and the entire gene were used to calculate the amount required for the overlap PCR. For a gene size of 2,631 base pairs, we used a set concentration of 263.1 ng for a 50 µL overlap PCR assembly. For example, fragment 1 cassette is 179 base pairs out of 2,631 base pairs, which is 6.8% of the entire assembled gene. Therefore, 1.79 µL of DNA was added from the 10 ng/µL working stock for fragment 1. The process was repeated for the remaining fragments. The remaining components added for the overlap PCR were 1x Q5 buffer, 0.4 mM dNTPs, 0.5 µL of Q5 DNA polymerase (0.02U/µL) in the absence of primers. The overlap PCR conditions were performed as followed: denature at 94°C for 30 sec, followed by 9 cycles of denature at 94°C for 15 sec, extend at 72°C for 1.5 min with a decrease of 0.5°C per cycle. Then 5 cycles of denature at 94°C for 15 sec, anneal for 67.5°C for 15 sec with a decrease of 0.5°C per cycle, extend at 72°C for 1.5 min, with a final polishing step of 72°C for 2 min.

To amplify full length genes, 4 µL of the overlap PCR assembly mixture was combined with 1x Q5 buffer, 0.4 mM dNTPs, 0.5 µL of Q5 DNA polymerase and the presence of 0.5 µM

Fragment1-Fwd and 0.5 μM Fragment11-Rvs primers. The PCR conditions for 40 cycles were as followed: denature at 94°C for 30 sec, followed by 17 cycles of denature at 94°C for 15 sec, anneal at 67°C for 15 sec with a decrease of 0.5°C per cycle, extend at 72°C for 1.5 min. Then 23 cycles of denature at 94°C for 15 sec, anneal at 59°C for 15 sec, extend at 72°C for 1.5 min, with a final polishing step of 72°C for 2 min. After PCR, samples (2 μL) were analyzed on a 1% ethidium-bromide agarose gel at 120 V for 45 min.

Once full-length amplicons were generated, the PCR product was purified by 1% ethidium-bromide agarose gel at 70 V for 120 min. The corresponding band was excised and dissolved with Zymo agarose dissolving buffer (100 mg of agarose to 300 μL Zymo agarose dissolving buffer) by incubating at 37°C for 1 h. Dissolved solution was loaded into a Zymo IC column and spun at 14,000 rcf for 30 sec. DNA wash buffer (500 μL) was added to the column, mixed by inversion, and removed. The interior wash was repeated for a total of 2x, followed by exterior washes to the outside of the column by rinsing with 1 mLx4 of DNA wash buffer. Following exterior column washes, the column was placed into a fresh collection tube, and 500 μLx4 of DNA wash buffer was spun through the column at 14,000 rcf for 30 sec. The column was further dried at 14,000 rcf for 3 min. DNA was eluted with 10 μLx2 of nuclease free water and UV quantified.

Gibson assembly was utilized to construct the expression vector using the newly generated homologous recombination library (100 ng) and the previously generated vector backbone (100 ng). Transformation was subsequently performed as described above. Colony PCR and Sanger sequencing was performed to validate the generation of the recombined library and to monitor parental plasmid contamination. Upon validation, the Gibson assembled material was scaled for

a total of 5 transformations into NEB 5-alpha competent *E. coli*. Colonies were scrapped from agar plates and purified using the Plasmid Midiprep II kit following the manufacturer's recommended instructions. Plasmid DNA was eluted in nuclease free water, UV quantified and stored for future transformation into XL1-blue *E. coli* for DrOPS sorting and polymerase expression.

Preparing *E. coli* for encapsulation in droplets

A detailed description of photolithography and microfluidic device fabrication for DrOPS sorting has been previously described [3,4]. Cells expressing the recombinant polymerase library were prepared by transforming 300 ng of the plasmid library into 20 µL of XL1-blue *E. coli*. DNA was incubated with cells for 30 min on ice, heat shocked at 42°C for 30 sec and immediately placed on ice for 2 min. SOC (1 mL) was added to the cells and placed into a shaker at 37°C with shaking at 225 rpm for 1 h. Recovered cells were then used to inoculate 50 mL of LB-carbenicillin (50 µg/mL) liquid medium in a 250 mL baffled flask. The liquid culture was grown to confluency overnight at 37°C with shaking at 225 rpm. Overnight (500 µL, 1:100 v/v) was used to inoculate 50 mL of LB-carbenicillin (50 µg/mL) liquid medium in a 250 mL baffled flask and grown at 37°C with shaking at 225 rpm. At an OD600 of ~0.6 au, the culture was cooled to 25°C, induced with 1 mM IPTG, and incubated overnight at 25°C with shaking at 225 rpm. After expression, 1 mL of cell culture was collected and centrifuged for 5 min at 3,220 rcf, and the supernatant was discarded. The cells were washed with 1 mL of commercial 1x ThermoPol buffer (20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1% TritonX-100, pH 8.8) for a total of four washes (4 mL). The bacterial pellet was suspended in 2 mL of 1x ThermoPol buffer and the absorbance was measured at 600 nm. Cells were diluted to an OD of 0.05 to enable encapsulation at occupancies of 0.1 cells per droplet according to a Poisson

distribution. Just prior to emulsification, the cells were mixed with the reagents for the polymerase activity assay (PAA). The polymerase activity assay consists of 1 μM of a self-priming hairpin template labeled with Cy3 at the 5' end (30merHP.V2, Supplementary Table 2), 2 μM of a 3' end labeled Iowa Black quencher sequence (QP08.Iowa, Supplementary Table 2), and 100 μM of TNA triphosphates (tNTPs) in 1x ThermoPol buffer.

All aqueous and oil solutions were sealed as individual 1.5 mL plastic micro-centrifuge tubes and controlled via pressure driven flow with custom LabVIEW software (National Instruments). Two lengths of Tygon tubing were inserted through holes drilled into the caps of the micro-centrifuge tubes and glued into place to create an airtight seal. One length of tubing remained in the pressure headspace above the reagent and was connected at the other end to a SMC ITV0011-2UMS digital pressure regulator). Another length of tubing was submerged in the reagent solution with the other end connected to the appropriate inlet of the microfluidic device. By manually applying a positive pressure head to the reagent vial via the SMC digital regulator, fluid was driven through the channels of the microfluidic device. A length of Tygon tubing was also inserted in the outlet and placed in a micro-centrifuge tube for droplet collection.

All emulsions were produced using custom PDMS chips [3,4]. Single emulsions were formed utilizing a flow focusing geometry. The aqueous phase containing the polymerase activity assay and *E. coli* cells was sheared by a continuous phase consisting of a low-viscosity fluorinated oil (HFE-7500) containing 1% (w/w) Pico-Surf surfactant. Pressures were maintained to achieve droplet diameters of ~20-22 μm and production rates of 30-35 kHz, allowing 110-125 x 10^6 droplets to be produced every hour. Single emulsions were collected under a layer of mineral oil

in 1.5 mL plastic micro-centrifuge tubes and incubated for 5 mins at 95°C to lyse the cells, followed by incubation at 55°C for various times.

FADS sorting of single emulsion droplets

A detailed schematic and operation of fluorescence activated droplet sorter (FADS) instrumentation was previously described [3,4]. In brief, following incubation, droplets were injected into a FADS device capable of droplet sorting. Incident light from a 552 nm laser was focused through a 20x plan apochromatic objective (Motic, Hong Kong) where droplets pass in single file. Emitted light was led through a 562 nm Quad Band Dichroic into an optical train through a series of long-pass dichroics to a photomultiplier tube (PMT). The sample was illuminated with blue light to not overlap with the spectral properties of Cy3 and was imaged with a high-speed camera at 35,000 frames per second (fps). The digital signals generated by the PMT were analyzed by a field-gated programmable array (FPGA) controlled with custom LabView software. Droplets falling within a user-defined threshold triggered the FPGA to send a square-wave pulse (50 kHz, 50% duty cycle, 60 μs), amplified to 600 V by a high-voltage amplifier, to the salt-water electrode (4 M NaCl) of the sorting chip. The resulting nonuniform electric field generated a dielectrophoretic (DEP) force that polarized and deflected the droplet into a collection channel.

Recovery of sorted DNA and library regeneration

Plasmid DNA was recovered from the population of positively sorted droplets present in the collection tube by extraction with Pico-Break following the manufacturers recommended protocol. DNA samples were recovered from sorted emulsions by extraction with 2 volumes of

Pico-Break, which contains 1H,1H,2H,2H-perfluorooctanol (PFO). After addition of Pico-Break 1, the samples were vortexed and centrifuged (60 sec, 2,000 xg) to attain phase separation. The aqueous layer (top) containing the plasmid DNA was recovered. The bottom layer was extracted a second time with 1 volume of nuclease free water to improve recovery yields. The combined aqueous layers containing the plasmid DNA were concentrated using a Zymo IC column, washed with 500 µLx2 DNA wash buffer and eluted with nuclease free water (20 µl). The gene was amplified in 1x Q5 buffer, 0.5 µM fragment1-Fwd (Supplementary Table 2), 0.5 µM fragment11-Rvs 0.4 mM dNTPs, 4 µL of recovered DNA, and 0.5 µL Q5 DNA polymerase (0.02 U/µL) in a total volume of 50 µL. PCR amplification was performed as followed: denature at 95°C for 2.5 min followed by 30 cycles: denature at 95°C for 30 sec, anneal at 70°C for 45 sec, extend at 72°C for 2.5 min followed by a polishing step of 72°C for 1 min. The PCR reaction (2 µL) was run on a 1% ethidium-bromide agarose gel for amplicon validation.

Upon validation of amplification, samples were pulled and mixed with 3x Zymo DNA binding buffer (i.e., 100 µL PCR reaction to 300 µL Zymo DNA binding buffer) and added to a Zymo IC column. DNA was cleaned with 500 µLx4 DNA wash buffer at 14,000 rcf for 30 sec. The Zymo IC column was further dried at 14,000 rcf for 3 min prior to elution. DNA was eluted from column with 20 µLx2 of nuclease free water at 14,000 rcf for 1 min. The addition of 5 µL NEB rCutSmart buffer and 5 µL of NEB DpnI was mixed with DNA to digest methylated and hemi-methylated parent plasmid for 4 h at 37°C followed by heat inactivation of the enzyme at 80°C for 20 min. Post DpnI treatment, 15 µL of NEB 6x Blue Gel loading dye was mixed with the DpnI treated sample and loaded into a 1% ethdium-bromide agarose gel run at 70 volts for 120 min to separate gene amplicon and digested parental plasmid.

The gene amplicon band was excised and dissolved in Zymo agarose dissolve buffer at 37°C for a minimum of 1 h. The dissolved gel solution was added to a Zymo IC column and spun at 14,000 rcf for 30 sec. The columns interior and exterior were rinsed with DNA wash buffer as previously mentioned. Columns were washed with addition 4x300 µL of DNA wash buffer. Samples were dried at 14,000 rcf for 3 min prior to the addition of 20 µL of nuclease free water and spun at 14,000 rcf for 1 min. DNA was UV quantified on a NanoDrop instrument.

For the Gibson assembly, equimass quantities of each (100 ng) vector and insert DNA are used following the manufacturer's recommended protocol with the exception of an extended incubation time (1 h). Post assembly, reaction mixture was mixed with 3x Zymo DNA binding buffer, loaded into Zymo IC column and spun at 14,000 rcf for 30 sec. The Zymo IC colum wash with 4x500 µL DNA washed buffer at 14,000 rcf for 30 sec. Prior to elution the Zymo IC column was further dried at 14,000 rcf for 3 min. DNA was eluted with 20 µL of nuclease free water at 14,000 rcf for 1 min. A transformation was performed as previously mentioned using 1 µL of Gibson assembled material.

Construction of mutagenic library

Mutagenic PCR was performed using Kod-WT (exo-) DNA polymerase, which harbors two exonuclease silencing mutations (D141A, E143A) and is preferred for blunt ended amplification. The polymerase was expressed and purified from *E. coli* as described previously [5]. The activity of Kod-WT (exo-) polymerase was empirically determined by serial diluting the enzyme in PCR reactions, from which a resulting 0.013 µM concentration of polymerase was determined suitable for PCR. The mutagenic PCR was performed with biased dNTP ratios, and supplemented $MnCl_2$

and $MgCl_2$ as described previously [6,7]. The PCR reaction was performed with the following final concentration: 1x ThermoPol, 1 μM forward primer,1 μM reverse primer, 1 mM TTP, 1 mM dCTP, 0.2 mM dGTP, 0.2 mM dATP, 5.5 mM $MgCl_2$, 0.013 μM Kod-WT (exo-) polymerase, 100 ng of 5-270 plasmid and 2 fold serial dilution of $MnCl_2$ from 500-31.25 μM. PCR cycles were performed as followed: denature at 95°C for 2 min followed by 30 cycles of denature at 95°C for 1 min, anneal at 60°C for 1 min, extend at 72°C for 2.5 min, and followed by agarose gel analysis (Image Lab 6.1, Bio-Rad). Successful PCR reactions are consolidated for a PCR cleanup, DpnI treatment, and agarose gel purification. The purified DNA and the corresponding vector backbone were combined in a Gibson assembly reaction, column purified, and transformed as described above. Single colonies were picked, grown, miniprepped and sequenced by Sanger sequencing. Mutagenic PCR can be performed with any combination of forward and reverse primer set as examples: Fragment9-Fwd and Fragment11-Rvs for the thumb subdomain and Fragment6-Fwd and Fragment8-Rvs for the palm and finger subdomains.

Colony picking and polymerase activity screen

Protocol for lysate activity screen was performed as described [8]. In brief, Gibson assembly material from post regeneration were transformed into XL1 Blue *E. coli*, recovered and plated onto agar plates containing 50 ng/μL carbenicillin. Single colonies were picked and placed into 4 mL of LB media containing 50 ng/μL carbenicillin and grown in a shaking incubator at 225 rpm, overnight at 37°C. Overnight culture (40 μL) was inoculated into fresh 4 mL LB media containing 50 ng/μL carbenicillin and grown to an OD_{600} ~0.6. Cells were induced with 1 mM IPTG and expressed at 25°C, 225 rpm for 18-20 h. Cells were harvested, centrifuged (1,500 x g, 25°C, 10 min) and suspended in 10 mM Tris-HCl, pH 8.0, 500 mM NaCl and 10% glycerol and

transferred to 1.5 mL centrifuge tubes. Tubes were placed into 70°C Eppendorff ThermoMixer with heated lid for 1 h, cooled for 1 h on ice, and centrifuged (16,300 x g, 4°C, 1 h). Clarified supernatant was assessed with Bradford dye against 10, 5, and 0 μM of purified parent polymerase.

Polymerase activity assay was performed with the following final concentrations: 1x ThermoPol, 0.5 μM of template 4, 0.5 μM of 5′-IR800-primer PBS8.20mer, 100 μM tNTPs, and 1 or 2 μL of lysate in a final volume of 20 μL. Primer-template were annealed in 1x ThermoPol at 90°C for 5 min and immediately placed on ice for 5 min, followed by the addition of tNTPs and lysate on ice. Reactions were placed into thermal cycler at 55°C. At designated time points 1 μL of reaction was removed, mixed with 39 μL of quenching buffer (95% formamide, 25 mM EDTA) and denatured at 95°C for 10 min. Denatured samples were loaded (10 μL) into a 15% urea denaturing PAGE and run at 10 W for 1.5 h and visualized on Li-Cor Odyssey CLx Imager, Image Studio Lite v5.2.

Recombinant polymerase expression and purification of novel TNA polymerases

Protein expression and purification for was performed as previously described [5]. In brief, pGDR11 vector containing the variant of interest was transformed into XL1 Blue *E. coli*, grown, induced for expression, and harvested. Cells were suspended in buffer (10 mM Tris-HCl, pH 8.0, 500 mM NaCl, 10% glycerol), sonicated, heat treated at 70°C for 1 h, and cooled on ice for 1 h. The solution was centrifuged (20000 rpm, 4°C, 20 min) and clarified supernatant was treated with a final concentration of 0.5% (v/v) PEI for 15 min. Treated supernatant was centrifuged (20000 rpm, 4°C, 20 min) to precipitate nucleic acids. Ammonium sulfate precipitation was

performed (final concentration: 60% (w/v)), centrifuged (20000 rpm, 4°C, 20 min), and the protein pellet was resuspended in equilibration buffer (10 mM Tris-HCl, pH 8.0 50 mM NaCl, 10% glycerol). Polymerase was purified by hand on a 5 mL heparin affinity column and eluted by stepwise addition of buffers containing: 10 mM Tris-HCl, pH 8.0, 10% glycerol and increasing concentration of NaCl (50, 100, 250, 500, 750 mM).

Thermal stability challenge

Purified polymerase was quantified to 10 μM and aliquoted (100 μL) into 6, 1.5 mL Eppendorf tubes and placed into an Eppendorff ThermoMixer set 90°C with a thermal lid to prevent evaporation. At the designated time (1-6 h) a tube was removed and placed on ice until activity evaluation. The collected tubes were spun at 16,3000 x g for 5 min at 4°C and UV quantified using a NanoDrop instrument. The polymerase was assayed for activity of TNA synthesis across a DNA template. Polymerase activity screen was performed with the following final concentrations: 1 μM primer-template duplex, 100 μM tNTPs, 1 μM heat-treated Polymerase in a 20 μL reaction. After 30 minutes of incubation, a 1 μL aliquot of the reaction was quenched by the addition of 39 μL of quenching buffer and visualized by denaturing polyacrylamide gel electrophoresis with fluorescent imaging on a Li-Cor Odyssey CLx Imager, Image Studio Lite v5.2.

Substrate recognition

TNA polymerases were evaluated for the ability to recognize and incorporated C5′-modified uracil TNA triphosphates (Trp and Phe) previously generated by chemical synthesis [9]. Polymerase activity screen was performed as described above with the exception of 2

equivalents of polymerase and combination of the C-5 modified tUTP substrate with standard tCTP, tGTP, and tATP substrates for a final concentration of 100 μM.

Synthesis and preparation of TNA aptamers for BLI analysis

Aptamer generation for BLI analysis was performed as described[10]. In brief, the DNA templates encoding the TNA aptamers were ordered with the addition of the PBS8 primer binding site on the 3′ termini. The DNA template (1.2 μM) was annealed to a biotinylated-PBS8 primer (1 μM) in 1x ThermoPol at 90°C for 5 min and immediately snap cooled, followed by the addition of 100 μM of respective TNA base mixes and 2 μM of 10-92 for a total volume of 750 μL. Primer extension was performed at 55°C for 2 h. The reaction was quenched by freezing the sample, vacuum concentrating to a volume of <100 μL and adding 250 μL of quenching buffer. The quenched sample was heated to 90°C for 5 min and loaded onto a 15% urea-denaturing polyacrylamide gel and ran at 15W for 2.5 hrs. The product band was excised, electroeluted, buffer exchanged into water and UV quantified. For BLI analysis, the aptamer was diluted to 100 nM final concentration (1 mL) with BLI binding buffer (10mM HEPES pH 7.0, 150 mM NaCl, 3 mM EDTA, 0.05% Tween-20) heated to 90°C for 5 minutes and snap cooled. The protein target was prepared by performing 2-fold serial dilutions starting from 100 nM to obtain the following concentrations: 25, 12.5, 6.25, and 3.125 nM.

BLI analysis of selected TNA aptamers

Aptamer sequences of interest were characterized with four different concentrations of target and one buffer only sensor to determine the background. Prior to testing, all sensors were equilibrated in BLI binding buffer for at least 30 min. After aptamer folding, the BLI run was

performed with the following steps: a buffer only baseline 60 s to equilibrate sensors, loading the aptamer for 200 s, a second buffer only baseline for 200 s, an association phase with the target protein for 600 s, and a dissociation phase for 600 s. The plate and reagents were incubated at 30°C for the duration of the experiment and for 10 min prior to each run. Data was acquired and analyzed using the Octet Data Acquisition software v12.0 and Analysis HT software v12.0.1.55, respectively. For full kinetic measurements, the buffer only baseline sample was used to subtract the background from all samples before applying Savitzky-Golay filtering and fitting both the association and dissociation curves together and applying a global fit to determine K_D and other metrics.

TNA polymerase fidelity

Fidelity analysis of the novel 10-92 TNA polymerase was performed as previously described [11]. In brief, a primer extension reaction was performed using a DNA primer (1 μM Extra_PBS37) containing a two-nucleotide mismatch (AA-AA) annealed to a DNA template (1 μM CMM595). The primer extension reaction was performed for 1 h at 55°C in 1x Thermopol buffer, 100 μM tNTPs, and 2 μM 10-92 TNA polymerase in a 25 μL reaction. The extension product was purified by denaturing PAGE and reverse transcribed into cDNA. The reverse transcription reaction was performed in 1x ThermoPol with 2 μM Bst-LF DNA polymerase, 2 μM PBS33, 0.4 mM dNTPs in a final volume of 20 μL and incubated for 4 h at 50°C. The cDNA product was PCR amplified with Taq DNA polymerase (PBS33 and PBS38) and agarose gel purified. The purified DNA was ligated into a TOPO vector and transformed into *E. coli* DH5-α competent cells. Colony PCR was performed on individual colonies to verify inserted amplicon, grown in liquid media, miniprepped and sequenced by Sanger sequencing with the M13F universal primer.

DNA sequences were aligned to the template and analyzed using CLC Main Workbench 23.0.5 software.

TNA versus DNA primed TNA synthesis

The DNA template (EM619, 1 μM) was annealed to either a TNA or DNA PBS8 primer (PBS8, 0.5 μM) in 1x ThermoPol at 95°C for 5 min and immediately snap cooled, followed by the addition of 10-92 polymerase (0.5 μM). The reaction was initiated with the addition of 100 μM of tNTPs for a total volume of 20 μL. Primer extension was performed at 55°C and quenched at designated timepoints with the addition of quenching buffer (95% Formamide, 25 mM EDTA). The quenched samples were heated to 95°C for 5 min and loaded onto a 15% urea-denaturing polyacrylamide gel and ran at 15W for 1.5 hrs.

10-92 for reversion analysis

Ten single point reversions were introduced into the backbone of 10-92, creating ten new constructs. Forward primers were designed to contain the desired reversion mutation (Supplementary Table 1) and reverse primers were ordered for whole plasmid PCR. The resulting PCR reaction was seeded into a KLD treatment and transformed into NEB 5alpha *E. coli* for colony picking. The resulting constructs were validated by Sanger sequencing, expressed in *E. coli*, purified as described above, validated by SDS-PAGE, and challenged in a primer extension.

Cloning Kod-RSGA with directed evolution mutations from rounds 6-10

The 10 mutations acquired by directed evolution were cloned into the backbone of Kod-RSGA,

creating the construct Kod-RSGA+DE mutations. A gBlock was ordered to contain the linear dsDNA sequence encompassing the 10 mutations (Supplementary Table 1, Kod-RSGA+DE mutations). Corresponding primers (Supplementary Table 2, Vector Fwd KodRSGA+ DE mutations and Vector Rvs KodRSGA+ DE mutations) were ordered to facilitate the amplification of the remaining gene and expression vector from the original pGDR11-*kod*-RSGA construct. The resulting vector portion was used after a DNA clean up, DpnI treatment, and agarose gel purification. The insert and vector were ligated through Gibson assembly and transformed into NEB 5alpha *E. coli* for colony picking. The resulting construct was validated by Sanger sequencing, expressed in *E. coli*, and purified as described above, and challenged in a primer extension.

Polymerase kinetics

TNA polymerase kinetic measurements were performed by monitoring the reaction progress over time as the DNA primer-template (P-T) duplex (PBS8 & EM619, Supplementary Table 2) was extended with tNTPs. For each time point, 3 individual reaction replicates (15 µL) from a single master mix poised under single-turnover conditions with equimolar concentrations of polymerase (0.2 µM) and pre-annealed P-T duplex (0.2 µM each, heated at 90°C for 5 min, cooled on ice) in ThermoPol buffer (NEB, 2X) were pre-equilibrated for 5 minutes at 55°C. The reactions were then initiated by the addition of a preheated tNTPs mixture (200 µM of each tNTP). The reactions were stopped at designated time points by plunging the reaction vessel into powdered dry ice. Once all the time points were collected, the frozen reactions were thawed on a cold plate by adding 15 µL (6 µM) EvaGreen® dye, previously identified as the optimal intercalating dye to monitor synthesis of TNA on a DNA template[12]. Each reaction (25 µL) was

transferred to a clear V-bottom 96-well plate and fluorescence intensity was measured (ex: 487/20 nm, em: 528/20 nm) using a Synergy Neo2 plate reader with the Gen5 v3.14.03 software (BioTek).

Baseline fluorescence was measured using Kod-WT (exo-), the parent polymerase with limited TNA synthesis ability [13], which was used to collect a 0 second time point by freezing the reaction immediately after the addition of the tNTP solution. Additionally, the maximum fluorescence value was obtained from a 2 hour reaction performed with each polymerase. Because Kod-RI does not reach full-length product, the maximum fluorescence value of Kod-RS was also used to for Kod-RI. The raw fluorescence data from the plate reader was normalized by subtracting the baseline and then dividing by the difference of the maximum (2 hour reaction) and minimum (baseline) fluorescence values. The fluorescence data was then converted to nucleotides per polymerase as previously described [12] by multiplying data with the conversion factor F = [Template] * L / [Polymerase], where L is the length of extension in bases, then plotted over time. Rates in nucleotides per minute are extracted as the slope of the linear range of each curve. Data were processed using Microsoft Excel v16.66.

Cloning of truncated polymerases for crystallography

The 10-92 gene was PCR amplified from pGDR11-10-92 using the 10-92-Fwd and 10-92-Rvs primers (Supplementary Table 2) containing *NdeI* and *NotI* restriction enzyme sites, respectively. Two 10-92 C-terminal truncations, 10-92_760 and 10-92_750, were PCR amplified by substituting the 10-92-Rvs primer with either 10-92-Rvs_760 or 10-92-Rvs_750 primer (Supplementary Table 2), both containing the *NotI* restriction enzyme site. All three gene

constructs additionally consist of two mutations (D141A and E143A) to inactivate exonuclease activity. Purified PCR product of each gene construct and pET21 were digested with *NdeI* and *NotI* restriction enzymes and ligated, resulting in pET21-10-92, pET21-10-92_760, and pET21-10-92_750. The three constructs were verified by Sanger sequencing.

Sample preparation and crystallization

The DNA template was purchased from IDT as an HPLC purified sample. The DNA primer was also ordered from IDT with standard desalting. The primer-template (P/T) duplex was prepared by combining equal parts of the primer and template strands in Kod buffer (50 mM Tris-HCl pH 8.5, 200 mM NaCl, 0.1 mM EDTA, 1 mM DTT) and supplemented with 20 mM $MgCl_2$, heating at 95°C for 5 min and slow cooling to 10°C over 10 min. An initial binary complex was prepared by incubating 10-92_760 (6 mg mL^{-1}) with 1.2 M equivalents of the annealed P/T duplex at 37°C for 30 min. 3 M excess of dtTTP was added and incubated at 37°C for 30 min. The resulting n+1 binary complex was further incubated with 10 M excess tATP at 37°C for another 30 min to yield the final ternary complex.

Next, the 10-92 ternary complex was screened against ~900 conditions in a hanging-drop format using a Mosquito crystallization robot (SPT LabTech). A positive crystal hit was initially identified in condition C9 in a NeXtal PACT Screen and was further optimized in 24-well hanging drop trays over a range of pH and PEG concentrations, with each drop consisting of 1 µL of sample mixed with 1 µL of mother liquor over 500 µL mother liquor in every well. Trays were stored in the dark at room temperature and crystals typically grew between 3-4 days. The 10-92 ternary complex crystal used for structure determination was grown in 0.2 M lithium

chloride, 0.1 M HEPES pH 6.5, and 35 % polyethylene glycol 6000.

Structure determination

A 2.73Å dataset was collected at the Advanced Light Source (Lawrence Berkeley National Laboratory, Berkeley, CA) from a single crystal. Images were indexed, integrated, and merged using XDS [14,15]. The initial model was determined by molecular replacement using Phaser [16] with PDB structure 5VU8 as the search model, and the final model was determined using iterative rounds of manual building through Coot [17] and refinement with phenix [18]. The stereochemistry and geometry of all structures were validated with Molprobity [19] with the final refinement parameters summarized in Supplementary Table 4. Final coordinates and structure factors have been deposited in the Protein Data Bank. All molecular graphics were prepared with PyMOL [20].

Supplementary Table 1. DNA gBlocks. Oligonucleotide sequences ordered from IDT written in the 5′→3′ direction.

Name	DNA Sequence (5′ → 3′)
Kod-RSGA_gene	ATGATCCTCGACACTGACTACATAACCGAGGATGGAAAGCCTGTCATAAGAATTTTCA AGAAGGAAAACGGCGAGTTTAAGATTGAGTACGACCGGACTTTTGAACCCTACTTCTA CGCCCTCCTGAAGGACGATTCTGCCATTGAGGAAGTCAAGAAGATAACCGCCGAGAGG CACGGGACGGTTGTAACGGTTAAGCGGGTTGAAAAGGTTCAGAAGAAGTTCCTCGGGA GACCAGTTGAGGTCTGGAAACTCTACTTTACTCATCCGCAGGACGTCCCAGCGATAAG GGACAAGATACGAGAGCATCCAGCAGTTATTGACATCTACGAGTACGACATACCCTTC GCCAAGCGCTACCTCATAGACAAGGGATTAGTGCCAATGGAAGGCGACGAGGAGCTGA AAATGCTCGCCTTCGCGATTGCGACTCTCTACCATGAGGGCGAGGAGTTCGCCGAGGG GCCAATCCTTATGATAAGCTACGCCGACGAGGAAGGGGCCAGGGTGATAACTTGGAAG

	AACGTGGATCTCCCCTACGTTGACGTCGTCTCGACGGAGAGGGAGATGATAAAGCGCT TCCTCCGTGTTGTGAAGGAGAAAGACCCGGACGTTCTCATAACCTACAACGGCGACAA CTTCGACTTCGCCTATCTGAAAAAGCGCTGTGAAAAGCTCGGAATAAACTTCGCCCTC GGAAGGGATGGAAGCGAGCCGAAGATTCAGAGGATGGGCGACAGGTTTGCCGTCGAAG TGAAGGGACGGATACACTTCGATCTCTATCCTGTGATAAGACGGACGATAAACCTGCC CACATACACGCTTGAGGCCGTTTATGAAGCCGTCTTCGGTCAGCCGAAGGAGAAGGTT TACGCTGAGGAAATAACCACAGCCTGGGAAACCGGCGAGAACCTTGAGAGAGTCGCCC GCTACTCGATGGAAGATGCGAAGGTCACATACGAGCTTGGGAAGGAGTTCCTTCCGAT GGAGGCCCAGCTTTCTCGCTTAATCGGCCAGTCCCTCTGGGACGTCTCCCGCTCCAGC ACTGGCAACCTCGTTGAGTGGTTCCTCCTCAGGAAGGCCTATGAGAGGAATGAGCTGG CCCCGAACAAGCCCGATGAAAAGGAGCTGGCCAGAAGACGGCAGAGCTATGAAGGAGG CTATGTAAAAGAGCCCGAGAGAGGGTTGTGGGAGAACATAGTGTACCTAGATTTTAGA TCCCTGTACCCCTCAATCATCATCACCCACAACGTCTCGCCGGATACGCTCAACAGAG AAGGATGCAAGGAATATGACGTTGCCCCACAGGTCGGCCACCGCTTCTGCAAGGACTT CCCAGGATTTATCCCGAGCCTGCTTGGAGACCTCCTAGAGGAGAGGCAGAAGATAAAG AAGAAGATGAAGGCCACGATTGACCCGATCGAGAGGAAGCTCCTCGATTACAGGCAGA GGCGTATCAAGATCCTGGCAAGCAGCTACTACGGTTACTACGGCTATGCAAGGGCGCG CTGGTACTGCAAGGAGTGTGCAGAGAGCGTAACGGCCTGGGGAAGGGAGTACATAACG ATGACCATCAAGGAGATAGAGGAAAAGTACGGCTTTAAGGTAATCTACAGCGACACCG ACGGATTTTTTGCCACAATACCTGGAGCCGATGCTGAAACCGTCAAAAAGAAGGCTAT GGAGTTCCTCAAGTATATCAACGCCAAACTTCCGGGCGCGCTTGAGCTCGAGTACGAG GGCTTCTACAAACGCGGCTTCTTCGTCACGAAGAAGAAGTATGCGGTGATAGACGAGG AAGGCAAGATAACAACGGGTGGACTTGAGATTGTGAGGCGTGACTGGAGCGAGATAGC GAAAGAGACGCAGGCGAGGGTTCTTGAAGCTTTGCTAAAGGACGGTGACGTCGAGAAG GCCGTGAGGATAGTCAAAGAAGTTACCGAAAAGCTGAGCAAGTACGAGGTTCCGCCGG AGAAGCTGGTGATCCACGAGCAGATAACGAGGGATTTAAAGGACTACAAGGCAACCGG TCCCCACGTTGCCGTTGCCAAGAGGTTGGCCGCGAGAGGAGTCAAAATACGCCCTGGA ACGGTGATAAGCTACATCGTGCTCAAGGGCTCTGGGAGGATAGGCGACAGGGCGATAC

	CGTTCGACGAGTTCGACCCGGCGAAGCACAAGTACGACGCCGAGTACTACATTGAGAA CCAGGTTCTCCCAGCCGTTGAGAGAATTCTGAGAGCCTTCGGTTACCGCAAGGAAGAC CTGCGCTACCAGAAGACGAGACAGGTTGGTTTGAGTGCTTGGCTGAAGCCGAAGGGAA CTTGA
Tgo_RSGA_ gBlock	AGCGGCCTGGTGCCGCGCGGCAGCATGATCCTCGACACTGACTACATAACCGAGGATG GAAAGCCTGTCATAAGAATTTTCAAGAAGGAAAACGGCGAGTTTAAGATTGACTACGA CCGGAACTTTGAACCCTACATCTACGCCCTCCTGAAGGACGATTCTGCCATTGAGGAC GTCAAGAAGATAACCGCCGAGAGGCACGGGACGACCGTAAGGGTTGTCCGGGCCGAAA AGGTTAAGAAGAAGTTCCTCGGGAGACCAATAGAGGTCTGGAAACTCTACTTTACTCA TCCGCAGGACGTCCCAGCGATAAGGGACAAGATAAAGGAGCATCCAGCAGTTGTGGAC ATCTACGAGTACGACATACCCTTCGCCAAGCGCTACCTCATAGACAAGGGATTAATCC CAATGGAAGGCGACGAGGAGCTGAAAATGCTCGCCTTCGCGATTGCGACTCTCTACCA TGAGGGCGAGGAGTTCGCCGAGGGGCCAATCCTTATGATAAGCTACGCCGACGAGGAA GGGGCCAGGGTGATAACTTGGAAGAACATCGATCTCCCCTACGTTGACGTCGTCTCGA CGGAGAAGGAGATGATAAAGCGCTTCCTCAAGGTTGTGAAGGAGAAAGACCCGGACGT TCTCATAACCTACAACGGCGACAACTTCGACTTCGCCTATCTGAAAAAGCGCTCCGAA AAGCTCGGAGTCAAGTTCATCCTCGGAAGGGAAGGAAGCGAGCCGAAGATTCAGAGGA TGGGCGACAGGTTTGCCGTCGAAGTGAAGGGACGGATACACTTCGATCTCTATCCTGT GATAAGACGGACGATAAACCTGCCCACATACACGCTTGAGGCCGTTTATGAAGCCATC TTCGGTCAGCCGAAGGAGAAGGTTTACGCTGAGGAAATAGCGCAGGCCTGGGAAACCG GCGAGGGACTTGAGAGAGTCGCCCGCTACTCGATGGAAGATGCGAAGGTCACATACGA GCTTGGGAAGGAGTTCTTCCCGATGGAGGCCCAGCTTTCTCGCTTAGTAGGCCAGTCC CTCTGGGACGTCTCCCGCTCCAGCACTGGCAACCTCGTTGAGTGGTTCCTCCTCAGGA AGGCCTATGAGAGGAATGAGCTGGCCCCGAACAAGCCCGATGAAAGGGAGCTGGCCAG AAGACGGGAGAGCTATGCGGGAGGCTATGTAAAAGAGCCCGAGAGAGGGTTGTGGGAG AACATAGTGTACCTAGATTTTAGATCCCTGTACCCCTCAATCATCATCACCCACAACG TCTCGCCGGATACGCTCAACAGAGAAGGATGCGAGGAATATGACGTTGCCCCACAGGT CGGCCACAAGTTCTGCAAGGACTTCCCAGGATTTATCCCGAGCCTGCTTGGAGACCTC

	CTAGAGGAGAGGCAGAAGGTAAAGAAGAAGATGAAGGCCACGATTGACCCGATCGAGA AGAAGCTCCTCGATTACAGGCAGAGGCGTATCAAGATCCTGGCAAGCAGCTTCTACGG TTACTACGGCTATGCAAAGGCGCGCTGGTACTGCAAGGAGTGTGCAGAGAGCGTAACG GCCTGGGGAAGGCAGTACATAGAGACCACCATCAGGGAGATAGAGGAAAAGTTTGGCT TTAAGGTACTCTACGCGGACACCGACGGATTTTTTGCCACAATACCTGGAGCCGATGC TGAAACCGTCAAAAAGAAGGCTAAGGAGTTCCTCGACTATATCAACGCCAAACTTCCG GGCCTGCTTGAGCTCGAGTACGAGGGCTTCTACAAACGCGGCTTCTTCGTCACGAAGA AGAAGTATGCGGTGATAGACGAGGAAGACAAGATAACAACGGGTGGACTTGAGATTGT GAGGCGTGACTGGAGCGAGATAGCGAAAGAGACGCAGGCGAGGGTTCTTGAAGCTATA CTAAAGCACGGTGACGTCGAGGAAGCCGTGAGGATAGTCAAAGAAGTTACCGAAAAGC TGAGCAAGTACGAGGTTCCGCCGGAGAAGCTGGTGATCTACGAGCAGATAACGAGGGA TTTAAAGGACTACAAGGCAACCGGTCCCCACGTTGCCGTTGCCAAGAGGTTGGCCGCG AGAGGAATAAAAATACGCCCTGGAACGGTGATAAGCTACATCGTGCTCAAGGGCTCTG GGAGGATAGGCGACAGGGCGATACCGTTCGACGAGTTCGACCCGGCGAAGCACAAGTA CGACGCCGAGTACTACATTGAGAACCAGGTTCTCCCAGCCGTTGAGAGAATTCTGAGA GCCTTCGGTTACCGCAAGGAAGACCTGCGCTACCAGAAGACGAGACAGGTTGGTTTGG GGGCTTGGCTGAAGCCGAAGACATGAGTCGACCTGCAGCCAAGCTTA
DV_RSGA_ gBlock	AGCGGCCTGGTGCCGCGCGGCAGCATGATCCTCGACGCTGACTACATAACCGAGGATG GAAAGCCTATTATAAGAATTTTCAAGAAGGAAAACGGCGAGTTTAAGGTTGAGTACGA CCGGAACTTTAGACCCTACATTTACGCCCTCCTGAAGGACGATTCTCAGATTGATGAA GTCAGGAAGATAACCGCCGAGAGGCACGGGAAGATAGTAAGAATTATAGATGCCGAAA AGGTTAGGAAGAAGTTCCTCGGGAGACCAATTGAGGTCTGGAGGCTCTACTTTGAACA TCCGCAGGACGTCCCAGCGATAAGGGACAAGATACGAGAGCATTCCGCAGTTATTGAC ATCTTTGAGTACGACATACCCTTCGCCAAGCGCTACCTCATAGACAAGGGATTAATTC CAATGGAAGGCGACGAGGAGCTGAAATTGCTCGCCTTCGCGATTGCGACTCTCTACCA TGAGGGCGAGGAGTTCGCCAAGGGGCCAATCATAATGATAAGCTACGCCGACGAGGAA GAAGCCAAAGTGATAACTTGGAAGAAGATCGATCTCCCCTACGTTGAGGTCGTCTCGA GCGAGAGGGAGATGATAAAGCGCTTCCTCAAGGTTATAAGGGAGAAAGACCCGGACGT

	TATAATAACCTACAACGGCGACTCTTTCGACCTTCCCTATCTGGTTAAGCGCGCCGAA AAGCTCGGAATAAAGCTACCCCTCGGAAGGGATGGAAGCGAGCCGAAGATGCAGAGGC TTGGCGACATGACAGCCGTCGAAATAAAGGGACGGATACACTTCGATCTCTATCACGT GATAAGACGGACGATAAACCTGCCCACATACACGCTTGAGGCCGTTTATGAAGCCATC TTCGGTAAGCCGAAGGAGAAGGTTTACGCTCACGAAATAGCTGAGGCCTGGGAAACCG GCAAGGGACTTGAGAGAGTCGCCAAGTACTCGATGGAAGATGCGAAGGTCACATACGA GCTTGGGAGGGAGTTCTTCCCGATGGAGGCCCAGCTTTCTCGCTTAGTCGGCCAGCCC CTCTGGGACGTCTCCCGCTCCAGCACTGGCAACCTCGTTGAGTGGTACCTCCTCAGGA AGGCCTATGAGAGGAATGAGCTGGCCCCGAACAAGCCCGATGAAAGGGAGTACGAGAG AAGGCTACGGGAGAGCTATGCTGGAGGCTATGTAAAAGAGCCCGAGAAAGGGTTGTGG GAGGGGTTAGTGTCCCTAGATTTTAGATCCCTGTACCCCTCAATCATCATCACCCACA ACGTCTCGCCGGATACGCTCAACAGAGAAGGATGCAGGGAATATGACGTTGCCCCAGA GGTCGGCCACAAGTTCTGCAAGGACTTCCCAGGATTTATCCCGAGCCTGCTTAAGAGG CTCCTAGATGAGAGGCAGGAAATAAAGAGGAAGATGAAGGCCTCTAAAGACCCGATCG AGAAGAAGATGCTCGATTACAGGCAGAGGCGTATCAAGATCCTGGCAAGCAGCTACTA CGGTTACTACGGCTATGCAAAAGCGCGCTGGTACTGCAAGGAGTGTGCAGAGAGCGTA ACGGCCTGGGGAAGGGAGTACATAGAGTTCGTAAGGAAGGAGCTGGAGGAAAAGTTCG GCTTTAAGGTATTATACATAGACACCGACGGACTCTACGCCACAATACCTGGAGCCAA ACCCGAAGAGATAAAAAAGAAGGCTCTAGAGTTCGTAGATTATATCAACGCCAAACTT CCGGGCCTGCTTGAGCTCGAGTACGAGGGCTTCTACGTGCGCGGCTTCTTCGTCACGA AGAAGAAGTATGCGTTGATAGACGAGGAAGGCAAGATAATCACGGGTGGACTTGAGAT TGTGAGGCGTGACTGGAGCGAGATAGCGAAAGAGACGCAGGCGAAAGTTCTTGAAGCT ATCCTAAAGCATGGTAACGTCGAGGAGGCCGTGAAGATAGTCAAAGAAGTTACCGAAA AGCTGAGCAAGTACGAGATACCGCCGGAGAAGCTGGTGATCTACGAGCAGATAACGAG GCCCTTACACGAGTACAAGGCAATAGGTCCCCACGTTGCCGTTGCCAAGAGGTTGGCC GCGAGAGGAGTCAAAGTGCGCCCTGGAATGGTGATAGGGTACATCGTGCTCAGGGGCG ACGGGCCAATAAGCAAGAGGGCGATACTTGCAGAGGAGTTCGACCTCGCGAAGCACAA GTACGACGCCGAGTACTACATTGAGAACCAGGTTCTCCCAGCCGTTCTTAGAATTCTG

	GAGGCCTTCGGTTACCGCAAGGAAGACCTGCGCTGGCAGAAGACGAAACAGACAGGTT TGACGGCTTGGCTGAACATCAAGAAGAAGTGAGTCGACCTGCAGCCAAGCTTA
9°N_RSGA_ gBlock	AGCGGCCTGGTGCCGCGCGGCAGCATGATCCTCGACACTGACTACATAACCGAGAACG GAAAGCCTGTCATAAGAGTGTTCAAGAAGGAAAACGGCGAGTTTAAGATTGAGTACGA CCGGACTTTTGAACCCTACTTCTACGCCCTCCTGAAGGACGATTCTGCCATTGAGGAT GTCAAGAAGGTGACCGCCAAAAGGCACGGGACGGTTGTAAAAGTTAAGCGGGCGGAAA AGGTTCAGAAGAAGTTCCTCGGGAGACCAATTGAGGTCTGGAAACTCTACTTTAACCA TCCGCAGGACGTCCCAGCGATAAGGGACCGTATACGAGCGCATCCAGCAGTTGTGGAC ATCTACGAGTACGACATACCCTTCGCCAAGCGCTACCTCATAGACAAGGGATTAATTC CAATGGAAGGCGACGAGGAGCTGACCATGCTCGCCTTCGCGATTGCGACTCTCTACCA TGAGGGCGAGGAGTTCGGCACCGGGCCAATCCTTATGATAAGCTACGCCGACGGCAGC GAAGCCAGGGTGATAACTTGGAAGAAAATCGATCTCCCCTACGTTGACGTCGTCTCGA CGGAGAAAGAGATGATAAAGCGCTTCCTCCGTGTTGTGCGTGAGAAAGACCCGGACGT TCTCATAACCTACAACGGCGACAACTTCGACTTCGCCTATCTGAAAAAGCGCTGTGAA GAACTCGGAATAAAATTCACCCTCGGAAGGGATGGAAGCGAGCCGAAGATTCAGAGGA TGGGCGACAGGTTTGCCGTCGAAGTGAAGGGACGGATACACTTCGATCTCTATCCTGT GATAAGACGGACGATAAACCTGCCCACATACACGCTTGAGGCCGTTTATGAAGCCGTC TTCGGTAAACCGAAGGAGAAGGTTTACGCTGAGGAAATAGCGCAGGCCTGGGAAAGCG GCGAGGGCCTTGAGAGAGTCGCCCGCTACTCGATGGAAGATGCGAAGGTCACATACGA GCTTGGGCGTGAGTTCTTCCCGATGGAGGCCCAGCTTTCTCGCTTAATCGGCCAGTCC CTCTGGGACGTCTCCCGCTCCAGCACTGGCAACCTCGTTGAGTGGTTCCTCCTCAGGA AGGCCTATAAAAGGAATGAGCTGGCCCCGAACAAGCCCGATGAACGTGAGCTGGCCAG ACGTCGTGGCGGTTATGCGGGAGGCTATGTAAAAGAGCCCGAGAGAGGGTTGTGGGAT AACATAGTGTACCTAGATTTTAGATCCCTGTACCCCTCAATCATCATCACCCACAACG TCTCGCCGGATACGCTCAACAGAGAAGGATGCAAGGAATATGACGTTGCCCCAGAAGT CGGCCACAAATTCTGCAAGGACTTCCCAGGATTTATCCCGAGCCTGCTTGGAGACCTC CTAGAGGAGAGGCAGAAGATAAAGCGCAAGATGAAGGCCACGGTTGACCCGCTGGAGA AAAAGCTCCTCGATTACAGGCAGAGGCGTATCAAGATCCTGGCAAGCAGCTTCTACGG

	TTACTACGGCTATGCAAAAGCGCGCTGGTACTGCAAGGAGTGTGCAGAGAGCGTAACG GCCTGGGGAAGGGAGTACATAGAAATGGTGATCCGCGAGCTGGAGGAAAAGTTCGGCT TTAAGGTACTGTACGCGGACACCGACGGACTGCATGCCACAATACCTGGAGCCGATGC TGAAACCGTCAAAAAGAAGGCTAAAGAGTTCCTCAAGTATATCAACCCGAAACTTCCG GGCCTGCTTGAGCTCGAGTACGAGGGCTTCTACGTGCGCGGCTTCTTCGTCACGAAGA AGAAGTATGCGGTGATAGACGAGGAAGGCAAGATAACAACGGGTGGACTTGAGATTGT GAGGCGTGACTGGAGCGAGATAGCGAAAGAGACGCAGGCGAGGGTTCTTGAAGCTATT CTAAAGCATGGTGACGTCGAGGAAGCCGTGAGGATAGTCAAAGAAGTTACCGAAAAGC TGAGCAAGTACGAGGTTCCGCCGGAGAAGCTGGTGATCCACGAGCAGATAACGAGGGA TTTACGTGACTACAAGGCAACCGGTCCCCACGTTGCCGTTGCCAAGAGGTTGGCCGCG AGAGGAGTCAAAATACGCCCTGGAACGGTGATAAGCTACATCGTGCTCAAGGGCTCTG GGAGGATAGGCGACAGGGCGATACCGGCGGACGAGTTCGACCCGGCGAAGCACCGTTA CGACGCCGAGTACTACATTGAGAACCAGGTTCTCCCAGCCGTTGAGAGAATTCTGAAA GCCTTCGGTTACCGCAAGGAAGACCTGCGCTACCAGAAGACGAAACAGGTTGGTTTGG GCGCTTGGCTGAAGGTTAAGGGAAAATGAGTCGACCTGCAGCCAAGCTTA
Kod-RSGA + DE mutations (R6-10)	GAGTTCCTTCCGATGGAGGCCCAGCTTTCTCGCTTAATCGGCCATTCCCTCTGGGACG TCTCCCGCTCCAGCACTGGCAACCTCGTTGAGTGGTTCCTCCTCAGGAAGGCCTATGA GAGGAATGAGCTGGCCCCGAACAAGCCCGATGAAAAGGAGCTGGCCAGAAGACGGCAG AGCTATGAAGGAGGCTATGTAAAAGAGCCCGAGAGAGGGTTGTGGGAGAACATAGTGT ACCTAGATTTTAGATCCCTGTACCCCTCAATCATCATCACCCACAACGTCTCGCCGGA TACGCTCAACAGAGAAGGATGCAAGGAATATGACGTTGCCCCACAGGTCGGCCACCGC TTCTGCAAGGACTTCCCAGGATTTATCCCGAGCCTGCTTGGAGACCTCCTAGAGGAGA GGCAGAAGATAAAGAAGAAGATGAAGGCCACGATTGACCTGATCGAGAGGAAGCTCCT CGATTACAGGCAGAGGCGTATCAAGATCCTGGCAAGCGGCTACTACGGTTACTACGGC TATGCAAGGGCGCGCTGGTACTGCAAGGAGTGTGCAGAGAGCGTAACGGCCTGGGGAA GGGAGTACATAACGATGACCATCAAGGAGATAGAGGAAAAGTACGGCTTTAAGGTAAT CTACAGCGACACCGACGGATTTTTTGCCCCAATACATGGAGCCGATGCTGAAACCGTC AAAAAGAAGGCTATGGAGTTCCTCAAGTATATCAACGCCAAACTTCCGGGCGCGCTTG

	AGCTCGAGTACGAGGGCTTCTACAAACGCGGCTTCTTCGTCACGAAGAAGAAGTATGC GGTGATAGACGAGGAAGGCAAGATAACAACGGGTGGACTTGAGATTGTGAGGCGTGGC TGGAGCGAGATAGCGAAAGAGACGCAGGCGAGGGTTCTTGAAGCTTTGCTAAAGGACG GTGACGTCGAGAAGGCCGTGAGGATAGTCAAAGAAGTTACCGAAAAGCTGAGCAAGTA CGAGGTTCCGCCGGAGAAGCTGGTGATCCACGAGCAGATAACGAGGGATTTAAAGGAC TACAAGGCAACCGGTCCCCACGTTGCCGTTGCCAAGAGGTTGGCCGCGAGAGGAGTCA AAATACGCCCTGGAACGGTGATAAGCTACATCGTGCTCAAGGGCTCTGGGAGGATAGG CGACAGGGCGATATCGTTCGACGAGTTCGACCCGGCGAAGCACAAGTACGACGCCGAG TACTACATTGAGAACCAGGTTCTCCCACCCGTTGAGAGAATTCTGAGAGCCTGCGGTT ACCGCAAGGAAGACCTGCGCTACCAGAAGACGAGACAGGTTGGTTTGAGTGCTTGGCT GACGCCGAAGGGAACTTGAGTCGACCTGCAGCCAAGCTTAATTAGCTGAGCTTGGACT CCT

Supplementary Table 2. DNA oligonucleotides. Oligonucleotide sequences ordered from IDT written in the 5′→3′ direction. Fluorescent dyes or modifications are noted on the specified termini. Highlighted in red are mismatched positions for sequences used for fidelity experiments.

Vector-Fwd	TGAGTCGACCTGCAGCCAAGCTTAATTAGCTGAGC
Vector-Rvs	GCTGCCGCGCGGCACCAG
10-92-Fwd	ATCCATATGATCCTCGACACTGACTAC
10-92-Rvs	ACGCATGCGGCCGCTCAAGTTCCTTTCGGCGTCAG
10-92-Rvs_760	ACGCATGCGGCCGCTCACTTCTGGTAGCGCAGGTC
10-92-Rvs_750	ACGCATGCGGCCGCTCAACCGCAGGCTCTCAG
Fragment1-Fwd	ATCACCATACGGATCCAAGCGGC
Fragment2-Fwd	TACGCCCTCCTGAAGGACGATTCT
Fragment3-Fwd	CAGGACGTCCCAGCGATAAGGGAC
Fragment4-Fwd	GCCTTCGCGATTGCGACTCTCTA
Fragment5-Fwd	AAGGGACGGATACACTTCGATCTCTAT
Fragment6-Fwd	GAGGCCCAGCTTTCTCGCTTA
Fragment7-Fwd	CCGGATACGCTCAACAGAGAAGGATGC
Fragment8-Fwd	GGCGTATCAAGATCCTGGCAAGCAGC
Fragment9-Fwd	CTTGAGCTCGAGTACGAGGGCTTCTAC
Fragment10-Fwd	GAGATAGCGAAAGAGACGCAGGCG
Fragment11-Fwd	CCGCCGGAGAAGCTGGTGATC
Fragment1-Rvs	AGAATCGTCCTTCAGGAGGGCGTA
Fragment2-Rvs	GTCCCTTATCGCTGGGACGTCCTG
Fragment3-Rvs	TAGAGAGTCGCAATCGCGAAGGC
Fragment4-Rvs	ATAGAGATCGAAGTGTATCCGTCCCTT
Fragment5-Rvs	TAAGCGAGAAAGCTGGGCCTC
Fragment6-Rvs	GCATCCTTCTCTGTTGAGCGTATCCGG
Fragment7-Rvs	GCTGCTTGCCAGGATCTTGATACGCC

Fragment8-Rvs	GTAGAAGCCCTCGTACTCGAGCTCAAG
Fragment9-Rvs	CGCCTGCGTCTCTTTCGCTATCTC
Fragment10-Rvs	GATCACCAGCTTCTCCGGCGG
Fragment11-Rvs	TAAGCTTGGCTGCAGGTCGACTCA
30mer.V2HP.Cy3	/5Cy3/ACAACCATTATCTAGAGCGATTCGTATAGGTGGTATCCGAAAGGATACCACC
QP08.Iowa	ATGGTTGT/3IABkFQ/
PBS8.20mer	/5IRD680/GTCCCCTTGG
Template 4	CATACGCGAGATACACTGCCAGAAAACCGCGTGTCTACGAGGTGGTATCCCCAAGGGGAC
Random Library (30.5)	TCGTTCTCGGACGGAGAGATattgaa-N_{40}-GGTGGTATCCCCAAGGGGAC
Activity / Fidelity template EM619	CCCACACCCTCCTATCGCTAAACACACACTTAATAAAGTTGGTGGTATCC
Aptamer template HIV-RT 2-1008	AACAAAGTATCGAGAGACTGACGAAAAACATAAGAGTAGCGGTGGTATCC
Aptamer template Bn 23	GGCGAATTACTAACAGCGCGGGTGGTATCC
Aptamer template CMM595	ACGCGCTTCTTGGAAACTCACGCACCCATAATAATACTAAGGTGGTATCC
PBS37	ATTGCTTGCCGGAGCTTATCGG**AA**GACTCACAGGCCTCCG
PBS33	TCGTTCTCGGACGGAGAGAT
PBS38	ATTGCTTGCCGGAGCTTATC
CMM595.Fidelity	TCGTTCTCGGACGGAGAGATACGCGCTTCTTGGAAACTCACGCACCCATAATAATACT AACGGAGGCCTGTGAGTC**AA**CC/3InvdT/
VectorFwd_KodRSGA+ DE mutations	ATTAGCTGAGCTTGGACTCCTGTTGATAGATCCAGT

VectorRvs_KodRSGA+ DE mutations	GGCCTCCATCGGAAGGAACTCCTTCCC
10-92_H339Q_Fwd	CTTAGTCGGCcagCCCCTCTGGG
10-92_H339Q_Rvs	CGAGAAAGCTGGGCCTCCATC
10-92_L474P_Fwd	CACGGTTGACccgCTGGAGAAAA
10-92_L474P_Rvs	GCCTTCATCTTGCGC
10-92_G493S_Fwd	CCTGGCAAGCagcTACTACGGTT
10-92_G493S_Rvs	ATCTTGATACGCCTCTGCC
10-92_P548T_Fwd	ATTTTTTGCCacaATACATGGAGCCGATG
10-92_P548T_Rvs	CCGTCGGTGTCCGCG
10-92_H550P_Fwd	TGCCCCAATAcctGGAGCCGATG
10-92_H550P_Rvs	AAAAATCCGTCGGTGTC
10-92_G615D_Fwd	TGTGAGGCGTgacTGGAGCGAGA
10-92_G615D_Rvs	ATCTCAAGTCCACCCG
10-92_S717P_Fwd	CAGGGCGATAccgTTCGACGAGT
10-92_S717P_Rvs	TCCCCTATCCTCCCAGAG
10-92_P741A_Fwd	GGTTCTCCCAgccGTTGAGAGAATTC
10-92_P741A_Rvs	TGGTTCTCAATGTAGTACTC
10-92_C749F_Fwd	TCTGAGAGCCttcGGTTACCGCA
10-92_C749F_Rvs	ATTCTCTCAACGGGTGG
10-92_T771K_Fwd	TGCTTGGCTGaagCCGAAGGGAAC
10-92_T771K_Rvs	CTCAAACCAACCTGTCTC

Supplementary Table 3. Summary of droplet information and incubation conditions.

Abbreviations: Homologous Recombination (Recomb.), error prone PCR (epPCR)

Round	Library	Top Variant	Fluorescent Threshold (afu)	Incubation 55 °C (h)	Droplets Detected	Droplets Sorted	Colonies Screened
1	Recomb.	-	60	18	36,680,319	20,079	-
2	Recomb.	-	50	14	27,145,396	170,219	-
3	Recomb.	-	50	10	27,121,065	158,867	-
4	Recomb.	-	50	3	32,852,979	13,588	-
5	Recomb.	5-270	50	1	33,678,095	198,421	500
6	epPCR: Thumb	-	50	18	11,905,802	121,520	100
7	epPCR: Thumb	7-47	50	12	13,879,733	37,650	100
8	epPCR: Thumb	8.64	50	5	17,863,171	1,140	100
9	epPCR: Thumb	8.64	50	5	11,480,842	22,534	100
10	epPCR: Palm, Finger	10-92	50	5	16,515,310	5,691	100

Supplementary Table 4. Data collection and refinement statistics.

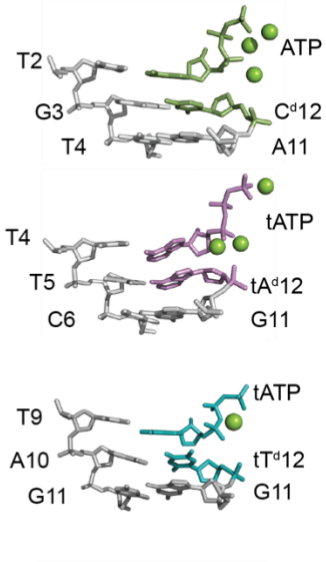
	10-92
Data Collection	
Space group	$P2_12_12_1$
Cell Dimensions	
a, *b*, *c* (Å)	76.869, 99.525, 110.727
α, β, γ (º)	90.0, 90.0, 90.0
Resolution (Å)	45.39-2.73 (2.828-2.73)
R_{merge}	0.1773 (0.97)
CC1/2	0.994 (0.766)
$I / \sigma I$	11.36 (2.24)
Completeness (%)	99.03 (98.20)
Redundancy	6.6 (6.8)
Refinement	
Resolution (Å)	2.73
No. reflections	22979 (2236)
R_{work}/R_{free}	0.2420/0.2966 (0.3490/0.4263)
No. atoms	6778
Protein/DNA	6686
tT/tATP	45
Solvent	47
B-factors	44.21
Protein/DNA	43.74
tT/tA	42.34
Solvent	41.13
R.m.s deviations	
Bond lengths (Å)	0.003
Bond angles (º)	0.53

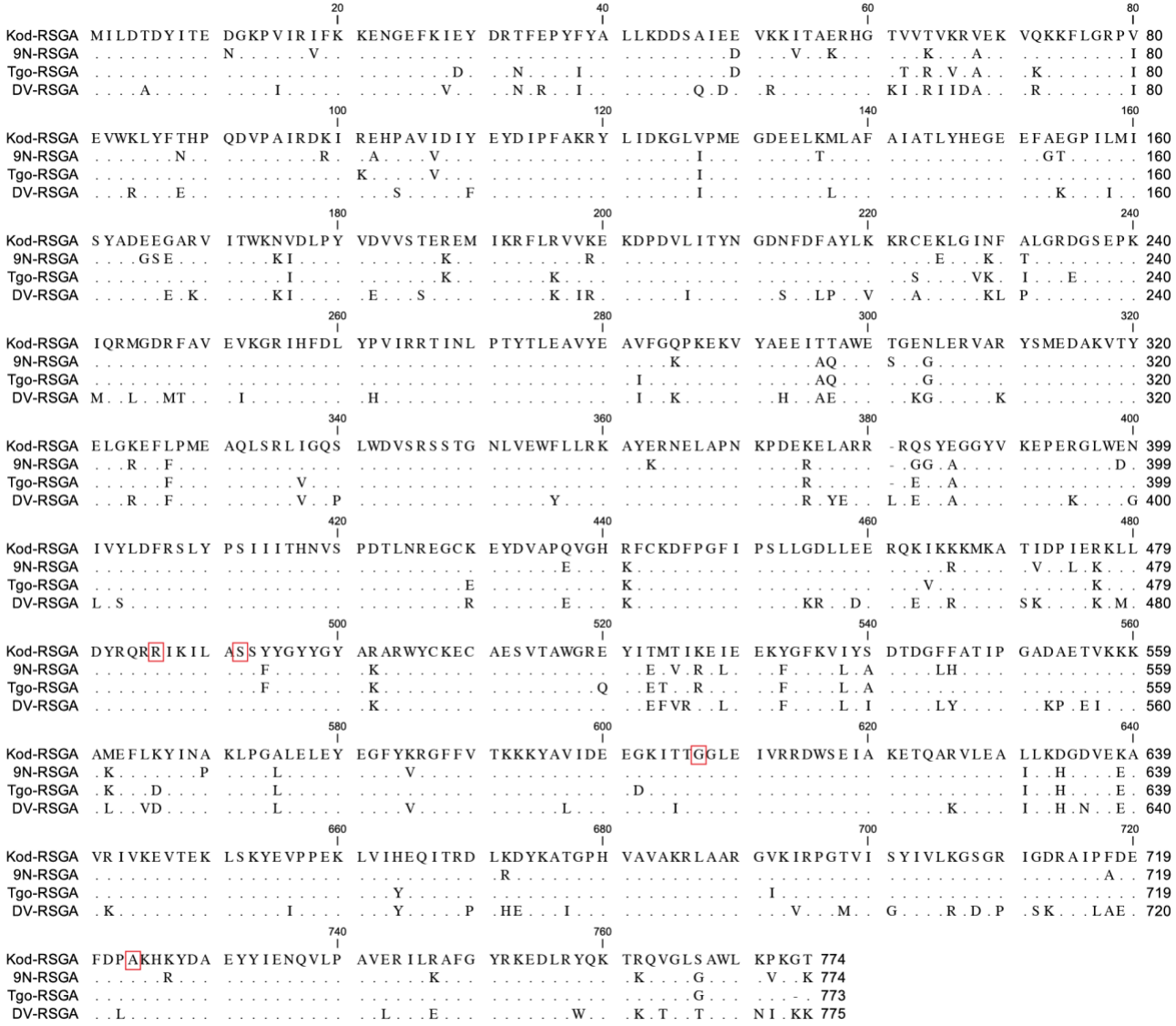
Number of TLS groups	3

* The statistics for the highest-resolution shell are shown in parenthesis.

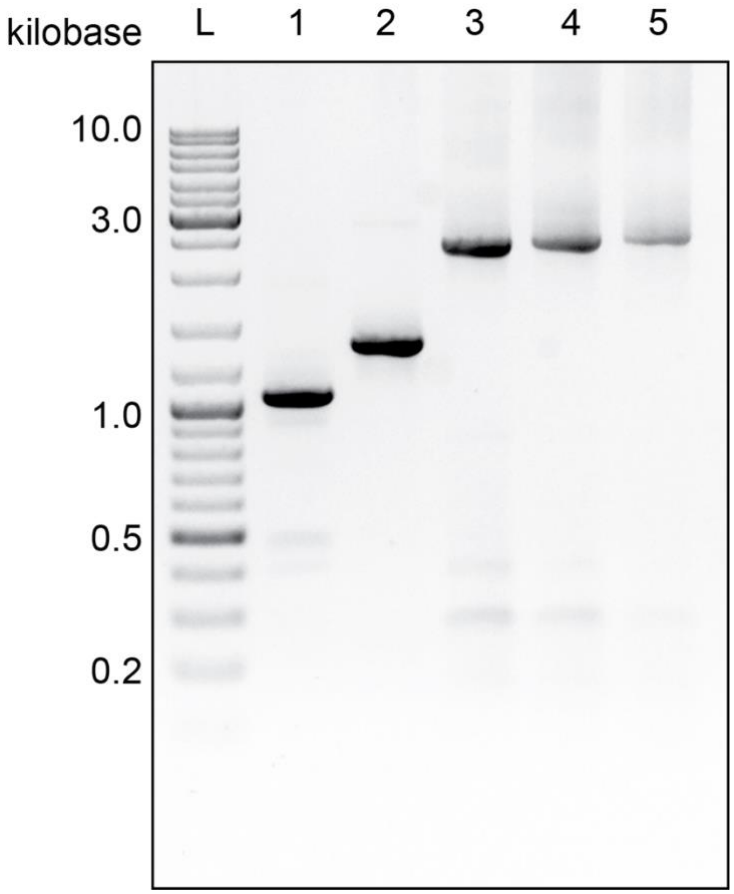
Supplementary Table 5. Local base-pair parameters[21].

		Kod WT	Kod-RI	10-92
T:dATP/ T:tATP	Shear (Å)	-0.01	0.02	0.44
	Stretch (Å)	-0.24	-0.71	-0.35
	Stagger (Å)	0.06	-1.83	-0.88
	Buckle (°)	-5.46	-20	-5.88
	Propeller(°)	0.75	-20.51	-1.40
	Opening (°)	2.85	-15.26	-1.63
C:G / T:tATP / A:tTTP	Shear (Å)	-0.27	-1.15	-0.11
	Stretch (Å)	-0.22	-0.1	0.06
	Stagger (Å)	0.01	-0.32	0.26
	Buckle (°)	-8.37	-32.93	-18.04
	Propeller(°)	-14.08	-8.41	-18.37
	Opening (°)	1.97	-6.13	5.27
T:A / C:G	Shear (Å)	-0.04	0.55	0.04
	Stretch (Å)	-0.04	-0.46	-0.15
	Stagger (Å)	-0.22	0.19	-0.07
	Buckle (°)	1.35	-12.39	-8.55
	Propeller(°)	0.96	8.37	-0.96
	Opening (°)	-2.94	4.84	-1.95

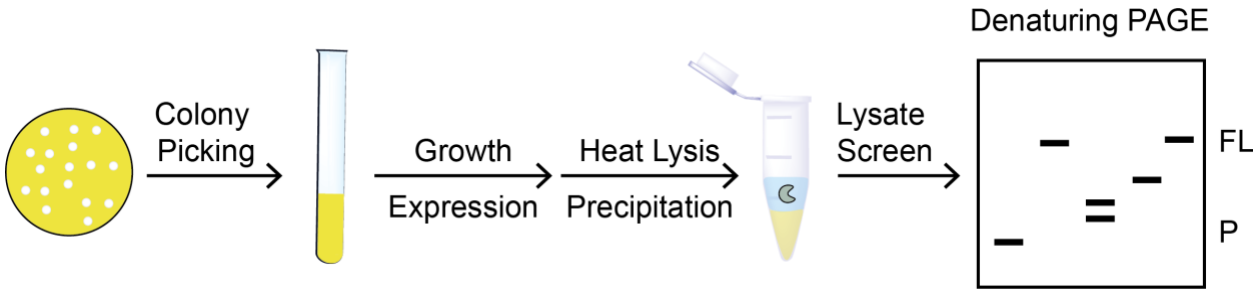




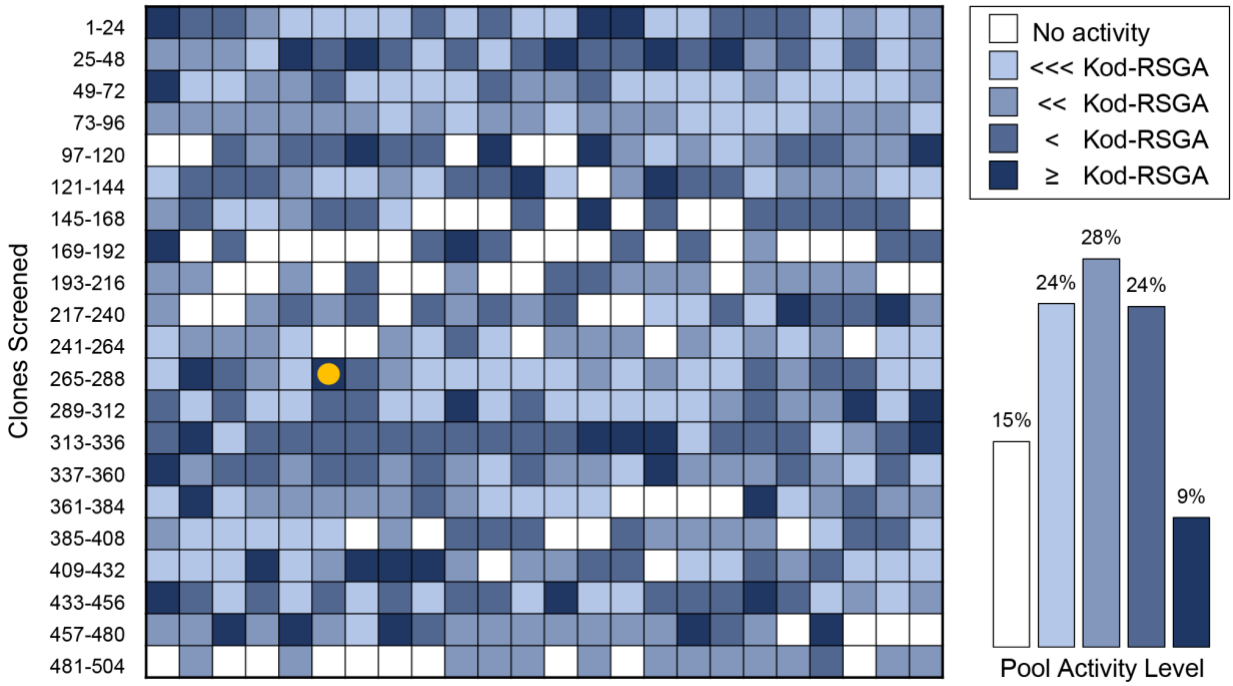
Supplementary Figure 1. Sequence alignment of archaeal RSGA DNA polymerase variants. *Red Boxes indicate the positions previously discovered in Kod-RSGA (A485R, N491S, R606G, T723A) that are known to enhance TNA synthesis. Polymerases contain the exonuclease silencing mutations D141A and E143A. Conserved residues are denoted as (.). Natural variation from the Kod scaffold is represented by the single letter amino acid abbreviation. Alignment was generated in CLC Main Workbench.*



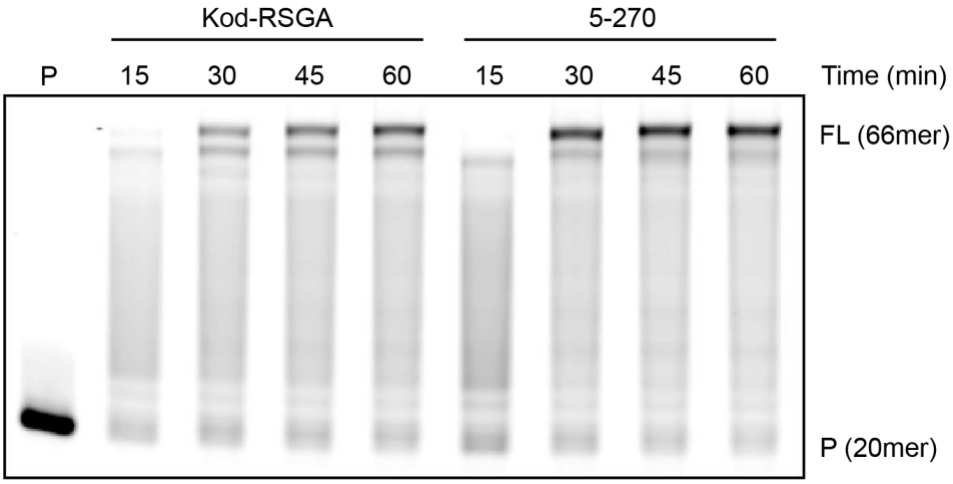
Supplementary Figure 2. Assembly of the DNA recombination library by overlap PCR. *Agarose gel showing assembly of the homologous recombination library. Lane L: DNA ladder. Lane 1: Kod-RSGA fragments 1—5. Lane 2: Kod-RSGA fragments 6-11. Lane 3: Kod-RSGA fragments 1-11. Lane 4: Kod-RSGA and Tgo-RSGA fragments 1-11. Lane 5: Kod-, Tgo-, DV-, and 9°N-RSGA fragments 1-11. The agarose gel was run once to confirm library assembly*.



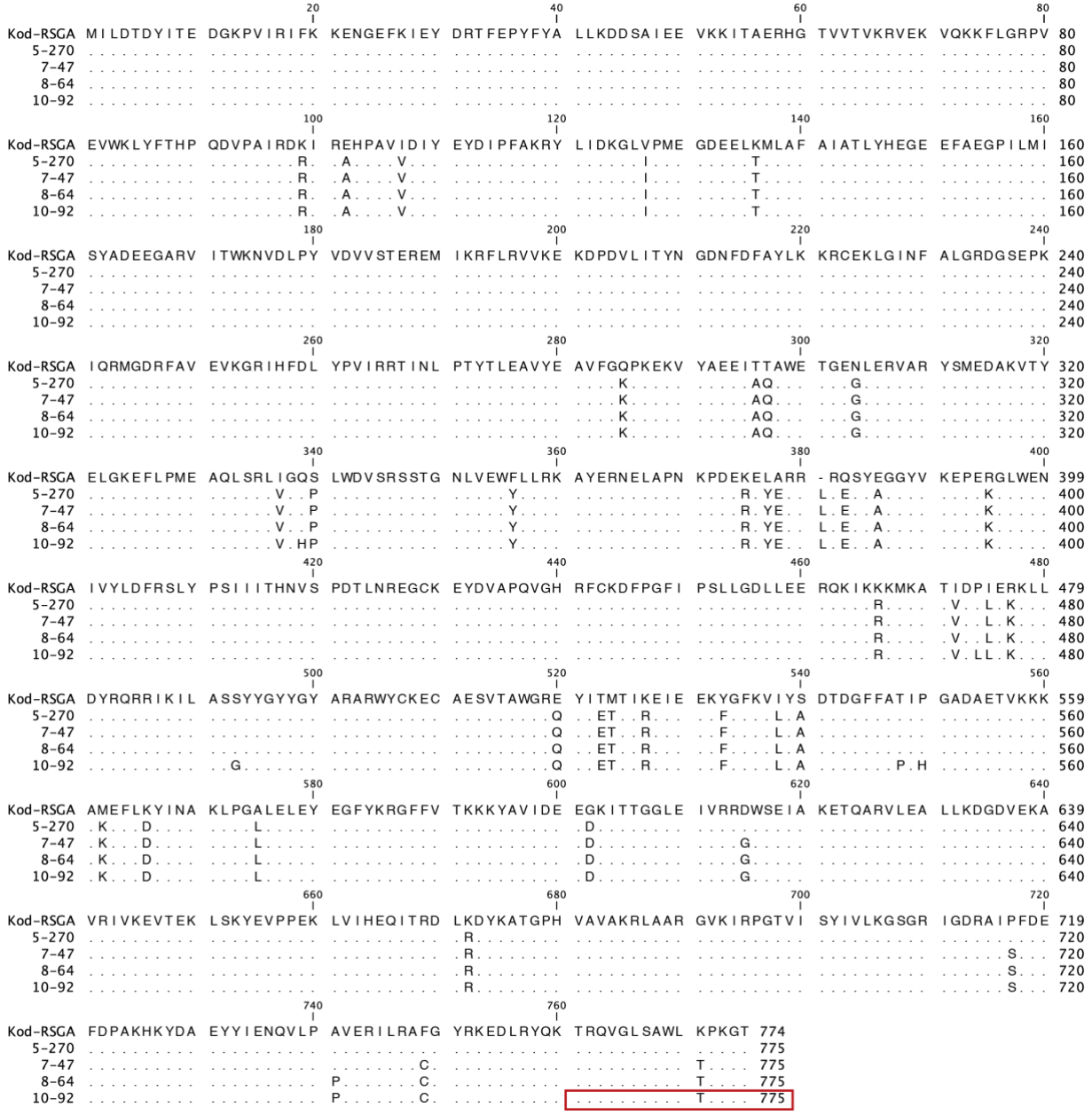
Supplementary Figure 3. Schematic of the lysate screen. *Regenerated libraries are transformed into XL1 Blue E. coli for expression of enriched variants obtained from DrOPS. Single isolated colonies are picked, grown, and induced to express a variant of interest. Cells are harvested, suspended in buffer, subjected to heat lysis, and cold incubation to precipitate endogenous E. coli proteins. The cell debris is clarified by centrifugation. Lysate supernatant is added to a polymerase activity assay and quenched at designated time points. The reaction is quenched, denatured, and analyzed by denaturing PAGE. High performing variants are purified and characterized for activity.*



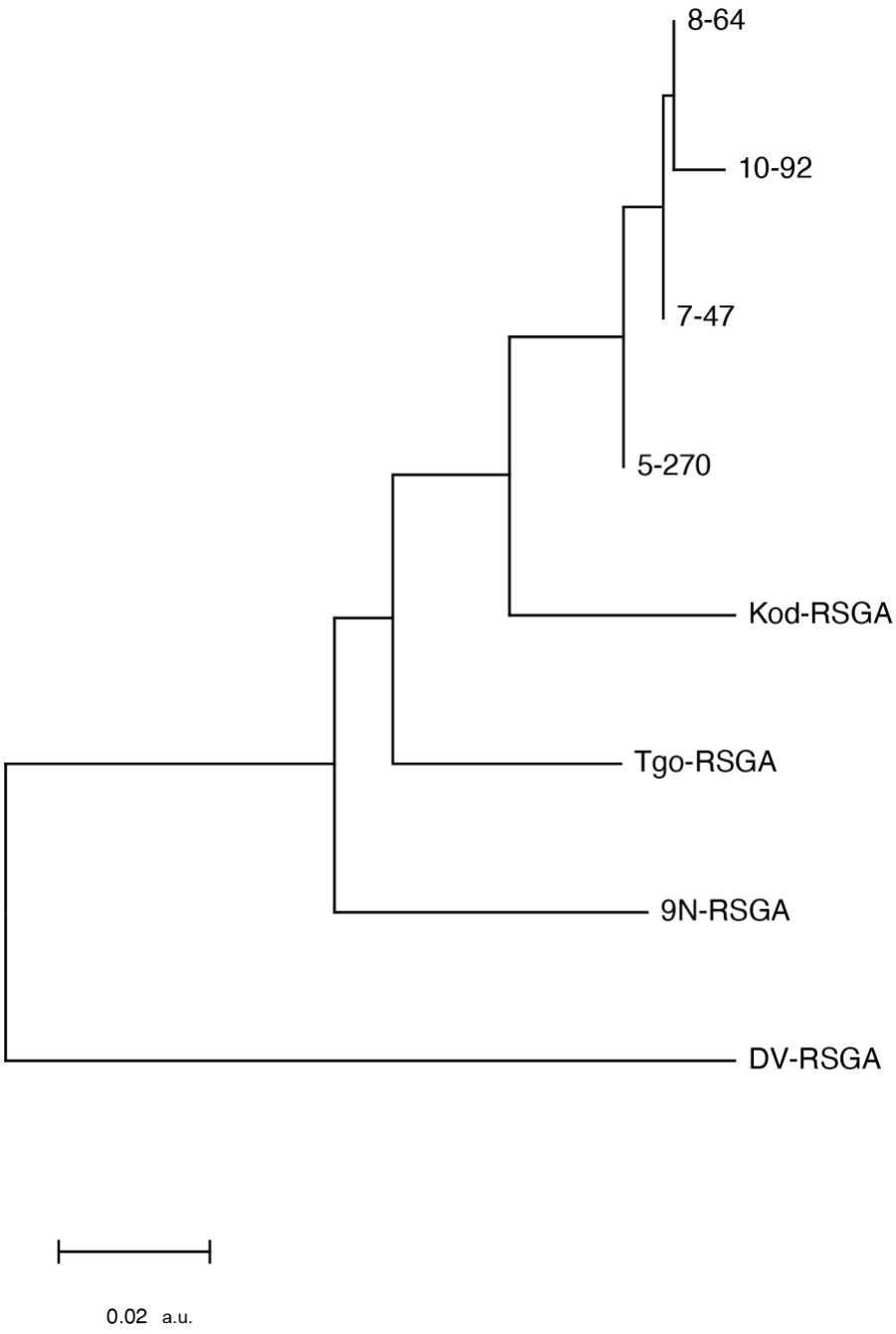
Supplementary Figure 4. Lysate activity screen following five rounds of selection. *Library members present in the pool after five rounds of in vitro selection were individually screened for TNA synthesis activity in a primer-extension assay. A qualitative assessment of primer extension efficiency relative to Kod-RSGA is represented as a heatmap. Clone 5-270 is shown with a gold dot. The percent of overall activity (i.e., the sum of each rank) is represented on a bar graph.*



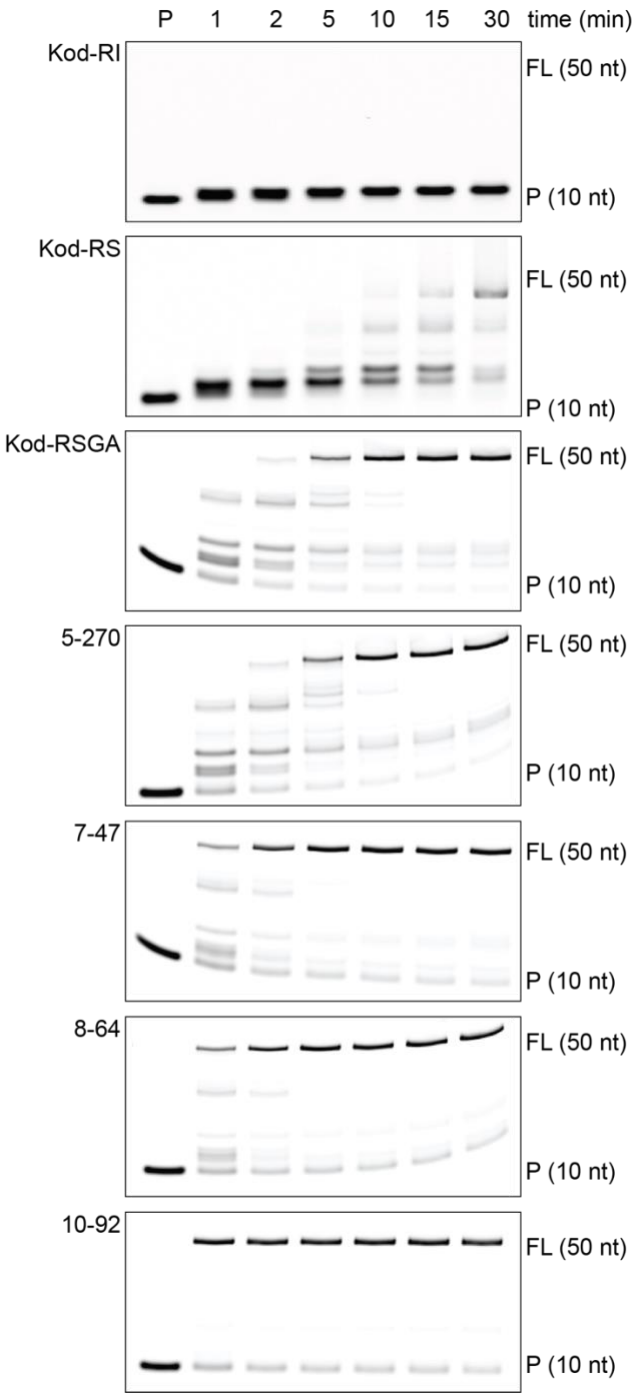
Supplementary Figure 5. TNA synthesis activity of clone 5-270.
Engineered polymerases Kod-RSGA and 5-270 are challenged to incorporate TNA substrates in a primer extension assay, templated by a random DNA aptamer library (L30.5).



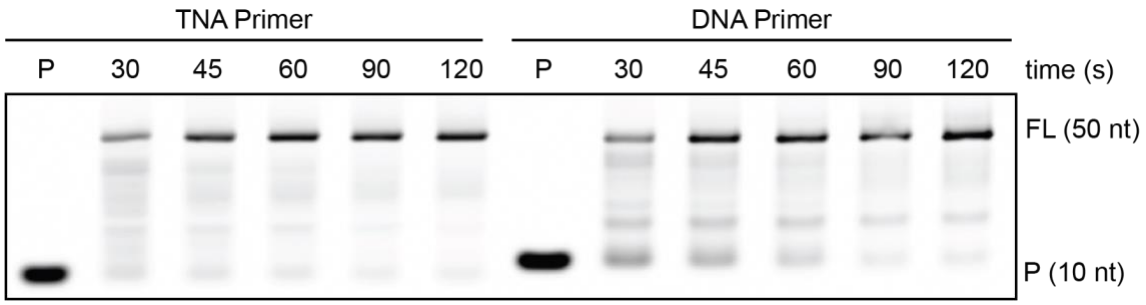
Supplementary Figure 6. Sequence alignment of newly evolved TNA polymerases. *TNA polymerases identified in this study are compared to Kod-RSGA from our previous study. Conserved residues are denoted as (.), while mutations are denoted by their single letter amino acid abbreviations. The red box denotes amino acid residues that are removed for crystallography. Alignment was generated in CLC Main Workbench.*



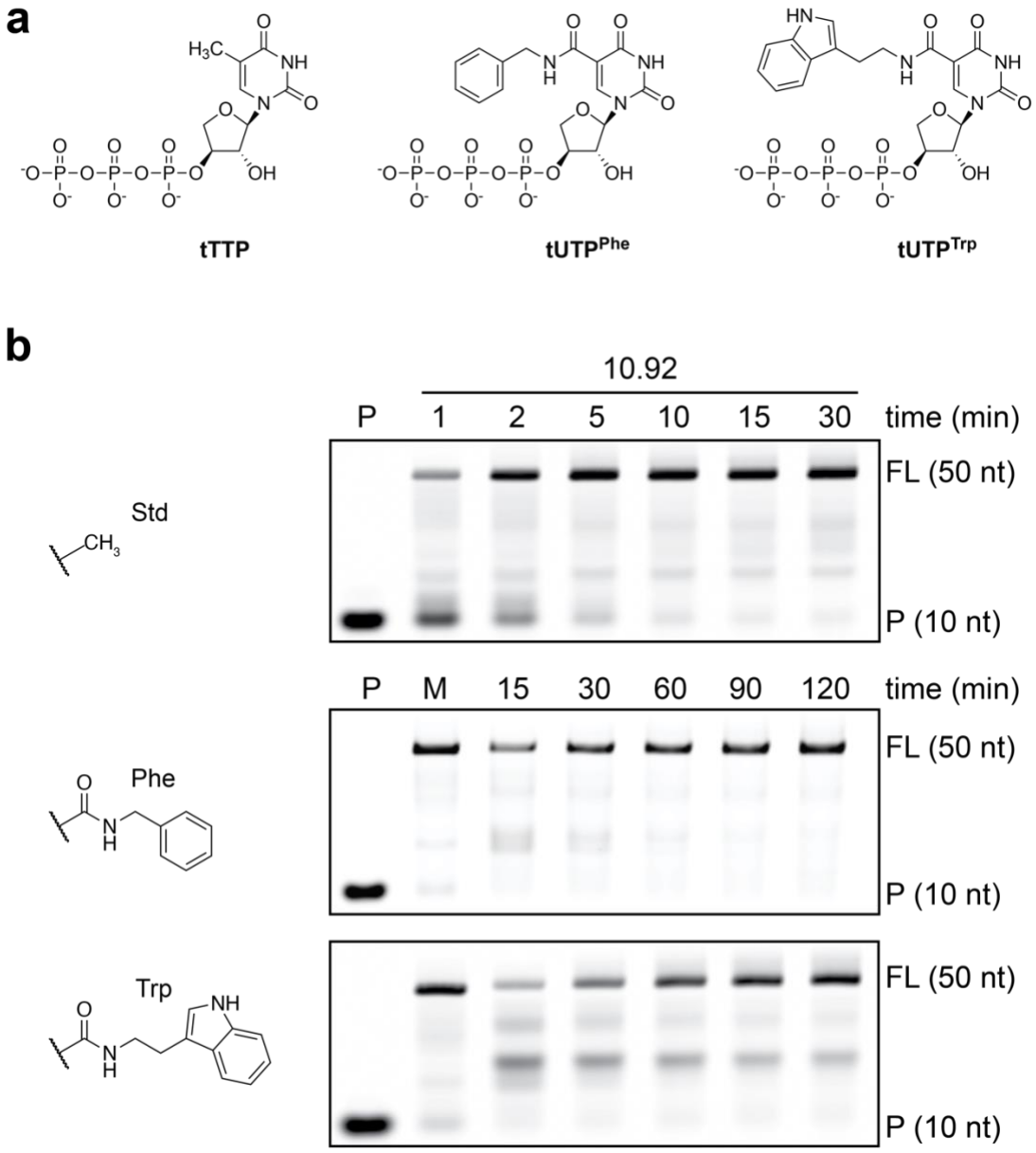
Supplementary Figure 7. Phylogenetic analysis of the evolved TNA polymerases. *MEGA was used to generate a phylogenetic tree by the neighbor-joining method of the newly evolved TNA polymerases relative to their starting parent sequences.*



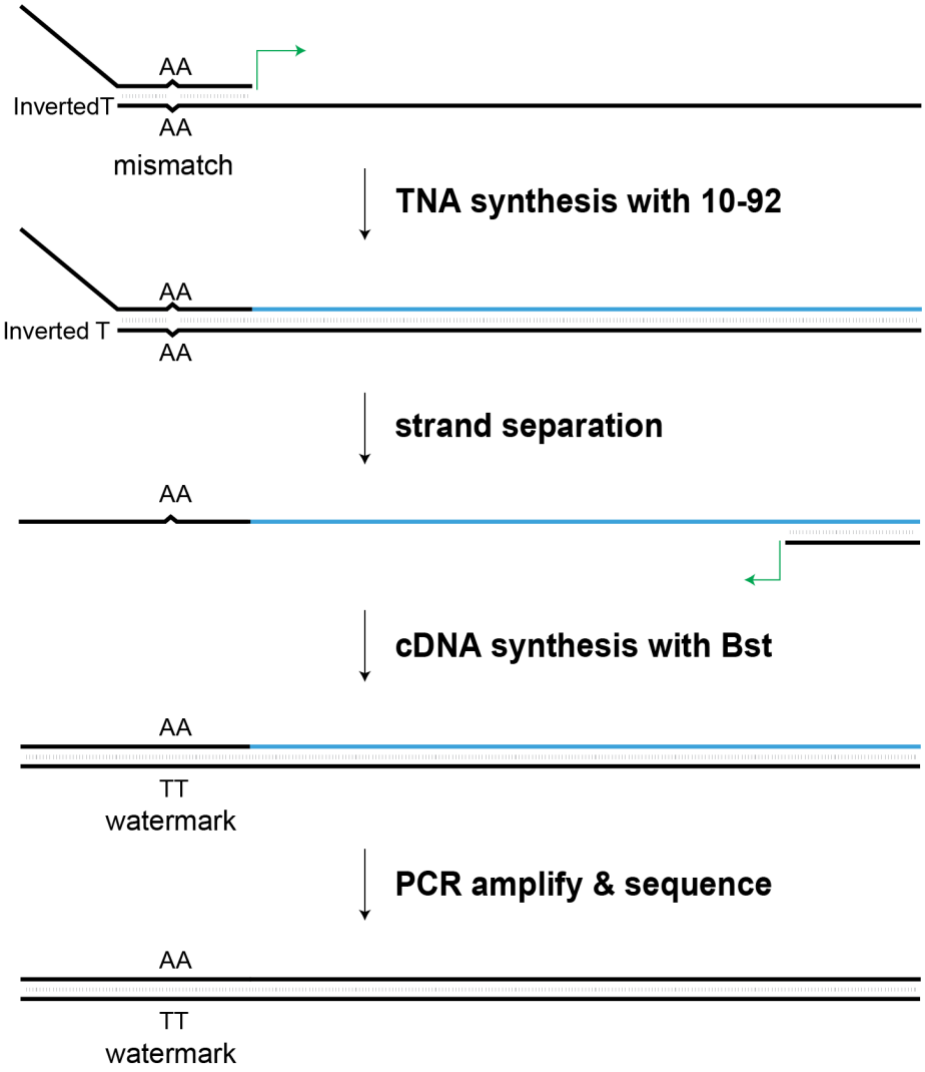
Supplementary Figure 8. Comparative activity analysis of ancestral TNA polymerases. *Each polymerase was challenged to extend a DNA primer with 40 consecutive TNA residues. The reactions were quenched at defined time points and analyzed by denaturing polyacrylamide gel electrophoresis with fluorescent imaging on a LiCOR imager. P; primer, FL; full-length product. Representative gels of two independent experiments are shown.*



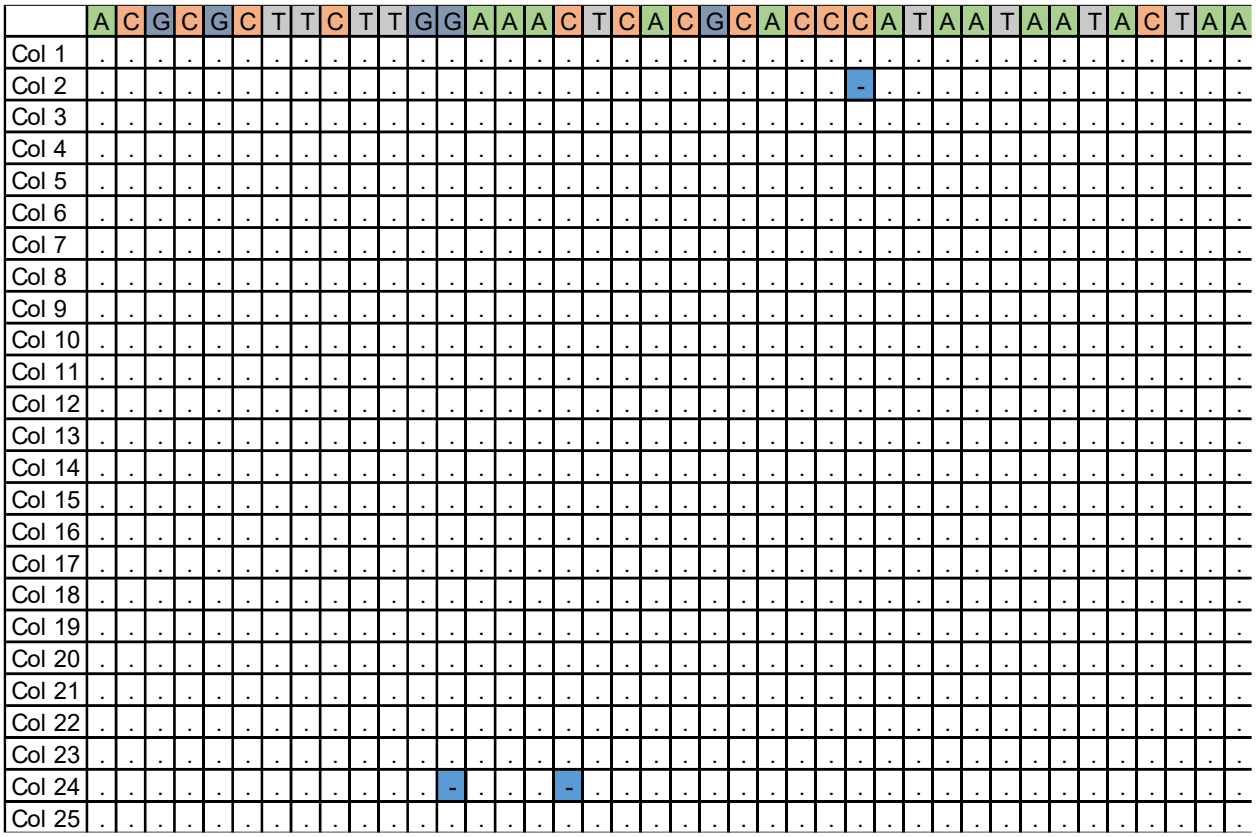
Supplementary Figure 9. TNA versus DNA primed TNA synthesis. *The 10-92 polymerase was challenged to copy a DNA template (EM619) by extending either a TNA or DNA primer with 40 consecutive TNA residues. The reactions were quenched at defined times and analyzed by denaturing polyacrylamide gel electrophoresis with fluorescent imaging on a LiCOR imager. P; primer, FL; full-length product. A representative gel of two independent experiments is shown.*



Supplementary Figure 10. Functional analysis of 10-92. *(a) Chemical structures of base-modified tUTP analogs. (b) Enzymatic synthesis of base-modified TNA. Time course analysis was performed with 10-92 and tNTP mixtures containing standard bases as well as base-modified tUTP analogs. Full length product with standard base tNTPs (top) can be observed after 30 sec, compared to 15 min for C5'-modifed tUTPs phenylalanine (Phe) (middle) or tryptophan (Trp) side chains (bottom). Representative gels of two independent experiments are shown.*

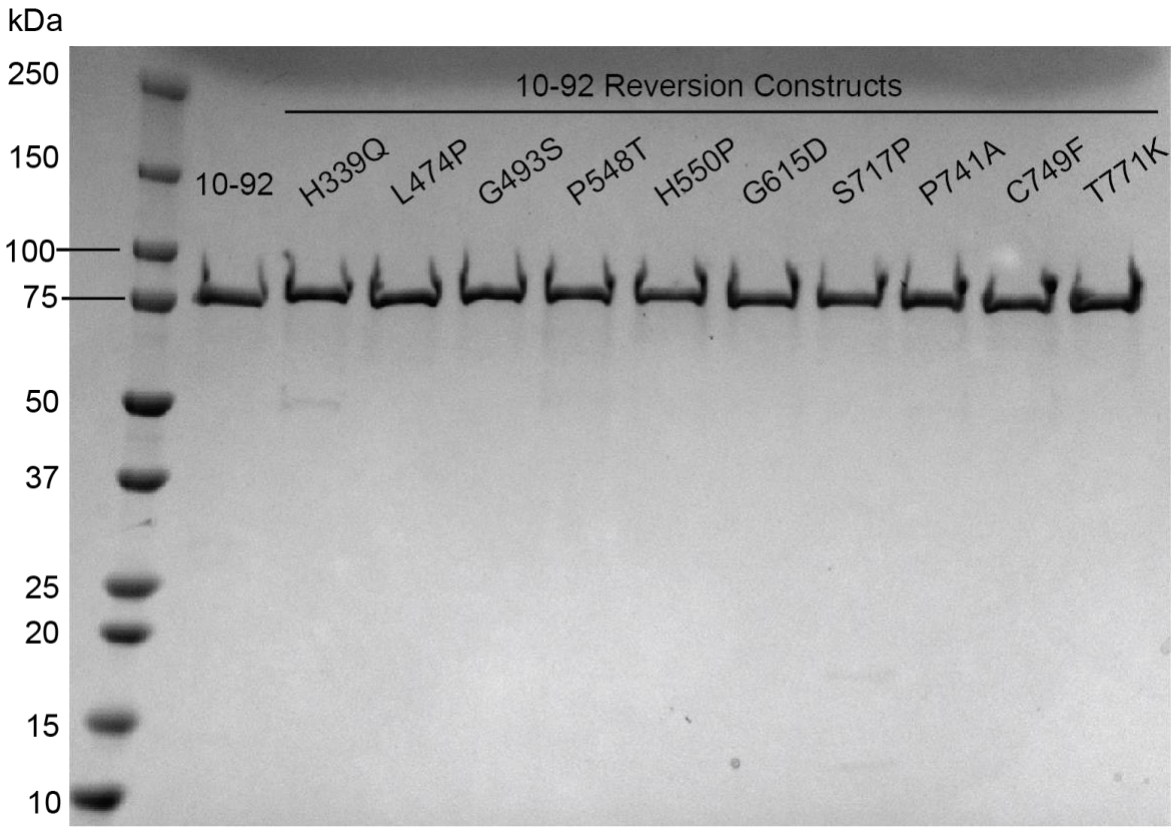


Supplementary Figure 11. Schematic of the fidelity assay. *An overhang primer is designed with an AA-AA mismatch in the primer region of the DNA template (black) with a 3' inverted deoxy-T. The 3' inverted deoxy-T prevents extension into the overhang region. The TNA strand (blue) is synthesized across the DNA template by 10-92 TNA polymerase. The DNA-TNA strand is separated from the DNA template and reverse transcribed with Bst. The cDNA is amplified utilizing the overhang region as a primer binding site, TOPO cloned and sequenced. Sequences containing the TT-AA watermark demonstrates a round of writing and reading TNA.*

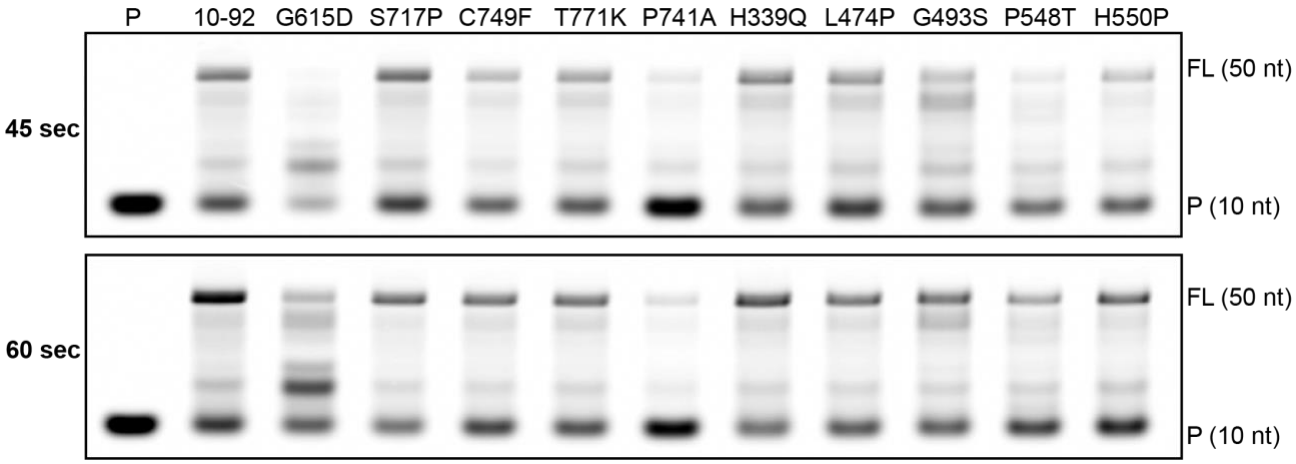


	A	C	G	C	G	C	T	T	C	T	T	G	G	A	A	A	C	T	C	A	C	G	C	A	C	C	C	A	T	A	A	T	A	A	T	A	C	T	A	A
Col 1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Col 2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.
Col 3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Col 4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Col 5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Col 6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Col 7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Col 8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Col 9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Col 10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Col 11	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Col 12	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Col 13	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Col 14	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Col 15	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Col 16	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Col 17	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Col 18	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Col 19	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Col 20	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Col 21	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Col 22	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Col 23	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Col 24	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Col 25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

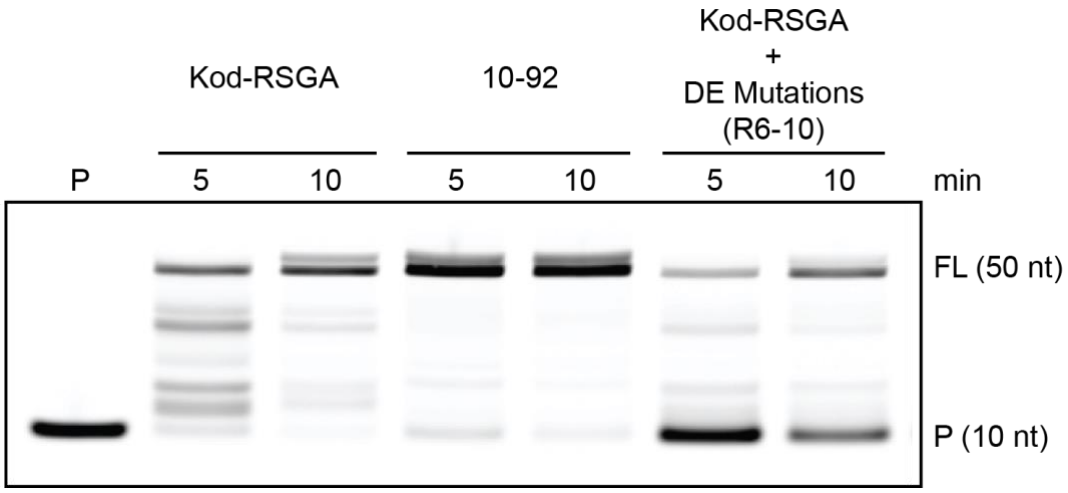
Supplementary Figure 12. Polymerase fidelity data. *Raw alignment of sangar sequencing results from a cycle of TNA replication by TNA polymerase 10-92 with standard base TNA triphosphates. The expected DNA sequence (colored) is written 5′→3′. Correct nucleotide reads are denoted as (.) and deleted nucleotides are highlighted and denoted as (-). From a study of 1000 nucleotide insertion events, the fidelity of the 10-92 TNA polymerase is 99.7% (0 misincorporations, 0 insertions, 3 deletions).*



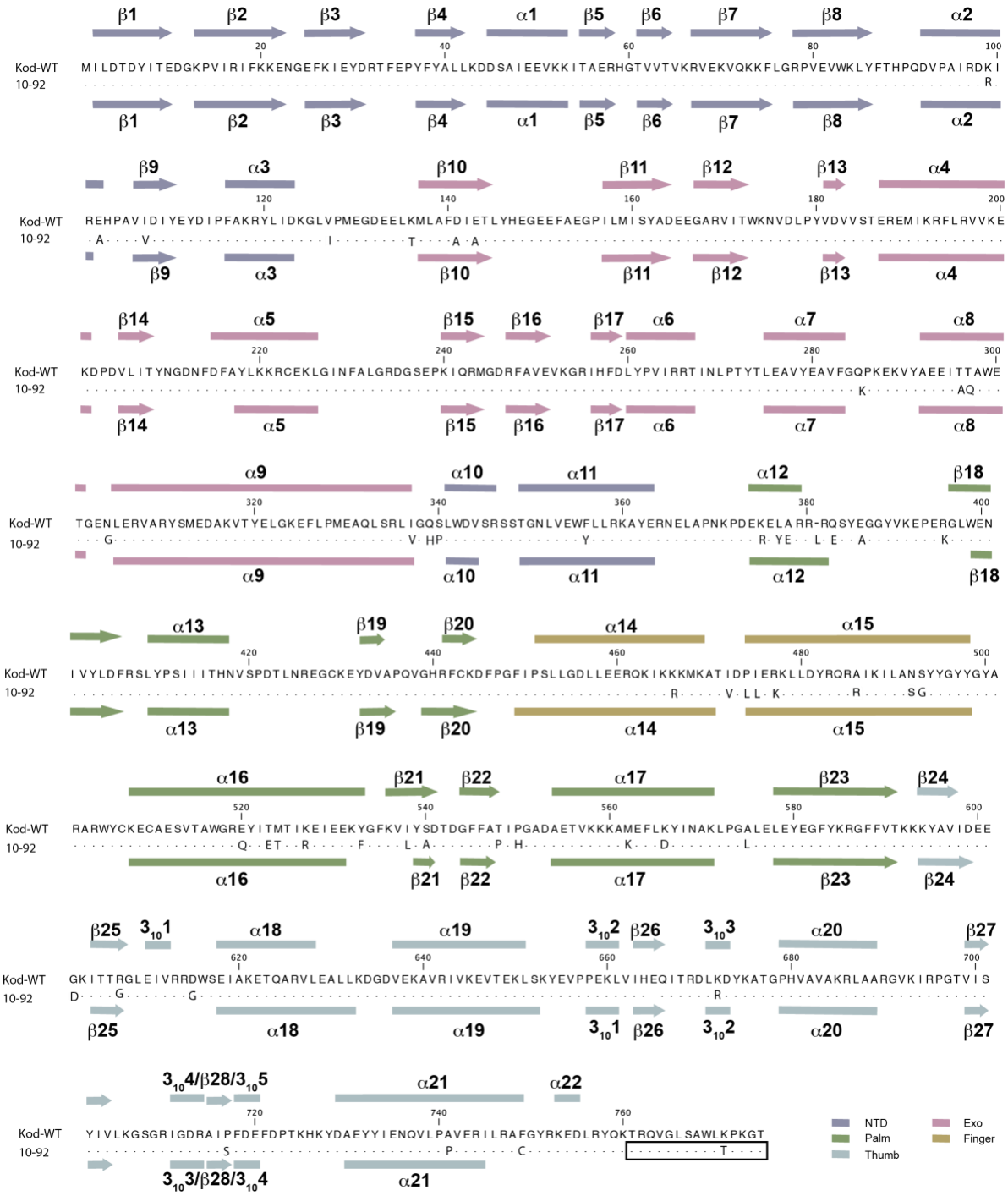
Supplementary Figure 13. SDS-PAGE analysis of 10-92 reversion proteins.
A single-point reversion study was performed in which each of the 10 mutations acquired by directed evolution of 5-270 to 10-92 were individually reverted back to their wild-type residue. The resulting constructs were grown in E. coli, expressed, purified by anion exchange chromatography, and validated with SDS-PAGE. Fractions from the 500 mM NaCl elution's were quantified (1.5 μg), and denatured at 95°C for 10 min in the presence of SDS-Loading Dye (1x) and DTT (16 mM) and run at 200V for 35 min. SDS-PAGE was run once to confirm the protein purity.



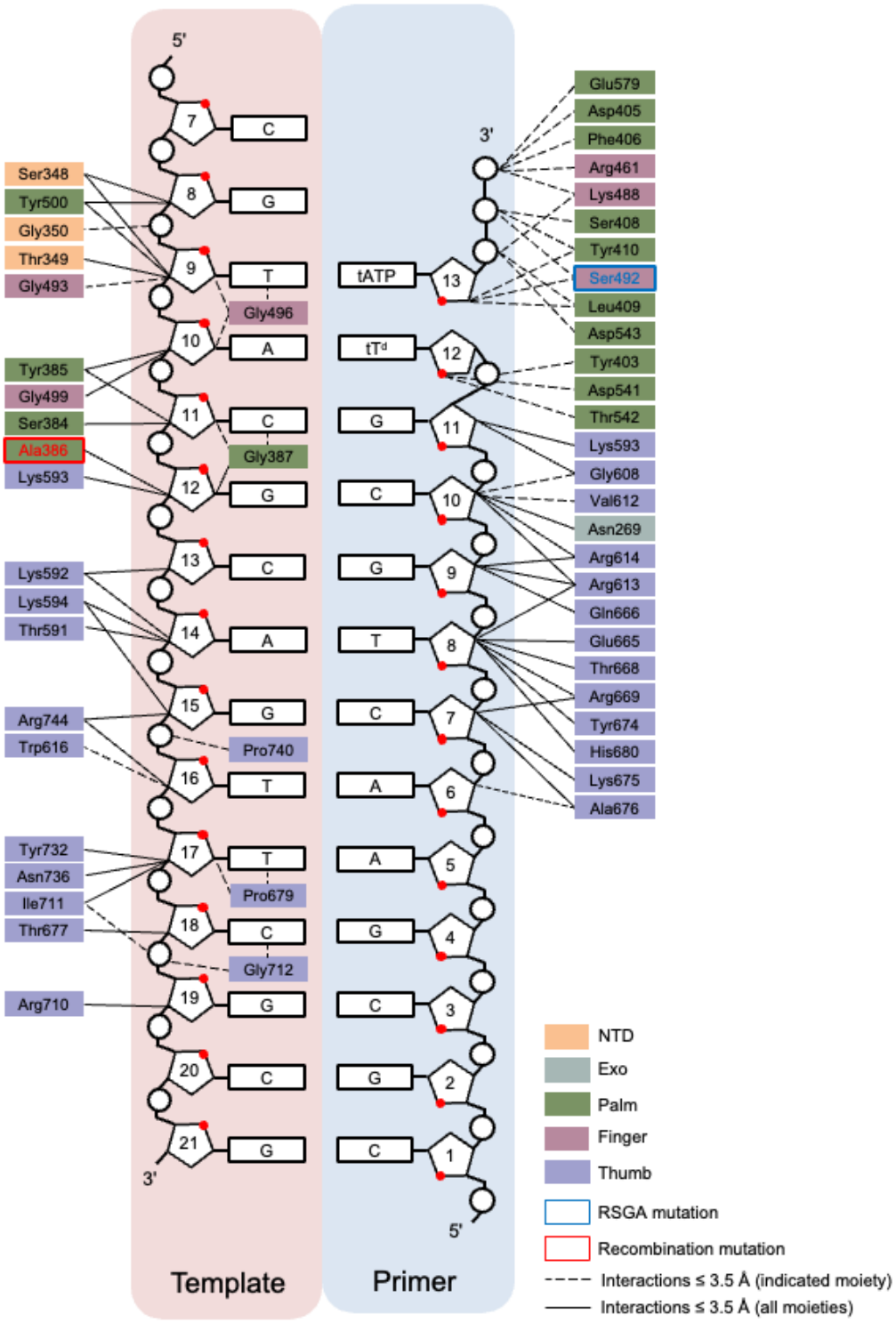
Supplementary Figure 14. TNA synthesis activity of 10-92 reversions.
Single point reversions introduced into the backbone of the 10-92 TNA polymerase. The mutations acquired in the evolution of 10-92 from 5-270 were reverted to the WT residue and are listed in the order they were discovered (left to right). Primer extension assays were performed using purified protein. The reactions were quenched at defined time points of 45 and 60 seconds and analyzed by denaturing polyacrylamide gel electrophoresis with fluorescent imaging on a LiCOR imager. P; primer, FL; full-length product. Representative gels of two independent experiments are shown.



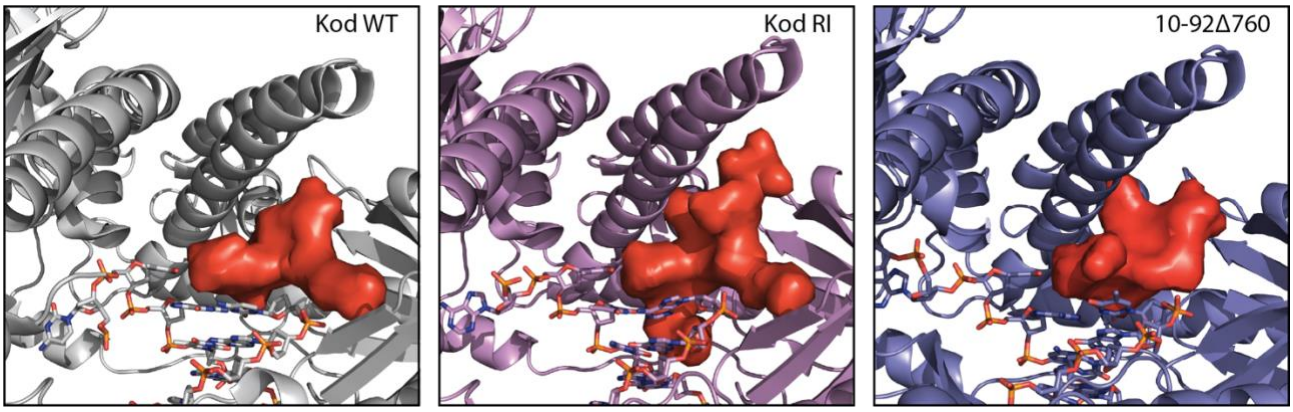
Supplementary Figure 15. Activity of Kod-RSGA with directed evolution mutations. *Kod-RSGA and 10-92 are compared to an engineered form of Kod-RSGA carrying the 10 beneficial mutations acquired by directed evolution that are found in the 10-92 sequence. Each protein was challenged to copy a DNA template by extending TNA substrates to a DNA primer strand within 5 or 10 min of incubation at 55°C. Reactions were analyzed by denaturing polyacrylamide gel electrophoresis with fluorescent imaging on a LiCOR imager. P; primer, FL; full-length product. A representative gel of two independent experiments is shown.*



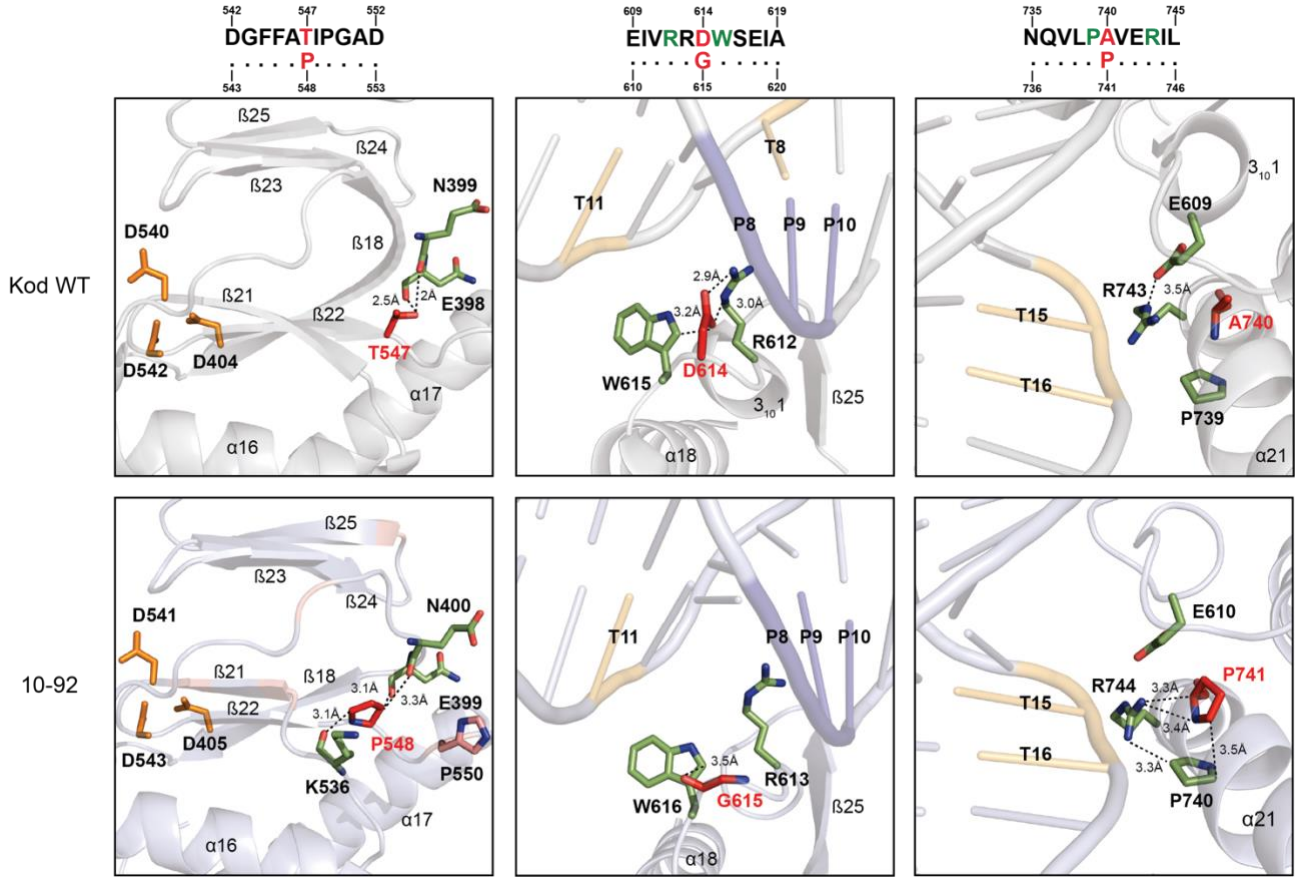
Supplementary Figure 16. Secondary structure map of the 10-92 TNA polymerase. *Secondary structure comparison of Kod-WT (top) and 10-92 (bottom), as determined by DSSP and PyMol. Matching residues are shown as dots. Residues removed for crystallography are boxed in black. Secondary structural elements are colored by domain.*



Supplementary Figure 17. Active site interaction map for the 10-92 TNA polymerase structure.



Supplementary Figure 18. Active site pocket volume. *Pocket volumes were calculated using PyVOL, searching all pockets over a minimum volume threshold (500Å) without partitioning. Probe min and max radii were set to 1.4 and 2.8, respectively. Metals, incoming tNTP triphosphates, and solvents were removed for measurement. The pocket of Kod-WT DNA polymerase (PDB: 5OMF), Kod-RI TNA polymerase (PDB: 5VU8), and 10-92Δ760 TNA polymerase (PDB: 8T3X) measure at 746 Å^3 ,1018 Å^3, and 684 Å^3, respectively*.



Supplementary Figure 19. Structural analysis of key reversion mutations. *Structural comparisons of three directed evolution mutations, Thr547 to Pro548 (left), Asp614 to Gly615 (middle), and Ala740 to Pro741 (right) between Kod-WT (top) and 10-92 (bottom) closed ternary complexes. Mutation residues are colored red while residues affected by the mutations are colored green. Mutations in close proximity to Pro548 are colored light pink. Catalytic residues are depicted in orange. Nucleotides in the primer and template strands that interact with residues are colored lilac and salmon, respectively.*

References for Appendix A

1 Sau, S. P., Fahmi, N. E., Liao, J.-Y., Bala, S. & Chaput, J. C. A scalable synthesis of α-L-threose nucleic acid monomers. *J. Org. Chem.* 81, 2302-2307, (2016).

2 Liao, J.-Y., Bala, S., Ngor, A. K., Yik, E. J. & Chaput, J. C. P(V) Reagents for the Scalable Synthesis of Natural and Modified Nucleoside Triphosphates. *J. Am. Chem. Soc.* 141, 13286-13289, (2019).

3 Vallejo, D., Nikoomanzar, A., Paegel, B. M. & Chaput, J. C. Fluorescence-Activated Droplet Sorting for Single-Cell Directed Evolution. *ACS Synth. Biol.* 8, 1430-1440, (2019).

4 Vallejo, D., Nikoomanzar, A. & Chaput, J. C. Directed evolution of custom polymerases using droplet microfluidics. *Methods Enzymol.* 644, 227-253, (2020).

5 Nikoomanzar, A., Dunn, M. R. & Chaput, J. C. Engineered polymerases with altered substrate specificity: expression and purification. *Curr. Protoc. Nucleic Acid Chem.* 69, 4.75, (2017).

6 Cadwell, R. C. & Joyce, G. F. Randomization of genes by PCR mutagenesis. *PCR Methods Appl.* 2, 28-33, (1992).

7 Chaput, J. C. & Szostak, J. W. Evolutionary optimization of a nonbiological ATP binding protein for improved folding stability. *Chem. Biol.* 11, 865-874, (2004).

8 Nikoomanzar, A., Vallejo, D., Yik, E. J. & Chaput, J. C. Programmed allelic mutagenesis of a DNA polymerase with single amino acid resolution. *ACS Synthic Biology* 9, 1873-1881, (2020).

9 Li, Q. *et al.* Synthesis and Polymerase Recognition of Threose Nucleic Acid Triphosphates Equipped with Diverse Chemical Functionalities. *J. Am. Chem. Soc.* 143, 17761-17768, (2021).

10 McCloskey, C. M. *et al.* Evolution of Functionally Enhanced alpha-l-Threofuranosyl Nucleic Acid Aptamers. *ACS Synth. Biol.* 10, 3190-3199, (2021).

11 Nikoomanzar, A., Vallejo, D. & Chaput, J. C. Elucidating the Determinants of Polymerase Specificity by Microfluidic-Based Deep Mutational Scanning. *ACS Synth. Biol.* 8, 1421-1429, (2019).

12 Nikoomanzar, A., Dunn, M. R. & Chaput, J. C. Evaluating the rate and substrate specificity of laboratory evolved XNA polymerases. *Anal. Chem.* 89, 12622-12625, (2017).

13 Medina, E., Yik, E. J., Herdewijn, P. & Chaput, J. C. Functional Comparison of Laboratory-Evolved XNA Polymerases for Synthetic Biology. *ACS Synth. Biol.* 10, 1429-1437, (2021).

14 Kabsch, W. Xds. *Acta Crystallogr. D Biol. Crystallogr.* 66, 125-132, (2010).

15 Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* 66, 133-144, (2010).

16 McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658-674, (2007).

17 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486-501, (2010).

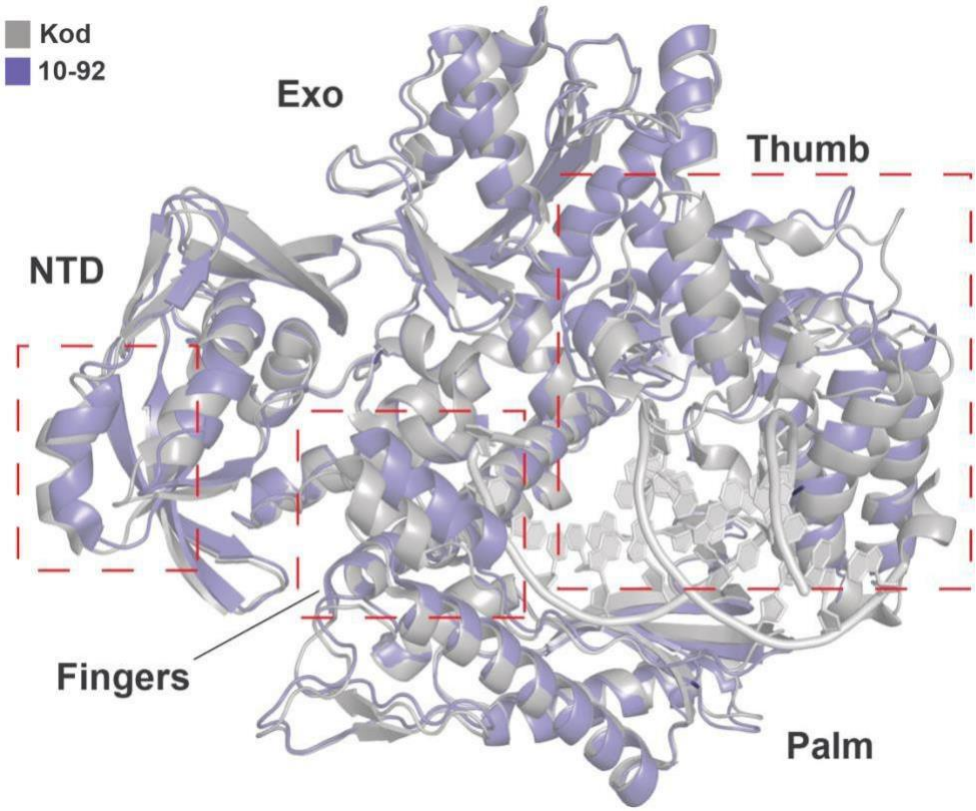
18 Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* 68, 352-367, (2012).

19 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 66, 12-21, (2010).

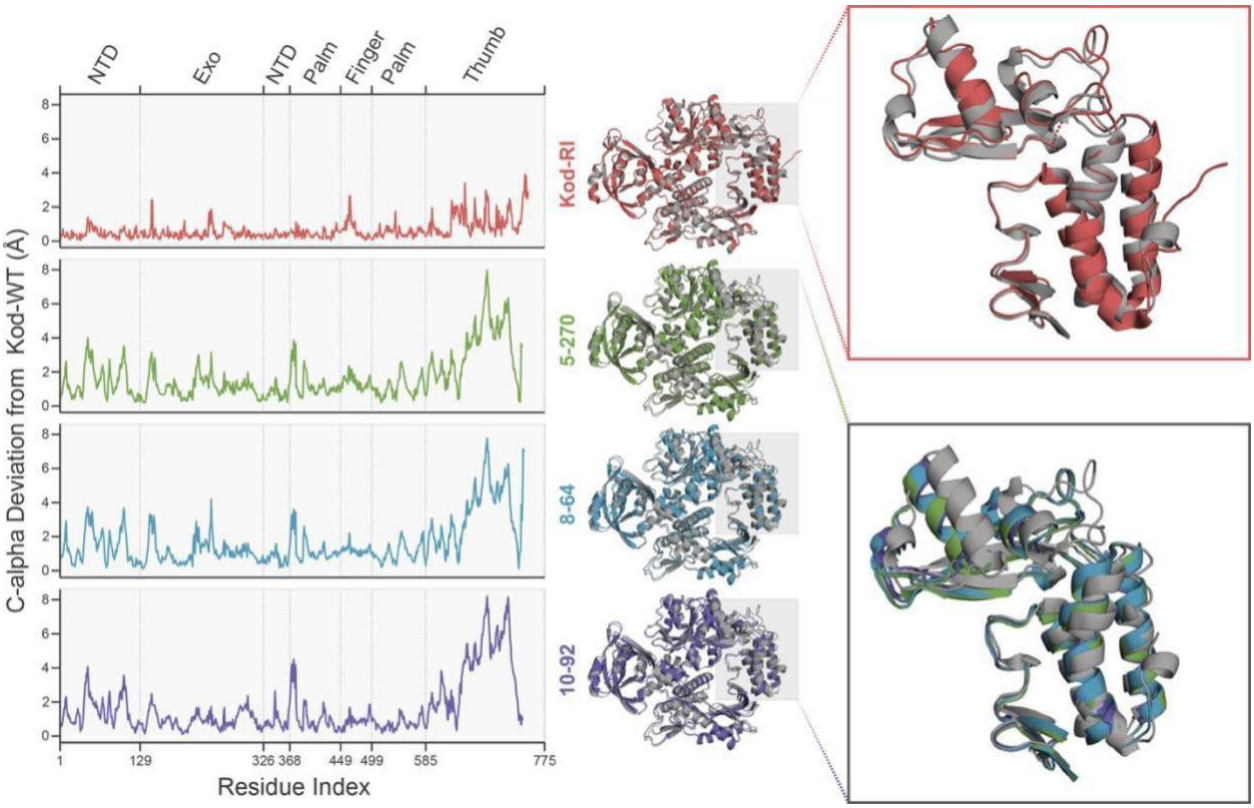
20 The PYMOL molecular graphics program (DeLano Scientific, San Carlos, CA, USA, 2002).

21 Zheng, G., Lu, X. J. & Olson, W. K. Web 3DNA--a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res.* 37, W240-246, (2009).

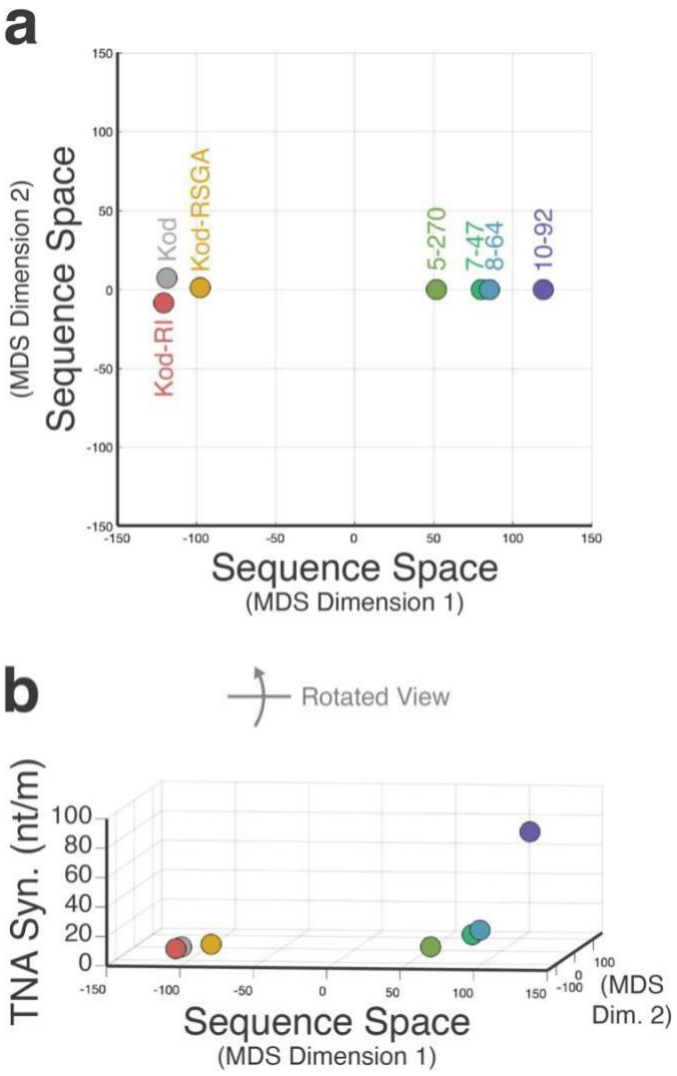
APPENDIX B: Supplementary Data of Chapter 3



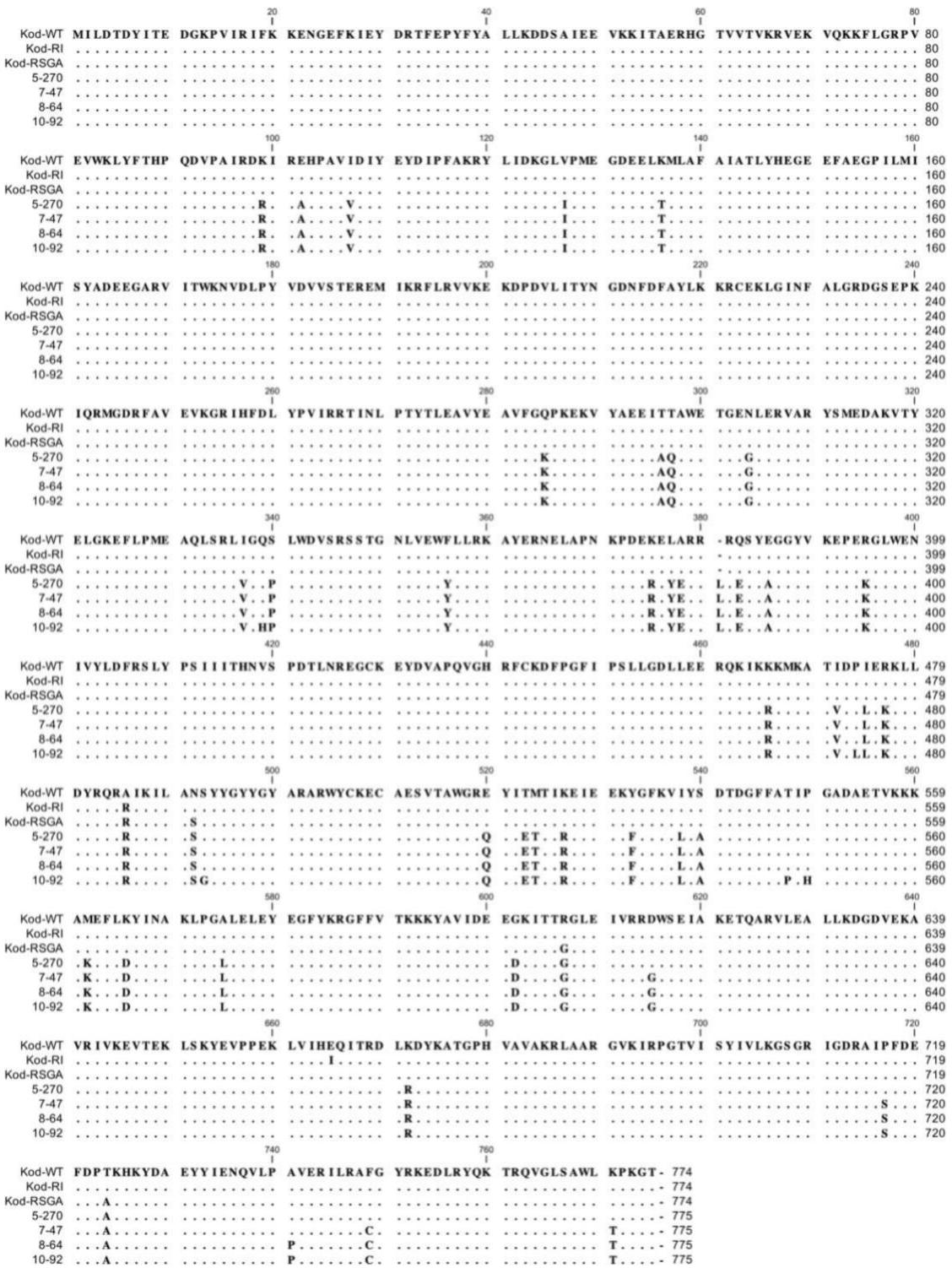
Supplementary Figure 1*. *Structural overlay of Kod and 10-92 closed ternary conformations. *Red dashed boxes depict areas of large conformational difference between Kod (transparent grey, PDB: 5OMF) and 10-92 (transparent purple, PDB: 8T3X). For clarity, only the Kod-bound duplex is displayed.*



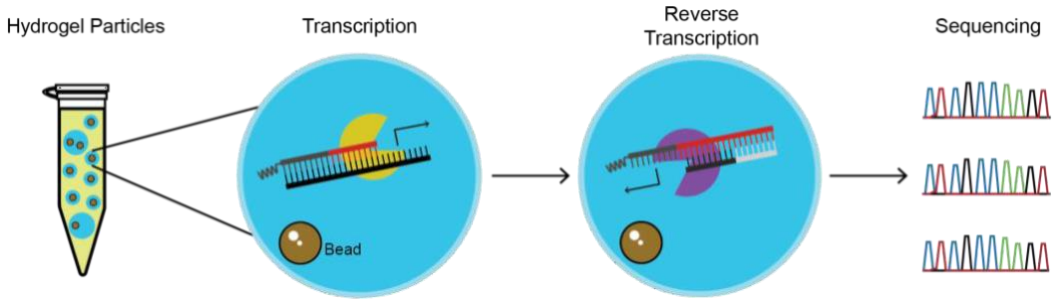
Supplementary Figure 2. ***Conformational rearrangement of the thumb subdomain.*** *Left, the distance in angstroms from all resolved C-α positions in the ternary structures of Kod-RI (red), 5-270 (green), 8-64 (blue), and 10-92 (purple) relative to their position in Kod. Middle, structural overlays of TNAPs aligned to Kod (gray). Right, magnified view of the thumb subdomain showing the structural similarity between Kod-RI and Kod and structural differences between 5-270, 8-64, and 10-92 relative to Kod.*



Supplementary Figure 3. Sequence homology-based fitness landscape of TNAP variants. a*, Two-dimensional representation of polymerase relationships derived from pairwise sequence alignment scores calculated with the EMBOSS Needle global alignment tool. Pairwise scores were assembled into a distance matrix and analyzed by classical multidimensional scaling (MDS). Each circle represents a polymerase variant, with closer positions indicating higher sequence homology.* ***b****, Rotated view of the same projection with an additional z-axis corresponding to TNA synthesis activity (nt min^{-1}) as reported in Supplementary Figure 7.*

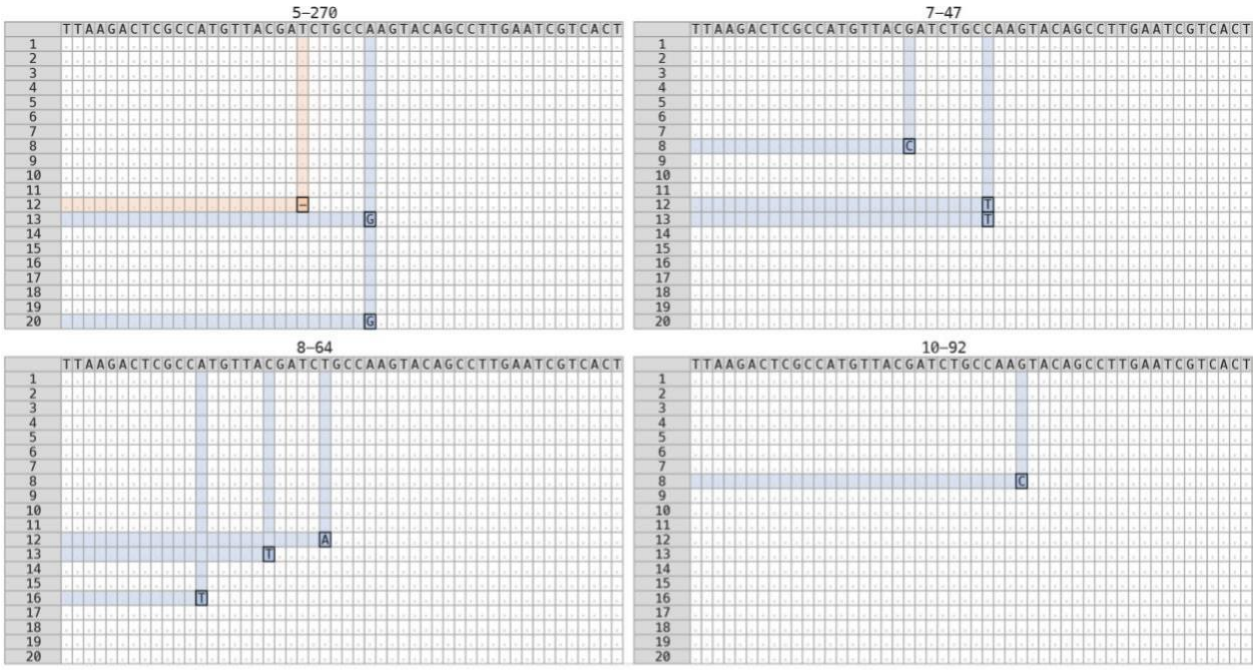


Supplementary Figure 4. Sequence alignment of TNAP generations. *Conserved residues are denoted as (.), while mutations are listed as single letter amino acid abbreviations. Alignment was generated in CLC Main Workbench.*



***Supplementary Figure 5**. **Workflow for TNAP fidelity measurements in hydrogel particles.** A single-stranded DNA template is hybridized to a DNA primer crosslinked to the hydrogel matrix and extended with a TNAP and tNTPs for 1 hour at 55°C. The DNA template is then stripped to display the TNA product strand. The displayed TNA strand is hybridized to an overhang DNA primer and reverse transcribed with Bst DNA polymerase and dNTPs for 4 hours at 50°C. cDNA is amplified by PCR, TOPO cloned, and sequenced. TT mismatches in the primer and template are used to confirm that each sequence passed through a round of TNA replication (DNA→TNA→DNA).*

a



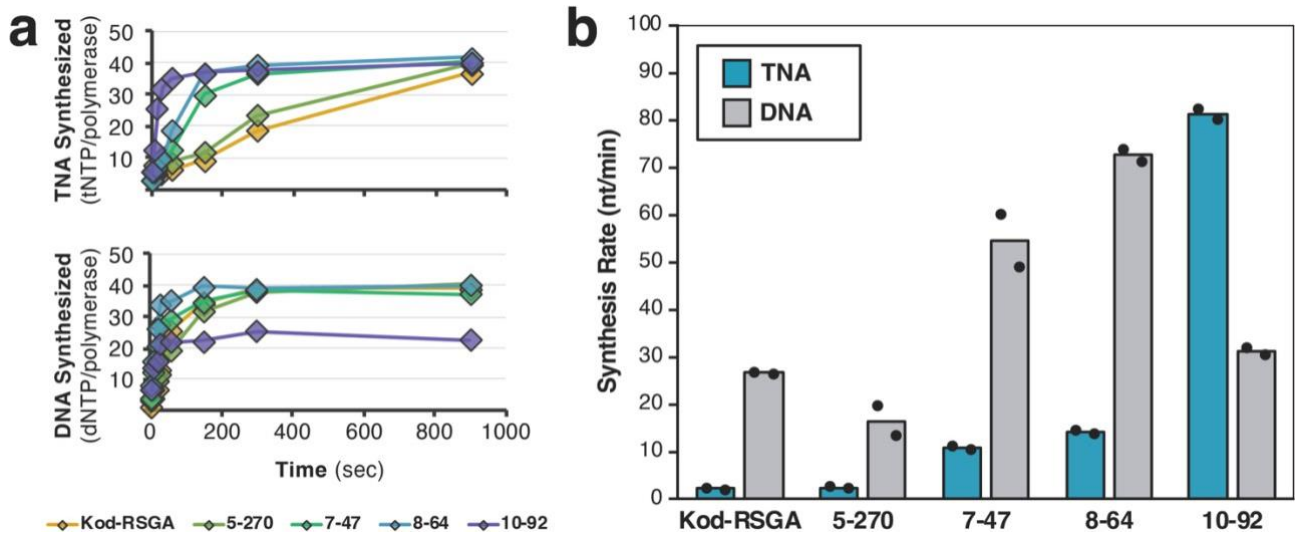
b

Experiment		(Substitution)			(Substitution + In/Del)		
TNAP	Reads (nt)	Errors	Error Rate	Fidelity (%)	Errors	Error Rate	Fidelity (%)
Kod–RSGA	1677	16	0.010	99.046	24	0.014	98.569
5–270	1000	2	0.002	99.800	3	0.003	99.700
7–47	1000	3	0.003	99.700	3	0.003	99.700
8–64	1000	3	0.003	99.700	3	0.003	99.700
10–92	1000	1	0.001	99.900	1	0.001	99.900

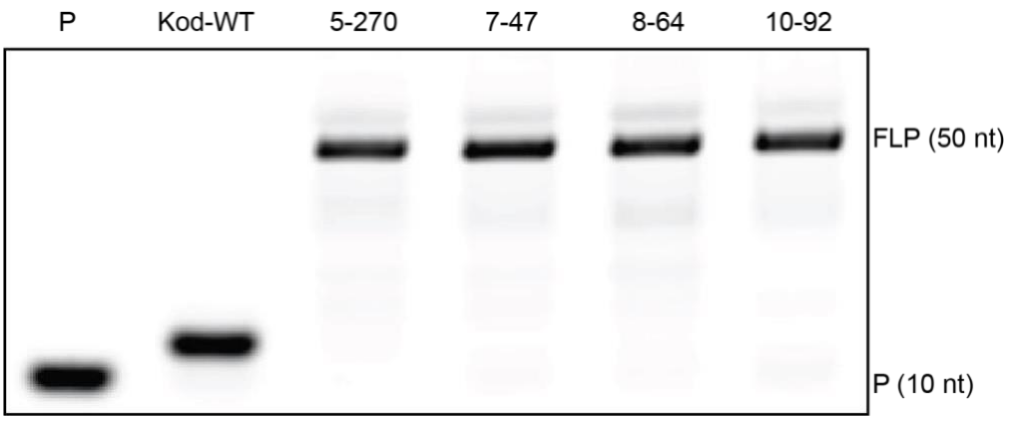
Binomial Two–Proportion Z–Test (Substitution)				
Comparison	$\Delta_{Fidelity}$ (pp)	z	p–Value	Statistacal Significance
Kod–RSGA → 5–270	0.754	2.309	0.0209	Significant (p<0.05)
5–270 → 7–47	–0.100	–0.448	0.6543	Not Significant
7–47 → 8–64	0.000	0.000	1.0000	Not Significant
8–64 → 10–92	0.200	1.001	0.3168	Not Significant

Binomial Two–Proportion Z–Test (Substitution + In/Del)				
Comparison	$\Delta_{Fidelity}$ (pp)	z	p–Value	Statistacal Significance
Kod–RSGA → 5–270	1.131	2.833	0.0046	Significant (p<0.05)
5–270 → 7–47	0.000	0.000	1.0000	Not Significant
7–47 → 8–64	0.000	0.000	1.0000	Not Significant
8–64 → 10–92	0.200	1.001	0.3168	Not Significant

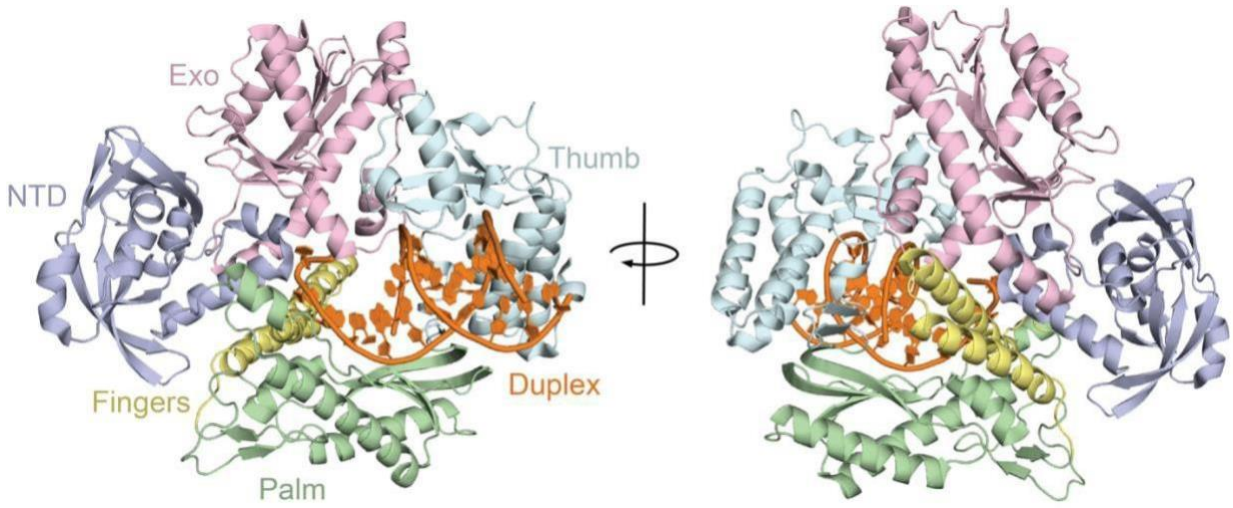
Supplementary Figure 6. TNAP fidelity raw data and statistical analysis. *(a) Raw Sanger sequencing alignments from a single cycle of TNA replication by the indicated TNAP variants. The expected DNA sequence is shown 5′→3′. Correct base incorporations are marked with (.), and substitutions or insertion/deletion events are outlined in black boxes with guiding blue or orange highlighting. (b) Fidelity values represent the aggregate accuracy determined from at least 1,000 nucleotides per enzyme, calculated both with and without indels. Statistical significance of fidelity changes between successive TNAP generations was assessed using a binomial z-test to show the significance between Kod-RSGA and 5-270.*



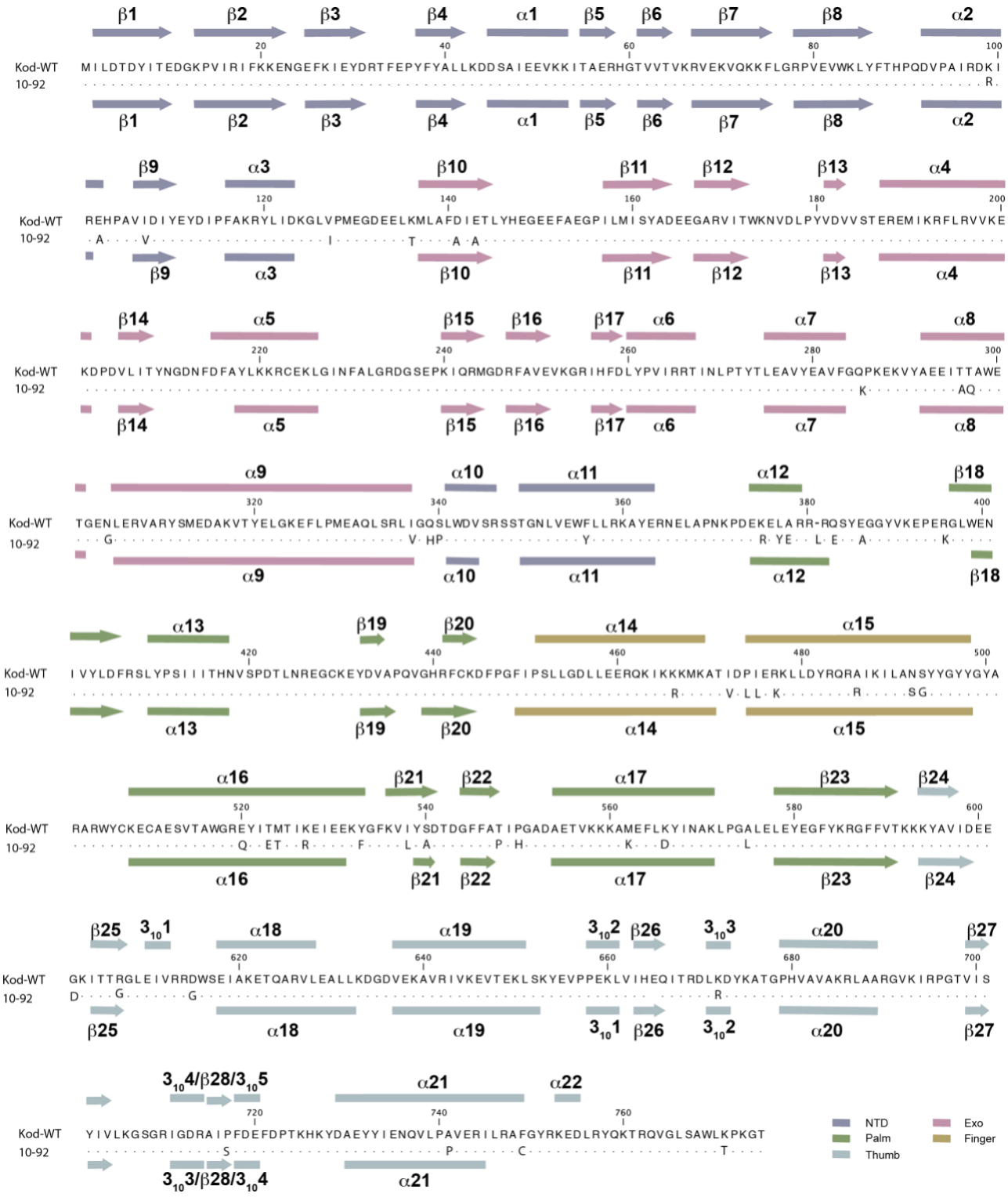
Supplementary Figure 7. TNAP kinetics and substrate specificity. a*, Kinetic time-course of primer-extension reactions catalyzed by Kod-RSGA, 5-270, 7-47, 8-64, and 10-92 using TNA (tNTP) or DNA (dNTP) triphosphates, average of n=2. The slope of the linear range corresponds to the rate of synthesis and is plotted in panel* ***b*** *as nucleotides per minute.*



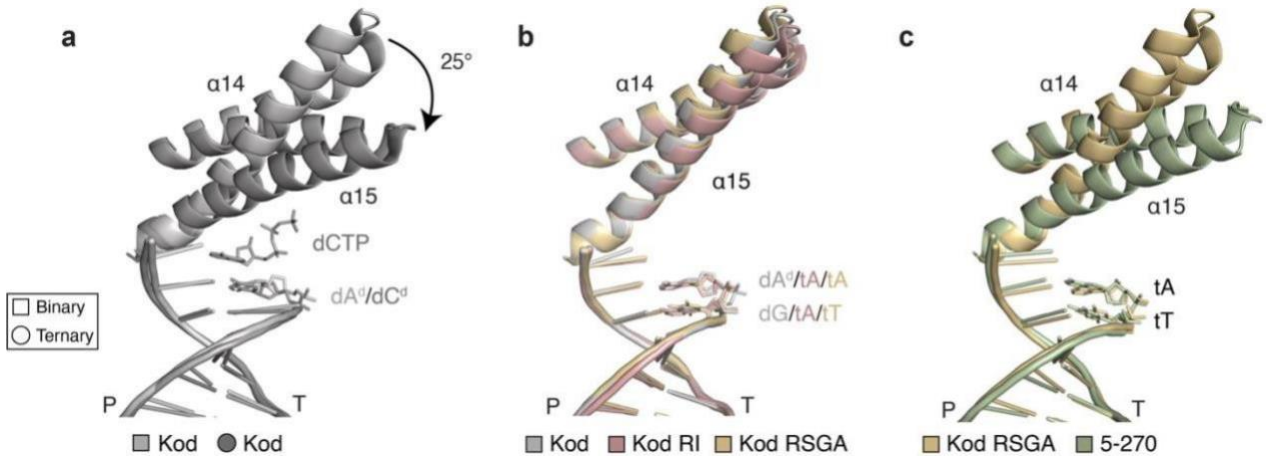
Supplementary Figure 8. TNAP thermal challenge*. TNA synthesis activity was assessed after exposing the polymerases to 6 hours of incubation at 90°C. Reactions were performed by incubating a 5' IR680-labelled DNA primer–template duplex (1 μM) with 1 μM polymerase and 100 μM tNTPs in 1x ThermoPol buffer for 30 minutes at 55°C. P, primer; FLP, full-length product.*



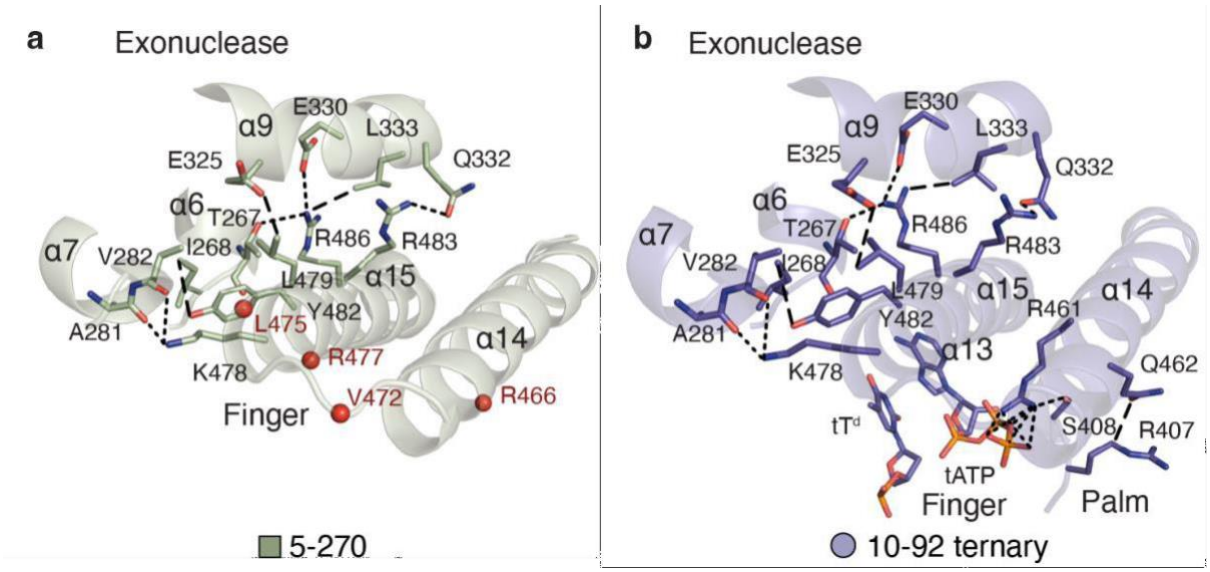
Supplementary Figure 9. Polymerase domains. *The architecture of B-family polymerases colored by domain, mapped on the closed ternary structure of wildtype Kod (PDB ID: 5OMF).*



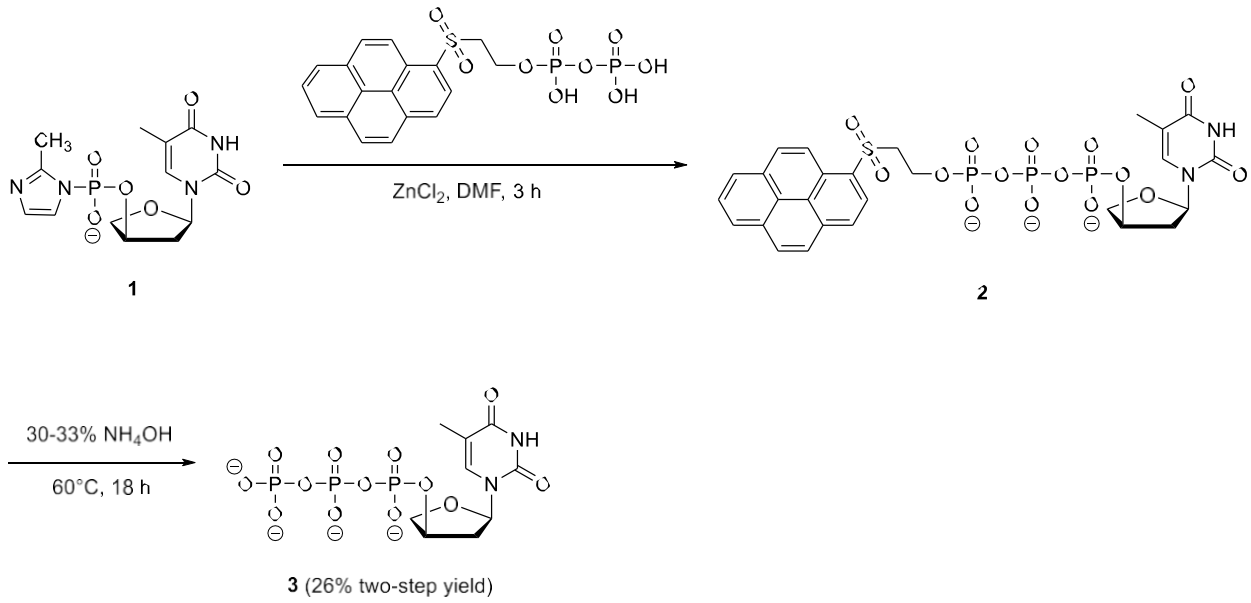
Supplementary Figure 10. Sequence and secondary structure alignment of Kod and 10-92. *Conserved residues are denoted as (.), while mutations are listed as single letter amino acid abbreviations. Secondary structures of Kod and 10-92, as determined by DSSP and PyMol, are depicted above and below the sequence alignment, respectively. Secondary structural elements are colored by domains, which are color-matched to Supplementary Figure 9.*



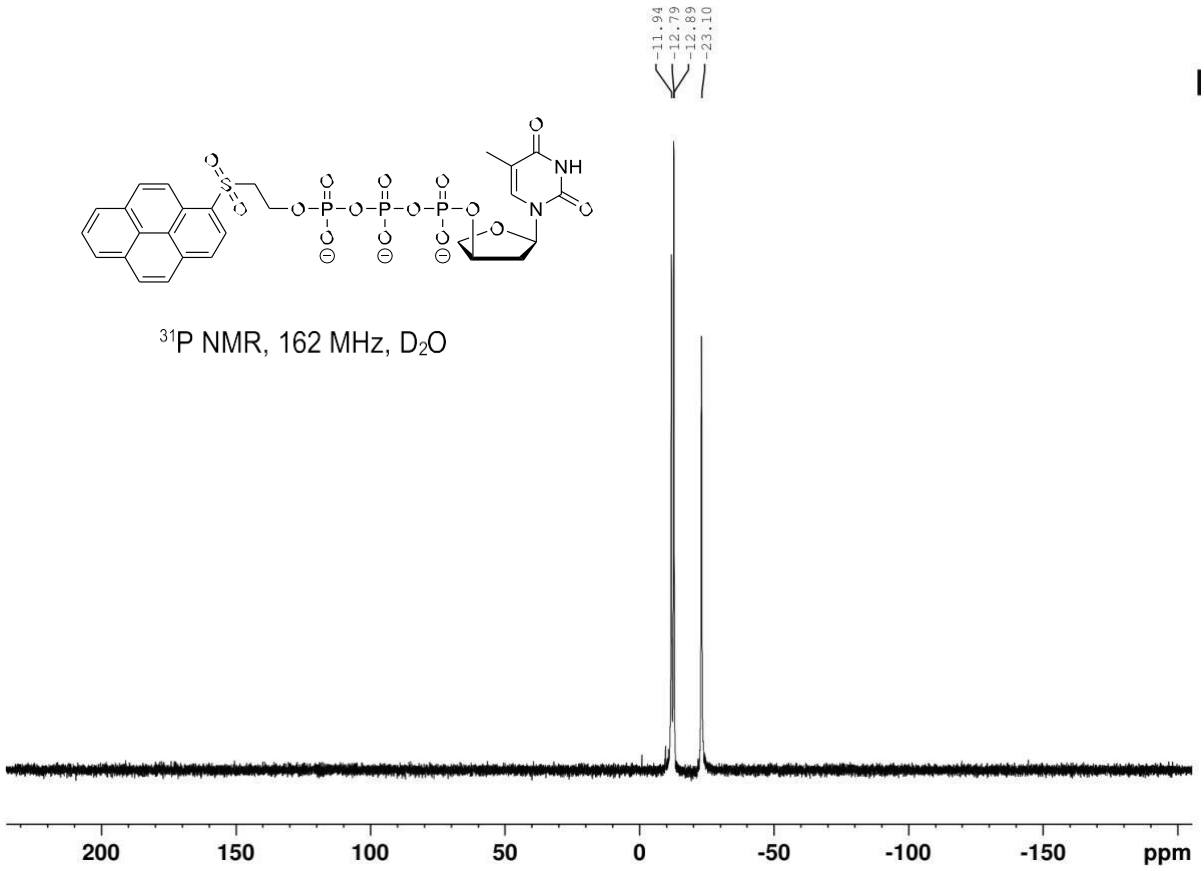
Supplementary Figure 11. Structural comparison of the finger subdomain of Kod, Kod-RI, Kod-RSGA, and 5-270 in the binary complex. ***a****, Overlay of the finger subdomains of Kod binary (light gray, PDB: 4K8Z) and closed ternary (dark gray, PDB: 7OMB) complexes.* ***b****, Overlay of the finger subdomains of Kod binary (light gray), Kod-RI binary (pink, PDB: 5VU9), and Kod-RSGA binary (yellow, PDB: 7RSU) complexes.* ***c****, Overlay of the finger subdomains of Kod RSGA and 5-270 (green) binary complexes. Structural overlays reveal the binary complex of 5-270 is structurally distinct from the binary complex of Kod, Kod-RI, and Kod-RSGA.*



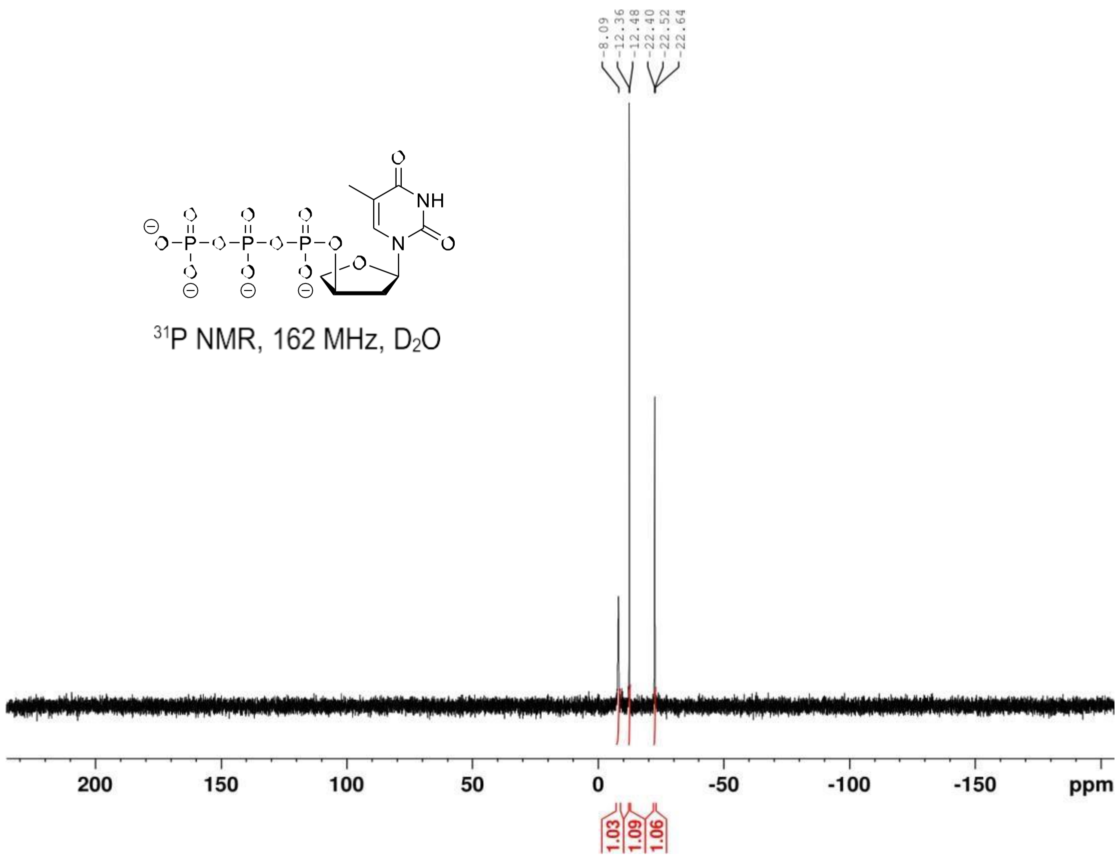
Supplementary Figure 12. Active site comparison. *The active site of* ***a****, the 5-270 binary structure reveals a closed finger conformation similar to* ***b****, the 10-92 closed ternary structure. Short dashed lines denote electrostatic interactions, long dashed lines show hydrophobic interactions. Residue labels: natural residues found in Kod DNAP (black); acquired mutations (red spheres, labels).*



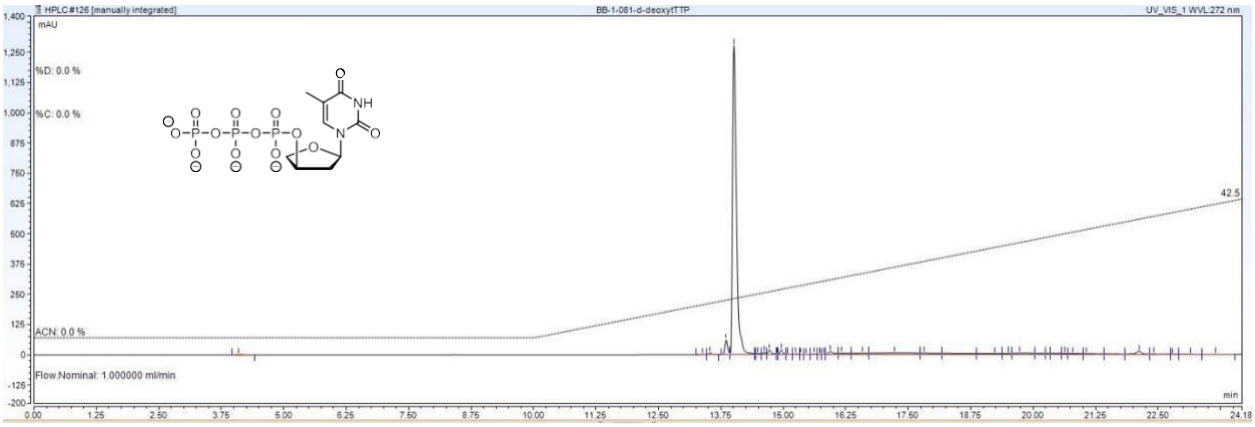
Supplementary Figure 13. ***Synthesis of 2'-deoxy-α-L-threofuranosyl thymine triphosphate (dtTTP) by pyrene pyrophosphate method.*** *Modified synthesis of 2'-deoxy-α-L-threofuranosyl thymine triphosphate (dtTTP) using pyrene pyrophosphate to incorporate the β- and γ-phosphates*[2]. *The synthesis of compound 1 was reported previously*[3].



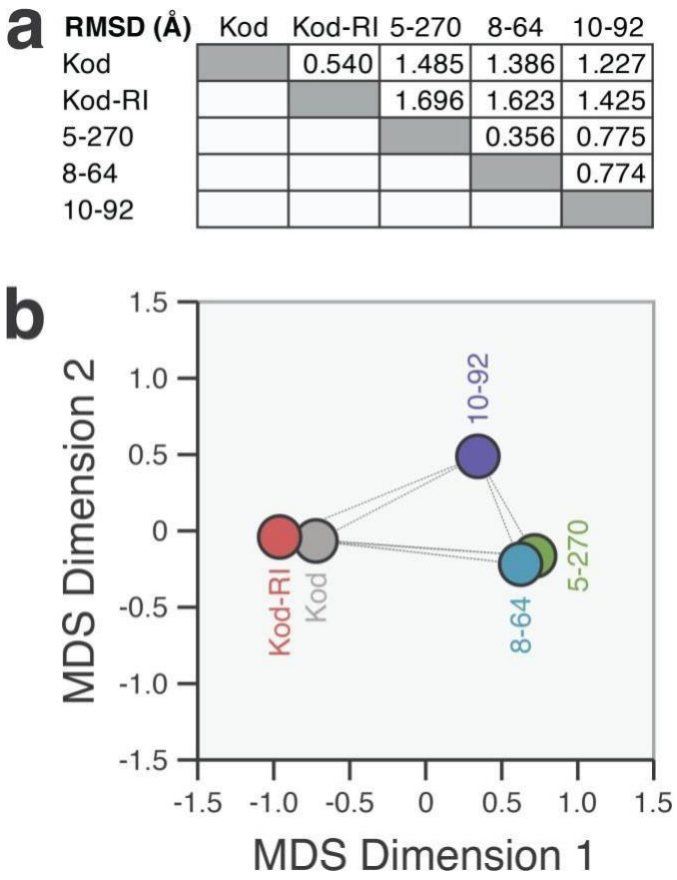
Supplementary Figure 14. NMR spectrum. ^{31}P NMR spectrum of compound 2.



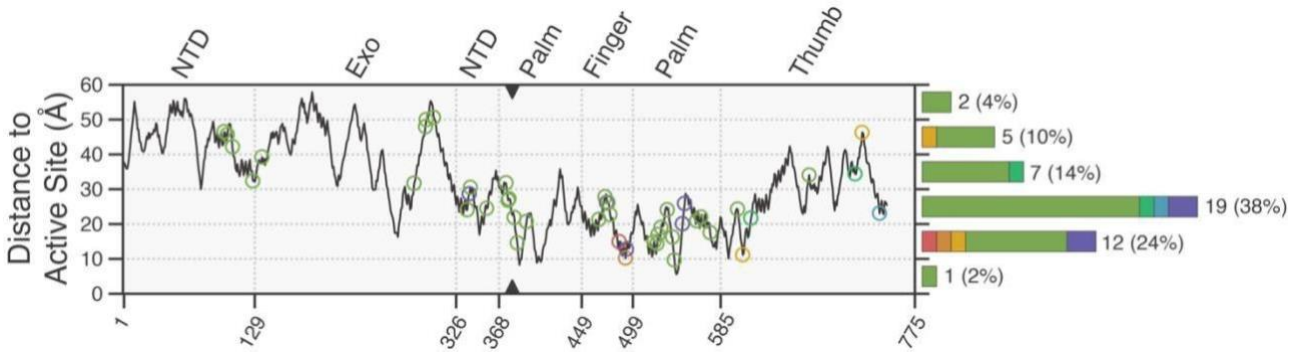
Supplementary Figure 15. NMR spectrum. 31P NMR spectrum of compound 3.



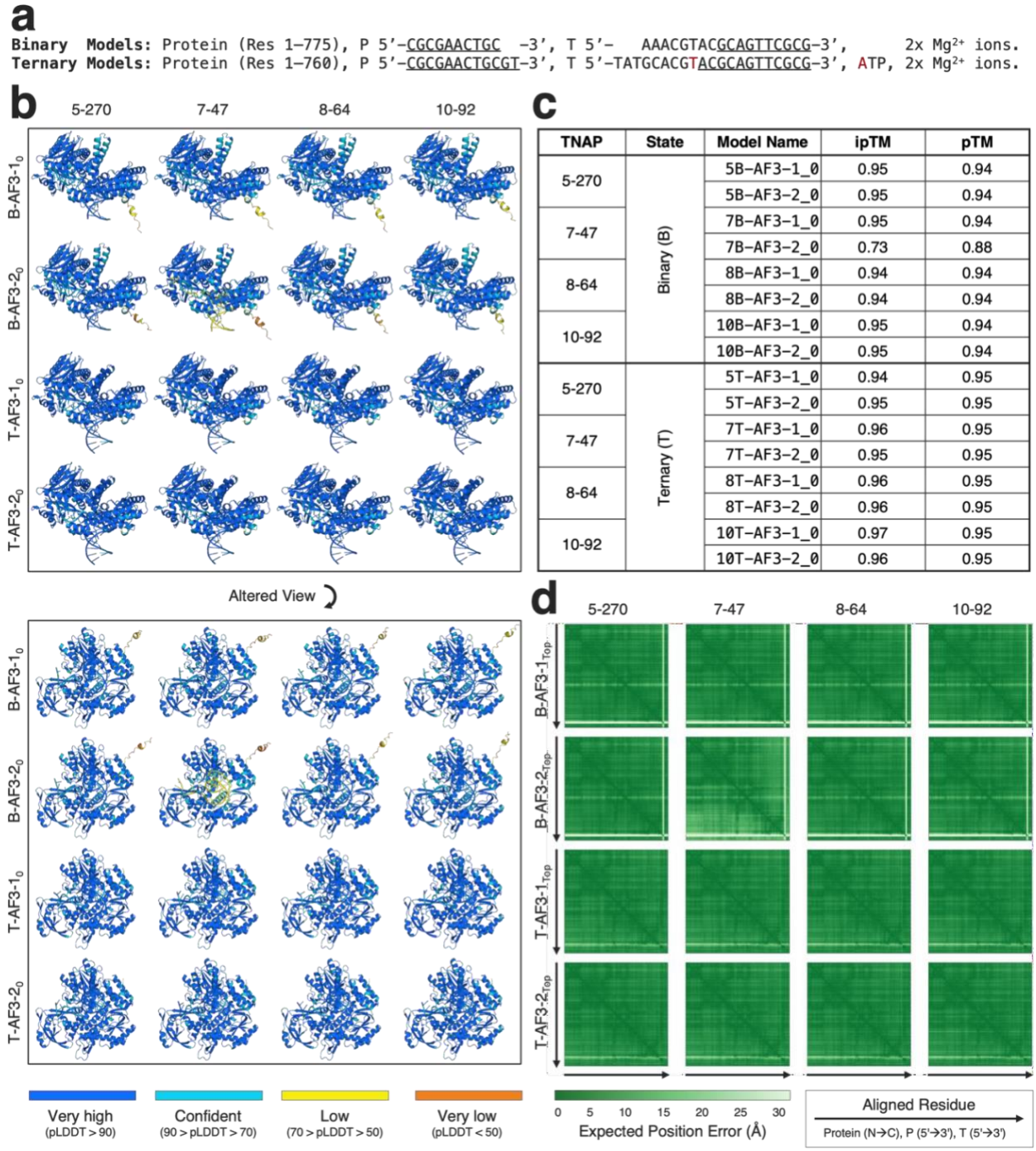
Supplementary Figure 16. HPLC spectrum. HPLC analysis of compound 3.



Supplementary Figure 17. Multidimensional analysis of the closed ternary complex of TNAP structures based on RMSD. *(a) Pairwise root mean square deviation (RMSD, Å) values for ternary complexes, calculated over Cα atoms following global alignment in PyMOL. Lower RMSD values indicate greater structural similarity. (b) Two-dimensional projection of polymerase relationships derived from the RMSD matrix using classical multidimensional scaling (MDS; double centering + eigendecomposition) in Python (v3.11) using NumPy. Each polymerase is shown as a colored circle. Shorter distances between points on the plot correspond to higher similarity in three-dimensional structure.*

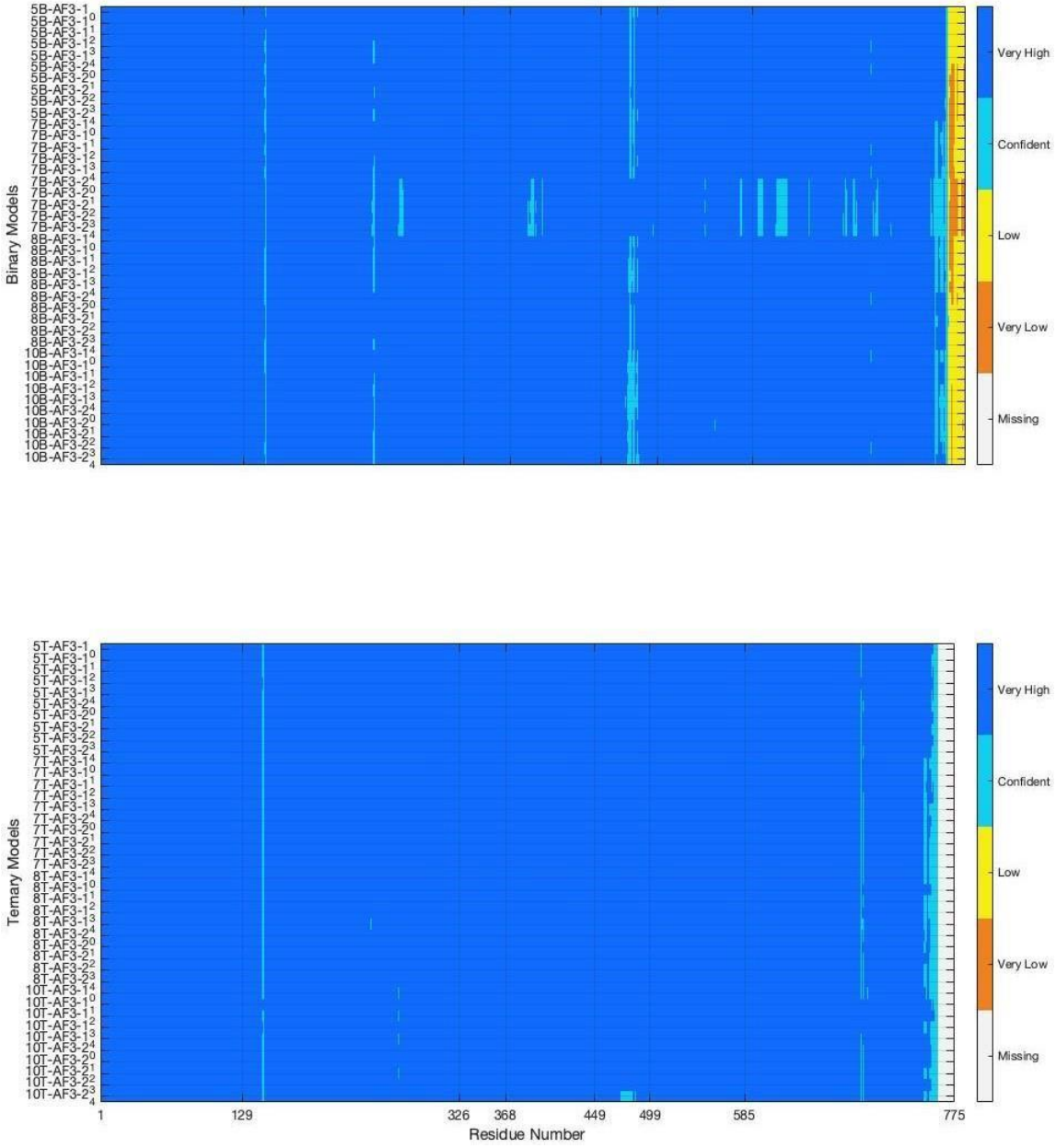


Supplementary Figure 18. Distance of mutations to active site. *The distance between the alpha-carbon of each residue in the 10-92 ternary structure to the active site (C2' of the terminal primer TNA residue) is measured in angstroms. Mutations compared to Kod are indicated as open circles and color-matched to Figure 1, with total counts provided in 10 Å shells.*

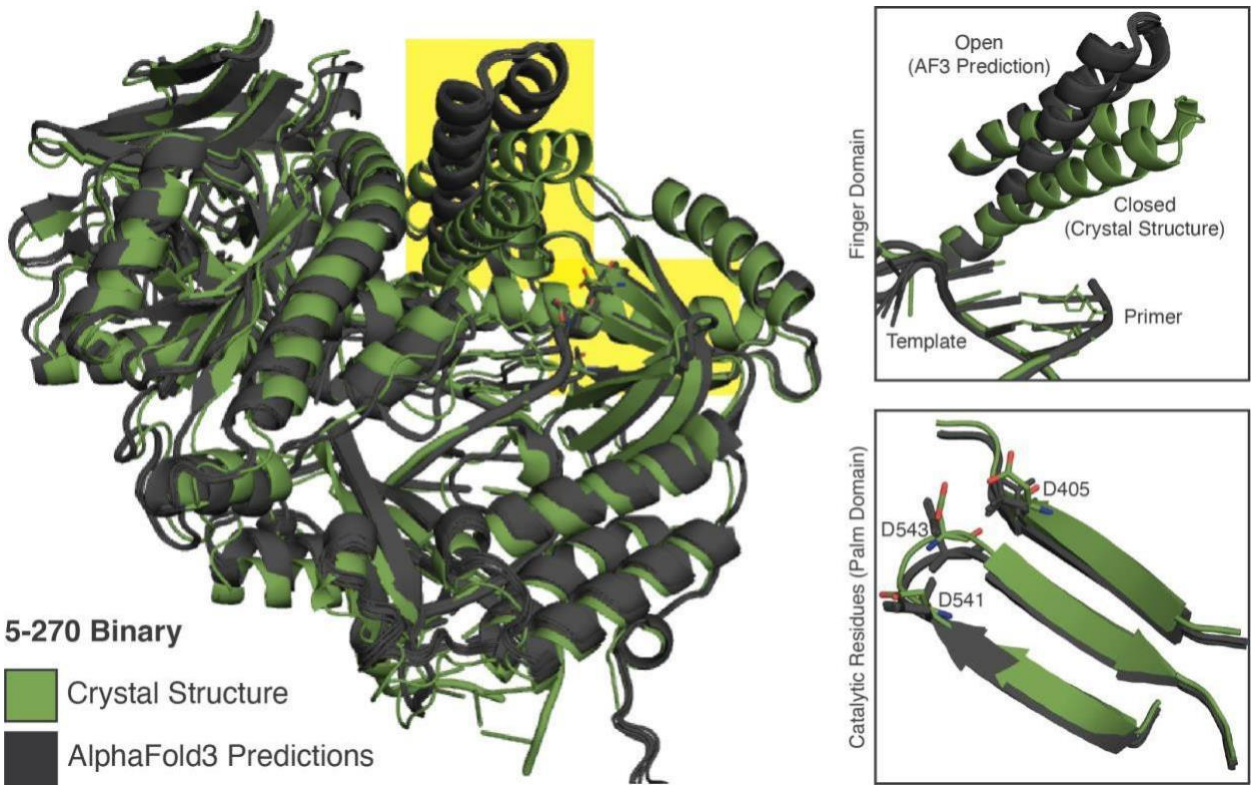


TNAP	State	Model Name	ipTM	pTM
5-270	Binary (B)	5B-AF3-1_0	0.95	0.94
		5B-AF3-2_0	0.95	0.94
7-47		7B-AF3-1_0	0.95	0.94
		7B-AF3-2_0	0.73	0.88
8-64		8B-AF3-1_0	0.94	0.94
		8B-AF3-2_0	0.94	0.94
10-92		10B-AF3-1_0	0.95	0.94
		10B-AF3-2_0	0.95	0.94
5-270	Ternary (T)	5T-AF3-1_0	0.94	0.95
		5T-AF3-2_0	0.95	0.95
7-47		7T-AF3-1_0	0.96	0.95
		7T-AF3-2_0	0.95	0.95
8-64		8T-AF3-1_0	0.96	0.95
		8T-AF3-2_0	0.96	0.95
10-92		10T-AF3-1_0	0.97	0.95
		10T-AF3-2_0	0.96	0.95

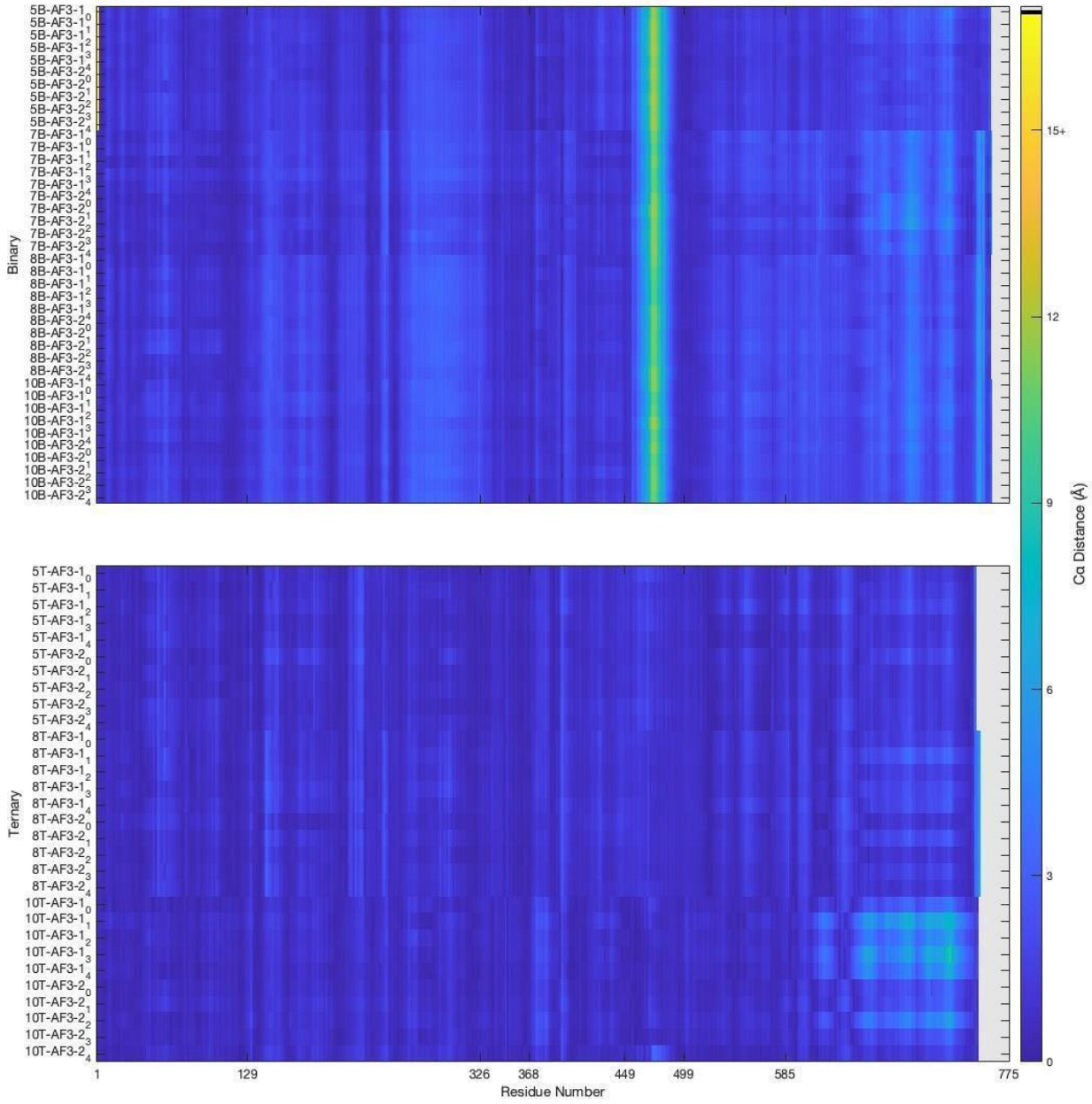
Supplementary Figure 19. AlphaFold3 prediction overview. *(**a**) Input specifications for AlphaFold3 binary and ternary complex predictions. Two independent predictions were run for each structure (denoted AF3-1 and AF3-2) with each seed yielding five models (0–4). (**b**) Representative predicted models (model 0) shown in two orientations and colored by per-residue predicted local distance difference test (pLDDT) confidence scores. (**c**) Summary table of predicted interfacial template modeling score (ipTM) and global template modeling score (pTM) for each representative model. (**d**) Predicted alignment error (PAE) matrices shown in Ångströms for the top-ranked model of each run.*



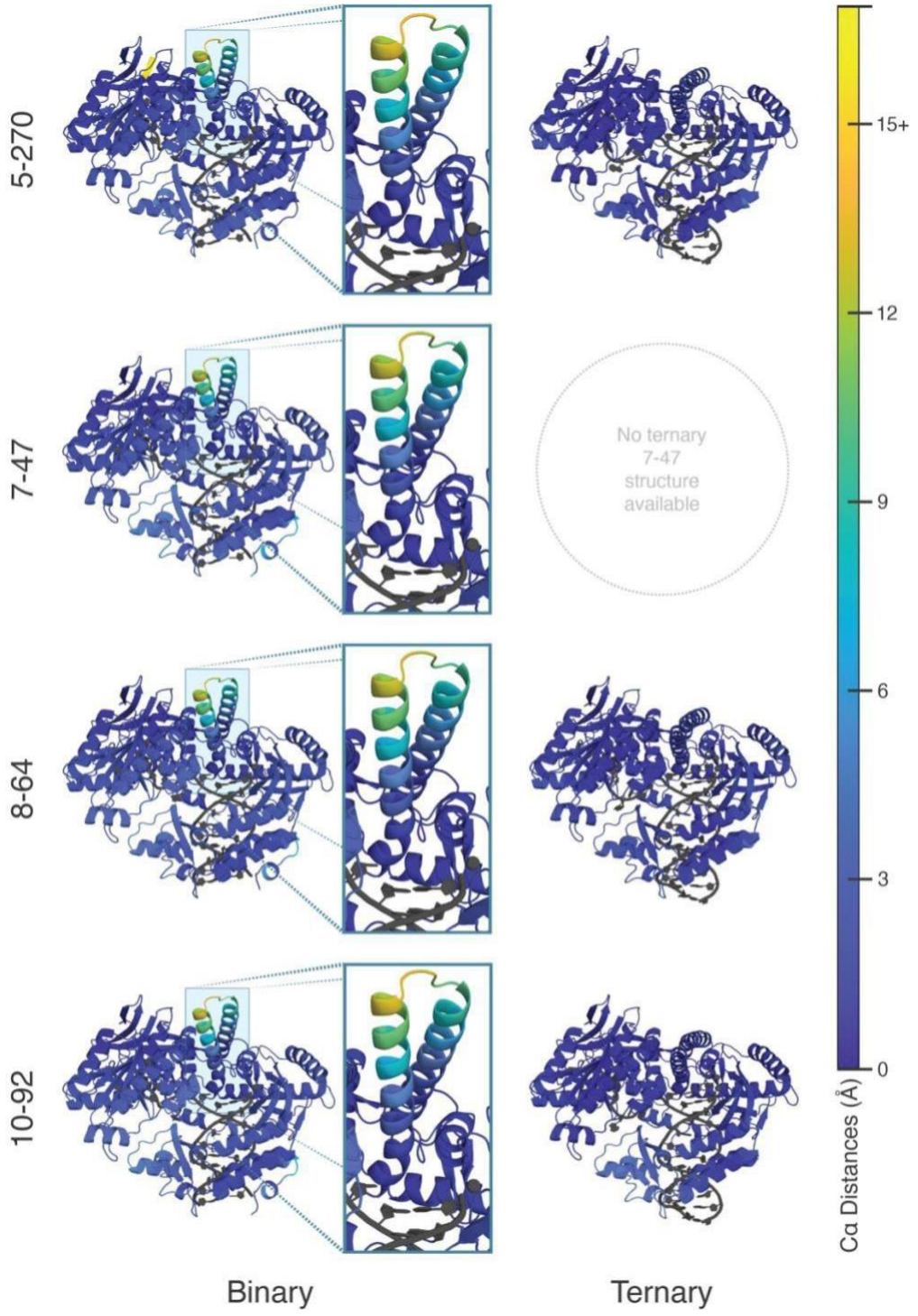
Supplementary Figure 20. AlphaFold3 prediction confidence. *AlphaFold3 model confidence (plDDT) heatmaps for binary (top) and ternary (bottom) complexes. Each row represents a single AlphaFold3 model. Two independent AlphaFold3 prediction runs were performed for each structure (denoted AF3-1 and AF3-2), each yielding five models (models 0–4, indicated by subscripts). Residue positions are shown along the x-axis, indicating domain boundaries. Confidence scores were binned as follows: Very high (plDDT > 90, dark blue), Confident (90 > plDDT > 70, light blue), Low (70 > plDDT > 50, yellow), and Very low (plDDT < 50, orange). Residues 761-775 were not predicted for the ternary structures.*



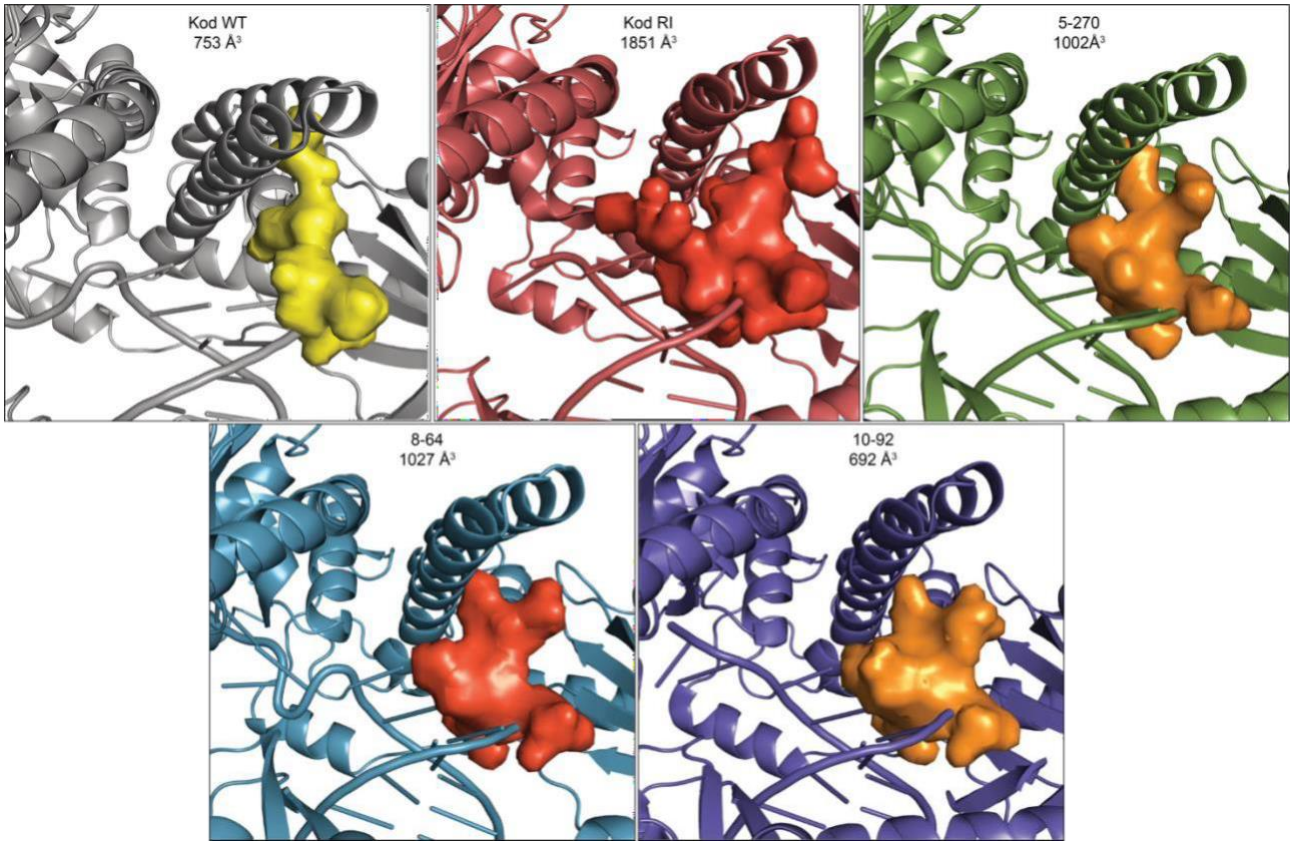
Supplementary Figure 21. AlphaFold3 prediction of the 5-270 binary structure. *Overlay of the experimentally determined binary structure (5-270, green) with five AlphaFold3 models (AF3-1$_{0\text{-}4}$, black). Overall structural RMSD = 1.430, 1.461, 1.449, 1.386, and 1.408 Å. AlphaFold3 does not predict the finger domain conformation or the rotamers of the catalytic triad.*



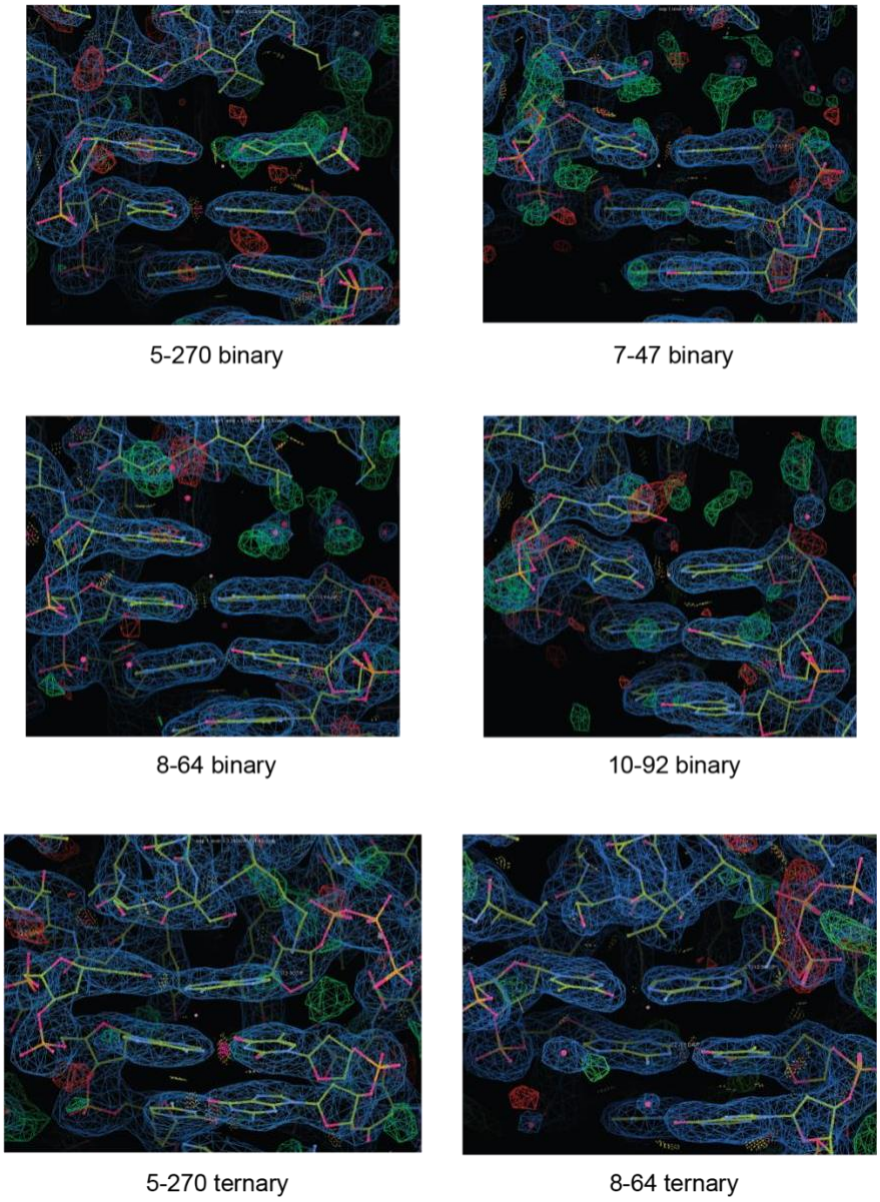
Supplementary Figure 22. Comparison of AlphaFold3 predictions to experimental structures. *The top panel shows AlphaFold3 predictions of the binary complexes (5-270, 7-47, 8-64, and 10-92), while the bottom panel shows predictions of ternary complexes (5-270, 8-64, and 10-92). Two AlphaFold3 runs were performed per model (denoted AF3-1 and AF3-2), producing 5 models per run (models 0 – 4, denoted by subscripts). Model predictions (10 per protein) are viewed as a heatmap across the amino acid sequence. Numbering along the backbone denotes domain and subdomain boundaries. Distances are measured between the predicted and crystallographic Cα position for each amino acid residue along the backbone.*



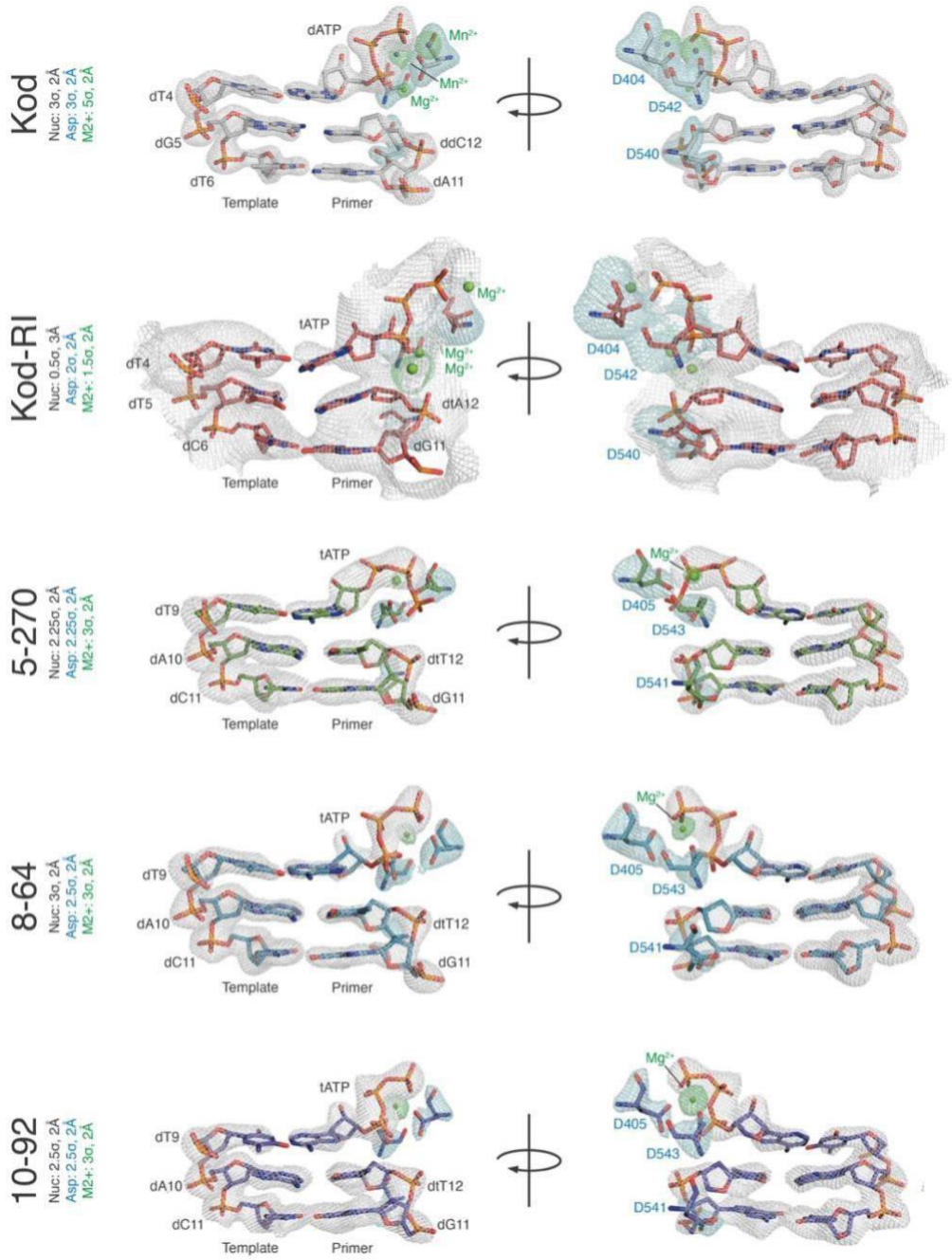
Supplementary Figure 23. Visualizing AlphaFold3 Predictions Against Experimental Structures. *Structural validation of AlphaFold3 predictions (AF3-1$_0$) relative to crystal structures for each crystallized binary and ternary complex (5-270, 7-47, 8-64, and 10-92). AlphaFold3 predictions are shown as cartoons, with residues colored according to the Cα distance (Å) between the predicted and experimentally resolved structures, following the indicated color scale. The zoomed-in view focuses on the binary finger domains. No experimental crystal structure is available for the ternary 7-47 complex for comparison.*



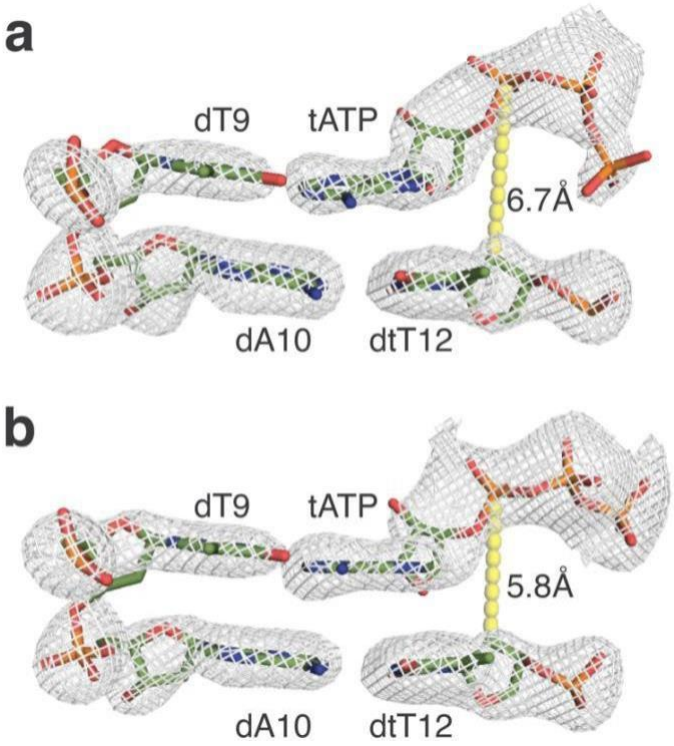
Supplementary Figure 24. Active site pocket volume. *Pocket volumes were calculated using PyVol, searching all pockets over a minimum volume threshold (500 Å) without partitioning. Probe min and max radii were set to 1.4 Å and 2.8 Å, respectively. Metals, incoming tNTP triphosphates, and solvents were removed for measurements*.



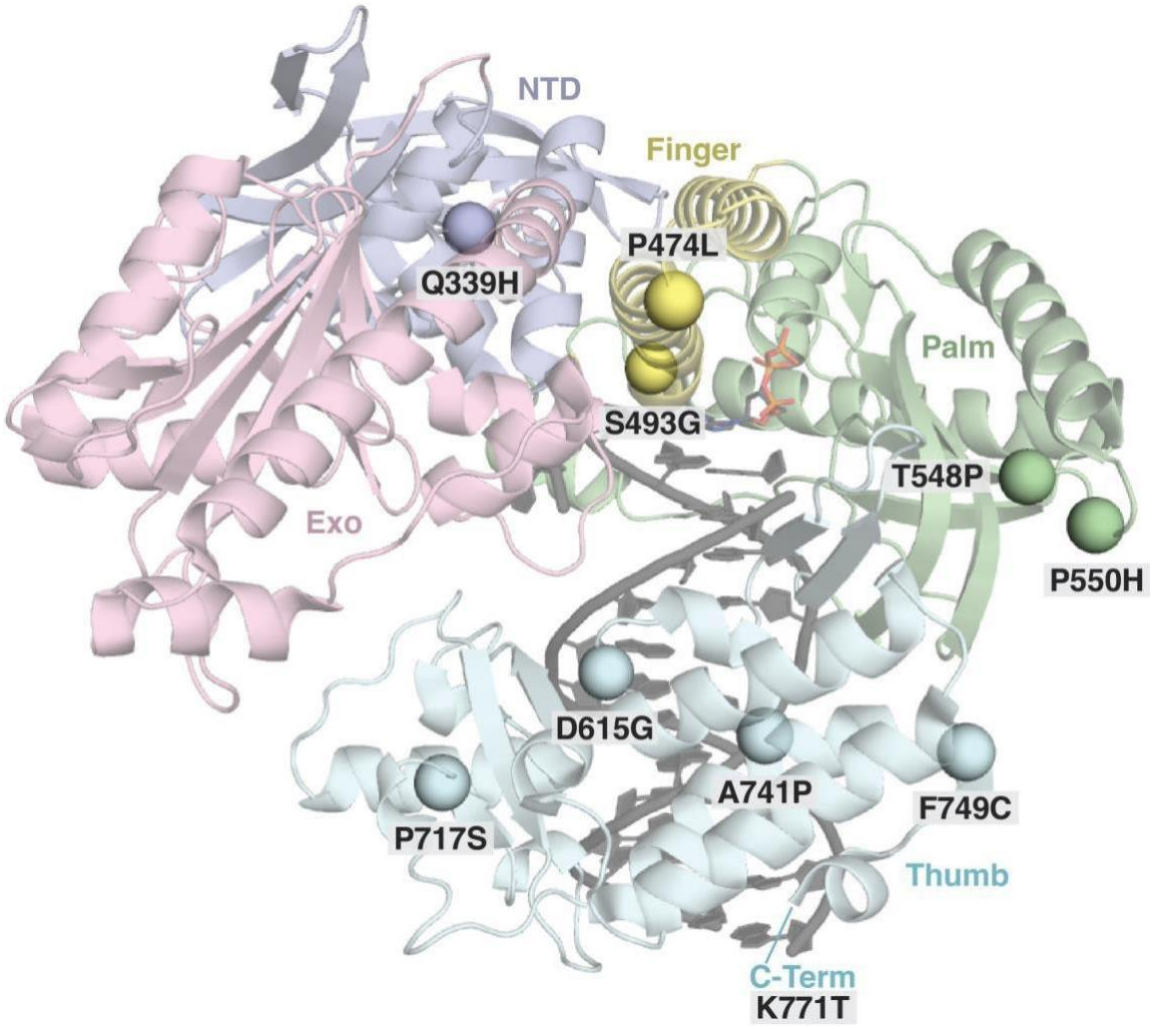
Supplementary Figure 25. 2Fo-Fc electron density maps of the active site. *For the newly reported structures, active site 2Fo-Fc electron density maps, contoured at 1.5σ, are shown.*



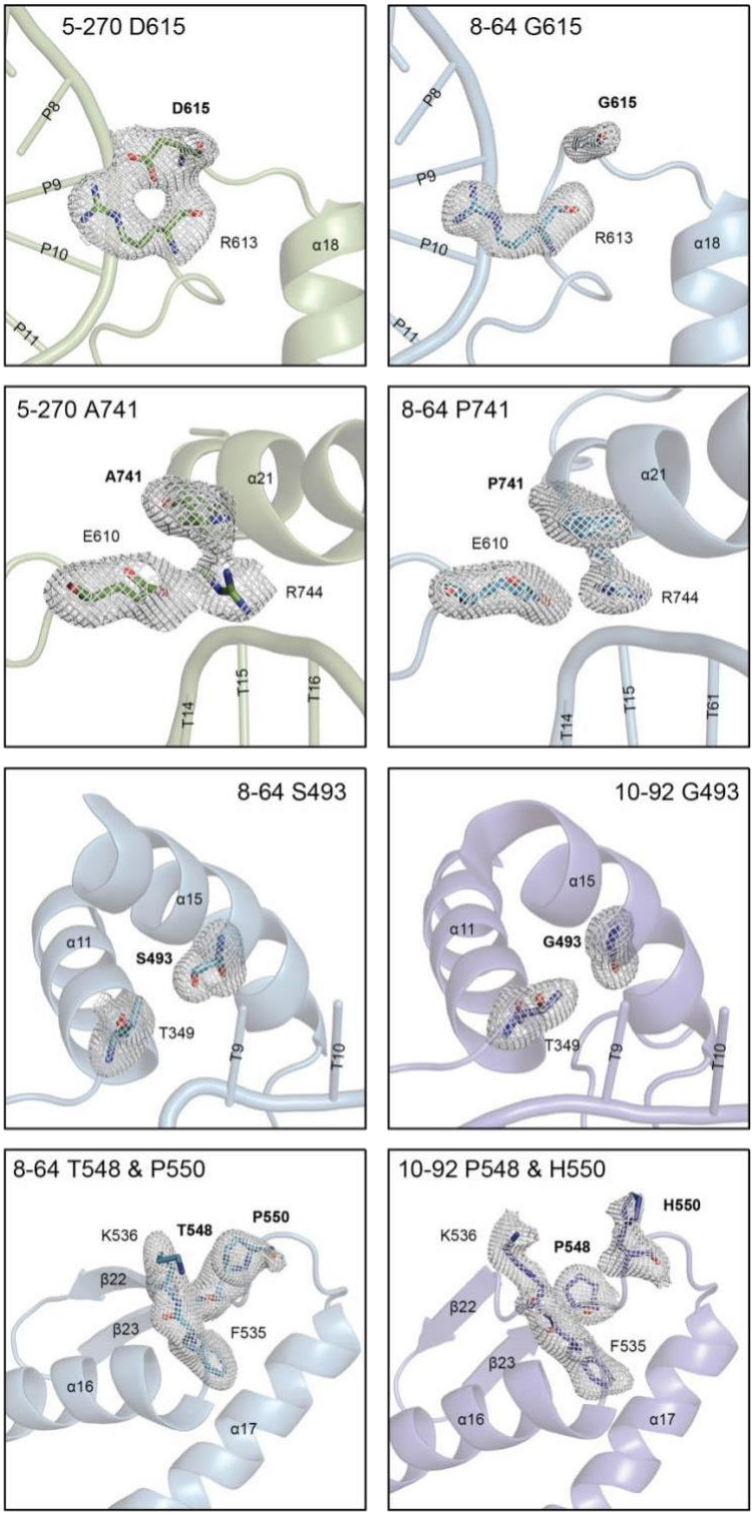
Supplementary Figure 26. Polder maps of key residues in the active site. *The incoming triphosphate, neighboring nucleotides, catalytic aspartates (D404/D405, D540/D541, and D542/D543), and divalent metal ions within the active sites of closed ternary complexes are shown. Polder maps (mFo–DFc) are displayed (nucleic acids, Nuc, gray; catalytic aspartates, Asp, aqua; divalent metals, M^{2+}, green), contoured at the indicated σ levels and carved to the specified distances. A 180° rotation about the y-axis provides an alternative view with side chains shown on the right.*



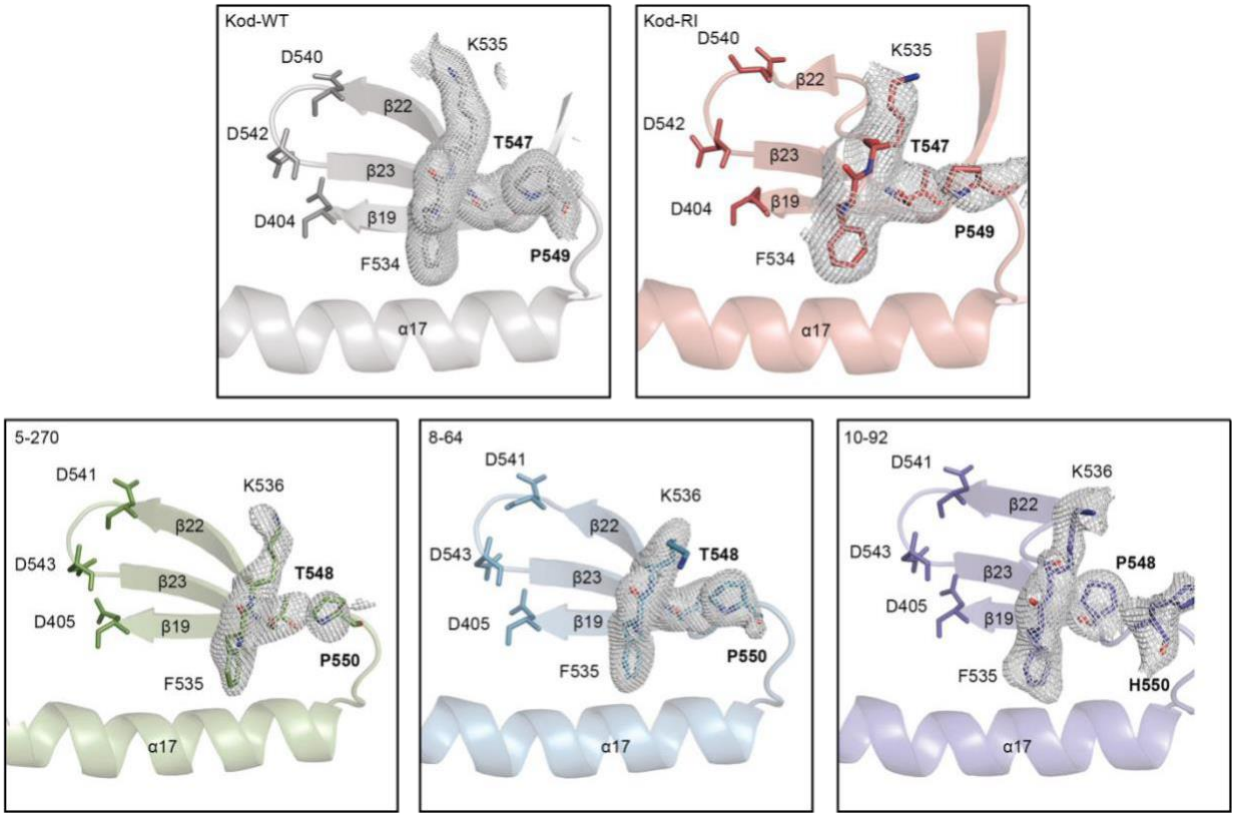
***Supplementary Figure 27**. **Active site geometry comparing two independently solved structures of 5-270 in a closed ternary complex.** Reaction centers are shown for (**a**) the original crystallographic dataset G7 and (**b**) an alternative independent dataset G3. The distance from the primer to the α-phosphate of the incoming nucleotide is indicated by a dashed yellow line. mFo–DFc polder maps support the modeled positions, contoured at 3.0σ and carved to 2.0 Å. The large pocket volume accounts for variability in the orientation of the triphosphate tail.*



Supplementary Figure 28. Catalysis-enhancing mutations mapped to Kod DNA polymerase. *The Kod^{exo-} scaffold (PDB ID 5OMF) is colored by domain. Spheres mark the Cα atoms of the ten residues mutated during error-prone evolution from 5-270 to 10-92. The K771T mutation near the C-terminus was not resolved in the crystal structure and is therefore not depicted. Domains are color-matched to Supplementary Figures 9 and 10.*



Supplementary Figure 29. Polder maps of mutated residues in 5-270, 8-64, and 10-92. *Polder maps (mFo–DFc) are displayed for E610, R613, D615/G615, A741/P741, and R744 in 5-270 and 8-64, and T349, S493/G493 F535, K536, T548/P548, and P550/H550 in 8-64 and 10-92. Residues shown in red indicates key residues that were mutated. S493 and T349 in 8-64 are contoured at 3.5σ and carved to 2.0 Å, while the rest of the maps are contoured at 3.0σ and carved to 2.0 Å.*



Supplementary Figure 30. Polder maps of mutated residues in DNAP and TNAPs relative to the catalytic aspartate residues. *Kod-WT, Kod-RI, 5-270, 8-64, and 10-92 structures are shown in individual panels with polder omit maps (mFo–DFc) for K535/K536, T547/T548/H548, P549/P550/H550. H550 in 10-92 is contoured at 3.5σ and carved to 2.0 Å, 8-64 maps are contoured at 4.0σ and carved to 2.0 Å, and the rest of the maps are contoured at 3.0σ and carved to 2.0 Å.*

.

	Oligo name	Oligo sequence
Kinetics	PBS8 primer	GTCCCCTTGGGGATACCACC
	EM619 template	CCCACACCCTCCTATCGCTAAACACACACTTAATAAAGTTGGTGGTATCCCCAAGGGGAC
Fidelity	Acrydite PBS8 primer	/5Acryd/GTCCCCTTGGGGATACCACC
	4NT.9G Fidelity template	GGATCGTCAGTGCAAAGAGATTAAGACTCGCCATGTTACGATCTGCCAAGTACAGCCTTGA ATCGTCACTGGTGGTATCCCCTTGGGGAC
	PBS7.PBS9 overhang primer	CTTTTAAGAACCGGACGAACGGATCGTCAGTGCATTGAGA
	PBS8 primer	GTCCCCTTGGGGATACCACC
	PBS9 primer	CTTTTAAGAACCGGACGAAC
Thermal challenge	PBS8_short primer	/5IRD680/GTCCCCTTGG
	EM619 template	CCCACACCCTCCTATCGCTAAACACACACTTAATAAAGTTGGTGGTATCCCCAAGGGGAC
Cloning	TNAP-Fwd	ATC*CATATG*ATCCTCGACACTGACTAC
	TNAP-Rvs	ACGCATG*CGGCCGC*TCAAGTTCCTTTCGGCGTCAG
	TNAP-Rvs_760	ACGCATG*CGGCCGC*TCACTTCTGGTAGCGCAGGTC
	8-64-Rvs_760	ACGCATG*CGGCCGC*TCAAGTTCCTTTCGGCGTCAG
Crystallography	P	CGCGAACTGCG
	T1	/5Cy5/AAACGTACGCAGTTCGCG
	T2	TATGCACGTACGCAGTTCGCG

Supplementary Table 1. DNA oligonucleotides organized by category. *Oligonucleotide sequences ordered from IDT written in the 5'→3' direction. Fluorescent dyes or modifications are noted on the specified termini. Highlighted in cyan and green are NdeI and NotI recognition sequences and in red are mismatched positions for sequences used in fidelity experiments*

Supplementary Table 2. Data collection and refinement statistics for binary structures

	5-270	7-47	8-64	10-92
Data Collection				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell Dimensions				
a, *b*, *c* (Å)	86.8, 110.5, 124.3	87.2, 110.6, 124.3	86.4, 108.0, 123.6	86.4, 108.3, 123.8
α, β, γ (º)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	45.95-2.56 (2.652-2.56)	50.51-2.18 (2.258-2.18)	49.5-2.5 (2.589-2.5)	33.3-2.172 (2.25-2.172)
R_{merge}	0.093 (0.863)	0.070 (0.265)	0.1368 (0.8707)	0.1051 (2.163)
CC1/2	0.998 (0.773)	0.99 (0.772)	0.992 (0.667)	0.998 (0.365)
$I / \sigma I$	15.08 (2.09)	10.94 (4.46)	8.39 (2.97)	10.73 (0.96)
Completeness (%)	99.42 (99.87)	98.23 (99.39)	99.91 (99.55)	99.79 (99.13)
Redundancy	6.5 (6.5)	3.4 (3.5)	5.2 (4.3)	6.8 (7.2)
Refinement				
Resolution (Å)	2.56	2.18	2.5	2.17
No. reflections	38976 (3841)	62205 (6183)	40682 (3980)	61779 (6048)
R_{work}/R_{free}	0.2181/0.2682 (0.2874/0.3524)	0.1887/0.2259 (0.2174/0.2723)	0.2295/0.2556 (0.3503/0.3649)	0.2167/0.2474 (0.3120/0.578)
No. atoms	6897	7358	6982	6966
Protein/DNA	6874	6873	6875	6878
Ligand	18	45	3	4
Solvent	5	440	104	84
B-factors	60.95	27.13	54.18	61
Protein/DNA	59.33	26.40	52.77	59.21
Ligand	78.70	36.86	47.08	63
Solvent	61.14	27.77	49.26	51.96
R.m.s deviations				
Bond lengths (Å)	0.002	0.003	0.002	0.002
Bond angles (º)	0.46	1.01	0.52	0.47

*Values in parentheses are for the highest-resolution shell.

Supplementary Table 3. Data collection and refinement statistics for ternary structures

	5-270	8-64
Data Collection		
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell Dimensions		
a, *b*, *c* (Å)	77.4, 101.7, 110.8	77.7, 101.5, 110.7
α, β, γ (º)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	48.67-3.03 (3.14-3.03)	48.59-2.38 (2.47-2.38)
R_{merge}	0.1628 (1.017)	0.0734 (0.7658)
CC1/2	0.995 (0.769)	0.999 (0.842)
$I / \sigma I$	12.57 (2.31)	19.45 (2.27)
Completeness (%)	99.61 (99.71)	99.87 (99.86)
Redundancy	6.6 (7.0)	6.6 (6.8)
Refinement		
Resolution (Å)	3.03	2.38
No. reflections	17539 (1729)	35755 (3519)
R_{work}/R_{free}	0.2240/0.2774 (0.2979/0.3730)	0.2159/0.2655 (0.2640/0.3218)
No. atoms	6783	6875
Protein/DNA	6743	6775
Ligands	41	32
Solvent	5	68
B-factors	73.16	59.48
Protein/DNA	71.59	57.79
Ligands	70.26	61.07
Solvent	47.47	45.98
R.m.s deviations		
Bond lengths (Å)	0.002	0.003
Bond angles (º)	0.39	0.53

*Values in parentheses are for the highest-resolution shell.

Kod-WT	Shear	Stretch	Stagger	Buckle	Prop-Tw	Opening	Shift	Slide	Rise	Tilt	Roll	Twist
T-a	-0.01	-0.236	0.064	-5.456	0.746	2.845	0	0	0	0	0	0
G-C	-0.271	-0.217	0.012	-8.368	-14.078	1.971	-0.605	-1.704	3.301	3.044	5.375	29.301
T-A	-0.041	-0.072	-0.218	1.349	0.959	-2.94	-0.741	-1.059	3.181	3.988	3.466	30.229
G-C	-0.374	-0.207	-0.253	-4.65	-1.784	2.365	0.375	-0.76	3.618	-3.476	0.867	37.528
G-C	-0.161	-0.058	-0.284	-12.984	-5.434	2.706	0.19	-1.105	3.634	1.913	3.746	30.239
C-G	-0.003	-0.181	0.283	-4.669	-7.263	-2.415	-1.05	-0.114	3.237	-5.455	1.91	32.406
C-G	0.444	-0.198	0.172	6.097	-8.909	2.69	0.322	0	3.125	1.314	9.18	29.577
G-C	-0.12	-0.238	0.582	6.321	-20.192	0.946	-0.973	0.556	3.334	-7.284	4.185	33.328
T-A	-0.22	-0.229	0.685	-7.786	-15.669	0.753	-0.196	-0.659	3.661	-0.31	-1.082	34.762
G-C	0.337	-0.07	0.356	8.874	-7.834	0.832	0.782	0.587	2.988	3.285	-0.55	36.967
G-C	-0.459	-0.047	0.702	-2.255	-18.78	4.552	0.18	-0.546	3.416	-3.949	3.155	31.512
T-A	0.108	-0.157	-0.043	-3.203	-16.523	1.076	-0.764	-0.964	3.225	6.71	0.883	35.98
C-G	0.125	-0.241	-0.341	6.525	-2.221	-2.727	0.796	-0.399	3.142	3.83	1.518	35.085
Kod-RI	Shear	Stretch	Stagger	Buckle	Prop-Tw	Opening	Shift	Slide	Rise	Tilt	Roll	Twist
T-a	0.025	-0.713	-1.831	-20.004	-20.512	-15.264	0	0	0	0	0	0
T-a	-1.15	-0.098	-0.315	-32.927	-8.414	-6.13	-0.248	-0.412	4.146	-9.214	4.829	22.767
C-G	0.548	-0.465	0.194	-12.395	8.368	4.839	0.292	0.169	3.187	-7.214	10.287	33.488
G-C	0.219	-0.089	-0.245	-8.262	-5.838	6.681	-0.417	-0.654	3.479	4.25	-6.934	35.888
C-G	-0.074	0.94	0.649	-11.417	-8.149	25.954	1.297	-1.113	3.282	-3.717	3.475	29.268
A-T	1.05	-0.208	-0.295	12.854	-2.334	-2.286	-1.941	0.944	2.978	6.559	0.132	33.195
G-C	-0.095	-0.298	1.19	8.051	4.134	10.399	1.033	-1.502	3.404	-14.843	8.355	24.019
T-A	-0.54	-0.412	0.415	6.18	-20.716	-0.975	-1.07	-0.709	3.533	9.595	-2.63	34.032
T-A	0.682	-0.426	0.416	1.218	-15.606	-1.074	0.382	-0.319	3.342	1.396	3.136	38.543
C-G	0.346	-0.028	-0.264	20.634	-9.77	6.291	1.103	0.339	2.729	12.053	5.085	32.118
C-G	-0.998	1.108	-1.466	18.676	-42.087	-49.397	-1.416	0.893	6.23	7.057	0.978	57.702
7-47	Shear	Stretch	Stagger	Buckle	Prop-Tw	Opening	Shift	Slide	Rise	Tilt	Roll	Twist
T-a	0.581	-0.161	-0.632	-14.01	3.651	3.729	0	0	0	0	0	0
A-t	-0.33	-0.316	0.264	-12.001	-15.878	2.545	-1.226	-1.453	3.454	-4.866	7.898	20.418
C-G	-0.281	-0.279	0.006	-5.25	0.447	-1.935	-0.664	-0.735	3.2	4.482	5.76	30.569
G-C	-0.338	-0.137	-0.001	-6.306	-16.368	5.687	0.297	-0.228	3.497	-1.32	10.256	35.117
C-G	0.501	-0.301	0.122	-10.195	2.838	3.516	0.03	-0.151	3.435	-1.238	0.731	34.881
A-T	0.344	-0.195	0.611	6.511	-2.481	-8.074	-1.278	0.429	3.038	-5.326	-3.241	33.223
G-C	0.101	-0.098	0.339	10.127	-4.82	9.088	1.287	-0.677	3.402	1.064	8.808	27.68
T-A	-0.367	-0.281	0.277	6.647	-16.741	-0.137	-1.096	-0.549	3.46	0.829	1.387	31.864
T-A	-0.307	-0.271	0.207	5.235	-16.374	-4.05	-0.099	-0.262	3.305	1.078	2.785	37.91
C-G	0.267	-0.345	-0.027	-0.78	-12.303	-2.07	0.486	-0.059	3.408	4.085	6.795	36.894
G-C	0.007	-0.32	-0.246	-2.119	-7.847	0.003	0.344	0.451	3.396	3.327	7.686	32.664
C+G	-0.108	-5.304	0.071	4.607	-0.38	81.91	-2.188	3.315	-0.894	174.391	-12.494	-138.25
G-C	-0.971	-1.128	-2.394	-34.185	-21.806	12.434	2.203	3.984	-0.606	-122.9	93.499	-113.624
8-64	Shear	Stretch	Stagger	Buckle	Prop-Tw	Opening	Shift	Slide	Rise	Tilt	Roll	Twist
T-a	0.552	-0.225	-0.205	-13.85	1.559	2.082	0	0	0	0	0	0
A-t	-0.052	-0.301	0.03	-10.223	-19.876	8.119	-0.812	-1.39	3.417	1.555	8.403	23.644
C-G	-0.005	-0.165	-0.383	-4.068	-2.314	0.284	-0.666	-0.339	3.261	6.477	4.187	30.737
G-C	0.025	0.072	0.117	-6.053	-16.221	6.737	-0.021	0.13	3.527	-4.558	17.572	35.08
C-G	0.339	-0.237	-0.339	-4.062	1.166	4.925	0.408	-0.184	3.315	2.719	-1.819	32.071
A-T	-0.016	-0.097	0.207	5.26	-10.203	-2.972	-1.429	0.469	3.24	-5.885	0.557	32.074
G-C	-0.109	-0.219	-0.216	-2.347	-11.83	5.865	0.705	-0.399	3.728	0.517	11.798	29.169
T-A	-0.387	-0.332	-0.394	7.021	-17.831	2.441	-1.074	-0.471	3.044	-0.022	4.633	29.771
T-A	-0.437	-0.317	-0.071	5.587	-21.638	-0.932	-0.207	0.012	3.3	-1.832	5.356	37.027
C-G	0.658	-0.388	-0.137	1.485	-11.061	1.957	0.645	0.167	3.38	1.752	4.017	38.519
G-C	-0.858	-0.418	-0.444	-15.041	-16.907	5.539	0.239	0.684	3.668	3.817	12.95	31.843
C-G	0.6	-0.244	-0.226	6.755	-3.96	-1.913	-0.275	0.364	2.753	-4.524	5.717	31.987
G-C	-1.193	-0.897	-2.257	-35.235	-27.081	9.202	0.705	1.646	4.511	15.653	15.696	33.765
10-92	Shear	Stretch	Stagger	Buckle	Prop-Tw	Opening	Shift	Slide	Rise	Tilt	Roll	Twist
T-a	0.437	-0.352	-0.884	-5.879	-1.4	1.627	0	0	0	0	0	0
A-t	-0.107	0.062	0.258	-18.035	-18.371	5.27	-0.329	-1.356	3.743	-4.419	15.703	24.875
C-G	0.035	-0.152	0.067	-8.551	-0.961	-1.952	-0.594	-0.46	3.188	5.43	5.21	31.003
G-C	0.065	-0.003	-0.097	-18.373	-10.071	7.339	0.26	-0.094	3.684	0.568	10.385	36.021
C-G	0.531	0.081	-0.366	-3.079	-6.454	9.812	0.646	0.23	2.957	5.087	4.259	31.317
A-T	0.112	-0.073	0.493	0.763	-5.571	-6.963	-1.663	0.642	3.345	-11.641	3.443	32.906
G-C	-0.175	0.293	0.501	2.177	-8.584	12.314	0.807	-0.827	3.286	-2.303	5.462	27.833
T-A	-0.104	-0.199	0.404	1.139	-11.845	-2.138	-1.442	-0.842	3.446	-0.208	3.022	31.388
T-A	-0.184	-0.164	-0.402	9.206	-26.265	-2.821	-0.231	0.003	2.909	5.907	-0.747	36.739
C-G	0.181	-0.041	0.496	-3.26	-18.176	3.541	0.626	-0.19	3.547	-3.181	10.773	34.391
G-C	0.55	0.157	0.168	-9.408	-28.969	-4.303	-0.405	0.423	3.499	5.018	11.005	38.193
C-G	0.425	-0.062	0.057	-16.986	2.121	6.174	1.471	0.568	3.542	1.889	3.763	34.259
G-C	0.304	-0.117	0.091	4.935	-10.879	-1.552	-0.678	1.743	3.147	-2.838	-0.161	35.274

Supplementary Table 4. Base pair parameters. *Base pair parameters of duplexes in ternary structures computed using Web x3DNA.*

Mutation	Domain	TNAP Generation	BLOSUM62	FoldX ΔΔG (kcal mol-1)	Cα Distance to Active Site (Å)			ABS SASA (Å^2)			REL SASA (% of max)			% Activity if reverted	
					Kod	5-270	10-92	Kod	5-270	10-92	Kod	5-270	10-92	45-sec Reaction	60-sec Reaction
K99R	NTD	5-270	2	0.061	47.5	46.8	46.6	103.1	130.5	144.0	50.3	54.8	60.5		
E102A	NTD	5-270	-1	0.016	46.5	45.2	45.8	132.3	86.3	76.4	75.9	79.3	70.2		
I107V	NTD	5-270	3	0.650	43.4	42.6	42.3	97.5	73.3	68.1	55.5	48.2	44.8		
V127I	NTD	5-270	3	-0.640	32.6	33.3	32.4	31.9	55.1	36.9	21.0	31.4	21.0		
K136T	Exo	5-270	-1	1.173	39.8	39.8	39.3	95.6	67.0	51.9	46.6	47.7	36.9		
D141A	Exo	Kodexo-	-2	N/A	38.7	39.1	38.6	13.4	12.0	11.8	12.3	11.1	10.8		
E143A	Exo	Kodexo-	-1	N/A	42.6	42.8	42.4	5.2	8.8	9.1	4.8	8.0	8.4		
Q285K	Exo	5-270	1	-0.323	31.8	31.2	31.7	103.1	97.2	124.9	57.7	47.4	60.9		
T296A	Exo	5-270	0	1.125	47.7	47.9	48.0	9.3	34.5	11.9	6.6	31.7	11.0		
T297Q	Exo	5-270	-1	-0.260	49.8	49.9	50.1	77.1	85.2	130.0	54.8	47.6	72.7		
N304G	Exo	5-270	0	-0.815	51.2	50.7	50.7	97.5	38.6	26.5	67.3	47.6	32.7		
I337V	NTD	5-270	3	0.916	24.6	25.4	24.0	0.2	3.2	1.8	0.1	2.1	1.2		
Q339H	NTD	33878	0	1.633	29.2	29.9	28.7	0.5	2.7	2.7	0.3	1.5	1.5	90.20	93.62
S340P	NTD	5-270	-1	-1.031	30.7	31.8	30.5	1.3	2.6	0.2	1.1	1.9	0.1		
F356Y	NTD	5-270	3	0.779	24.5	25.1	24.5	8.0	42.1	40.4	4.0	19.6	18.9		
K375R	Palm	5-270	2	-0.275	33.4	31.6	31.8	160.1	150.4	144.1	78.1	63.1	60.5		
L377Y	Palm	5-270	-1	1.387	28.1	26.7	26.8	25.0	45.7	51.1	13.9	21.3	23.9		
A378E	Palm	5-270	-1	-0.214	29.1	27.1	27.3	74.4	56.7	103.7	68.4	32.5	59.6		
+381L	Palm	5-270	N/A	N/A	N/A	22.1	22.6	N/A	102.3	94.7	N/A	57.0	52.7		
Q383E	Palm	5-270	2	0.076	22.5	21.7	21.9	100.3	41.8	43.4	56.1	24.0	24.9		
E386A	Palm	5-270	-1	0.822	15.0	14.2	14.5	96.8	54.8	52.5	55.5	50.4	48.3		
R395K	Palm	5-270	2	-0.061	18.9	20.8	20.6	132.6	116.1	110.0	55.7	56.6	53.7		
K466R	Finger	5-270	2	0.087	20.9	21.1	21.3	151.5	146.1	163.4	73.9	61.3	68.6		
I472V	Finger	5-270	3	0.066	28.1	28.1	27.8	167.9	146.0	130.7	95.5	96.1	86.0		
P474L	Finger	33878	-3	0.109	25.4	25.9	26.1	80.1	106.7	153.5	58.4	77.7	85.5	69.86	75.83
I475L	Finger	5-270	2	-0.359	25.7	26.3	26.2	38.5	58.0	61.7	21.9	32.3	34.4		
R477K	Finger	5-270	2	0.303	22.5	22.5	22.7	85.4	74.4	96.1	35.9	36.3	46.9		
A485R	Finger	Kod-RI	-1	-2.007	14.8	15.3	14.9	30.9	31.4	17.9	28.4	13.2	7.5		
N491S	Finger	Kod-RS	1	0.376	10.6	10.5	10.1	48.8	35.1	37.0	33.6	29.7	31.3		
S493G	Finger	33878	0	1.666	13.2	13.2	12.7	0.0	0.7	0.0	0.0	0.6	0.0	47.71	74.31
E520Q	Palm	5-270	2	-0.031	15.2	14.9	14.8	88.1	79.7	70.6	50.6	44.6	39.5		
T523E	Palm	5-270	-1	-0.277	14.5	14.3	14.3	55.8	75.0	77.7	39.7	43.1	44.6		
M524T	Palm	5-270	-1	2.050	17.0	17.0	16.9	49.7	65.3	68.7	25.7	46.5	48.9		
K527R	Palm	5-270	2	-0.239	19.2	19.1	18.9	123.8	121.5	148.2	60.4	51.0	62.2		
Y533F	Palm	5-270	3	0.284	24.1	24.3	24.1	54.8	34.7	51.4	25.6	17.3	25.7		
I538L	Palm	5-270	2	-0.088	15.7	16.0	16.1	0.0	0.0	0.0	0.0	0.0	0.0		
S540A	Palm	5-270	1	-0.331	9.3	9.7	9.6	0.5	0.0	0.4	0.4	0.0	0.4		
T548P	Palm	33878	-1	6.284	19.7	19.8	20.0	3.6	0.5	3.2	2.6	0.4	2.3	25.85	47.91
P550H	Palm	33878	-2	1.468	25.1	25.1	25.8	91.1	87.2	169.9	66.4	63.6	92.9	60.16	68.26
M562K	Palm	5-270	-1	0.480	21.5	21.5	20.9	100.8	107.5	104.6	52.2	52.4	51.0		
K566D	Palm	5-270	-1	0.524	22.4	22.3	22.0	168.5	108.4	105.6	82.2	76.0	73.9		
A575L	Palm	5-270	-1	-1.199	17.7	17.6	17.4	25.4	84.7	77.9	23.4	47.2	43.4		
G602D	Thumb	5-270	-1	3.856	23.4	24.6	24.2	26.7	74.6	71.0	32.9	52.2	49.7		
R606G	Thumb	Kod-RSGA	-2	2.597	11.0	11.1	11.1	94.9	1.9	1.1	39.8	2.3	1.4		
D615G	Thumb	17349	-1	1.425	21.9	21.4	21.6	17.9	26.1	22.7	12.5	18.3	28.0	11.66	31.51
K672R	Thumb	5-270	2	-1.366	35.3	33.5	34.1	178.3	184.2	229.5	87.0	77.3	96.4		
P717S	Thumb	17349	-1	3.514	35.8	34.3	34.4	17.1	12.4	7.3	12.5	9.0	6.2	99.04	90.40
T723A	Thumb	Kod-RSGA	0	-0.079	47.7	46.0	46.4	118.9	96.5	92.2	84.5	88.7	84.8		
A741P	Thumb	23590	-1	3.358	23.1	23.2	23.0	0.2	0.8	1.3	0.2	0.8	0.9	17.28	19.32
F749C	Thumb	17349	-2	3.387	28.9			60.5			30.2			69.81	73.12
K771T	Thumb	17349	-1											73.05	74.55

Supplementary Table 5. Integrated mutational analysis. *For each mutation, the polymerase domain, generation of appearance during TNAP evolution, BLOSUM62 substitution score, predicted stability change (FoldX ΔΔG, kcal mol^{-1}), Cα distance to the active site (Å), and absolute and relative solvent-accessible surface area (ABS SASA, Å^2; REL SASA, % of maximum) are reported for the ternary structures of Kod^{exo-}, 5-270, and 10-92. Functional impact is summarized as residual activity (% of 10-92) upon reversion to 5-270, extracted from prior data[1]. N/A = non-applicable.*

References for Appendix B

1. Maola, V. A. *et al.* Directed evolution of a highly efficient TNA polymerase achieved by homologous recombination. *Nature Catalysis* 7, 1173-1185 (2024).

2. Liao, J.-Y., Bala, S., Ngor, A. K., Yik, E. J. & Chaput, J. C. P(V) Reagents for the Scalable Synthesis of Natural and Modified Nucleoside Triphosphates. *J. Am. Chem. Soc.* 141, 13286-13289 (2019).

3. Bala, S. *et al.* Synthesis of 2'-Deoxy-alpha-l-threofuranosyl Nucleoside Triphosphates. *J. Org. Chem.* 83, 8840-8850 (2018).