

UC Irvine

UC Irvine Electronic Theses and Dissertations

Title

Directed Evolution and Structural Analysis of Engineered Polymerases

Permalink

<https://escholarship.org/uc/item/0kr6352v>

ISBN

9798191649474

Author

Hajjar, Mohammad

Publication Date

2026-09-21

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Directed Evolution and Structural Analysis of Engineered Polymerases

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Pharmacological Sciences

by

Mohammad Hajjar

Dissertation Committee:
Professor John C. Chaput, Chair
Professor Brian M. Paegel
Assistant Professor Benjamin R. Morehouse
Assistant Professor Thomas F. Martínez

2026

Portions of Chapter 1 © 2022 Mohammad Hajar (Oxford University Press)
Portions of Chapter 2 © 2024 Mohammad Hajar (Springer Nature Limited)
Portions of Chapter 3 © 2025 Mohammad Hajar (Springer Nature)
Portions of Chapter 4 © 2026 Mohammad Hajar (Springer Nature America, Inc.)
All other materials © 2026 Mohammad Hajar

DEDICATION

Bismillāh—ir-Raḥmān—ir-Raḥeem.
Al-ḥamdu lillāh; waṣ-ṣalātu wa-s-salāmu ‘alā Rasūlillāh.

I thank God for His endless blessings and favors.

Then I thank:

My treasured
mother, father,
& sister

My
beloved
wife

My
family
& friends

For your love, generosity, sacrifice, support, guidance, companionship,
and encouragement to uphold truth & maintain patience.

TABLE OF CONTENTS

	Page
LIST OF FIGURES	vi
ACKNOWLEDGMENTS	viii
VITA.....	x
ABSTRACT OF THE DISSERTATION	xi
INTRODUCTION	1
0.1 REFERENCES	14
CHAPTER 1: Crystallographic Analysis of Engineered Polymerases Synthesizing Phosphonomethylthreosyl Nucleic Acid.....	20
1.1 ABSTRACT.....	21
1.2 INTRODUCTION	22
1.3 MATERIAL AND METHODS.....	25
1.4 RESULTS	30
1.5 DISCUSSION.....	37
1.6 CONCLUSION.....	41
1.7 REFERENCES	42
1.8 FIGURES.....	46
CHAPTER 2: Directed Evolution of Highly Efficient TNA Polymerase Achieved by Homologous Recombination	53
2.1 ABSTRACT.....	54
2.2 INTRODUCTION	55
2.3 RESULTS	58
2.4 DISCUSSION.....	68
2.5 METHODS	71

2.6 REFERENCES	81
2.7 FIGURES	87
CHAPTER 3: Directed Evolution of a TNA Polymerase Identifies Independent Paths to Fidelity and Catalysis	95
3.1 ABSTRACT	97
3.2 INTRODUCTION	98
3.3 RESULTS	100
3.4 DISCUSSION	110
3.5 METHODS	113
3.6 REFERENCES	125
3.7 FIGURES	129
CHAPTER 4: Rapid Evolution of a Highly Efficient RNA Polymerase by Homologous Recombination	135
4.1 ABSTRACT	136
4.2 INTRODUCTION	137
4.3 RESULTS	140
4.4 DISCUSSION	150
4.5 METHODS	154
4.6 REFERENCES	168
4.7 FIGURES	174
4.8 EXTENDED DATA FIGURES	181
CHAPTER 5: Summary and Conclusions	185
5.1 REFERENCES	198
APPENDIX A: Supplementary Information for Chapter 1	200
APPENDIX B: Supplementary Information for Chapter 2	212

APPENDIX C: Supplementary Information for Chapter 3	281
APPENDIX D: Supplementary Information for Chapter 4	319

LIST OF FIGURES

	Page
Figure 1.1 pTNA Synthesis by Kod-RI TNA Polymerase	47
Figure 1.2 Active Site View of Kod-RI TNA Polymerase Showing pTNA Incorporation	48
Figure 1.3 Active Site View of the DNA, TNA, and pTNA Adducts	49
Figure 1.4 Pseudorotation Wheel and Sugar Puckering	50
Figure 1.5 Engineered Mutations in Kod-RSGA.....	51
Figure 1.6 Structural Consequences of the Engineered Mutations.....	52
Figure 2.1 Homologous recombination as a path to engineered XNA polymerases	88
Figure 2.2 Directed evolution of a highly efficient TNA polymerase.....	89
Figure 2.3 Functional analysis of TNA polymerases.....	90
Figure 2.4 Kinetic analysis of TNA synthesis	91
Figure 2.5 Enhanced functional activity	92
Figure 2.6 Structural topology of the 10-92 TNA polymerase.....	93
Figure 2.7 Active site of the 10-92 TNA polymerase.....	94
Figure 3.1 Evolution of the 10-92 TNA polymerase	129
Figure 3.2 The finger subdomain of TNA polymerases	130
Figure 3.3 Stepwise changes in global and local structural conformation	131
Figure 3.4 Structural optimization of the nascent base pair and reaction center	132
Figure 3.5 Integrated analysis of 51 mutations acquired during the evolution of 10-92	133
Figure 3.6 Structural analysis of key mutations driving improved catalysis.....	134
Figure 4.1 Library generation and selection strategy.....	175
Figure 4.2 Evolution of a DNA-dependent RNA polymerase.....	176

Figure 4.3 Functional analysis of RNA polymerase C28	177
Figure 4.4 Fidelity and template composition requirements of RNA polymerase C28.....	178
Figure 4.5 Modified RNA synthesis by RNA polymerase C28.....	179
Figure 4.6 Crystallographic analysis of the engineered C28 RNA polymerase	180
Extended Data Figure 4.1 C28 Reversion Analysis.....	181
Extended Data Figure 4.2 C28 Activity Under Fidelity Assay Conditions	182
Extended Data Figure 4.3 C28 Extends Primers Despite Template Mismatches.....	183
Extended Data Figure 4.4 Mapping mutations acquired by homologous recombination.....	184

ACKNOWLEDGMENTS

I thank my advisor, Professor Chaput, for his mentorship, patience, and support throughout my doctoral training. I am grateful for the opportunities he provided, the lessons I learned from him, and the guidance that made this research possible.

I extend my thanks to the current and former members and visiting researchers of the Chaput Lab: Dr. Chim for his scientific enthusiasm, constructive discussions, and teaching me crystallography; Arlene for facilitating the logistics of our research; Victoria, Esau, and Eric for teaching me molecular biology and microfluidics techniques; Ali and Derek for establishing foundational protocols and systems; and Alex and Eloi for productive discussions. I am grateful to Victoria, Joy, Dr. Chim, and the crystallography teams, including Katherine, Grace, Ethan, Pranav, Salaar, Elionel, Jarred, Emme, Claudia, Riley, Tiffany, Manuel, Jenny, Calvin, Ashley, Morgan, Roman, Aditya, Kai, Tina, Betty, and others. I also thank the chemists who synthesized oligonucleotide monomers – Clark, Sergei, Biju, Daisy, Bhawna, Javeena, Qingfeng, and Masahito; those who synthesized polymers – Noah and Erica; and the application-focused teams, including Kim, Nandini, Adriana, Jaicy, Turnee, Kefan, Daniel, Kelly, Yutong, Keer, and Cailen. I thank Sabrina for laboratory support, Jessica for purchasing, and the rotation students, Pun and Rose, for their contributions to my work. To all, thank you for contributing to a collaborative and intellectually engaging research environment.

I thank Professors Paegel, Morehouse, and Martínez for their time, guidance, and thoughtful feedback as members of my doctoral committee; Professor Herdewijn of KU Leuven for generously sharing his extensive knowledge and perspective on nucleic acid chemistry; Professors Yajun and Butts for their collaborations; Professors Paegel and Sjogren for sharing laboratory equipment; Val and Jan for their insights and technical advice; and the faculty and members of the Department of Pharmaceutical Sciences. I am particularly grateful to Professor Mobley for his patience and guidance during my rotation and to his students, Ana and AK, for teaching me the basics of computation. I am grateful to Dana for coordinating departmental affairs.

I also thank Professors Chaput, Spitale, Trader, and Paegel for their mentorship and for the opportunity to teach alongside them as a teaching assistant in medicinal chemistry. I am especially grateful to Professor Paegel for taking the time to discuss scientific presentation, teaching styles, and active-learning strategies. I also thank the students, whose questions and thoughtful discussions taught me a great deal in return.

More broadly, I extend my gratitude to all my teachers and professors, and to everyone who has taught me something beneficial, offered guidance, or simply wished me well along the way.

Portions of Chapter 1 are adapted from “Crystallographic analysis of engineered polymerases synthesizing phosphonomethylthreosyl nucleic acid.” *Nucleic Acids Research* (2022), 50, 9663–9674. Used with permission from Oxford University Press under CC BY 4.0. The coauthors are Nicholas Chim, Chao Liu, Piet Herdewijn, and John C. Chaput. John C. Chaput directed and supervised research which forms the basis for the dissertation.

Portions of Chapter 2 are adapted from “Directed evolution of a highly efficient TNA polymerase achieved by homologous recombination.” *Nature Catalysis* (2024), 7, 1173–1185. Used with permission from Springer Nature Limited. The coauthors are Victoria A. Maola, Eric J. Yik, Joy J. Lee, Manuel J. Holguin, Riley N. Quijano, Calvin K. Nguyen, Katherine L. Ho, Jenny V. Medina, Nicholas Chim, and John C. Chaput. John C. Chaput directed and supervised research which forms the basis for the dissertation.

Portions of Chapter 3 are adapted from “Directed evolution of a TNA polymerase identifies independent paths to fidelity and catalysis.” *Nature Communications* (2026), 17, 925. Used with permission from Springer Nature under CC BY 4.0. The coauthors are Victoria A. Maola, Joy J. Lee, Manuel J. Holguin, Riley N. Quijano, Calvin K. Nguyen, Katherine L. Ho, Jenny V. Medina, Elionel Botello-Cornejo, Bhawna Barpuzary, Nicholas Chim, and John C. Chaput. John C. Chaput directed and supervised research which forms the basis for the dissertation.

Portions of Chapter 4 are adapted from “Rapid evolution of a highly efficient RNA polymerase by homologous recombination.” *Nature Chemical Biology* (2026), 22, 960–968. Used with permission from Springer Nature America, Inc. The coauthors are Esau L. Medina, Victoria A. Maola, Grace K. Ko, Ethan J. Ho, Alexandria R. Horton, Nicholas Chim, and John C. Chaput. John C. Chaput directed and supervised research which forms the basis for the dissertation.

The research presented in this dissertation was supported in part by the National Science Foundation (NSF) under awards 2001434, MCB-1946312, and CHE-2433788; the Belgian National Fund for Scientific Research under grant FWO-G085321N; and the University of California. Award CHE-2433788 was administered through the NSF Divisions of Chemistry of Life Processes and Genetic Mechanisms. The NSF funded open-access publication of Chapter 1.

I acknowledge UCI’s Summer Undergraduate Research Program (SURP) and Undergraduate Research Opportunities Program (UROP) for supporting my undergraduate research; the Interdisciplinary Research Teams Program, facilitated by UCI and the California Institute for Telecommunications and Information Technology (Calit2), for its mentorship opportunities; and Vertex Pharmaceuticals for support through the Vertex Fellowship.

I acknowledge the Advanced Light Source (ALS) at Lawrence Berkeley National Laboratory and the Stanford Synchrotron Radiation Lightsource (SSRL) at SLAC National Accelerator Laboratory for access to synchrotron facilities and for support during data collection.

I thank the UCI Libraries for providing resources and support throughout my graduate studies.

I thank Jan van IJken for permission to screen embryonic development footage during my defense.

VITA

Mohammad Hajjar

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

FIELD OF STUDY

Chemical Biology; Molecular and Structural Biology; Biochemistry; Directed Evolution;
Polymerase Engineering; Nucleic Acid Chemistry

PUBLICATIONS

Hajjar, M., Chim, N., Liu, C., Herdewijn, P. and Chaput, J.C. (2022) Crystallographic analysis of engineered polymerases synthesizing phosphonomethylthreosyl nucleic acid. *Nucleic Acids Res*, **50**, 9663.

Hajjar, M., Horton, A.R., Maola, V.A., Chim, N. and Chaput, J.C. (2026) How custom polymerases are driving innovation in synthetic biology. *Curr Opin Chem Biol*, **92**.

Hajjar, M., Maola, V.A., Lee, J.J., Holguin, M.J., Quijano, R.N., Nguyen, K.K., Ho, K.L., Medina, J.V., Botello-Cornejo, E., Barpuzary, B., Chim, N. and Chaput, J.C. (2026) Directed evolution of a TNA polymerase identifies independent paths to fidelity and catalysis. *Nat Commun*, **17**.

Kundu, N., McCloskey, C.M., Hajjar, M. and Chaput, J.C. (2023) Parameterizing the binding properties of XNA aptamers isolated from a low stringency selection. *Biochemistry*, **62**, 3245.

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

Lee, J.J., Kundu, N., Lozoya-Collinas, A., Lee, E.M., Hajjar, M., Setterholm, N.A., Chim, N. and Chaput, J.C. (2026) A scalable route to validated sequence–function datasets for AI-driven XNA aptamer discovery. Manuscript submitted.

Maola, V.A., Yik, E.J., Hajjar, M., Lee, J.J., Holguin, M.J., Quijano, R.N., Nguyen, K.K., Ho, K.L., Medina, J.V., Chim, N. and Chaput, J.C. (2024) Directed evolution of a highly efficient TNA polymerase achieved by homologous recombination. *Nature Catalysis*, **7**.

Medina, E.L., Maola, V.A., Hajjar, M., Ko, G.K., Ho, E.J., Horton, A.R., Chim, N. and Chaput, J.C. (2026) Rapid evolution of a highly efficient RNA polymerase by homologous recombination. *Nature Chemical Biology*, **22**.

Vincent, E., Maola, V.A., Lopez, C., Borie-Guichot, M., Dalla Pozza, M., Hajjar, M., Sieskind, R., Lucas-Staat, S., Navaza, R., Mulard, L., Hollenstein, M., Chaput, J.C. and Delarue, M. (2026) Polymerase evolution enables access to glycosylated xeno-nucleic acids with expanded chemical functionality. *bioRxiv*, 2026.07.20.739590.

ABSTRACT OF THE DISSERTATION

Directed Evolution and Structural Analysis of Engineered Polymerases

by

Mohammad Hajjar

Doctor of Philosophy in Pharmacological Sciences

University of California, Irvine, 2026

Professor John Chaput, Chair

Modern chemical biology seeks genetic polymers with enhanced pharmacological and physicochemical properties for applications in therapeutics, diagnostics, and biotechnology. Reprogramming DNA polymerases offers a controlled, scalable, and surprisingly versatile method of synthesis, paving the way for the selection of functional polymers. However, the precision that makes DNA polymerases finely tuned, fast, and faithful molecular machines also limits their ability to accept noncognate substrates. This thesis examines the adaptability of hyperthermophilic archaeal B-family DNA polymerases toward polymers with unique benefits and challenges: phosphonomethylthreosyl nucleic acid (pTNA), threose nucleic acid (TNA), and RNA.

pTNA is an orthogonal polymer intended to remain insulated from cellular DNA and RNA, raising fundamental questions about how it can be copied from a DNA template. Crystallographic analysis showed that an engineered DNA polymerase can organize canonical Watson-Crick base pairing between nascent pTNA and templating DNA, stabilizing an otherwise unfavorable DNA–pTNA heteroduplex.

TNA, by contrast, can exchange genetic information with DNA and resists nuclease degradation, making it valuable for molecular evolution and biotechnology. By combining homologous recombination and random mutagenesis with high-throughput microfluidic screening, this work reports the directed evolution of TNA polymerase 10-92, which synthesizes TNA at ~1

nt s⁻¹ with >99% fidelity. Its utility is demonstrated by efficient synthesis of TNA aptamers, including molecules containing C5-modified uracil residues.

RNA is a highly versatile natural genetic polymer, and its synthesis can benefit from a thermostable, primer-dependent polymerase. Homologous recombination and microfluidic screening yielded RNA polymerase C28, which operates at ~3 nt s⁻¹ with >99% fidelity and supports long-range RNA synthesis, reverse transcription, chimeric DNA–RNA PCR amplification, and incorporation of several 2'-fluoro and base-modified ribonucleotides.

Structures of 10-92 and its evolutionary intermediates revealed extensive, long-range remodeling that progressively refined the active site to accommodate TNA. Fidelity and catalytic rate followed distinct evolutionary trajectories, demonstrating that these properties can be improved through separable mechanisms. Mutational analysis further suggested that a highly stable scaffold permitted the accumulation of numerous substitutions predicted to be stability-neutral, whereas a smaller set of distal mutations, predicted to be destabilizing, fine-tuned catalysis by altering flexibility and rigidity. Structural analysis of C28 also shows that efficient RNA synthesis arose from steric-gate removal and altered electrostatics, and likely from distributed interactions that remodeled substrate recognition.

Together, this work establishes 10-92 and C28 as practical tools for synthetic genetics and reveals how polymerases can stabilize otherwise unfavorable base pairing, reshape catalytic geometry through distant sequence changes, and enhance fidelity through mechanisms separable from catalytic rate.

INTRODUCTION

Polymerases deserve contemplation. Their function sustains one of the defining features of biological life: the maintenance and propagation of genetic information within cells. What other molecular machinery bears a greater responsibility? And yet, every polymerase begins as a humble sequence of amino acids, folding into an architecture of astonishing order and catalytic precision.

The very geometric and chemical requirements that make a polymerase so well suited to its natural function can obscure its remarkable plasticity. Changes in the amino acid sequence can reshape activity, specificity, and substrate tolerance, enabling polymerases to extend beyond DNA and RNA into a far broader chemical space. This expanded repertoire includes nucleic acids with altered sugars, backbones, and nucleobases that possess properties unavailable to their natural counterparts and offer considerable potential in diagnostics, therapeutics, and biotechnology.

The tension between catalytic precision and functional plasticity forms the central subject of this dissertation. Through laboratory-directed evolution coupled with biochemical and structural characterization, the work presented here examines how polymerases can be reprogrammed, how sequence changes alter molecular geometry, and how those structural changes reshape catalytic function. It further asks how far a polymerase can depart from its natural activity while retaining the defining features of an efficient enzyme. In doing so, this work seeks both to develop new tools for biotechnology and medicine and to illuminate the principles governing polymerase function.

The Logic and Architecture of Replicative DNA Synthesis

Consider a replicative DNA polymerase. Bespoke to the chemistry it performs, the enzyme is sculpted around a primer-template junction: a DNA duplex that ends asymmetrically, with one

strand, the primer, providing the reactive terminus from which synthesis proceeds, and the other, the template, extending beyond it to dictate the sequence yet to be written. The template conveys this information through the complementary pairing of adenine (A) with thymine (T) and cytosine (C) with guanine (G) (1). At the primer-template junction, the polymerase arranges the primer 3'-hydroxyl, templating base, incoming deoxyribonucleoside triphosphate (dNTP), catalytic metal ions, and conserved residues into a single reaction center. These components must be positioned with ångström-scale precision; even subtle displacement can separate productive chemistry from arrest. In the catalytically competent complex, the primer 3'-hydroxyl is aligned for in-line nucleophilic attack on the α -phosphate of the incoming dNTP. A phosphodiester bond forms, extending the primer by one nucleotide, and pyrophosphate is released (2-4).

During nucleotide addition, the polymerase moves through a coordinated series of conformational states within a fold often likened to a right hand (5). The polymerase domain comprises palm, thumb, and fingers subdomains. The palm contains the conserved catalytic residues, the thumb positions the duplex and contributes to translocation and processivity, and the fingers organize the templating base and incoming nucleotide. Upon recognition of a cognate dNTP, the fingers close to complete the catalytically competent reaction center. Many replicative polymerases contain an N-terminal domain and a 3'-5' proofreading exonuclease domain that contribute to substrate surveillance and accurate synthesis (6,7).

Each incorporation is one step in a repeated catalytic cycle. The polymerase translocates along the duplex by one register, positions the next templating base within the active site, and catalyzes phosphodiester bond formation before reopening for the next round of synthesis (8). It must repeat this cycle while remaining associated with the growing nucleic acid across many successive incorporations. This capacity, known as processivity, describes the number of

nucleotides incorporated during a single polymerase–DNA binding event and enables long stretches of DNA to be copied without repeated dissociation and rebinding.

Not all DNA polymerases are adapted for sustained genome replication. Different polymerase families share common catalytic principles but differ in domain organization, nucleotide-addition rate, processivity, fidelity, and substrate tolerance, reflecting their cellular roles. Some carry out rapid chromosome replication, whereas others specialize in DNA repair, lesion bypass, or short-patch synthesis (9,10). This dissertation focuses on hyperthermophilic archaeal B-family polymerases (11). Their thermostability, high processivity and fidelity, and comparatively broad tolerance of modified nucleotides make them useful models of replicative synthesis and productive starting scaffolds for polymerase engineering (12).

KOD DNA polymerase, from the hyperthermophilic archaeon *Thermococcus kodakarensis*, is a B-family polymerase that exemplifies the remarkable catalytic rate of a replicative polymerase, extending DNA at approximately 200 nucleotides per second under optimal conditions (13,14). This rate is more remarkable because it is coupled to high fidelity. At each templating position, a faithful polymerase must select the correct deoxyribonucleotide from three incorrect alternatives (15). Base pairing contributes to this discrimination but does not determine it alone (16). Kinetic studies of other high-fidelity DNA polymerases show how induced fit can partition bound substrates between chemistry and release: correct dNTPs favor conformational states committed to incorporation, whereas mismatches favor poorly aligned states from which dissociation is likely (17,18). Misincorporated nucleotides that escape initial selection may subsequently be excised by the 3'–5' proofreading exonuclease.

Polymerases must also discriminate among chemically similar nucleotide substrates. Ribonucleotides differ from deoxyribonucleotides by a 2'-hydroxyl group in place of hydrogen on the sugar. They are present at substantially higher concentrations within cells, yet replicative DNA polymerases largely exclude them through a conserved steric-gate residue and additional structural checkpoints (19,20). Successful synthesis therefore depends not only on nucleobase pairing, but also on recognition of the sugar and triphosphate, accommodation of the primer terminus, and maintenance of a duplex geometry compatible with continued extension.

The precision required for rapid and faithful DNA synthesis consequently restricts the enzyme's chemical range. Polymerases are not general catalysts for joining nucleotide-like molecules; they are integrated systems tuned to particular dimensions, stereochemistry, electrostatics, and backbone geometries. Even modest substrate modifications can disrupt binding, active-site assembly, phosphodiester-bond formation, translocation, or continued extension. Reprogramming a polymerase therefore involves more than simply enlarging a binding pocket or removing an isolated steric clash. It often requires remodeling interactions within and beyond the active site so that a noncognate triphosphate can be recognized and incorporated, and the resulting primer terminus can support subsequent rounds of extension while preserving template-directed synthesis.

Beyond Natural Nucleic Acid Chemistry

Harnessing polymerases in the laboratory transformed nucleic acid synthesis into a programmable technology. Their controlled use enabled the polymerase chain reaction (PCR), DNA sequencing, molecular cloning, mutagenesis, and many other methods that now define molecular biology (21-24). Early engineering largely optimized polymerases for natural or near-

natural reactions, improving properties such as thermal stability, catalytic rate, processivity, fidelity, and resistance to reaction inhibitors (25-31). Even when modified substrates were introduced, the resulting polymers generally retained the chemical framework of DNA or RNA. The rapid expansion of chemical and synthetic biology introduced more ambitious challenges, including altering the chemistry of the polymer itself. Each nucleotide contains three chemically addressable components: a nucleobase, a sugar, and an internucleotide linkage, and each component can be modified independently or in combination. Nucleobase modifications can introduce new pairing interactions, expand the genetic alphabet, or add functional groups beyond those available to natural nucleotides. Changes to the backbone alter the repeating scaffold more fundamentally, thereby affecting the polymer's dimensions, charge, flexibility, and conformational preferences. DNA and RNA are a part of a much larger conceivable space of information-bearing molecules (32,33).

Many such polymers are described as xeno-nucleic acids, or XNAs. The term was initially introduced for nucleic acids in which an alternative scaffold replaces the ribose or deoxyribose sugar, although it is sometimes used more broadly to refer to artificial genetic polymers (34). This chemical space spans closely modified forms of DNA and RNA to major changes in sugar, ring size, backbone length, stereochemistry, and internucleotide connectivity (32,33).

Changing the repeating scaffold alters more than the chemical composition of the polymer. It changes the spacing and orientation of successive nucleobases and can influence sugar conformation, base stacking, helical geometry, groove dimensions, duplex stability, and molecular recognition. Depending on the system, these changes can confer increased resistance to nucleases, enhanced chemical stability, altered hybridization strength, or access alternative folding landscapes. Many XNAs retain sequence-specific pairing with natural nucleic acids, allowing

information to pass between DNA, RNA, and XNA (35). In contrast, others form orthogonal pairing systems that do not readily cross-pair with natural polymers (36,37). This combination of chemical diversity and sequence programmability has made synthetic nucleic acids attractive for biotechnology and medicine. Increased biological stability can support affinity reagents and catalysts that remain functional in environments where DNA or RNA would be rapidly degraded. Modification can also produce binding surfaces and catalytic folds distinct from those available to natural nucleic acids. Engineered systems have consequently enabled the selection of biologically stable aptamers and XNA enzymes, commonly called XNAzymes, and have opened related opportunities in therapeutics, diagnostics, information storage, and molecular materials (24,33,38-42).

However, for these polymers to become evolvable molecular systems, chemical synthesis alone is insufficient. Solid-phase synthesis can produce oligonucleotides of defined sequence. Still, polymerase-mediated synthesis is needed to generate large, sequence-diverse libraries and to copy selected molecules through repeated cycles of selection and amplification. A complete write-read cycle generally requires an enzyme that can synthesize the modified polymer from a DNA template and a reverse transcriptase that can recover its sequence in DNA for amplification and analysis. As such, polymerases connect the chemical invention of a synthetic nucleic acid to its replication, selection, and evolution (35).

Natural polymerases are poorly suited to this task because their structural and kinetic checkpoints are adapted to natural substrates. A modified triphosphate may bind but fail to induce productive active-site closure or achieve the geometry required for phosphodiester-bond formation. It may be incorporated once yet produce a primer terminus that cannot be extended efficiently. As modified residues accumulate, changes in duplex geometry can disrupt primer-

template positioning, translocation, and subsequent nucleotide addition, leading to stalling and reduced processivity. Acceptance of a single modified nucleotide is therefore fundamentally different from sustained synthesis of a fully modified polymer. Engineering a polymerase for this purpose requires coordinated adaptation of the catalytic cycle so that a noncognate substrate can be selected, incorporated, and repeatedly extended with useful rate, processivity, and fidelity. It is to make unfamiliar chemistry proceed with something approaching the fluency once reserved for the natural one.

From Pentose to Tetrose: TNA and pTNA

Among synthetic genetic polymers, α -L-threofuranosyl nucleic acid (TNA) is a striking departure from DNA and RNA. TNA replaces their five-carbon sugars and 3'–5' linkages with a four-carbon threose scaffold with 3'–2' phosphodiester linkages. Its backbone repeat consequently contains one fewer atom than those of natural nucleic acids. Despite this contraction, TNA forms stable antiparallel Watson-Crick duplexes with complementary TNA, DNA, and RNA (43). Sequence information can therefore be transferred from DNA into TNA and recovered by reverse transcription into DNA (44-48). This exchange enables TNA libraries to undergo repeated cycles of synthesis, selection, amplification, and molecular evolution. TNA has substantial nuclease resistance, stability under simulated physiological conditions, and resistance to acid-mediated degradation (49,50), properties that have supported the development of biologically stable aptamers, catalysts, and polymers for diverse applications (51-59).

These structural differences present substantial challenges for polymerases. A TNA nucleoside triphosphate (tNTP) differs from a dNTP in sugar composition, connectivity, and conformational preference. Following incorporation, the enzyme must extend a threose-terminated

primer while accommodating a growing TNA-DNA heteroduplex. Early engineered polymerases established that DNA-templated TNA synthesis and TNA reverse transcription were possible, but generally exhibited low catalytic rates, limited full-length synthesis, and dependence on permissive conditions such as manganese ions, which can compromise fidelity (47,48).

Phosphonomethylthreosyl nucleic acid (pTNA) extends this challenge by modifying the internucleotide linkage. Relative to TNA, pTNA inserts a methylene unit between the threose 3' position and phosphorus, producing a 3'–2' phosphonomethyl backbone (36). This restores a six-atom backbone repeat comparable in length to that of DNA while retaining a scaffold distinct from both DNA and TNA. Unlike TNA, pTNA does not form stable antiparallel duplexes with DNA or RNA and is therefore considered an orthogonal genetic polymer. This orthogonality promises alternate applications but also complicates synthesis, requiring the polymerase to maintain a nascent pTNA-DNA complex that is not intrinsically stable in solution (37,60). TNA and pTNA in related but distinct aspects test the substrate scope and push the limits of polymerase adaptation for synthetic biology.

A Natural but Noncognate Polymerase Substrate: RNA

RNA occupies a distinct position among the polymers considered here. It is a natural nucleic acid synthesized throughout biology by dedicated RNA polymerases, yet ribonucleoside triphosphates (NTPs) are strongly noncognate substrates for the replicative DNA polymerases used as engineering scaffolds in this dissertation.

RNA differs from DNA, and it has distinct steric features and conformational preference of the sugar. A ribose carries a 2'-hydroxyl group, and uracil (U) replaces thymine (T), lacking its 5-methyl group. Because intracellular NTP concentrations can substantially exceed those of

dNTPs by up to 100-fold (61), replicative DNA polymerases use several structural checkpoints to prevent ribonucleotide incorporation and preserve genome integrity, including a bulky steric-gate residue obstructing the ribose 2'-hydroxyl (19,62,63). Engineering a DNA polymerase for efficient RNA synthesis stretches beyond identifying and mutating steric gates, as substrate selectivity is distributed across the enzyme (42,64,65). Efficient synthesis requires extension from a ribose-terminated primer, accommodation of the growing RNA–DNA heteroduplex, and sustained translocation and nucleotide addition. Rather than initiating synthesis *de novo* at a promoter, an engineered DNA polymerase extends a preassembled primer-template junction, allowing additional control over the sequence and 5' region.

An efficient DNA-dependent RNA polymerase also provides a platform for testing base- and sugar-modified NTPs, including 2'-substituted analogs used in therapeutic oligonucleotides. Full-length synthesis is more demanding than isolated incorporation because every addition creates a modified primer terminus and further alters the product duplex.

Navigating Polymerase Sequence Space

Rational design offers the most direct route to modifying polymerase function. Structural information can identify steric gates and residues that contact the incoming nucleotide or primer–template duplex, allowing targeted mutations to broaden substrate acceptance or improve extension from noncognate primer termini. When the principal barrier is local and structurally apparent, a few substitutions may produce gains in activity. Rational design becomes less reliable, however, when the desired phenotype depends on interactions distributed throughout the catalytic cycle. Static structures cannot fully capture conformational dynamics, long-range interactions, or the effects of epistasis and sequence context (66). In a large protein, like a B-family polymerase,

distal residues may influence packing, electrostatics, domain motion, or duplex positioning. Complex polymerase functions are often difficult to predict or construct one substitution at a time. Directed evolution addresses this uncertainty by searching experimentally rather than requiring a complete mechanistic model in advance. Large libraries of related polymerases are screened or selected for the desired phenotype, and productive variants are recovered, diversified, and validated. Repeated rounds can accumulate mutations that transform weak activity into efficient noncognate synthesis, after which lead variants are validated biochemically.

Different diversification methods explore different regions of sequence space (67). Homologous recombination constructs mosaic genes from related proteins that share a common fold but differ at positions distributed throughout their sequences (68,69). Because the exchanged sequence blocks originate from functional parent enzymes, recombinant libraries can sample broad combinations while retaining a reasonable probability of producing folded proteins. Recombination may also transfer groups of interacting residues whose collective effects would be difficult to reproduce through sequential point mutation (70).

Error-prone PCR instead introduces randomly dispersed point mutations and explores the local sequence space surrounding an existing scaffold (71). It can uncover beneficial substitutions at positions that would not have been targeted by rational design. Most random substitutions, however, are neutral or deleterious, and increasing the mutation load raises the likelihood that useful changes will be accompanied by defects in folding, DNA binding, or catalysis (70).

These methods are particularly effective when combined. Recombination can generate new functional scaffolds, after which error-prone PCR and targeted mutagenesis can refine activity within a promising background. This strategy underlies the work presented in this dissertation; B-

family sequence diversity was used to access noncognate polymerase function, followed by progressive optimization and biochemical and structural characterization. Directed evolution identifies functional solutions before mechanistic characterization suggests how those solutions were achieved.

Artificial intelligence has already produced major advances in biomolecular structure prediction and de novo protein design (72,73). Polymerases remain a more difficult target because their behavior emerges from a dynamic, multistep catalytic cycle rather than a single structure, but polymerase engineering is expected to become increasingly AI-guided as richer functional, kinetic, and structural datasets become available.

Compartmentalized Screening of Polymerase Function

Polymerase libraries are useful only if each enzyme's activity can be linked to the gene that encodes it. In a bulk reaction, enzymes and products diffuse freely, making it impossible to identify the variant responsible for a successful synthesis event. Compartmentalization preserves this genotype–phenotype link by isolating individual polymerase variants and their encoding plasmids within discrete reaction volumes (74-76).

Droplet-based optical polymerase sorting (DrOPS) achieves this isolation in water-in-oil microdroplets, typically about 20 μm in diameter (77). A polymerase library is expressed in bacterial cells, with each cell carrying a plasmid encoding one variant. Cells are diluted to low occupancy and encapsulated with a fluorescent primer–template reporter, a complementary quencher strand, nucleotide triphosphates, and the remaining reaction components. Heat lysis releases the polymerase and its plasmid into the droplet. Productive full-length primer extension

prevents quencher binding and produces a fluorescent signal while the corresponding gene remains confined to the same compartment (78).

The droplets are then analyzed by fluorescence-activated sorting. Fluorescent droplets are collected, and their plasmid DNA is recovered for characterization or further diversification. This approach allows millions of polymerase variants to be screened rapidly for the desired synthetic activity.

Scope of This Dissertation

This dissertation examines how hyperthermophilic archaeal B-family DNA polymerases can be adapted to synthesize pTNA, TNA, and RNA. Across four studies, directed evolution, biochemical characterization, and X-ray crystallography are used to investigate how changes in polymerase sequence alter substrate recognition, catalytic activity, and fidelity. The first study examines engineered polymerases that synthesize pTNA. Crystallographic and biochemical analyses compare polymerase complexes containing pTNA, TNA, and DNA products to investigate how an orthogonal phosphonomethyl backbone is accommodated during synthesis. The second study describes the evolution of 10–92 through homologous recombination, droplet-based screening, and mutagenesis. The resulting polymerase synthesizes TNA at approximately 1 nucleotide per second with 99% fidelity. The third study follows the evolutionary trajectory from an early TNA polymerase to 10–92. Biochemical and structural analyses of key intermediates examine how catalytic activity, fidelity, thermostability, and substrate preference changed during successive stages of evolution. The final study applies homologous recombination and droplet-based screening to RNA synthesis. This work produced C28, a DNA-dependent RNA polymerase that synthesizes RNA at approximately 3 nucleotides per second with greater than 99% fidelity,

and then characterized the evolved enzyme biochemically and structurally. Together, these studies examine how polymerase function can be redirected toward distinct noncognate substrates and how those new activities emerge during laboratory evolution.

0.1 REFERENCES

1. Watson, J.D. and Crick, F.H.C. (1953) Molecular Structure of Nucleic Acids - a Structure for Deoxyribose Nucleic Acid. *Nature*, **171**, 737-738.
2. Steitz, T.A. (1998) Structural biology - A mechanism for all polymerases. *Nature*, **391**, 231-232.
3. Steitz, T.A., Smerdon, S.J., Jager, J. and Joyce, C.M. (1994) A Unified Polymerase Mechanism for Nonhomologous DNA and Rna-Polymerases. *Science*, **266**, 2022-2025.
4. Johnson, K.A. (2010) The kinetic and chemical mechanism of high-fidelity DNA polymerases. *Bba-Proteins Proteom*, **1804**, 1041-1048.
5. Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G. and Steitz, T.A. (1985) Structure of Large Fragment of Escherichia-Coli DNA-Polymerase-I Complexed with Dtmp. *Nature*, **313**, 762-766.
6. Doubl  , S., Tabor, S., Long, A.M., Richardson, C.C. and Ellenberger, T. (1998) Crystal structure of a bacteriophage T7 DNA replication complex at 2.2   resolution. *Nature*, **391**, 251-258.
7. Kiefer, J.R., Mao, C., Braman, J.C. and Beese, L.S. (1998) Visualizing DNA replication in a catalytically active DNA polymerase crystal. *Nature*, **391**, 304-307.
8. Chim, N., Meza, R.A., Trinh, A.M., Yang, K.F. and Chaput, J.C. (2021) Following replicative DNA synthesis by time-resolved X-ray crystallography. *Nature Communications*, **12**.
9. Yang, W. and Gao, Y. (2018) Translesion and Repair DNA Polymerases: Diverse Structure and Mechanism. *Annu Rev Biochem*, **87**, 239-261.
10. Braithwaite, D.K. and Ito, J. (1993) Compilation, Alignment, and Phylogenetic-Relationships of DNA-Polymerases. *Nucleic Acids Res*, **21**, 787-802.
11. Hashimoto, H., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., Inoue, T. and Kai, Y. (2001) Crystal structure of DNA polymerase from hyperthermophilic archaeon KOD1. *J Mol Biol*, **306**, 469-477.
12. Nikoomanzar, A., Chim, N., Yik, E.J. and Chaput, J.C. (2020) Engineering polymerases for applications in synthetic biology. *Q Rev Biophys*, **53**.
13. Nikoomanzar, A., Dunn, M.R. and Chaput, J.C. (2017) Evaluating the Rate and Substrate Specificity of Laboratory Evolved XNA Polymerases. *Anal Chem*, **89**, 12622-12625.
14. Griep, M.A., Kotera, C.A., Nelson, R.M. and Viljoen, H.J. (2006) Kinetics of the DNA polymerase. *Chem Eng Sci*, **61**, 3885-3892.

15. Kunkel, T.A. and Bebenek, R. (2000) DNA replication fidelity. *Annual Review of Biochemistry*, **69**, 497-529.
16. Kool, E.T. (2002) Active site tightness and substrate fit in DNA replication. *Annual Review of Biochemistry*, **71**, 191-219.
17. Johnson, K.A. (2008) Role of induced fit in enzyme specificity: A molecular forward/reverse switch. *J Biol Chem*, **283**, 26297-26301.
18. Wong, I., Patel, S.S. and Johnson, K.A. (1991) An Induced-Fit Kinetic Mechanism for DNA-Replication Fidelity - Direct Measurement by Single-Turnover Kinetics. *Biochemistry-U.S.*, **30**, 526-537.
19. Astatke, M., Ng, K.M., Grindley, N.D.F. and Joyce, C.M. (1998) A single side chain prevents DNA polymerase I (Klenow fragment) from incorporating ribonucleotides. *P Natl Acad Sci USA*, **95**, 3402-3407.
20. Brown, J.A. and Suo, Z.C. (2011) Unlocking the Sugar "Steric Gate" of DNA Polymerases. *Biochemistry-U.S.*, **50**, 1135-1142.
21. Takagi, M., Nishioka, M., Kakihara, H., Kitabayashi, M., Inoue, H., Kawakami, B., Oka, M. and Imanaka, T. (1997) Characterization of DNA polymerase from *Pyrococcus* sp. strain KOD1 and its application to PCR. *Appl Environ Microb*, **63**, 4504-4510.
22. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA Sequencing with Chain-Terminating Inhibitors. *P Natl Acad Sci USA*, **74**, 5463-5467.
23. Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. and Arnheim, N. (1985) Enzymatic Amplification of Beta-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle-Cell Anemia. *Science*, **230**, 1350-1354.
24. Taylor, A.I., Pinheiro, V.B., Smola, M.J., Morgunov, A.S., Peak-Chew, S., Cozens, C., Weeks, K.M., Herdewijn, P. and Holliger, P. (2015) Catalysts from synthetic genetic polymers. *Nature*, **518**, 427-430.
25. Ishino, S. and Ishino, Y. (2014) DNA polymerases as useful reagents for biotechnology - the history of developmental research in the field. *Front Microbiol*, **5**.
26. Aschenbrenner, J. and Marx, A. (2017) DNA polymerases and biotechnological applications. *Curr Opin Biotech*, **48**, 187-195.
27. Coulther, T.A., Stern, H.R. and Beuning, P.J. (2019) Engineering Polymerases for New Functions. *Trends Biotechnol*, **37**, 1091-1103.
28. Davidson, J.F., Fox, R., Harris, D.D., Lyons-Abbott, S. and Loeb, L.A. (2003) Insertion of the T3 DNA polymerase thioredoxin binding domain enhances the processivity and fidelity of

DNA polymerase. *Nucleic Acids Res*, **31**, 4702-4709.

29. Wang, Y., Prosen, D.E., Mei, L., Sullivan, J.C., Finney, M. and Vander Horn, P.B. (2004) A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance. *Nucleic Acids Res*, **32**, 1197-1207.

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

31. Baar, C., d'Abbadie, M., Vaisman, A., Arana, M.E., Hofreiter, M., Woodgate, R., Kunkel, T.A. and Holliger, P. (2011) Molecular breeding of polymerases for resistance to environmental inhibitors. *Nucleic Acids Res*, **39**.

32. Anosova, I., Kowai, E.A., Dunn, M.R., Chaput, J.C., Van Horn, W.D. and Egli, M. (2016) The structural diversity of artificial genetic polymers. *Nucleic Acids Res*, **44**, 1007-1021.

33. Chaput, J.C., Egli, M. and Herdewijn, P. (2025) The XNA alphabet. *Nucleic Acids Res*, **53**.

34. Chaput, J.C. and Herdewijn, P. (2019) What Is XNA? *Angew Chem Int Edit*, **58**, 11570-11572.

35. Pinheiro, V.B., Taylor, A.I., Cozens, C., Abramov, M., Renders, M., Zhang, S., Chaput, J.C., Wengel, J., Peak-Chew, S.Y., McLaughlin, S.H. *et al.* (2012) Synthetic Genetic Polymers Capable of Heredity and Evolution. *Science*, **336**, 341-344.

36. Liu, C., Cozens, C., Jaziri, F., Rozenski, J., Maréchal, A., Dumbre, S., Pezo, V., Marlière, P., Pinheiro, V.B., Groaz, E. *et al.* (2018) Phosphonomethyl Oligonucleotides as Backbone-Modified Artificial Genetic Polymers. *J Am Chem Soc*, **140**, 6690-6699.

37. Chaput, J.C., Herdewijn, P. and Hollenstein, M. (2020) Orthogonal Genetic Systems. *Chembiochem*, **21**, 1408-1411.

38. Taylor, A.I., Wan, C.J.K., Donde, M.J., Peak-Chew, S.Y. and Holliger, P. (2022) A modular XNAzyme cleaves long, structured RNAs under physiological conditions and enables allele-specific gene silencing. *Nat Chem*, **14**, 1295-+.

39. Nguyen, K., Malik, T.N. and Chaput, J.C. (2023) Chemical evolution of an autonomous DNAzyme with allele-specific gene silencing activity. *Nature Communications*, **14**.

40. Schuder, D.N., Lu, N.D. and Chaput, J.C. (2024) Revealr-Based Diagnostic Panel for Rapid Detection of Acute Respiratory Infections. *Acs Synth Biol*, **13**, 4202-4208.

41. Lee, E.M., Nguyen, K., Setterholm, N.A., Malik, T.N. and Chaput, J.C. (2026) Allele-specific knockdown by an engineered DNAzyme capable of RNase H1 evasion. *Nucleic Acids Res*, **54**.

42. Cozens, C., Pinheiro, V.B., Vaisman, A., Woodgate, R. and Holliger, P. (2012) A short adaptive path from DNA to RNA polymerases. *P Natl Acad Sci USA*, **109**, 8067-8072.

43. Schöning, K.U., Scholz, P., Guntha, S., Wu, X., Krishnamurthy, R. and Eschenmoser, A. (2000) Chemical etiology of nucleic acid structure: The α -threofuranosyl-(3'→2') oligonucleotide system. *Science*, **290**, 1347-1351.
44. Chaput, J.C. and Szostak, J.W. (2003) TNA synthesis by DNA polymerases. *J Am Chem Soc*, **125**, 9274-9275.
45. Dunn, M.R. and Chaput, J.C. (2016) Reverse Transcription of Threose Nucleic Acid by a Naturally Occurring DNA Polymerase. *Chembiochem*, **17**, 1804-1808.
46. Chaput, J.C., Ichida, J.K. and Szostak, J.W. (2003) DNA polymerase-mediated DNA synthesis on a TNA template. *J Am Chem Soc*, **125**, 856-857.
47. Nikoomanzar, A., Vallejo, D., Yik, E.J. and Chaput, J.C. (2020) Programmed Allelic Mutagenesis of a DNA Polymerase with Single Amino Acid Resolution. *Acs Synth Biol*, **9**, 1873-1881.
48. Dunn, M.R., Otto, C., Fenton, K.E. and Chaput, J.C. (2016) Improving Polymerase Activity with Unnatural Substrates by Sampling Mutations in Homologous Protein Architectures. *Acs Chem Biol*, **11**, 1210-1219.
49. Culbertson, M.C., Temburnikar, K.W., Sau, S.P., Liao, J.Y., Bala, S. and Chaput, J.C. (2016) Evaluating TNA stability under simulated physiological conditions. *Bioorg Med Chem Lett*, **26**, 2418-2421.
50. Lee, E.M., Setterholm, N.A., Hajjar, M., Barpuzary, B. and Chaput, J.C. (2023) Stability and mechanism of threose nucleic acid toward acid-mediated degradation. *Nucleic Acids Res*, **51**, 9542-9551.
51. Dunn, M.R., McCloskey, C.M., Buckley, P., Rhea, K. and Chaput, J.C. (2020) Generating Biologically Stable TNA Aptamers that Function with High Affinity and Thermal Stability. *J Am Chem Soc*, **142**, 7721-7724.
52. Li, X.T., Li, Z. and Yu, H.Y. (2020) Selection of threose nucleic acid aptamers to block PD-1/PD-L1 interaction for cancer immunotherapy. *Chem Commun*, **56**, 14653-14656.
53. Yang, K.F., McCloskey, C.M. and Chaput, J.C. (2020) Reading and Writing Digital Information in TNA. *Acs Synth Biol*, **9**, 2936-2942.
54. Zhang, L. and Chaput, J.C. (2020) In Vitro Selection of an ATP-Binding TNA Aptamer. *Molecules*, **25**.
55. Wang, Y., Wang, Y.Y., Song, D.F., Sun, X., Zhang, Z., Li, X.T., Li, Z. and Yu, H.Y. (2021) A Threose Nucleic Acid Enzyme with RNA Ligase Activity. *J Am Chem Soc*, **143**, 8154-8163.
56. Wang, Y.J., Nguyen, K., Spitale, R.C. and Chaput, J.C. (2021) A biologically stable DNzyme that efficiently silences gene expression in cells. *Nat Chem*, **13**.

57. Wang, F., Liu, L.S., Li, P., Leung, H.M., Tam, D.Y. and Lo, P.K. (2022) Biologically stable threose nucleic acid-based probes for real-time microRNA detection and imaging in living cells. *Mol Ther-Nucl Acids*, **27**, 787-796.
58. Matsuda, S., Bala, S., Liao, J.Y., Datta, D., Mikami, A., Woods, L., Harp, J.M., Gilbert, J.A., Bisbe, A., Manoharan, R.M. *et al.* (2023) Shorter Is Better: The α -(1)-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**, 19691-19706.
59. Li, P., Zhu, C.Y., Liu, L.S., Han, C.T.J., Chu, H.C., Li, Z.H., Mao, Z.W., Wang, F. and Lo, P.K. (2024) Ultra-stable threose nucleic acid-based biosensors for rapid and sensitive nucleic acid detection and in vivo imaging. *Acta Biomater*, **177**, 472-485.
60. Schmidt, M. (2010) Xenobiology: A new form of life as the ultimate biosafety tool. *Bioessays*, **32**, 322-331.
61. Traut, T.W. (1994) Physiological Concentrations of Purines and Pyrimidines. *Mol Cell Biochem*, **140**, 1-22.
62. Bonnin, A., Lázaro, J.M., Blanco, L. and Salas, M. (1999) A single tyrosine prevents insertion of ribonucleotides in the eukaryotic-type ϕ 29 DNA polymerase. *J Mol Biol*, **290**, 241-251.
63. Gardner, A.F. and Jack, W.E. (1999) Determinants of nucleotide sugar recognition in an archaeon DNA polymerase. *Nucleic Acids Res*, **27**, 2545-2553.
64. Freund, N., Taylor, A.I., Arangundy-Franklin, S., Subramanian, N., Peak-Chew, S.Y., Whitaker, A.M., Freudenthal, B.D., Abramov, M., Herdewijn, P. and Holliger, P. (2023) A two-residue nascent-strand steric gate controls synthesis of 2'-methyl- and 2'-(2-methoxyethyl)-RNA. *Nat Chem*, **15**, 91-+.
65. Xia, G., Chen, L.J., Sera, T., Fa, M., Schultz, P.G. and Romesberg, F.E. (2002) Directed evolution of novel polymerase activities: Mutation of a DNA polymerase into an efficient RNA polymerase. *P Natl Acad Sci USA*, **99**, 6597-6602.
66. Chim, N., Shi, C.H., Sau, S.P., Nikoosmanzar, A. and Chaput, J.C. (2017) Structural basis for TNA synthesis by an engineered TNA polymerase. *Nature Communications*, **8**.
67. Zeymer, C. and Hilvert, D. (2018) Directed Evolution of Protein Catalysts. *Annu Rev Biochem*, **87**, 131-157.
68. Minshull, J. and Stemmer, W.P.C. (1999) Protein evolution by molecular breeding. *Curr Opin Chem Biol*, **3**, 284-290.
69. Stemmer, W.P.C. (1994) Rapid Evolution of a Protein in-Vitro by DNA Shuffling. *Nature*, **370**, 389-391.
70. Romero, P.A. and Arnold, F.H. (2009) Exploring protein fitness landscapes by directed evolution. *Nat Rev Mol Cell Bio*, **10**, 866-876.

71. Cadwell, R.C. and Joyce, G.F. (1994) Mutagenic Pcr. *Pcr Meth Appl*, **3**, S136-S140.
72. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583-+.
73. Watson, J.L., Juergens, D., Bennett, N.R., Trippe, B.L., Yim, J., Eisenach, H.E., Ahern, W., Borst, A.J., Ragotte, R.J., Milles, L.F. *et al.* (2023) De novo design of protein structure and function with RFdiffusion. *Nature*, **620**.
74. Ghadessy, F.J., Ong, J.L. and Holliger, P. (2001) Directed evolution of polymerase function by compartmentalized self-replication. *P Natl Acad Sci USA*, **98**, 4552-4557.
75. Larsen, A.C., Dunn, M.R., Hatch, A., Sau, S.P., Youngbull, C. and Chaput, J.C. (2016) A general strategy for expanding polymerase function by droplet microfluidics. *Nature Communications*, **7**.
76. Tawfik, D.S. and Griffiths, A.D. (1998) Man-made cell-like compartments for molecular evolution. *Nat Biotechnol*, **16**, 652-656.
77. Vallejo, D., Nikoomanzar, A., Paegel, B.M. and Chaput, J.C. (2019) Fluorescence-Activated Droplet Sorting for Single-Cell Directed Evolution. *Acs Synth Biol*, **8**, 1430-1440.
78. Yik, E.J., Maola, V.A. and Chaput, J.C. (2023) Engineering TNA polymerases through iterative cycles of directed evolution. *Method Enzymol*, **691**, 29-59.

CHAPTER 1: Crystallographic Analysis of Engineered Polymerases

Synthesizing Phosphonomethylthreosyl Nucleic Acid

PUBLICATION INFORMATION

This chapter is adapted from the following publication: Hajjar, M., Chim, N., Liu, C., Herdewijn, P. and Chaput, J.C. (2022) *Nucleic Acids Res*, **50**, 9663-9674, <https://doi.org/10.1093/nar/gkac792>

© The Author(s) 2022. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Data Availability

Atomic coordinates and structure factors for the reported crystal structures of Kod-RI/pTNA, Kod-RI/DNA and Kod-RSGA/pTNA have been deposited with the Protein Data Bank under accession numbers 7RSR, 7RSS and 7TQW, respectively.

Supplementary Data

Supplementary data associated with the publication are available through *Nucleic Acids Research* and are reproduced in Appendix A.

Acknowledgments

The authors thank members of the Chaput laboratory for their helpful comments on the manuscript.

Funding

National Science Foundation [2001434 and 1946312 to J.C.C.]; Belgian National Fund for Scientific Research [FWO-G085321N]. Funding for open access charge: National Science Foundation. The authors declare no conflicts of interest.

1.1 ABSTRACT

Xeno-nucleic acids (XNAs) are synthetic genetic polymers with backbone structures composed of non-ribose or non-deoxyribose sugars. Phosphonomethylthreosyl nucleic acid (pTNA), a type of XNA that does not base pair with DNA or RNA, has been suggested as a possible genetic material for storing synthetic biology information in cells. A critical step in this process is the synthesis of XNA episomes using laboratory-evolved polymerases to copy DNA information into XNA. Here, we investigate the polymerase recognition of pTNA nucleotides using X-ray crystallography to capture the post-catalytic complex of engineered polymerases following the sequential addition of two pTNA nucleotides onto the 3'-end of a DNA primer. High-resolution crystal structures reveal that the polymerase mediates Watson-Crick base pairing between the extended pTNA adducts and the DNA template. Comparative analysis studies demonstrate that the sugar conformation and backbone position of pTNA are structurally more similar to threose nucleic acid (TNA) than DNA even though pTNA and DNA share the same six-atom backbone repeat length. Collectively, these findings provide new insight into the structural determinants that guide the enzymatic synthesis of an orthogonal genetic polymer, and may lead to the discovery of new variants that function with enhanced activity.

1.2 INTRODUCTION

All organisms store their genetic information in DNA and propagate that information through a process of genome replication and cell division. In a test tube, this process of information storage and propagation has been extended to xeno-nucleic acids (XNAs), artificial genetic polymers in which the natural deoxyribose sugar found in DNA has been replaced with an alternative sugar or sugar-like moiety (1). In a typical XNA replication cycle, XNA templates are reverse transcribed into DNA, amplified using the polymerase chain reaction (PCR), and transcribed back into XNA (2). This process of copying information back and forth between DNA and XNA requires engineered polymerases that are carefully evolved in the laboratory to recognize specific XNA substrates (3). Although XNA replication systems have been used to evolve nucleic acid aptamers and catalysts with backbone structures composed entirely of XNA (4-13), others have contemplated the potential for XNA to serve as a repository for synthetic biology information that is maintained inside a cell (14). Such XNA-modified organisms could overcome the biocontainment problem facing current genetically-modified organisms by providing a genetic firewall to separate synthetic biology information from the natural DNA information of the cell (15). XNA-modified organisms that escape containment would lose their synthetic biology information by being deprived of critical XNA building block materials that are not provided by nature but required to maintain a replicating XNA chromosome (16).

Phosphonomethylthreosyl nucleic acid (pTNA, also abbreviated PMT and tPhoNA, Figure 1a) is considered a possible XNA candidate for information storage in cellular systems (17). pTNA is an unusual example of an XNA; it contains modifications in both the sugar and phosphodiester moieties with α -L-threofuranosyl sugars connected by 3'→2' phosphonomethyl backbone linkages. This chemical configuration makes pTNA stable to nucleases that degrade

natural genetic polymers but prevents pTNA from forming stable antiparallel duplex structures with complementary strands of DNA and RNA (17). As such, pTNA is viewed as an orthogonal genetic polymer, as it does not cross-pair with the genetic polymers of life and, therefore, would not interfere with the normal nucleic acid processes of the cell.

Efforts to advance pTNA as an orthogonal genetic polymer have focused on developing engineered polymerases that can faithfully copy DNA sequences into pTNA. The most recent example is TgoT-EPFLH, an engineered polymerase that derives from a replicative B-family DNA polymerase isolated from the archaeal hyperthermophilic species *Thermococcus gorgonarius* (Tgo) (17). In the presence of mutagenic manganese ions, TgoT-EPFLH has been shown to synthesize pTNA oligonucleotides of lengths up to 57 pTNA residues. This finding, as well as related work on glycerol nucleic acid (GNA) (18), demonstrate that stable Watson-Crick base pairing between the primer and template strands is not an essential criterion for the enzymatic synthesis of XNA polymers on DNA templates.

Here we investigate the structural determinants that enable engineered polymerases to synthesize short segments of an orthogonal genetic polymer on a DNA template. Since we were unable to crystalize Tgo-EPFLH, we focused our analysis on Kod-RI, a laboratory-evolved polymerase that was previously developed to synthesize β -L-threose nucleic acid (TNA), a close structural analog of pTNA (19). X-ray crystal structures of the post-catalytic complex showing Kod-RI bound to a DNA substrate that had been extended with either two pTNA adenosine residues (ptA) or two 2'-deoxyadenosine residues (dA) were solved to high resolution. Both structures were compared to an identical structure previously solved by extending Kod-RI with two TNA adenosine residues (tA) (20). Additionally, we solved the crystal structure for the post-catalytic complex of Kod-RSGA, a further evolved TNA polymerase relative to Kod-RI (21),

bound to the same primer-template duplex following extension with two pTNA residues. Close inspection of the active site region of each polymerase reveals features that are unique to each system. Together, this work expands our understanding of the mechanisms that natural and engineered polymerases utilize to recognize their cognate substrates. We postulate that this information could be used to create new variants with improved XNA synthesis activity.

1.3 MATERIAL AND METHODS

Oligonucleotides and nucleotides

DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT) (Supplementary Table S1). Primers used for polymerase extension assays were purchased as a 5'-hexynyl derivative, then labeled with an IR-fluorescent dye via click chemistry using a CuBr solution and IR Dye Azide 680 or 800 CW (Glen Research). The template used for X-ray crystallography was 5'-Cy5-labeled and HPLC purified by IDT. TNA and pTNA nucleotides were chemically synthesized as previously described (17,22,23).

Protein expression and purification

Kod-RI and Kod-RSGA were cloned into pET21a (Novagen) and expressed and purified from *E. coli* as previously described (20). Briefly, the gene for each polymerase (Kod-RI or Kod-RSGA) was PCR amplified from a pGDR11 expression vector using primers P_f and P_r containing *Nde*I and *Not*I restriction enzyme sites, respectively. Digestion using *Nde*I and *Not*I restriction enzymes (New England Biolabs) and ligation introduced the amplicons into pET21a (Novagen). The new plasmids were sequence-verified (Retrogen) prior to use. Then, each plasmid was transfected into Acella[®] cells (Edge BioSystems). The cells were grown aerobically at 37°C in LB media containing 100 mg mL⁻¹ ampicillin to an OD₆₀₀ of 0.8 and induced with 1 mM isopropyl b-D-thiogalactoside (Gold Biotechnology) at 18°C for 20 hr. Cells were harvested by centrifugation (3315×g, 20 min, 4°C).

Harvested cells were lysed in lysis buffer (Table S2) by sonication, and the lysate was centrifuged (23,708×g, 30 min, 4°C). The clear supernatant was heat-treated (70°C, 20 min) to

precipitate endogenous *E. coli* proteins by centrifugation (23,708×g, 30 min, 4°C). The supernatant was filtered through a 0.2-microns sterilized filter, then loaded onto 5 mL HiTrap Q HP in tandem with a 5 mL heparin HP column (GE Healthcare) and washed with low salt buffer. The Q column was removed before eluting the polymerase in a high salt buffer (Table S2) with a linear gradient. Individual fractions were visualized by SDS-PAGE, pooled, and concentrated using a 30 kDa cutoff Amicon centrifugal filter (Millipore). The concentrated protein was purified by size exclusion chromatography (Superdex 200 HiLoad 16/600, GE Healthcare) pre-equilibrated with Kod buffer (Table S2) and concentrated to 10 mg mL⁻¹.

Primer extension assays

Primer extension assays were performed in 20-mL reaction volumes containing 1X ThermoPol® reaction buffer (New England Biolabs, Table S2), 0.5 mM IR-labelled primer (P₁), and 0.5 mM template (T₁) (Table S1). The primer and template strands were annealed by incubating at 95°C for 5 minutes and cooling on ice for 5 minutes, and the duplex was combined with 100 mM of the appropriate nucleotides. The reactions were initiated by adding 0.5 mM engineered polymerase pre-incubated with 1 mM MnCl₂ and incubated at 55°C for the specified times. DNA full-length markers were extended by Kod-RI or Kod-RSGA without MnCl₂. The reactions were quenched with stop buffer (Table S2), denatured for 10 minutes at 95°C, and analyzed by 15% denaturing polyacrylamide gel electrophoresis (10W, 1.5 hours). Gels were imaged using a LI-COR Odyssey CLx imager.

Protein crystallization

Laboratory-evolved Kod polymerases were crystallized with the identical DNA primer-template duplex following the addition of two adenosine residues for pTNA or DNA. We

refer to these complexes as Kod-RI/pTNA, Kod-RI/DNA, and Kod-RSGA/pTNA. For each complex, equal amounts of the primer P_c and the Cy5-labelled template T_c (Table S1) were annealed in Kod buffer (Table S2) supplemented with 20 mM MgCl₂ (Honeywell, Fluka) by heating to 95°C for 5 min and cooling for 10 min at 10°C. Next, 1.5 molar equivalents of the duplex were incubated with 5 mg mL⁻¹ polymerase at 37°C for 30 min to form the binary complex in solution. Finally, 5 molar equivalents of the appropriate nucleotide (dATP or ptADP) were incubated with the binary complexes at 37°C for 30 min to form the post-catalytic complex used for protein crystallization.

Optimal protein crystallization conditions were identified by screening commercial protein crystallization kits (Hampton Research and Qiagen) using a pipetting robot (Mosquito, TTP Techlab Ltd) to set up 0.5 mL hanging drops of equal volumes of sample and crystallization solutions in 96-well format containing 100 mL of crystallization solutions. Successful conditions were optimized over pH and precipitant concentration ranges using 24-well trays with 2 mL drops and 500 mL crystallization solutions. Crystallization reagents were purchased from commercial suppliers and were of analytical grade. Stock solutions of sodium acetate (BioWorld) and sodium citrate tribasic dihydrate (BioXtra, Sigma-Aldrich) were filtered before use. The crystallization condition of the reported Kod-RI/pTNA and Kod-RI/DNA structures was 0.2 M sodium acetate, 0.1 M sodium citrate pH 4.5-5.5, 7-10% (w/v) polyethylene glycol 4000, and the crystallization solution for the reported Kod-RSGA/pTNA structure was 0.1 M Mg(OAc)₂ tetrahydrate, 0.1 M MOPS pH 7, and 7-10% PEG 8K. Crystal trays were stored in the dark at 24°C, and crystals grew in ~4 weeks. Crystals of the reported structures were cryoprotected and dehydrated in 25% ethylene glycol and preserved in liquid nitrogen before data collection at the synchrotron.

Data collection, processing, and structural determination

The Kod-RI/pTNA dataset was collected at the Advanced Light Source (ALS) at the Lawrence Berkeley National Laboratory on 2021-05-21 at 100 K using a single wavelength protocol (1.0 Å) on beamline 8.2.1 equipped with a CCD detector, ADSC Quantum 315r. The dataset was reduced, integrated, and scaled by XDS (24). Pointless (25) and aimless (26) were run for further space group determination and final data merging. Molecular replacement was performed using Phaser (27) using the previously published Kod-RI/TNA model (PDB ID: 5VU9) (20). The two TNA nucleotides were replaced by pTNA ligands created using eLBOW (28). Iterative rounds of manual building using Coot (29) followed by refinement using phenix.refine (30) were performed in Phenix 1.13-2998 (31). Residues 757 until the C-terminus of the 774-residue Kod-RI and the template's 5' terminal nucleotide were not built due to poor electron density. Final rounds of refinement employed TLS parameters, where Kod-RI was partitioned into 5 TLS groups (residues 1-110, 111-214, 215-337, 338-552, and 553-756). Two additional TLS groups were the template and the primer. The stereochemistry and geometry of all structures were validated with Molprobity (32), and the final refinement parameters are summarized in Supplementary Table S3. Final coordinates and structure factors have been deposited into the Protein Data Bank, PDB ID: 7RSR. Polder maps (33) were created using the collected dataset and the final refinement of the model.

The Kod-RI/DNA dataset was collected at the Stanford Synchrotron Radiation Lightsource (SSRL) at the National Accelerator Laboratory on 2021-07-21 at 100 K using a single wavelength protocol (1.0 Å) on beamline BL 12-1 using a Pixel detector, Dectris Eiger X 16M. The indexing, integration, and merging were done using iMOSFLM (34), and molecular replacement was performed using Phaser on an early refinement of the Kod-RI/pTNA model.

The same refinement strategy was employed, and the same residues were built. The TLS groups were residues 1-225, 226-304, 305-552, 553-765, template, and primer. The structure was deposited with PDB ID: 7RSS.

The Kod-RSGA/pTNA dataset was collected at ALS on 2021-12-20 at 100 K using a single wavelength protocol (1.0 Å) on beamline 5.0.1 equipped with a pixel Dectris PILATUS3 2M detector. The dataset was processed using XDS, and molecular replacement was performed using Phaser on the Kod-RI/pTNA model and followed the same refinement strategy. The TLS groups were residues 1-110, 111-214, 215-337, 338-552, and 553-756, template, and primer. The structure was deposited with PDB ID: 7TQW.

Structural measurements, calculations, and visualization

Endocyclic torsion angles were extracted using PROSIT (35). Exocyclic torsion angles, base-pair parameters, and base-pair step parameters were extracted using Web 3DNA (36). Torsion angles for unnatural moieties not recognized by Web 3DNA and phosphate to phosphate (or phosphonate) distances were measured in PyMOL by Schrödinger (<https://pymol.org/2/>). Protein RMSD measurements were based on the alpha carbons and calculated in PyMOL. All structural alignments and molecular graphics were prepared using PyMOL; alignments in Figure 3 were based on the six nucleotides displayed for each genetic system (T4, T5, C6, G11, A12, and A13).

1.4 RESULTS

pTNA synthesis by Kod-RI TNA polymerase

Kod-RI is an engineered TNA polymerase that was evolved from a replicative B-family DNA polymerase isolated from the archaeal hyperthermophilic species *Thermococcus kodakarensis* (Kod) (19). The polymerase harbors the A485R and E664I mutations required for TNA synthesis on DNA templates, as well as the 3',5'-exonuclease silencing mutations (D141A, E143A) that prevent the removal of newly added TNA residues. The ability for Kod-RI to synthesize TNA has been extensively studied by X-ray crystallography and structures now exist for each of the major intermediates in the TNA synthesis pathway, including the apo, binary, open and closed ternary complexes of the pre-catalytic state as well as the translocated post-catalytic product (20). Recognizing that pTNA is a close structural analog of TNA (Figure 1a), we wondered if Kod-RI might also facilitate pTNA synthesis by recognizing chemically prepared ptNDP nucleotides (17,37), the diphosphate monophosphonate building blocks of pTNA, in a template-dependent primer extension reaction (Figure 1a). Our findings indicate that Kod-RI is capable of synthesizing trace amounts of full-length pTNA after 24 hours of incubation at 55°C (Figure 1b). Although inadequate for most synthetic biology applications (38), previous studies have shown that this level of activity can be improved by directed evolution using polymerase engineering techniques that have been established to evolve DNA polymerases for the ability to transcribe and reverse transcribe other types of XNA polymers (3).

Crystal structure capturing Kod-RI incorporating pTNA

Kod-RI is a large (90 kDa), highly dynamic engineered polymerase with a disc-shaped architecture that encompasses the N-terminal, exonuclease, and catalytic domains of the protein.

Similar to other DNA polymerases, the catalytic domain resembles a right hand consisting of the thumb, fingers, and palm subdomains (Figure S1a). To allow direct comparison to our previously solved X-ray crystal structure of the post-catalytic complex of Kod-RI bound to a DNA duplex that had been extended by two TNA adenosine residues (tA) (20), crystals of Kod-RI capturing the extended product of pTNA synthesis were grown from primer-extension reactions that were performed in buffered solutions (Figure 2a). A parallel set of reactions performed with 2'-deoxyadenosine triphosphate (dATP) allowed us to obtain the DNA extended product as a control for understanding how Kod-RI recognizes the natural all-DNA duplex. Crystals obtained from both experiments resolved to resolutions of 1.98 Å and 2.71 Å for pTNA and DNA, respectively, and the structures were solved by molecular replacement (Table S3) using coordinates from our previous Kod-RI/TNA structure (PDB ID: 5VU9). Both of the new Kod-RI structures are highly homologous to the original Kod-RI/TNA polymerase structure with calculated Ca RMSD values of 0.567 Å and 0.578 Å for the Kod-RI/pTNA and Kod-RI/DNA complexes, respectively. $2F_o - F_c$ polder maps contoured at 7.5 σ and 7.0 σ for pTNA and DNA, respectively, unambiguously confirm the presence of the ptA residues at positions 12 and 13 of the primer strand as well as the equivalent dA nucleotides on the control all-DNA structure (Figure 2b-c).

The nucleotide addition products reside in the enzyme active site of Kod-RI as canonical Watson-Crick base pairs with their complementary thymine bases on the DNA template. Close inspection of the newly added residues failed to identify any differences in the hydrogen bonding pattern of the Watson-Crick base pair even though pTNA does not form a stable duplex with DNA in solution. In fact, heavy atom distances (2.7 - 3.0 Å) observed for the ptA-dT base pairs at positions 12 and 13 of the primer strand are nearly identical to the hydrogen bonding distances

observed for the tA-dT and dA-dT base pairs in the equivalent Kod-RI structures produced by extending the primer with either tATP or dATP, respectively (Figure 2d,e, Figure S2). No signs of duplex destabilization were observed in the pTNA structure, indicating that Kod-RI likely forces pTNA-DNA hybridization by acting as a large molecular clamp to hold the complementary pTNA and DNA strands together.

To better understand the molecular interactions responsible for pTNA synthesis, we examined the amino acid residues in the polymerase active site that directly contact the pTNA nucleotides. In the Kod-RI structure of the pTNA extended primer, we observe a novel alternative conformation for R606 (labeled R606^B) that is not present in the Kod-RI structures solved for the DNA or TNA extended primers (Figures 2 and S2a). Located on b26 of the thumb subdomain (Figure S1b), the positively charged guanidino group of R606^B forms a strong ion pair interaction with the negatively charged phosphonate group connecting the newly added ptA12 and ptA13 nucleotides (Figure 2d). In the Kod-RI structure of the DNA extended primer, the analogous phosphate group forms a hydrogen bond to Y594 on b25 of the thumb subdomain (Figure 2e). Interestingly, neither of these interactions are observed in the previously solved structure of Kod-RI that was extended with TNA (Figure S2), which could be due to the fact that TNA has a backbone repeat unit that is one atom shorter than that of pTNA or natural DNA.

Other structural differences observed in the polymerase active site include a bridging water interaction between Y594 and the phosphate group linking ptA12 and ptA13 of the pTNA extended structure (Figure S3). This interaction suggests that Y594 could play an important role in coordinating the phosphodiester backbone during oligonucleotide synthesis, since it forms a direct contact to the DNA extended structure but only an indirect contact in the pTNA extended structure. Another novel interaction observed in the pTNA extended structure is a weak

hydrogen bond between K592 and the O4' atom on the sugar moiety of ptA13 (Figure 2d). Additionally, we noticed that the catalytic aspartate D542 was located further away from the terminal pTNA residue than its position in the all-DNA extended structure of Kod-RI, suggesting that the incoming ptNDP substrate may adopt a suboptimal geometry in the enzyme active site. The same observation was also made in the previously solved structure of Kod-RI containing a TNA extended primer (20). Collectively, these structural differences likely contribute to the reduced catalytic activity of pTNA synthesis by the Kod-RI TNA polymerase.

Alignment of the three crystal structures of Kod-RI enabled a direct comparison of the backbone conformation of the pTNA, DNA, and TNA extended products. TNA exhibits a 5.5 Å P-P distance between adjacent phosphate groups (Figure 3), which is consistent with TNA having a five-atom backbone repeat unit rather than the more common six-atom repeat unit found in pTNA and DNA (39). By comparison, the analogous P-P distances in pTNA and DNA are 6.1 Å and 6.3 Å, respectively, which is consistent with a standard DNA duplex (40). Interestingly, pTNA adopts a backbone conformation that is strikingly similar to that of TNA, as evidenced from the overlays comparing pTNA to DNA and pTNA to TNA (Figure 3, S4). In the pTNA—TNA overlay, the backbone residues occupy positions that are closer to the central axis of the primer-template duplex than the equivalent DNA extended product. Additionally, the phosphate group linking adenosine nucleotides 12 and 13 resides at the exact position for both the pTNA and TNA containing structures. By comparison, all of the atoms in the DNA backbone are positioned further away from the central axis with the phosphate group linking dA₁₂ and dA₁₃ shifted by ~0.5 Å relative to its position in pTNA and TNA. The local base-pair and base-pair step parameters are reported in Supplementary Tables S4-S5.

We evaluated the torsion angles and pseudorotation parameters of the newly added adenosine residues in each of the three structures of Kod-RI (Figure 4, Figure S5, and Table S6). The sugar conformation of TNA has been reported to be in the range of *C3'-endo* to *C4'-exo* and is closer to *C3'-endo* in the active site of Kod-RI ($P_{tA12} = 13^\circ$ and $P_{tA13} = 25^\circ$) (41). By contrast, the threose ring of pTNA adopts a sugar conformation that is *C3'-endo* ($P_{ptA13} = 37^\circ$) and *C4'-exo* ($P_{ptA12} = 47^\circ$) in the enzyme active site of Kod-RI. The extent of puckering versus planarity is described by v_{max} , which does not vary appreciably between the threose sugars of pTNA and TNA. Interestingly, the pTNA extended structure supports a computationally predicted model of a single 2'-deoxythreose phosphonate nucleotide which produced a sugar pucker in the range of the *C3'-endo* to *C4'-exo* conformations as investigated in the active site of HIV-1 reverse transcriptase (42). Similar analysis of the DNA extended structure reveals that the 2'-deoxyadenosine sugar at position 12 of the primer strand adopts a favorable *C2'-endo* conformation ($P_{dA12} = 181^\circ$) which transitions to a *C4'-exo* ($P_{dA13} = 60^\circ$) conformation at position 13 of the primer strand. This result is consistent with previous structure sugar conformations observed previously in crystal structures of DNA polymerases (43).

Analysis of the c and d torsion angle covariance (Figure S6) indicates that the second DNA incorporation shifts to a lower c value and exhibits a considerable 54° drop in the d angle, moving towards an A-form helical geometry, while staying within the boundary of B-form structures. In contrast, the first pTNA and TNA nucleotides incorporated onto the 3' end of the DNA primer ($A_{12}; n+1$) exhibit slightly higher d values relative to the canonical B-form geometry of DNA and observe only a slight decrease in c while maintaining their d values for the second incorporation ($A_{13}; n+2$). In other reported polymerase structures, a similar conformational change for DNA nucleotides adopting a 'northern' A-form conformation has been

reported with a higher crystallographic resolution for DNA nucleotides in Bst DNA polymerase (A-family, *Bacillus stearothermophilus*) incorporating DNA (43). Likewise, a structure of wildtype Kod bound to a DNA duplex without any incorporations shows the terminal primer nucleotide adopting a similar 'northern' A-form conformation (44). The observed changes in the sugar pucker of the DNA adducts may be necessary for positioning the 3'-terminal hydroxyl group for the subsequent nucleotide incorporation steps in the DNA polymerase cycle.

Kod-RSGA mutations increase pTNA synthesis

To expand our findings with Kod-RI, we evaluated the ability of Kod-RSGA to direct the synthesis of pTNA on DNA templates. Kod-RSGA is a more efficient TNA polymerase that was recently discovered by programmed allelic mutagenesis (21). Similar to Kod-RI, Kod-RSGA contains the exonuclease silencing mutations (D141A and E143A) and shares the A485R gain-of-function mutation necessary for TNA synthesis. Additional mutations include N491S located in the finger subdomain, as well as the mutations R606G and T723A, both of which are located in the thumb subdomain (Figure 5a,b). Since Kod-RSGA has been reported to exhibit a 20-fold improvement in the rate of TNA synthesis relative to Kod-RI (21), we investigated the impact of these additional mutations on the rate of pTNA synthesis. Primer extension assays comparing Kod-RI and Kod-RSGA as polymerases for pTNA synthesis on a DNA template reveal that Kod-RSGA functions with higher overall activity than Kod-RI (Figure 5c). However, despite the enhanced activity of Kod-RSGA as a novel pTNA polymerase, this enzyme is still unable to generate full-length product. This is not an unexpected outcome, given that Kod-RSGA was originally evolved for TNA synthesis activity (21).

To better understand the activity of Kod-RSGA as a pTNA polymerase, we solved the X-ray crystal structure of Kod-RSGA in the post-catalytic state following two cycles of pTNA addition to the 3' end of the DNA primer. The structure resolved to a resolution of 3.01 Å (Table S3), enabling a detailed comparison with the equivalent structure of Kod-RI TNA polymerase. As expected, the Kod-RI/pTNA and Kod-RSGA/pTNA structures are highly homologous with a calculated Ca RMSD value of 0.561. The conformation of the pTNA nucleotides in Kod-RSGA strongly resembles the conformation of pTNA in Kod-RI in terms of torsion angles of the backbone and sugar pucker (Table S6, Figures S7).

The biggest difference between the Kod-RI and Kod-RSGA structures extended with pTNA is the structural rearrangement observed as a result of the R606G mutation found in Kod-RSGA. As discussed previously, the R606 residue observed in Kod-RI forms a novel alternative conformation that leads to a strong ion pair interaction with the negatively charged phosphonate group connecting the newly added ptA12 and ptA13 nucleotides (Figure 2d). In Kod-RSGA, this position is mutated to a glycine residue, which allows Y594 to form a strong hydrogen bond to the same phosphate group previously contacted by the R606. This contact is predicted to be important for synthesis, as it is also observed in the structure of wild type Kod DNA polymerase bound to an all-DNA substrate (44).

1.5 DISCUSSION

Understanding the structural determinants that allow engineered polymerases to transcribe and reverse transcribe synthetic genetic polymers with orthogonal base pairing properties is an important step towards the creation of a genetic firewall that can separate the synthetic biology information produced by humanity from the natural biological information of the cell (15). Such endeavors will almost certainly be aided by the development of more efficient XNA polymerases, as the length constraints imposed by synthetic biology vastly exceed what is currently possible by solid-phase XNA synthesis. At a minimum, XNA polymerases will be needed to copy synthetic biology information from DNA into XNA, replicate XNA independent of DNA, and reverse transcribe XNA back into DNA. Although we are still years away from replicating XNA polymers in a model cellular organism, the process of generating the enzymes required to achieve this goal will provide important new insights into the molecular recognition of XNA substrates by laboratory evolved XNA polymerases.

Our study represents an important first step in the direction of understanding how engineered polymerases are able to copy genetic information from DNA into XNA when the XNA system being synthesized does not cross-pair with DNA. To our knowledge, all previous crystal structures of XNA polymerases describe enzymes that recognize XNA substrates that are known to exchange genetic information with DNA through the formation of stable antiparallel Watson-Crick duplex structures that can occur free in solution (20,45-47). However, pTNA, like GNA before it, are examples of orthogonal genetic polymers that do not base pair with DNA, but for reasons that are unclear, are still able to transfer information from DNA into XNA, or vice versa, using engineered polymerases that can synthesize the complementary strand (17,18). How an engineered polymerase is able to recognize an artificial genetic polymer that does not form a

stable duplex with the templating strand is an interesting question that could offer new insights into the mechanism of polymerase-mediated oligonucleotide synthesis. The current work focuses on the crystal structure of the post-translocated complex in which a DNA duplex has been extended by two pTNA residues. This intermediate in the pTNA synthesis pathway was viewed as a logical starting point, as similar complexes have previously yielded high resolution structures (20).

Our structures reveal that pTNA bears a remarkable likeness to TNA when bound as a substrate to the active site of an engineered polymerase. TNA, a chemical analog of pTNA, has been extensively studied as a substrate for polymerase recognition, both as a template and as a nucleoside triphosphate (19,21,48-50). Relative to TNA, which has a five-atom backbone repeat unit consisting of threofuranosyl nucleic acids connected by 3'→2' phosphodiester linkages, pTNA comprises a six-atom backbone repeat unit in which a methyl group has been inserted between the 3' hydroxyl group and the phosphorous atom (Figure 1). This chemical change transitions TNA from a synthetic genetic polymer that can base pair with DNA and RNA to one that cannot, implying that pTNA should be structurally different than TNA. However, despite this prediction, our crystal structures reveal that both XNA systems adopt similar sugar puckers with conformations in the *C3'-endo* and *O4'-endo* range and give rise to nucleic acid backbones with their phosphate groups located at identical positions. One possible interpretation of this observation is that the polymerase is forcing the substrates to adopt a specific conformation that is required for XNA synthesis and the threose sugars are adjusting to meet these constraints. Since it is difficult to answer this question with the current data, an important future goal will be to determine the structures of pTNA and TNA homoduplexes in the absence of a bound polymerase. A related future goal will be to solve crystal structures of the polymerase bound to

longer pTNA extension products to see at which point the pTNA-DNA strands begin to dissociate from one another.

Although the backbone conformations of pTNA and TNA are structurally similar, they are distinct enough to form different interactions with the polymerase. These differences may be required in order to maintain Watson-Crick base pairing with the DNA template. The conformational transition observed for the 3' terminal deoxyribose sugar in the control DNA structure is likely a paradigmatic example of how natural wild-type polymerases function as compared to the more rigid conformations observed for the threose sugars in the pTNA and TNA bound structures. Other notable differences include contacts to pTNA when the substrate is bound to Kod-RI versus Kod-RSGA. For example, the Kod-RSGA structure restores the hydrogen bonding interaction between Y594 and the phosphate group linking ptA12 to ptA13, which is also present in the DNA control structure and thought to be important for DNA synthesis (Figure 6). We postulate that this interaction may also be important for pTNA synthesis, since Kod-RSGA is a more efficient pTNA polymerase than Kod-RI.

Finally, it is interesting to note that the polymerase mutations previously known to be beneficial for TNA synthesis also appear to benefit pTNA synthesis. The growing body of information coming from polymerase engineering studies is helping to identify the structural determinants of XNA synthesis. As these studies continue, it will be interesting to see how known hot spots, like V93, L408, A485, N491, I521, and E664, correlate to improved activity when additional gain-of-function mutations are discovered elsewhere in the polymerase. Are these mutations merely reducing the stringency of substrate recognition or are they driving the polymerase toward a new chemical function? As more structures are solved for XNA polymerases, and additional XNA polymerases are better characterized biochemically, it will be

interesting to assess how easy or difficult it is for a natural DNA polymerase to erase its memory for DNA in favor of a synthetic genetic material that may bear little chemical resemblance to the genetic polymers of life. Indeed, pTNA with modifications in both the sugar and phosphate backbone could be an important step in this new direction.

1.6 CONCLUSION

We describe the first crystal structures of an orthogonal genetic polymer bound to an engineered polymerase. Our results indicate that engineered polymerases recognize pTNA differently than TNA even though pTNA appears to be structurally similar to TNA when bound to the enzyme active site. These findings highlight the challenges of developing new XNA polymerases by directed evolution, as even structurally related substrates require unique intermolecular interactions in order to allow for efficient XNA synthesis on DNA templates.

1.7 REFERENCES

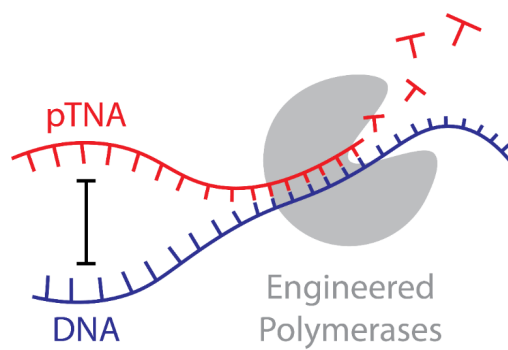
1. Chaput, J.C. and Herdewijn, P. (2019) What Is XNA? *Angew. Chem. Int. Ed. Engl.*, **58**, 11570-11572.
2. Joyce, G.F. (2012) Toward an alternative biology. *Science*, **336**, 307-308.
3. Nikoomanzar, A., Chim, N., Yik, E.J. and Chaput, J.C. (2020) Engineering polymerases for applications in synthetic biology. *Q. Rev. Biophys.*, **53**, e8.
4. Yu, H., Zhang, S. and Chaput, J.C. (2012) Darwinian evolution of an alternative genetic system provides support for TNA as an RNA progenitor. *Nat. Chem.*, **4**, 183-187.
5. Pinheiro, V.B., Taylor, A.I., Cozens, C., Abramov, M., Renders, M., Zhang, S., Chaput, J.C., Wengel, J., Peak-Chew, S.Y., McLaughlin, S.H. *et al.* (2012) Synthetic genetic polymers capable of heredity and evolution. *Science*, **336**, 341-344.
6. Rangel, A.E., Chen, Z., Ayele, T.M. and Heemstra, J.M. (2018) In vitro selection of an XNA aptamer capable of small-molecule recognition. *Nucleic Acids Res.*, **46**, 8057-8068.
7. Dunn, M.R., McCloskey, C.M., Buckley, P., Rhea, K. and Chaput, J.C. (2020) Generating biologically stable TNA aptamers that function with high affinity and thermal stability. *J. Am. Chem. Soc.*, **142**, 7721-7724.
8. Ereemeeva, E., Fikatas, A., Margamuljana, L., Abramov, M., Schols, D., Groaz, E. and Herdewijn, P. (2019) Highly stable hexitol based XNA aptamers targeting the vascular endothelial growth factor. *Nucleic Acids Res.*, **47**, 4927-4939.
9. McCloskey, C.M., Li, Q., Yik, E.J., Chim, N., Ngor, A.K., Medina, E., Grubisic, I., Co Ting Keh, L., Poplin, R. and Chaput, J.C. (2021) Evolution of Functionally Enhanced alpha-l-Threofuranosyl Nucleic Acid Aptamers. *ACS Synth. Biol.*, **10**, 3190-3199.
10. Taylor, A.I., Pinheiro, V.B., Smola, M.J., Morgunov, A.S., Peak-Chew, S., Cozens, C., Weeks, K.M., Herdewijn, P. and Holliger, P. (2015) Catalysts from synthetic genetic polymers. *Nature*, **518**, 427-430.
11. Wang, Y., Ngor, A.K., Nikoomanzar, A. and Chaput, J.C. (2018) Evolution of a general RNA-cleaving FANA enzyme. *Nat. Commun.*, **9**, 5067.
12. Wang, Y., Wang, Y., Song, D., Sun, X., Zhang, Z., Li, X., Li, Z. and Yu, H. (2021) A Threose Nucleic Acid Enzyme with RNA Ligase Activity. *J. Am. Chem. Soc.*, **143**, 8154-8163.
13. Wang, Y., Wang, Y., Song, D., Sun, X., Li, Z., Chen, J.Y. and Yu, H. (2022) An RNA-cleaving threose nucleic acid enzyme capable of single point mutation discrimination. *Nat. Chem.*, **14**, 350-359.

14. Herdewijn, P. and Marliere, P. (2009) Toward safe genetically modified organisms through the chemical diversification of nucleic acids. *Chem. Biodivers.*, **6**, 791-808.
15. Chaput, J.C., Herdewijn, P. and Hollenstein, M. (2020) Orthogonal Genetic Systems. *Chembiochem*, **21**, 1408-1411.
16. Schmidt, M. (2010) Xenobiology: a new form of life as the ultimate biosafety tool. *BioEssays*, **32**, 322-331.
17. Liu, C., Cozens, C., Jaziri, F., Rozenski, J., Marechal, A., Dumbre, S., Pezo, V., Marliere, P., Pinheiro, V.B., Groaz, E. *et al.* (2018) Phosphonomethyl Oligonucleotides as Backbone-Modified Artificial Genetic Polymers. *J. Am. Chem. Soc.*, **140**, 6690-6699.
18. Tsai, C.-H., Chen, J. and Szostak, J.W. (2007) Enzymatic synthesis of DNA on glycerol nucleic acid templates without stable duplex formation between the product and template. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 14598-14603.
19. Larsen, A.C., Dunn, M.R., Hatch, A., Sau, S.P., Youngbull, C. and Chaput, J.C. (2016) A general strategy for expanding polymerase function by droplet microfluidics. *Nat. Commun.*, **7**, 11235.
20. Chim, N., Shi, C., Sau, S.P., Nikoomanzar, A. and Chaput, J.C. (2017) Structural basis for TNA synthesis by an engineered TNA polymerase. *Nat. Commun.*, **8**, 1810.
21. Nikoomanzar, A., Vallejo, D., Yik, E.J. and Chaput, J.C. (2020) Programmed allelic mutagenesis of a DNA polymerase with single amino acid resolution. *ACS Synth. Biol.*, **9**, 1873-1881.
22. Sau, S.P., Fahmi, N.E., Liao, J.-Y., Bala, S. and Chaput, J.C. (2016) A scalable synthesis of α -L-threose nucleic acid monomers. *J. Org. Chem.*, **81**, 2302-2307.
23. Liao, J.-Y., Bala, S., Ngor, A.K., Yik, E.J. and Chaput, J.C. (2019) P(V) Reagents for the Scalable Synthesis of Natural and Modified Nucleoside Triphosphates. *J. Am. Chem. Soc.*, **141**, 13286-13289.
24. Kabsch, W. (2010) Xds. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 125-132.
25. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 72-82.
26. Evans, P.R. and Murshudov, G.N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.*, **69**, 1204-1214.
27. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658-674.

28. Moriarty, N.W., Grosse-Kunstleve, R.W. and Adams, P.D. (2009) electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D Biol. Crystallogr.*, **65**, 1074-1080.
29. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486-501.
30. Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H. and Adams, P.D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.*, **68**, 352-367.
31. Liebschner, D., Afonine, P.V., Baker, M.L., Bunkoczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.W., Jain, S., McCoy, A.J. *et al.* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol*, **75**, 861-877.
32. Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. and Richardson, D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12-21.
33. Liebschner, D., Afonine, P.V., Moriarty, N.W., Poon, B.K., Sobolev, O.V., Terwilliger, T.C. and Adams, P.D. (2017) Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr D Struct Biol*, **73**, 148-157.
34. Battye, T.G., Kontogiannis, L., Johnson, O., Powell, H.R. and Leslie, A.G. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 271-281.
35. Sun, G., Voigt, J.H., Filippov, I.V., Marquez, V.E. and Nicklaus, M.C. (2004) PROSIT: pseudo-rotational online service and interactive tool, applied to a conformational survey of nucleosides and nucleotides. *J Chem Inf Comput Sci*, **44**, 1752-1762.
36. Li, S., Olson, W.K. and Lu, X.J. (2019) Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res.*, **47**, W26-W34.
37. Luo, M., Wu, S., Kalkeri, R., Ptak, R.G., Zhou, T., Van Mellaert, L., Wang, C., Dumbre, S.G., Block, T., Groaz, E. *et al.* (2020) Scalable Synthesis, In Vitro cccDNA Reduction, and In Vivo Antihepatitis B Virus Activity of a Phosphonomethoxydeoxythreosyl Adenine Prodrug. *J Med Chem*, **63**, 13851-13860.
38. Chaput, J.C. (2021) Redesigning the Genetic Polymers of Life. *Acc. Chem. Res.*, **54**, 1056-1065.
39. Schöning, K.U., Scholz, P., Guntha, S., Wu, X., Krishnamurthy, R. and Eschenmoser, A. (2000) Chemical etiology of nucleic acid structure: the a-threofuranosyl-(3'-->2') oligonucleotide system. *Science*, **290**, 1347-1351.

40. Pallan, P.S., Wilds, C.J., Wawrzak, Z., Krishnamurthy, R., Eschenmoser, A. and Egli, M. (2003) Why does TNA cross-pair more strongly with RNA than with DNA? An answer from X-ray analysis. *Angew. Chem. Int. Ed.*, **42**, 5893-5895.
41. Wilds, C.J., Wawrzak, Z., Krishnamurthy, R., Eschenmoser, A. and Egli, M. (2002) Crystal structure of a B-form DNA duplex containing (L)-a-threofuranosyl (3'-2') nucleosides: a four-carbon sugar is easily accommodated into the backbone of DNA. *J. Am. Chem. Soc.*, **124**, 13716-13721.
42. Wu, T., Froeyen, M., Kempeneers, V., Pannecouque, C., Wang, J., Busson, R., De Clercq, E. and Herdewijn, P. (2005) Deoxythreosyl phosphonate nucleosides as selective anti-HIV agents. *J. Am. Chem. Soc.*, **127**, 5056-5065.
43. Kiefer, J.R., Mao, C., Braman, J.C. and Beese, L.S. (1998) Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. *Nature*, **391**, 304-307.
44. Bergen, K., Betz, K., Welte, W., Diederichs, K. and Marx, A. (2013) Structures of KOD and 9°N DNA polymerases complexed with primer template duplex. *ChemBioChem*, **14**, 1058-1062.
45. Jackson, L.N., Chim, N., Shi, C. and Chaput, J.C. (2019) Crystal structures of a natural DNA polymerase that functions as an XNA reverse transcriptase. *Nucleic Acids Res.*, **47**, 6973-6983.
46. Samson, C., Legrand, P., Tekpinar, M., Rozenski, J., Abramov, M., Holliger, P., Pinheiro, V.B., Herdewijn, P. and Delarue, M. (2020) Structural Studies of HNA Substrate Specificity in Mutants of an Archaeal DNA Polymerase Obtained by Directed Evolution. *Biomolecules*, **10**.
47. Li, Q., Maola, V.A., Chim, N., Hussain, J., Lozoya-Colinas, A. and Chaput, J.C. (2021) Synthesis and Polymerase Recognition of Threose Nucleic Acid Triphosphates Equipped with Diverse Chemical Functionalities. *J. Am. Chem. Soc.*, **143**, 17761-17768.
48. Chaput, J.C., Ichida, J.K. and Szostak, J.W. (2003) DNA polymerase-mediated DNA synthesis on a TNA template. *J. Am. Chem. Soc.*, **125**, 856-857.
49. Chaput, J.C. and Szostak, J.W. (2003) TNA synthesis by DNA polymerases. *J. Am. Chem. Soc.*, **125**, 9274-9275.
50. Yu, H., Zhang, S., Dunn, M.R. and Chaput, J.C. (2013) An efficient and faithful in vitro replication system for threose nucleic acid. *J. Am. Chem. Soc.*, **135**, 3583-3591.

1.8 FIGURES



Graphical Abstract

Structural insights into the enzymatic synthesis of an orthogonal genetic polymer that does not form antiparallel duplex structures with DNA and RNA.

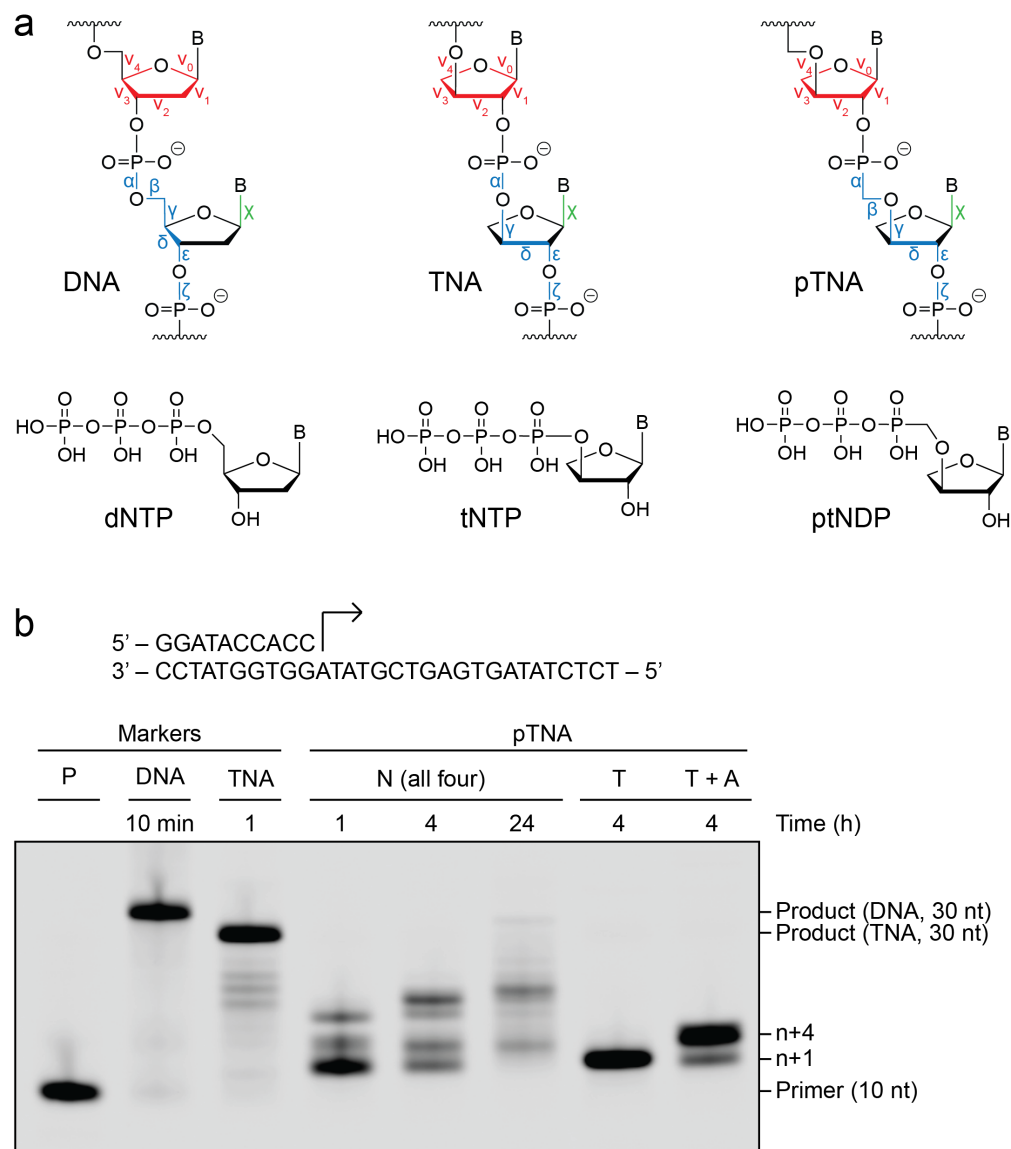


Figure 1.1 pTNA Synthesis by Kod-RI TNA Polymerase

(a) Molecular structures of the linearized backbones of DNA, TNA, and pTNA and their nucleotides. Backbone, sugar, and glycosidic torsion angles are colored blue, red, and green, respectively. (b) Kod-RI-mediated extension of a 5' IR800-labeled DNA primer annealed to a DNA template with ptNDP substrates. Full-length DNA and TNA products along with the starting primer are provided as markers. The reaction contained 0.5 μ M IR800-labeled primer–template duplex, 100 μ M of the specified nucleotides, and 0.5 μ M Kod-RI (pre-primed with 1 mM MnCl_2 for TNA and pTNA), in 1X ThermoPol® buffer for the specified time at 55°C.

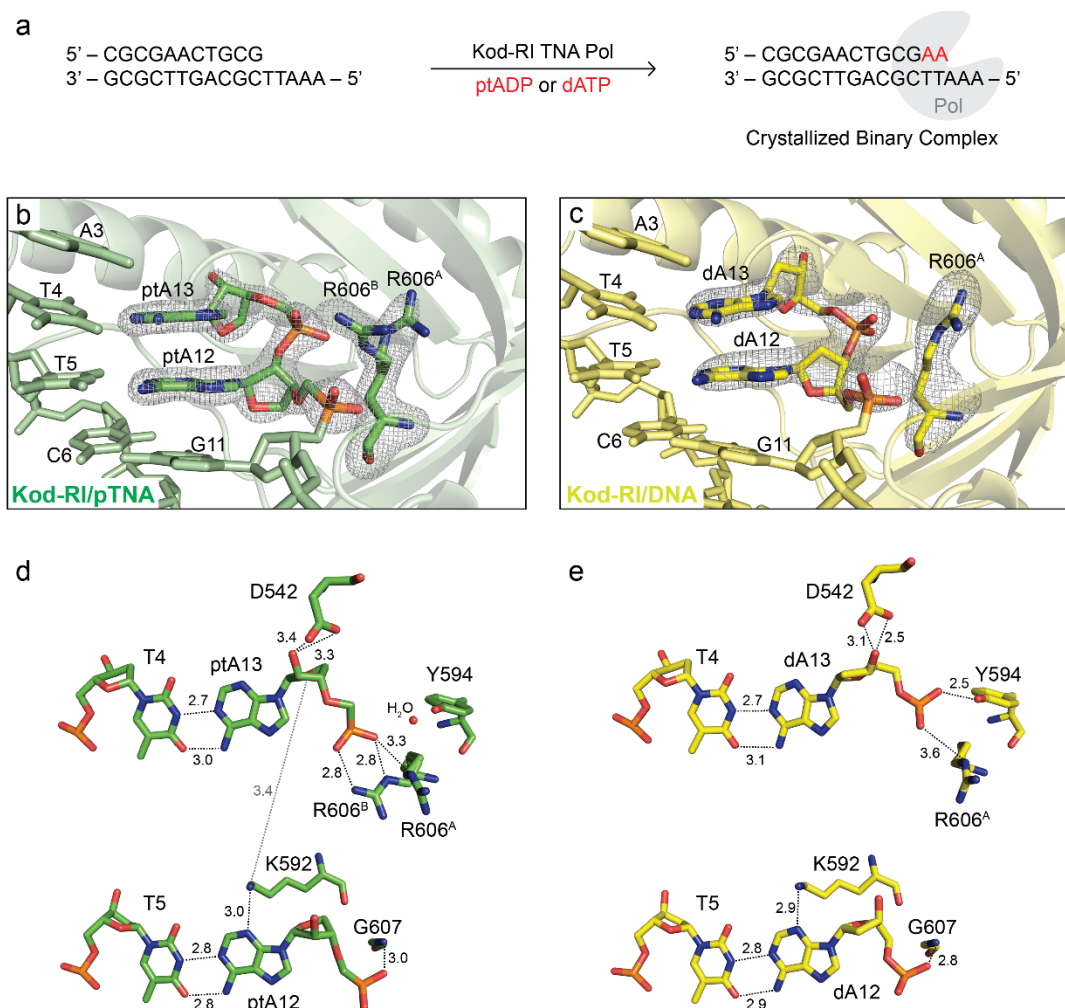


Figure 1.2 Active Site View of Kod-RI TNA Polymerase Showing pTNA Incorporation

(a) Schematic representation of the crystallized binary complex following the addition and translocation of two adenosine nucleotides. (b) View of the active site of Kod-RI incorporating pTNA with a $2F_o - F_c$ composite polder map contoured at 7.5σ for the incorporated ptA12 and ptA13, and another polder map contoured at 5.0σ for both conformations of arginine 606, labeled A and B. The exonuclease domain is hidden to avoid obstruction of the view. (c) Kod-RI incorporating DNA with polder maps contoured at 7σ for DNA (dA12 and dA13) and 4σ for the single conformation of arginine 606. (d, e) The intermolecular interactions observed in the active site with distances reported in angstroms. Color scheme: structure obtained for the pTNA extended primer (green) and DNA extended primer (yellow).

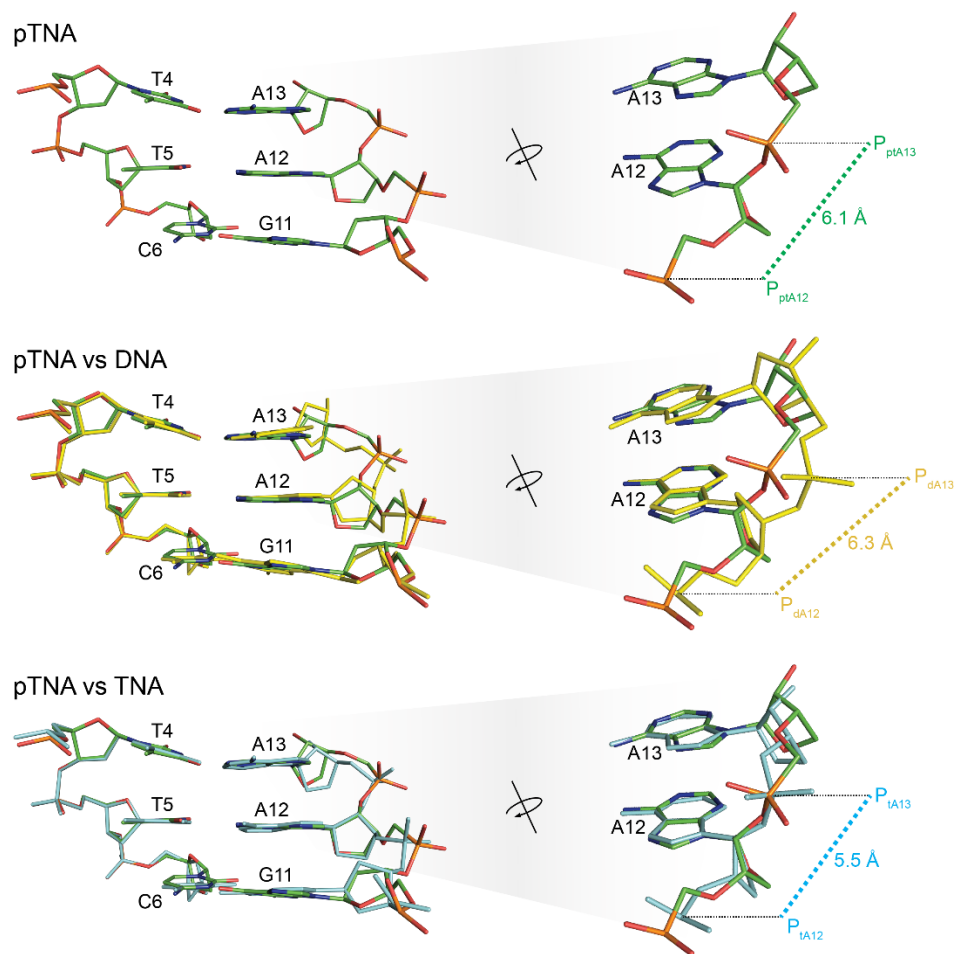


Figure 1.3 Active Site View of the DNA, TNA, and pTNA Adducts

Structural comparison of the backbone moiety of the DNA (yellow), TNA (cyan), and pTNA (green) residues following primer extension with Kod-RI. Images on the right provide an elevated and rotated view of the overlay shown on the left in the Watson-Crick base pairing format with the templating strand. The P-P distances in are provided in units of angstroms.

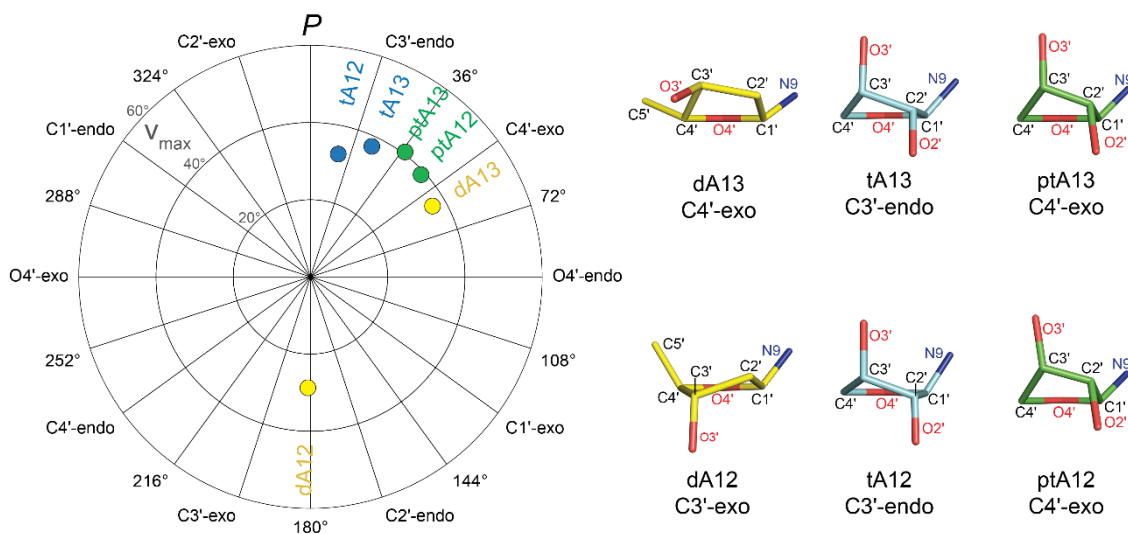


Figure 1.4 Pseudorotation Wheel and Sugar Puckering

A pseudorotation wheel plotting the pseudorotation phase angle P (degrees) and the maximum sugar torsion angle $v_{\max}$ (degrees) for the incorporated pTNA nucleotides (ptA12 and ptA13) and the analogous DNA (dA12 and dA13) and TNA nucleotides (tA12 and tA13). The puckered sugar moiety of each nucleotide is displayed in uniform orientation, where the C4'-O4'-C1' atoms are positioned in the back.

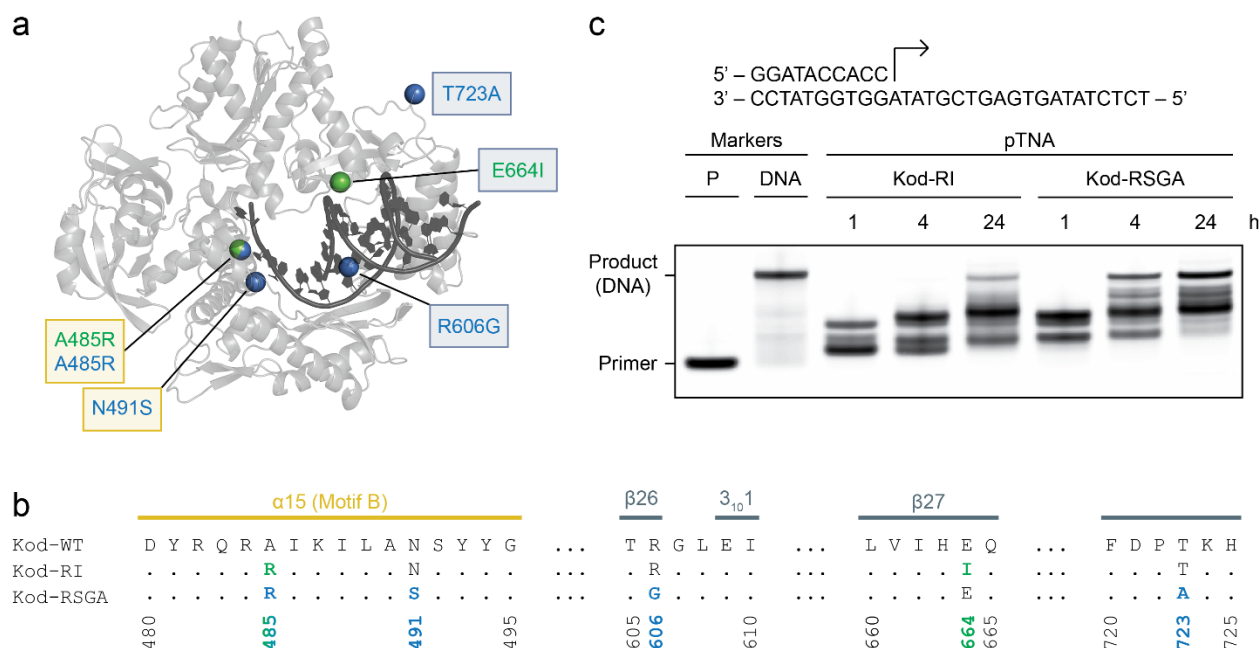


Figure 1.5 Engineered Mutations in Kod-RSGA

(a) The overall architecture of Kod-RSGA TNA polymerase (gray) bound to a DNA duplex (black) is shown with the positions of the amino acid mutations observed in Kod-RI and Kod-RSGA highlighted as green and blue spheres, respectively. Mutations boxed in yellow and gray occur in the finger and thumb subdomains, respectively. (b) A sequence alignment highlights the secondary structures and positions of the engineered mutations. Both mutations in the finger are part of a conserved sequence known as Motif B. (c) A time course demonstrates a faster synthesis rate for the Kod-RSGA. Both enzymes extended a 5' IR680-labeled DNA primer annealed to a DNA template with ptNDP substrates. A DNA primer (11 nt) and a full-length DNA marker (18 nt) are provided. The reaction contained 0.5 μ M IR800-labelled primer-template duplex, 100 μ M ptNDPs, and 0.5 μ M Kod-RI or Kod-RSGA (pre-primed with 1 mM MnCl_2), in 1X ThermoPol® buffer for the 15-minute intervals at 55°C.

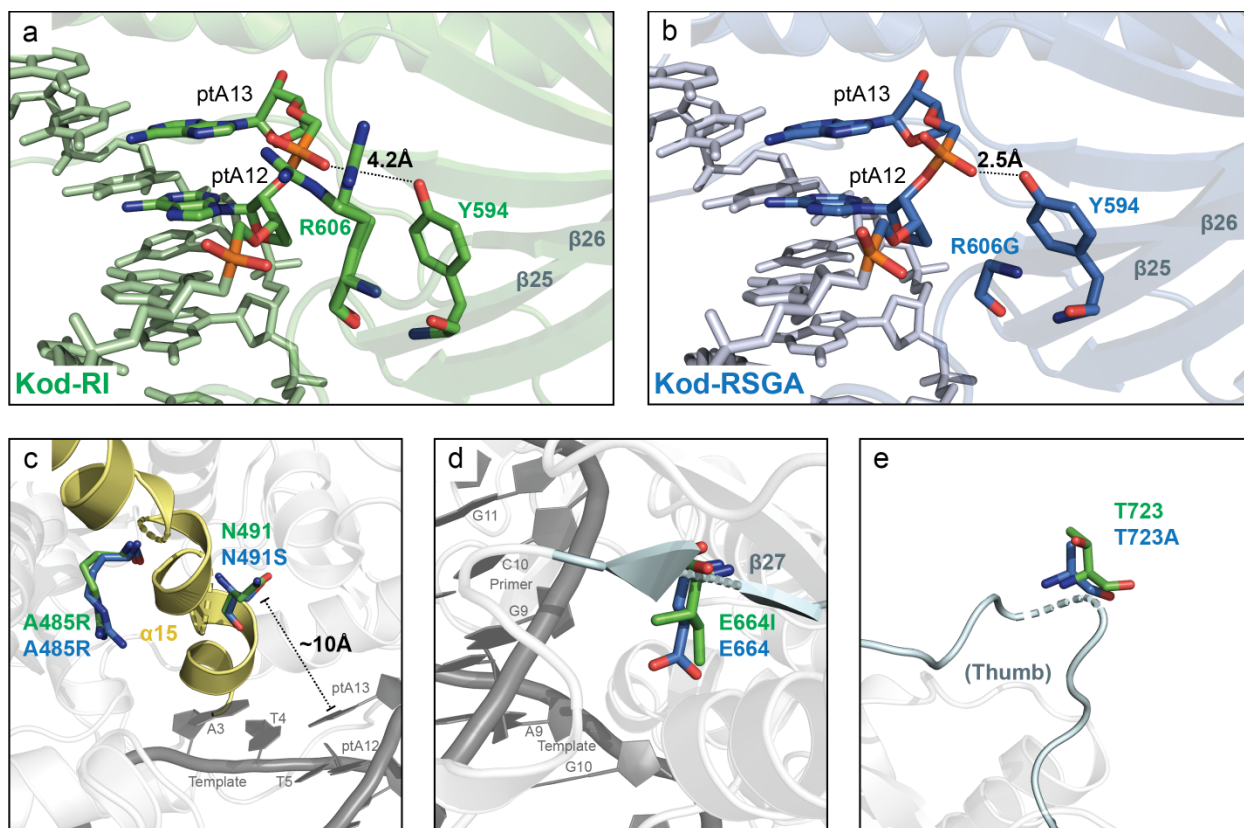


Figure 1.6 Structural Consequences of the Engineered Mutations

Panels (a) and (b) show the active sites of Kod-RI and Kod-RSGA, respectively, highlighting the interactions of the pTNA backbone with residues 606 and Y594. Panels (c), (d), and (e) show an overlay of the engineered mutations in Kod-RI (green) and Kod-RSGA (blue). The $\alpha 15$ helix of the fingers subdomain (yellow) and the $\beta 27$ sheet and the loop containing position 723 in the thumb subdomain (cyan) are highlighted.

CHAPTER 2: Directed Evolution of Highly Efficient TNA Polymerase

Achieved by Homologous Recombination

PUBLICATION INFORMATION

This chapter is adapted from the following publication: Maola, V.A., Yik, E.J., Hajjar, M., Lee, J.J., Holguin, M.J., Quijano, R.N., Nguyen, K.K., Ho, K.L., Medina, J.V., Chim, N. *et al.* (2024), *Nat Catal*, 7, <https://doi.org/10.1038/s41929-024-01233-1>

© The Author(s), under exclusive licence to Springer Nature Limited

2024

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. JC is the corresponding author.

Data Availability

The crystal structure of 10-92 was deposited in the Protein Data Bank (8T3X). Source Data and the Reporting Summary files are available through *Nature Catalysis*.

Supplementary Data

Supplementary data associated with the publication are available through *Nature Catalysis* and are reproduced in Appendix B.

Peer Review

Nature Catalysis thanks Masayasu Kuwahara, Vitor Pinheiro and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Funding

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

Competing Interests

J.C.C., V.A.M., E.J.Y. 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. The other authors declare no competing interests.

2.1 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.

2.2 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 medicine¹⁻³. 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 pathway⁶ and dynamic motion⁷ 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 triphosphates⁸⁻¹⁴. 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 polymerases⁸⁻¹⁰, 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 substrate²¹. As these studies were performed using polymerase libraries generated by saturation mutagenesis and deep mutational scanning⁸⁻¹⁰, 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 sequence²⁴. 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 sequencing²⁵⁻²⁷.

TNA was chosen as the target for polymerase activity to allow comparison with previous efforts⁸⁻¹⁰ 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. 1a)²⁸. Unlike other potentially natural derivatives of RNA²⁹, TNA is capable of forming stable antiparallel duplex structures with itself and with complementary strands of DNA and RNA²⁸. It's unusual 2',3'-linked phosphodiester backbone is recalcitrant to nucleolytic enzymes³⁰ and stable against acid-mediated degradation³¹, making it an attractive genetic system for therapeutic applications³². In vitro selection experiments have shown that TNA is capable of folding into shapes with ligand binding and catalytic activity³³⁻³⁸, and its recognition by Bst DNA polymerase, a TNA reverse transcriptase, allows digital information stored in TNA to be recovered and read by DNA sequencing³⁹. Finally, TNA has been evaluated as a building block for improving the biostability and activity of gene silencing

applications involving DNazymes⁴⁰, antisense oligonucleotides⁴¹, and RNA interference pathways⁴².

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.

2.3 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 space⁴³. 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 classes⁴⁴. 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 (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 orthologs⁴⁵. Additionally, the library contained 10 conserved regions (Fig. 1c) that were required for synthetic recombination using the polymerase chain reaction (PCR). Homologous recombination (Fig. 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 (Supplementary Fig. 2) to create a pool of >4 million polymerase variants⁴⁶. 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 libraries⁸⁻¹⁰, this population of sequences was expected to explore a larger range of sequence space due to the combinatorial nature of the assembly process⁴⁷.

Droplet-Based Polymerase Evolution

Iterative rounds of directed evolution were accomplished using the massively parallel screening technique of droplet-based optical polymerase sorting (DrOPS, 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. 1e), including chemically synthesized tNTPs⁴⁹. The droplet population ($\sim 10^7$) 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. 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 (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. 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. 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. 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. 2a,b). Comparative sequence analysis shows that 10-92 is a highly chimeric enzyme comprised of pieces from all four parent polymerases (Fig. 2c, 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. 3a,b, Supplementary Figs. 7-8), capable of extending a DNA primer to full-length product with 40 nucleotide additions occurring in 45 seconds (Fig. 3c). TNA synthesis activity appears to be slightly more efficient from a TNA primer than a DNA primer (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. 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. 3f, Supplementary Fig. 10), which are valuable substrates for functionally enhanced TNA aptamers, known as threomers³⁵. 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 (Supplementary Figs. 11-12). This is a notable improvement over previous generations, which suffered from G-G mismatches during TNA synthesis⁵⁰. 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. 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 primer⁵², PKPro provides a more holistic view of XNA synthesis by measuring the average rate of nucleotide incorporation across a template⁵¹. 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 duplexes⁵¹, 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. 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 (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 (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 (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 (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 library¹⁰. 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-2⁵⁴, and the full-length S1 protein from SARS-CoV-2³⁵. 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. 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. 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 (Supplementary Table 4)²¹.

The 10-92 TNA polymerase adopts a disk-shaped architecture (Fig. 6) that is consistent with all known archaeal B-family DNA polymerases⁵⁶. Superposition of the structure against the natural Kod DNA polymerase⁵⁷ reveals major conformational changes to the catalytic domain as well as smaller changes to the N-terminal (NTD) and exonuclease (Exo) domains (Fig. 6, 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 (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. 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. 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 α 12-helix to shift by 3Å relative to its position in Kod DNA polymerase (Fig 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 snugly fit tATP substrate in the active site pocket formed by the palm and finger subdomains (Fig. 7a, 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 (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 7b). Consistent with fast enzyme kinetics, the templating base (dT8) forms a favorable coplanar base pair with the tATP substrate (Fig 7c, Supplementary Table 5), thus correcting a suboptimal base pair geometry previously observed in the ancestral Kod-RI TNA polymerase structure²¹. 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 polymerase⁵⁷. 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 (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 (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} helix (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.

2.4 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 backbones¹⁻³. 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 polymerases⁵⁸. 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 incorporation⁶¹.

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 activity⁸⁻¹⁰, 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 triphosphate⁶². 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 recognition¹⁰. 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 activity²⁵⁻²⁷. 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 library⁶³. 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 polymerase⁵⁷, 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.

2.5 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 (Supplementary Table 1), were disassembled and reassembled by synthetic DNA shuffling⁴⁶ using a collection of primers (Supplementary Table 2). For extensive details, see 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 μ L of XL1-blue *E. coli*. Recovered cells were then used to inoculate 50 mL of LB-carbenicillin (50 μ g/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 μ L, 1:100 v/v) was used to inoculate 50 mL of LB-carbenicillin (50 μ g/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 (NH₄)₂SO₄, 10 mM KCl, 2 mM MgSO₄, 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 μ M of a self-priming hairpin template labeled with Cy3 at the 5' end (30merHP.V2, Supplementary Table 2), 2 μ M of a 3' end labeled Iowa Black quencher sequence (QP08.Iowa, Supplementary Table 2), and 100 μ M 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 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 MnCl₂ and MgCl₂ 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 MgCl₂, 0.013 μM Kod-WT (exo-) polymerase, 100 ng of 5-270 plasmid and 2 fold serial dilution of MnCl₂ 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 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 OD₆₀₀ ~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 methods⁶⁶. For extensive details, see Supplementary Methods.

Thermostability 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 synthesis⁶⁹. 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 μ M.

BLI validation of TNA aptamers

DNA templates (1.2 μ M) encoding known TNA aptamers were annealed to a biotinylated-PBS8 primer (1 μ M) in 1x ThermoPol at 90°C for 5 min and snap cooled, followed by the addition of 100 μ M of respective TNA base mixes and 2 μ M of 10-92 for a total volume of 750 μ L. Primer extension was performed at 55°C. The reaction was quenched by freezing the sample, vacuum concentrating to a volume of <100 μ L and adding 250 μ L 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 cycle⁹ and analyzing the Sanger sequences using CLC Main Workbench 23.0.5. For extensive details, see Supplementary Methods.

TNA versus DNA primed TNA synthesis

The DNA template (EM619, 1 μ M) was annealed to either a TNA or DNA PBS8 primer (PBS8, 0.5 μ M) in 1x ThermoPol at 95°C for 5 min and immediately snap cooled, followed by the addition of 10-92 polymerase (0.5 μ M). The reaction was initiated with the addition of 100 μ M of tNTPs for a total volume of 20 μ L. 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 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 (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 μ L) from a single master mix poised under single-turnover conditions with equimolar concentrations of polymerase (0.2 μ M) and pre-annealed P-T duplex (0.2 μ M 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 μ M 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 μ L (6 μ M) EvaGreen® dye, previously identified as the optimal intercalating dye to monitor synthesis of TNA on a DNA template⁵¹. Each reaction (25 μ L) 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 ability⁷⁰, 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 described⁵¹ by multiplying data with the conversion factor $F = [\text{Template}] * L / [\text{Polymerase}]$, 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 (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. For extensive details, see 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 MgCl₂, 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. The 10-92 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 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 Phaser⁷³ with PDB structure 5VU8 as the search model, and the final model was determined using iterative rounds of manual building through Coot⁷⁴ and refinement with phenix⁷⁵. The stereochemistry and geometry of all structures were validated with Molprobity⁷⁶ 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 PyMOL⁷⁷.

2.6 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 Doublet, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å 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 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).
- 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 polymerases 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 α-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).

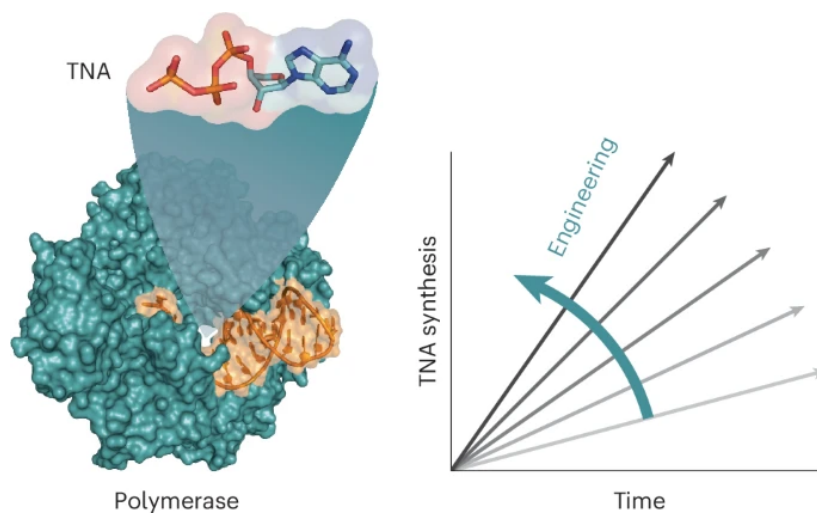
- 31 Lee, E. M., Setterholm, N. A., Hajjar, M., Barpuzary, B. & Chaput, J. C. Stability and mechanism of threose nucleic acid toward acid-mediated degradation. *Nucleic Acids Res.* **51**, 9542-9551 (2023).
- 32 Dunn, M. R., Jimenez, R. M. & Chaput, J. C. Analysis of aptamer discovery and technology. *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. *et al.* Shorter is better: the alpha-(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- α -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^oN-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).

2.7 FIGURES



TOC Figure

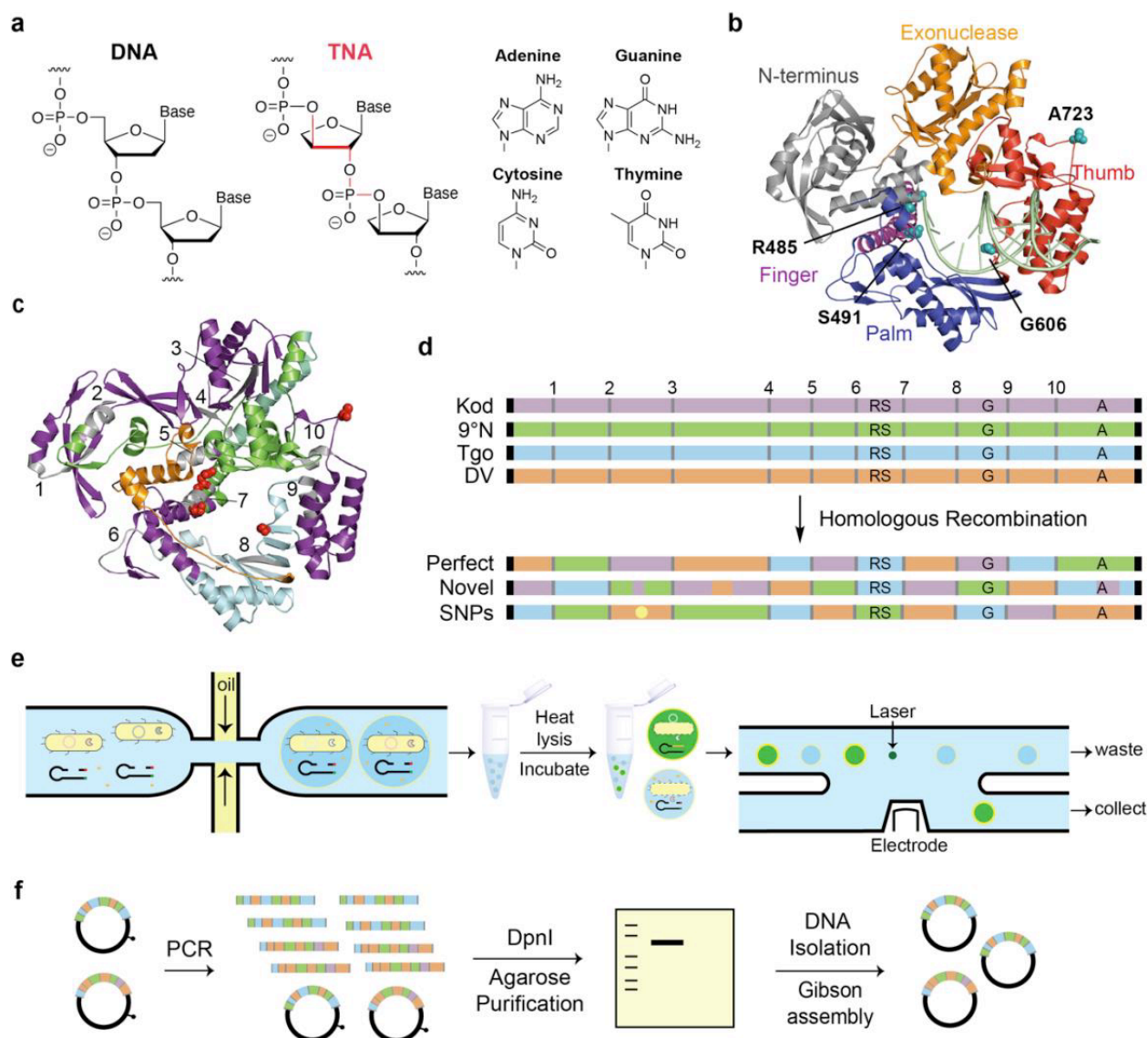


Figure 2.1 Homologous recombination as a path to engineered XNA polymerases

(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.

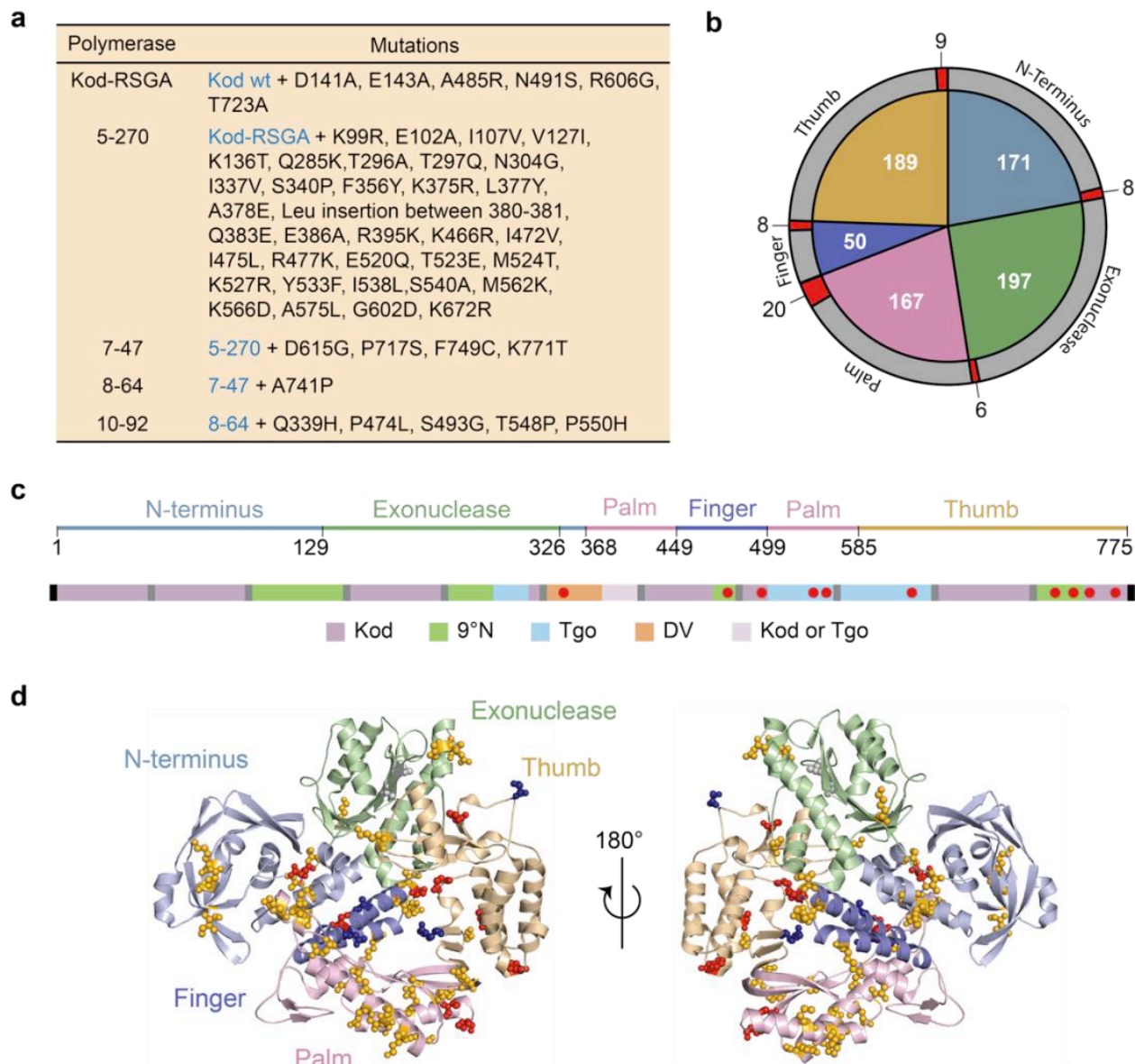


Figure 2.2 Directed evolution of a highly efficient TNA polymerase

(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).

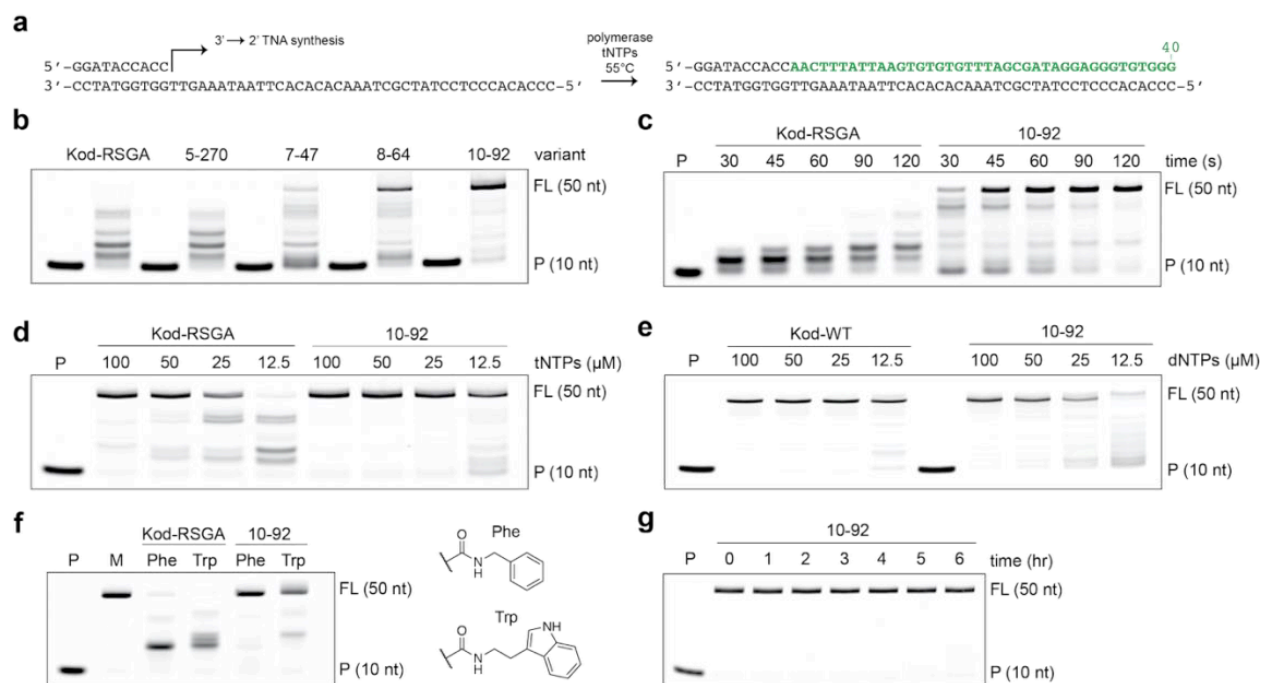


Figure 2.3 Functional analysis of TNA polymerases

(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 μM) with 1 μM polymerase and 100 μM tNTPs in 1X ThermoPol buffer at 55°C. C5-modified tNTP mixtures contain 2 μM 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.

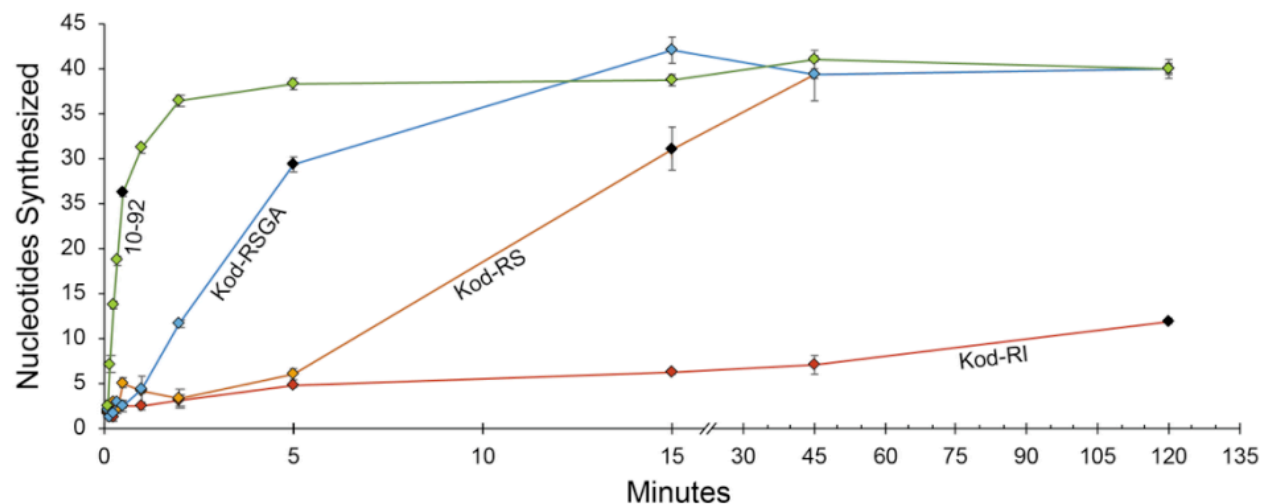


Figure 2.4 Kinetic analysis of TNA synthesis

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.

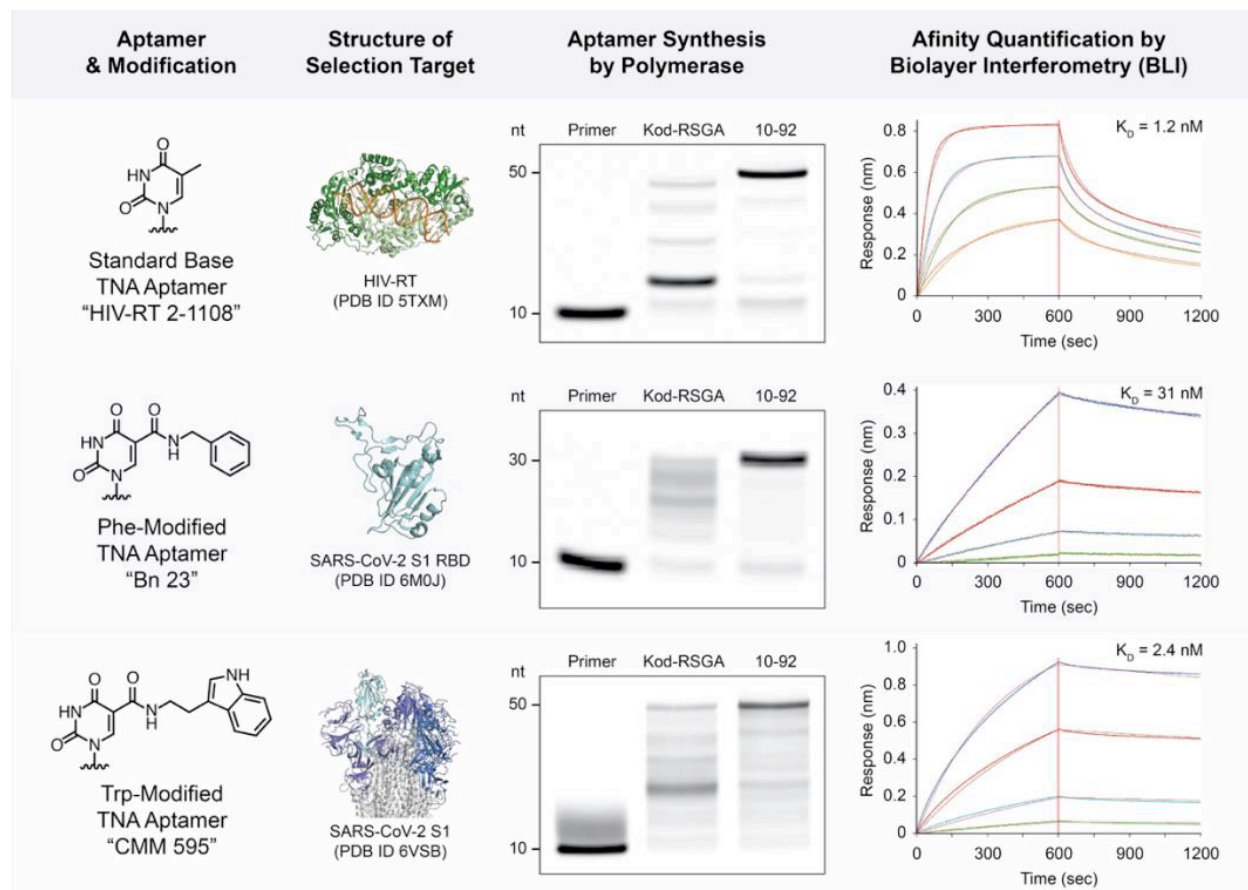


Figure 2.5 Enhanced functional activity

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.

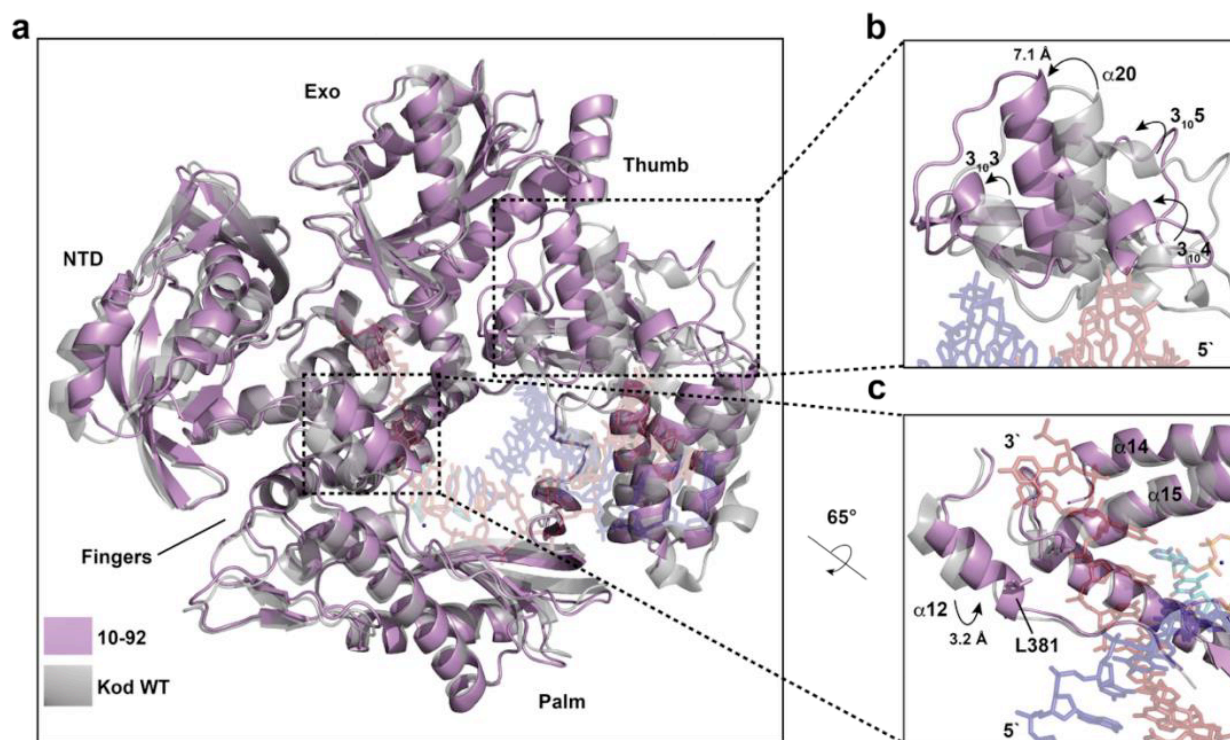


Figure 2.6 Structural topology of the 10-92 TNA polymerase

(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.

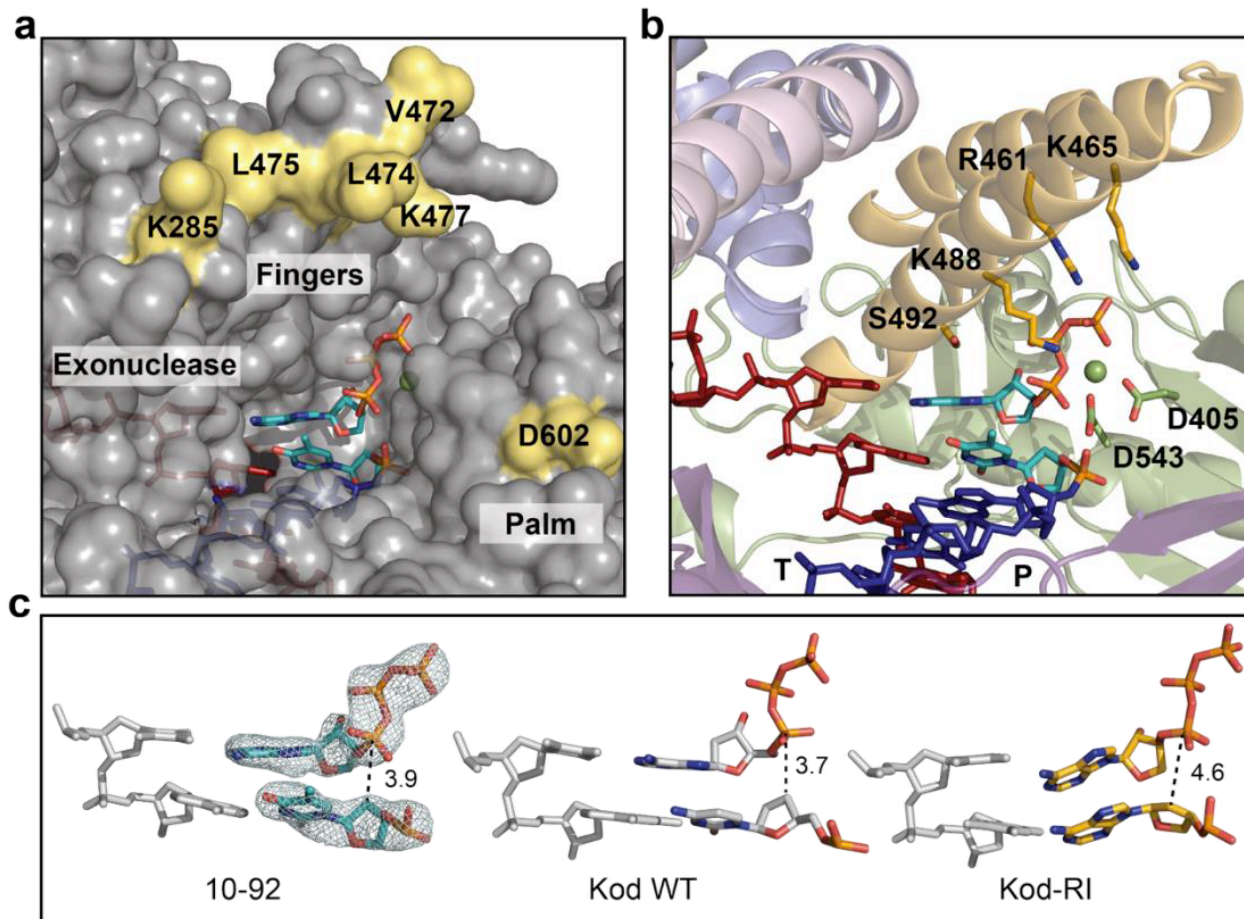


Figure 2.7 Active site of the 10-92 TNA polymerase

(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 4.0 σ .

CHAPTER 3: Directed Evolution of a TNA Polymerase Identifies

Independent Paths to Fidelity and Catalysis

PUBLICATION INFORMATION

This chapter is adapted from the following publication: Hajjar, M., Maola, V.A., Lee, J.J., Holguin, M.J., Quijano, R.N., Nguyen, K.K., Ho, K.L., Medina, J.V., Botello-Cornejo, E., Barpuzary, B. *et al.* (2025), *Nature Communications*, **17**, <https://doi.org/10.1038/s41467-025-67652-1>

© The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Author Contributions

J.C. and N.C. conceived the study and designed the research plan. M.H., V.M., J.L., M.J.H., R.Q., K.N., K.H., J.M., E.B., and B.B. performed the experiments and analyzed the data. J.C. and N.C. wrote the manuscript with input from all authors. M.H., V.M., and J.L. contributed equally. JC and NC are the corresponding authors.

Data Availability

The crystallographic data generated in this study have been deposited in the Protein Data Bank under accession codes 9OAT, 9OAU, 9OAV, 9OAW, 9OAX, and 9OAY. The Protein Data Bank furthermore contains additional structures referenced in this work: 4K8Z, 5OMF, 7OMB, 8T3X. The kinetics, fidelity, and thermostability data generated in this study are provided in the Supplementary Information. Source Data and the Reporting Summary files are available through *Nature Communications*.

Supplementary Data

Supplementary data associated with the publication are available through *Nature Communications* and are reproduced in Appendix C.

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.

Peer Review

Nature Communications thanks Vitor Pinheiro and the other anonymous reviewers for their contribution to the peer review of this work. A peer review file is available online.

Funding

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.

Competing Interests

J.C., V.M., 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.

3.1 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 medicine¹⁻³. Through recursive cycles of diversification and selection, populations of molecules adapt to novel biochemical functions that lie outside the sphere of activities found in nature⁴. Though effective, most protein evolution studies are unable to match the catalytic efficiency of natural enzymes⁵. 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 campaigns⁶⁻⁹, 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 form¹⁰. 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. 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. 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. 1b, 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 (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 (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. 1c and Supplementary Fig. 4), the most active variants identified in rounds 5, 7, and 8, respectively, and the parent sequences used for further diversification¹¹.

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. 1d, 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 10-92¹¹, 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 time¹⁷⁻²². 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. 1d and Supplementary Fig. 8). This result highlights the exceptional structural integrity of the variants and challenges the view that directed evolution inevitably compromises protein folding stability⁵. Such trade-offs are particularly relevant in enzyme engineering campaigns, where enhanced activity often comes at the expense of reduced thermostability²³. Remarkably, the 10-92 TNAP, with its 51 amino acid mutations relative to its closest natural homolog, Kod DNAP (Fig. 1c), exhibits no measurable loss of thermostability (Supplementary Fig. 8). This observation supports the view that protein stability promotes evolvability²⁴, which may be a result of the thermal lysis step of the DrOPS selection protocol²⁵.

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 (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 (Supplementary Table 2). Consistent with all known archaeal B-family DNAPs, the enzymes adopt a disk-shaped architecture (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 subdomains²⁶. In all cases, the duplex is well resolved and tightly bound between the palm and thumb subdomains (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 $\alpha 14$ and $\alpha 15$ are shifted by $\sim 22^\circ$ relative to their position in the open binary structure of wild-type Kod DNAP (Fig. 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 triphosphate²⁸. 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. 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-RI³¹ and

Kod-RSGA³² (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. 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. 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. 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 substate-specific contacts (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-RSGA³² (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 DNAPs²⁰. However, upon correct substrate binding and the ensuing conformational change¹⁹, 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 (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 (Supplementary Table 3).

By combining this data with previously solved structures for Kod-RI³¹ and 10-92¹¹, 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. 3a,e and Supplementary Figure 17). The backbone C α 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. 1d). Surprisingly, most of the 51 acquired mutations found in 10-92 are located >20 Å from the catalytic core (Fig. 3a, and Supplementary Fig. 18), emphasizing the importance of long-range structural interactions in fine-tuning the active site geometry³⁵.

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 (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 synthesis¹¹. 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 Å³ observed in 10-92, which is approximately 10% smaller than the starting enzyme (Fig. 3c–d; 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 DNA³⁶.

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. 3e,f and 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. 4; 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 (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. 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 (Figure 5 and 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 mutagenesis¹¹. Structural analyses show that mutations occur across both surface exposed and buried positions, with most mutations residing far from the catalytic center (Figure 5), again underscoring the importance of long-range interactions in shaping new protein functions³⁵. 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 (Figure 5), highlighting the trade-off between structural integrity and catalytic enhancement²³. Limited reversion analysis of later-stage mutations demonstrates that at least one function-driving mutation is present in each of the key intermediates (Figure 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 10-92¹¹. 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 (Figure 5,6). 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 (Figure 6 and 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 (Figure 6 and Supplementary Figs. 28-30), allowing for enhanced catalytic efficiency. Finally, the S493G mutation observed in 10-92, is located in the active site pocket (Figure 6 and 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 evolution¹¹, 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 evolution¹⁵⁻¹⁷.

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 complexes¹⁶.

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 evolvability²⁴

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 (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 previously³⁸. 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, Supplementary Table 1) was extended with dNTPs or tNTPs to measure DNA or TNA synthesis, respectively, over time³⁹. 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 μ M) and pre-annealed duplex (0.2 μ M each, heated at 90°C for 5 min, cooled on ice) in 1x ThermoPol buffer [20 mM Tris-HCl, 10 mM (NH₄)₂S₄, 10 mM KCl, 2 mM MgSO₄, 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 μ M 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 μ L (6 μ M) EvaGreen® dye, previously identified as the optimal intercalating dye to monitor the synthesis on a DNA template³⁹. Each reaction (25 μ L) 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 polymerase³⁹ by multiplying data with the conversion factor $F = [\text{Template}] * L / [\text{Polymerase}]$, 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 particles⁴⁰. In brief, an aliquot of ~12 million hydrogel particles displaying the acrydite PBS8 primer (GTCCCC**TT**GGGGATACCACC, Supplementary Table 1) was placed into a 1.5 mL and washed twice with 100 μ L 1x ThermoPol buffer. The hydrogels were then resuspended in 100 μ L 1x ThermoPol buffer containing 2 μ M 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 μ L of tNTPs (5 mM stock) and 20 μ L of TNAP (10 μ M 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 μ L 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 μ L 100 mM NaOH (37°C, 3 min) to strip the template, and neutralized with 100 μ L breaking buffer. The hydrogels were resuspended in 100 μ L 1x ThermoPol buffer containing 2 μ M PBS7.PBS9 overhang primer (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 μ L of dNTPs (5 mM stock) and 20 μ L of Bst-LF (10 μ M 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 (Supplementary Table 1). PCR material was ligated into a TOPO vector using a TOPO-TA kit and cloned into DH5 α 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 $\mu_{\text{exp} \rightarrow \text{obs}} = (\# \text{observed} / \# \text{expected}) \times 1000$. 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/GTCCCCCTTGG & EM619: CCCACACCCTCCTATCGCTAAACACACACTTAATAAAGTTGGTGGTATCCCCAAGGG GAC, 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 (ACGCATGCGGCCGCTCAAGTTCCTTTCGGCGTCAG), 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 (ACGCATGCGGCCGCTCACTTCTGGTAGCGCAGGTC, Supplementary Table 1);

8-64-Rvs_760: ACGCATGCGGCCGCTCAAGTTCCTTTCGGCGTCAG 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 purified³¹. Briefly, BL21(DE3) cells harboring pET21-*tnap* were grown aerobically at 37 °C in LB medium containing 100 µg mL⁻¹ ampicillin. At an OD₆₀₀ 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 (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 XDS⁴¹. Data collection statistics are summarized in Tables S2 and S3. Initial models were determined by molecular replacement using Phaser⁴²; 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 Coot⁴³ and refinement with phenix⁴⁴. The stereochemistry and geometry of all structures were validated with Molprobity⁴⁵ 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 PyMOL⁴⁷. Local base-pair and base-pair step parameters are computed using Web 3DNA⁴⁸. 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 (Supplementary Fig. 3) were performed using the EMBOSS Needle algorithm⁴⁹ on the EMBL–EBI platform⁵⁰. Conservation was quantified with BLOSUM62 substitution scores⁵¹. Functional effects of reversions to 5-270 were quantified from the 45-sec primer extension assay as previously reported¹¹. Band intensities were measured by densitometry in ImageJ⁵² 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 FoldX⁵³ 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 FreeSASA⁵⁴ with the default parameters. Values were reported as relative exposure (%) in ternary complexes of Kod, 5-270, and 10-92.

AlphaFold3 predictions

AlphaFold3⁵⁵ 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 Mg²⁺ 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 Mg²⁺ ions.

2'-deoxy- α -L-threofuranosyl thymidine-3'-triphosphate (dTTP) 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 μ m particle size) for most of the crude reaction mixture. Yields are reported as isolated yields of pure compounds. ¹H, ¹³C, and ³¹P NMR spectra were analysed on 400 MHz NMR spectrometer (Bruker Avance NEO). ¹H values are reported in parts per million relative to Me₄Si or corresponding deuterium solvents used as an internal standard. ¹³C values are reported in parts per million relative to corresponding deuterium solvents used as an internal standards. ³¹P 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 ZnCl₂ (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% H₂O 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% H₂O/isopropanol containing 1% DIPEA) by liquid loading (2 mL of 2% H₂O/isopropanol containing 1% DIPEA) with eluents 5% to 10% H₂O/isopropanol containing 1% DIPEA and then 5% H₂O 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**. ³¹P NMR (162 MHz, D₂O) 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 NH₄OH 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 μ m syringe filter and dropwise added to the forty times volume of acetone (80 mL) at room temperature containing NaClO₄ (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). ³¹P NMR (162 MHz, D₂O) δ -8.09, -12.42 (d, J = 19.5 Hz), -22.52 (t, J = 19.5 Hz, 1H).

3.6 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. Doublié, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å 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- α -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 α -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).

3.7 FIGURES

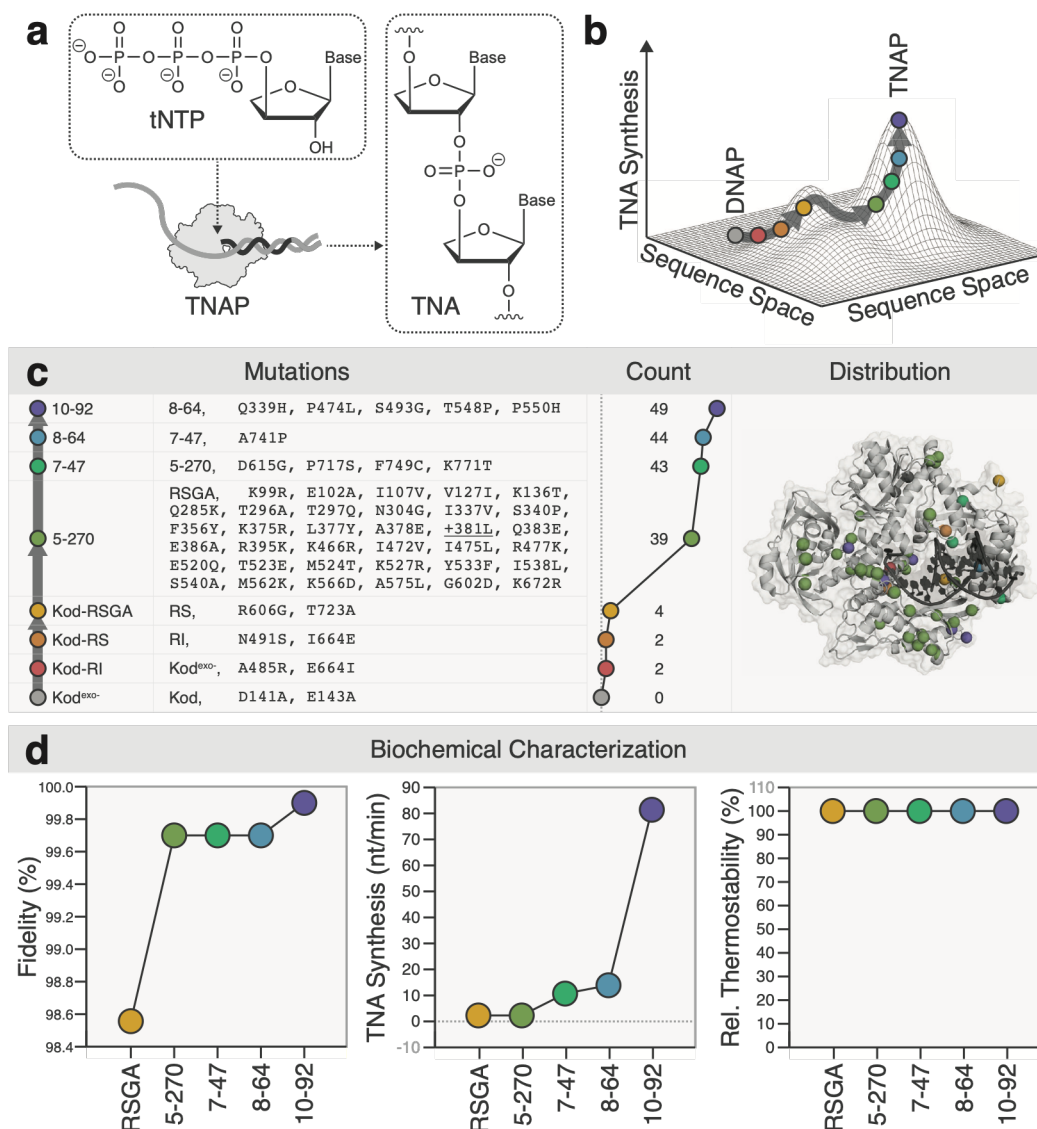


Figure 3.1 Evolution of the 10-92 TNA polymerase

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).

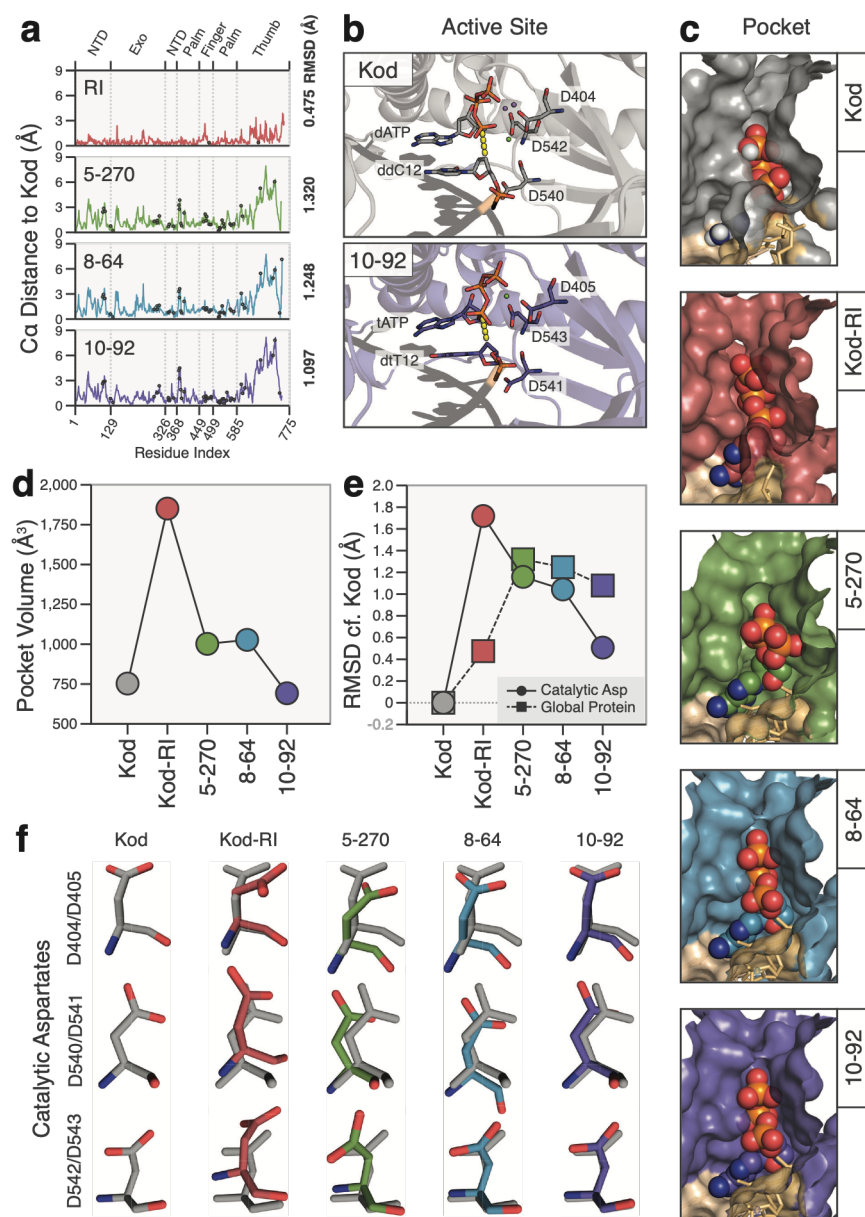


Figure 3.3 Stepwise changes in global and local structural conformation

a, The distance in angstroms from all resolved C- α 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 (Å³) for Kod, Kod-RI, 5-270, 8-64, and 10-92. **e**, Root mean square deviation (RMSD) of C- α 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.

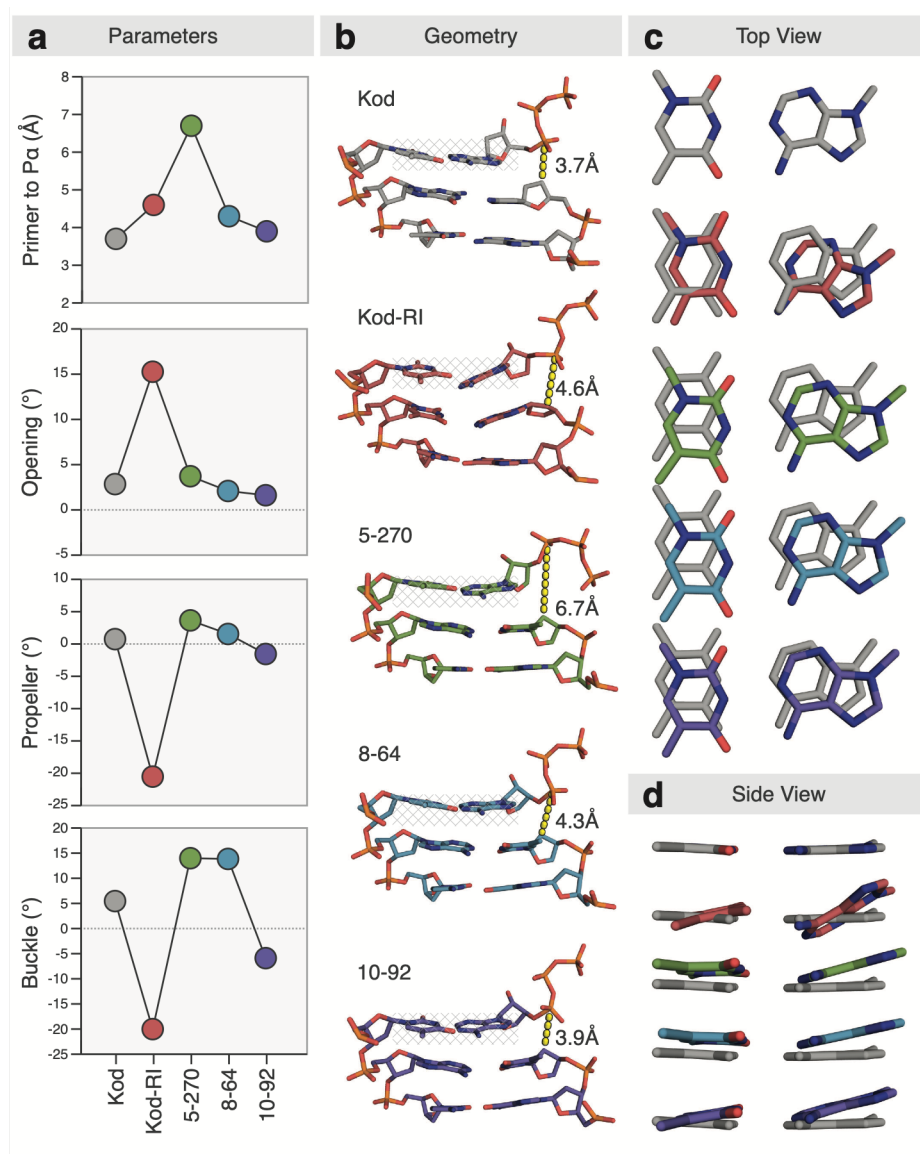


Figure 3.4 Structural optimization of the nascent base pair and reaction center

a, Nucleic acid structural parameters plotted across generations of TNAP. Measurements reflect changes in the distance from the primer to the α -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 α -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.

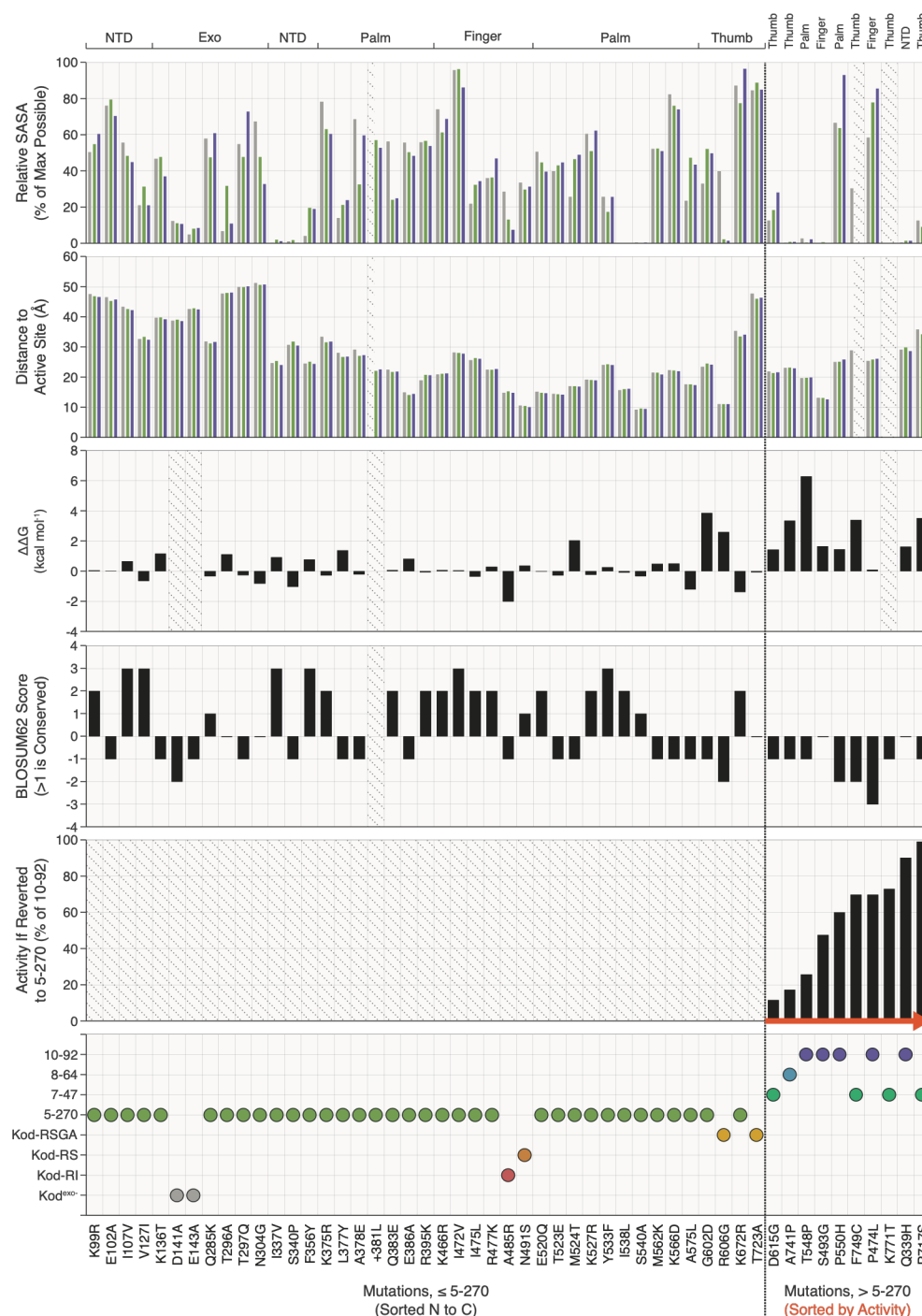


Figure 3.5 Integrated analysis of 51 mutations acquired during the evolution of 10-92

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 data¹¹), BLOSUM62 score, predicted stability change ($\Delta\Delta G$, 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.

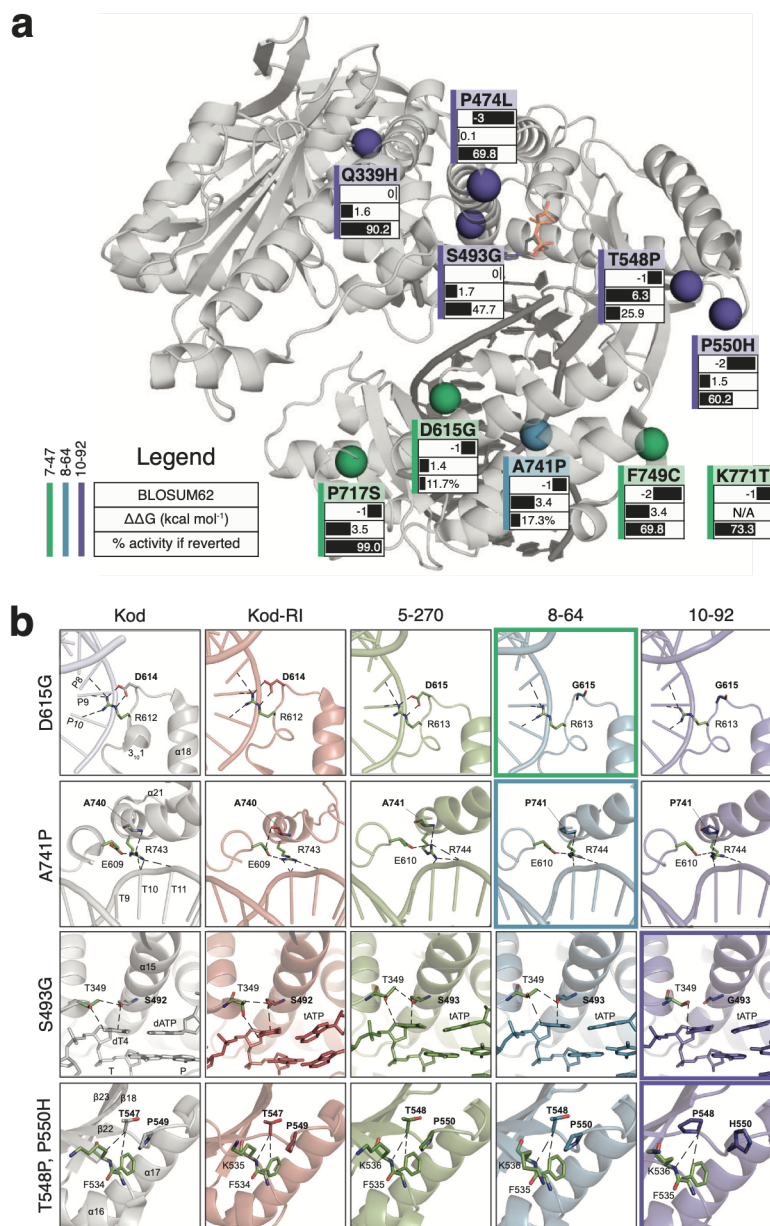


Figure 3.6 Structural analysis of key mutations driving improved catalysis

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 $\Delta\Delta G$ (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: Rapid Evolution of a Highly Efficient RNA Polymerase by Homologous Recombination

PUBLICATION INFORMATION

This chapter is adapted from the following publication: edina, E.L., Maola, V.A., Hajjar, M., Ko, G.K., Ho, E.J., Horton, A.R., Chim, N. and Chaput, J.C. (2026), *Nature Chemical Biology*, **22**, <https://doi.org/10.1038/s41589-025-02124-7>

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2026

Author Contributions

J.C.C., N.C. and E.L.M. conceived the project and designed the experiments. E.L.M. and V.A.M. evolved the enzyme. E.L.M., V.A.M., M.H. and A.R.H. performed the biochemical characterization. M.H., G.K.K. and E.J.H. crystallized the enzyme. J.C.C., E.L.M., V.A.M. and M.H. wrote the manuscript. All authors reviewed and commented on the manuscript. JC is the corresponding author.

Data Availability

The atomic coordinates and structure factors for the crystal structure of C28 have been deposited with the Research Collaboratory for Structural Bioinformatics (PDB ID: 9MYE). Source Data and the Reporting Summary files are available through *Nature Chemical Biology*.

Supplementary Data

Supplementary data associated with the publication are available through *Nature Chemical Biology* and are reproduced in Appendix D.

Peer Review

Nature Chemical Biology thanks Jack Szostak, Yi Shi for their contribution to the peer review of this work.

Funding

This work was supported by the CLP and GM Divisions of the National Science Foundation (CHE-2433788 to J.C.C.) and the University of California, Irvine.

Competing Interests

J.C.C., E.L.M., V.A.M. and the University of California, Irvine, have filed a patent application on the composition and activity of the C28 RNA polymerase. The other authors declare no competing interests.

4.1 ABSTRACT

Engineering DNA polymerases to efficiently synthesize artificial or noncognate nucleic acids remains an essential challenge in synthetic biology. Here we describe an evolutionary campaign designed to convert a family of highly selective DNA polymerases into an unnatural homolog with strong RNA synthesis activity. Starting from a homologous recombination library, a short evolutionary path was achieved using a single-cell droplet-based microfluidic selection strategy to produce C28, a newly engineered polymerase that can synthesize RNA with a rate of $\sim 3 \text{ nt s}^{-1}$ and of $>99\%$ fidelity. C28 is capable of long-range RNA synthesis, reverse transcription and chimeric DNA–RNA amplification using the PCR. Despite strong discrimination against other genetic systems, C28 readily accepts several 2'F and base-modified RNA analogs. Together, these findings highlight the power of directed evolution as an approach for reprogramming DNA polymerases with activities that could help drive future applications in biotechnology and medicine.

4.2 INTRODUCTION

Replicative DNA polymerases are a class of highly faithful enzymes responsible for transferring genetic information from parent to daughter strands during cell division. These enzymes follow a complex reaction mechanism in which each cycle of nucleotide addition involves a set of dynamic conformational changes that allow the enzyme to form a covalent phosphodiester bond between the growing daughter strand and the incoming 2'-deoxyribonucleoside triphosphate (dNTP)¹⁻³. While the intrinsic sequence fidelity of the enzyme is governed by conformational dynamics and kinetic partitioning⁴⁻⁶, the ability for DNA polymerases to discriminate against ribonucleoside triphosphates (NTPs), present in the cell at levels up to 100-fold in excess of dNTPs⁷, has been attributed to a series of structural checkpoints that maintain the integrity of the genome by ensuring proper dNTP selection, chemical bond formation and continued DNA synthesis⁸⁻¹⁰.

Unraveling the mystery of how natural DNA polymerases can achieve high levels of substrate specificity remains an essential goal in synthetic biology, as such knowledge is expected to help drive the development of custom polymerases for future applications in biotechnology and medicine¹¹. This problem has been intensely studied in the context of RNA discrimination by DNA polymerases, such as *Escherichia coli* (*E. coli*) DNA polymerase I, which discriminates against NTPs by a factor of 100k-fold¹². Structural and biochemical studies have shown that this remarkable level of substrate specificity is achieved through a 'steric gate' residue, which blocks NTPs from entering the active site by packing against the 2' sugar position of incoming nucleotides¹²⁻¹⁴. However, mutations made to the steric gate position and other active site positions identified by structure-based, in vivo and in vitro selection strategies have yielded variants with limited RNA synthesis activity¹⁵⁻¹⁹. One exception is Tgo-QGLK²⁰, a designed polymerase

isolated from the archaeal species *Thermococcus gorgonarius* (Tgo; Fig. 1a), bearing mutations at the uracil stalling (V93Q)²¹, steric gate (Y409G)¹² and Terminator (A485L)²² positions, as well as an auxiliary site (E664K) responsible for enhancing the synthesis of xeno-nucleic acids (XNA)²³. The use of Tgo-QGLK in the laboratory led to the generation of an RNA-containing bacterial genome²⁴ and the synthesis of hypermodified RNA²⁵.

Recognizing that polymerases exert stringent geometric control over the shape and orientation of the incoming substrate^{8,26}, many groups have suggested that a precise set of determinants regulates polymerase specificity^{18,20,27}. An alternative view is that polymerase specificity is governed by long-range interactions within the protein structure, which are required to optimize the active-site pocket for efficient nucleic acid synthesis. Support for this model can be found in a recently evolved polymerase that synthesizes threose nucleic acid (TNA) at a rate of 1 nt s⁻¹ with 99.9% fidelity²⁸, which is within 5–200-fold of DNA synthesis rates observed in the parental enzymes²⁹. Notably, the crystal structure of the 10–92 polymerases caught in a catalytically relevant state revealed large rearrangements in the thumb and palm subdomains that markedly improved the geometry of the TNA triphosphate in the active site²⁸. This observation suggests that the evolutionary path from natural DNA polymerases to highly specialized XNA enzymes may be farther than previously assumed, as active-site refinement could necessitate topological changes that are not easily achieved by rational design.

Here we explore the problem of substrate specificity using directed evolution to identify the amino acid changes required to transition a family of highly selective DNA polymerases into an unnatural homolog with robust RNA synthesis activity. After a short evolutionary path, a newly engineered polymerase, termed C28, was identified that can synthesize RNA with a rate of ~3 nt s⁻¹ and >99% fidelity. The C28 RNA polymerase accepts several RNA analogs, and facilitates

long-range synthesis, reverse transcription (RT) and chimeric DNA–RNA amplification using the PCR. A crystal structure of the polymerase reveals interesting electrostatic changes near the primer–template (P–T) duplex that contribute to the enhanced RNA synthesis activity of the enzyme. Together, these findings illustrate the importance of directed evolution as a strategy for reprogramming DNA polymerases with new activities.

4.3 RESULTS

Library diversification

We chose homologous recombination as the diversification strategy for producing our starting library³⁰. This technique, which recombines protein fragments from related species, generates diverse library repertoires with a higher frequency of functional proteins than error-prone strategies³¹ and has been used previously to engineer DNA polymerases for the ability to bypass lesions, function under inhibitory conditions and incorporate fluorescence tags³²⁻³⁴. Building from prior work in the field³⁵, we computationally designed a polymerase library (Fig. 1b) based on four naturally occurring B-family DNA polymerases isolated from the hyperthermophilic archaeal species Tgo, *Pyrococcus* Deep Vent (DV), *Thermococcus* 9°N (9°N) and *T. kodakarensis* (Kod). These enzymes are highly homologous (89–96% sequence identity) and generally known to be more accepting of modified nucleotides than A-family polymerases³⁶. The four 2.4-kb genes were designed to contain 11 conserved regions needed for recombination by PCR, two exonuclease-silencing (exo) mutations (D141A and E143A) and four gain-of-function mutations (V93Q, Y409G, A485L and E664K) known to enhance RNA synthesis activity²⁰.

Each gene was individually prepared as linear double-stranded DNA by high-fidelity gBlock assembly (Supplementary Figs. 1 and 2 and Supplementary Tables 1 and 2). Homologous recombination was then achieved by DNA shuffling, wherein PCR was used to disassemble the genes into 48 distinct fragments (12 segments $\times$ 4 genes) that were subsequently reassembled in a single PCR reaction (Fig. 1b and Supplementary Figs. 3 and 4) to produce a polymerase library with a theoretical diversity of 412 variants³⁵. The library was ligated into a protein expression

vector, cloned and sequenced (Fig. 1c). DNA sequencing of representative members revealed a rich pool of highly diverse variants that comprised the expected recombinants as well as new recombinants generated by template switching during thermocycling and spontaneous random single-nucleotide mutations that together increased the diversity of the library beyond the theoretical limit defined by recombination.

Single-cell directed evolution

Polymerase evolution was accomplished using the massively parallel, single-cell, microfluidics approach of droplet-based optical polymerase sorting³⁷. Droplet-based optical polymerase sorting is a two-step microfluidic strategy (Fig. 1d) in which a population of *E. coli*, each expressing a different polymerase variant, is encapsulated in water-in-oil droplets along with the biochemical reagents needed to perform a polymerase activity assay (PAA). Polymerases are released into the water-in-oil compartment by thermal lysis and challenged to extend a self-priming template designed as an optical sensor for polymerase activity³⁷. Enzymes that can extend the primer to full length generate a fluorescence signal by disrupting a fluorophore–quencher pair located at the 3' end of the template. Functional polymerases are then separated from the larger population of inactive variants by passing the droplets through a custom-built fluorescence-activated droplet sorting (FADS) instrument at a rate of ~1 kHz³⁸. Flowing droplets pass through a focused laser line located upstream of a salt-water electrode and Y-junction that divides the channel into the waste and collection lines. Droplets that achieve a user-defined level of laser-induced fluorescence are positively sorted into the collection bin and the encoding DNA is recovered, amplified by PCR, reinserted in an expression vector and transformed back into *E. coli* for another round of selection (Supplementary Table 3).

After two rounds of selection (Fig. 2a), a screen of ~100 members identified five promising variants with activities for RNA synthesis that were equivalent to or slightly better than that of Tgo-QGLK (Supplementary Fig. 5). The protein-coding region of the five variants was then shuffled to create a new polymerase library that was subjected to a third round of selection (Fig. 2b). A second screen of ~100 variants isolated by droplet-based polymerase screening identified clone C28 as a highly efficient DNA-dependent RNA polymerase. Comparative sequence analysis shows that C28 differs from its closest natural homolog, Tgo DNA polymerase, by 45 amino acid mutations, including the two exonuclease-silencing mutations (D141A and E143A) and four gain-of-function mutations (V93Q, Y409G, A485L and E664K) present in the starting library (Supplementary Fig. 6 and Supplementary Table 4). The 39 acquired mutations are distributed throughout the protein and comprise segments from each of the four ancestral proteins (Fig. 2c,d). A reversion analysis performed in segments reveals greater importance of mutations in the catalytic domain than in the N-terminal and exonuclease domains (Extended Data Fig. 1).

Biochemical characterization

We characterized the activity of the C28 RNA polymerase using standard biochemical assays. The polymerase was challenged by extending a DNA primer with 40 ribonucleotides (nt; Fig. 3a). After incubation, the reactions were quenched, and the primer-extended products were analyzed by denaturing PAGE. The results demonstrate that C28 is a more efficient RNA polymerase than Tgo-QGLK, capable of completely extending a DNA primer with 40 nt in 10 s as compared to the trace amounts of full-length product observed for Tgo-QGLK after 30 s (Fig. 3b). This level of activity is only slightly diminished under RNA priming conditions, whereby the reaction requires an additional 5 s to convert the primer to full-length product (Supplementary Fig.

7) and extends to long-read templates with coding regions of ~1,000-nt positions (Supplementary Fig. 8). Consistent with the extreme thermal stability of the parent enzymes, C28 maintains an optimum reaction temperature of 50–70 °C and exhibits a level of RNA synthesis activity that is undiminished after 6 h of prolonged heating at 90 °C (Fig. 3c,d). This feature of the polymerase may be due to the thermal lysis step in the selection protocol, which requires the enzyme population to withstand 5 min of heating at 95 °C.

Next, we evaluated the substrate specificity of C28 by challenging the enzyme to recognize a set of natural and artificial nucleic acids (XNAs)³⁹. The set of substrates included nucleoside triphosphates for the genetic classes of DNA, RNA, 2'-fluoroarabino nucleic acid (FANA), arabino nucleic acid (ANA), hexitol nucleic acid (HNA), TNA and phosphonomethylthreosyl nucleic acid (PMT). Analysis of the resulting gel indicates that C28 strongly discriminates against ANA, HNA, TNA and PMT after permissive extension conditions of 1-h incubation at 55 °C, with rapid and definitive chain termination occurring after a limited number of nucleotide incorporations (Fig. 3e). By comparison, the enzyme shows comparable activity for DNA and FANA, as evidenced by the presence of similar product bands after a brief incubation of 1 min at 55 °C (Fig. 3e). FANA synthesis was considered unsurprising due to its close structural similarity to DNA—a single hydrogen-to-fluorine substitution at the 2' carbon position of the sugar—and known recognition by the natural parent polymerases of Tgo, DV, 9°N and Kod⁴⁰. Notably, the strong substrate specificity observed for C28 against ANA, HNA and TNA contrasts with the weaker specificity observed previously for Tgo-QGLK whereby ANA, HNA and TNA synthesis yielded longer extension products⁴¹. Together, these findings support a change in the checkpoint behavior of the enzyme, with C28 being more selective for RNA substrates than its synthetic homolog Tgo-QGLK.

It is widely known that engineered polymerases can have intrinsic template-sequence biases that limit their activity in practical applications⁴². This problem is often overlooked when polymerases are evaluated against individual templates designed to be unstructured and missing motifs that may be difficult to copy. To examine the potential for template-sequence bias in our assays, we repeated the primer-extension assays using a set of six different degenerate DNA libraries (L1–L6), each containing a central random region of 40 contiguous nt positions with AT ratios of either 60:40 or 50:50 that were flanked with fixed-sequence primer binding sites (Fig. 3f). PAGE analysis of the resulting products reveals a striking difference in the RNA synthesis activity of the C28 polymerase relative to Tgo-QGLK (Fig. 3f). While C28 is able to extend nearly all of the primer to full-length product in all cases tested, Tgo-QGLK yielded markedly reduced levels of full-length RNA with noticeably diminished activity against the libraries prepared with an equal distribution of all four nucleotides (Fig. 3f). These findings support a more efficient C28 enzyme that functions with reduced template bias.

Given the fast catalytic rate observed for the C28 RNA polymerase, we decided to measure the rate of RNA synthesis more precisely using the technique of polymerase kinetic profiling⁴³. Unlike traditional single-nucleotide incorporation assays, which measure the rate of one nucleotide addition to the 3' end of a DNA primer⁴⁴, polymerase kinetic profiling measures the average rate of nucleotide synthesis across a DNA template⁴³. In doing so, the values represent a more holistic view of the catalytic cycle, which includes substrate recognition, catalysis and continued synthesis. Time-course experiments were performed by incubating an unlabeled P–T duplex with the C28 polymerase and NTPs at 55 °C (Supplementary Figs. 9–10). Individual reactions were stopped at designated time points by plunging the reaction vessel into powdered dry ice. The reactions were thawed on a cold plate in the presence of LCGreen Plus, previously identified as an optimal

quantitative intercalating dye for DNA–RNA duplexes⁴³, and the rates were evaluated by plotting the fluorescence values as a function of time. Results from background subtracted plots provide an RNA synthesis rate of $\sim 3.3 \text{ nt s}^{-1}$ for C28, which is noticeably faster than Tgo-QGLK ($\sim 0.6 \text{ nt s}^{-1}$).

Next, we evaluated the functional activity of the C28 RNA polymerase in various formats. We began by measuring the fidelity of RNA synthesis using a hydrogel particle-display approach to facilitate the replication cycle ($\text{DNA} \rightarrow \text{RNA} \rightarrow \text{DNA}$)⁴⁵. As illustrated in Fig. 4a, this assay uses hydrogel-coated magnetic beads prepared with a DNA primer that becomes covalently cross-linked throughout the gel matrix (Supplementary Fig. 11). The favorable solution-like environment allows reagents and enzymes to flow freely in and out of the hydrogel matrix, enabling RNA transcription and RT to occur without the need for separate purification steps, as templates are removed by washing with appropriate buffers. After replication, the cDNA is recovered, amplified, cloned and sequenced. Using this strategy, we measured the aggregate fidelity of RNA replication by sequencing $\geq 1,000$ -nt incorporation events from a full replication cycle using controls to eliminate PCR artifacts (Supplementary Fig. 12). Our analysis reveals that C28 functions with an observed fidelity of $\sim 99.8\%$ when paired with Bst DNA polymerase as a separate enzyme for RT due to its ability to synthesize DNA on RNA templates⁴⁶ (Fig. 4b and Supplementary Fig. 13).

Recognizing that the rapid RNA-extension assay described in Fig. 3b shows a small fraction of lighter bands below the full-length product, we sought to ensure that our fidelity measurements were not inadvertently biased by nucleotide misincorporation events that could be missed in the sequencing analysis due to premature termination. Gratifyingly, these bands disappear when the RNA-extension assay is repeated under the conditions used for the fidelity

assay, which involved a much longer incubation time of 1 h (Extended Data Fig. 2). As a second control, we evaluated the efficiency of mismatch extension under the conditions used for the fidelity assay. Although some mismatches are more challenging to read through than others, the data provide convincing evidence that the time frame used for the fidelity experiment is sufficiently long as to allow most mismatches to be extended to full-length product (Extended Data Fig. 3). Taken together, we conclude that the small fraction of lighter intensity bands observed during the rapid extension times of 10–60 s are primers caught in the process of extension, rather than premature truncation events caused by nucleotide misincorporation.

Next, we assessed the ability of C28 to synthesize DNA on templates composed of DNA, RNA and FANA. The results indicate that C28 can copy all three templates into DNA, albeit with reduced efficiency for the RNA template (Fig. 4c). Given the ability for C28 to copy genetic information back and forth between DNA and RNA, we chose to perform a second hydrogel-based fidelity assay using C28 for both enzymatic steps. Our results indicate that C28 functions with an aggregate fidelity of ~99.6% when used as the same enzyme to transcribe and reverse transcribe information between DNA and RNA (Fig. 4d and Supplementary Fig. 14).

Encouraged by this result, we explored the potential of C28 to incorporate RNA nucleotides into double-stranded DNA via PCR. Although Romesberg has previously shown that SFM4-3 can amplify oligonucleotides of mixed RNA and DNA composition, that study required unusually long extension times (>10 h)⁴⁷. Using C28, we observe the formation of chimeric DNA–RNA amplicons under standard thermocycling times and reaction conditions (Fig. 4e). The reaction allows for all single-nucleotide substitutions of dNTPs (dA, dC, dG, dT) for their corresponding NTP (A, C, G, U) and double nucleotide substitutions of purines (rA and rG) and

pyrimidines (rU and rC). However, further combinations, such as mixed purine–pyrimidine substitutions (rA–rU or rC–rG pairs) fail to produce the desired amplicon, indicating that C28 strongly discriminates against NTPs when the complementary ribonucleotide is present in the template.

Given the importance of chemical modifications in RNA biology and medicine, we chose to examine the ability of C28 to recognize RNA analogs found in biomedical applications. Common examples include 5-methylcytidine and pseudouridine, both found in the SARS-CoV-2 mRNA vaccines developed by Moderna and Pfizer, and 2'-modified RNA analogs present in many FDA-approved oligonucleotide therapeutics⁴⁸. Our screen of several sugar-modified and base-modified RNA analogs (Fig. 5a), used in place of the standard UTP or CTP monomers, indicates that C28 is widely accepting of diverse RNA substrates (Fig. 5b). We also tested the ability of C28 to recognize 2'-O-methyl-RNA substrates by sequentially replacing each NTP with its corresponding 2'-O-methyl-NTP. However, these reactions produced mostly truncated products, indicating that additional structural changes are needed for C28 to recognize bulky 2' substituents (Supplementary Fig. 15).

Structural analysis

The robust catalytic activity of the C28 RNA polymerase motivated us to examine the structural impact of the selected mutations by X-ray crystallography. Structural insights were obtained by solving a binary crystal structure of the polymerase in a postcatalytic state bound to a P–T duplex extended with a single ribonucleotide (U12) complementary to a templating deoxyadenosine (dA7) residue. Diffraction-quality crystals were obtained from a screen of ~900

conditions performed in a hanging-drop format. The best crystal resolved to a diffraction limit of 2.22 Å resolution and was solved by molecular replacement (Supplementary Table 5).

The protein structure reveals a disk-shaped architecture (Fig. 6a), consistent with all known archaeal B-family DNA polymerases⁴⁹. The structure is well-ordered and strikingly similar to natural Kod DNA polymerase⁵⁰ with a full backbone (C α) root-mean-square deviation (RMSD) of 0.916 Å between the two structures. Inspection of the active site reveals unambiguous electron density for the U12 ribonucleotide added to the 3' end of the DNA primer as well as the steric gate mutation at position G410 (Fig. 6a). The incorporated U12 residue exhibits canonical Watson–Crick base pairing with the templating adenine base and adopts a C2'-endo sugar pucker, commonly associated with DNA⁵¹ (Supplementary Figs. 16–19 and Supplementary Tables 6–8). Interestingly, this structure offers crystallographic evidence of a steric gate mutation in a B-family polymerase, showing how the tyrosine to glycine mutation eliminates steric bulk at this position (Fig. 6a).

Recognizing that electrostatic forces have an important role in protein folding and function and that their long-range interactions are often critical to complex multistep reactions⁵², we decided to compare the electrostatic surface potential of C28 to that of natural Kod DNA polymerase (Fig. 6b,c). The resulting analysis shows a continuous positive patch at the duplex interface of C28 that is partially disrupted in Kod, thereby increasing coordination of the duplex by the C28 polymerase. The change in surface potential at this region is largely driven by the E664K mutation present in the starting library, and originally discovered to enhance the synthesis of XNA nucleotides²³. Given the close structural homology of C28 to Kod, coupled with the substantial gain in RNA synthesis activity of the C28 polymerase, we postulate that the functional

role of the selected mutations is to improve the dynamic behavior of the protein through long-range interactions (Extended Data Fig. 4) in the catalytic domain that contribute to RNA synthesis (Extended Data Fig. 1).

4.4 DISCUSSION

DNA polymerases have a crucial role in preserving genome integrity during cell division by using a set of checkpoints that prevent the misincorporation of RNA nucleotides into DNA chromosomes⁵³. These safety checks allow the polymerase to recognize and stall when NTPs are mistakenly added to the growing DNA strand, showcasing a highly selective mechanism of substrate specificity. This process has become a model for exploring the underlying determinants of substrate discrimination, with the aim of establishing rules that can be used to reprogram DNA polymerases for the ability to recognize new, user-defined substrates. Researchers have identified several key mutations, including the steric gate^{12–14}, as well as local and distal mutations, that reduce the barrier to NTP discrimination^{15–19}. However, apart from one example²⁰, these sequence alterations have resulted in enzymes with only limited RNA synthesis activity^{12–19}, indicating that this problem is more complex than initially assumed.

These observations suggest that the evolutionary gap required to convert natural polymerases into highly efficient enzymes with new catalytic activities may be larger than expected, necessitating extensive mutations in protein architecture to refine the active-site pocket for compatibility with alternative sugar chemistries. To explore this hypothesis, we used a homologous recombination strategy to examine a broad range of enzyme variants that sampled different regions of the fitness landscape³⁰. This library approach, inspired by principles of mammalian immune diversity, has frequently been used to isolate new enzymes by directed evolution⁵⁴. It offers a high proportion of functional starting sequences by reducing the likelihood of deleterious stop codons commonly associated with error-prone methods⁵⁵. To maximize library sampling, we used a massively parallel screening platform that evaluates microfluidic droplets at a rate of 1 kHz, allowing us to sample 108 droplets in ~10 h⁵⁶. The library design and screening

process, combined with an iterative recombination strategy, resulted in the discovery of an engineered polymerase that can synthesize RNA with near-natural catalytic activity ($\sim 3 \text{ nt s}^{-1}$, $>99\%$ fidelity). For example, the parent enzymes have activities with natural DNA substrates that are only 2–65-fold faster than RNA synthesis with C28 (ref 29). Consistent with our hypothesis, the molecular evolution process required the accumulation of 39 new amino acid mutations to achieve a heightened state of catalytic activity for RNA synthesis. Although most of these activities appear to be driven by mutations in the catalytic domain (Extended Data Fig. 1), this result was surprising, given the close structural relationship between DNA and RNA, which differs only by the presence of a 2'-hydroxyl group.

Analysis of the X-ray crystal structure of the C28 polymerase reveals the importance of long-range interactions as a mechanism for optimizing the catalytic activity of engineered polymerases. In this case, all 39 mutations lie outside the active site pocket, most occupying regions of the enzyme between 35 and 60 Å from the 3' hydroxyl group of the DNA primer. Recognizing the close structural homology ($\text{RMSD} < 1 \text{ Å}$) between C28 and its natural homolog, we speculate that the selected mutations improve the catalytic activity of the enzyme by optimizing a network of long-range interactions that are critical to the dynamic motion of the enzyme, which undergoes large conformational changes during each cycle of nucleotide addition. Given the difficulty in modeling enzymes with complex reaction mechanisms involving dynamic long-range interactions, it is unlikely that these mutations could have been identified by advanced computational methods.

The C28 RNA polymerase overcomes several key challenges in RNA biotechnology by providing an efficient and cost-effective approach to producing natural and modified RNA oligonucleotides with certain inherent advantages over current chemical and enzymatic strategies⁵⁷. Solid-phase synthesis based on phosphoramidite chemistry is restricted by its

scalability, sustainability and length⁵⁸. Iterative cycles of coupling, capping, oxidation and deblocking produce large quantities of hazardous waste that require special disposal. The high cost and reduced coupling efficiency of RNA synthesis relative to DNA synthesis are resource demanding and limit the length of synthetic RNAs to 50–60 nt, respectively⁵⁹. By comparison, C28 offers an environmentally friendly biocatalytic alternative to solid-phase synthesis with no hazardous waste and RNA lengths up to 1 kb. Unlike T7 RNA polymerase, which restricts the 5' terminus to the GGG initiator motif, C28 follows a primer-extension mechanism that can accept nearly any sequence combination and chemical modification in the 5' region. Additionally, C28 operates at elevated temperatures, making it easier to read through structured templates, and enzymatic strategies are available to site-specifically label RNA nucleotides by selectively replacing the template at intermediate steps in the synthesis^{25,60–62}. Together, these advantages offer a versatile tool to explore RNA biochemistry in vitro, and possibly in vivo²⁴.

Strong substrate recognition of the C28 polymerase toward known RNA modifications offers the potential to overcome barriers that currently limit the production of RNA therapeutics⁵⁸. This could include, for example, the synthesis of 2'-fluoro-modified RNA found in antisense oligonucleotides and small interfering RNA therapeutics⁴⁸, which could be produced by enzymatic cascade strategies that remove the 5'-primer region⁶³. The long-read synthesis capability, coupled with its recognition of a suite of base modifications, offers a path for expanding the chemistry of mRNA vaccines, while avoiding unwanted immunostimulatory byproducts, such as short dsRNA, that can accumulate as a byproduct of T7 RNA transcription⁶⁴. Finally, the ability of C28 to synthesize chimeric DNA–RNA oligonucleotides by PCR provides an exciting opportunity to generate interesting molecules that may be challenging to produce using current chemical and enzymatic pathways due to limitations in either cost or substrate compatibility.

In conclusion, we demonstrate that an engineered polymerase can achieve high catalytic activity by exploring diverse regions of the fitness landscape. The approach taken is generalizable and could be used to dramatically expand the synthetic biology toolbox by producing new examples of enzymes that can synthesize diverse classes of nucleic acid molecules.

4.5 METHODS

Reagents

Reagents were obtained from the following sources: DNA oligonucleotides and gBlocks from IDT; RNA, FANA and ANA triphosphates from TriLink Biotechnologies; TNA triphosphates chemically synthesized⁶⁵; HNA and PMT triphosphates provided by Piet Herdewijn; the FANA template was synthesized on an ABI3400 DNA synthesizer using FANA phosphoramidites and chemical synthesis reagents from Glen Research; 10× ThermoPol buffer (20 mM Tris–HCl, 10 mM (NH₄)₂SO₄, 10 mM KCl, 2 mM MgSO₄, 0.1% Triton X-100, pH 8.8), Q5 high-fidelity DNA polymerase, Taq DNA polymerase, Gibson assembly, DpnI, T4 DNA ligase, Monarch DNA Gel Extraction Kit, NEB 5-α and BL21 (DE3) competent *E. coli* cells from New England Biolabs; XL10-gold competent cells from Agilent; TOPO-TA cloning, ethylenediaminetetraacetic acid (EDTA), Tris, Dynabeads M-270 carboxylic acid 2.8 μM, isopropyl alcohol, ethanol, N,N,N',N'-tetramethylethane-1,2-diamine (TEMED), Thermo Scientific GeneRuler from Thermo Fisher Scientific; silicone emulsifiers KF-6012, KF-6038 and DM-FLUID-A-6cs from Shin-Etsu Chemical; ammonium persulfate, mineral oil and Triton X-100 from Sigma-Aldrich; acrylamide (40%), bis-acrylamide 19:1, SDS from Bio-Rad; plasmid mini-prep kit from Biomiga; Zymo PCR clean-up kit from Zymo; all purification columns from GE Healthcare; polydimethylsiloxane base, curing agent from Dow Corning; SU-8 2025 photoresist from Thermo Fisher Scientific; fluorinated oil HFE-7500 from 3M Novec; Pico-Surf, Pico-Glide, Pico-Break from Sphere Fluidics; Tygon tubing (EW-06419-01) from Cole-Parmer; SMC ITV0011-2UMS digital pressure regulator from Automation Distribution; 552-nm laser from Coherent OBIS LS; apochromatic objective from Motic; field-gated programmable array (FPGA) from National Instruments; high-voltage amplifier from Trek 2210; crystallization screening kits

Morpheus I, Morpheus II and Morpheus III from Molecular Dimensions; Crystal Screen, Index, PEG/ION from Hampton Research; PACT, ProComplex and JCSG+ are from NeXtal; crystallography UTP from Sigma-Aldrich; and adenosine-5'-((α,β)-methyleno)triphosphate (ApCpp) from Jena Biosciences.

The sequences of oligonucleotides are listed in Supplementary Table 2.

Design and construction of the recombination library

Genes of B-family DNA polymerases (*T. gorgonarius* (Tgo; GenBank: KP682507), *Thermococcus* sp. 9°N (9°N; KP682506), *Pyrococcus* species GB-D DV (KP682509), *T. kodakarensis* (Kod; KP682508)), designed with 3'–5' exonuclease-silencing mutations (D141A and E143A) and RNA synthesis mutations (V93Q, Y409G, A485L and E664K), had 11 regions of high homology (~6–12 amino acids) that were identified and codon-swapped to ensure perfect homology for DNA shuffling. The designed Tgo-QGLK, 9°N-QGLK, DV-QGLK and Kod-QGLK genes were ordered as gBlocks.

Q5 PCR amplification (which, unless stated otherwise, is a 50- μ l reaction, 1 $\times$ Q5 buffer, 1 μ M forward primer, 1 μ M reverse primer, 0.4 mM dNTPs and 1 μ l Q5 DNA polymerase (0.02 U μ l⁻¹); 95 °C for 2.5 min, 25 $\times$ (95 °C for 30 s, 68 °C for 45 s, 72 °C for 2 min), 72 °C for 1 min) was used to generate pGDR11 vector backbone from 10 ng pGDR11–Kod–WT plasmid. The PCR reaction mix was cleaned up (30 μ l) and treated with DpnI according to the manufacturer's protocols. The amplified vector backbone was purified from an agarose gel and quantified by UV absorbance.

Each of the four parent genes were fragmented into 12 DNA fragments (totaling 48) by Q5 PCR amplification using appropriate primers for internal regions of each gene. Each PCR reaction was mixed with loading buffer and purified on a 2% agarose gel. Correctly sized amplicons were excised, dissolved and column purified, then eluted in 10 μl of nuclease-free water and quantified by UV. The 48 gene fragments were DNA shuffled by overlap PCR. Each fragment was diluted to 10 ng μl^{-1} stocks, and a 0.5 μl aliquot of each was combined for Q5 amplification (100 μl reaction; 1 $\times$ Q5 buffer, 0.4 mM dNTPs, 2 μl Q5 DNA polymerase). Cycling conditions were as follows: nine cycles of 94 $^{\circ}\text{C}$ for 15 s and 72 $^{\circ}\text{C}$ for 65 s (-0.5°C per cycle) followed by five cycles of 94 $^{\circ}\text{C}$ for 15 s, 67.5 $^{\circ}\text{C}$ for 15 s (-0.5°C per cycle) and 72 $^{\circ}\text{C}$ for 75 s. The full-length recombined genes were amplified by taking 4 μl of the overlap PCR reaction to seed a new Q5 PCR amplification using recomb fragment 1 forward and recomb fragment 12 reverse primers (20 $\times$ (95 $^{\circ}\text{C}$ 30 s, 69 $^{\circ}\text{C}$ 45 s, 72 $^{\circ}\text{C}$ 4 min)). A reaction aliquot (2 μl) was visualized on a 1% agarose gel before the reaction (50 μl) was mixed with loading buffer (10 μl) and purified on a 2% gel, excised, dissolved, column purified, then eluted in 20 μl of nuclease-free water.

A total of 300 ng of the generated library inserted and 100 ng of the vector backbone were added to 10- μl Gibson assembly (total volume of 20 μl), incubated 50 $^{\circ}\text{C}$ for 1 h, then 4 μl were heat-shock transformed into 50- μl NEB 5- α cells (42 $^{\circ}\text{C}$ for 30 s), recovered with SOC at 37 $^{\circ}\text{C}$ for 1 h at 300 rpm, and streaked on carbenicillin (50 $\mu\text{g ml}^{-1}$) agar plates and incubated overnight at 37 $^{\circ}\text{C}$.

Individual colonies were picked and added to a colony PCR reaction containing 1 $\times$ ThermoPol, 0.4 mM dNTPs, Taq DNA polymerase (0.03 U μl^{-1}), 1 μM recomb fragment 1 forward primer and 1 μM recomb fragment 12 reverse primer. Cycling conditions were as follows: 95 $^{\circ}\text{C}$ for 2 min, followed by 30 cycles of 95 $^{\circ}\text{C}$ for 15 s, 60 $^{\circ}\text{C}$ for 15 s and 72 $^{\circ}\text{C}$ for 2 min, with

a final extension at 72 °C for 1 min. An aliquot (2 µl) were visualized on 1% agarose gel to verify the amplicon size. Full plasmid sequencing verified the sequence of the recombined genes. The Gibson assembled material was transformed into NEB 5- α cells and colonies were scrapped and plasmid library was purified, UV-quantified and stored for future transformations.

Encapsulating single *E. coli* in w/o droplets

A total of 300 ng of the plasmid library were transformed into 20 µl of XL10-gold cells and SOC-recovered cells were used to inoculate 50 ml of Luria–Bertani (LB)-carbenicillin (50 µg ml⁻¹) at 37 °C, 250 rpm. The overnight culture was used to inoculate a fresh 50 ml of LB-carbenicillin (500 µl, 1:100 vol/vol) at 37 °C, 250 rpm. At an OD600 of ~0.6, the culture was induced with 1 mM IPTG at 25 °C and incubated overnight, 250 rpm. After expression, 1 ml of culture was collected and centrifuged for at 3,220g, 5 min. The cells were washed four times with 1 ml 1× ThermoPol and centrifuged after every wash. The pellet was resuspended in 2 ml of 1× ThermoPol and diluted to OD600 of 0.05 to enable encapsulation at occupancies of 0.1 cells per droplet according to a Poisson distribution. Cells were mixed with PAA reagents—1 µM of a 5'-Cy3-labeled self-priming hairpin template (30merHP.V2), 2-µM 3'-labeled Iowa black quencher (QP08.Iowa), 100 µM of RNA triphosphates (NTPs), 1× ThermoPol.

Aqueous and oil solutions were sealed in individual 1.5-ml microcentrifuge tubes and controlled through pressure-driven flow with custom LabVIEW software (National Instruments). Two lengths of Tygon tubing were inserted through holes drilled into the caps of the tubes and glued 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 an SMC ITV0011-2UMS digital pressure regulator. Another length of tubing was submerged in the reagent with the other end connected to the

appropriate inlet of the microfluidic device. By manually applying a positive pressure head to the reagent vial through 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 microcentrifuge tube for droplet collection. All emulsions were produced using custom polydimethylsiloxane chips⁶⁶. Single emulsions were formed using a flow focusing geometry. The aqueous phase containing the PAA and *E. coli* was sheared by a continuous phase consisting of a low-viscosity fluorinated oil (HFE-7500) containing 1% (wt/wt) Pico-Surf. Pressures were maintained to achieve droplet diameters of $\sim 20\text{--}22\text{ }\mu\text{m}$ and production rates of 30–35 kHz, allowing $110\text{--}125 \times 10^6$ droplets to be produced every hour. Single emulsions were collected under a layer of mineral oil in 1.5-ml tubes and incubated for 5 min at 95 °C to lyse cells, followed by incubation at 55 °C for various times.

FADS sorting of w/o droplets

After incubation, droplets were injected into an FADS66. Incident light from a 552-nm laser was focused through a $\times 20$ plan-apochromatic objective, through which droplets passed in a single file. The emitted light was directed through a 562-nm Quad Band Dichroic into an optical train comprising a series of long-pass dichroics, then to a photomultiplier tube. The sample was illuminated with blue light to avoid spectral overlap with Cy3 and imaged with a high-speed camera at 35,000 frames per s. The digital signals generated by the photomultiplier tube were analyzed by an FPGA controlled with custom LabVIEW software. Droplets 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 of the sorting chip. The resulting nonuniform electric field generated a dielectrophoretic force that polarized and deflected the droplet into a collection channel.

DNA recovery from sorted droplets and library regeneration

Plasmid DNA was recovered from positively sorted droplets by extraction with Pico-Break following the manufacturer's protocol, where samples were vortexed and centrifuged (60 s, 2,000g) and the top DNA-containing aqueous layer was recovered. The bottom layer was extracted a second time with nuclease-free water to improve yields. The combined aqueous layers containing plasmid DNA were concentrated using a Zymo IC column, washed with 500 μ l DNA wash buffer twice and eluted with nuclease-free water (20 μ l) before Q5 amplification using 1 μ M recomb fragment 1 forward and 0.5 μ M recomb fragment 12 reverse primers, 4 μ l of recovered DNA and otherwise aforementioned Q5 amplification conditions. The PCR reaction amplicons were visualized on 1% agarose gels.

Samples were pooled and mixed with 5 $\times$ Zymo DNA binding buffer (that is, 100 μ l PCR reaction with 500 μ l binding buffer) and cleaned using a Zymo IC column and eluted with 20 μ l of nuclease-free water. The purified DNA was then DpnI-treated and gel purified. Gibson assembly using 100 ng vector and 200 ng insert DNA was performed for 1-h incubation and otherwise following the manufacturer's protocol. The reaction mixture was mixed with 3 $\times$ Zymo DNA binding buffer and purified over a Zymo IC column, eluted with 20 μ l of nuclease-free water, and 4 μ l was transformed into cells.

Activity screening from clarified lysate

Plasmids encoding polymerase variants were transformed into XL1 Blue cells and plated onto carbenicillin-agar plates. Single colonies were picked and grown in 4 ml LB-carbenicillin at 37 °C, 225 rpm, overnight. Overnight cultures (40 μ l) were used to inoculate fresh 4-ml cultures grown to an OD₆₀₀ ~ 0.6, before 1 mM IPTG induction and expression at 25 °C, 225 rpm for 18–

20 h. Cells were collected by centrifugation (1,500g, 25 °C, 10 min) and resuspended in 10 mM Tris-HCl, pH 8.0, 500 mM NaCl and 10% glycerol and transferred to 1.5 ml centrifuge tubes. Tubes were heated at 70 °C for 1 h, cooled for 1 h on ice and centrifuged (16,300g, 4 °C, 1 h). The clarified supernatant concentration was assessed using the Bradford dye.

Activity assays were performed with 1× ThermoPol, 0.5 μM of EM619 Apt template and IR680 PBS8 primer, 100 μM NTPs and 1 μl of lysate in a final volume of 20 μl. Primer-template complexes were annealed in 1× ThermoPol at 90 °C for 5 min and immediately placed on ice for 5 min. NTPs and lysate were added and reactions were incubated at 55 °C. At designated time points, 1 μl of the reaction mixture was added to 39 μl of quenching buffer (95% formamide, 25 mM EDTA) and denatured at 95 °C for 10 min. A total of 10 μl of sample was loaded onto a 15% urea denaturing PAGE gel, run at 10 W for 1.5 h and visualized using a Li-Cor Odyssey CLx imager with Image Studio Lite (v.5.2).

C28 reversion analysis

Reversion constructs were generated by Gibson assembly using a PCR-amplified region of the Tgo-QGLK gene (reversion construct 1, recomb fragment 1 forward and 4 reverse; reversion construct 2, recomb fragment 6 forward and 7 reverse; reversion construct 3, recomb fragment 10 forward and 12 reverse) with PCR-amplified C28 plasmid (reversion construct 1, recomb fragment 5 forward and pGDR11 vector reverse; reversion construct 2, recomb fragment 8 forward and 5 reverse; reversion construct 3, pGDR11 vector forward and recomb fragment 9 reverse). The resulting constructs were sequence verified, expressed in XL10-gold E. coli and purified.

Primer-extension assays were performed in 20 μ l reaction volumes containing 0.5 μ M/0.8 μ M of the pre-annealed EM619/IR680 PBS8 duplex, 0.8 μ M polymerase, 25–100 μ M dNTPs and 1 $\times$ ThermoPol buffer at 55 $^{\circ}$ C for 15, 30 or 60 s of incubation.

C28 expression and purification for biochemical analysis

Protocols for the expression and purification of thermostable recombinant polymerases were previously described⁶⁷. DH5 α or XL10-gold competent *E. coli* cells were transformed with a pGDR11 vector encoding the recombinant polymerase, subsequently grown and expressed in LB media supplemented with carbenicillin. Cells were centrifuged, resuspended in lysis/equilibration buffer (50 mM NaCl, 10 mM Tris-HCl, pH 7.5, 10% glycerol), heat shocked (70 $^{\circ}$ C, 1 h), precipitated on ice (>30 min) and centrifuged to pellet the cell debris. Supernatant was collected and treated with 10% (vol/vol) polyethyleneimine for 15 min. Lysate was centrifuged to remove nucleic acids precipitated by the polyethyleneimine. Protein was then precipitated with ammonium sulfate overnight. After incubation, the protein pellet was resuspended in lysis/equilibration buffer. Polymerases were purified on a heparin affinity column using a concentration gradient of NaCl starting at 250–1,000 mM (10 mM Tris-HCl, pH 7.5, 10% glycerol). All polymerases were buffer exchanged into 500 mM NaCl storage buffer (10 mM Tris-HCl, pH 7.5, 10% glycerol) and concentrated using a 3K Amicon Ultra-0.5 spin column.

C28 biochemical characterization

Unless stated otherwise, all primer-extension assays were performed as in the cell lysate activity screening. For RNA synthesis, reactions with 1 μ M C28 or Tgo-QGLK were incubated at 55 $^{\circ}$ C for specified times. Optimal C28 activity temperature was assayed across a temperature range of 20 $^{\circ}$ C to 70 $^{\circ}$ C for 10 s. Thermal stability was assessed where six C28 aliquots (50 μ l each,

10 μ M) were heated for 1, 2, 3, 4, 5 or 6 h in storage buffer at 90 °C, cooled to 4 °C, then challenged to a primer-extension reaction. Substrate specificity assays varied across nucleoside triphosphate classes. The template-sequence biases of Tgo-QGLK and C28 were performed with six different DNA libraries (libraries L1–L6) of varying AT:GC composition using the PBS8 20-mer primer for 10 min. C28 was evaluated for its ability to synthesize DNA on templates composed of DNA, RNA and FANA using the 30-mer template series and the IR680 PBS8 short primer for 1 h. C28 was evaluated using different sugar-modified and base-modified RNA analogs after 10 min of incubation.

For mismatch primer-extension controls, 1.5 μ M of pre-annealed P–T duplex (IR680-labeled PBS8 primer with either EM619 or a mismatch variant), 1 $\times$ ThermoPol buffer, 100 μ M each NTP and 1.5 μ M C28 were incubated at 55 °C for 1 h.

Chimeric DNA–RNA PCR

C28 was evaluated for the ability to amplify mixed DNA and RNA backbone by PCR (94 °C 15 s, 20 $\times$ (95 °C 15 s, 58 °C 15 s, 68 °C 30 s), 72 °C 1 min). An RNA apt template (20 nM) was amplified in 1 $\times$ ThermoPol, 2 μ M DNA PBS8 primer (forward), 2 μ M DNA PBS7 primer (reverse), 0.5 mM d/rNTPs and 0.5 μ M C28 in a total volume of 50 μ l.

C28 quantitative kinetics

Primer-extension reaction progress was monitored over time, where two reaction replicates per polymerase (15 μ l reaction, 0.2 μ M polymerase, 0.2 μ M pre-annealed PBS8 and EM619 duplex, 2 $\times$ ThermoPol) were preheated for 5 min at 55 °C before initiation with 100 μ M NTP preheated to 55 °C. The reactions were quenched at designated time points by plunging the reaction

tubes into powdered dry ice. Once all time points were frozen, they were thawed on a cold plate (4 °C) by adding 15 µl of 6X LCGreen Plus dye, previously identified as the optimal intercalating dye for tracking the progress of RNA synthesis along a DNA template⁴³. An aliquot (25 µl) of each reaction was transferred to a transparent V-bottom 96-well plate, and fluorescence intensity was measured using a BioTek Synergy Neo2 multimode reader.

For each polymerase, baseline and maximum fluorescence were defined as 0-s and 15-min time points, respectively. Fluorescence was normalized by subtracting the baseline, then dividing by the difference between the maximum and minimum values. Data were converted to nucleotides per polymerase as described⁴³ by multiplying data by the conversion factor: $F = (\text{template}) \times (\text{extension length}) / (\text{polymerase})$. Rates were extracted as the slope of the linear range of each curve. Data were processed using Microsoft Excel.

C28 fidelity

In a 100 µl RNA transcription reaction, ~12 million DNA primer-functionalized hydrogel particles⁴⁵ were resuspended in nuclease-free water, 1× ThermoPol buffer (10 µl, 10×), single-stranded DNA template (2 µl, 100 µM), 1% KF-6012 (1 µl) 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 RNA synthesis, 2 µl of NTPs (5 mM stock) and 20 µl of XNA polymerase (10 µM stock) were added, mixed and then incubated at 55 °C for 1 h. Hydrogels were then washed with breaking buffer (10 mM Tris–HCl, pH 7.5, 100 mM NaCl, 1 mM EDTA, 1% (vol/vol) SDS, 1% (vol/vol) Triton X-100) twice, incubated with 100 µl of 100 mM NaOH (37 °C, 3 min) to strip the template, and neutralized with an equivalent of breaking buffer. Hydrogel particles displaying transcribed RNA oligomers were washed twice with 1× ThermoPol buffer (100 µl).

In a 100 μ l RNA RT reaction, hydrogels were resuspended in nuclease-free water, 1 $\times$ ThermoPol buffer (10 μ l, 10 $\times$), PBS7. PBS9 overhang primer (2 μ l, 100 μ M) containing a double nucleotide mismatch (TT–TT), 0.01% KF-6012 (1 μ l, 1%), 3 mM magnesium sulfate (3 μ l, 100 mM) and then heated at 95 $^{\circ}$ C for 3 min. The reaction mixture was then snap-cooled on ice for 5 min to promote primer/template annealing. To initiate the RT reaction, 10 μ l of dNTPs (5 mM stock) and 20 μ l of Bst-LF (10 μ M stock) were added, mixed and then incubated at 50 $^{\circ}$ C for 4–8 h.

To collect and purify cDNA, hydrogels were pulled down on a magnetic stand, supernatant removed and incubated with 100 μ l of 100-mM NaOH (37 $^{\circ}$ C, 3 min). Hydrogels were pulled down again (1 min), supernatant collected in a 1.5-ml centrifuge tube and neutralized with an equivalent of 3 M NaOAc (pH 6–7). cDNA was precipitated with cold 100% isopropyl alcohol (three equivalents) and incubated on dry ice (>15 min). Precipitated cDNA was centrifuged for 1 h (16,363g), washed with cold 70% ethanol and centrifuged for 20 min (16,363g) twice, and air-dried. The concentration of the resuspended pellet (10- μ l nuclease-free water) was measured using a nanodrop.

The cDNA was PCR-amplified with Taq DNA polymerase, using a PBS8/PBS9 primer pair. PCR material was ligated into a TOPO vector using a TOPO-TA kit and cloned into DH5 α competent cells. Individual clones were picked, grown in liquid LB media and sequenced (Primordium Labs). DNA sequences were aligned and analyzed with the template using MEGA11 and CLC Main software. Error rates for each base substitution were determined as $\mu_{exp \rightarrow obs} = \left(\frac{n_{obs}}{n_{exp}} \right) \times 1,000$. The total error rate was determined by summing the error rate for each substitution.

For C28 primer extension under fidelity conditions, reactions were performed containing 0.5 μ M/0.8 μ M of pre-annealed EM619/IR680 PBS8 duplex, 0.8 μ M polymerase, 100 μ M NTPs and 1 $\times$ ThermoPol buffer at 55 °C for 60 min.

For synthesis of an authentic standard control for size, 1 μ M of 5' IR680-labeled DNA primer was ligated to 1 μ M of a 5' phosphorylated RNA oligonucleotide through splinted ligation (IR680 PBS8, 5' Phos/cEM619_RNA, EM619) with T4 DNA ligase in 1 $\times$ T4 ligation buffer at 23 °C for 2 h.

C28 expression and purification for crystallography

The c28 gene and pET21b vector were PCR-amplified using respective primer sets and cloned using Gibson assembly, and the resulting construct was sequence verified (Plasmidsaurus). BL21(DE3) cells harboring pET21b-c28 were grown aerobically at 37 °C in LB media supplemented with 100 μ g ml⁻¹ carbenicillin. At an OD₆₀₀ of 0.8, C28 expression was induced with 1 mM isopropyl β -D-thiogalactoside at 18 °C for 20 h. Cells were collected by centrifugation (3,315g, 20 min, 4 °C), then resuspended in a low salt buffer (10 mM Tris-HCl, pH 7.0, 0.1 M NaCl, 0.1 mM EDTA, 1 mM DTT, 10% glycerol, 5 mg egg hen lysozyme) before lysis by sonication. The lysate was centrifuged (23,703g, 30 min, 4 °C), heat treated (70 °C, 1 h) to precipitate endogenous E. coli proteins and centrifuged again (23,703g, 30 min, 4 °C). The supernatant was filtered via a 0.2 μ m filter, loaded onto a 5-ml HiTrap Q HP column in tandem with a 5-ml heparin HP column and washed with low salt buffer. The Q HP column was removed before elution with a linear gradient reaching high-salt buffer (10 mM Tris-HCl, pH 7.0, 1 M NaCl, 0.1 mM EDTA, 1 mM DTT, 10% glycerol). Fractions containing C28 were collected and concentrated using a 30-kDa Amicon centrifugal filter before running through a HiLoad 16/600

Superdex 200-pg gel filtration column with medium salt buffer (50 mM Tris-HCl, pH 7.5, 200 mM NaCl, 0.1 mM EDTA, 1 mM DTT). Purified C28 was concentrated to ~25 mg ml⁻¹.

C28 crystallization, structure determination and analysis

Crystallography samples were prepared by incubating 6 mg ml⁻¹ purified C28 with 1.2 molar equivalents of a pre-annealed DNA P-T duplex in buffer (0.5 M Tris-HCl, pH 7.4, 2 M NaCl, 20 mM MgCl₂) at 37 °C for 30 min to yield an initial binary complex. A ternary complex was then prepared by sequentially adding five molar equivalents of UTP and five molar equivalents of nonhydrolyzable ApCpp to the binary complex with incubation at 37 °C for 30 min.

The resulting sample was screened against 864 commercial crystallization conditions in a hanging-drop format using the Mosquito crystallization robot (SPT LabTech). A crystal grown from Morpheus II G10 (100-mM amino acids II, 0.1 M Buffer System 6 pH 8.5, 32.5% (vol/vol) Precipitant Mix 6) yielded a dataset that was collected at the Advanced Light Source, Lawrence Berkeley National Laboratory. The dataset was indexed, refined and integrated by iMOSFLM⁶⁸ with a high-resolution limit of 2.22 Å, scaled and merged using Pointless and Aimless^{69,70}. Molecular replacement was performed using Phenix-PhaserMR71 using Protein Data Bank (PDB) ID 8T3X as a search model, followed by iterative rounds of building, refinement and validation using Coot, Phenix-Refine and MolProbity^{72,73,74,75}. Crystallographic programs were accessed through the PHENIX Suite⁷⁶. Sufficient electron density enabled the building of residues 1–747 of the protein, 5–17 of the template and 2–12 of the primer. No observable density was found for ApCpp, and, thus, a model of the binary n + 1 complex was built. The dataset structure factors and coordinates were deposited in the Research Collaboratory for Structural Bioinformatics (PDB ID: 9MYE).

Molecular visualizations, RMSD and distance measurements were performed using PyMOL⁷⁷. Electron density $2F_o - F_c$ maps were created in PHENIX as polder maps⁷⁸. Electrostatic potential surface maps were created using APBS and PDB2PQR as PyMOL plug-ins⁷⁹. Local base-pair and base-pair step parameters are computed using web 3DNA⁸⁰. Sugar puckers and torsion angles were computed using PROSIT⁸¹. Polymerase-duplex contact maps were computed using MAPIYA⁸².

4.6 REFERENCES

1. 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).
2. Doublié, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* **391**, 251–258 (1998).
3. 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).
4. Johnson, K. A. Role of induced fit in enzyme specificity: a molecular forward/reverse switch. *J. Biol. Chem.* **283**, 26297–26301 (2008).
5. 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).
6. Berezina, S. Y., Gill, J. P., Lamichhane, R. & Millar, D. P. Single-molecule Förster resonance energy transfer reveals an innate fidelity checkpoint in DNA polymerase I. *J. Am. Chem. Soc.* **134**, 11261–11268 (2012).
7. Traut, T. W. Physiological concentrations of purines and pyrimidines. *Mol. Cell. Biochem.* **140**, 1–22 (1994).
8. Kool, E. T. Active site tightness and substrate fit in DNA replication. *Annu. Rev. Biochem.* **71**, 191–219 (2002).
9. Loakes, D. & Holliger, P. Polymerase engineering: towards the encoded synthesis of unnatural polymers. *Chem. Commun. (Camb.)* **27**, 4619–4631 (2009).
10. Chen, T. & Romesberg, F. E. Directed polymerase evolution. *FEBS Lett.* **588**, 219–229 (2014).
11. Nikoobazar, A., Chim, N., Yik, E. J. & Chaput, J. C. Engineering polymerases for applications in synthetic biology. *Q. Rev. Biophys.* **53**, e8 (2020).
12. Astatke, M., Ng, K., Grindley, N. D. F. & Joyce, C. M. A single side chain prevents *Escherichia coli* DNA polymerase I (Klenow fragment) from incorporating ribonucleotides. *Proc. Natl Acad. Sci. USA* **95**, 3402–3407 (1998).
13. Bonnin, A., Lazaro, J. M., Blanco, L. & Salas, M. A single tyrosine prevents insertion of ribonucleotides in the eukaryotic-type phi 29 DNA polymerase. *J. Mol. Biol.* **290**, 241–251 (1999).
14. Brown, J. A. & Suo, Z. Unlocking the sugar ‘steric gate’ of DNA polymerases. *Biochemistry* **50**, 1135–1142 (2011).

15. Gao, G. X., Orlova, M., Georgiadis, M. M., Hendrickson, W. A. & Goff, S. P. Conferring RNA polymerase activity to a DNA polymerase: a single residue in reverse transcriptase controls substrate selection. *Proc. Natl Acad. Sci. USA* **94**, 407–411 (1997).
16. Staiger, N. & Marx, A. A DNA polymerase with increased activity for ribonucleotides and C5-modified deoxyribonucleotides. *ChemBioChem* **11**, 1963–1966 (2010).
17. Patel, D. H. & Loeb, L. A. Multiple amino acid substitutions allow DNA polymerases to synthesize RNA. *J. Biol. Chem.* **275**, 40266–40272 (2000).
18. Xia, G. et al. 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).
19. Ong, J. L., Loakes, D., Jaroslowski, S., Too, K. & Holliger, P. Directed evolution of DNA polymerase, RNA polymerase and reverse transcriptase activity in a single polypeptide. *J. Mol. Biol.* **361**, 537–550 (2006).
20. Cozens, C., Pinheiro, V. B., Vaisman, A., Woodgate, R. & Holliger, P. A short adaptive path from DNA to RNA polymerases. *Proc. Natl Acad. Sci. USA* **109**, 8067–8072 (2012).
21. Fogg, M. J., Pearl, L. H. & Connolly, B. A. Structural basis for uracil recognition by archaeal family B DNA polymerases. *Nat. Struct. Biol.* **9**, 922–927 (2002).
22. Gardner, A. F. et al. Terminator DNA polymerase: modified nucleotides and unnatural substrates. *Front. Mol. Biosci.* **6**, 28 (2019).
23. Pinheiro, V. B. et al. Synthetic genetic polymers capable of heredity and evolution. *Science* **336**, 341–344 (2012).
24. Mehta, A. P. et al. Bacterial genome containing chimeric DNA-RNA sequences. *J. Am. Chem. Soc.* **140**, 11464–11473 (2018).
25. Brunderova, M. et al. Expedient production of site specifically nucleobase-labelled or hypermodified RNA with engineered thermophilic DNA polymerases. *Nat. Commun.* **15**, 3054 (2024).
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. Gardner, A. F. & Jack, W. E. Determinants of nucleotide sugar recognition in an archaeon DNA polymerase. *Nucleic Acids Res.* **27**, 2545–2553 (1999).
28. Maola, V. A. et al. Directed evolution of a highly efficient TNA polymerase achieved by homologous recombination. *Nat. Cat.* **7**, 1173–1185 (2024).
29. Hogrefe, H. H., Cline, J., Lovejoy, A. E. & Nielson, K. B. DNA polymerases from hyperthermophiles. *Methods Enzymol.* **334**, 91–116 (2001).

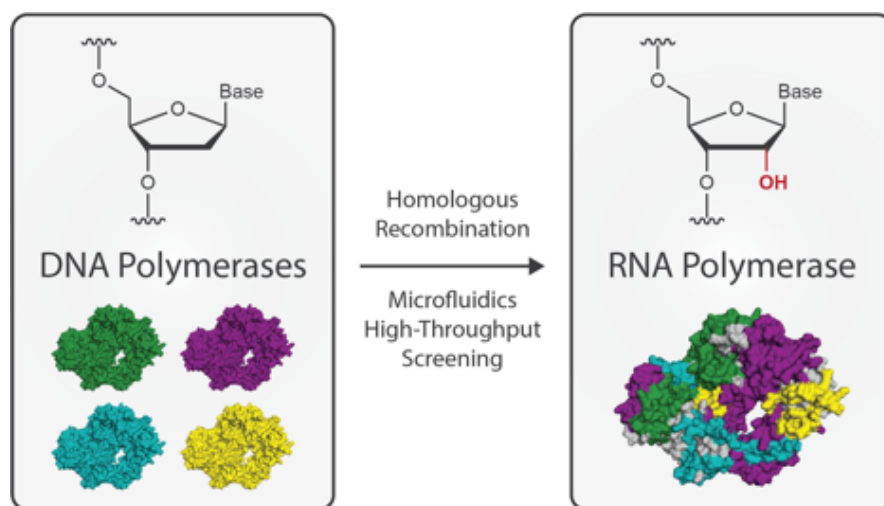
30. Ness, J. E. et al. Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. *Nat. Biotechnol.* **20**, 1251–1255 (2002).
31. Cramer, A., Raillard, S.-A., Bermudez, E. & Stemmer, W. P. C. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**, 288–291 (1998).
32. d’Abbadie, M. et al. Molecular breeding of polymerases for amplification of ancient DNA. *Nat. Biotechnol.* **25**, 939–943 (2007).
33. Baar, C. et al. Molecular breeding of polymerases for resistance to environmental inhibitors. *Nucleic Acids Res.* **39**, e51 (2011).
34. Leconte, A. M. et al. Directed evolution of DNA polymerases for next-generation sequencing. *Angew. Chem. Int. Ed. Engl.* **49**, 5921–5924 (2010).
35. Yik, E. J., Maola, V. A. & Chaput, J. C. Engineering TNA polymerases through iterative cycles of directed evolution. *Methods Enzymol.* **691**, 29–59 (2023).
36. Hottin, A. & Marx, A. Structural insights into the processing of nucleobase-modified nucleotides by DNA polymerases. *Acc. Chem. Res.* **49**, 418–427 (2016).
37. Larsen, A. C. et al. A general strategy for expanding polymerase function by droplet microfluidics. *Nat. Commun.* **7**, 11235 (2016).
38. Vallejo, D., Nikoobazar, A., Paegel, B. M. & Chaput, J. C. Fluorescence-activated droplet sorting for single-cell directed evolution. *ACS Synth. Biol.* **8**, 1430–1440 (2019).
39. Chaput, J. C. & Herdewijn, P. What Is XNA?. *Angew. Chem. Int. Ed. Engl.* **58**, 11570–11572 (2019).
40. Wang, Y., Ngor, A. K., Nikoobazar, A. & Chaput, J. C. Evolution of a general RNA-cleaving FANA enzyme. *Nat. Commun.* **9**, 5067 (2018).
41. 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).
42. Ross, M. G. et al. Characterizing and measuring bias in sequence data. *Genome Biol.* **14**, R51 (2013).
43. Nikoobazar, A., Dunn, M. R. & Chaput, J. C. Evaluating the rate and substrate specificity of laboratory evolved XNA polymerases. *Anal. Chem.* **89**, 12622–12625 (2017).
44. Boosalis, M. S., Petruska, J. & Goodman, M. F. DNA polymerase insertion fidelity. Gel assay for site-specific kinetics. *J. Biol. Chem.* **262**, 14689–14696 (1987).
45. Medina, E. & Chaput, J. C. Measuring XNA polymerase fidelity in a hydrogel particle format. *Nucleic Acids Res.* **53**, gkaf038 (2025).

46. Jackson, L. N., Chim, N., Shi, C. & Chaput, J. C. Crystal structures of a natural DNA polymerase that functions as an XNA reverse transcriptase. *Nucleic Acids Res.* **47**, 6973–6983 (2019).
47. Chen, T. & Romesberg, F. E. Polymerase chain transcription: exponential synthesis of RNA and modified RNA. *J. Am. Chem. Soc.* **139**, 9949–9954 (2017).
48. Egli, M. & Manoharan, M. Chemistry, structure and function of approved oligonucleotide therapeutics. *Nucleic Acids Res.* **51**, 2529–2573 (2023).
49. 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).
50. 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).
51. Wing, R. et al. Crystal structure analysis of a complete turn of B-DNA. *Nature* **287**, 755–758 (1980).
52. Zhou, H. X. & Pang, X. Electrostatic interactions in protein structure, folding, binding, and condensation. *Chem. Rev.* **118**, 1691–1741 (2018).
53. Nick McElhinny, S. A. et al. Genome instability due to ribonucleotide incorporation into DNA. *Nat. Chem. Biol.* **6**, 774–781 (2010).
54. Minshull, J. & Stemmer, W. P. Protein evolution by molecular breeding. *Curr. Opin. Chem. Biol.* **3**, 284–290 (1999).
55. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10**, 866–876 (2009).
56. Agresti, J. J. et al. Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc. Natl Acad. Sci. USA* **107**, 4004–4009 (2010).
57. Obexer, R., Nassir, M., Moody, E. R., Baran, P. S. & Lovelock, S. L. Modern approaches to therapeutic oligonucleotide manufacturing. *Science* **384**, eadl4015 (2024).
58. Andrews, B. I. et al. Sustainability challenges and opportunities in oligonucleotide manufacturing. *J. Org. Chem.* **86**, 49–61 (2021).
59. Flamme, M., McKenzie, L. K., Sarac, I. & Hollenstein, M. Chemical methods for the modification of RNA. *Methods* **161**, 64–82 (2019).
60. Liu, Y. et al. Synthesis and applications of RNAs with position-selective labelling and mosaic composition. *Nature* **522**, 368–372 (2015).

61. Chen, D., Han, Z., Liang, X. & Liu, Y. Engineering a DNA polymerase for modifying large RNA at specific positions. *Nat. Chem.* **17**, 382–392 (2025).
62. Haslecker, R. et al. Extending the toolbox for RNA biology with SegModTeX: a polymerase-driven method for site-specific and segmental labeling of RNA. *Nat. Commun.* **14**, 8422 (2023).
63. Moody, E. R., Obexer, R., Nickl, F., Spiess, R. & Lovelock, S. L. An enzyme cascade enables production of therapeutic oligonucleotides in a single operation. *Science* **380**, 1150–1154 (2023).
64. Kariko, K., Muramatsu, H., Ludwig, J. & Weissman, D. Generating the optimal mRNA for therapy: HPLC purification eliminates immune activation and improves translation of nucleoside-modified, protein-encoding mRNA. *Nucleic Acids Res.* **39**, e142 (2011).
65. 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).
66. Vallejo, D., Nikoomanzar, A. & Chaput, J. C. Directed evolution of custom polymerases using droplet microfluidics. *Methods Enzymol.* **644**, 227–253 (2020).
67. Nikoomanzar, A., Dunn, M. R. & Chaput, J. C. Engineered polymerases with altered substrate specificity: expression and purification. *Curr. Protoc. Nucleic Acid Chem.* **69**, 4.75.1–4.75.20 (2017).
68. Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 271–281 (2011).
69. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution?. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
70. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006).
71. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
72. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
73. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).
74. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D Struct. Biol.* **74**, 531–544 (2018).

75. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
76. 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).
77. DeLano, W. L. The PYMOL molecular graphics system. api.semanticscholar.org/CorpusID:60136037 (2002).
78. Liebschner, D. et al. Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr. D Struct. Biol.* **73**, 148–157 (2017).
79. Jurrus, E. et al. Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **27**, 112–128 (2018).
80. 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).
81. Sun, G., Voigt, J. H., Filippov, I. V., Marquez, V. E. & Nicklaus, M. C. PROSIT: pseudo-rotational online service and interactive tool, applied to a conformational survey of nucleosides and nucleotides. *J. Chem. Inf. Comput. Sci.* **44**, 1752–1762 (2004).
82. Badaczewska-Dawid, A. E., Nithin, C., Wroblewski, K., Kurcinski, M. & Kmiecik, S. MAPIYA contact map server for identification and visualization of molecular interactions in proteins and biological complexes. *Nucleic Acids Res.* **50**, W474–W482 (2022).

4.7 FIGURES



TOC Figure

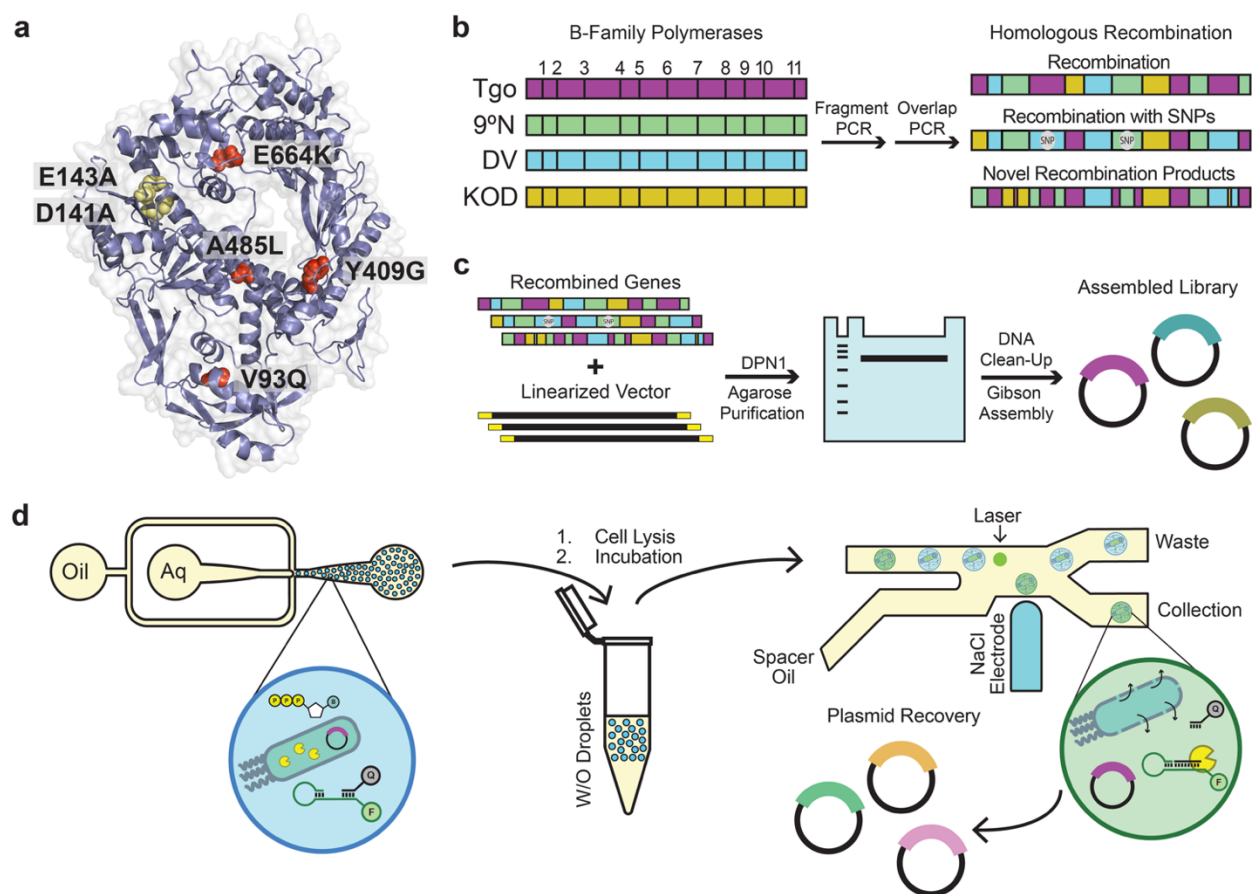


Figure 4.1 Library generation and selection strategy

a, Crystal structure of Tgo DNA polymerase (PDB ID: 1TGO) showing the locations of the QGLK mutations (red) and exonuclease-silencing mutations (yellow). b, Library design using homologous recombination to shuffle the sequences of B-family archaeal DNA polymerases (Tgo, 9°N, Kod and DV) containing the QGLK and exonuclease-silencing mutations. DNA shuffling was performed by PCR. c, Library members were inserted into the expression plasmids by Gibson assembly. d, Schematic view of DrOPS. This high-throughput screening technique uses microfluidic devices to encapsulate single *E. coli* cells expressing library members in w/o droplets and to sort droplets containing active members based on laser-induced fluorescence from an optical sensor. Aq denotes aqueous reagents used for droplet production (NTPs, buffer, sensor and *E. coli*). DrOPS, droplet-based optical polymerase sorting; SNPs, single-nucleotide mutations; w/o, water-in-oil.

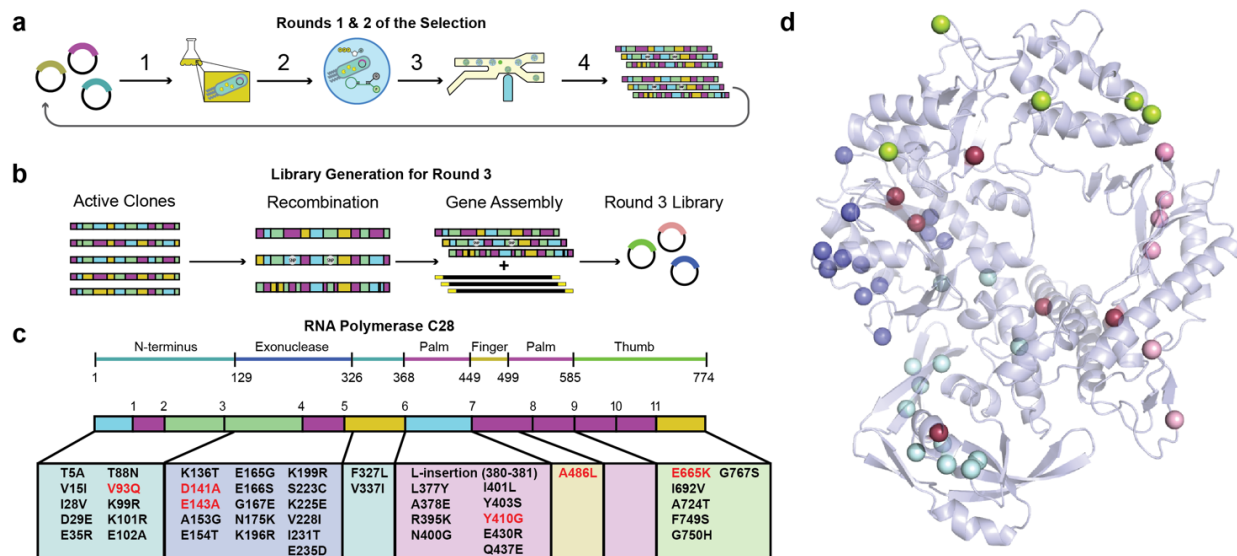


Figure 4.2 Evolution of a DNA-dependent RNA polymerase

a, Schematic representation of DrOPS selection for rounds 1 and 2. The library was (1) transformed into competent *E. coli* cells, (2) encapsulated into w/o droplets, (3) incubated and sorted based on activity and (4) active clones were recovered and amplified. b, Recombination of high-activity clones. The five most active clones identified from round 2 were recombined to create a new pool of polymerase variants. c, Schematic illustration of the sequence composition of clone C28, the most active variant from round 3, showing the selected mutations in black and exonuclease inactivation and QGLK mutations in red. d, Crystal structure of wild-type Tgo DNA polymerase (PDB ID: 1TGO) showing the location of the amino acid mutations observed in C28. The selected mutations appear as spheres, colored by domain—N terminus (cyan), exonuclease domain (purple), palm (pink) and thumb (green). Exonuclease inactivation and QGLK mutations appear in red.

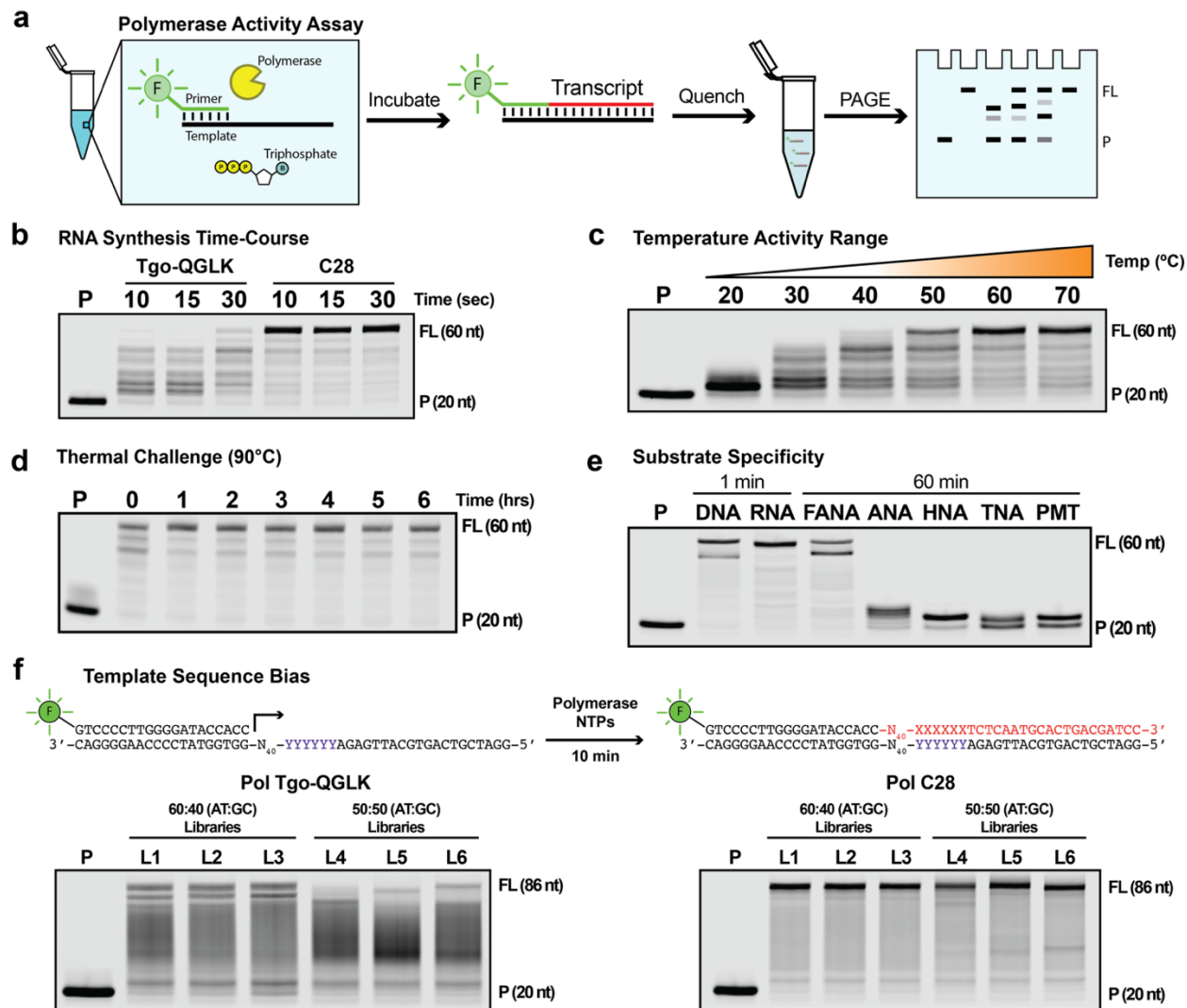


Figure 4.3 Functional analysis of RNA polymerase C28

a, Schematic illustration of the PAA. Extension of a labeled DNA primer annealed to a DNA template. b, Short time-course analysis of RNA synthesis comparing C28 to Tgo-QGLK over 30 s. c, Temperature optimization of C28 in a 10-s RNA synthesis reaction. d, Thermal stability of C28. RNA synthesis was evaluated in a 1-min RNA synthesis reaction after incubating the polymerase for 1–6 h at 90 °C. e, Substrate specificity. C28 was challenged to synthesize DNA, RNA and FANA in a 1-min reaction and ANA, HNA, TNA and PMT in a 1-h reaction. f, Template-sequence bias. DNA libraries with 60:40 (L1–L3) or 50:50 (L4–L6) AT:GC compositions were copied into RNA using either Tgo-QGLK or C28 in a 10-min reaction. PAAs were performed by incubating IR680 PBS8 primer annealed to an EM619 template (1 μ M) with NTPs (100 μ M each) in 1 $\times$ ThermoPol buffer, incubated at 55 °C unless otherwise stated. Reactions were analyzed by denaturing PAGE and imaged using a Li-COR imager. Representative gels of two independent experiments are shown. P, primer. FL, full length; P, primer.

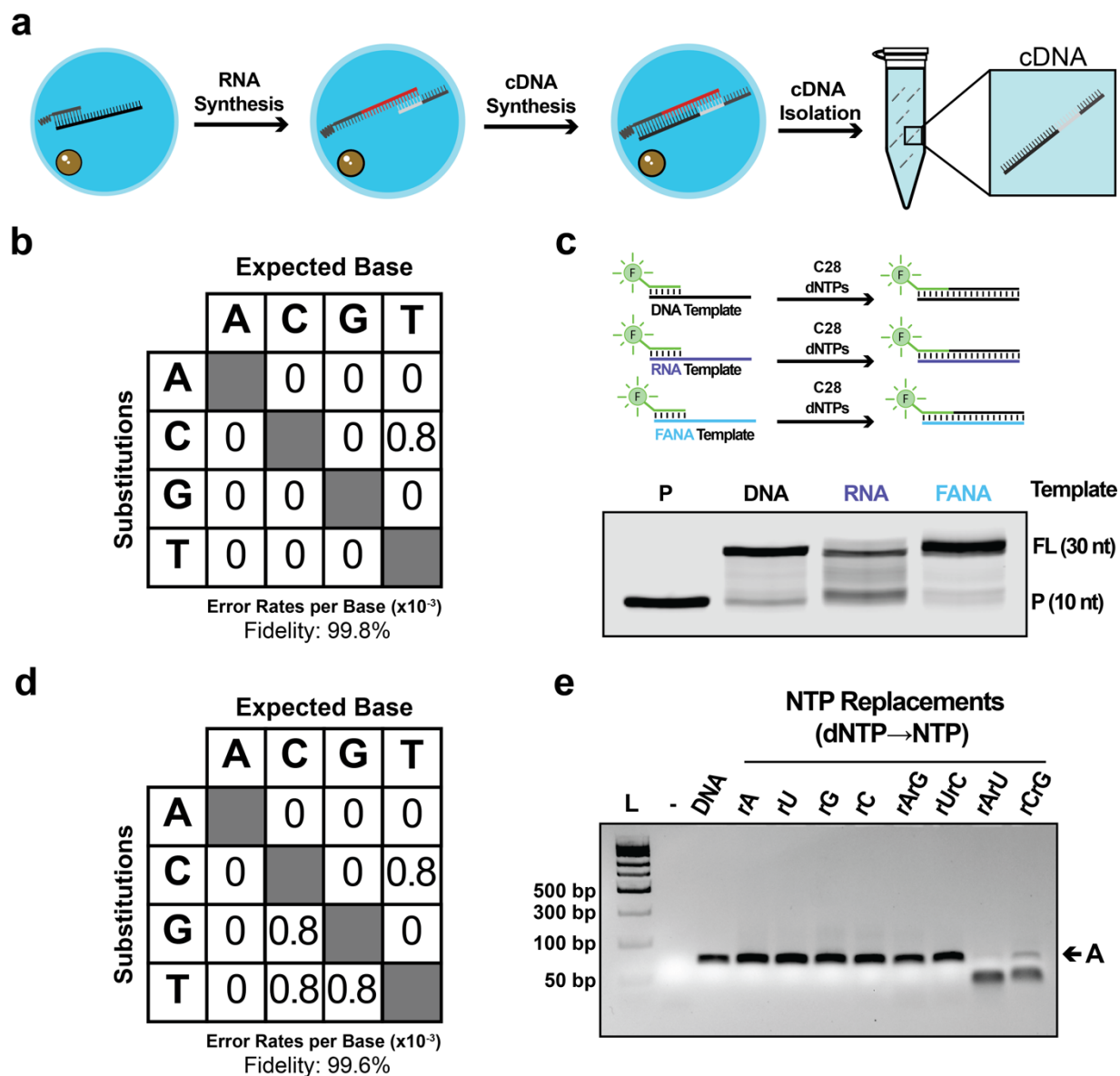
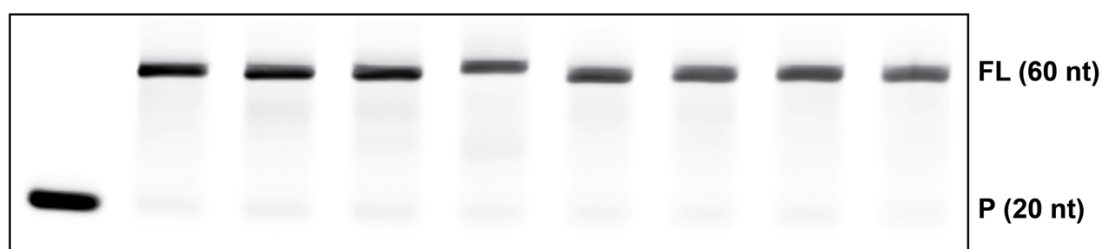


Figure 4.4 Fidelity and template composition requirements of RNA polymerase C28

a, Polymerase fidelity was measured by transcribing and reverse transcribing a DNA template in hydrogel-coated magnetic bead particles. **b**, Mutational profile observed for C28 and Bst DNA polymerase ($n > 1,000$ nucleotide incorporations). **c**, DNA synthesis activity observed on DNA, RNA and FANA templates. **d**, Mutational profile observed using C28 for RNA synthesis and RT ($n > 1,000$ nucleotide incorporations). **e**, PCR amplification of varying dNTP/NTP compositions with 90-s extension times per cycle. The RNA nucleotide substitutions are indicated above each gel lane. Representative gels of two independent experiments are shown. L, ladder; -, negative control; A, amplicon.



P	RNA	1	2	3	4	5	6	7
---	-----	---	---	---	---	---	---	---



a, Chemical structure of RNA compared to common RNA analogs. **b**, Synthesis of chemically modified RNA by primer extension. Reactions were performed for 10 min at 55 °C. UTP analogs (lanes 1–5), CTP analogs (lanes 6–7). Representative gels of two independent experiments are shown.

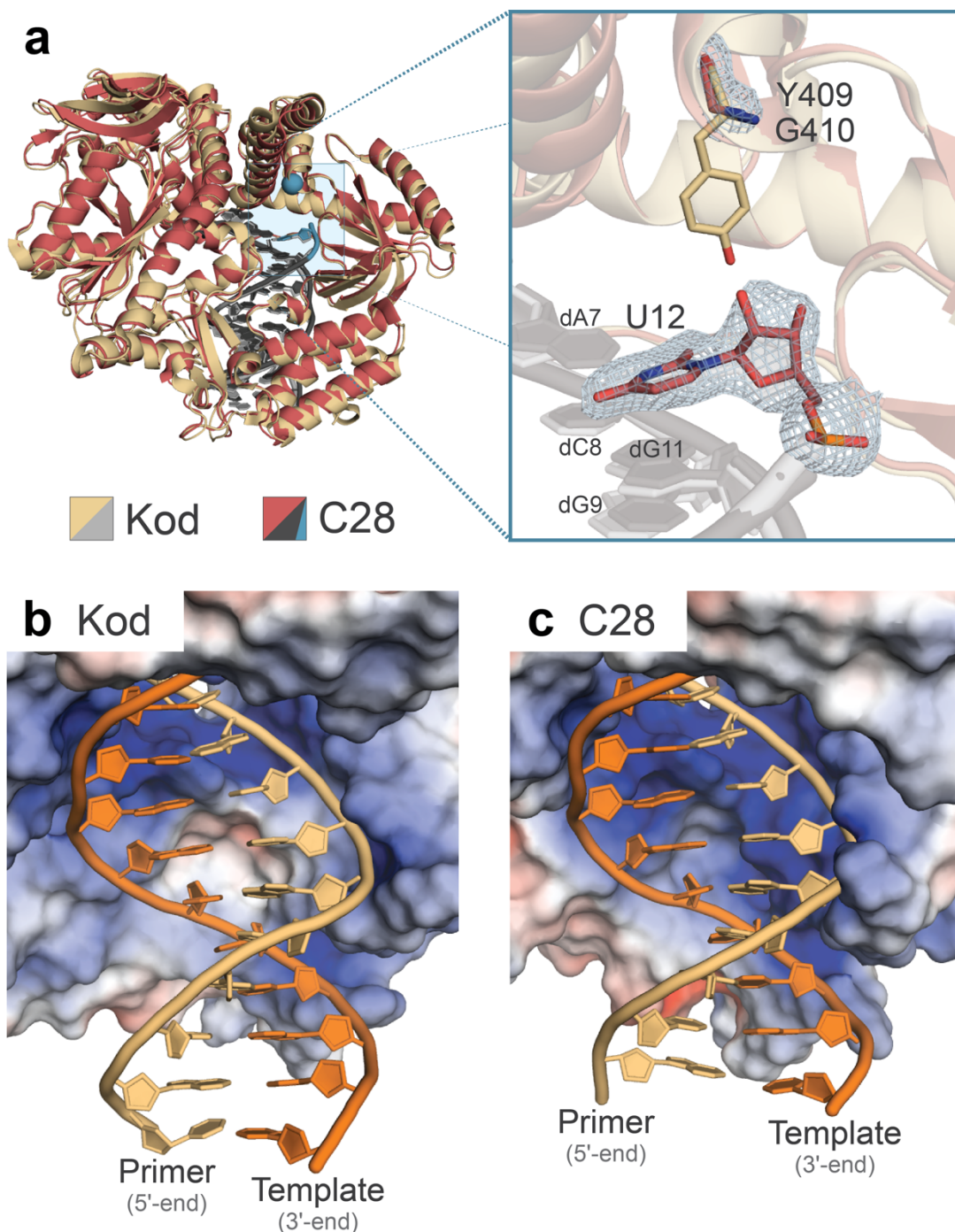
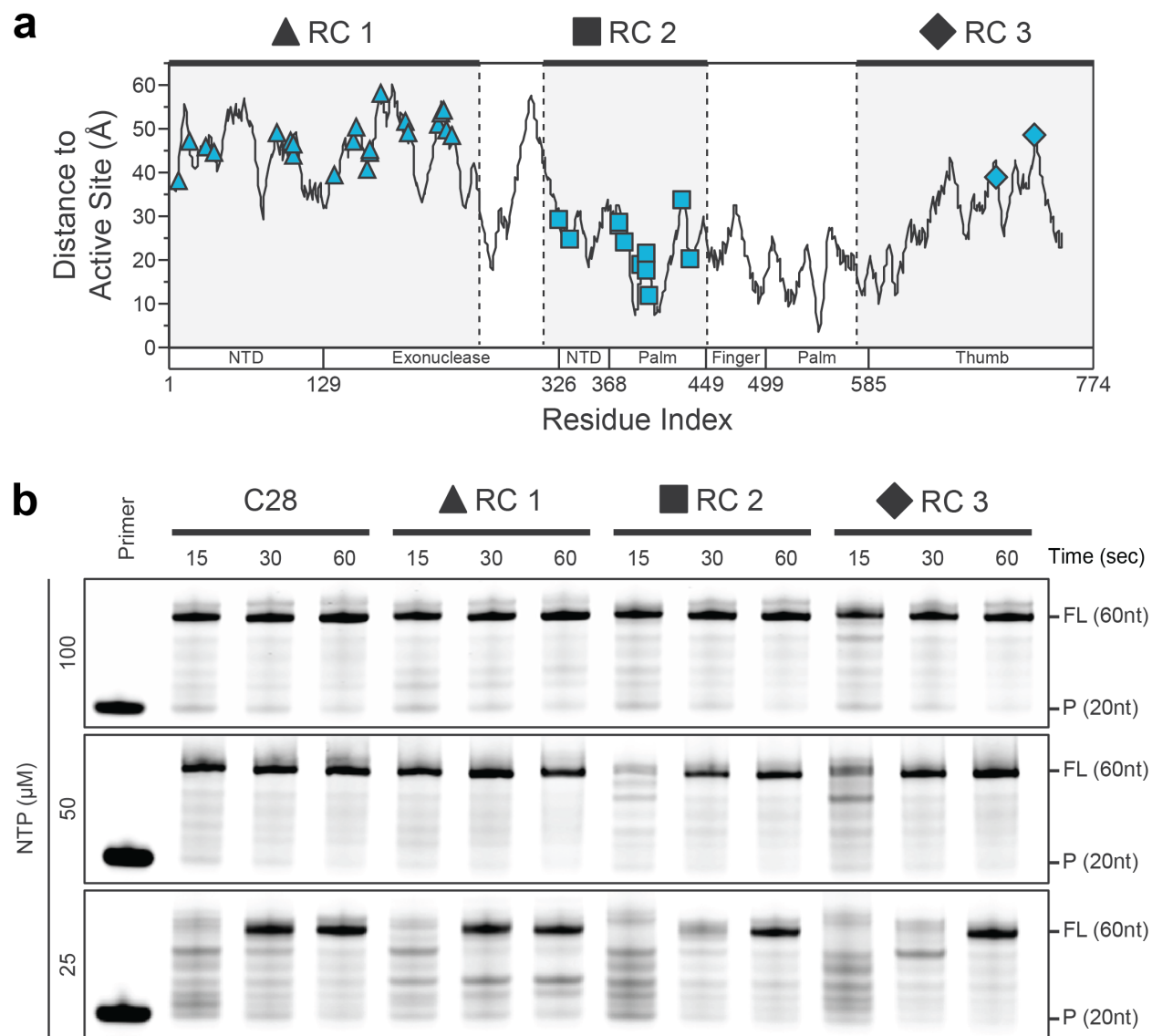


Figure 4.6 Crystallographic analysis of the engineered C28 RNA polymerase

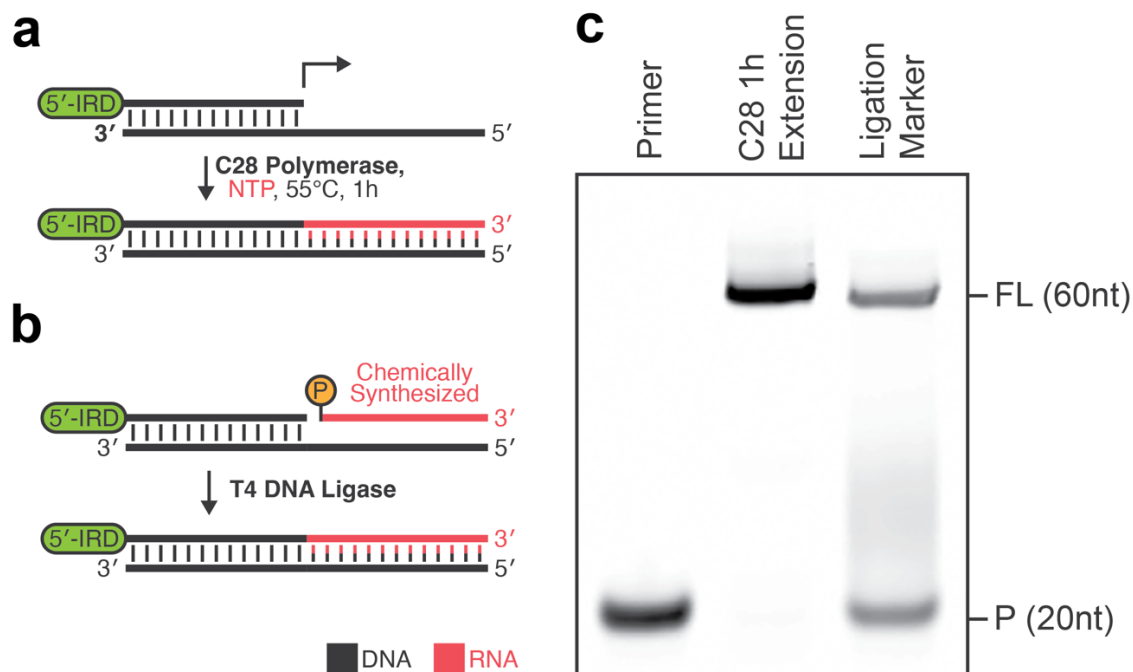
a, Structure of C28 bound to a P–T duplex (red and dark gray) overlaid with natural Kod DNA polymerase (yellow and light gray; PDB ID: 4K8Z). Boxed area highlights the U12 ribonucleotide incorporation and position of the steric gate mutation, both shown in a close-up view of the active site as stick models (Kod Y409 is yellow, C28 G410 and U12 are red). The $2F_o - F_c$ polder maps for U12 and G410 are contoured at 4.5σ with a carve radius of 2. **b,c**, Electrostatic surface maps calculated for the duplex interface of Kod (**b**) and C28 (**c**).

4.8 EXTENDED DATA FIGURES



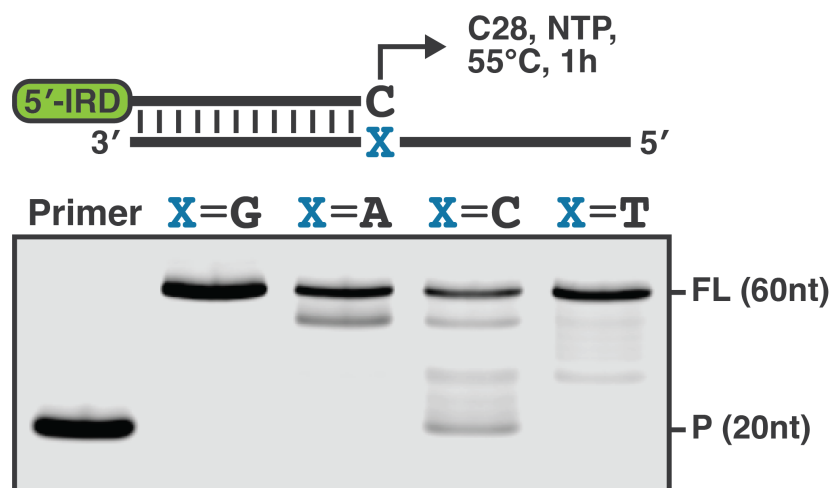
Extended Data Figure 4.1 C28 Reversion Analysis

a, Three segments of C28 harboring acquired mutations were reverted back to the parental Tgo-QGLK to create individual reversion constructs, RC1 (triangle), RC2 (square), and RC3 (diamond). RC3 contains five mutations, three of which are not observed in the crystal structure. **b**, Comparison of C28 activity against RC1, RC2, and RC3. Product formation was evaluated after 15, 30, or 60 seconds of incubation at 55°C in an RNA primer extension assay with decreasing substrate concentrations.



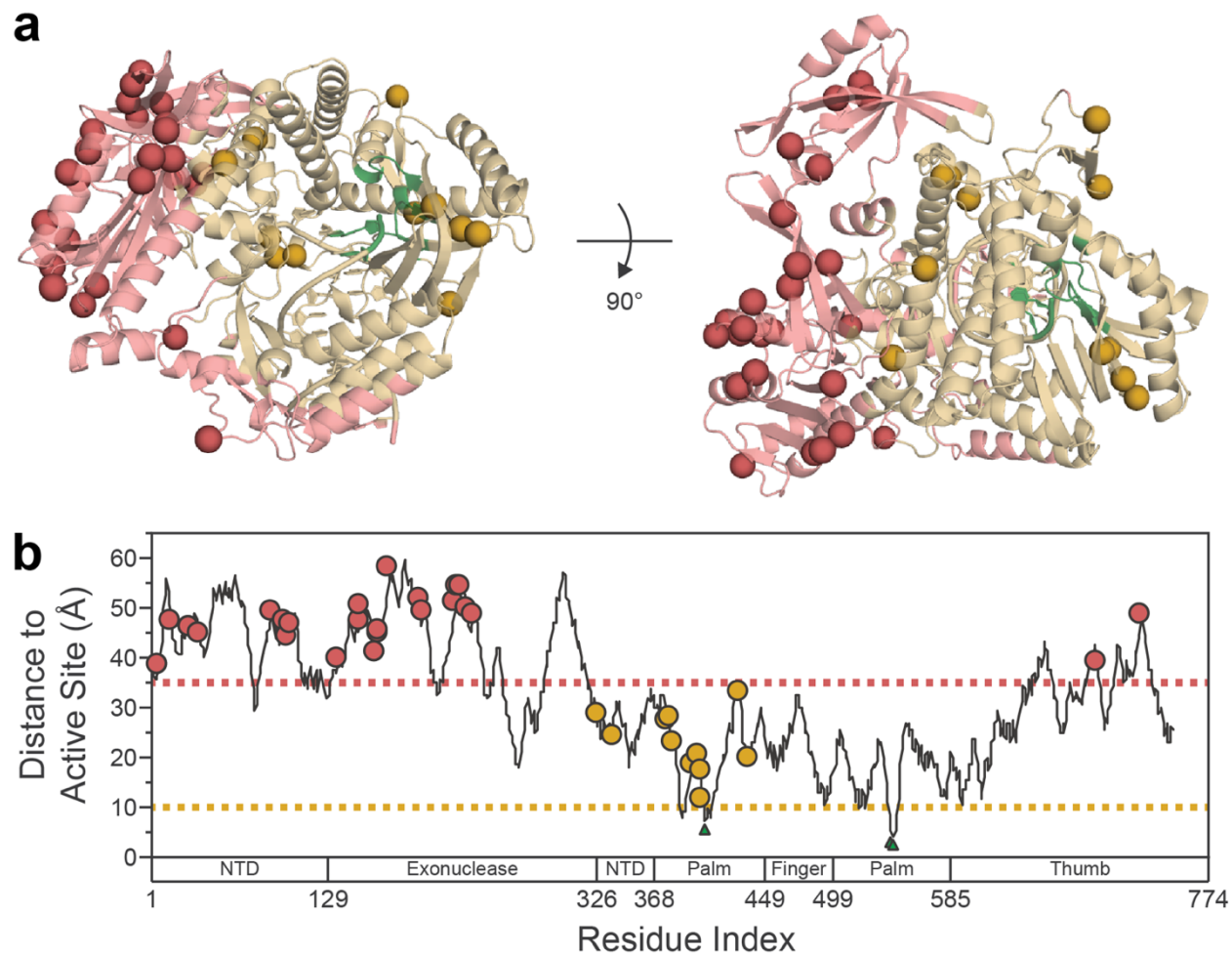
Extended Data Figure 4.2 C28 Activity Under Fidelity Assay Conditions

a, Schematic depicting the standard primer extension assay using a 5' IR680-labeled primer annealed to a DNA template. **b**, Schematic depicting the ligation of a 5' IR680-labeled primer to a chemically synthesized 5' phosphorylated RNA oligonucleotide to generate an authentic standard for the product of the primer extension reaction. **c**, Denaturing PAGE analysis of C28 product formation after 1-hour of incubation at 55°C along with an authentic standard of primer-extended product.



Extended Data Figure 4.3 C28 Extends Primers Despite Template Mismatches

A schematic depicts the 5' IR680-labeled primer terminating in a 3' cytosine, paired with a DNA template in which the complementary position (X) is either a matched guanine or one of the three mismatched bases (adenine, cytosine, or thymine). Comparison of C28 product formation reveals full-length product in all primer-duplex pairs following a 1-hour incubation at 55°C.



Extended Data Figure 4.4 Mapping mutations acquired by homologous recombination

a, The crystal structure of C28 is colored according to the distance to the active site, which is defined as the 3'-hydroxyl of U12. Residues within a 10 Å radius are green, residues between 10 and 35 Å are yellow, and residues outside 35 Å are red. Spheres indicate the alpha carbons of residues in C28 that differ from Tgo-QGLK (exo-). A rotated view provides clarity and perspective. **b**, The distance of each residue to the active site is measured in Å and mapped over the polymerase domains. Spheres are color coordinated to mutations shown in panel a. Three green triangles indicate the catalytic aspartates in the active site. Data excludes three C-terminal mutations not observed in the crystal structure.

CHAPTER 5: Summary and Conclusions

Replicative DNA polymerases are finely tuned molecular machines. Their active sites select the correct nucleotide, position the primer terminus, coordinate catalytic metal ions, and maintain the geometry required for rapid and accurate synthesis. The same constraints that support natural function also limit the acceptance of substrates that differ from DNA. This dissertation examined how those constraints can be altered and what polymerase reprogramming reveals about the catalytic function.

Using phosphonomethylthreosyl nucleic acid (pTNA), threose nucleic acid (TNA), and RNA as distinct substrates, the work progressed from structural accommodation to functional adaptation. The pTNA study showed how an engineered polymerase organizes an orthogonal genetic polymer that does not form a stable duplex with DNA in free solution (1). The TNA studies examined whether a DNA polymerase could achieve near-natural activity on a substantially altered backbone and how that function emerged during directed evolution (2,3). The RNA study then tested whether the same evolutionary framework could be extended to a substrate that closely resembles DNA but is strongly excluded by natural DNA polymerases (4). Together, these studies show that polymerase specificity is stringent but adaptable, and that new activity emerges through coordinated changes within the active site and across the broader protein architecture.

Polymerase Recognition of pTNA

pTNA provided an unusual test of polymerase adaptability. Its phosphonomethyl backbone prevents stable pairing with DNA or RNA in free solution. Nevertheless, engineered polymerases can transfer genetic information from DNA into pTNA. The structures described in Chapter 1 showed that the polymerase enables this process by organizing nascent pTNA residues into

canonical Watson–Crick base pairs with the DNA template. Although pTNA and DNA have similar nominal backbone repeat lengths, polymerase-bound pTNA adopts sugar and backbone conformations more closely resembling TNA. The enzyme therefore does more than catalyze bond formation; it stabilizes the geometry required for an otherwise unfavorable heteroduplex.

Comparison of Kod-RI and Kod-RSGA showed that R606G reorganizes local contacts with pTNA and restores a Y594 interaction resembling one observed during DNA synthesis. Because glycine at this position also supports TNA synthesis, the result suggests a shared role in accommodating threose-containing substrates. Together with the TNA-like conformation of polymerase-bound pTNA, this finding highlights the influence of the threose sugar on substrate geometry (1).

These structures capture short postcatalytic products rather than extended synthesis or the complete catalytic cycle. Structures of longer pTNA products and ternary complexes containing an incoming pTNA substrate would help determine how this geometry is maintained during nucleotide addition and translocation.

Directed Evolution of Efficient TNA Synthesis

Chapter 2 examined whether a DNA polymerase could be evolved to synthesize TNA with near-natural efficiency. Earlier TNA polymerases supported information transfer but remained slow and, in the case of Kod-RI, were frequently evaluated under Mn^{2+} -containing conditions, which can compromise fidelity. Their structures exhibited imperfect base-pair recognition and a suboptimal distance between the TNA primer terminus and the incoming triphosphate (5), suggesting that improved activity would require extensive refinement of the reaction center.

Homologous recombination sampled combinations of residues from B-family polymerases while likely retaining a greater proportion of folded variants than random mutagenesis alone. High-throughput screening by DrOPS, followed by error-prone mutagenesis and further selection, produced 10-92, which synthesizes TNA at approximately 1 nt s^{-1} with greater than 99% fidelity. This represented a substantial improvement over Kod-RI, Kod-RS, and Kod-RSGA, which synthesized TNA at approximately 0.08, 1.9, and 5.7 nt min^{-1} , respectively. At approximately 58 nt min^{-1} , 10-92 brought TNA synthesis into the nucleotide-per-second regime (2).

The structure of 10-92, more thoroughly examined in Chapter 3, showed that this gain in activity involved changes throughout the polymerase rather than a localized alteration of the substrate-binding pocket. Rearrangements in the thumb, palm, and surrounding regions improved base-pair geometry, primer-substrate positioning, catalytic residue alignment, and contacts with the nascent duplex. Efficient TNA synthesis therefore emerged through coordinated changes across the polymerase architecture (2,3).

Mutational Landscape of TNA Polymerase Evolution

Reconstruction of the 10-92 lineage provided an opportunity to examine how acquired mutations collectively produced a new polymerase function. These mutations arose through two distinct diversification mechanisms. Homologous recombination produced broad architectural change, whereas later rounds of error-prone PCR introduced fewer mutations associated with the largest gains in catalytic rate.

The transition from Kod-RSGA to 5-270 introduced 35 substitutions inherited from naturally occurring B-family polymerases. Approximately half were conservative by BLOSUM62, and their predicted effects on stability were generally small and balanced between stabilizing and

destabilizing changes. The substitutions were distributed throughout the enzyme, with no clear relationship to active site distance or solvent exposure. The increase in fidelity could not be intuitively explained by single mutations alone. Instead, the recombined sequence altered multiple structural elements, improved the geometry of the nascent TNA-DNA base pair, and shifted the fingers toward a more preorganized conformation. The fidelity gain therefore appears to reflect distributed architectural remodeling rather than a discrete active site mutation.

This interpretation is consistent with induced-fit models of polymerase specificity, in which conformational changes help determine whether a bound nucleotide proceeds to chemistry (6). The altered fingers conformation in 5-270 may improve organization of the TNA substrate while preserving discrimination against incorrect nucleotides. However, static structures cannot conclusively determine whether this reflects ground-state preorganization or altered conformational dynamics. The early increase in fidelity, therefore, points to improved structural organization rather than to acceleration of the chemical step itself, but its kinetic basis remains unresolved.

The later transition from 5-270 to 10-92 was different. Only ten mutations were acquired, but several had large functional effects. Reversion of D615G, A741P, and T548P reduced full-length product formation by approximately 88%, 83%, and 74%, respectively. An impactful mutation appeared in each later generation, yet the largest increase in rate occurred only after several had accumulated, suggesting that their effects were not simply additive. D615G illustrates this pattern; it produced little immediate gain when it first appeared in 7-47 but became important in the final 10-92 background. This result is consistent with epistasis, although the relevant interactions will require further experimentation and analysis.

The identities and locations of the latter mutations suggest that they altered local structural mechanics. The three substitutions with the largest reversion effects are buried and located approximately 20 Å from the active site. All ten are nonconservative by BLOSUM62, and seven introduce or remove glycine or proline, residues with strong effects on backbone conformation.

FoldX predicts that most function-driving mutations destabilize the parent scaffold, suggesting a possible trade-off between stability and activity. The stability of the Kod scaffold may have allowed these substitutions to accumulate without compromising folding or function. Stable proteins can tolerate beneficial but destabilizing mutations without falling below the stability threshold required for activity (7,8). Consistent with this model, 10-92 retained activity after prolonged exposure to 90°C despite its extensive sequence divergence. The heat-lysis step used in DrOPS may also have maintained selection for thermal robustness. Stability may therefore have provided the capacity needed to alter polymerase structure and improve TNA synthesis.

Independent Development of Fidelity and Catalysis

Structural analysis of the evolutionary intermediates showed that fidelity and catalytic rate improved at different stages. Fidelity increased sharply during the transition to 5-270, alongside restoration of Watson-Crick geometry and stabilization of the nascent base pair. The largest gains in rate occurred later, as the primer-to- α -phosphate distance shortened and the catalytic aspartates progressively approached the arrangement observed in wild-type Kod. These results indicate that accurate recognition and rapid catalysis have partly distinct structural requirements.

This separation clarifies what selection for full-length synthesis rewards. A mismatch can distort the primer terminus and hinder subsequent nucleotide addition, reducing the probability that an inaccurate polymerase completes the template within the selection window (9,10).

Selection for rapid completion can therefore favor accuracy by avoiding mismatch-induced stalls, even when fidelity itself is not directly selected (11). The 10-92 lineage showed that a speed-based selection improved fidelity before producing the structural refinements required for rapid catalysis.

Generalization to RNA Synthesis

Chapter 4 tested whether homologous recombination and DrOPS could be extended beyond TNA. RNA differs from DNA by a 2'-hydroxyl group, yet replicative DNA polymerases strongly exclude ribonucleotides. Starting from B-family DNA polymerases carrying four known RNA-synthesis mutations (12), the selection produced C28 (4).

C28 synthesizes RNA at approximately 3 nt s⁻¹ with >99% fidelity. It supports extending RNA across long templates, reverse transcription, chimeric DNA-RNA PCR, and the incorporation of several 2'-fluoro and base-modified nucleotides. These activities establish C28 as a useful RNA polymerase and demonstrate that the evolutionary strategy for TNAs can be applied to another class of substrates.

C28 reached this activity through a different structural path from 10-92. Its structure remains close to Kod, with an RMSD of 0.916 Å, and all mutations acquired during recombination lie outside the active site pocket, most between 35 and 60 Å from the primer terminus. The pre-existing Y409G steric-gate mutation removes bulk near the ribose 2' position, while E664K contributes to a continuous positive surface at the duplex interface. Because both mutations were already present in the starting library, the additional activity of C28 reflects the broader recombined sequence. Segmental reversions further indicate that mutations in the catalytic domain contribute most strongly to this gain.

The contrast with 10-92 shows that efficient synthesis of noncognate substrates does not require a common structural solution. TNA synthesis involved extensive, readily visible remodeling, whereas RNA synthesis improved substantially with little change in the global structure. The static similarity of C28 to a natural polymerase does not imply a simple or exclusively local mechanism. In both cases, high activity required mutations distributed across the polymerase architecture.

Wild-Type Precision as a Design Principle

One of the striking themes to emerge from this dissertation is that the directed evolution of new polymerase functions did not proceed by abandoning wild-type precision. Instead, directed evolution repeatedly reconstructed its underlying catalytic logic around a different substrate. This pattern is most evident in the 10-92 lineage. Homologous recombination first drove the enzyme away from the wild-type architecture, producing extensive structural rearrangements that opened access to a new region of functional sequence space. Yet this departure was not the endpoint. Within the globally divergent 5-270 scaffold, the nascent TNA-DNA base pair had already recovered a more canonical geometry, coinciding with the early increase in fidelity. Subsequent point mutations progressively refined the reaction center: the catalytic aspartates returned toward their wild-type arrangement, the primer terminus and incoming α -phosphate approached the distance observed during DNA synthesis, and the globally divergent intermediates partially converged toward a more Kod-like structure. The final enzyme remained substantially altered and specialized for TNA, but it was structurally closer to Kod than the intermediates from which it arose. This partial reconvergence suggests that the initial architectural disruption created evolutionary opportunity, whereas later refinement recovered the structural order required for efficient catalysis.

The active site pocket illustrates the same principle at a local level. Its volume first expanded dramatically during the transition from Kod to Kod-RI by ~250%, then progressively decreased until it was ~10% smaller than that of Kod, thereby appropriately matching the reduced dimensions of the four-carbon TNA nucleotide. The endpoint is therefore not a reproduction of the DNA active site. Rather, it is the recovery of its governing principle enabling close geometric complementarity between enzyme and cognate substrate. Directed evolution changed the identity of the preferred substrate not by discarding the physical requirements of polymerase catalysis, but by rebuilding them around a different chemical system.

This principle is also evident in the structural recognition of pTNA. In Kod-RI, R606 adopts an alternative conformation and forms an ion-pair interaction with the phosphonate linkage of the extended pTNA strand. The R606G substitution in Kod-RSGA removes this interaction, permitting Y594 to form a strong hydrogen bond at the same backbone position. Notably, the corresponding Y594 interaction is also observed when wild-type Kod DNA polymerase is bound to an all-DNA substrate. The more active pTNA polymerase therefore recovers a mode of backbone organization used during natural DNA synthesis, even though pTNA contains modifications to both its sugar and internucleotide linkage. This does not represent global structural convergence, but it provides a local example in which improved activity coincides with the restoration of an interaction characteristic of the natural system.

The C28 RNA polymerase provides a complementary expression of this principle. Unlike 10-92, C28 does not display extensive static remodeling relative to Kod; its backbone architecture remains largely preserved. Within the familiar framework, the incorporated ribonucleotide forms a canonical Watson-Crick base pair and adopts a C2'-endo sugar pucker commonly associated with DNA. The steric-gate mutation removes a direct obstacle to the ribose 2'-hydroxyl, while distal

substitutions alter the electrostatic surface of the duplex interface and are proposed to retune the long-range interaction networks that govern polymerase dynamics. Thus, even RNA synthesis proceeds within an architecture that continues to impose many of the structural conventions of DNA polymerization.

These observations showcase the elegance and precision of the wild-type DNA polymerase, but they should not be interpreted as a general call for literal reversion to wild type when engineering polymerases for new chemistry. Engineered polymerases contain interaction networks, domain arrangements, and substrate-specific contacts that are absent from Kod. D615G disrupts a hydrogen-bonding network and may increase the flexibility needed during TNA translocation, while A741P reorganizes contacts with the TNA-DNA heteroduplex. The 10-92 pocket is smaller than the natural DNA pocket and, more decisively, has inverted its substrate preference, synthesizing TNA more rapidly than DNA. The final enzyme is not a restored DNA polymerase. It is a new polymerase that has recovered native-like catalytic order within a substantially adapted architecture.

In this sense, the natural polymerase provides more than the starting scaffold for directed evolution. It provides an enduring model of what successful catalysis requires. Its value lies not in the expectation that every engineered polymerase should preserve the same sequence or conformation, but in the admirable precision with which it performs substrate selection, base-pair organization, molecular alignment, metal coordination, chemical bond formation, and continued synthesis.

The implication for rational design is not that engineered polymerases should simply be made more wild-type-like. In most cases, that objective is too vague to be actionable. The relevant

features of wild-type precision are distributed across the enzyme, emerge from long-range and epistatic interactions, and may be inseparable from the substrate-specific adaptations that distinguish the polymerase. It is therefore often difficult to identify a mutation that will intentionally restore a wild-type-like parameter without disrupting the new function. Nevertheless, wild-type structures provide a valuable reference when comparison reveals a discrete and chemically interpretable defect, such as a displaced catalytic residue, an excessive primer-phosphate distance, a lost backbone contact, or an improperly sized active site pocket. In the comparatively rare cases where that is obvious, rational design should consider whether a targeted change could recover a productive geometric relationship while preserving the adaptations required for the noncognate substrate. More often, however, directed evolution is needed to discover the distributed and context-dependent route by which catalytic order can be rebuilt.

Artificial Intelligence and Polymerase Engineering

Artificial intelligence (AI) could offer a practical route engineering polymerases. Structure-prediction systems can now predict the structures of biomolecules, while protein language models can rank mutations or generate sequences from patterns learned across natural protein diversity (13-16). These tools could help predict structures enabling rational design, prioritize mutations, design compact libraries, estimate stability, and evaluate polymerase-substrate geometry before experimental screening (17).

The 10-92 lineage illustrates a major remaining challenge, particularly for large proteins with dynamic catalytic cycles. In Chapter 3, AlphaFold 3 was used to predict 10-92 and its evolutionary intermediates before the corresponding crystallographic structures were publicly released. The model recovered the canonical B-family polymerase fold but failed to reproduce the

preorganized closed fingers conformation and broader conformational rearrangements observed experimentally. It also did not accurately reproduce the positioning of key active site residues, including the catalytic aspartates (3).

These discrepancies do not diminish the value of structure prediction. Rather, they show that the relevant phenotype extended beyond preservation of the native fold to include sequence-dependent changes in substrate-bound architecture and conformational state. Predictive polymerase engineering may therefore require models that describe structural ensembles and transitions between states, rather than relying on a single structure. Experimental structures of evolutionary intermediates could provide useful benchmarks for developing such models (17).

AI-guided engineering has already shown promise across diverse proteins, and applications to polymerases are beginning to emerge (18,19). A practical near-term strategy would integrate AI into an iterative design-build-test-learn cycle, where models could be coupled with DrOPS or a similar method to measure the activities of positive and negative variants, and a task-specific model could use those data to select the next experimental round (20-22). Its success would be measured not only by the best enzyme recovered, but also by improved hit rates and fewer experiments required to reach a desired function.

Limitations and Future Work

The conclusions of this dissertation are constrained by the molecular states that could be observed experimentally. Crystal structures provide detailed atomic models, but they capture only conformations compatible with crystallization and do not directly report conformational populations, transition rates, or the sequence of events during nucleotide addition. Future work could combine structural analysis with methods that probe conformational motion and kinetic

partitioning. These approaches could test whether the closed binary state observed in the TNA polymerase lineage is populated in solution and determine how its formation influences substrate binding, chemistry, and translocation.

A related limitation is that the proposed effects of distal mutations proposed to influence catalysis despite minimal change in the local static structure remain unexplored. Their contributions could be examined through targeted reversions, additional structural characterization, and measurements or simulation of conformational dynamics. Such experiments would help distinguish mutations that alter specific interaction networks from those that act through broader changes in stability or conformational dynamics.

The TNA lineage also provides an opportunity for systematic analysis of epistasis. Pairwise and higher-order combinations of D615G, A741P, T548P, S493G, and P550H could establish whether the late increase in catalytic activity reflects additive effects, direct physical coupling, or dependence on the broader 5-270 background. Introducing these substitutions into earlier intermediates or excluding them from 10-92, analyzing NGS data from directed evolution, and additional structural studies would be insightful.

Further directed evolution could determine whether 10-92 and C28 are approaching functional limits with diminishing returns or whether substantial gains remain accessible. An equally important direction is to transfer the principles identified here to other systems. This could include repurposing the C28 scaffold to further accommodate modified RNA, the 10-92 scaffold to accommodate pTNA, or engineering polymerases for novel nucleic acid chemistries. The resulting sequence-function and structural data would also provide valuable training and validation sets for future predictive models.

Ultimately, the value of these enzymes will depend on their integration into practical synthesis workflows. Translation beyond laboratory-scale primer extension could enable programmable enzymatic manufacture of modified oligonucleotides with reduced dependence on multistep chemical synthesis and its associated waste (23,24). Such platforms would extend the significance of polymerase engineering from mechanistic discovery to scalable applications in therapeutics, diagnostics, and biotechnology.

5.1 REFERENCES

1. Hajjar, M., Chim, N., Liu, C., Herdewijn, P. and Chaput, J.C. (2022) Crystallographic analysis of engineered polymerases synthesizing phosphonomethylthreosyl nucleic acid. *Nucleic Acids Res*, **50**, 9663-9674.
2. Maola, V.A., Yik, E.J., Hajjar, M., Lee, J.J., Holguin, M.J., Quijano, R.N., Nguyen, K.K., Ho, K.L., Medina, J.V., Chim, N. *et al.* (2024) Directed evolution of a highly efficient TNA polymerase achieved by homologous recombination. *Nat Catal*, **7**.
3. Hajjar, M., Maola, V.A., Lee, J.J., Holguin, M.J., Quijano, R.N., Nguyen, K.K., Ho, K.L., Medina, J.V., Botello-Cornejo, E., Barpuzary, B. *et al.* (2025) Directed evolution of a TNA polymerase identifies independent paths to fidelity and catalysis. *Nature Communications*, **17**.
4. Medina, E.L., Maola, V.A., Hajjar, M., Ko, G.K., Ho, E.J., Horton, A.R., Chim, N. and Chaput, J.C. (2026) Rapid evolution of a highly efficient RNA polymerase by homologous recombination. *Nature Chemical Biology*, **22**.
5. Chim, N., Shi, C.H., Sau, S.P., Nikoomanzar, A. and Chaput, J.C. (2017) Structural basis for TNA synthesis by an engineered TNA polymerase. *Nature Communications*, **8**.
6. Johnson, K.A. (2008) Role of induced fit in enzyme specificity: A molecular forward/reverse switch. *J Biol Chem*, **283**, 26297-26301.
7. Bloom, J.D., Labthavikul, S.T., Otey, C.R. and Arnold, F.H. (2006) Protein stability promotes evolvability. *P Natl Acad Sci USA*, **103**, 5869-5874.
8. Tokuriki, N., Stricher, F., Serrano, L. and Tawfik, D.S. (2008) How Protein Stability and New Functions Trade Off. *Plos Comput Biol*, **4**.
9. Johnson, S.J. and Beese, L.S. (2004) Structures of mismatch replication errors observed in a DNA polymerase. *Cell*, **116**, 803-816.
10. Batra, V.K., Beard, W.A., Pedersen, L.C. and Wilson, S.H. (2016) Structures of DNA Polymerase Mismatched DNA Termini Transitioning to Pre-catalytic Complexes Support an Induced-Fit Fidelity Mechanism. *Structure*, **24**, 1863-1875.
11. Ravasio, R., Husain, K., Evans, C.G., Phillips, R., Ribezzi-Crivellari, M., Szostak, J.W. and Murugan, A. (2026) Evolution of error correction through a need for speed. *Science*, **391**.
12. Cozens, C., Pinheiro, V.B., Vaisman, A., Woodgate, R. and Holliger, P. (2012) A short adaptive path from DNA to RNA polymerases. *P Natl Acad Sci USA*, **109**, 8067-8072.
13. Madani, A., Krause, B., Greene, E.R., Subramanian, S., Mohr, B.P., Holton, J.M., Olmos, J.J.r., Xiong, C.M., Sun, Z.Z., Socher, R. *et al.* (2023) Large language models generate functional protein sequences across diverse families. *Nat Biotechnol*, **41**, 1099-+.

14. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J. *et al.* (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, **630**.
15. Hayes, T., Rao, R., Akin, H., Sofroniew, N.J., Oktay, D., Lin, Z.M., Verkuil, R., Tran, V.Q., Deaton, J., Wiggert, M. *et al.* (2025) Simulating 500 million years of evolution with a language model. *Science*, **387**, 850-858.
16. Notin, P., Rollins, N., Gal, Y., Sander, C. and Marks, D. (2024) Machine learning for functional protein design. *Nat Biotechnol*, **42**, 216-228.
17. Listgarten, J. and Jiang, H.L. (2026) How artificial intelligence is reengineering protein engineering. *Science*, **392**, 159-166.
18. Paik, I., Ngo, P.H.T., Shroff, R., Diaz, D.J., Maranhao, A.C., Walker, D.J.F., Bhadra, S. and Ellington, A.D. (2023) Improved Bst DNA Polymerase Variants Derived via a Machine Learning Approach. *Biochemistry-US*, **62**, 410-418.
19. Jiang, F., Kang, L.Q., Li, M.C., Zhong, B.Z.T., Chen, X.X., Wu, B.H., Li, M.R., Yu, Y.X. and Hong, L. (2026) Deep learning-guided dual-fitness evolution of T7 RNA polymerase for enhanced stability and activity. *Nucleic Acids Res*, **54**.
20. Romero, P.A., Krause, A. and Arnold, F.H. (2013) Navigating the protein fitness landscape with Gaussian processes. *P Natl Acad Sci USA*, **110**, E193-E201.
21. Larsen, A.C., Dunn, M.R., Hatch, A., Sau, S.P., Youngbull, C. and Chaput, J.C. (2016) A general strategy for expanding polymerase function by droplet microfluidics. *Nature Communications*, **7**.
22. Landwehr, G.M., Bogart, J.W., Magalhaes, C., Hammarlund, E.G., Karim, A.S. and Jewett, M.C. (2025) Accelerated enzyme engineering by machine-learning guided cell-free expression. *Nature Communications*, **16**.
23. Andrews, B.I., Antia, F.D., Brueggemeier, S.B., Diorazio, L.J., Koenig, S.G., Kopach, M.E., Lee, H., Olbrich, M. and Watson, A.L. (2021) Sustainability Challenges and Opportunities in Oligonucleotide Manufacturing. *J Org Chem*, **86**, 49-61.
24. Moody, E.R., Obexer, R., Nickl, F., Spiess, R. and Lovelock, S.L. (2023) An enzyme cascade enables production of therapeutic oligonucleotides in a single operation. *Science*, **380**, 1150-1154.

APPENDIX A: Supplementary Information for Chapter 1

TABLE OF CONTENTS

Table S1. Oligonucleotide Sequences

Table S2. Buffer Compositions

Table S3. Data collection and refinement statistics

Table S4. Local Base-Pair Parameters

Table S5. Local Base-Pair Step Parameters

Table S6. Torsion Angles and Pseudorotation Parameters

Figure S1. Kod Structural Overview

Figure S2. Kod-RI incorporating TNA and Active Site Overlays

Figure S3. Kod-RI/pTNA: Y594 Water Coordination

Figure S4. Backbone Comparison: Comprehensive Overlay

Figure S5. Torsion Angles Definition

Figure S6 Incorporated Nucleotides (χ, δ) Covariance

Figure S7. pTNA's Sugar Puckering in Kod-RSGA

Supplementary Table S1. Oligonucleotide Sequences

Name	Oligonucleotide sequence (5' – 3')	Comments
P _f	ATCCATATGATCCTCGACACTGACTAC	Cloning
P _r	ACGCATGCGGCCGCTCAAGTTCCTTCGGCTTCAGCCA	
P ₁	/5IRD/ GGATACCACC	Primer extensions
T ₁	TCTCTATAGTGAGTCGTATAGGTGGTATCC	
P _c	CGCGAACTGCG	Crystal complex
T _c	/Cy5/ AAATTCGCAGTTCGCG	

Supplementary Table S2. Buffer Compositions

Name	Buffer Composition	Comments
Kod low salt/lysis buffer	10 mM Tris.Cl pH 7.5, 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 10% glycerol, 0.125 mg mL ⁻¹ egg hen lysozyme when lysing cells	Purification
Kod high salt buffer	10 mM Tris.Cl pH 7.5, 1 M NaCl, 0.1 mM EDTA, 1 mM DTT, 10% glycerol	
Kod buffer	50 mM Tris.Cl pH 8.5, 200 mM NaCl, 0.1 mM EDTA, 1 mM DTT	Protein storage; Crystallization
Thermopol® (NEB)	20 mM Tris-HCl, 10 mM (NH ₄) ₂ SO ₄ , 10 mM KCl, 2 mM MgSO ₄ , 0.1% Triton® X-100 pH 8.8	Primer extensions
Stop Buffer	95% formamide, 25 mM EDTA, pH 8.0	

Supplementary Table S3. Data Collection and Refinement Statistics. The statistics for the highest-resolution shell are shown in parentheses.

	Kod-RI DNA	Kod-RI pTNA	Kod-RSGA pTNA
Data Collection			
Space group	<i>P 2 2₁ 2₁</i>	<i>P 2 2₁ 2₁</i>	<i>P 2 2₁ 2₁</i>
Cell Dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	66.3, 112.0, 149.2	65.8, 111.6, 149.0	64.1, 111.5, 144.3
<i>a</i> , <i>b</i> , <i>g</i> (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	57.03-2.71 (2.81-2.71)	40.92-1.98 (2.05-1.98)	44.04-3.01 (3.12-3.01)
Total reflections	61392 (6062)	552026 (55933)	68272 (6983)
Unique reflections	30746 (3037)	77125 (7617)	19530 (1955)
Redundancy	2.0 (2.0)	7.2 (7.3)	3.5 (3.6)
Completeness (%)	99.36 (99.80)	99.90 (99.97)	91.96 (94.51)
<i>I</i> / <i>σI</i>	13.14 (4.74)	17.86 (2.46)	10.33 (2.44)
Wilson B-factor	47.10	36.59	54.99
<i>R</i> _{merge}	0.026 (0.137)	0.074 (1.204)	0.1458 (0.7314)
CC1/2	0.998 (0.949)	0.999 (0.741)	0.991 (0.810)
CC*	1 (0.987)	1 (0.923)	0.998 (0.946)
Refinement			
Resolution (Å)	2.71	1.98	3.01
No. of reflections	30731 (3037)	77092 (7615)	19492 (1944)
<i>R</i> _{work} / <i>R</i> _{free}	0.161/0.226 (0.188/0.254)	0.176/0.212 (0.249/0.243)	0.176/0.270 (0.258/0.365)
No. atoms	6883	7052	6777
Protein/DNA	6798	6767	6735
ptA 12 & 13	-	42	42
Water	85	243	-
B-factors	52.35	52.71	54.24
Protein/DNA	52.47	52.91	54.22
ptA 12 & 13	N/A	34.71	55.84
Water	42.42	50.37	-
Clashscores	5.38	4.48	12.09
RMS dev. (Bonds; Å)	0.008	0.008	0.010
RMS dev. (Angles; °)	1.22	1.16	1.32
PDB ID	7RSS	7RSR	7TQW

Supplementary Table S4. Local Base-Pair Parameters.

BASE PAIR	Shear (Å)	Stretch (Å)	Stagger (Å)	Buckle (°)	Propeller (°)	Opening (°)
RI/DNA						
dA13 - dT4	-0.21	-0.12	0.35	-24.29	-10.10	8.20
dA12 - dT5	0.20	-0.19	-0.20	-3.79	-2.24	0.36
dG11 - dC6	0.12	-0.21	0.16	-1.57	-13.73	4.38
RI/TNA						
tA13 - dT4	-0.25	-0.37	0.29	-16.69	-15.14	1.70
tA12 - dT5	-0.30	-0.20	-0.32	-4.28	-4.75	1.76
dG11 - dC6	0.37	-0.25	-0.22	-5.48	-4.14	1.47
RI/pTNA						
ptA13 - dT4	-0.15	-0.20	0.24	-11.36	-15.97	3.96
ptA12 - dT5	-0.04	-0.15	-0.38	2.77	-3.40	-1.06
dG11 - dC6	0.31	-0.21	0.34	-10.99	-13.44	3.51
RSGA/pTNA						
ptA13 - dT4	-0.19	-0.15	0.29	-16.7	-25.13	1.18
ptA12 - dT5	-0.44	-0.10	0.42	-8.42	-13.20	-6.01
dG11 - dC6	0.55	-0.15	-0.01	-9.00	-0.43	4.82

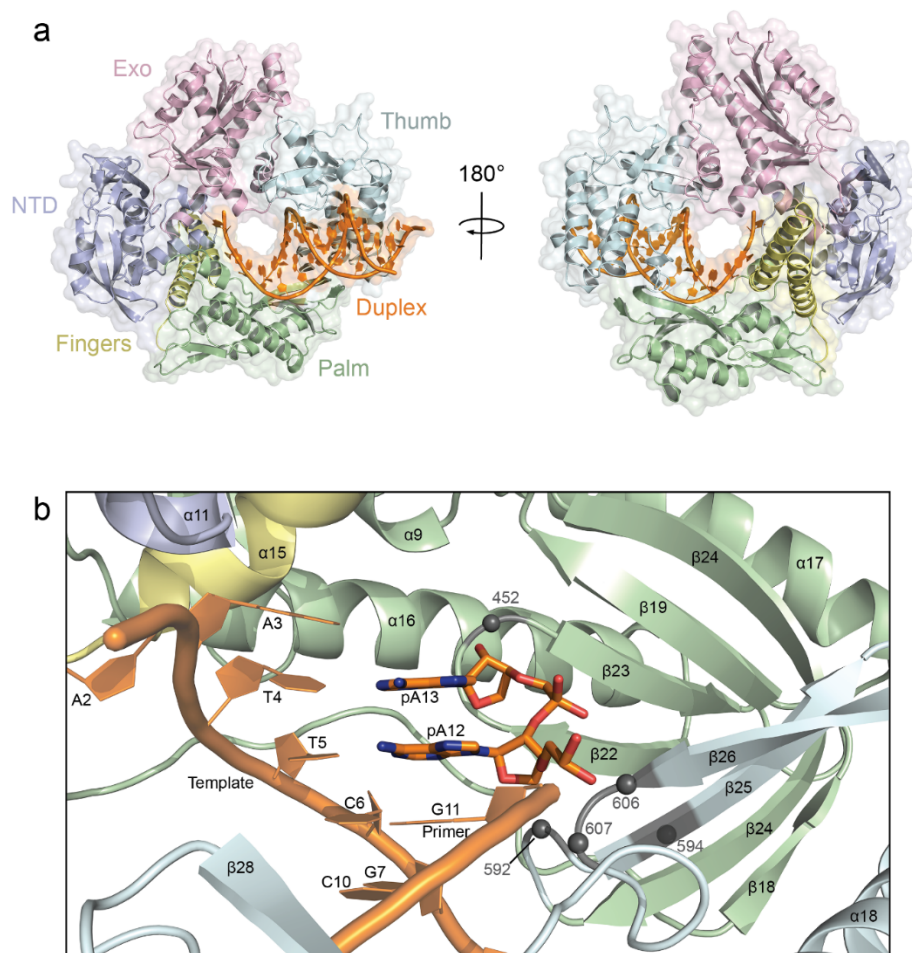
Supplementary Table S5. Local Base-Pair Step Parameters.

STEP	Shift (Å)	Slide (Å)	Rise (Å)	Tilt (°)	Roll (°)	Twist (°)
RI/DNA						
dT4 dT5 / dA12 dA13	-0.50	-0.08	2.83	5.82	8.19	29.75
dT5 dC6 / dG11 dA12	0.66	-0.55	3.31	-4.09	7.12	40.23
RI/TNA						
dT4 dT5 / tA12 tA13	-0.29	-0.41	3.02	6.36	5.35	28.08
dT5 dC6 / dG11 tA12	0.41	-0.29	3.39	0.33	0.77	43.30
RI/pTNA						
dT4 dT5 / ptA12 ptA13	-0.19	-0.30	2.97	6.93	6.15	31.35
dT5 dC6 / dG11 ptA12	0.82	-0.69	3.64	-4.63	0.07	42.36
RSGA/pTNA						
dT4 dT5 / ptA12 ptA13	-0.70	-0.48	3.16	-0.28	10.07	30.50
dT5 dC6 / dG11 ptA12	1.04	-0.57	3.47	-0.37	-9.14	44.77

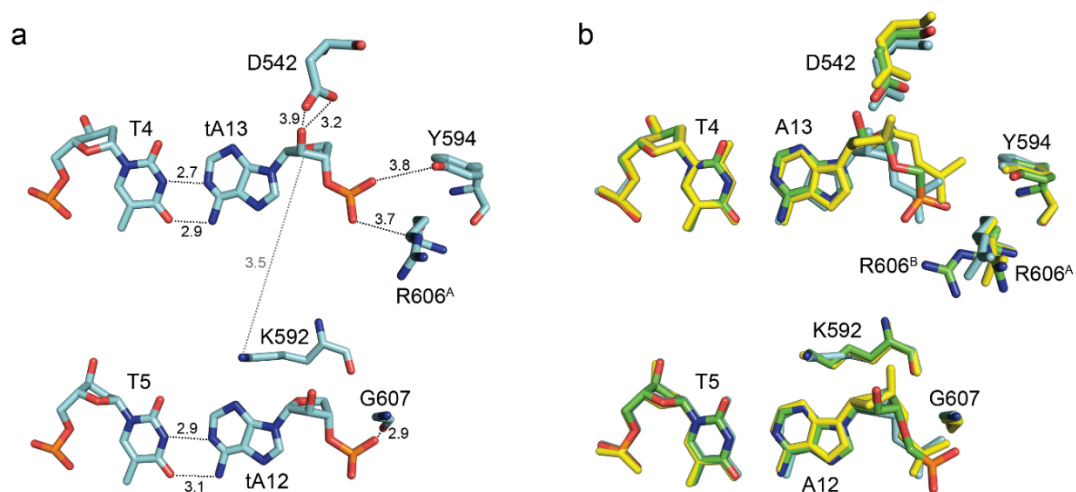
Supplementary Table S6. Torsion Angles and Pseudorotation Parameters.

	α	β	γ	δ	ϵ	ζ	χ	ν_0	ν_1	ν_2	ν_3	ν_4	$\nu_{\max}$	P
RI/DNA														
dA13	-77.7	167.4	51.5	91.2	N/A	N/A	-121.7	-24.9	2.9	18.4	-33.1	36.5	36.6	59.9
dA12	-71.7	-151.6	42.0	145.0	-167.5	-92.3	-103.5	-8.9	23.6	-28.8	24.5	-10.0	28.8	181.3
dG11	84.1	-118.6	-160.3	94.2	-150.5	-67.4	-172.3	-9.3	-7.7	20.7	-26.5	22.4	26.4	38.5
RI/TNA														
tA13	-88.9	N/A	168.3	159.0	N/A	N/A	-101.1	-5.4	-17.7	33.8	-36.6	24.7	37.3	25.3
tA12	126.5	N/A	-135	161.4	-108.4	-83.2	-85.3	2.6	-21.3	31.88	-30.0	16.1	32.6	12.8
dG11	-56.6	161.9	49.0	130.1	177.1	-88.8	-126.3	-35.5	40.7	-30.3	10.9	15.1	40.0	139.3
RI/pTNA														
ptA13	66.0	-149.9	176.4	162.0	N/A	N/A	-119.1	-14.4	-12.7	32.3	-41.1	32.6	40.5	37.2
ptA12	-39.2	-147.4	-63.4	163.2	-84.6	-160.3	-92.7	-20.0	-4.9	26.5	-37.8	35.1	39.0	47.2
dG11	86.0	-124.8	-157.0	113.6	-134.6	-75.3	-166.7	-22.4	17.8	-7.0	-5.8	17.7	21.8	108.6
RSGA/pTNA														
ptA13	68.9	-160.1	-177.8	166.8	N/A	N/A	-107.2	-17.3	-11.7	34.0	-44.6	36.1	44.1	39.5
ptA12	-36.2	-148.4	-68.9	145.5	-93.3	-149.2	-78.2	-11.3	-6.4	20.3	-27.1	22.7	27.0	41.2
dG11	95.1	-125.6	-170.5	129.9	-135.5	-87.3	-154.9	-26.8	32.9	-26.3	11.7	9.4	32.2	144.7

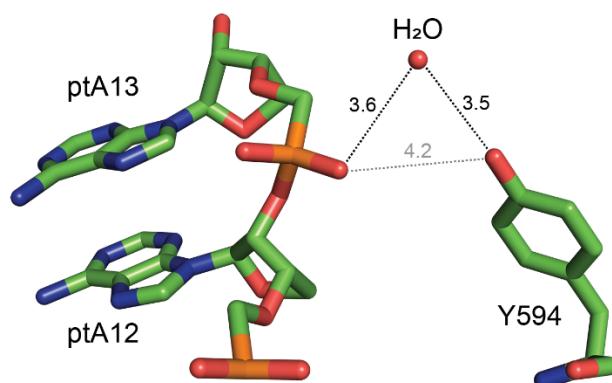
The torsion angles are reported in degrees for the two incorporated adenosine nucleotides, A12 and A13, and the preceding nucleotide of the primer, dG11. A pictorial definition of the torsion angles is provided in SI Fig 2. The angles α , β , γ , δ , ϵ , and ζ are assigned to the exocyclic backbone of the genetic polymers. The β angle is omitted for TNA. The ϵ and ζ angles non-computable for the terminal A13 nucleotides. The χ angle is torsion angle about the glycosidic bond. The angles ν_0 through 4 describe the sugar moiety's endocyclic torsion angles. The $\nu_{\max}$ is the maximum torsion angle and is reflective of the degree of puckering; a $\nu_{\max}$ of 0° reflects a planar furanose ring. The pseudorotation phase angle, P , ranges from 0° to 360° and $P = 0^\circ$ is defined such that the torsion angle ν_2 is maximally positive.



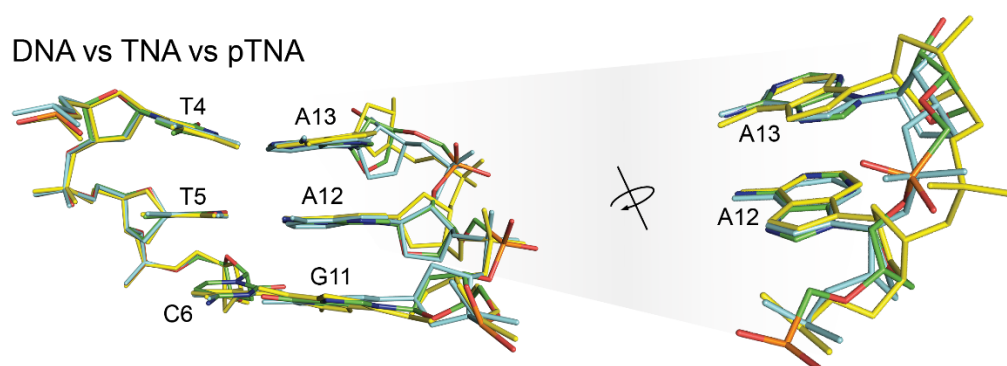
Supplementary Figure S1. Kod Structural Overview (a) The global architecture of Kod bound to a primer-template duplex. The front and back view showing a surface rendering overlayed on a cartoon representation is colored by protein subdomain. The primer-template duplex is shown in orange. (b) A view of the active site is shown with the exonuclease domain hidden to avoid view obstruction. Positions with residues that interact with the incorporated pTNA are shown as gray spheres.



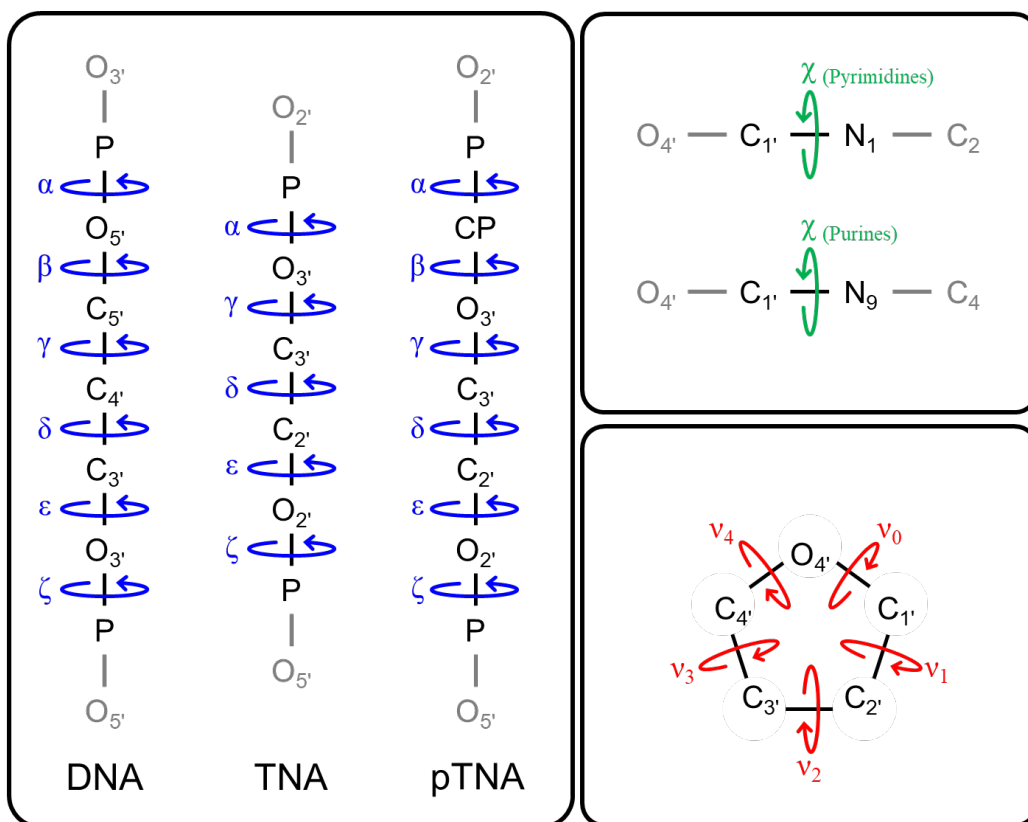
Supplementary Figure S2. Kod-RI Incorporating TNA and Active Site Overlays. (a) A panel equivalent to main text Figure 2b-d for TNA incorporations showing base pairing and protein interactions in angstroms. (b) An overlay of the incorporated DNA, TNA, pTNA is shown, highlighting the backbone variation in context of the active site.



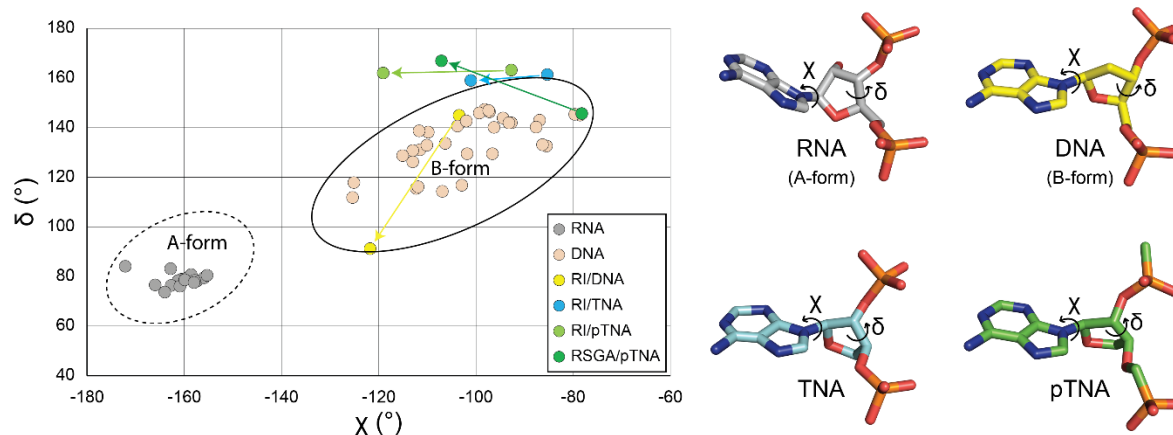
Supplementary Figure S3. Tyrosine 594 Forms a Bridging Water Interaction. In the structure of Kod-RI incorporating pTNA, tyrosine 594 is 4.2 Å away from the phosphonate backbone. Instead of hydrogen bonding, compensatory water coordination is observed.



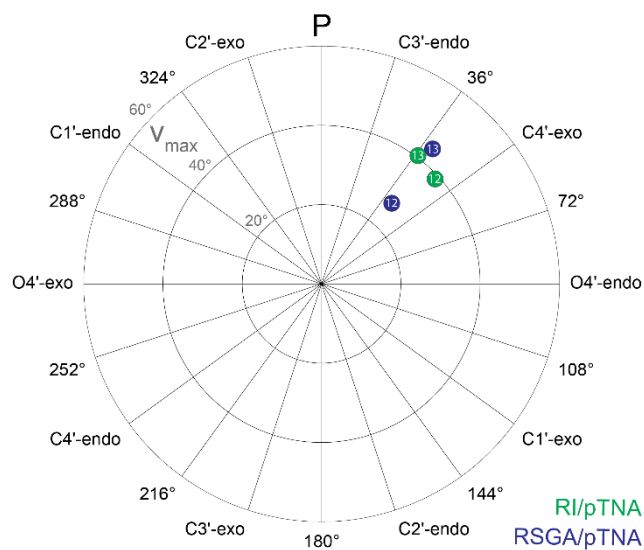
Supplementary Figure 4. A Backbone Comparison of Newly Added Nucleotides . A view identical to that of Figure 3, overlaying RI/pTNA (green, colored by element), RI/DNA (solid yellow), and RI/TNA (solid cyan) at once.



Supplementary Figure 5. Torsion Angles Definition. A torsion angle about a bond is defined by 4 atoms, the two bonding atoms, and an additional atom bonded to each. Pictorial definitions of the exocyclic backbone torsion angles in the backbones of DNA, TNA, and pTNA are shown in blue. TNA's shortened backbone lacks the β angle. For all genetic systems, the torsion angles about the glycosidic bond for pyrimidines and purines are defined in green, and the sugar endocyclic torsion angles are defined in red.



Supplementary Figure 6. Incorporated Nucleotides (χ, δ) Covariance. The (χ, δ) covariance is used to analyze the helical geometry of the TNA and pTNA nucleotides with canonical A-form RNA (PDB ID: 3ND4) and B-form DNA (PDB ID: 3BSE) nucleotides as a reference. Arrows drawn on the plot represent the shift in (χ, δ) covariance between the first (A12; $n+1$) and the second incorporations (A13; $n+2$). On the right, representative monomers, that is, A12 for DNA, TNA, and pTNA, are shown in comparison to a canonical A-form RNA monomer.



Supplementary Figure 7. pTNA's Sugar Puckering in Kod-RSGA. The pseudorotation parameters of pTNA in RI/pTNA (green) and RSGA/pTNA (blue) are plotted, depicting the slight increase in the extent of puckering ($v_{\max}$).

APPENDIX B: Supplementary Information for Chapter 2

TABLE OF CONTENTS

Supplementary Tables

Supplementary Table 1. DNA gBlocks

Supplementary Table 2. DNA oligonucleotides

Supplementary Table 3. Summary of droplet information and incubation conditions

Supplementary Table 4. Data collection and refinement statistics

Supplementary Table 5. Local base-pair parameters

Supplementary Figures

Supplementary Figure 1. Sequence alignment of archaeal RSGA DNA polymerase variants

Supplementary Figure 2. Assembly of the DNA recombination library by overlap PCR

Supplementary Figure 3. Schematic of the lysate screen

Supplementary Figure 4. Lysate activity screen following five rounds of selection

Supplementary Figure 5. TNA synthesis activity of clone 5.270

Supplementary Figure 6. Sequence alignment of newly evolved TNA polymerases

Supplementary Figure 7. Phylogenetic analysis of the evolved TNA polymerases

Supplementary Figure 8. Comparative analysis of ancestral TNA polymerases

Supplementary Figure 9. TNA versus DNA primed TNA synthesis

Supplementary Figure 10. Functional analysis of 10-92

Supplementary Figure 11. Schematic of the fidelity assay

Supplementary Figure 12. Polymerase fidelity data

Supplementary Figure 13. SDS-PAGE analysis of the 10-92 reversion proteins

Supplementary Figure 14. TNA synthesis activity of the 10-92 reversions

Supplementary Figure 15. Activity of Kod-RSGA with directed evolution mutations

Supplementary Figure 16. Secondary structure map of the 10-92 TNA polymerase

Supplementary Figure 17. Active site interaction map

Supplementary Figure 18. Active site pocket volume

Supplementary Figure 19. Structural analysis of key reversion mutations

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-SurfTM 1 surfactant, Pico-GlideTM 1, and Pico-BreakTM 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), respectively. 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 μ M Vector-Fwd (Supplementary Table 2), 0.5 μ M Vector-Rvs, 5 ng of pGDR11 9°N-WT, 0.4 mM dNTPs, 0.5 μ L Q5 DNA polymerase (0.02 U/ μ L). 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 μ L) 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 μ L of PCR reaction was mixed with 300 μ L 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 μ L 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 μ L x2 of nuclease free water and UV quantified.

gBlocks were reconstituted at the recommended volume with nuclease free water, mixed with 100 ng (1 μ L) 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 μ L of Gibson reaction was transformed into 50 μ L NEB 5-alpha competent *E. coli* and incubated on ice for 30 min. Prior to heat shock, 28 μ M 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 μ L) 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 μ L 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 μ L of DNA was added from the 10 ng/ μ L 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 μ L of Q5 DNA polymerase (0.02U/ μ L) 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 μ L of the overlap PCR assembly mixture was combined with 1x Q5 buffer, 0.4 mM dNTPs, 0.5 μ L of Q5 DNA polymerase and the presence of 0.5 μ M Fragment1-Fwd and 0.5 μ M 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 μ L) 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 μ L 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 μ L) 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 μ Lx4 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 μ Lx2 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 μ L 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 (NH₄)₂SO₄, 10 mM KCl, 2 mM MgSO₄, 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 $110\text{-}125 \times 10^6$ 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 μ s), 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 μ Lx2 DNA wash buffer and eluted with nuclease free water (20 μ L). The gene was amplified in 1x Q5 buffer, 0.5 μ M fragment1-Fwd (Supplementary Table 2), 0.5 μ M fragment11-Rvs 0.4 mM dNTPs, 4 μ L of recovered DNA, and 0.5 μ L Q5 DNA polymerase (0.02 U/ μ L) in a total volume of 50 μ L. 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 μ L) 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 μ L PCR reaction to 300 μ L Zymo DNA binding buffer) and added to a Zymo IC column. DNA was cleaned with 500 μ Lx4 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 μ Lx2 of nuclease free water at 14,000 rcf for 1 min. The addition of 5 μ L NEB rCutSmart buffer and 5 μ L 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 μ L of NEB 6x Blue Gel loading dye was mixed with the DpnI treated sample and loaded into a 1% ethidium-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 μ L of DNA wash buffer. Samples were dried at 14,000 rcf for 3 min prior to the addition of 20 μ L 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 column was washed with 4x500 μ L 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 μ L of nuclease free water at 14,000 rcf for 1 min. A transformation was performed as previously mentioned using 1 μ L 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 previously⁵. The activity of Kod-WT (exo-) polymerase was empirically determined by serial diluting the enzyme in PCR reactions, from which a resulting 0.013 μ M 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 μ M forward primer, 1 μ M reverse primer, 1 mM TTP, 1 mM dCTP, 0.2 mM dGTP, 0.2 mM dATP, 5.5 mM $MgCl_2$, 0.013 μ M Kod-WT (exo-) polymerase, 100 ng of 5-270 plasmid and 2 fold serial dilution of $MnCl_2$ from 500-31.25 μ M. 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, minipreped 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 described⁸. 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 OD₆₀₀ ~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).

Thermostability challenge

Purified polymerase was quantified to 10 μ M and aliquoted (100 μ L) 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 μ M primer-template duplex, 100 μ M tNTPs, 1 μ M heat-treated Polymerase in a 20 μ L reaction. After 30 minutes of incubation, a 1 μ L aliquot of the reaction was quenched by the addition of 39 μ L 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 synthesis⁹. 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 μ M.

Synthesis and preparation of TNA aptamers for BLI analysis

Aptamer generation for BLI analysis was performed as described¹⁰. 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 μ M) was annealed to a biotinylated-PBS8 primer (1 μ M) in 1x ThermoPol at 90°C for 5 min and immediately snap cooled, followed by the addition of 100 μ M of respective TNA base mixes and 2 μ M of 10-92 for a total volume of 750 μ L. 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 μ L and adding 250 μ L 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 μ M Extra_PBS37) containing a two-nucleotide mismatch (AA-AA) annealed to a DNA template (1 μ M CMM595). The primer extension reaction was performed for 1 h at 55°C in 1x Thermopol buffer, 100 μ M tNTPs, and 2 μ M 10-92 TNA polymerase in a 25 μ L 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 μ M Bst-LF DNA polymerase, 2 μ M PBS33, 0.4 mM dNTPs in a final volume of 20 μ L 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, miniprepmed 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 μ M) was annealed to either a TNA or DNA PBS8 primer (PBS8, 0.5 μ M) in 1x ThermoPol at 95°C for 5 min and immediately snap cooled, followed by the addition of 10-92 polymerase (0.5 μ M). The reaction was initiated with the addition of 100 μ M of tNTPs for a total volume of 20 μ L. 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 5 α *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 μ L) from a single master mix poised under single-turnover conditions with equimolar concentrations of polymerase (0.2 μ M) and pre-annealed P-T duplex (0.2 μ M 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 μ M 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 μ L (6 μ M) EvaGreen® dye, previously identified as the optimal intercalating dye to monitor synthesis of TNA on a DNA template¹². Each reaction (25 μ L) 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 ability¹³, 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 described¹² by multiplying data with the conversion factor $F = [\text{Template}] * L / [\text{Polymerase}]$, 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 MgCl₂, 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 Molprobit 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 AGAAGGAAAACGGCGAGTTTAAAGATTGAGTACGACCGGACTTTTGAACCCTACTTCTA CGCCCTCCTGAAGGACGATTCTGCCATTGAGGAAGTCAAGAAGATAACCGCCGAGAGG CACGGGACGGTTGTAACGGTTAAGCGGGTTGAAAAGGTTTCAAGAAGTTCTCTCGGGA GACCAGTTGAGGTCTGGAACTCTACTTTACTCATCCGCAGGACGTCCCAGCGATAAG GGACAAGATACGAGAGCATCCAGCAGTTATTGACATCTACGAGTACGACATACCCTTC GCCAAGCGCTACCTCATAGACAAGGGATTAGTGCCAATGGAAGGCGACGAGGAGCTGA AAATGCTCGCCTTCGCGATTGCGACTCTCTACCATGAGGGCGAGGAGTTTCGCCGAGGG GCCAATCCTTATGATAAGCTACGCCGACGAGGAAGGGGCCAGGGTGATAAATTGGAAG AACGTGGATCTCCCCTACGTTGACGTCGTCTCGACGGAGAGGGAGATGATAAAGCGCT TCCTCCGTGTTGTGAAGGAGAAAGACCCGGACGTTCTCATAACCTACAACGGCGACAA CTTCGACTTCGCCTATCTGAAAAAGCGCTGTGAAAAGCTCGGAATAAACTTCGCCCTC GGAAGGGATGGAAGCGAGCCGAAGATTGAGAGGATGGGCGACAGGTTTGCCGTGGAAG TGAAGGGACGGATACACTTCGATCTCTATCCTGTGATAAGACGGACGATAAACCTGCC CACATACACGCTTGAGGCCGTTTATGAAGCCGTCTTCGGTCAGCCGAAGGAGAAGGTT TACGCTGAGGAAATAACCAACAGCCTGGGAAACCGGCGAGAACCTTGAGAGAGTCGCCC GCTACTCGATGGAAGATGCGAAGGTCACATACGAGCTTGGAAGGAGTTCTTCCGAT GGAGGCCAGCTTTCTCGCTTAATCGGCCAGTCCCTCTGGGACGTCTCCCGCTCCAGC ACTGGCAACCTCGTTGAGTGTTTCTCCTCAGGAAGGCCTATGAGAGGAATGAGCTGG CCCCGAACAAGCCCGATGAAAAGGAGCTGGCCAGAAGACGGCAGAGCTATGAAGGAGG CTATGTAAAAGAGCCCGAGAGAGGGTTGTGGGAGAACATAGTGACCTAGATTTTAGA TCCCTGTACCCCTCAATCATCATCACCCACAACGTCTCGCCGGATACGCTCAACAGAG AAGGATGCAAGGAATATGACGTTGCCCCACAGGTCGGCCACCGCTTCTGCAAGGACTT

	CCCAGGATTTATCCCGAGCCTGCTTGGAGACCTCCTAGAGGAGAGGCAGAAGATAAAG AAGAAGATGAAGGCCACGATTGACCCGATCGAGAGGAAGCTCCTCGATTACAGGCAGA GGCGTATCAAGATCCTGGCAAGCAGCTACTACGGTTACTACGGCTATGCAAGGGCGCG CTGGTACTGCAAGGAGTGTGCAGAGAGCGTAACGGCCTGGGGAAGGGAGTACATAACG ATGACCATCAAGGAGATAGAGGAAAAGTACGGCTTTAAGGTAATCTACAGCGACACCG ACGGATTTTTTGGCCACAATACCTGGAGCCGATGCTGAAACCGTCAAAAAGAAGGCTAT GGAGTTCCTCAAGTATATCAACGCCAACTTCCGGGCGCGCTTGAGCTCGAGTACGAG GGCTTCTACAAACGCGGCTTCTTCGTACGAAGAAGAAGTATGCGGTGATAGACGAGG AAGGCAAGATAACAACGGGTGGACTTGAGATTGTGAGGCGTGAAGTGGAGCGAGATAGC GAAAGAGACGCAGGCGAGGGTTCTTGAAGCTTTGCTAAAGGACGGTGACGTCGAGAAG GCCGTGAGGATAGTCAAAGAAGTTACCGAAAAGCTGAGCAAGTACGAGGTTCCGCCGG AGAAGCTGGTGATCCACGAGCAGATAACGAGGGATTTAAAGGACTACAAGGCAACCGG TCCCCACGTTGCCGTTGCCAAGAGGTTGGCCGCGAGAGGAGTCAAATACGCCCTGGA ACGGTGATAAGCTACATCGTGCTCAAGGGCTCTGGGAGGATAGGCGACAGGGCGATAC CGTTCGACGAGTTCGACCCGGCGAAGCACAAAGTACGACGCCGAGTACTACATTGAGAA CCAGGTTCTCCCAGCCGTTGAGAGAATTCTGAGAGCCTTCGGTTACCGCAAGGAAGAC CTGCGCTACCAGAAGACGAGACAGGTTGGTTTGAGTGCTTGGCTGAAGCCGAAGGGAA CTTGA
Tgo_RSGA_ gBlock	AGCGGCCTGGTGCCGCGCGGCAGCATGATCCTCGACACTGACTACATAACCGAGGATG GAAAGCCTGTCTATAAGAATTTTCAAGAAGGAAAACGGCGAGTTTAAGATTGACTACGA CCGGAACCTTTGAACCTTACATCTACGCCCTCCTGAAGGACGATTCTGCCATTGAGGAC GTCAAGAAGATAACCGCCGAGAGGCACGGGACGACCGTAAGGGTTGTCCGGGCCGAAA AGGTTAAGAAGAAGTTCCTCGGGAGACCAATAGAGGTCTGGAACTCTACTTTACTCA TCCGCAGGACGTCCCAGCGATAAGGGACAAGATAAAGGAGCATCCAGCAGTTGTGGAC ATCTACGAGTACGACATACCCTTCGCCAAGCGCTACCTCATAGACAAGGGATTAATCC CAATGGAAGGCGACGAGGAGCTGAAAATGCTCGCCTTCGCGATTGCGACTCTCTACCA TGAGGGCGAGGAGTTCCGCCAGGGGCCAATCCTTATGATAAGCTACGCCGACGAGGAA GGGGCCAGGGTGATAACTTGGAAGAACATCGATCTCCCCTACGTTGACGTCGTCTCGA

	CGGAGAAGGAGATGATAAAGCGCTTCCTCAAGGTTGTGAAGGAGAAAGACCCGGACGT TCTCATAACCTACAACGGCGACAACCTTCGACTTCGCCTATCTGAAAAAGCGCTCCGAA AAGCTCGGAGTCAAGTTCATCCTCGGAAGGGAAGGAAGCGAGCCGAAGATTTCAGAGGA TGGGCGACAGGTTTGCCGTCGAAGTGAAGGGACGGATACACTTCGATCTCTATCCTGT GATAAGACGGACGATAAACCTGCCCACATACACGCTTGAGGCCGTTTATGAAGCCATC TTCGGTCAGCCGAAGGAGAAGGTTTACGCTGAGGAAATAGCGCAGGCCTGGGAAACCG GCGAGGGACTTGAGAGAGTCGCCCCTACTCGATGGAAGATGCGAAGGTCACATACGA GCTTGGAAGGAGTTCTTCCCGATGGAGGCCAGCTTTCTCGCTTAGTAGGCCAGTCC CTCTGGGACGTCTCCCGCTCCAGCACTGGCAACCTCGTTGAGTGTTTCTCCTCAGGA AGGCCTATGAGAGGAATGAGCTGGCCCCGAACAAGCCCGATGAAAGGGAGCTGGCCAG AAGACGGGAGAGCTATGCGGGAGGCTATGTAAAAGAGCCCGAGAGAGGGTTGTGGGAG AACATAGTGTACCTAGATTTTAGATCCCTGTACCCCTCAATCATCATCACCCACAACG TCTCGCCGGATACGCTCAACAGAGAAGGATGCGAGGAATATGACGTTGCCCCACAGGT CGGCCACAAGTTCTGCAAGGACTTCCCAGGATTTATCCCGAGCCTGCTTGAGACCTC CTAGAGGAGAGGCAGAAGGTAAAGAAGAAGATGAAGGCCACGATTGACCCGATCGAGA AGAAGCTCCTCGATTACAGGCAGAGGCGTATCAAGATCCTGGCAAGCAGCTTCTACGG TTACTACGGCTATGCAAAGGCGCGCTGGTACTGCAAGGAGTGTGCAGAGAGCGTAACG GCCTGGGGAAGGCAGTACATAGAGACCACCATCAGGGAGATAGAGGAAAAGTTTGGCT TTAAGGTACTCTACGCGGACACCGACGGATTTTTTGGCCACAATACCTGGAGCCGATGC TGAAACCGTCAAAAAGAAGGCTAAGGAGTTCTCTCGACTATATCAACGCCAAACTTCCG GGCCTGCTTGAGCTCGAGTACGAGGGCTTCTACAAACGCGGCTTCTTCGTACGAAGA AGAAGTATGCGGTGATAGACGAGGAAGACAAGATAACAACGGGTGGACTTGAGATTGT GAGGCGTGA CTGGAGCGAGATAGCGAAAGAGACGCAGGCGAGGGTTCTTGAAGCTATA CTAAAGCACGGTGACGTCGAGGAAGCCGTGAGGATAGTCAAAGAAGTTACCGAAAAGC TGAGCAAGTACGAGGTTCCGCCGGAAGCTGGTGATCTACGAGCAGATAACGAGGGA TTTAAAGGACTACAAGGCAACCGGTCCCCACGTTGCCGTTGCCAAGAGGTTGGCCGCG AGAGGAATAAAAATACGCCCTGGAACGGTGATAAGCTACATCGTGCTCAAGGGCTCTG GGAGGATAGGCGACAGGGCGATACCGTTTCGACGAGTTTCGACCCGGCGAAGCACAAGTA
--	---

	CGACGCCGAGTACTACATTGAGAACCAGGTTCTCCCAGCCGTTGAGAGAATTCTGAGA GCCTTCGGTTACCGCAAGGAAGACCTGCGCTACCAGAAGACGAGACAGGTTGGTTTGG GGGCTTGGCTGAAGCCGAAGACATGAGTCGACCTGCAGCCAAGCTTA
DV_RSGA_ gBlock	AGCGGCCTGGTGCCGCGCGGCAGCATGATCCTCGACGCTGACTACATAACCGAGGATG GAAAGCCTATTATAAGAATTTTCAAGAAGGAAAACGGCGAGTTTAAGGTTGAGTACGA CCGGAACCTTTAGACCCTACATTTACGCCCTCCTGAAGGACGATTCTCAGATTGATGAA GTCAGGAAGATAACCGCCGAGAGGCACGGGAAGATAGTAAGAATTATAGATGCCGAAA AGGTTAGGAAGAAGTTCCTCGGGAGACCAATTGAGGTCTGGAGGCTCTACTTTGAACA TCCGCAGGACGTCCCAGCGATAAGGGACAAGATACGAGAGCATTCCGCAGTTATTGAC ATCTTTGAGTACGACATACCCTTCGCCAAGCGCTACCTCATAGACAAGGGATTAATTC CAATGGAAGGCGACGAGGAGCTGAAATTGCTCGCCTTCGCGATTGCGACTCTCTACCA TGAGGGCGAGGAGTTCGCCAAGGGGCCAATCATAATGATAAGCTACGCCGACGAGGAA GAAGCCAAAGTGATAACTTGAAGAAGATCGATCTCCCCTACGTTGAGGTCGTCTCGA GCGAGAGGGAGATGATAAAGCGCTTCCTCAAGGTTATAAGGGAGAAAGACCCGGACGT TATAATAACCTACAACGGCGACTCTTTTCGACCTTCCCTATCTGGTTAAGCGCGCCGAA AAGCTCGGAATAAAGCTACCCCTCGGAAGGGATGGAAGCGAGCCGAAGATGCAGAGGC TTGGCGACATGACAGCCGTCGAAATAAAGGGACGGATACACTTCGATCTCTATCACGT GATAAGACGGACGATAAACCTGCCCCACATACACGCTTGAGGCCGTTTATGAAGCCATC TTCGGTAAGCCGAAGGAGAAGGTTTACGCTCACGAAATAGCTGAGGCCTGGGAAACCG GCAAGGGACTTGAGAGAGTCGCCAAGTACTCGATGGAAGATGCGAAGGTCACATACGA GCTTGGGAGGGAGTTCTTCCCGATGGAGGCCAGCTTTCTCGCTTAGTCGGCCAGCCC CTCTGGGACGTCTCCCGCTCCAGCACTGGCAACCTCGTTGAGTGGTACCTCCTCAGGA AGGCCTATGAGAGGAATGAGCTGGCCCCGAACAAGCCCGATGAAAGGGAGTACGAGAG AAGGCTACGGGAGAGCTATGCTGGAGGCTATGTAAAAGAGCCCGAGAAAGGGTTGTGG GAGGGGTTAGTGTCCCTAGATTTTAGATCCCTGTACCCCTCAATCATCATCACCCACA ACGTCTCGCCGGATACGCTCAACAGAGAAGGATGCAGGGAATATGACGTTGCCCCAGA GGTCGGCCACAAGTTCTGCAAGGACTTCCCAGGATTTATCCCAGCCTGCTTAAGAGG CTCCTAGATGAGAGGCAGGAAATAAAGAGGAAGATGAAGGCCTCTAAAGACCCGATCG

	AGAAGAAGATGCTCGATTACAGGCAGAGGCGTATCAAGATCCTGGCAAGCAGCTACTA CGGTTACTACGGCTATGCAAAAGCGCGCTGGTACTGCAAGGAGTGTGCAGAGAGCGTA ACGGCCTGGGGAAGGGAGTACATAGAGTTCGTAAGGAAGGAGCTGGAGGAAAAGTTCG GCTTTAAGGTATTATACATAGACACCGACGGACTCTACGCCACAATACCTGGAGCCAA ACCCGAAGAGATAAAAAAGAAGGCTCTAGAGTTCGTAGATTATATCAACGCCAACTT CCGGGCCTGCTTGAGCTCGAGTACGAGGGCTTCTACGTGCGCGGCTTCTTCGTACGA AGAAGAAGTATGCGTTGATAGACGAGGAAGGCAAGATAATCACGGGTGGACTTGAGAT TGTGAGGCGTGACTGGAGCGAGATAGCGAAAGAGACGCAGGCGAAAGTTCTTGAAGCT ATCCTAAAGCATGGTAACGTCGAGGAGGCCGTGAAGATAGTCAAAGAAGTTACCGAAA AGCTGAGCAAGTACGAGATACCGCCGGAGAAGCTGGTGATCTACGAGCAGATAACGAG GCCCTTACACGAGTACAAGGCAATAGGTCCCCACGTTGCCGTTGCCAAGAGGTTGGCC GCGAGAGGAGTCAAAGTGCGCCCTGGAATGGTGATAGGGTACATCGTGCTCAGGGGCG ACGGGCCAATAAGCAAGAGGGCGATACTTGCAGAGGAGTTCGACCTCGCGAAGCACAA GTACGACGCCGAGTACTACATTGAGAACCAGGTTCTCCAGCCGTTCTTAGAATTCTG GAGGCCTTCGGTTACCGCAAGGAAGACCTGCGCTGGCAGAAGACGAAACAGACAGGTT TGACGGCTTGGCTGAACATCAAGAAGAAGTGAGTCGACCTGCAGCCAAGCTTA
9°N_RSGA_ gBlock	AGCGGCCTGGTGCCGCGCGGCAGCATGATCCTCGACACTGACTACATAACCGAGAACG GAAAGCCTGTCATAAGAGTGTTCAAGAAGGAAAACGGCGAGTTTAAGATTGAGTACGA CCGGACTTTTGAACCCTACTTCTACGCCCTCCTGAAGGACGATTCTGCCATTGAGGAT GTCAAGAAGGTGACCGCCAAAAGGCACGGGACGGTTGTAAAAGTTAAGCGGGCGGAAA AGGTTCAGAAGAAGTTCCTCGGGAGACCAATTGAGGTCTGGAAACTCTACTTTAACCA TCCGCAGGACGTCCCAGCGATAAGGGACCGTATACGAGCGCATCCAGCAGTTGTGGAC ATCTACGAGTACGACATACCCTTCGCCAAGCGCTACCTCATAGACAAGGGATTAATTC CAATGGAAGGCGACGAGGAGCTGACCATGCTCGCCTTCGCGATTGCGACTCTCTACCA TGAGGGCGAGGAGTTCGGCACCGGGCCAATCCTTATGATAAGCTACGCCGACGGCAGC GAAGCCAGGGTGATAACTTGGAAGAAAATCGATCTCCCCTACGTTGACGTCGTCTCGA CGGAGAAAGAGATGATAAAGCGCTTCCCTCCGTGTTGTGCGTGAGAAAGACCCGGACGT TCTCATAACCTACAACGGCGACAACCTTCGACTTCGCCTATCTGAAAAAGCGCTGTGAA

	GAACTCGGAATAAAATTACCCCTCGGAAGGGATGGAAGCGAGCCGAAGATTCAGAGGA TGGGCGACAGGTTTGGCGTCGAAGTGAAGGGACGGATACACTTCGATCTCTATCCTGT GATAAGACGGACGATAAACCTGCCCACATACACGCTTGAGGCCGTTTATGAAGCCGTC TTCGGTAAACCGAAGGAGAAGGTTTACGCTGAGGAAATAGCGCAGGCCTGGGAAAGCG GCGAGGGCCTTGAGAGAGTCGCCCCGCTACTCGATGGAAGATGCGAAGGTCACATACGA GCTTGGGCGTGAGTTCTTCCCGATGGAGGCCAGCTTTCTCGCTTAATCGGCCAGTCC CTCTGGGACGTCTCCCGCTCCAGCACTGGCAACCTCGTTGAGTGTTCTCTCCTCAGGA AGGCCTATAAAAGGAATGAGCTGGCCCCGAACAAGCCCGATGAACGTGAGCTGGCCAG ACGTCGTGGCGGTTATGCGGGAGGCTATGTAAAAGAGCCCGAGAGAGGGTTGTGGGAT AACATAGTGTACCTAGATTTTAGATCCCTGTACCCCTCAATCATCATCACCCACAACG TCTCGCCGGATACGCTCAACAGAGAAGGATGCAAGGAATATGACGTTGCCCCAGAAGT CGGCCACAAATTCTGCAAGGACTTCCCAGGATTTATCCCGAGCCTGCTTGGAGACCTC CTAGAGGAGAGGCAGAAGATAAAGCGCAAGATGAAGGCCACGGTTGACCCGCTGGAGA AAAAGCTCCTCGATTACAGGCAGAGGCGTATCAAGATCCTGGCAAGCAGCTTCTACGG TTACTACGGCTATGCAAAAGCGCGCTGGTACTGCAAGGAGTGTGCAGAGAGCGTAACG GCCTGGGGAAGGGAGTACATAGAAATGGTGATCCGCGAGCTGGAGGAAAAGTTTCGGCT TTAAGGTACTGTACGCGGACACCGACGGACTGCATGCCACAATACCTGGAGCCGATGC TGAAACCGTCAAAAAGAAGGCTAAAGAGTTCTCTCAAGTATATCAACCCGAAACTTCCG GGCCTGCTTGAGCTCGAGTACGAGGGCTTCTACGTGCGCGGCTTCTTCGTACGAAGA AGAAGTATGCGGTGATAGACGAGGAAGGCAAGATAACAACGGGTGGACTTGAGATTGT GAGGCGTGACTGGAGCGAGATAGCGAAAGAGACGCAGGCGAGGGTTCTTGAAGCTATT CTAAAGCATGGTGACGTCGAGGAAGCCGTGAGGATAGTCAAAGAAGTTACCGAAAAGC TGAGCAAGTACGAGGTTCCGCCGGAGAAGCTGGTGATCCACGAGCAGATAACGAGGGA TTTACGTGACTACAAGGCAACCGGTCCCCACGTTGCCGTTGCCAAGAGGTTGGCCGCG AGAGGAGTCAAATACGCCCTGGAACGGTGATAAGCTACATCGTGCTCAAGGGCTCTG GGAGGATAGGCGACAGGGCGATACCGGCGGACGAGTTTCGACCCGGCGAAGCACCGTTA CGACGCCGAGTACTACATTGAGAACCAGGTTCTCCCAGCCGTTGAGAGAATTCTGAAA
--	--

	GCCTTCGGTTACCGCAAGGAAGACCTGCGCTACCAGAAGACGAAACAGGTTGGTTTGG GCGCTTGGCTGAAGGTTAAGGGAAAATGAGTCGACCTGCAGCCAAGCTTA
Kod-RSGA + DE mutations (R6-10)	GAGTTCCTTCCGATGGAGGCCAGCTTTCTCGCTTAATCGGCCATTCCCTCTGGGACG TCTCCCGCTCCAGCACTGGCAACCTCGTTGAGTGGTTCCTCCTCAGGAAGGCCTATGA GAGGAATGAGCTGGCCCCGAACAAGCCCGATGAAAAGGAGCTGGCCAGAAGACGGCAG AGCTATGAAGGAGGCTATGTAAAAGAGCCCGAGAGAGGGTTGTGGGAGAACATAGTGT ACCTAGATTTTAGATCCCTGTACCCCTCAATCATCATCACCCACAACGTCTCGCCGGA TACGCTCAACAGAGAAGGATGCAAGGAATATGACGTTGCCCCACAGGTCGGCCACCGC TTCTGCAAGGACTTCCCAGGATTTATCCCGAGCCTGCTTGGAGACCTCCTAGAGGAGA GGCAGAAGATAAAGAAGAAGATGAAGGCCACGATTGACCTGATCGAGAGGAAGCTCCT CGATTACAGGCAGAGGCGTATCAAGATCCTGGCAAGCGGCTACTACGGTTACTACGGC TATGCAAGGGCGCGCTGGTACTGCAAGGAGTGTGCAGAGAGCGTAACGGCCTGGGGAA GGGAGTACATAACGATGACCATCAAGGAGATAGAGGAAAAGTACGGCTTTAAGGTAAT CTACAGCGACACCGACGGATTTTTTGGCCCAATACATGGAGCCGATGCTGAAACCGTC AAAAAGAAGGCTATGGAGTTCCTCAAGTATATCAACGCCAAACTTCCGGGCGCGCTTG AGCTCGAGTACGAGGGCTTCTACAAACGCGGCTTCTTCGTCACGAAGAAGAAGTATGC GGTGATAGACGAGGAAGGCAAGATAACAACGGGTGGACTTGAGATTGTGAGGCGTGGC TGGAGCGAGATAGCGAAAGAGACGCAGGCGAGGGTTCTTGAAGCTTTGCTAAAGGACG GTGACGTCGAGAAGGCCGTGAGGATAGTCAAAGAAGTTACCGAAAAGCTGAGCAAGTA CGAGGTTCCGCCGGAGAAGCTGGTGATCCACGAGCAGATAACGAGGGATTTAAAGGAC TACAAGGCAACCGGTCCCCACGTTGCCGTTGCCAAGAGGTTGGCCGCGAGAGGAGTCA AAATACGCCCTGGAACGGTGATAAGCTACATCGTGCTCAAGGGCTCTGGGAGGATAGG CGACAGGGCGATATCGTTCGACGAGTTCGACCCGGCGAAGCACAAGTACGACGCCGAG TACTACATTGAGAACCAGGTTCTCCCACCCGTTGAGAGAATTCTGAGAGCCTGCGGTT ACCGCAAGGAAGACCTGCGCTACCAGAAGACGAGACAGGTTGGTTTGAGTGCTTGGCT GACGCCGAAGGGAACCTTGAGTCGACCTGCAGCCAAGCTTAATTAGCTGAGCTTGGACT 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	CCGCCGAGAGAAGCTGGTGATC
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/ACAACCATTATCTAGAGCGATTTCGTATAGGTGGTATCCGAAAGGATACCACC
QP08.Iowa	ATGGTTGT/3IABkFQ/
PBS8.20mer	/5IRD680/GTCCCCCTTG
Template 4	CATACGCGAGATACACTGCCAGAAAACCGCGTGTCTACGAGGTGGTATCCCCAAGGGG AC
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	ACGCGCTTCTTGGAACCTCACGCACCCATAATAATACTAAGGTGGTATCC
PBS37	ATTGCTTGCCGAGCTTATCGGAAGACTCACAGGCCTCCG
PBS33	TCGTTCTCGGACGGAGAGAT
PBS38	ATTGCTTGCCGAGCTTATC
CMM595.Fidelity	TCGTTCTCGGACGGAGAGATACGCGCTTCTTGGAACCTCACGCACCCATAATAATACT AACGGAGGCCTGTGAGTCAACC/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	ATTTTTTGGCCacaATACATGGAGCCGATG
10-92_P548T_Rvs	CCGTCGGTGTCCGCG
10-92_H550P_Fwd	TGCCCCAATAcctGGAGCCGATG
10-92_H550P_Rvs	AAAAATCCGTCCGGTGTC
10-92_G615D_Fwd	TGTGAGGCGTgacTGGAGCGAGA
10-92_G615D_Rvs	ATCTCAAGTCCACCCG
10-92_S717P_Fwd	CAGGGCGATAccgTTCGACGAGT
10-92_S717P_Rvs	TCCCCTATCCTCCAGAG
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

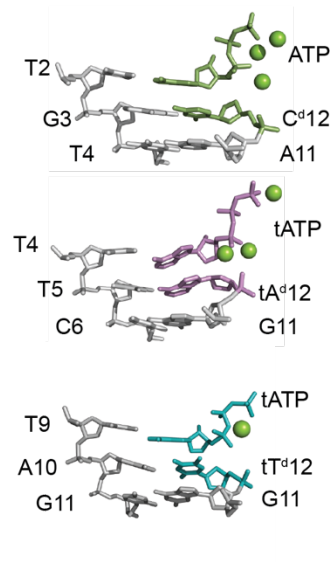
Supplementary Table 4. Data collection and refinement statistics.

10-92	
<i>Data Collection</i>	
Space group	$P2_12_12_1$
Cell Dimensions	
a, b, c (Å)	76.869, 99.525, 110.727
α, β, γ (°)	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)
<i>Refinement</i>	
Resolution (Å)	2.73
No. reflections	22979 (2236)
$R_{\text{work}}/R_{\text{free}}$	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 (°)	0.53
Number of TLS groups	3

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

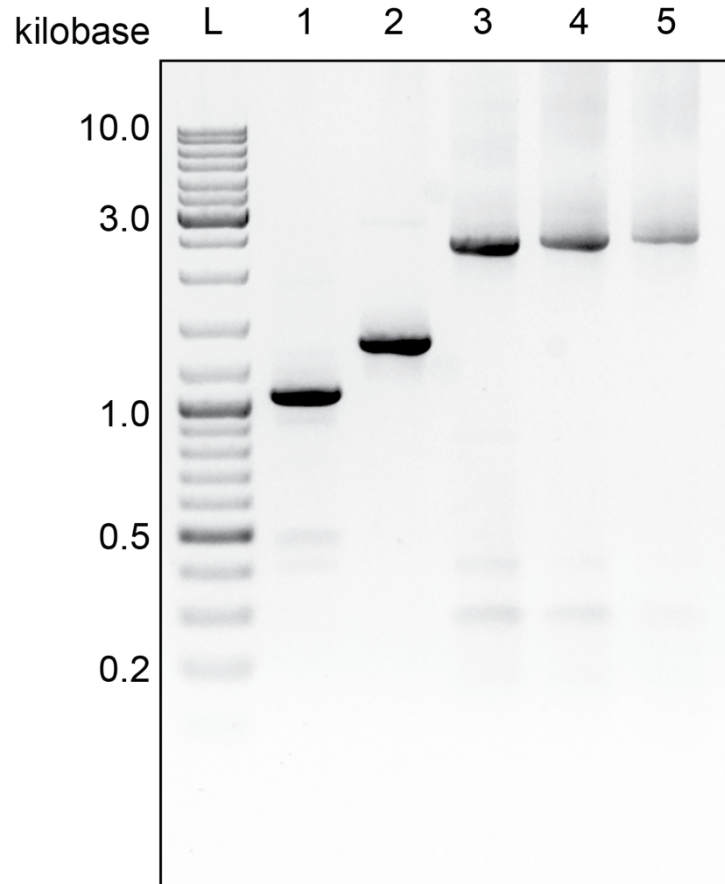
Supplementary Table 5. Local base-pair parameters²¹.

		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

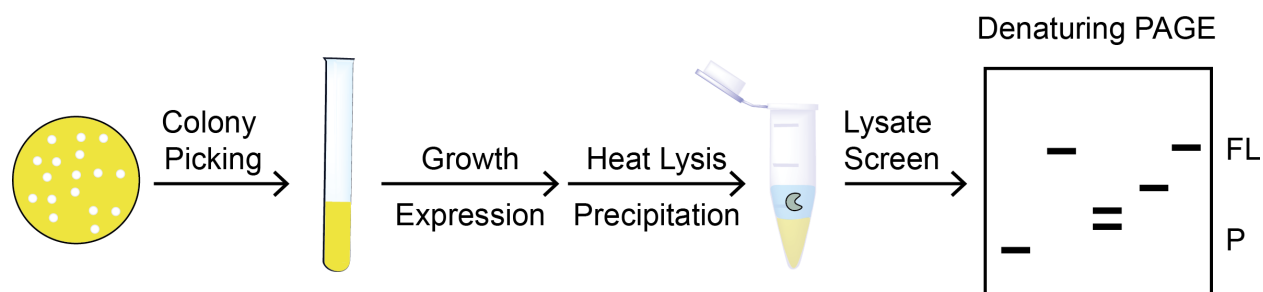


		20		40		60		80	
Kod-RSGA	MILDTDYITE	DGKPVIRIFK	KENGEFKIEY	DRTFEPYFYA	LLKDDSAIEE	VKKITAERHG	TVVTVKRVEK	VQKKFLGRPV	80
9N-RSGA	N.....V.....	D.....N.....I.....	D.....	D.....	V...K.....	K...A.....	K.....	I.....	80
Tgo-RSGA	A.....	I.....	V.....N.R.....I.....	Q.D.....	R.....	K.I.RIIDA.....	R.....	I.....	80
DV-RSGA	A.....	I.....	V.....N.R.....I.....	Q.D.....	R.....	K.I.RIIDA.....	R.....	I.....	80
		100		120		140		160	
Kod-RSGA	EVWKLTFTHP	QDVPAIRDKI	REHPAVIDIY	EYDIPFAKRY	LIDKGLVPME	GDEELKMLAF	AIATLYHEGE	EFAEGPILMI	160
9N-RSGA	N.....	R.....A.....V.....	I.....	I.....	T.....	K.....	GT.....	I.....	160
Tgo-RSGA	R.....E.....	K.....S.....V.....F.....	I.....	I.....	L.....	K.....I.....	I.....	I.....	160
DV-RSGA	R.....E.....	K.....S.....V.....F.....	I.....	I.....	L.....	K.....I.....	I.....	I.....	160
		180		200		220		240	
Kod-RSGA	SYADEEGARV	ITWKNVDLPY	VDVVS TEREM	IKRFLRVVKE	KDPDVLITYN	GDNFDFAYLK	KRCEKLGINF	ALGRDGSEPK	240
9N-RSGA	GSE.....	KI.....	K.....R.....	R.....	S.....E.....K.....T.....	S.....VK.....I.....E.....	I.....E.....	I.....	240
Tgo-RSGA	E.K.....	KI.....	E.....S.....	K.IR.....	I.....	S.LP.....V.....A.....KL.....P.....	P.....	P.....	240
DV-RSGA	E.K.....	KI.....	E.....S.....	K.IR.....	I.....	S.LP.....V.....A.....KL.....P.....	P.....	P.....	240
		260		280		300		320	
Kod-RSGA	IQRMGDRFAV	EVKGR IHFDL	YPVIRRTINL	PTYTLEAVYE	AVFGQPKEKV	YAE EITTAW	TGENLERVAR	YSMEDAKVTY	320
9N-RSGA					K.....	AQ.....	S.....G.....		320
Tgo-RSGA					I.....	AQ.....	G.....		320
DV-RSGA	M.L.L.MT.....	I.....	H.....	I.....K.....	H.....AE.....	K.G.....	K.....		320
		340		360		380		400	
Kod-RSGA	ELGKEFLPME	AQLSRLIGQS	LWDVSRSSSTG	NLVEWFLLRK	AYERNELAPN	KPDEKELARR	RQSYEGGYV	KEPERGLWEN	399
9N-RSGA	R.....F.....	F.....	F.....	F.....	K.....	R.....	GG.A.....	D.....	399
Tgo-RSGA	R.....F.....	F.....	F.....	F.....	K.....	R.....	E.A.....		399
DV-RSGA	R.....F.....	F.....	F.....	F.....	K.....	R.....	E.A.....		399
		420		440		460		480	
Kod-RSGA	IVYLDFRSLY	PSIIITHNVS	PDTLNREGCK	EYDVAPQVGH	RFCKDFPGFI	PSLLGDLLEE	RQKIKKKMKA	TIDPIERKLL	479
9N-RSGA				E.....	K.....	K.....	R.....	V.L.K.....	479
Tgo-RSGA				E.....	K.....	K.....	V.....	K.....	479
DV-RSGA	L.S.....		R.....	E.....	K.....	KR.D.....	E.R.....	SK.....K.M.....	480
		500		520		540		560	
Kod-RSGA	DYRQRRIKIL	ASSYGYGY	ARARWYCKEC	AESVTAWGRE	YITMTIKEIE	EKYGFKVIYS	DTDGGFFATIP	GADAETVKKK	559
9N-RSGA	F.....	K.....	K.....	Q.....	E.V.R.L.....	F.....L.A.....	LH.....		559
Tgo-RSGA	F.....	K.....	K.....	Q.....	ET.R.....	F.....L.A.....			559
DV-RSGA	F.....	K.....	K.....	Q.....	EFVR.L.....	F.....L.I.....	LY.....	K.P.E.I.....	560
		580		600		620		640	
Kod-RSGA	AMEFLKYINA	KLPGALELEY	EGFYKRGFFV	TKKKYAVIDE	EGKITIGGLE	IVRRDWSEIA	KETQARVLEA	LLKDG DVEKA	639
9N-RSGA	K.....P.....	L.....	V.....	L.....	D.....			I.H.....E.....	639
Tgo-RSGA	K.....D.....	L.....	V.....	L.....	D.....			I.H.....E.....	639
DV-RSGA	L.VD.....	L.....	V.....	L.....	D.....			I.H.N.....E.....	640
		660		680		700		720	
Kod-RSGA	VRIVKEVTEK	LSKYEVPEPK	LVIHEQITRD	LKDYKATGPH	VAVAKRLAAR	GVKIRPGTVI	SYIVLKGSGR	IGDRAIPFDE	719
9N-RSGA				R.....				A.....	719
Tgo-RSGA			Y.....			I.....			719
DV-RSGA	K.....	I.....	Y.....P.....	HE.....I.....		V.....M.....	G.....R.D.P.....	S.K.....LAE.....	720
		740		760					
Kod-RSGA	FDPAKHKYDA	EYYIENQVLP	AVERILRAFG	YRKEDLRYQK	TRQVGLSAWL	KPKGT			774
9N-RSGA	R.....	K.....	K.....	K.....	K.....G.....	V.....K.....			774
Tgo-RSGA	R.....	K.....	K.....	K.....	K.....G.....	V.....K.....			773
DV-RSGA	L.....	L.....E.....	W.....	K.T.T.....	N.I.KK.....				775

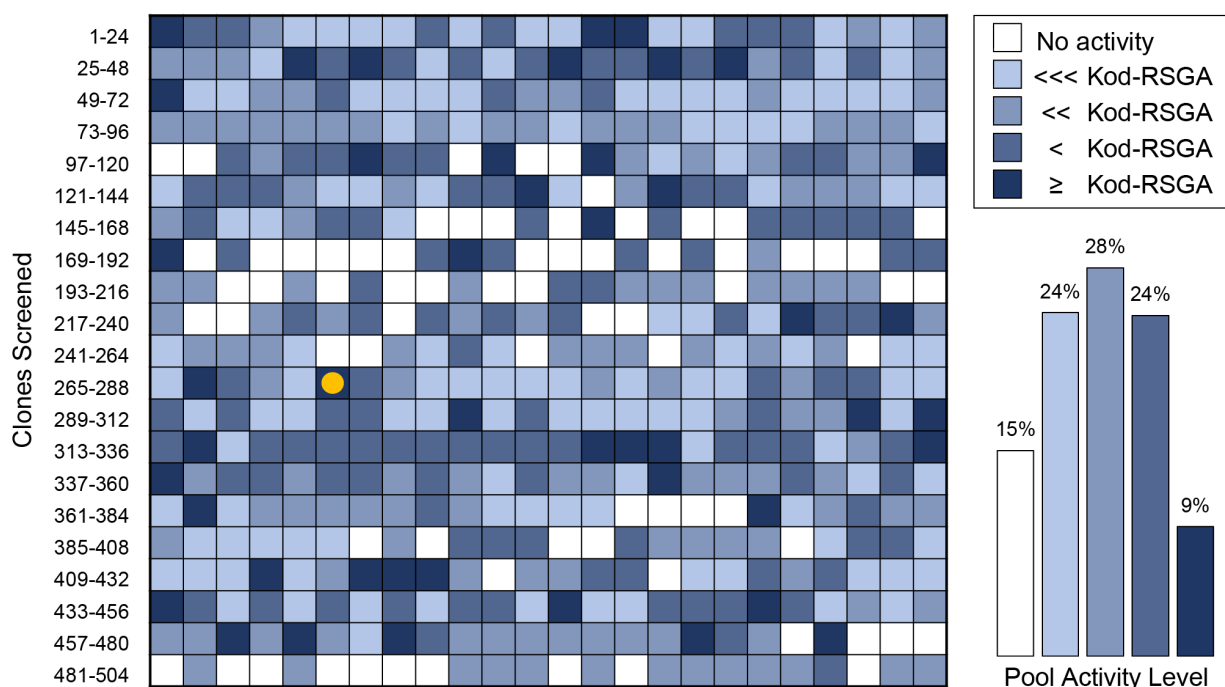
Supplementary Figure 1. Sequence alignment of archaeal RSGA DNA polymerase variants. Red Boxes indicate the positions previously discovered in Kod-RSGA (A485R, N491S, R606G, T723A) that are known to enhance TNA synthesis. Polymerases contain the exonuclease silencing mutations D141A and E143A. Conserved 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.



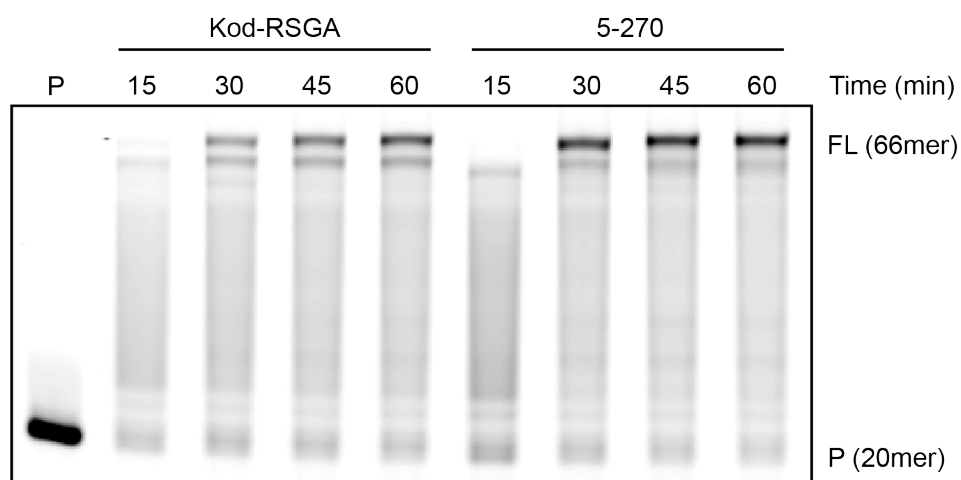
Supplementary Figure 2. Assembly of the DNA recombination library by overlap PCR. Agarose gel showing assembly of the homologous recombination library. Lane L: DNA ladder. Lane 1: Kod-RSGA fragments 1—5. Lane 2: Kod-RSGA fragments 6-11. Lane 3: Kod-RSGA fragments 1-11. Lane 4: Kod-RSGA and Tgo-RSGA fragments 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.



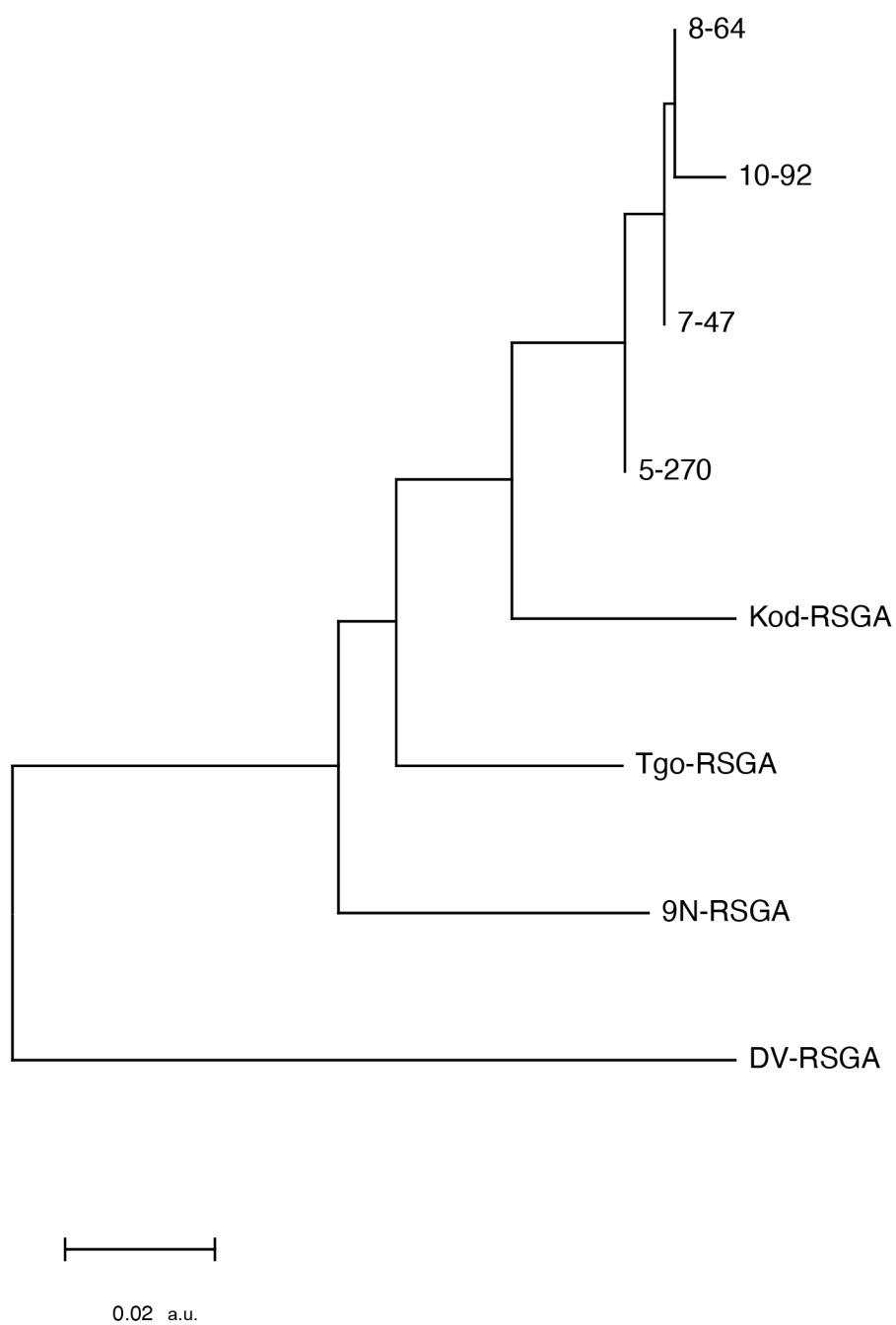
Supplementary Figure 4. Lysate activity screen following five rounds of selection. Library members present in the pool after five rounds of in vitro selection were individually screened for TNA synthesis activity in a primer-extension assay. A qualitative assessment of primer extension efficiency relative 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.



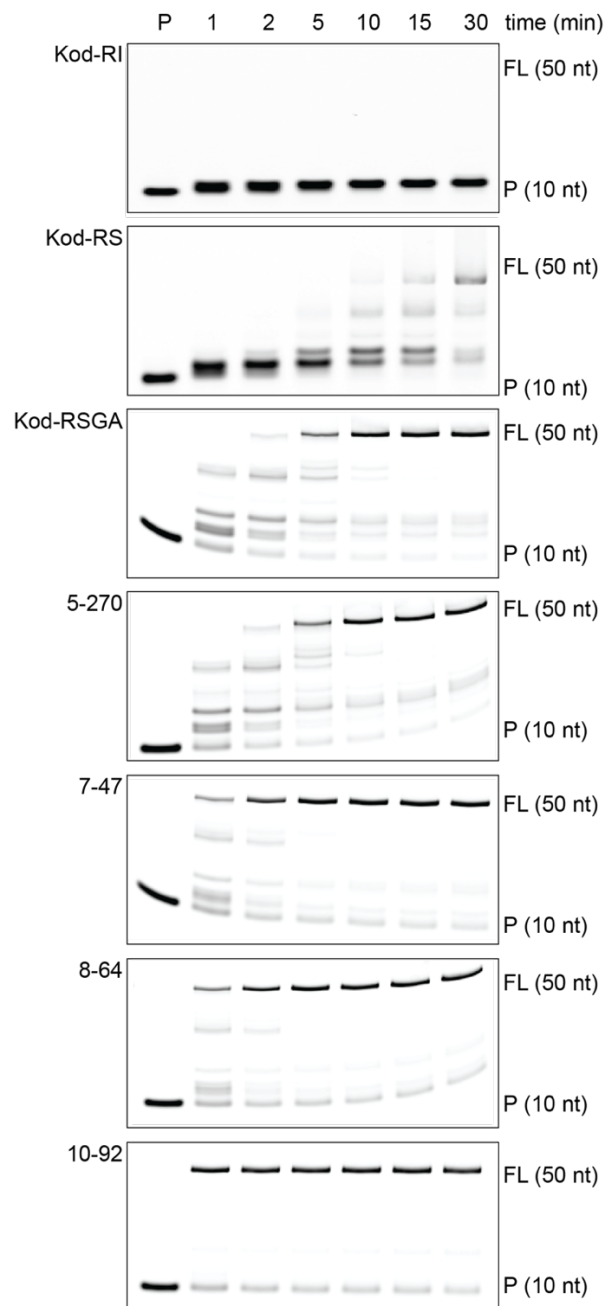
Supplementary Figure 5. TNA synthesis activity of clone 5-270. Engineered polymerases Kod-RSGA and 5-270 are challenged to incorporate TNA substrates in a primer extension assay, templated by a random DNA aptamer library (L30.5).

		20		40		60		80	
Kod-RSGA	MILDTDYITE	DGKPVIRIFK	KENGEFKIEY	DRTFEPYFYA	LLKDDSAIEE	VKKITAERHG	TVTVKRVEK	VQKKFLGRPV	80
5-270									80
7-47									80
8-64									80
10-92									80
		100		120		140		160	
Kod-RSGA	EVWKLYFTHP	QDVPAIRDKI	REHPAVIDIY	EYDIPFAKRY	LIDKGLVPME	GDEELKM LAF	AIATLYHEGE	EFAEGPI LMI	160
5-270		R. A. V			I. I	T. T			160
7-47		R. A. V			I. I	T. T			160
8-64		R. A. V			I. I	T. T			160
10-92		R. A. V			I. I	T. T			160
		180		200		220		240	
Kod-RSGA	SYADEEGARV	ITWKNVDLPY	VDVVSTEREM	IKRFLRVVKE	KDPDVLITYN	GDNFDFAYLK	KRCEKLGINF	ALGRDGSEPK	240
5-270									240
7-47									240
8-64									240
10-92									240
		260		280		300		320	
Kod-RSGA	IQRMGDRFAV	EVKGRIHFDL	YPVIRRTINL	PTYTLEAVYE	AVFGQPKEKV	YAEIITTAWE	TGENLERVAR	YSMEDAKVTY	320
5-270					K. K	AQ. G			320
7-47					K. K	AQ. G			320
8-64					K. K	AQ. G			320
10-92					K. K	AQ. G			320
		340		360		380		400	
Kod-RSGA	ELGKEFLPME	AQLSRLIGQS	LWDVSRSSSTG	NLVEWFLLRK	AYERNELAPN	KPDEKELARR	RQSYEGGYV	KEPERGLWEN	399
5-270		V. P		Y. Y		R. YE	L. E. A		400
7-47		V. P		Y. Y		R. YE	L. E. A		400
8-64		V. P		Y. Y		R. YE	L. E. A		400
10-92		V. HP		Y. Y		R. YE	L. E. A		400
		420		440		460		480	
Kod-RSGA	IVYLDFRSLY	PSIIITHNVS	PDTLNREGCK	EYDVAPQVGH	RFCKDFPGFI	PSLLGDLLEE	RQIKKKMKMA	TIDPIERKLL	479
5-270							R. V. L. K		480
7-47							R. V. L. K		480
8-64							R. V. L. K		480
10-92							R. V. LL. K		480
		500		520		540		560	
Kod-RSGA	DYRQRRIKIL	ASSYYGGYGY	ARARWYCKEC	AESVTAWGRE	YITMTIKEIE	EKYGFKVIYS	DTDGGFFATIP	GADAETVKKK	559
5-270				Q	ET. R	F. L. A			560
7-47				Q	ET. R	F. L. A			560
8-64				Q	ET. R	F. L. A			560
10-92		G		Q	ET. R	F. L. A		P. H	560
		580		600		620		640	
Kod-RSGA	AMEFLKYINA	KLPGALELEY	EGFYKRGFFV	TKKKYAVIDE	EGKITTTGGL	IVRRDWSEIA	KETQARVLEA	LLKGDGVEKA	639
5-270	K. D				D				640
7-47	K. D				D	G			640
8-64	K. D				D	G			640
10-92	K. D				D	G			640
		660		680		700		720	
Kod-RSGA	VRIVKEVTEK	LSKYEVPPEK	LVIHEQITRD	LKDYKATGPH	VAVAKRLAAR	GVKIRPGTVI	SYIVLKGSGR	IGDRAIPFDE	719
5-270				R					720
7-47				R				S	720
8-64				R				S	720
10-92				R				S	720
		740		760		774			
Kod-RSGA	FDPAKHKYDA	EYYIENQVLP	AVERILRAFG	YRKEDLRYQK	TRQVGLSAWL	KPKGT	774		
5-270							775		
7-47							775		
8-64							775		
10-92							775		

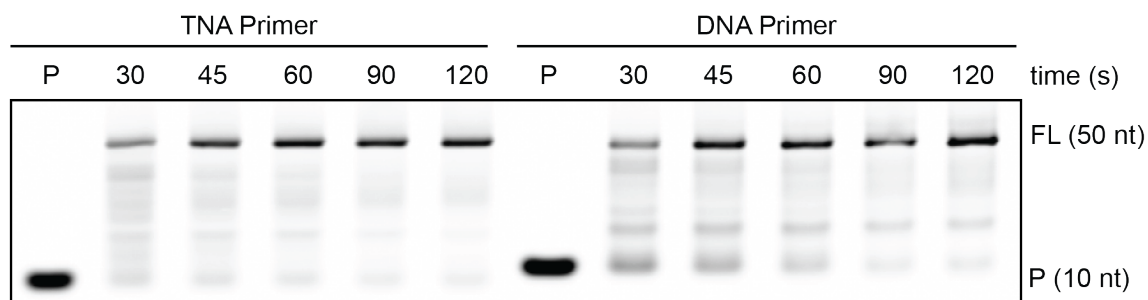
Supplementary Figure 6. Sequence alignment of newly evolved TNA polymerases. TNA polymerases identified in this study are compared to Kod-RSGA from our previous study. Conserved residues are denoted as (.), while mutations are denoted by their single letter amino acid abbreviations. The red box denotes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



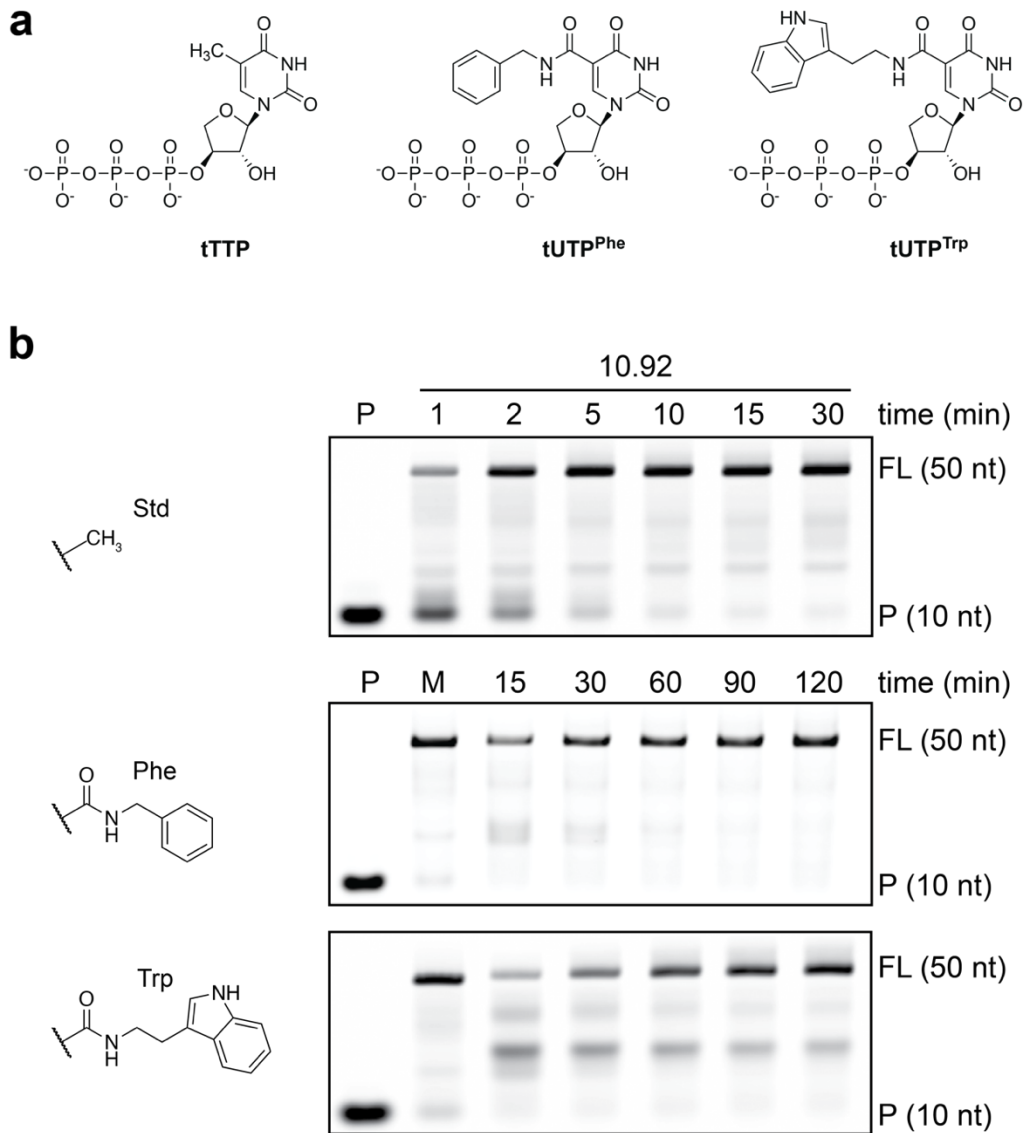
Supplementary Figure 7. Phylogenetic analysis of the evolved TNA polymerases. MEGA was used to generate a phylogenetic tree by the neighbor-joining method of the newly evolved TNA polymerases relative to their starting parent sequences.



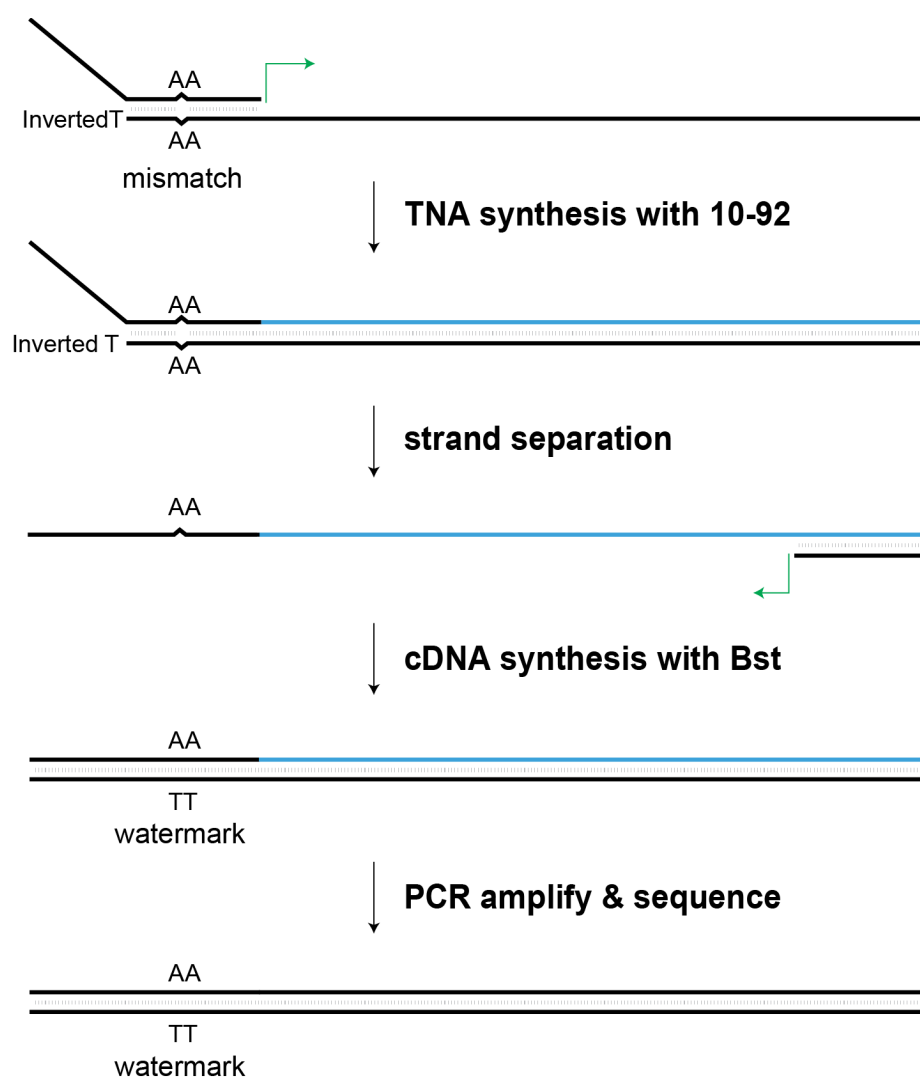
Supplementary Figure 8. Comparative activity analysis of ancestral TNA polymerases. Each polymerase was challenged to extend a DNA primer with 40 consecutive TNA residues. The reactions were quenched at defined time points and analyzed by denaturing polyacrylamide gel electrophoresis with fluorescent 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.



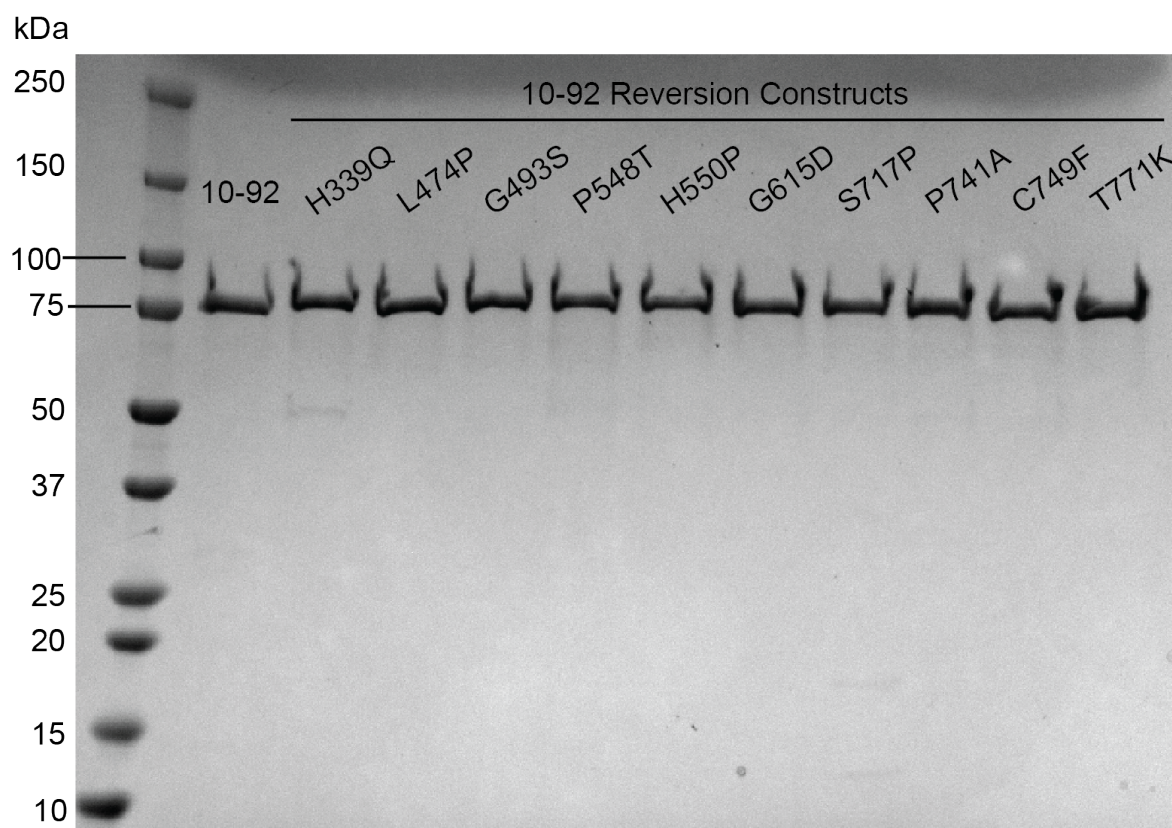
Supplementary Figure 10. Functional analysis of 10-92. (a) Chemical structures of base-modified tUTP analogs. (b) Enzymatic synthesis of base-modified TNA. Time course analysis was performed with 10-92 and tNTP mixtures containing standard bases as well as base-modified tUTP analogs. Full length product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



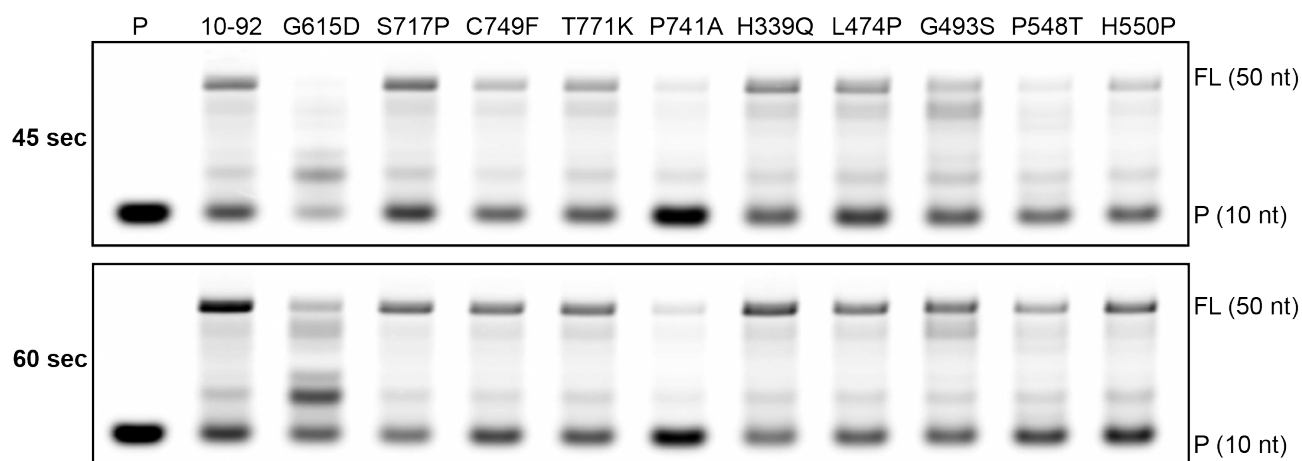
Supplementary Figure 11. Schematic of the fidelity assay. An overhang primer is designed with an AA-AA mismatch in the primer region of the DNA template (black) with a 3' inverted deoxy-T. The 3' inverted deoxy-T prevents extension into the overhang region. The TNA strand (blue) is synthesized across 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	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

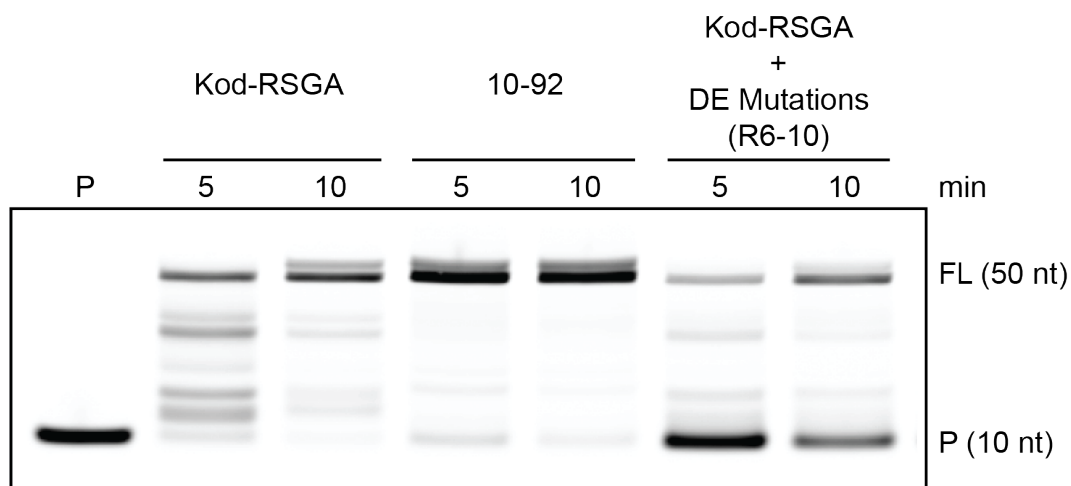
Supplementary Figure 12. Polymerase fidelity data. Raw alignment of sangar sequencing results from a cycle of TNA replication by TNA polymerase 10-92 with standard base TNA triphosphates. The expected DNA sequence (colored) is written 5'→3'. Correct nucleotide reads are denoted as (.) and deleted nucleotides 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).



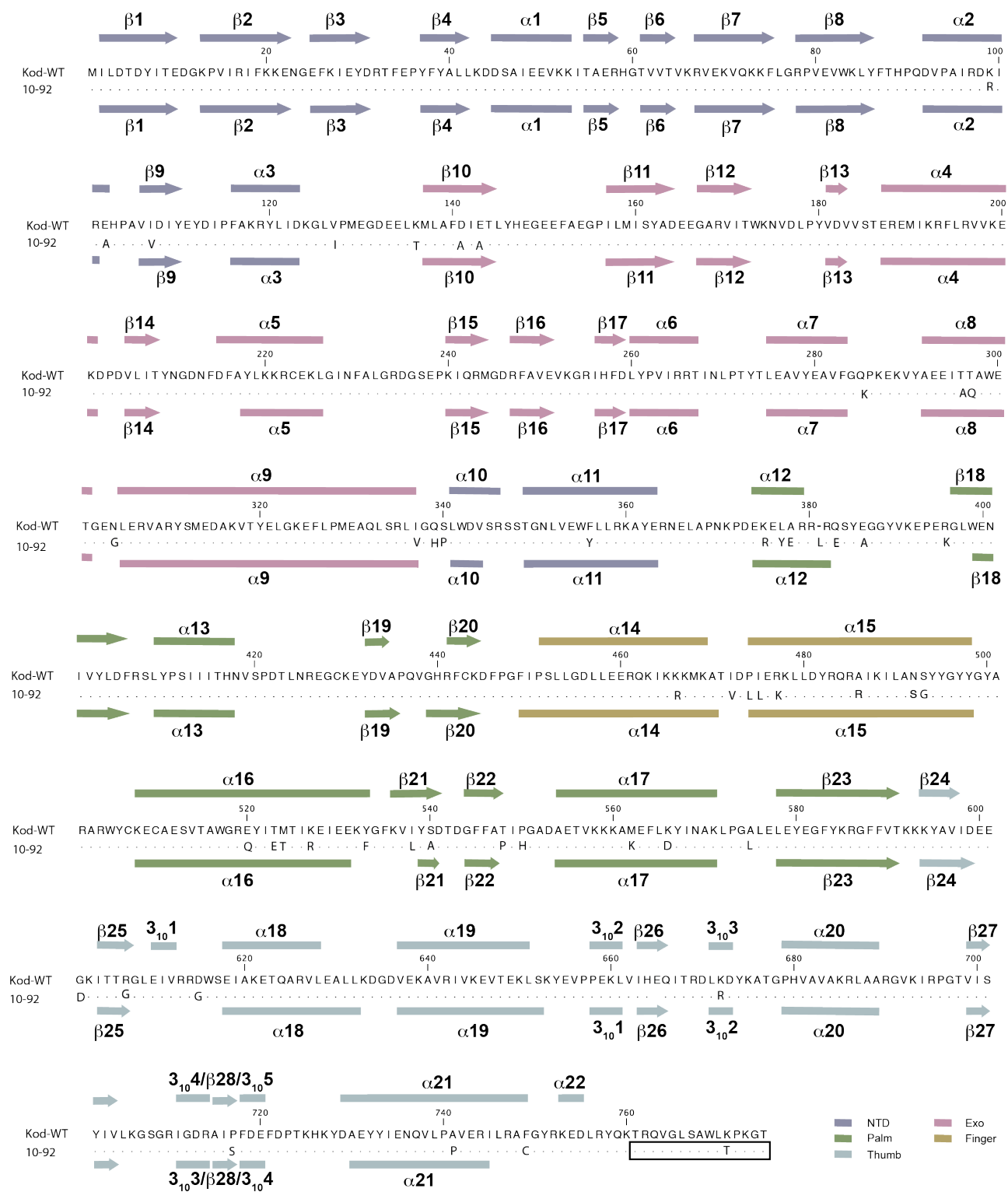
Supplementary Figure 13. SDS-PAGE analysis of 10-92 reversion proteins. A single-point reversion study was performed in which each of the 10 mutations acquired by directed evolution of 5-270 to 10-92 were individually reverted back to their wild-type residue. The resulting constructs were grown in *E. 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.



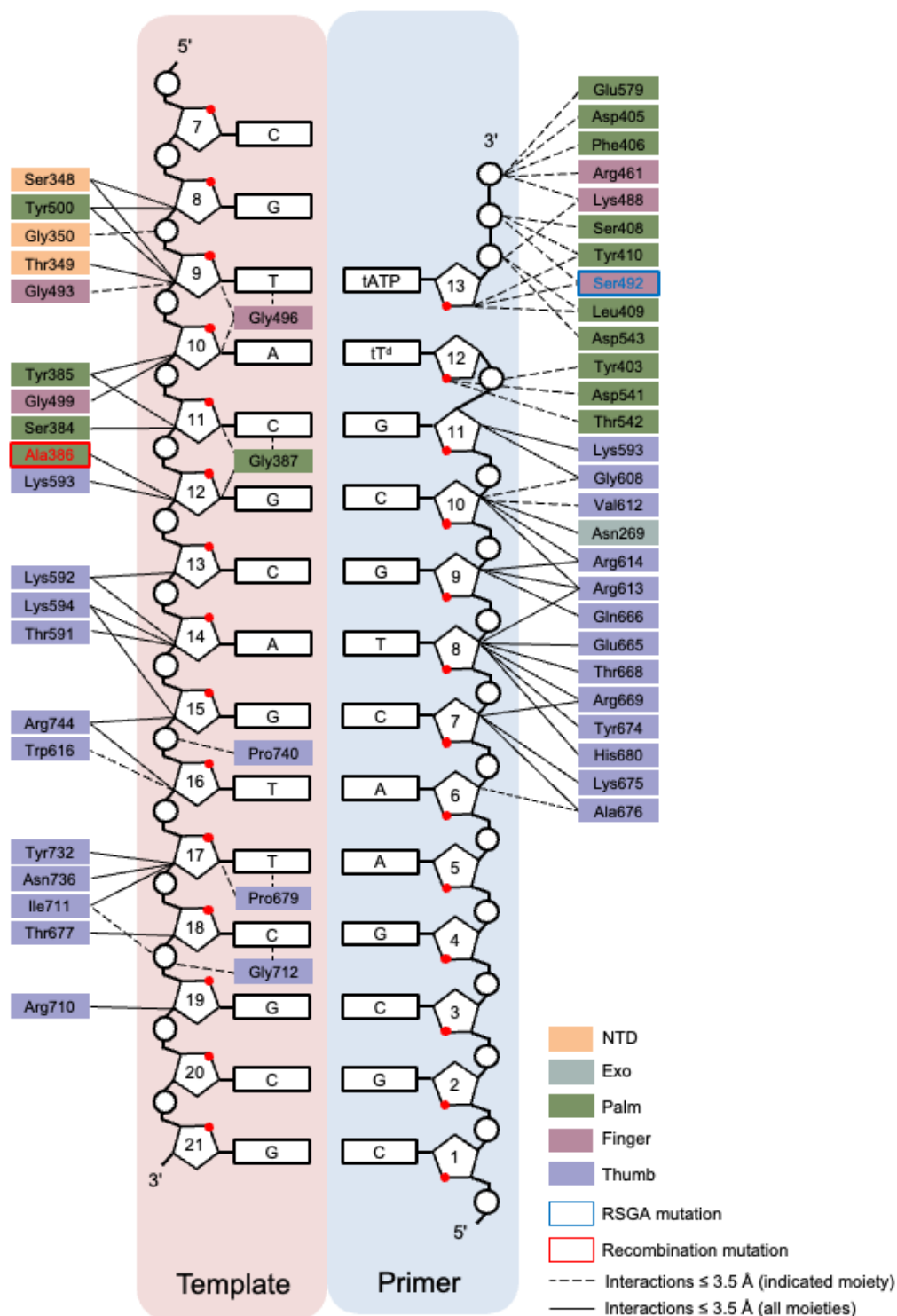
Supplementary Figure 14. TNA synthesis activity of 10-92 reversions. Single point reversions introduced into the backbone of the 10-92 TNA polymerase. The mutations acquired in the evolution of 10-92 from 5-270 were reverted to the WT residue and are listed in the order they were discovered (left to 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.



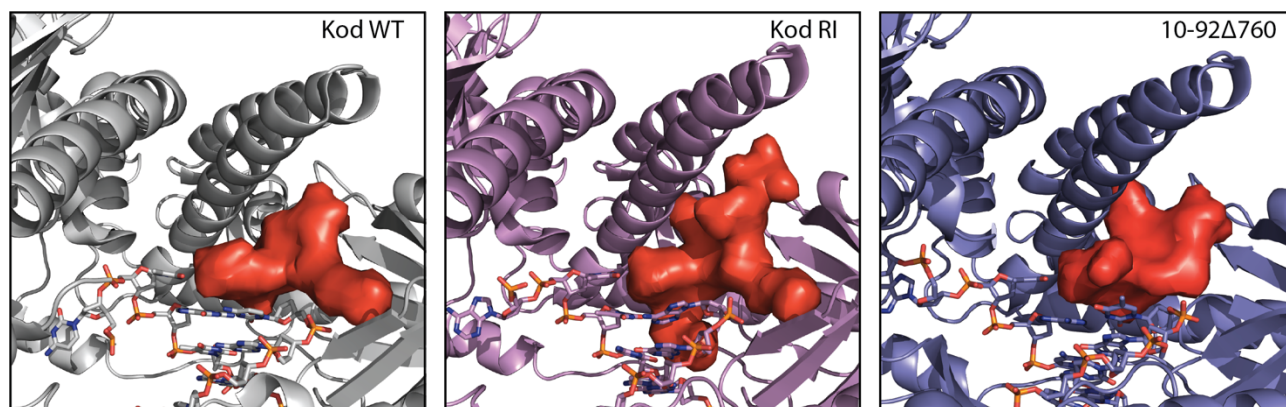
Supplementary Figure 15. Activity of Kod-RSGA with directed evolution mutations. Kod-RSGA and 10-92 are compared to an engineered form of Kod-RSGA carrying the 10 beneficial mutations acquired by directed evolution that are found in the 10-92 sequence. Each protein was challenged to copy a DNA template 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.



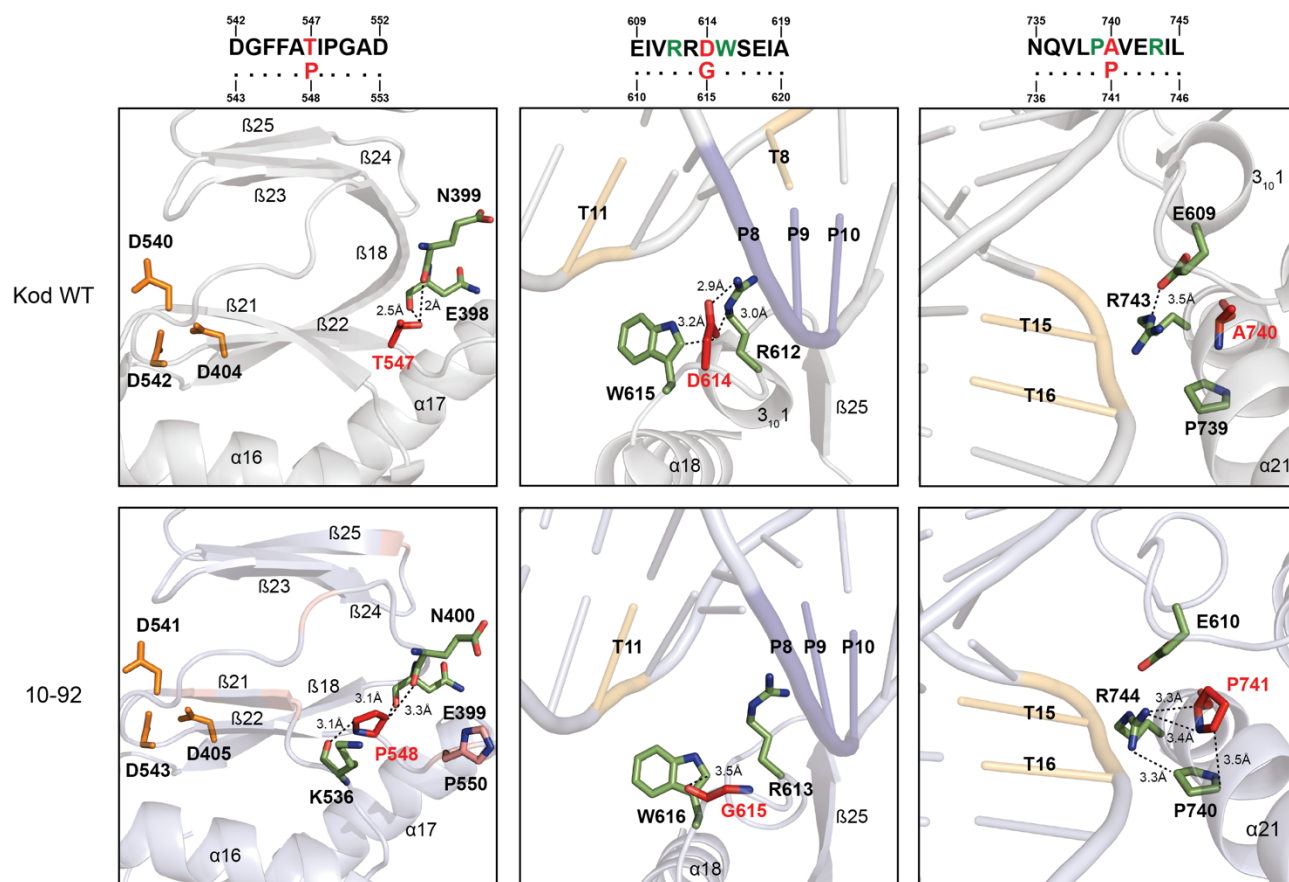
Supplementary Figure 16. Secondary structure map of the 10-92 TNA polymerase. Secondary structure comparison of Kod-WT (top) and 10-92 (bottom), as determined by DSSP and PyMol. Matching residues are shown as dots. Residues removed for crystallography are boxed in black. Secondary structural elements 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 ($500\text{\AA}^3$) 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 10-92 Δ 760 TNA polymerase (PDB: 8T3X) measure at $746\text{\AA}^3$, $1018\text{\AA}^3$, and $684\text{\AA}^3$, 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.

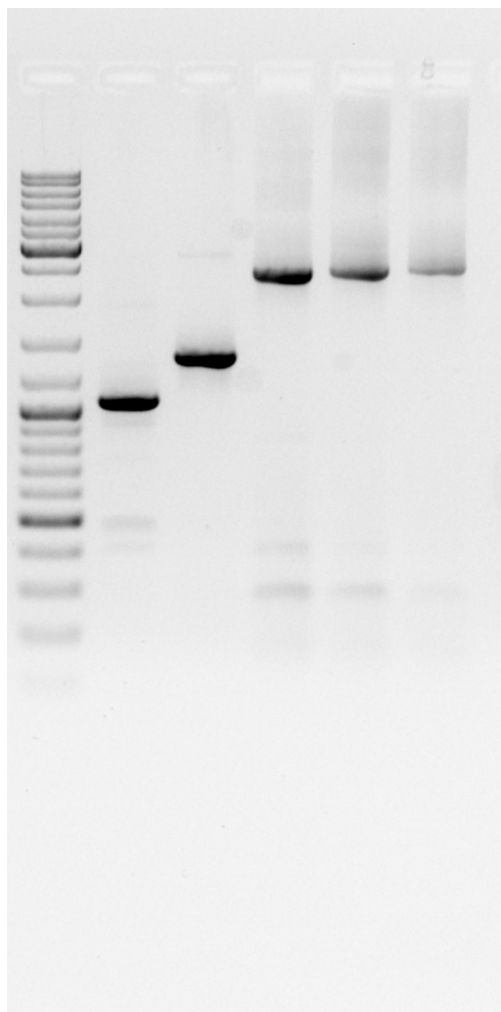
SUPPLEMENTARY REFERENCES

- 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 Synth. 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 α -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).

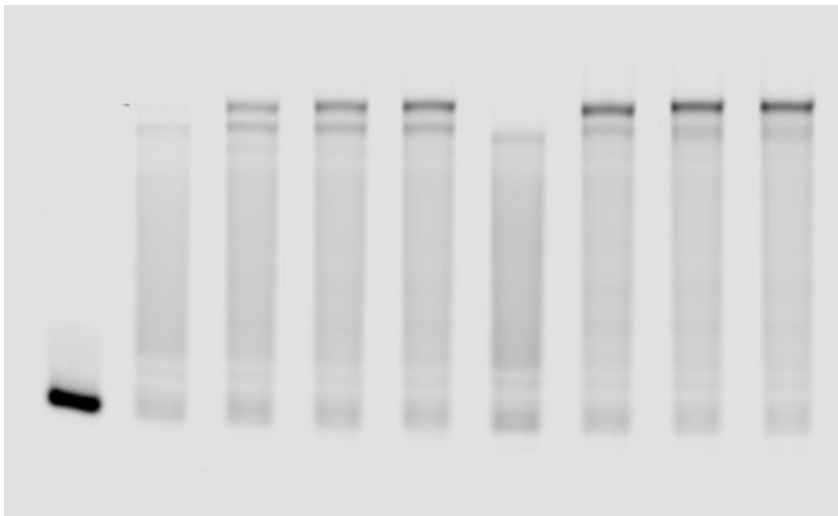
- 14 Kabsch, W. Xds. *Acta Crystallogr. D Biol. Crystallogr.* 66, 125-132, (2010).
- 15 Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* 66, 133-144, (2010).
- 16 McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* 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).

UNCROPPED GELS

Supplementary Figure 2:

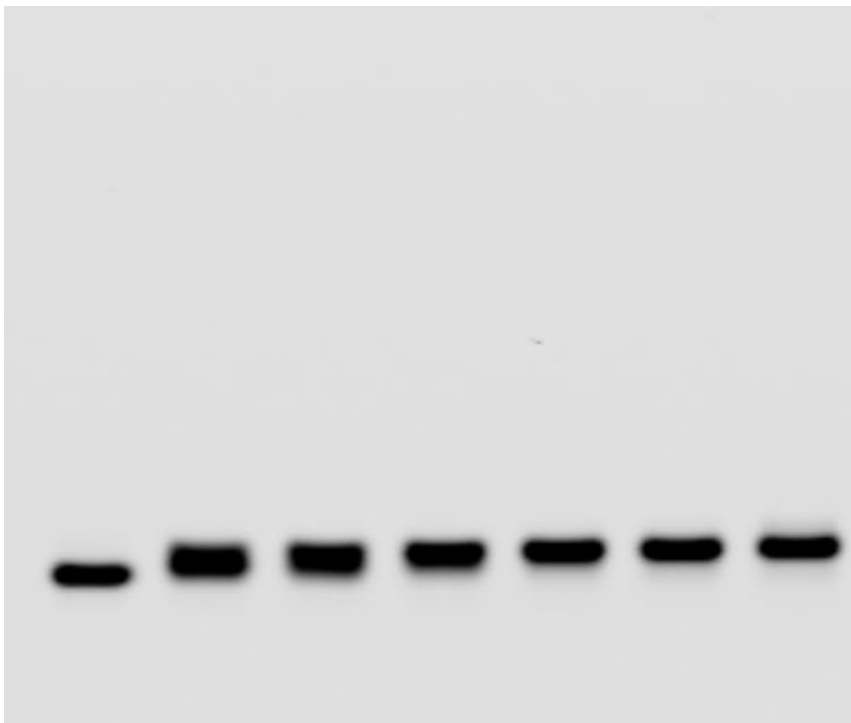


Supplementary Figure 5:

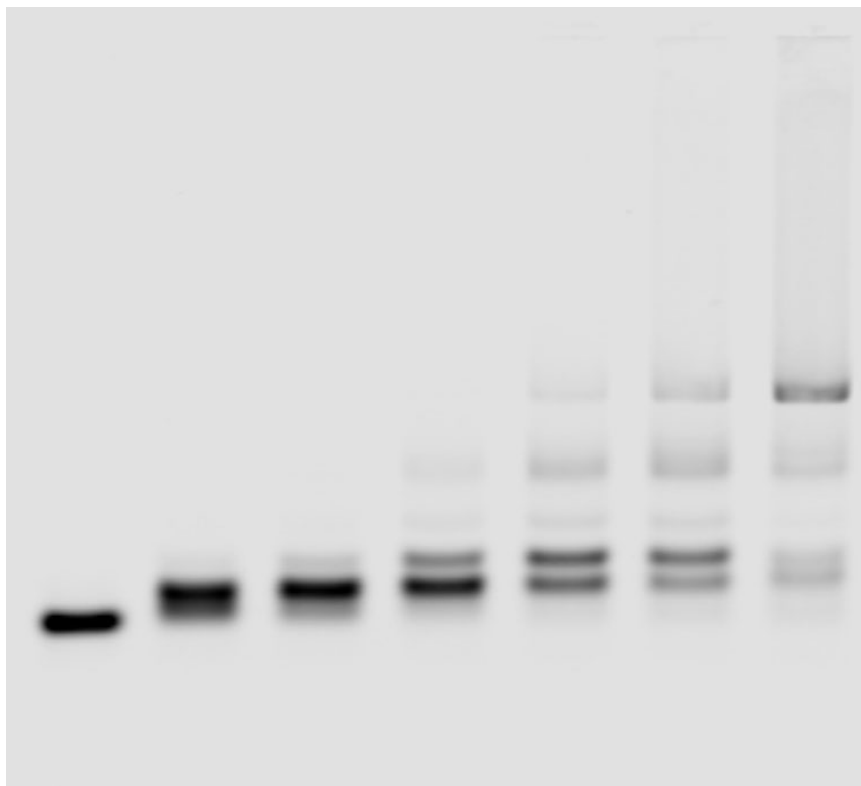


Supplementary Figure 8:

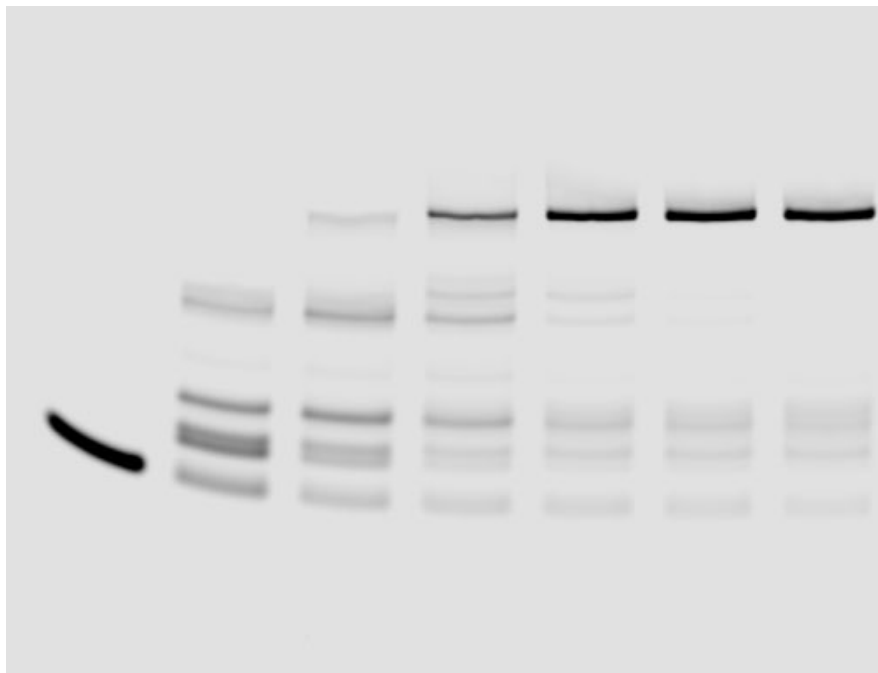
Kod RI



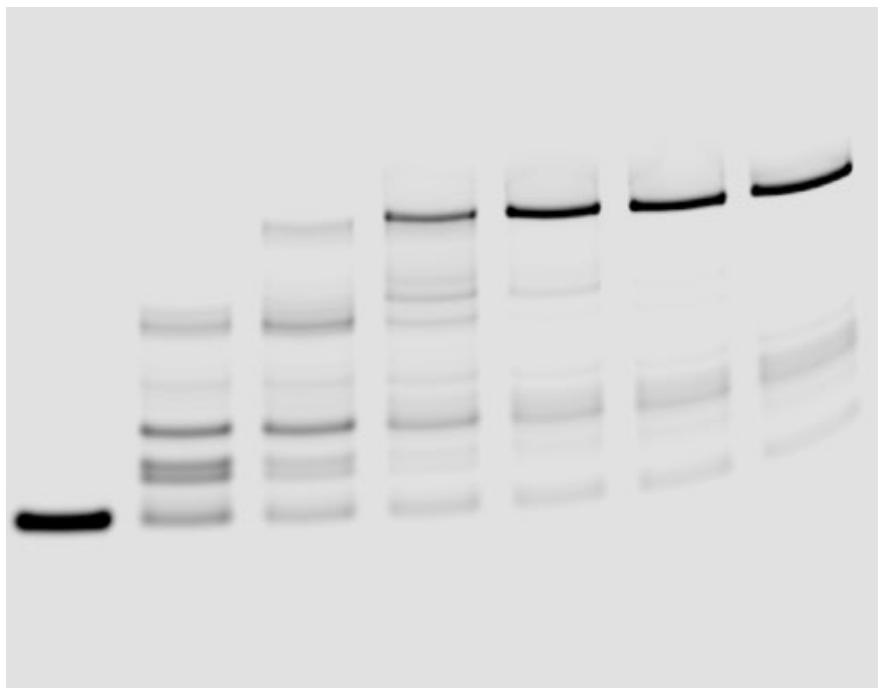
Kod RS



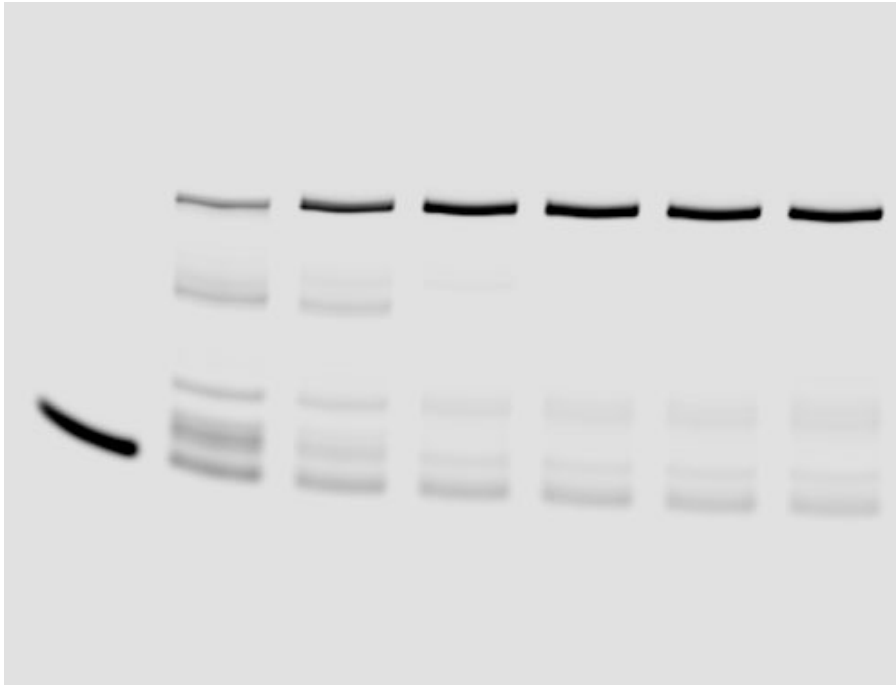
Kod RSGA



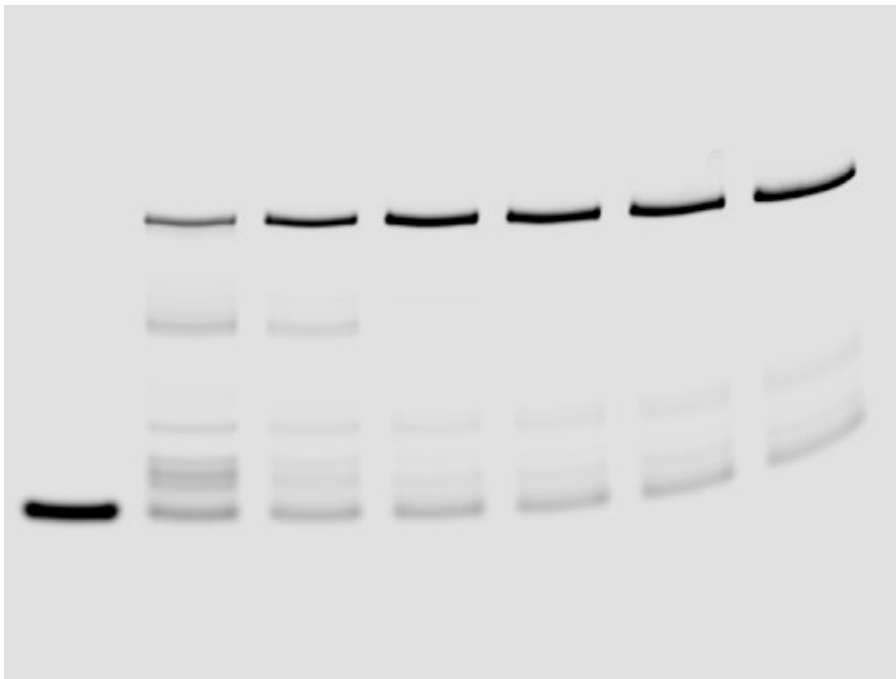
5-270



7-47



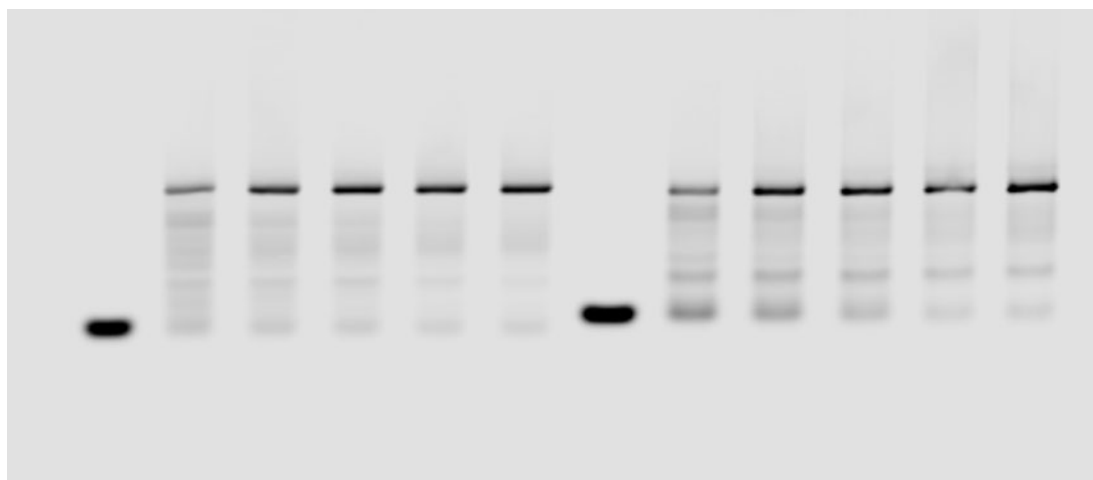
8-64



10-92

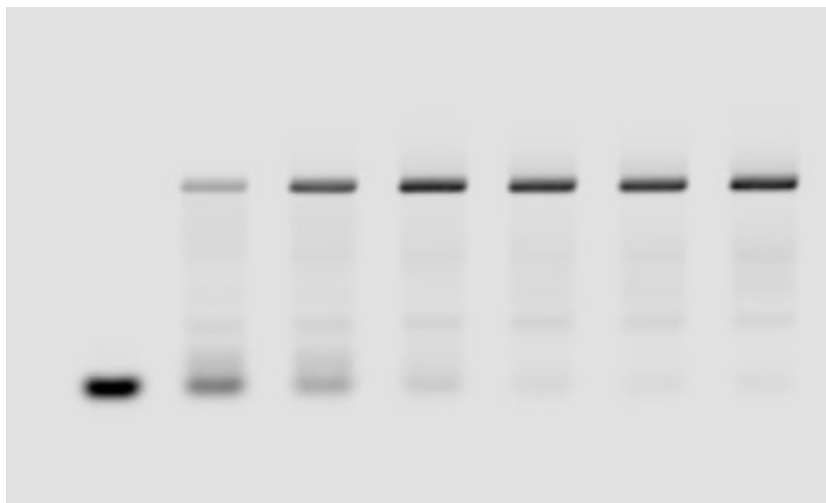


Supplementary Figure 9

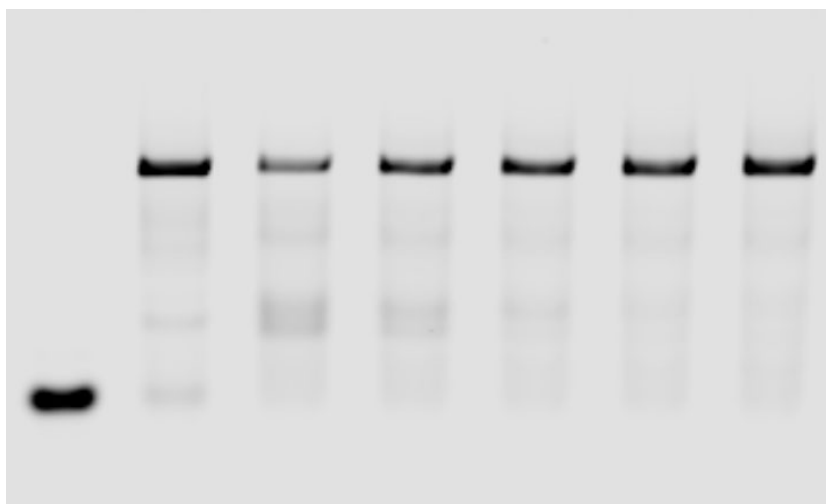


Supplementary Figure 10

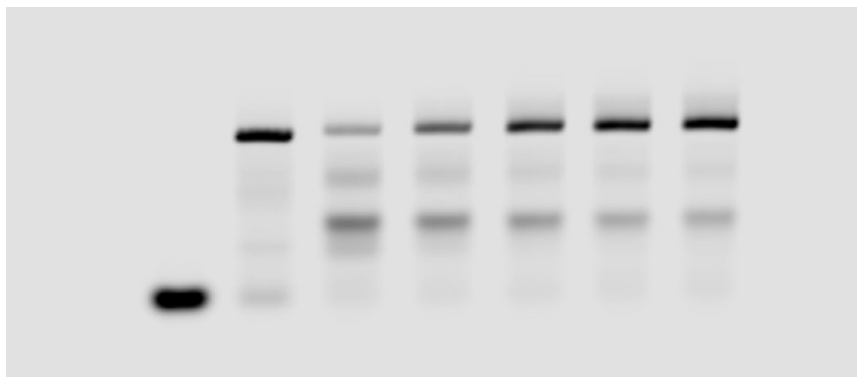
Std



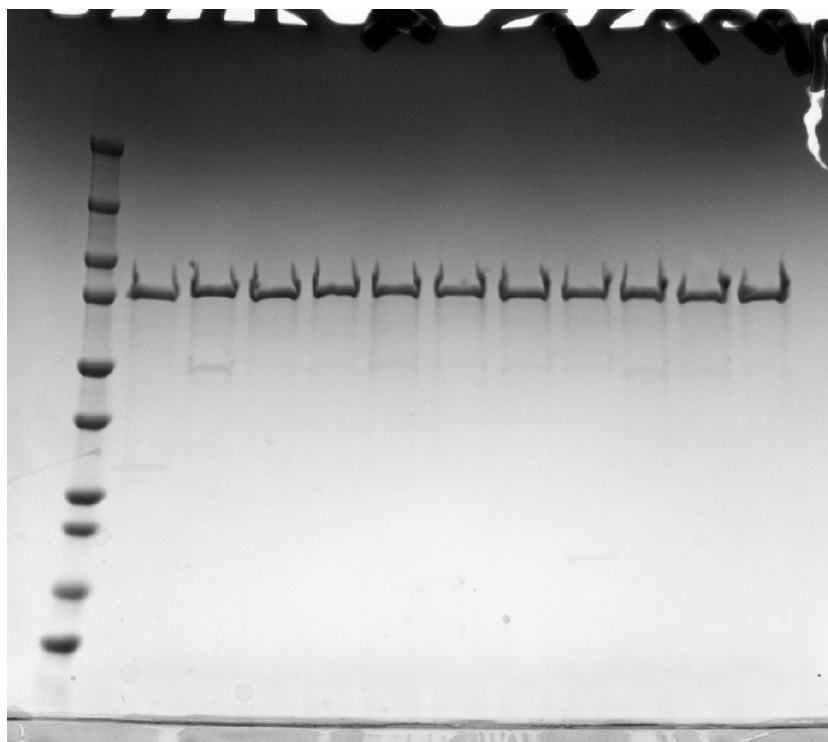
Phe



Trp

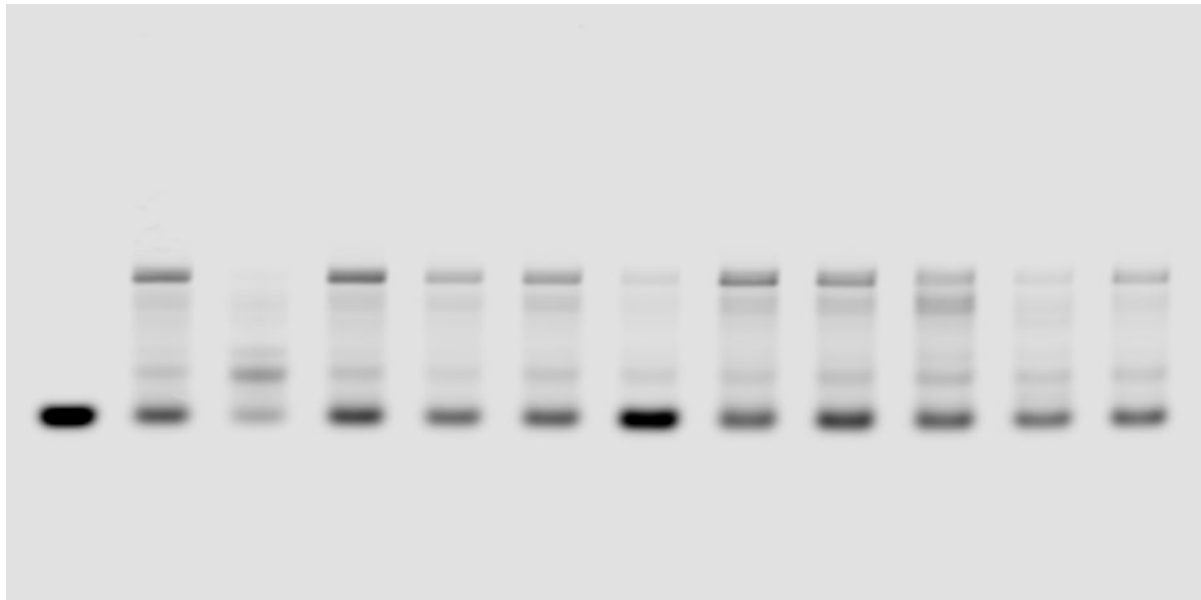


Supplementary Figure 13

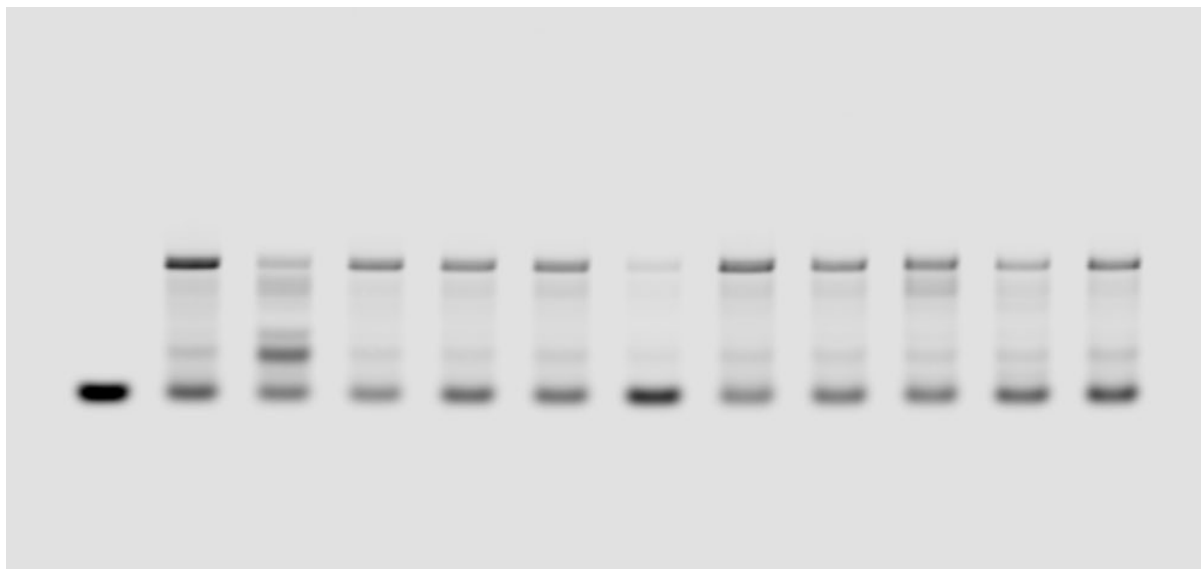


Supplementary Figure 14

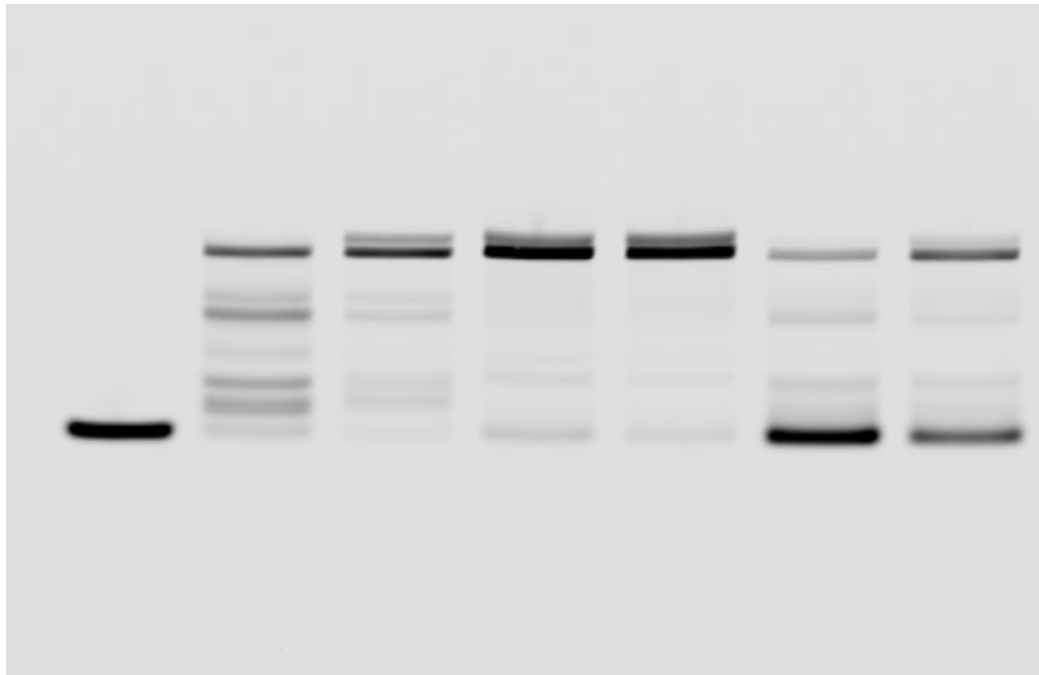
45s



60s



Supplementary Figure 15



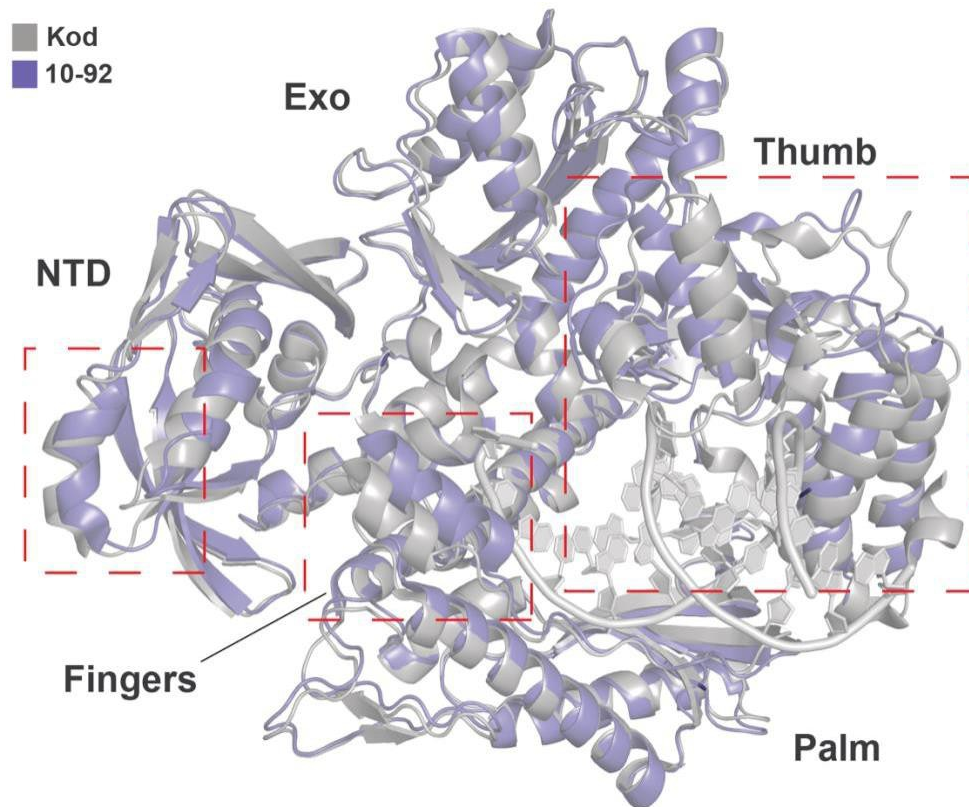
APPENDIX C: Supplementary Information for Chapter 3

Supplementary Figures

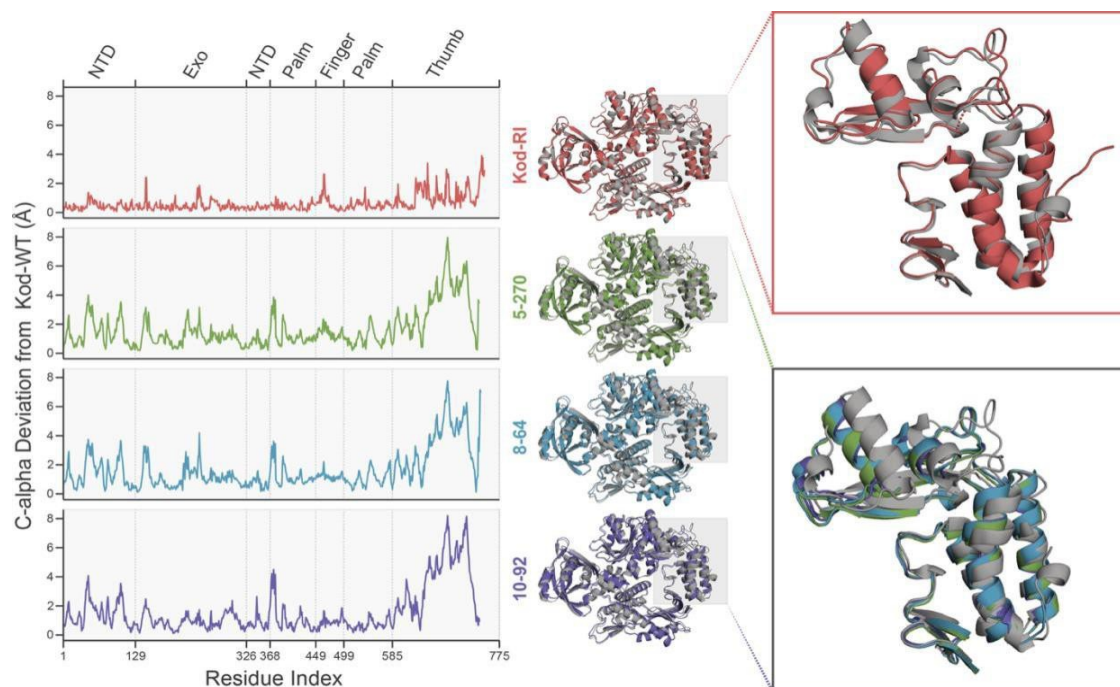
Supplementary Figure 1. Structural overlay of Kod and 10-92 closed ternary conformations.
Supplementary Figure 2. Conformational rearrangement of the thumb subdomain.
Supplementary Figure 3. Sequence homology-based fitness landscape of TNAP variants.
Supplementary Figure 4. Sequence alignment of TNAPs.
Supplementary Figure 5. Workflow for TNAP fidelity measurements in hydrogel particles.
Supplementary Figure 6. TNAP fidelity raw data and statistical analysis.
Supplementary Figure 7. TNAP kinetics and substrate specificity.
Supplementary Figure 8. TNAP thermal challenge.
Supplementary Figure 9. Polymerase domains.
Supplementary Figure 10. Sequence and secondary structure alignment of Kod and 10-92.
Supplementary Figure 11. Structural comparison of the finger subdomain of Kod, Kod-RI, Kod-RSGA, and 5-270 in the binary complex.
Supplementary Figure 12. Active site comparison.
Supplementary Figure 13. Synthesis of 2'-deoxy- α -L-threofuranosyl thymine triphosphate.
Supplementary Figure 14. ^{31}P NMR spectrum of compound 2.
Supplementary Figure 15. ^{31}P NMR spectrum of compound 3.
Supplementary Figure 16. HPLC analysis of compound 3.
Supplementary Figure 17. Multidimensional analysis of the closed ternary complex of TNAP structures based on RMSD.
Supplementary Figure 18. Distance of mutations to the active site.
Supplementary Figure 19. AlphaFold3 prediction overview.
Supplementary Figure 20. AlphaFold3 prediction confidence.
Supplementary Figure 21. AlphaFold3 prediction of the 5-270 binary structure.
Supplementary Figure 22. Comparison of AlphaFold3 predictions with experimental structures.
Supplementary Figure 23. Visualization of AlphaFold3 predictions against experimental structures.
Supplementary Figure 24. Active site pocket volume.
Supplementary Figure 25. 2Fo-Fc electron density maps of the active site.
Supplementary Figure 26. Polder maps of key residues in the active site.
Supplementary Figure 27. Active site geometry comparing two independently solved structures of 5-270 in a closed ternary complex.
Supplementary Figure 28. Catalysis-enhancing mutations mapped across polymerase domains.
Supplementary Figure 29. Polder maps of mutated residues in 5-270, 8-64, and 10-92.
Supplementary Figure 30. Polder maps of mutated residues in DNAP and TNAPs relative to the catalytic aspartate residues.

Supplementary Tables

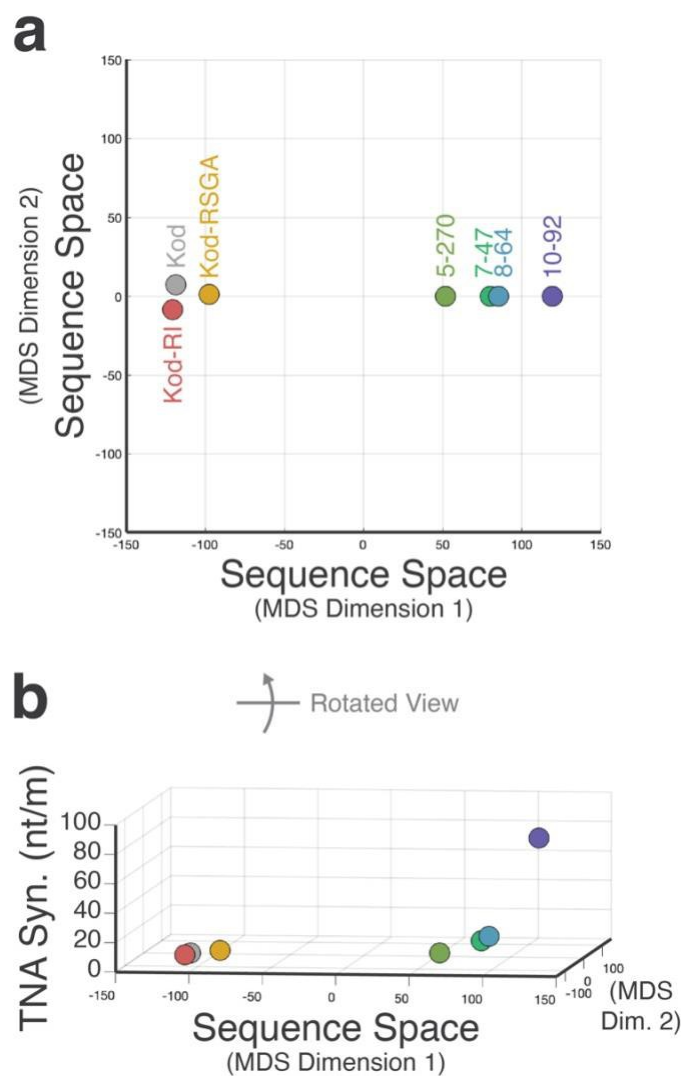
Supplementary Table 1. DNA oligonucleotides.
Supplementary Table 2. Data collection and refinement statistics for binary structures.
Supplementary Table 3. Data collection and refinement statistics for ternary structures.
Supplementary Table 4. Base pair parameters.
Supplementary Table 5. Integrated mutational analysis.



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.



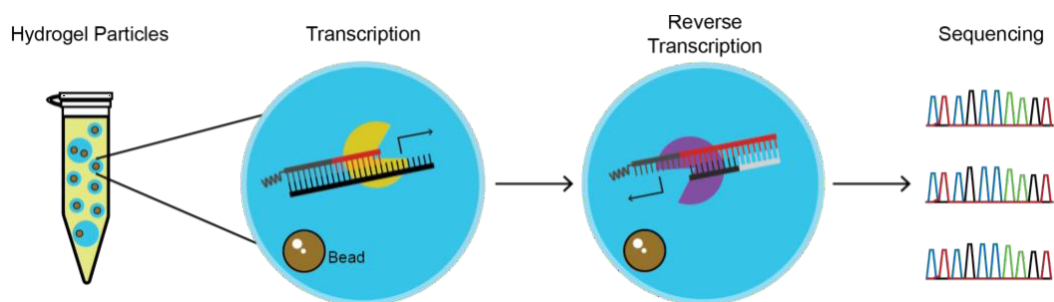
Supplementary Figure 2. Conformational rearrangement of the thumb subdomain. Left, the distance in angstroms from all resolved C- α positions in the ternary structures of Kod-RI (red), 5-270 (green), 8-64 (blue), and 10-92 (purple) relative to their position in Kod. Middle, structural overlays of TNAPs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

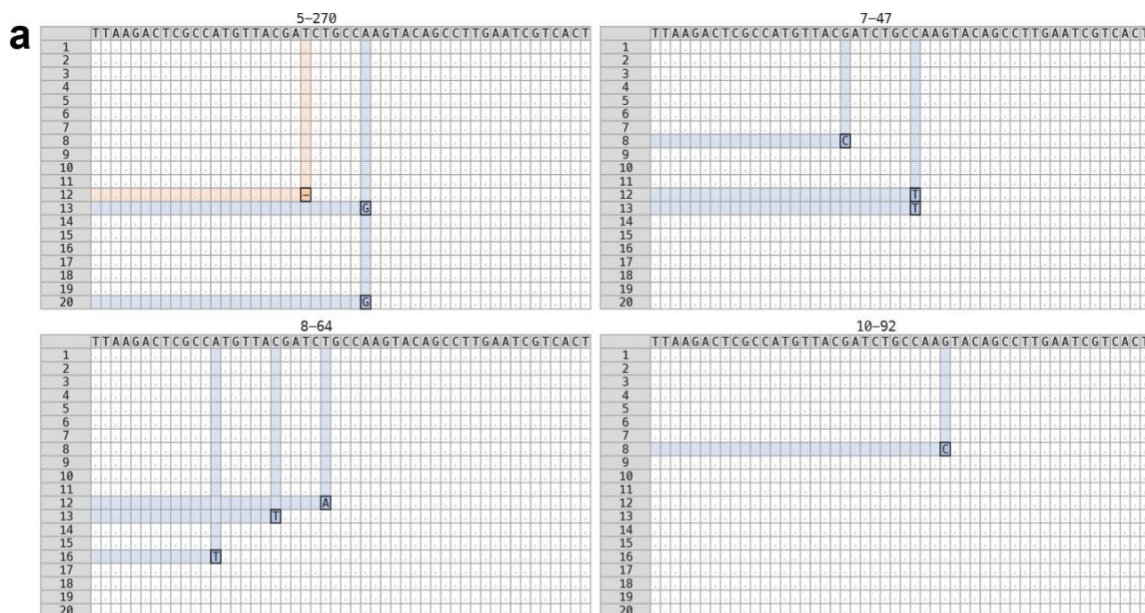
		20		40		60		80	
Kod-WT	MILDTDYITE	DGKPVIRIFK	KENGEFKIEY	DRTFEPYFYA	LLKDDSAIEE	VKKITAERHG	TVVTVKRVEK	VQKKFLGRPV	80
Kod-RI									80
Kod-RSGA									80
5-270									80
7-47									80
8-64									80
10-92									80
		100		120		140		160	
Kod-WT	EVWKLYFTHP	QDVPAIRDKI	REHPAVIDIY	EYDIPFAKRY	LIDKGLVPME	GDEELKMLAF	AIAITLYHEGE	EFAEGPILMI	160
Kod-RI									160
Kod-RSGA									160
5-270		.R.	.A.	.V.		.I.	.T.		160
7-47		.R.	.A.	.V.		.I.	.T.		160
8-64		.R.	.A.	.V.		.I.	.T.		160
10-92		.R.	.A.	.V.		.I.	.T.		160
		180		200		220		240	
Kod-WT	SYADEEGARV	ITWKNVDLPY	VDVVSSTEREM	IKRFLRVVKE	KDPDVLITYN	GDNFDFAYLK	KRCEKLGINF	ALGRDGESEPK	240
Kod-RI									240
Kod-RSGA									240
5-270									240
7-47									240
8-64									240
10-92									240
		260		280		300		320	
Kod-WT	IQRMGDRFAV	EVKGRIHFDL	YPVIRRTINL	PTYTLEAVYE	AVFGQPKEKV	YAEETITAWE	TGENLERVAR	YSMEDAKVTY	320
Kod-RI									320
Kod-RSGA									320
5-270					.K.	.AQ.	.G.		320
7-47					.K.	.AQ.	.G.		320
8-64					.K.	.AQ.	.G.		320
10-92					.K.	.AQ.	.G.		320
		340		360		380		400	
Kod-WT	ELGKEFLPME	AQLSRLIGQS	LWDVSRSSSTG	NLVEWFLLRK	AYERNELAPN	KPDEKELARR	.RQSYEGGYV	KEPERGLWEN	399
Kod-RI									399
Kod-RSGA									399
5-270		.V.	.P.	.Y.		.R.YE.	.L.E.A.	.K.	400
7-47		.V.	.P.	.Y.		.R.YE.	.L.E.A.	.K.	400
8-64		.V.	.P.	.Y.		.R.YE.	.L.E.A.	.K.	400
10-92		.V.	.P.	.Y.		.R.YE.	.L.E.A.	.K.	400
		420		440		460		480	
Kod-WT	IVYLDERSLY	PSIIITHNVS	PDTLNREGCK	EYDVAPQVGH	RFCKDFPGFI	PSLLGDLLEE	RQKIKKKMKA	TIDPIERKLL	479
Kod-RI									479
Kod-RSGA									479
5-270							.R.	.V.L.K.	480
7-47							.R.	.V.L.K.	480
8-64							.R.	.V.L.K.	480
10-92							.R.	.V.L.K.	480
		500		520		540		560	
Kod-WT	DYRQRAIKIL	ANSYYGYGY	ARARWYCKEC	AESVTAWGRE	YITMTIKEIE	EKYGFKVIYS	DTDGFFATIP	GADAETVKKK	559
Kod-RI	.R.								559
Kod-RSGA	.R.	.S.							559
5-270	.R.	.S.		.Q.	.ET.R.	.F.L.A.			560
7-47	.R.	.S.		.Q.	.ET.R.	.F.L.A.			560
8-64	.R.	.S.		.Q.	.ET.R.	.F.L.A.			560
10-92	.R.	.S.G.		.Q.	.ET.R.	.F.L.A.	.P.H.		560
		580		600		620		640	
Kod-WT	AMEFLKYINA	KLPGALELEY	EGFYKRGFFV	TKKKYAVIDE	EGKITTRGLE	IVRRDWEIA	KETQARVLEA	LLKDGDEVEKA	639
Kod-RI									639
Kod-RSGA					.G.				639
5-270	.K.	.D.	.L.		.D.	.G.			640
7-47	.K.	.D.	.L.		.D.	.G.			640
8-64	.K.	.D.	.L.		.D.	.G.			640
10-92	.K.	.D.	.L.		.D.	.G.			640
		660		680		700		720	
Kod-WT	VRIVKEVTEK	LSKYEVPEEK	LVIEHQITRD	LKDYKATGPH	VAVAKRLAAR	GVKIRPGTVI	SYIVLKGSGR	IGDRAIPFDE	719
Kod-RI			.I.						719
Kod-RSGA									719
5-270				.R.					720
7-47				.R.				.S.	720
8-64				.R.				.S.	720
10-92				.R.				.S.	720
		740		760		780		800	
Kod-WT	FDPTKHKYDA	EYYIENQVLP	AVERILRAFG	YRKEDLRYQK	TRQVGLSAWL	KPKGT	774		
Kod-RI							774		
Kod-RSGA	.A.						774		
5-270	.A.						775		
7-47	.A.					.T.	775		
8-64	.A.					.T.	775		
10-92	.A.					.T.	775		

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).



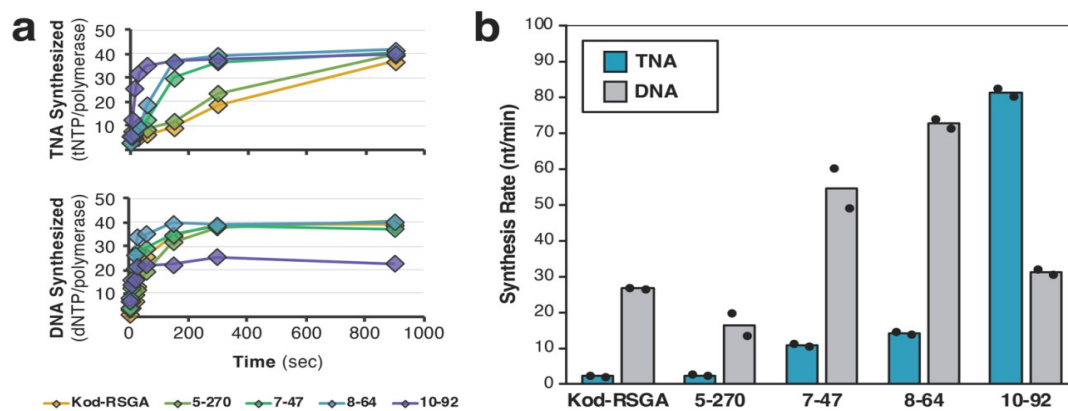
b

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

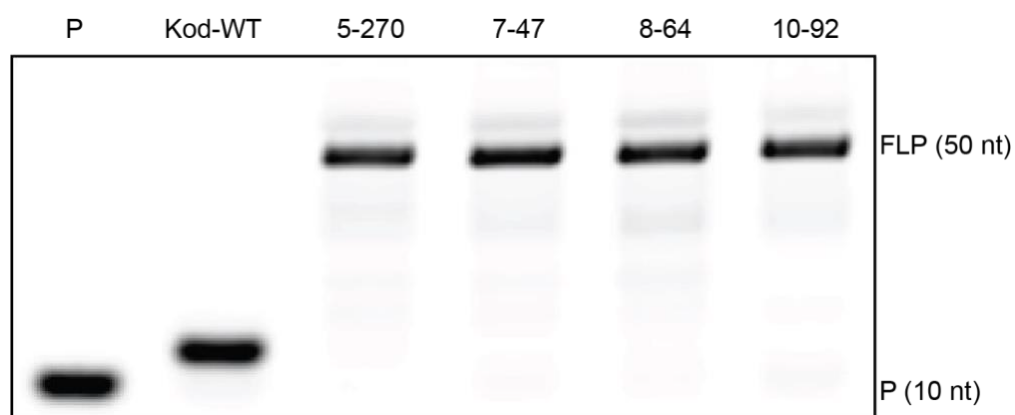
Binomial Two-Proportion Z-Test (Substitution)				
Comparison	Δ Fidelity (pp)	z	p-Value	Statistical Significance
Kod-RSGA → 5-270	0.754	2.309	0.0209	Significant ($p < 0.05$)
5-270 → 7-47	-0.100	-0.448	0.6543	Not Significant
7-47 → 8-64	0.000	0.000	1.0000	Not Significant
8-64 → 10-92	0.200	1.001	0.3168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

Binomial Two-Proportion Z-Test (Substitution + In/Del)				
Comparison	Δ Fidelity (pp)	z	p-Value	Statistical 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

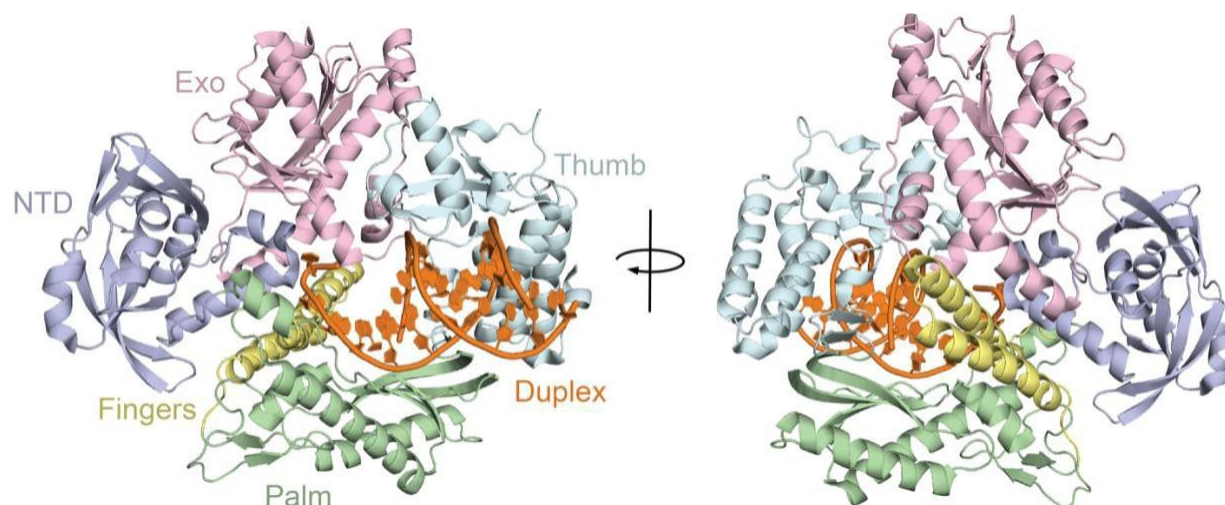
Supplementary Figure 6. TNAP fidelity raw data and statistical analysis. (a) Raw Sanger sequencing alignments from a single cycle of TNA replication by the indicated TNAP variants. The expected DNA sequence is shown 5'→3'. Correct base incorporations are marked with (.), and substitutions or insertion/deletion 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.



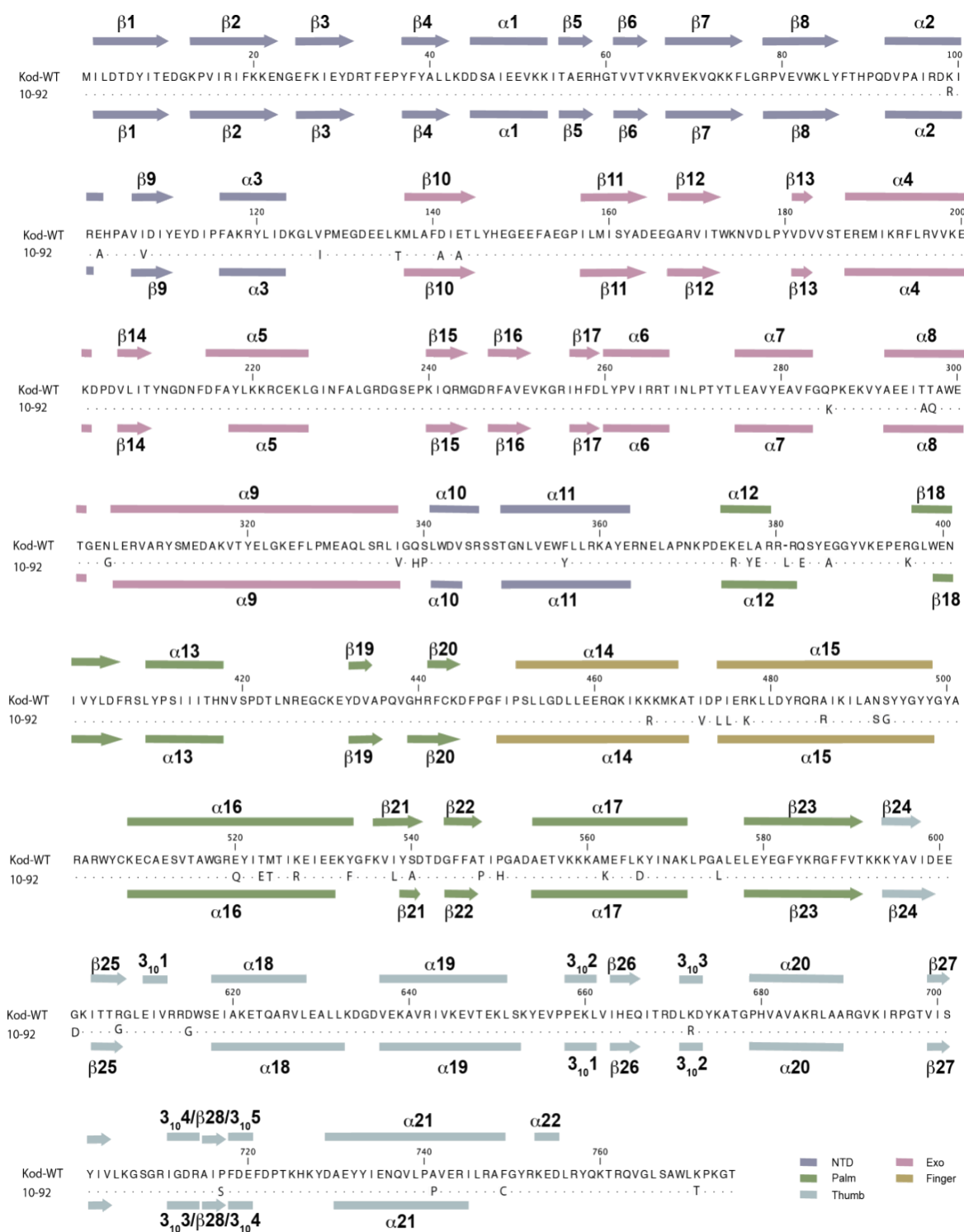
Supplementary Figure 7. TNAP kinetics and substrate specificity. **a**, Kinetic time-course of primer-extension reactions catalyzed by Kod-RSGA, 5-270, 7-47, 8-64, and 10-92 using TNA (tNTP) or DNA (dNTP) triphosphates, average of $n=2$. The slope of the linear range corresponds to the rate of synthesis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



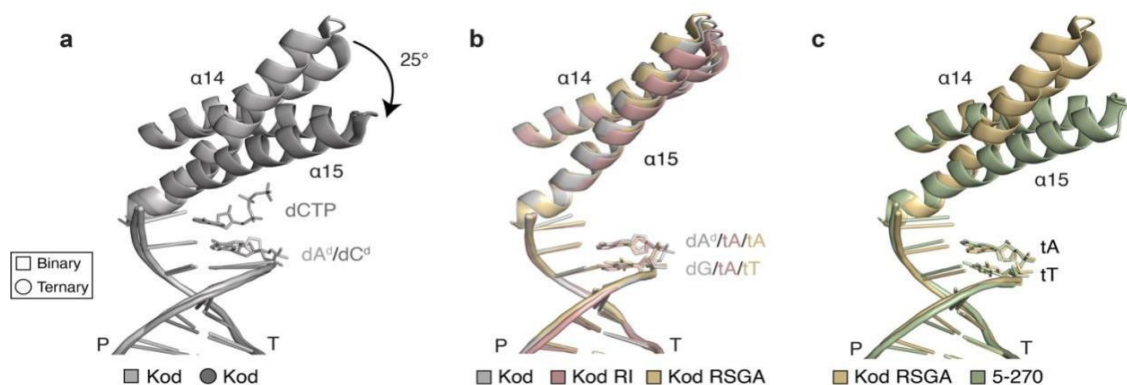
Supplementary Figure 8. TNAP thermal challenge. TNA synthesis activity was assessed after exposing the polymerases to 6 hours of incubation at 90°C. Reactions were performed by incubating a 5' IR680-labelled DNA primer-template duplex (1 μ M) with 1 μ M polymerase and 100 μ M tNTPs in 1x ThermoPol buffer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



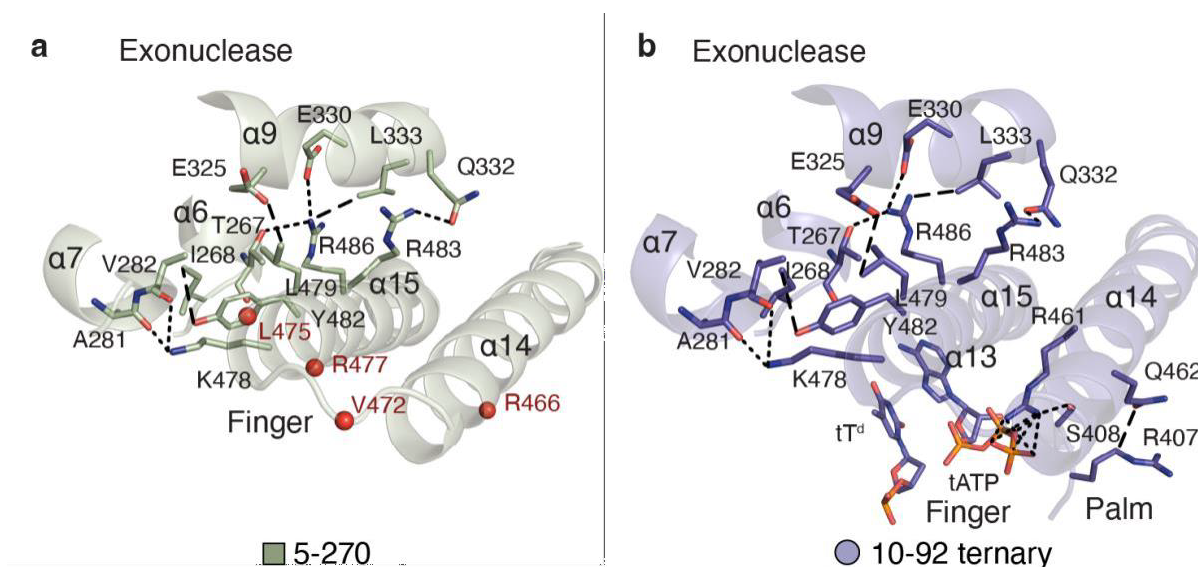
Supplementary Figure 9. Polymerase domains. The architecture of B-family polymerases colored by domain, mapped on the closed ternary structure of wildtype Kod (PDB ID: 5OMF).



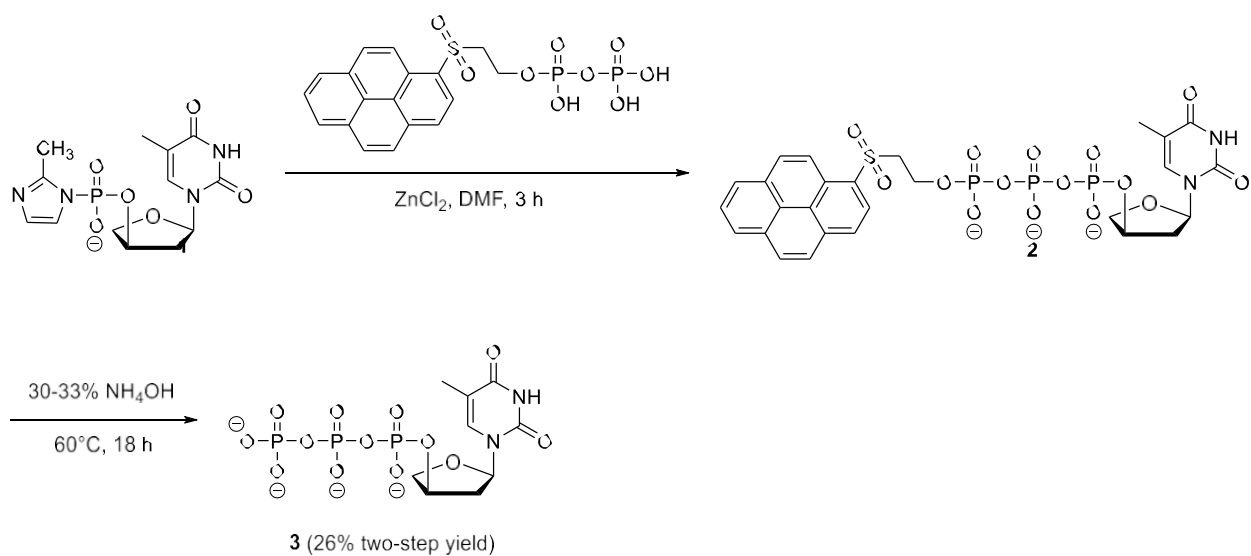
Supplementary Figure 10. Sequence and secondary structure alignment of Kod and 10-92. Conserved residues are denoted as (.), while mutations are listed as single letter amino acid abbreviations. Secondary structures of Kod and 10-92, as determined by DSSP and PyMol, are depicted above and below the sequence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



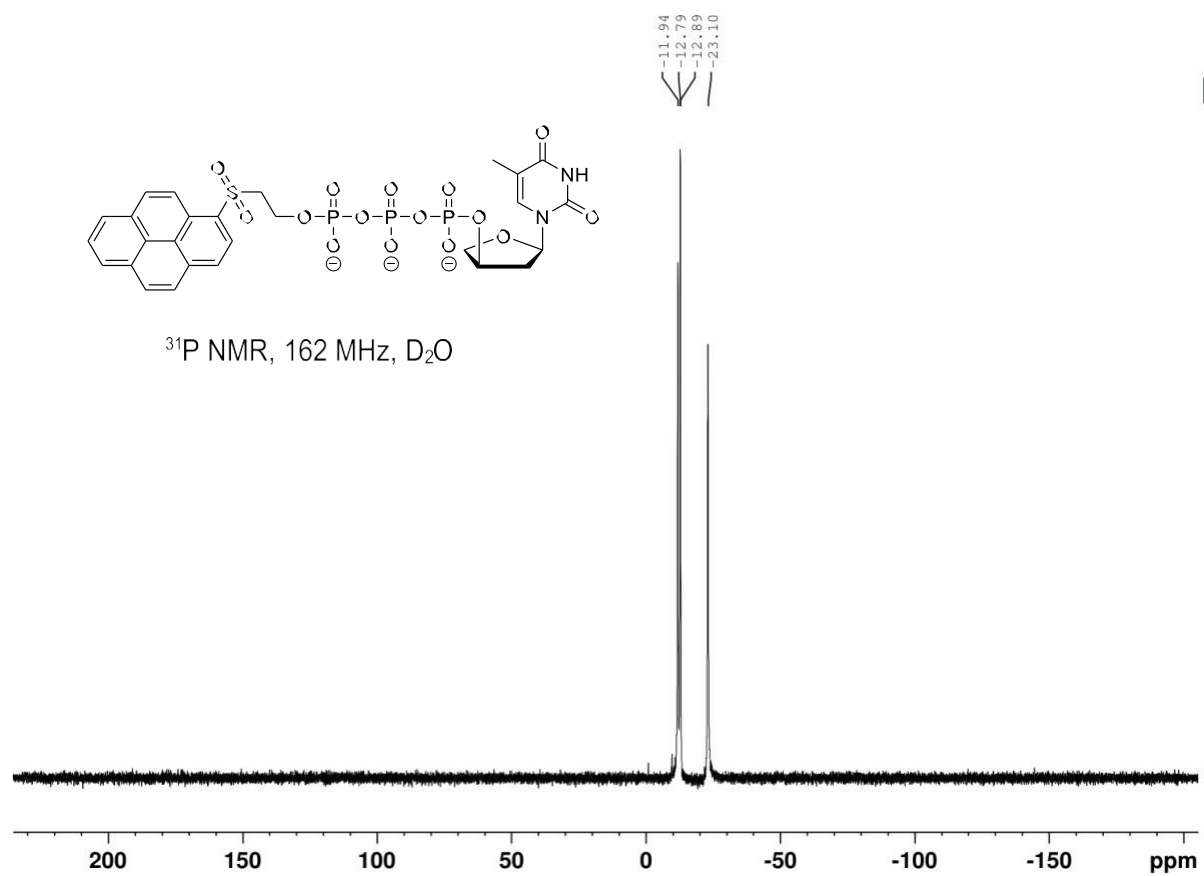
Supplementary Figure 11. Structural comparison of the finger subdomain of Kod, Kod-RI, Kod-RSGA, and 5-270 in the binary complex. **a**, Overlay of the finger subdomains of Kod binary (light gray, PDB: 4K8Z) and closed ternary (dark gray, PDB: 7OMB) complexes. **b**, Overlay of the finger subdomains 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.



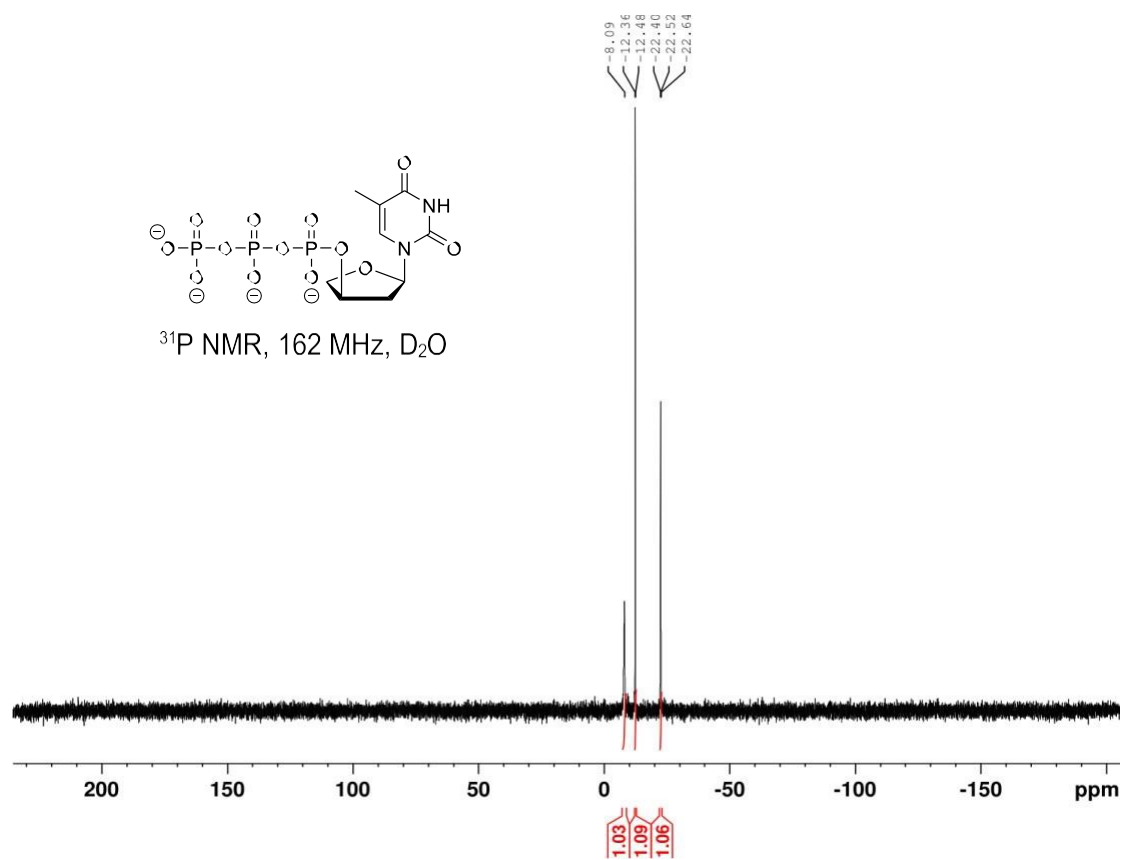
Supplementary Figure 12. Active site comparison. The active site of **a**, the 5-270 binary structure reveals a closed finger conformation similar to **b**, the 10-92 closed ternary structure. Short dashed lines denote electrostatic interactions, long dashed lines show hydrophobic interactions. Residue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



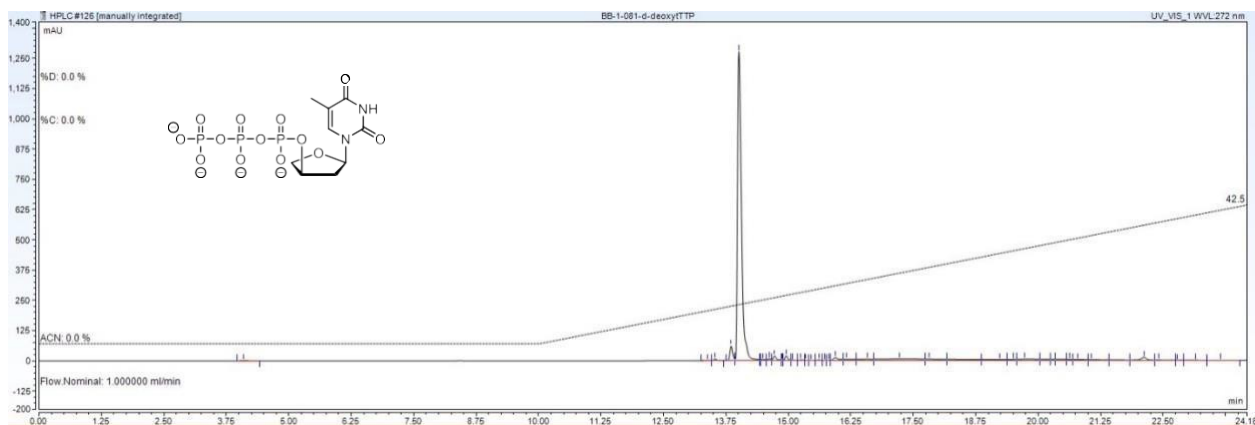
Supplementary Figure 13. Synthesis of 2'-deoxy- α -L-threofuranosyl thymine triphosphate (dtTTP) by pyrene pyrophosphate method. Modified synthesis of 2'-deoxy- α -L-threofuranosyl thymine triphosphate (dtTTP) using pyrene pyrophosphate to incorporate the β - and γ -phosphates². The synthesis of compound 1 was reported previously³.



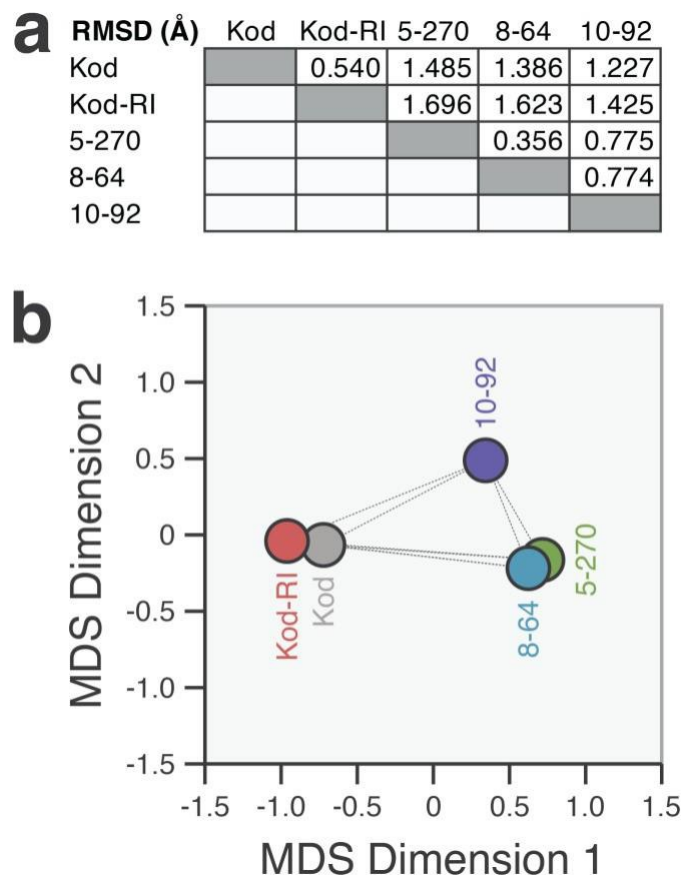
Supplementary Figure 14. NMR spectrum. ^{31}P NMR spectrum of compound 2.



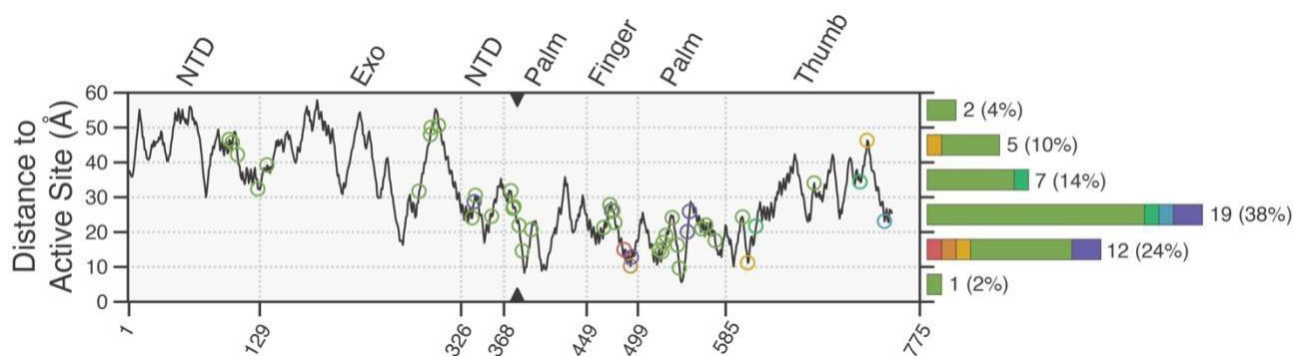
Supplementary Figure 15. NMR spectrum. ^{31}P NMR spectrum of compound 3.



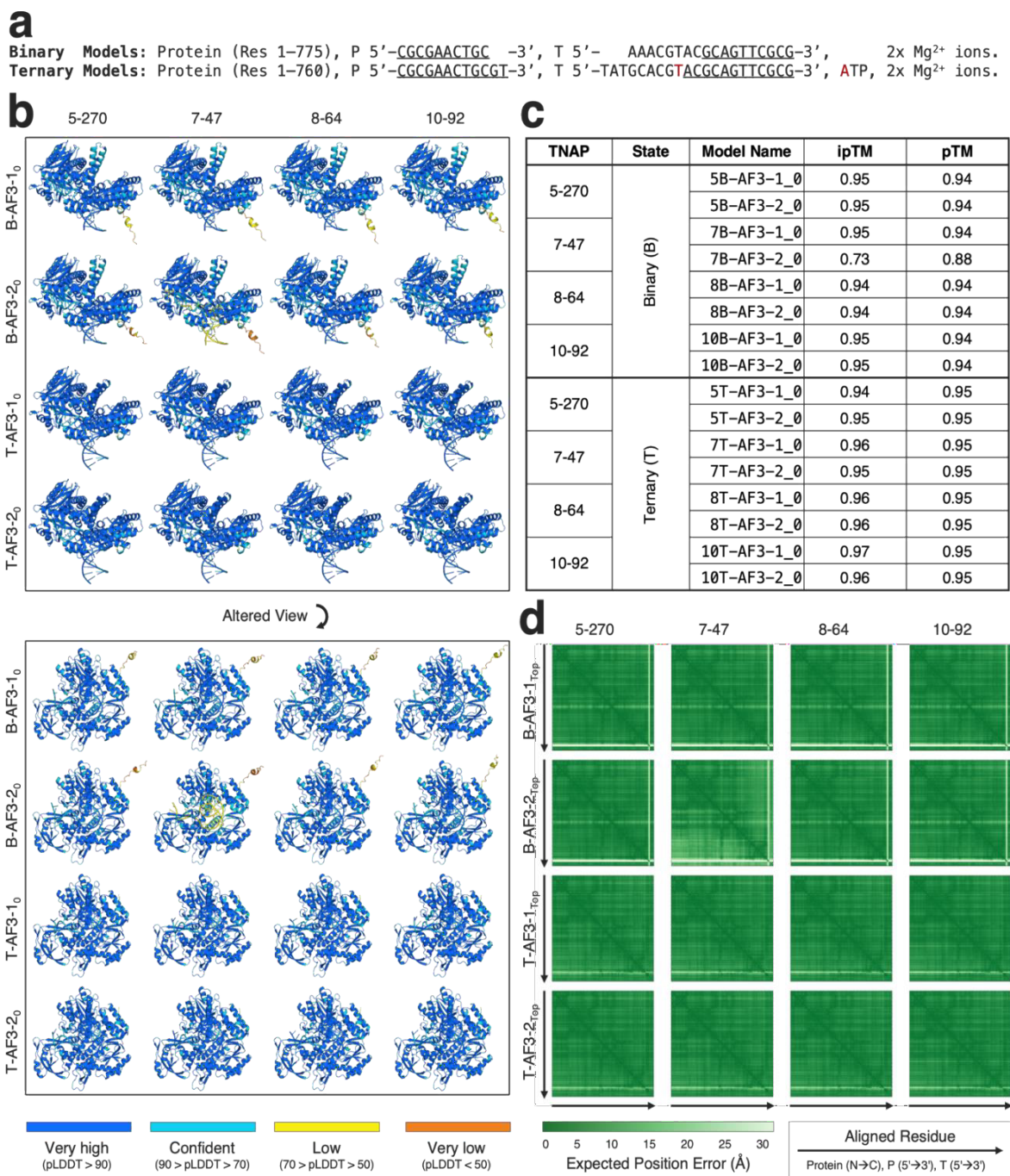
Supplementary Figure 16. HPLC spectrum. HPLC analysis of compound 3.



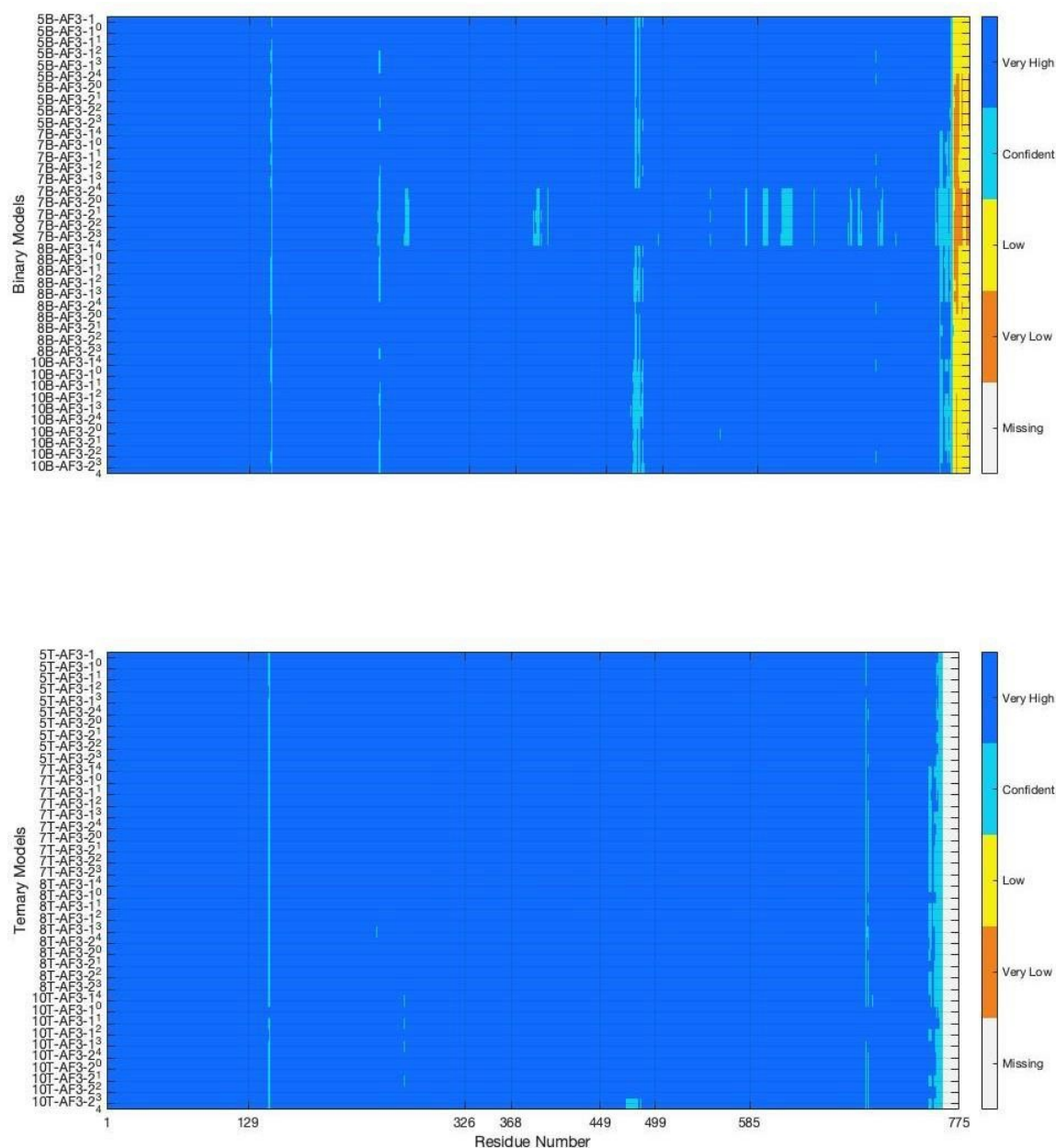
Supplementary Figure 17. Multidimensional analysis of the closed ternary complex of TNAP structures based on RMSD. (a) Pairwise root mean square deviation (RMSD, Å) values for ternary complexes, calculated over C α atoms following global alignment in PyMOL. Lower RMSD values indicate greater structural 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.



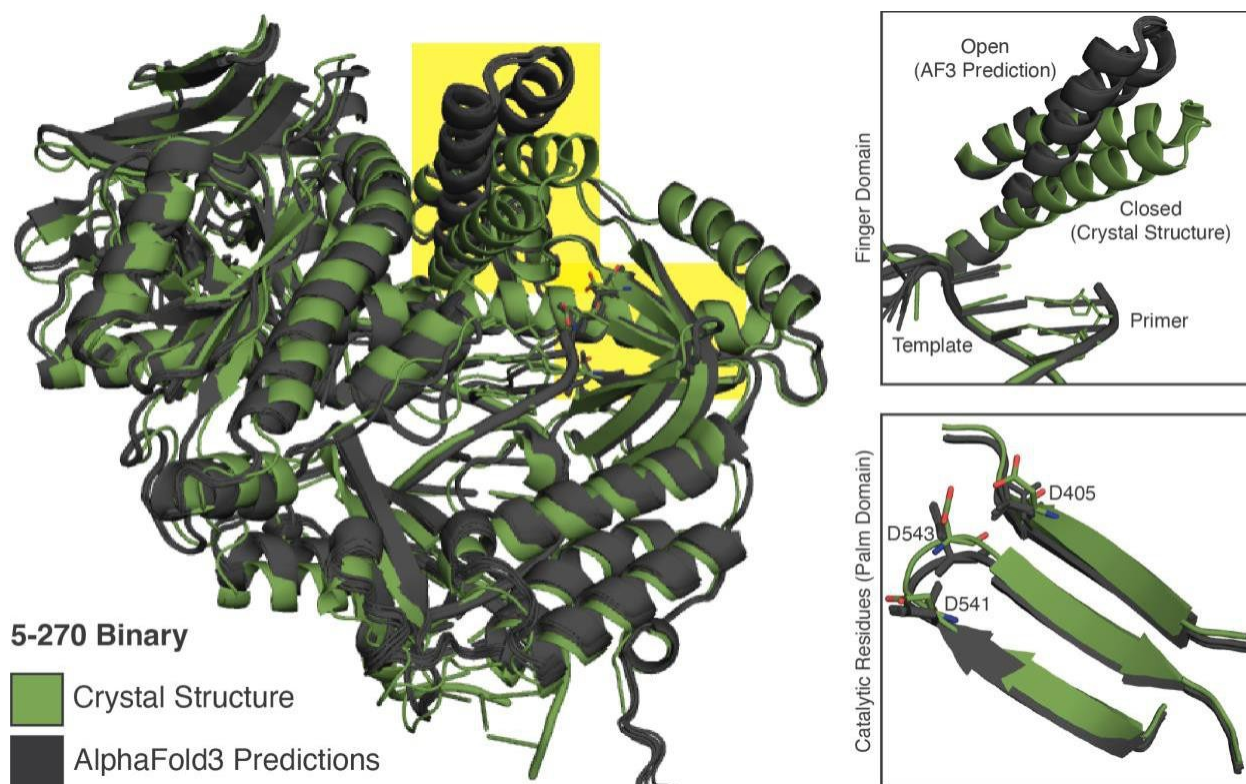
Supplementary Figure 18. Distance of mutations to active site. The distance between the alpha-carbon of each residue in the 10-92 ternary structure to the active site (C2' of the terminal primer TNA residue) is measured in angstroms. Mutations compared to Kod are indicated as open circles and color-matched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



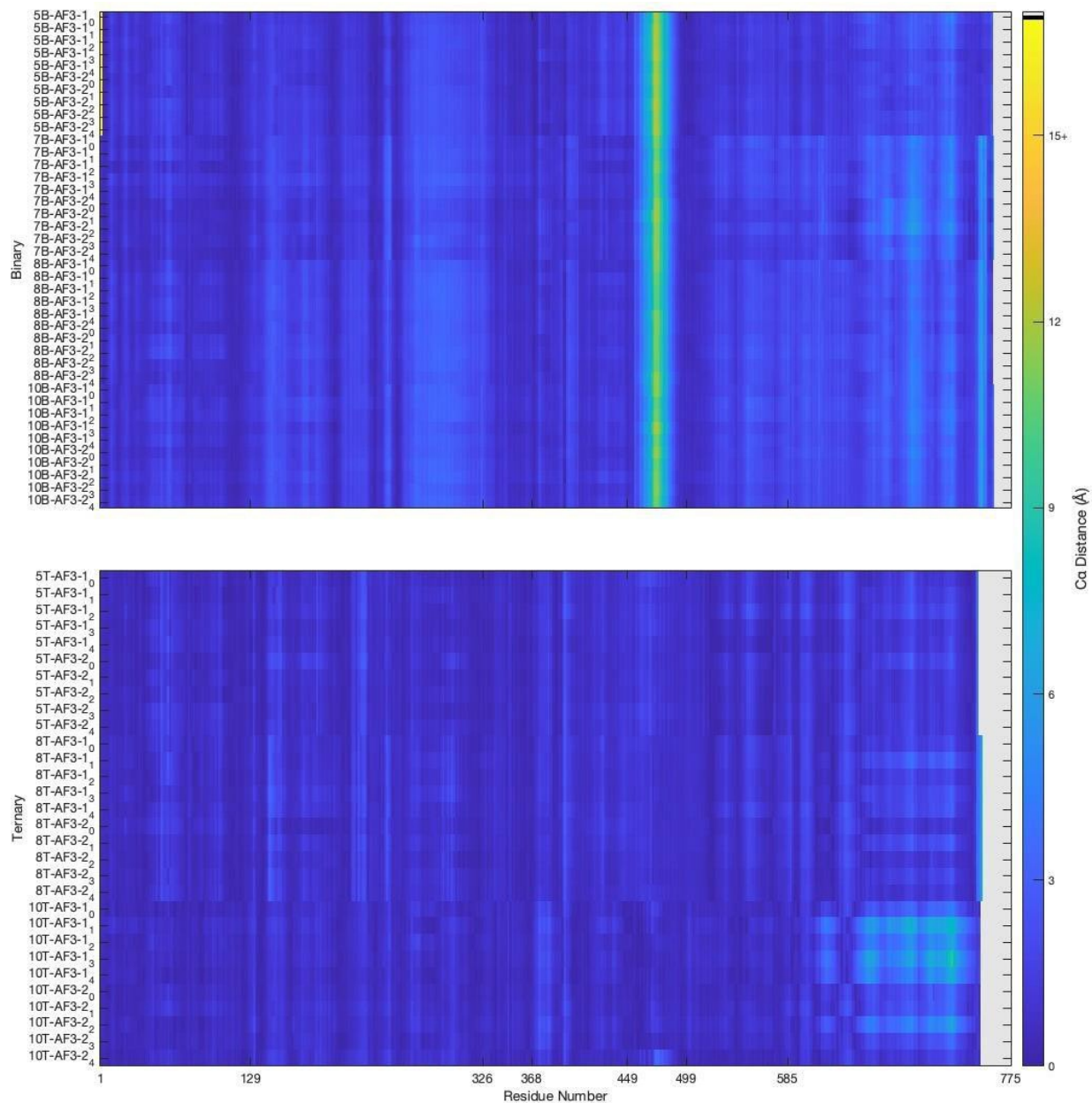
Supplementary Figure 19. AlphaFold3 prediction overview. (a) Input specifications for AlphaFold3 binary and ternary complex predictions. Two independent predictions were run for each structure (denoted AF3-1 and AF3-2) with each seed yielding five models (0-4). (b) Representative predicted models (model 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.



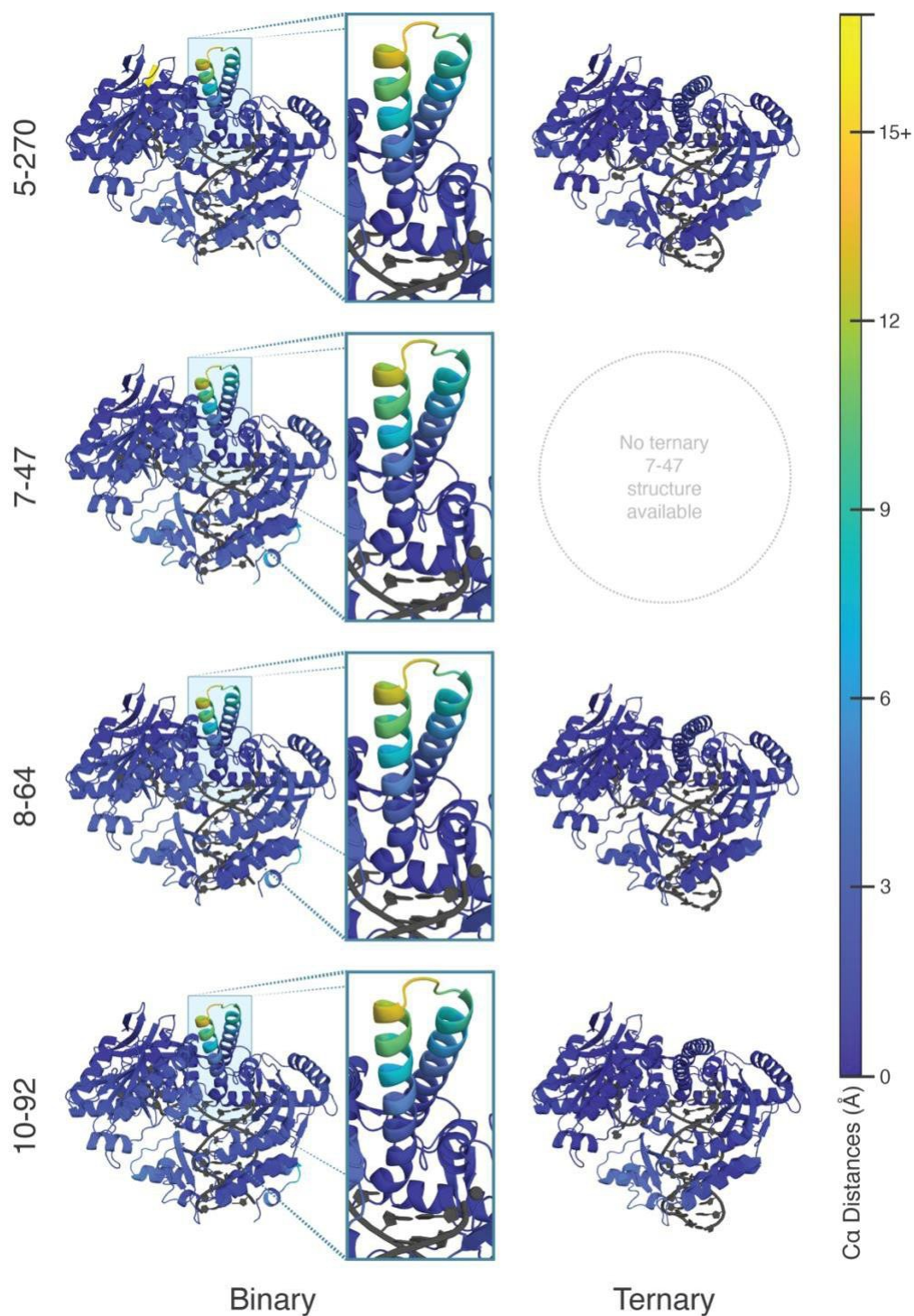
Supplementary Figure 20. AlphaFold3 prediction confidence. AlphaFold3 model confidence (pLDDT) heatmaps for binary (top) and ternary (bottom) complexes. Each row represents a single AlphaFold3 model. Two independent AlphaFold3 prediction runs were performed for each structure (denoted AF3-1 and AF3-2), 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 (AF3-1₀₋₄, 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.

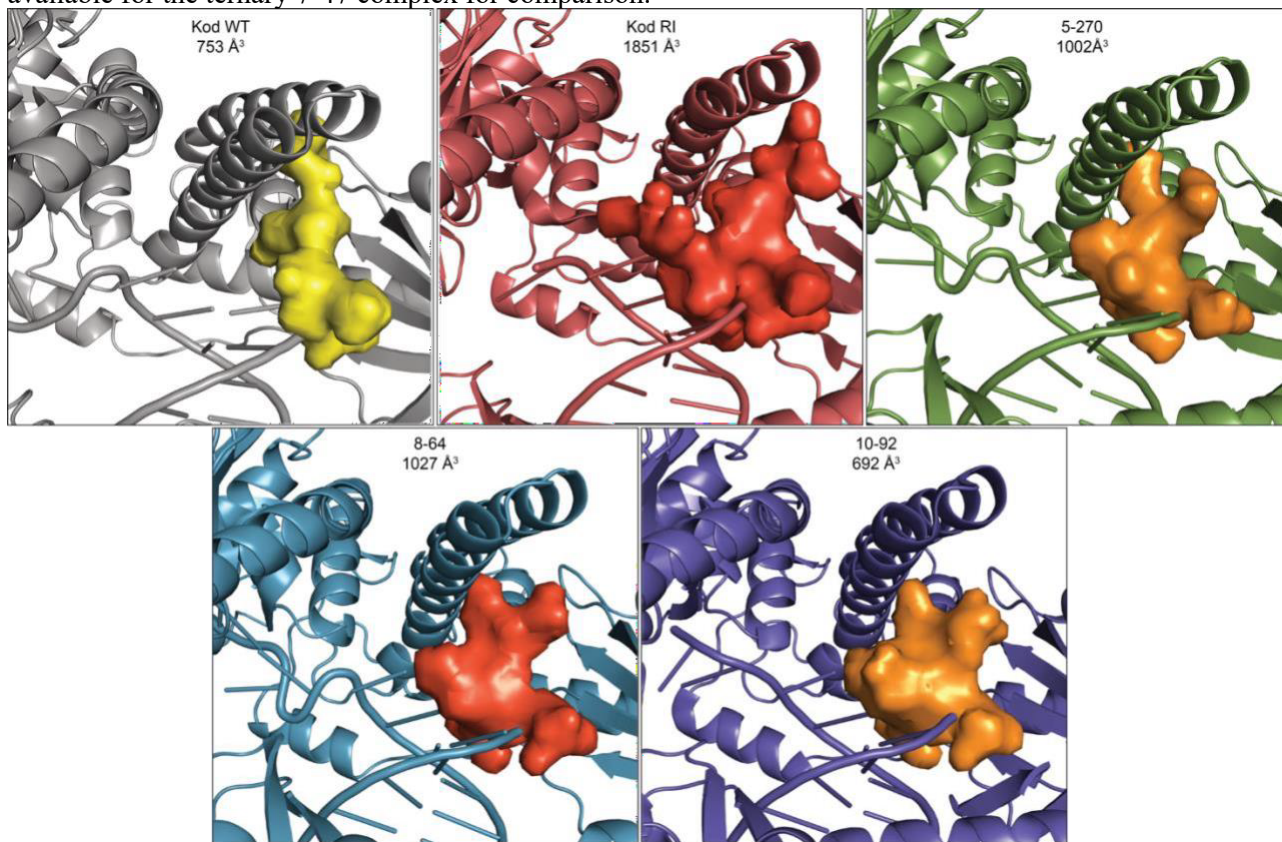


Supplementary Figure 22. Comparison of AlphaFold3 predictions to experimental structures. The top panel shows AlphaFold3 predictions of the binary complexes (5-270, 7-47, 8-64, and 10-92), while the bottom panel shows predictions of ternary complexes (5-270, 8-64, and 10-92). Two AlphaFold3 runs were 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 Ca position for each amino acid residue along the backbone.

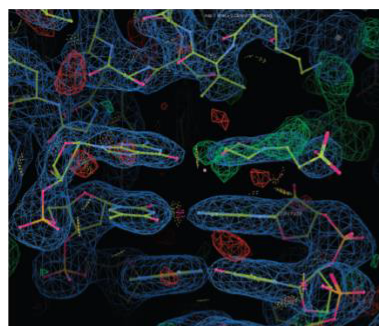


Supplementary Figure 23. Visualizing AlphaFold3 Predictions Against Experimental Structures. Structural validation of AlphaFold3 predictions (AF3-1₀) relative to crystal structures for each crystallized binary and ternary complex (5-270, 7-47, 8-64, and 10-92). AlphaFold3 predictions are shown as cartoons, with residues colored according to the C α 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 Å³) 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.



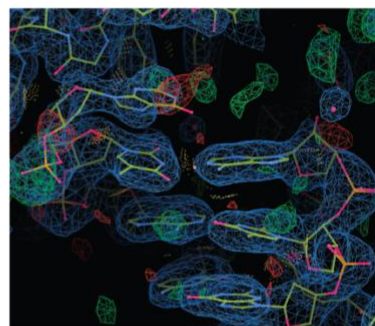
5-270 binary



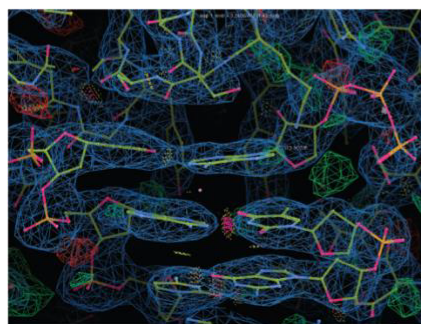
7-47 binary



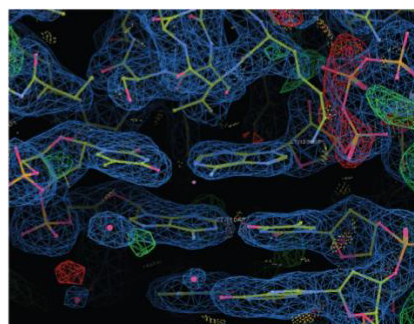
8-64 binary



10-92 binary

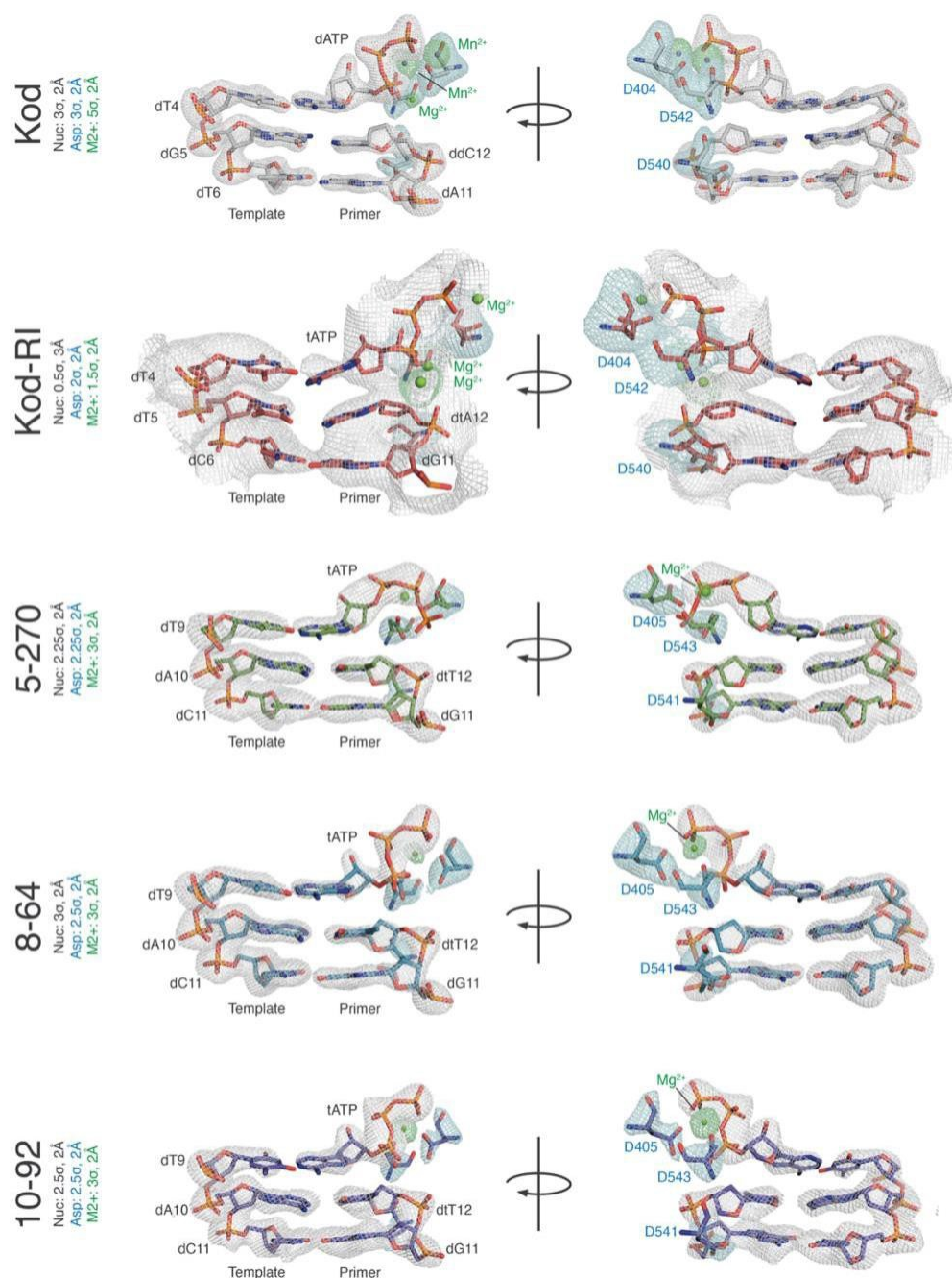


5-270 ternary

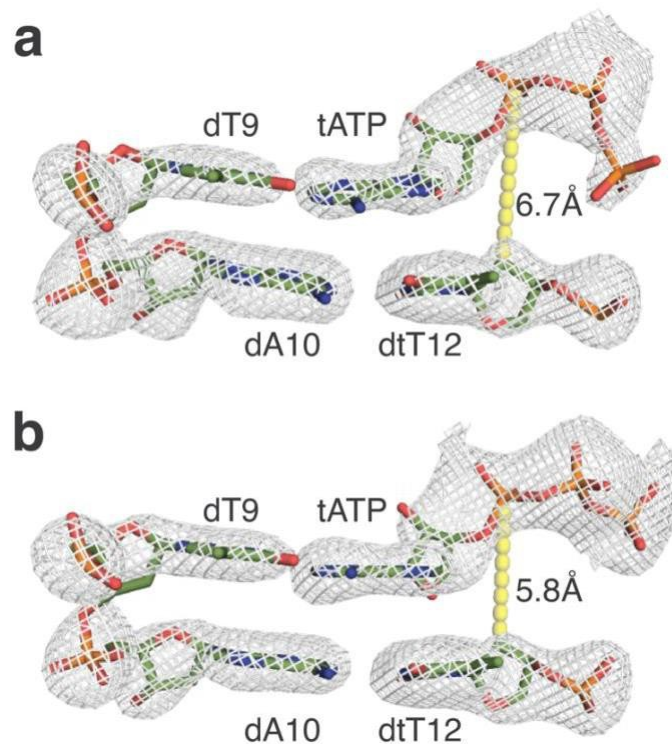


8-64 ternary

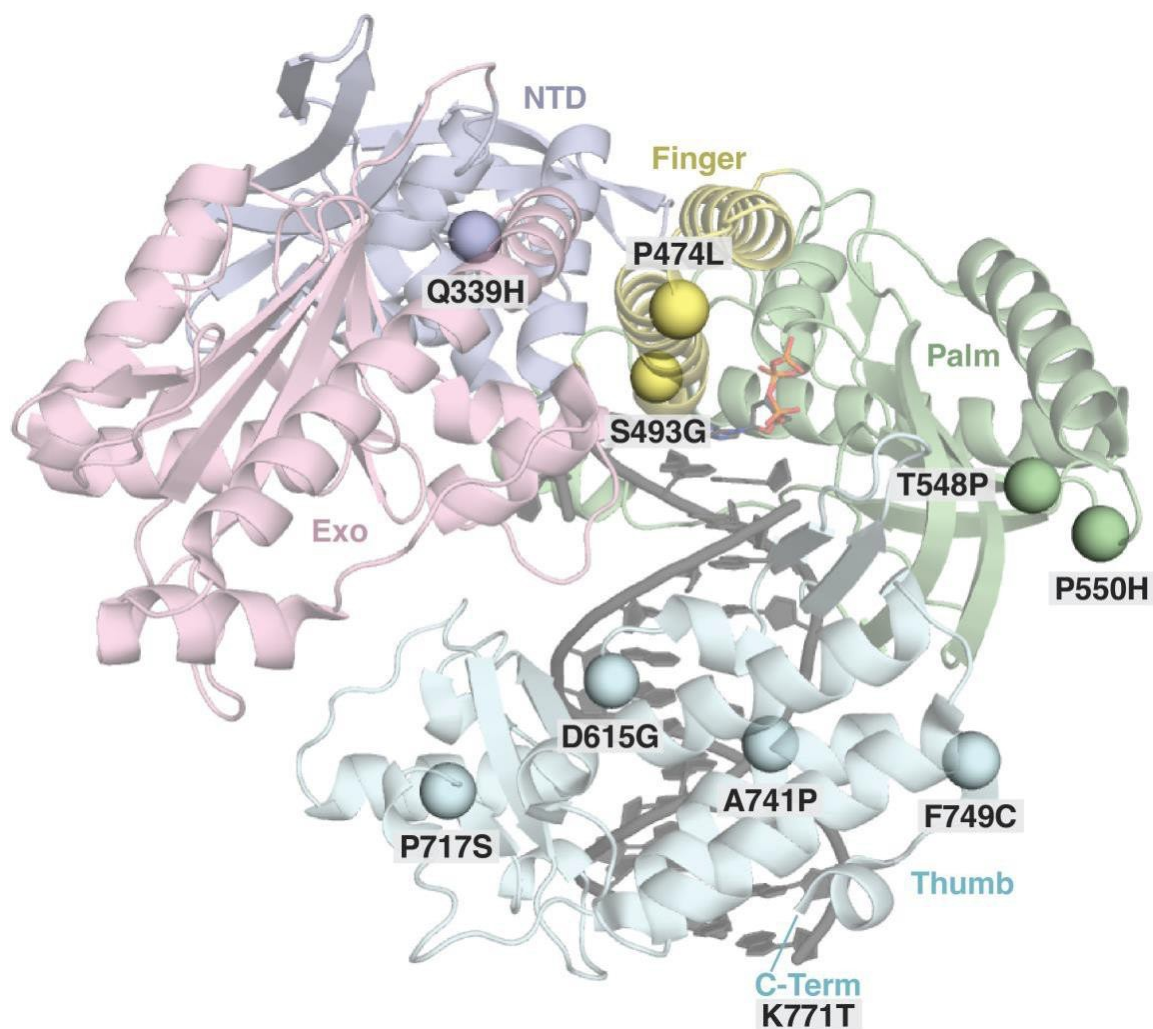
Supplementary Figure 25. 2Fo-Fc electron density maps of the active site. For the newly reported structures, active site 2Fo-Fc electron density maps, contoured at 1.5σ , are shown.



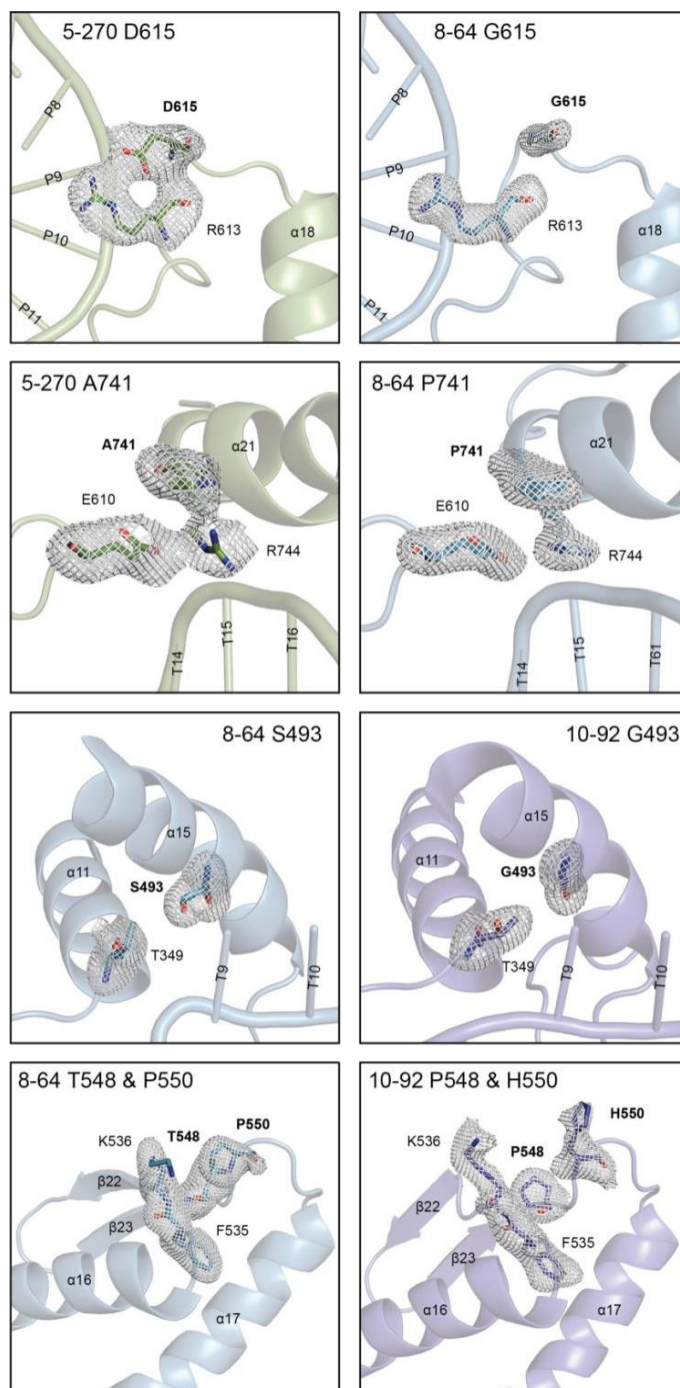
Supplementary Figure 26. Polder maps of key residues in the active site. The incoming triphosphate, neighboring nucleotides, catalytic aspartates (D404/D405, D540/D541, and D542/D543), and divalent metal ions within the active sites of closed ternary complexes are shown. Polder maps (*mFo*–*DFc*) are 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.



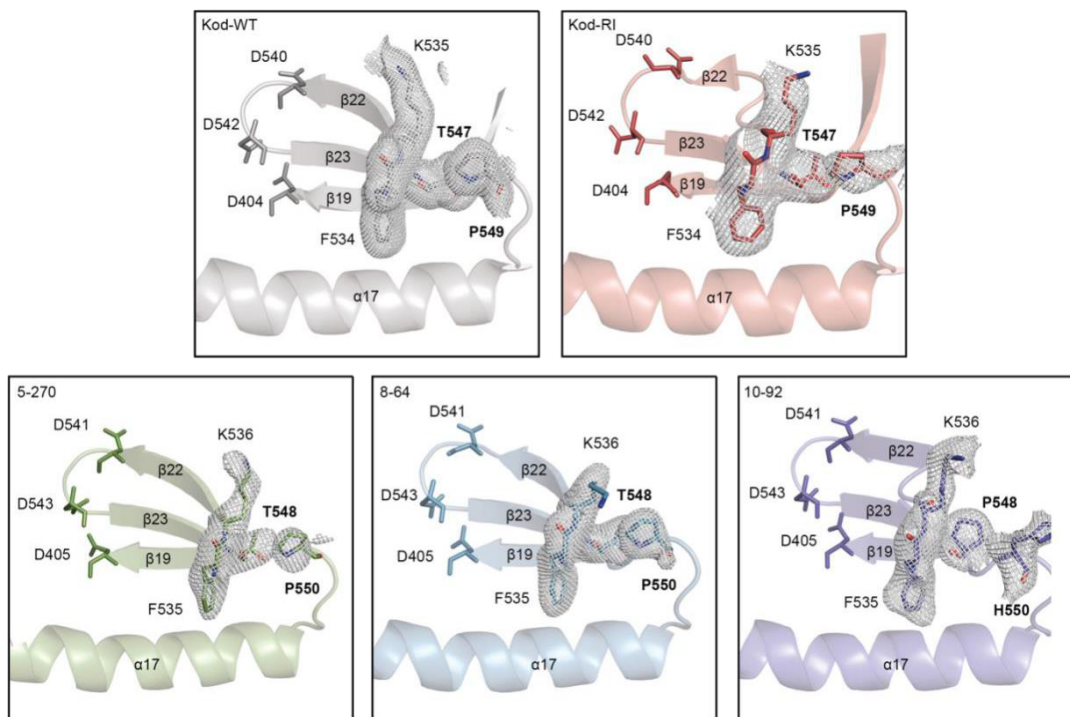
Supplementary Figure 27. Active site geometry comparing two independently solved structures of 5-270 in a closed ternary complex. Reaction centers are shown for (a) the original crystallographic dataset *G7* and (b) an alternative independent dataset *G3*. The distance from the primer to the α -phosphate 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 C α 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 $2.0 \text{ \AA}$, while the rest of the maps are contoured at 3.0σ and carved to $2.0 \text{ \AA}$.



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 $2.0\text{ \AA}$, 8-64 maps are contoured at 4.0σ and carved to $2.0\text{ \AA}$, and the rest of the maps are contoured at 3.0σ and carved to $2.0\text{ \AA}$.

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.

	Oligo name	Oligo sequence
Kinetics	PBS8 primer	GTCCCCCTGGGGATACCACC
	EM619 template	CCCACACCCTCCTATCGCTAAACACACACTTAATAAAGTTGGTGGTATCCCCAAGGGGAC
Fidelity	Acrydite PBS8 primer	/5Acryd/GTCCCCCTGGGGATACCACC
	4NT.9G Fidelity template	GGATCGTCAGTGCAAAGAGATTAAGACTCGCCATGTTACGATCTGCCAAGTACAGCCTTGA ATCGTCACTGGTGGTATCCCCCTGGGGAC
	PBS7.PBS9 overhang primer	CTTTTAAGAACCGGACGAACGGATCGTCAGTGCATTTGAGA
	PBS8 primer	GTCCCCCTGGGGATACCACC
	PBS9 primer	CTTTTAAGAACCGGACGAAC
Thermal challenge	PBS8_short primer	/5IRD680/GTCCCCCTGG
	EM619 template	CCCACACCCTCCTATCGCTAAACACACACTTAATAAAGTTGGTGGTATCCCCAAGGGGAC
Cloning	TNAP-Fwd	ATCCATATGATCCTCGACACTGACTAC
	TNAP-Rvs	ACGCATGCGGCCGCTCAAGTTCCTTTTCGGCGTCAG
	TNAP-Rvs_760	ACGCATGCGGCCGCTCACTTCTGGTAGCGCAGGTC
	8-64-Rvs_760	ACGCATGCGGCCGCTCAAGTTCCTTTTCGGCGTCAG
Crystallography	P	CGCGAACTGCG
	T1	/5Cy5/AAACGTACGCAGTTCGCG
	T2	TATGCACGTACGCAGTTCGCG

Supplementary Table 2. Data collection and refinement statistics for binary structures. *Values in parentheses are for the highest-resolution shell

	5-270	7-47	8-64	10-92
Data Collection				
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Cell Dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (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 (°)	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)
<i>R</i> _{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>I</i> / σ <i>I</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)
<i>R</i> _{work} / <i>R</i> _{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 (°)	0.46	1.01	0.52	0.47

Supplementary Table 3. Data collection and refinement statistics for ternary structures. *Values in parentheses are for the highest-resolution shell.

	5-270	8-64
Data Collection		
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell Dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	77.4, 101.7, 110.8	77.7, 101.5, 110.7
α , β , γ (°)	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)
<i>R</i> _{merge}	0.1628 (1.017)	0.0734 (0.7658)
CC1/2	0.995 (0.769)	0.999 (0.842)
<i>I</i> / σ <i>I</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)
<i>R</i> _{work} / <i>R</i> _{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 (°)	0.39	0.53

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

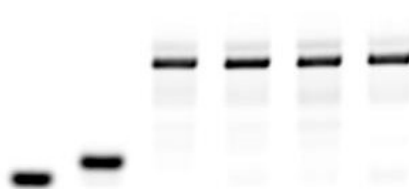
Kod-WT	Shear	Stretch	Stagger	Buckle	Prop-Tw	Opening	Shift	Slide	Rise	Tilt	Roll	Twist
T-a	-0.01	-0.236	0.064	-5.456	0.746	2.845	0	0	0	0	0	0
G-C	-0.271	-0.217	0.012	-8.368	-14.078	1.971	-0.605	-1.704	3.301	3.044	5.375	29.301
T-A	-0.041	-0.072	-0.218	1.349	0.959	-2.94	-0.741	-1.059	3.181	3.988	3.466	30.229
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 5. Integrated mutational analysis. For each mutation, the polymerase domain, generation of appearance during TNAP evolution, BLOSUM62 substitution score, predicted stability change (FoldX $\Delta\Delta G$, kcal mol⁻¹), C α distance to the active site (Å), and absolute and relative solvent-accessible surface area (ABS SASA, Å²; 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 data¹. N/A = non-applicable.

Mutation	Domain	TNAP Generation	BLOSUM62	FoldX $\Delta\Delta G$ (kcal mol ⁻¹)	C α 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

REFERENCES

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- α -l-threofuranosyl Nucleoside Triphosphates. *J. Org. Chem.* **83**, 8840-8850 (2018).



Uncropped gel for Supplementary Figure 8

APPENDIX D: Supplementary Information for Chapter 4

Supplementary Table 1. DNA gBlocks.

gBlock Name	DNA Sequence (5'→ 3')
Tgo-QGLK	ATGATCCTCGATACAGACTACATAACTGAGGATGGAAAGCCCGTCATCAGGATCTTCAA GAAGGAGAACGGCGAGTTCAAAATAGACTACGACAGAACTTTGAGCCATACATCTACG CGCTCTTGAAGGACGACTCTGCGATTGAGGACGTCAAGAAGATAACTGCCGAGAGGCAC GGCACTACCGTTAGGGTTGTCAGGGCCGAGAAAGTGAAGAAGAAGTTCCTAGGCAGGCC GATAGAGGTCTGGAAGCTCTACTTCACTCACCCCCAGGACCAGCCCGCAATCAGGGACA AGATAAAGGAGCATCCTGCCGTTGTGGACATCTACGAGTACGACATCCCCTTCGCGAAG CGCTACCTCATAGACAAAGGCTTAATCCCGATGGAGGGCGACGAGGAAGTTAAGATGCT CGCCTTCGCGATCGCGACGCTCTATCACGAGGGCGAGGAGTTCGCCGAAGGGCCTATCC TGATGATAAGCTACGCCGACGAGGAAGGGGCGCGCGTTATTACCTGGAAGAATATCGAC CTTCCCTATGTCGACGTCGTTTCCACCGAGAAGGAGATGATAAAGCGCTTCCTCAAGGT CGTCAAGGAAAAGGATCCCGACGTCCTCATAACCTACAACGGCGACAAGTTCTGACTTCG CCTACCTCAAGAAGCGCTCCGAGAAGCTCGGAGTCAAGTTCATCCTCGGAAGGGAAGGG AGCGAGCCGAAAATCCAGCGCATGGGCGATCGCTTTGCGGTGGAGGTCAAGGGAAGGAT TCACTTCGACCTCTACCCCGTCATTAGGAGAACGATTAACCTCCCCACTTACACCCTTG AGGCAGTATATGAAGCCATCTTTGGACAGCCGAAGGAGAAGGTCTACGCTGAGGAGATA GCGCAGGCCTGGGAAACGGGCGAGGGATTAGAAAGGGTGGCCCGCTACTCGATGGAGGA CGCAAAGGTAACCTATGAACTCGGAAAAGAGTTCTTCCCTATGGAAGCCCAGCTCTCGC GCCTCGTAGGCCAGAGCCTCTGGGATGTATCTCGCTCGAGTACCGGAAACCTCGTCGAG TGTTTTTTGCTGAGGAAGGCCTACGAGAGGAATGAACTTGACCAAACAAGCCGGACGA GAGGGAGCTGGCAAGAAGAAGGGAGAGCTACGCGGGTGGATACGTCAAGGAGCCCCGAAA GGGGACTGTGGGAGAACATCGTGTATCTGGACTTCCGCTCCCTGGGCCCTTCGATAATA

	ATCACCCATAACGTCTCCCCTGATACACTCAACAGGGAGGGTTGTGAGGAGTACGACGT GGCTCCTCAGGTAGGCCATAAGTTCTGCAAGGACTTCCCCGGCTTCATCCCAAGCCTCC TCGGAGACCTCTTGGAGGAGAGACAGAAGGTAAAGAAGAAGATGAAGGCCACTATAGAC CCAATCGAGAAGAACTCCTCGATTACAGGCAACGACTGATCAAAATCCTTGCTAATAG CTTCTACGGTTACTACGGCTATGCAAAGGCCCGCTGGTACTGCAAGGAGTGCGCCGAGA GCGTTACCGCTTGGGGCAGGCAGTACATCGAGACCACGATAAGGGGAAATAGAGGAGAAA TTTGGCTTTAAAGTCCTCTACGCGGACACAGATGGATTTTTCGCAACAATACCTGGAGC GGACGCCGAAACCGTCAAAAAGAAGGCAAAGGAGTTCCTGGACTACATCAACGCCAAAC TGCCCGGCCTGCTCGAACTCGAATACGAGGGCTTCTACAAGCGCGGCTTCTTCGTGACG AAGAAGAAGTACGCGGTTATAGACGAGGAGGACAAGATAACGACGCGCGGGCTTGAAAT AGTTAGGCGTGACTGGAGCGAGATAGCGAAGGAGACGCAGGCGAGGGTTCTTGAGGCGA TACTAAAGCACGGTGACGTTGAAGAAGCGGTAAGGATTGTCAAAGAGGTTACGGAGAAG CTGAGCAAGTACGAGGTTCCACCGGAGAAGCTGGTCATCTACAAACAGATAACCCGCGA CCTGAAGGACTACAAGGCCACCGGGCCGCATGTGGCTGTTGCAAAACGCCTCGCCGCAA GGGGGATAAAAATCCGGCCCGGAACGGTCATAAGCTACATCGTGCTCAAAGGCTCGGGA AGGATTGGGGACAGGGCTATACCCTTTGACGAATTTGACCCGGCAAAGCACAAGTACGA TGCAGAATACTACATCGAGAACCAGGTTCTTCCAGCTGTGGAGAGGATTCTGAGGGCCT TTGGTTACCGTAAAGAAGATTTAAGGTATCAGAAAACGCGGCAGGTTGGCTTGGGGGCG TGGCTAAAACCTAAGACA
9N-QLGK	ATGATTCTGGATACCGATTACATTACCGAAAACGGCAAACCGGTGATTCGTGTGTTCAA GAAGGAGAACGGCGAGTTCAAAATCGAATACGATCGTACCTTTGAACCGTATTTTTACG CGCTCTTGAAGGACGACTCTGCGATCGAAGATGTGAAAAAAGTGACCGCGAAACGTCAT GGCACCGTGGTGAAAGTGAAACGTGCGGAAAAAGTGCAGAAGAAGTTCCTAGGCAGGCC GATTGAAGTGTGGAAACTGTATTTCAACCATCCGCAGGATCAGCCGGCGATTTCGTGATC GTATTTCGTGCGCATCCGGCGGTGGTGGATATTTATGAATATGATATCCCCTTCGCGAAG

	CGCTATCTGATTGATAAAGGCCTGATTCCGATGGAAGGCGATGAAGAACTGACCATGCT GGCCTTTGCGATCGCGACGCTCTATCACGAGGGCGAAGAATTTGGCACCGGCCCGATT TGATGATTAGCTATGCGGATGGCAGCGAAGCGCGTGTGATTACCTGGAAAAAATCGAT CTGCCGTATGTGGATGTGGTGAGCACCGAAAAAGAAATGATCAAACGCTTTCTGCGTGT GGTGCGTGAAAAAGATCCGGATGTGCTGATTACCTATAACGGCGATAACTTTGATTTTCG CGTATCTGAAAAACGTTGCGAAGAACTGGGCATCAAATTTACCCTGGGCCGTGATGGT AGCGAACCGAAAATTCAGCGTATGGGCGATCGTTTTGCGGTGGAAGTGAAGGGAAGGAT TCACTTCGACCTCTACCCGGTGATTGCGCGTACCATTAACCTGCCGACCTATACCCTGG AAGCGGTGTATGAAGCGGTGTTTGGCAAACCGAAAGAAAAAGTGTACGCGGAAGAAAT GCGCAGGCGTGGGAAAGCGGCGAAGGCCTGGAACGTGTGGCGCGTTATAGCATGGAGGA CGCAAAGGTAACCTATGAACTCGGCCGTGAATTTTTCCCGATGGAAGCGCAGCTGTCTC GTCTGATTGGCCAGAGCCTCTGGGATGTATCTCGCTCGAGTACCGGCAACCTGGTGGAA TGGTTTCTGCTGCGTAAAGCGTATAAACGTAACGAACTTGCACCAAACAAGCCGGACGA ACGTGAACTGGCCCGTCGTCGTGGCGGTTATGCGGGCGGTTATGTGAAAGAACCGGAAC GTGGCCTGTGGGATAACATTGTGTATCTGGATTTTCGTAGCCTGGGCCCTTCGATAATA ATCACCCATAACGTCTCCCCGGATACCCTGAACCGTGAAGGCTGCAAAGAATATGATGT GGCGCCGGAAGTGGGCCATAAATTCTGCAAAGACTTCCCCGGCTTCATCCCAAGCCTGC TGGGCGATCTGCTGGAAGAACGCCAGAAAATCAAACGCAAAATGAAAGCGACCGTTGAT CCGCTGGAAAAAAACTGCTGGATTATCGTCAGCGCCTGATTAAATTTCTGGCCAACAG CTTCTATGGCTATTATGGTTATGCGAAAGCGCGTTGGTATTGCAAGGAGTGCGCCGAGA GCGTTACCGCGTGGGGCCGTGAATATATCGAAATGGTGATCCGCGAACTGGAAGAAAAA TTTCGGCTTCAAAGTGCTGTATGCGGATACCGATGGCCTGCATGCGACCATTCGGGGTGC GGATGCGGAAACCGTGAAAAAAAAGCGAAAGAATTCCTGAAATACATCAATCCGAAAC TGCCGGGCCTGCTCGAACTCGAATACGAGGGCTTCTACGTGCGTGGCTTTTTTCGTGACC AAAAAAAATACGCGGTGATCGATGAAGAAGGCAAAATTACCACCCGTGGGCTTGAAAT
--	--

	AGTTAGGCGTGACTGGAGCGAAATTGCGAAAGAAACCCAGGCGCGTGTGCTGGAAGCGA TTCTGAAACATGGCGATGTGGAAGAAGCGGTGCGTATTGTCAAAGAGGTTACGGAGAAG CTGAGCAAATATGAAGTTCCGCCGGAAAACTGGTGATTCTATAAACAAATTACCCGTGA TCTGCGTGATTATAAAGCGACCGGTCCGCATGTGGCGGTTGCAAACGCCTCGCCGCGC GTGGCGTGAAAATTCGTCCGGGCACCGTGATTAGCTATATTGTGCTGAAAGGCAGCGGC CGTATTGGCGATCGTGCGATTCCGGCGGATGAATTTGATCCGACCAAACATCGTTATGA TGCGGAATATTATATCGAGAACCAGGTTCTTCCAGCTGTGGAACGTATTCTGAAAGCGT TTGGCTATCGTAAAGAAGATCTGCGCTATCAGAAAACCAAACAGGTGGGCCTGGGCGCG TGGCTGAAAGTTAAAGGCAAA
DV-QGLK	ATGATACTTGACGCTGACTACATCACCGAGGATGGGAAGCCGATTATAAGGATTTTCAA GAAGGAGAACGGCGAGTTCAAGGTTGAGTACGACAGAACTTTAGACCTTACATTTACG CGCTCTTGAAGGACGACTCTCAGATTGATGAGGTTAGGAAGATAACCGCCGAGAGGCAT GGGAAGATAGTGAGAATTATAGATGCCGAAAAGGTAAGGAAGAAGTTCCTAGGCAGGCC GATTGAGGTATGGAGGCTGTACTTTGAACACCCTCAGGACCAGCCCGCAATAAGGGATA AGATAAGAGAGCATTCCGCAGTTATTGACATCTTTGAGTACGACATCCCCCTTCGCGAAG CGCTACCTAATAGACAAAGGCCTAATTCCAATGGAAGGCGATGAAGAGCTCAAGTTGCT CGCATTTGCGATCGCGACGCTCTATCACGAGGGGGAGGAGTTCGCGAAGGGGCCCATTA TAATGATAAGCTATGCTGATGAGGAAGAAGCCAAAGTCATAACGTGGAAAAAGATCGAT CTCCCGTACGTCGAGGTAGTTTCCAGCGAGAGGGAGATGATAAAGCGGTTCTCAAGGT GATAAGGGAGAAAGATCCCGATGTTATAATTACCTACAACGGCGATTCTTTCGACCTTC CCTATCTAGTTAAGAGGGCCGAAAAGCTCGGGATAAAGCTACCCCTGGGAAGGGACGGT AGTGAGCCAAAGATGCAGAGGCTTGGGGATATGACAGCGGTGGAGATAAAGGGAAGGAT TCACTTCGACCTCTACCACGTGATTAGGAGAACGATAAACCTCCCAACATACACCCTCG AGGCAGTTTATGAGGCAATCTTCGGAAAGCCAAAGGAGAAAGTTTACGCTCACGAGATA GCTGAGGCCTGGGAGACTGGAAAGGGACTGGAGAGAGTTGCAAAGTATTCAATGGAGGA

	CGCAAAGGTAACCTATGAACTCGGTAGGGAGTTCTTCCCAATGGAGGCCAGCTTTCAA GGTTAGTCGGCCAGCCCCCTCTGGGATGTATCTCGCTCGAGTACTGGCAACTTGGTGGAG TGGTACCTCCTCAGGAAGGCCTACGAGAGGAATGAACTTGCACCAAACAAGCCGGACGA GAGGGAGTACGAGAGAAGGCTAAGGGAGAGCTACGCTGGGGGATACGTTAAGGAGCCGG AGAAAGGGCTCTGGGAGGGGTTAGTTTCCCTAGATTTTCAGGAGCCTGGGCCCTTCGATA ATAATCACCCATAACGTCTCCCCGGATACGCTGAACAGGGAAGGGTGTAGGGAATACGA TGTCGCCCCAGAGGTTGGGCACAAGTTCTGCAAGGACTTCCCCGGCTTCATCCCAAGCC TGCTCAAGAGGTTATTGGATGAAAGGCAAGAAATAAAAAGGAAGATGAAAGCTTCTAAA GACCCAATCGAGAAGAAGATGCTTGATTACAGGCAACGGCTGATCAAAATCCTGGCGAA TTCCTATTATGGGTATTATGGGTACGCAAAGCCCGTTGGTACTGTAAGGAGTGCGCCG AGAGCGTTACGGCCTGGGGGAGGGAATATATAGAGTTCGTAAGGAAGGAAGTGGAGGAA AAGTTCGGGTTCAAAGTCTTATACATAGACACAGATGGACTCTACGCCACAATTCCTGG GGCAAACCCGAGGAGATAAAGAAGAAAGCCCTAGAGTTCGTAGATTATATAAACGCCA AGCTCCCAGGGCTGCTCGAACTCGAATACGAGGGCTTCTACGTGAGAGGGTTCTTCGTG ACGAAGAAGAAGTATGCGTTGATAGATGAGGAAGGGAAGATAATCACTAGGGGGCTTGA AATAGTTAGGCGTGACTGGAGCGAAATAGCCAAAGAAACCCAAGCAAAAGTCCTAGAGG CTATCCTAAAGCATGGCAACGTTGAGGAGGCAGTAAAGATAGTCAAAGAGGTTACGGAG AAGCTGAGCAAGTACGAAATACCTCCAGAAAAGCTAGTTATTTACAAACAGATCACGAG GCCCCTTCACGAGTACAAGGCTATAGGTCCGCACGTTGCCGTTGCAAAACGCCTCGCCG CTAGAGGAGTAAAGGTGAGGCCTGGCATGGTGATAGGGTACATAGTGCTGAGGGGAGAC GGGCCAATAAGCAAGAGGGCTATCCTTGCAAGGAGTTCGATCTCAGGAAGCATAAGTA TGACGCTGAGTATTACATCGAGAACCAGGTTCTTCCAGCTGTGCTTAGAATATTAGAGG CCTTTGGGTACAGGAAAGAAGACCTCAGGTGGCAGAAGACTAAACAGACAGGTCTTACG GCATGGCTTAACATCAAGAAGAAG
--	--

Kod-QGLK	ATGATCCTCGACACTGACTACATAACCGAGGATGGAAAGCCTGTCATAAGAATTTTCAA GAAGGAGAACGGCGAGTTCAAGATTGAGTACGACCGGACTTTTGAACCCTACTTCTACG CGCTCTTGAAGGACGACTCTGCCATTGAGGAAGTCAAGAAGATAACCGCCGAGAGGCAC GGGACGGTTGTAACGGTTAAGCGGGTTGAAAAGGTTTCAGAAGAAGTTCCTAGGCAGGCC GGTTGAGGTCTGGAACTCTACTTTACTCATCCGCAGGACCAGCCAGCGATAAGGGACA AGATACGAGAGCATCCAGCAGTTATTGACATCTACGAGTACGACATCCCCTTCGCGAAG CGCTACCTCATAGACAAGGGATTAGTGCCAATGGAAGGCGACGAGGAGCTGAAAATGCT CGCCTTCGCGATCGCGACGCTCTATCACGAGGGCGAGGAGTTCGCCGAGGGGCCAATCC TTATGATAAGCTACGCCGACGAGGAAGGGGCCAGGGTGATAACTTGGAAGAACGTGGAT CTCCCCCTACGTTGACGTCGTCTCGACGGAGAGGGAGATGATAAAGCGCTTCCTCCGTGT TGTGAAGGAGAAAGACCCGGACGTTCTCATAACCTACAACGGCGACAACCTTCGACTTCG CCTATCTGAAAAAGCGCTGTGAAAAGCTCGGAATAAACTTCGCCCTCGGAAGGGATGGA AGCGAGCCGAAGATTCAGAGGATGGGCGACAGGTTTGCCGTCGAAGTGAAGGGAAGGAT TCACTTCGACCTCTACCCTGTGATAAGACGGACGATAAACCTGCCCACATACACGCTTG AGGCCGTTTATGAAGCCGTCTTCGGTCAGCCGAAGGAGAAGGTTTACGCTGAGGAAATA ACCACAGCCTGGGAAACCGGCGAGAACCTTGAGAGAGTCGCCCCGCTACTCGATGGAGGA CGCAAAGGTAACCTATGAACTCGGGAAGGAGTTCCTTCCGATGGAGGCCCAGCTTTCTC GCTTAATCGGCCAGTCCCTCTGGGATGTATCTCGCTCGAGTACTGGCAACCTCGTTGAG TGTTTCCTCCTCAGGAAGGCCTATGAGAGGAATGAACTTGCACCAAACAAGCCGGACGA AAAGGAGCTGGCCAGAAGACGGCAGAGCTATGAAGGAGGCTATGTAAAAGAGCCCGAGA GAGGGTTGTGGGAGAACATAGTGTACCTAGATTTTAGATCCCTGGGCCCTTCGATAATA ATCACCCATAACGTCTCCCCGGATACGCTCAACAGAGAAGGATGCAAGGAATATGACGT TGCCCCACAGGTCGGCCACCGCTTCTGCAAGGACTTCCCCGGCTTCATCCCAAGCCTGC TTGGAGACCTCCTAGAGGAGAGGCAGAAGATAAAGAAGAAGATGAAGGCCACGATTGAC CCGATCGAGAGGAAGCTCCTCGATTACAGGCAGAGGCTGATCAAGATCCTGGCAAACAG
----------	--

	CTACTACGGTTACTACGGCTATGCAAGGGCGCGCTGGTACTGCAAGGAGTGCGCCGAGA GCGTTACGGCCTGGGGAAGGGAGTACATAACGATGACCATCAAGGAGATAGAGGAAAAG TACGGCTTTAAGGTAATCTACAGCGACACCGACGGATTTTTTGGCCACAATACCTGGAGC CGATGCTGAAACCGTCAAAAAGAAGGCTATGGAGTTCCTCAAGTATATCAACGCCAAAC TTCCGGGCGCGCTCGAACTCGAATACGAGGGCTTCTACAAACGCGGCTTCTTCGTCACG AAGAAGAAGTATGCGGTGATAGACGAGGAAGGCAAGATAACAACGCGCGGGCTTGAAAT AGTTAGGCGTGACTGGAGCGAGATAGCGAAAGAGACGCAGGCGAGGGTTCTTGAAGCTT TGCTAAAGGACGGTGACGTCGAGAAGGCCGTGAGGATAGTCAAAGAGGTTACGGAGAAG CTGAGCAAGTACGAGGTTCCGCCGGAGAAGCTGGTGATCCACAAACAGATAACGAGGGA TTTAAAGGACTACAAGGCAACCGGTCCCCACGTTGCCGTTGCAAAACGCCTCGCCGCGA GAGGAGTCAAATACGCCCTGGAACGGTGATAAGCTACATCGTGCTCAAGGGCTCTGGG AGGATAGGCGACAGGGCGATACCGTTTCGACGAGTTCGACCCGACGAAGCACAAGTACGA CGCCGAGTACTACATCGAGAACCAGGTTCTTCCAGCTGTGGAGAGAATTCTGAGAGCCT TCGGTTACCGCAAGGAAGACCTGCGCTACCAGAAGACGAGACAGGTTGGTTTGAGTGCT TGGCTGAAGCCGAAGGGAAC
--	---

Supplementary Table 2. DNA Oligonucleotides. Oligonucleotides are written in the 5'→3' direction. Modifications are denoted using IDT nomenclature. Templates used for fidelity and template mismatch experiments have their mismatch/watermark positions highlighted in **RED**.

Sequence Name	Sequence (5'→ 3')
Recomb Frag 1 Fwd	ATCACCATACGGATCCAAGCGGC
Recomb Frag 1 Rvs	AGAGTCGTCCTTCAAGAGCGCGTA
Recomb Frag 2 Fwd	TACGCGCTCTTGAAGGACGACTCT
Recomb Frag 2 Rvs	CGGCCTGCCTAGGAACCTCTT
Recomb Frag 3 Fwd	AAGAAGTTCCTAGGCAGGCCG
Recomb Frag 3 Rvs	CTCGTGATAGAGCGTCGCGATCGC
Recomb Frag 4 Fwd	GCGATCGCGACGCTCTATCACGAG
Recomb Frag 4 Rvs	GTAGAGGTCTGAAGTGAATCCTTCCCTT
Recomb Frag 5 Fwd	AAGGGAAGGATTCACTTCGACCTCTAC
Recomb Frag 5 Rvs	GAGTTCATAGGTTACCTTTGCGTCCTC
Recomb Frag 6 Fwd	GAGGACGCAAAGGTAACCTATGAACTC
Recomb Frag 6 Rvs	GTCCGGCTTGTTTGGTGCAAGTTC
Recomb Frag 7 Fwd	GAAGTTCACCAACAAGCCGGAC
Recomb Frag 7 Rvs	TGGGATGAAGCCGGGGAAGTC
Recomb Frag 8 Fwd	GACTTCCCCGGCTTCATCCCA
Recomb Frag 8 Rvs	AACGCTCTCGGCGCACTCCTT
Recomb Frag 9 Fwd	AAGGAGTGCGCCGAGAGCGTT
Recomb Frag 9 Rvs	GTAGAAGCCCTCGTATTGAGTTGAG
Recomb Frag 10 Fwd	CTCGAACTCGAATACGAGGGCTTCTAC
Recomb Frag 10 Rvs	GCTCAGCTTCTCCGTAACCTCTTTGAC
Recomb Frag 11 Fwd	GTCAAAGAGGTTACGGAGAAGCTGAGC
Recomb Frag 11 Rvs	CACAGCTGGAAGAACCTGGTTCTCGAT
Recomb Frag 12 Fwd	ATCGAGAACCAGGTTCTTCCAGCTGTG
Recomb Frag 12 Rvs	TAAGCTTGGCTGCAGGTGCACTCA
pGDR11 Vector Fwd	TGAGTCGACCTGCAGCCAAGCTTA
pGDR11 Vector Rvs	GCCGCTTGGATCCGTATGGTGAT

5' Phos/cEM619_RNA	/5Phos/rArArCrUrUrUrArUrUrArArGrUrGrUrGrUrGrUrUrArGrCrGrArUrArGrGrArGrGrUrGrUrGrGrG
pET21_fw	GATCCGGCTGCTAACAAAGCCC
pET21_rv	CCATATGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGGGA
pET21_5ovr_fw_C28-28	TTTTGTTTAACTTTAAGAAGGAGATATACATATGATACTTGACGCTGACTACATCACCGAG
pET21_3ovr_rv_C28-28	TTCCTTTCGGGCTTTGTTAGCAGCCGGATCTCATGTCTTAGGTTTTAGCCAAGCA
Crystallography Template	/5Cy5/AAACGTACGCAGTTCGCG
Crystallography Primer	CGCGAACTGCG

Supplementary Table 3: Droplet Information and Incubation Conditions.

Round	Library	Top Variant	Fluorescent Threshold (afu)	Incubation 55 °C (h)	Droplets Detected	Droplets Sorted	Colonies Screened
1	Recomb.	-	50	18	34,776,598	754,628	100
2	Recomb.	-	50	8	21,429,314	313,260	100
3	Recomb.	-	50	4	15,801,994	56,632	100

Supplementary Table 4: Selected C28 mutations relative to ancestral enzymes.

	Kod-QGLK (Exo⁻)	Tgo-QGLK (Exo⁻)	9N-QGLK (Exo⁻)	DV-QGLK (Exo⁻)
NTD	T5A, V15I, I28V, T33N, E35R, F38I, E50D, V62T, T64R, K66V, V68A, Q72K, V80I, T88N, K99R, E102A, I107V, V127I	T5A, V15I, I28V, D29E, E35R, T88N, K99R, K101R, E102A	T5A, N11D, V15I, V18I, I28V, T33N, E35R, F38I, V54I, K57E, V62T, K64R, K66V, Q72K	Q47A, D49E, E50D, R52K, K61T, I62T, I65V, I66V, D67R, R72K, R84K, E88N, K99R, E102A, S104P, I107V, F110Y
Exo	K136T, A153G, E154T, E165G, E166S, G167E, N175K, V176I, R188K, K199R, K225E, N229K, A231T, V282I, T296A, T297Q, N304G	K136T, A153G, E154T, E165G, E166S, G167E, N175K, K196R, K199R, S223C, K225E, V228I, I231T, E235D	V282I, K285Q, S301T, R324K	K136T, L137M, A153G, K154T, I158L, E165G, E166S, K169R, E182D, S186T, R188K, K196R, I198V, I206L, S213N, L216F, P217A, V220K, A223C, K225E, L230F, P231T, M241I, L244M, M247R, T248F, I252V, H262P, K285Q, H293E, E297Q, K303E, K310R, R324K
NTD		F327L, V337I	F327L, K363E	F327L, V337I, P340S, Y356F
Palm	K375R, L377Y, A378E, 380-381L, Q383E, E386A, R395K, N400G, I401L, Y403S, K430R, Q437E, R441K	L377Y, A378E, 380-381L, R395K, N400G, I401L, Y403S, E430R, Q437E	L377Y, A378E, 380-381 L, G383E, G384S, R395K, D399E, N400G, I401L, Y403S, K430R	
Finger	I464V, R477K, Y494F		I464V, R466K, V472I, L475I	K455G, R456D, D459E, E463K, I464V, R466K, S471T, K472I, M479L, Y494F
Palm	R502K, E520Q, T523E, M524T, K527R, Y533F, I538L, S540A, M562K, K566D, A575L		E520Q, M524T, V525T, L529I, L545F, H546F, K566D, P570A, V585K	E520Q, F524T, V525T, R526I, K527R, L529I, I540A, L545F, Y546F, K553D, P554A, E556T, I557V, L562K, V565L, V585K
Thumb	G602D, L631I, D634H, K639E, H664Y, F749S, G750H, G774del	I692V, A724T, F749S, G750H, G767S	G602D, H664Y, R672K, A718F, R727K, K747R, F749S, G750H, K762R, G767S, V772P G774del, K775T	L597V, G602D, I605T, K626R, N636D, K642R, I656V, P670D, H672K, E673D, I667T, V694I, M698T, G701S, R706K, D708S, P710R, S712G, K713D, L717P, A718F, E719D, L723P, R724T, L743E, E747R, F749S, G750H, W758Y, K762R, T764V, T767S, N771K, I772del, K773P, K775T

Supplementary Table 5. Crystallographic Data Collection and Refinement Statistics.

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

Data Collection	
Resolution range (Å)	64.03 - 2.22 (2.299 - 2.22)
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Unit cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	77.9698 102.34 112.23
<i>a</i> , <i>b</i> , <i>c</i> (°)	90.0, 90.0, 90.0
Total reflections	89976 (8829)
Unique reflections	45059 (4429)
Multiplicity	2.0 (2.0)
Completeness (%)	99.23 (99.84)
Mean <i>I</i> / σ <i>I</i>	5.21 (0.90)
Wilson B-factor	38.50
R _{merge}	0.07752 (0.817)
R _{meas}	0.1096 (1.155)
R _{pim}	0.07752 (0.817)
CC1/2	0.987 (0.308)
CC*	0.997 (0.687)
Refinement	
Reflections used in refinement	44719 (4423)
Reflections used for R _{free}	1998 (198)
R _{work}	0.2412 (0.3407)
R _{free}	0.2722 (0.3975)
CC _{work}	0.934 (0.568)
CC _{free}	0.921 (0.406)
Number of non-hydrogen atoms	6715
Macromolecules (Protein/DNA/U12)	6610
Solvent (Water)	105
Protein residues	747
RMS _{Bonds}	0.004
RMS _{Angles}	0.63
Ramachandran favored (%)	96.11
Ramachandran allowed (%)	3.89
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	1.24
Clashscore	9.65
Average B-factor	55.97
Macromolecules	55.97
Solvent	44.48
Number of TLS groups	7

Supplementary Table 6. Local Base-Pair Parameters.

Base Pair	Shear (Å)	Stretch (Å)	Stagger (Å)	Buckle (Å)	Propeller (°)	Opening (°)
dA7 - U12	0	-0.23	-0.2	-23.86	0.33	-1.07
dC8 - dG11	0	-0.1	-0.19	1.98	1.93	0.65
dG9 - dC10	-0.51	-0.29	-0.5	-12.4	-14.12	3.05
dC10 - dG9	0.31	-0.23	-0.05	-11.61	-1.66	2.67
dA11 - dT8	0.24	-0.32	0.24	6.23	-6.71	-1.64
dG12 - dC7	0.11	-0.21	0.05	1.74	-2.43	0.91
dT13 - dA6	-0.36	-0.36	-0.3	3.56	-21.9	0.77
dT14 - dA5	-0.34	-0.08	-0.68	10.33	-16.3	4.31
dC15 - dG4	-0.36	-0.17	-0.54	4.74	-16.61	5.49
dG16 - dC3	0.6	0.1	0.23	-6.93	-17.44	-8.19
dC17 - dG2	-0.25	0.07	0.95	-18.77	8.45	-3.37

Supplementary Table 7. Local Base-Pair Step Parameters.

Base Pair	Shift (Å)	Slide (Å)	Rise (Å)	Tilt (°)	Roll (°)	Twist (°)
dA7 - U12	-0.15	-0.28	2.71	3.43	7.18	25.01
dC8 - dG11	0.16	-0.4	3.7	-0.71	12.42	42.73
dG9 - dC10	0.16	-0.55	3.33	-1.54	-0.43	26.46
dC10 - dG9	-0.9	0.34	3.18	-1.45	-1.21	34.69
dA11 - dT8	0.4	-0.36	3.63	-1.58	12.03	28.27
dG12 - dC7	-0.92	-0.24	3.27	2.15	0.56	34.3
dT13 - dA6	-0.02	0.09	3.04	2.98	8.18	38.56
dT14 - dA5	0.41	0.5	3.28	1.94	3.87	34.33
dC15 - dG4	-0.93	0.91	3.78	-6.37	8.85	35.91
dG16 - dC3	0.85	0.27	3.83	-3.58	6.46	35.26

Supplementary Table 8. Torsion Angles and Pseudorotation Parameters.

Template

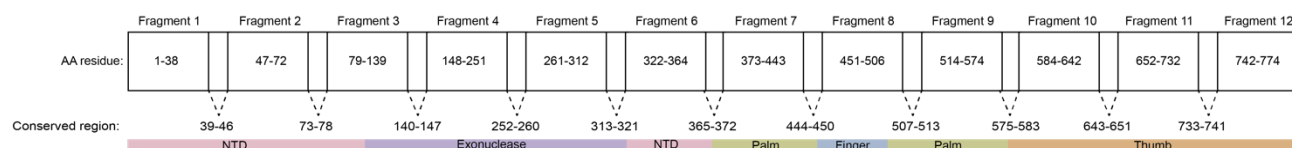
Nucleotide	ν_0	ν_1	ν_2	ν_3	ν_4	P	$\nu_{\max}$	χ	ψ	TYPE
dG5	-34.01	24.70	-7.31	-11.72	28.72	102.79	33.04	-81.61	152.53	Beta, O4'-endo
dT6	2.12	-24.53	36.13	-36.05	21.56	15.56	37.51	-160.11	56.19	Beta, C3'-endo
dA7	2.56	-23.50	34.17	-33.73	19.78	14.63	35.31	-159.17	57.35	Beta, C3'-endo
dC8	-21.52	36.72	-37.15	25.74	-2.97	165.52	38.37	-120.06	64.89	Beta, C2'-endo
dG9	-13.19	28.29	-31.76	24.96	-7.62	174.79	31.89	-111.84	45.51	Beta, C2'-endo
dC10	-39.60	29.91	-9.92	-12.41	32.70	104.91	38.54	-126.95	42.42	Beta, O4'-endo
dA11	-15.56	30.45	-32.88	24.84	-6.05	171.50	33.24	-106.14	51.49	Beta, C2'-endo
dG12	-38.51	44.55	-33.10	12.01	16.24	139.41	43.59	-108.25	38.94	Beta, C1'-exo
dT13	-28.51	39.34	-34.52	19.11	5.61	152.90	38.77	-114.72	44.36	Beta, C2'-endo
dT14	5.16	18.24	-33.13	37.06	-26.85	206.50	37.02	-91.73	-70.92	Beta, C3'-exo
dC15	-26.65	36.93	-32.69	18.22	5.03	153.39	36.56	-95.40	41.96	Beta, C2'-endo
dG16	-23.34	37.48	-36.70	24.11	-0.77	162.35	38.51	-111.01	54.36	Beta, C2'-endo
dC17	-27.69	37.15	-32.09	17.04	6.41	151.24	36.61	-95.37	-16.20	Beta, C2'-endo

Primer

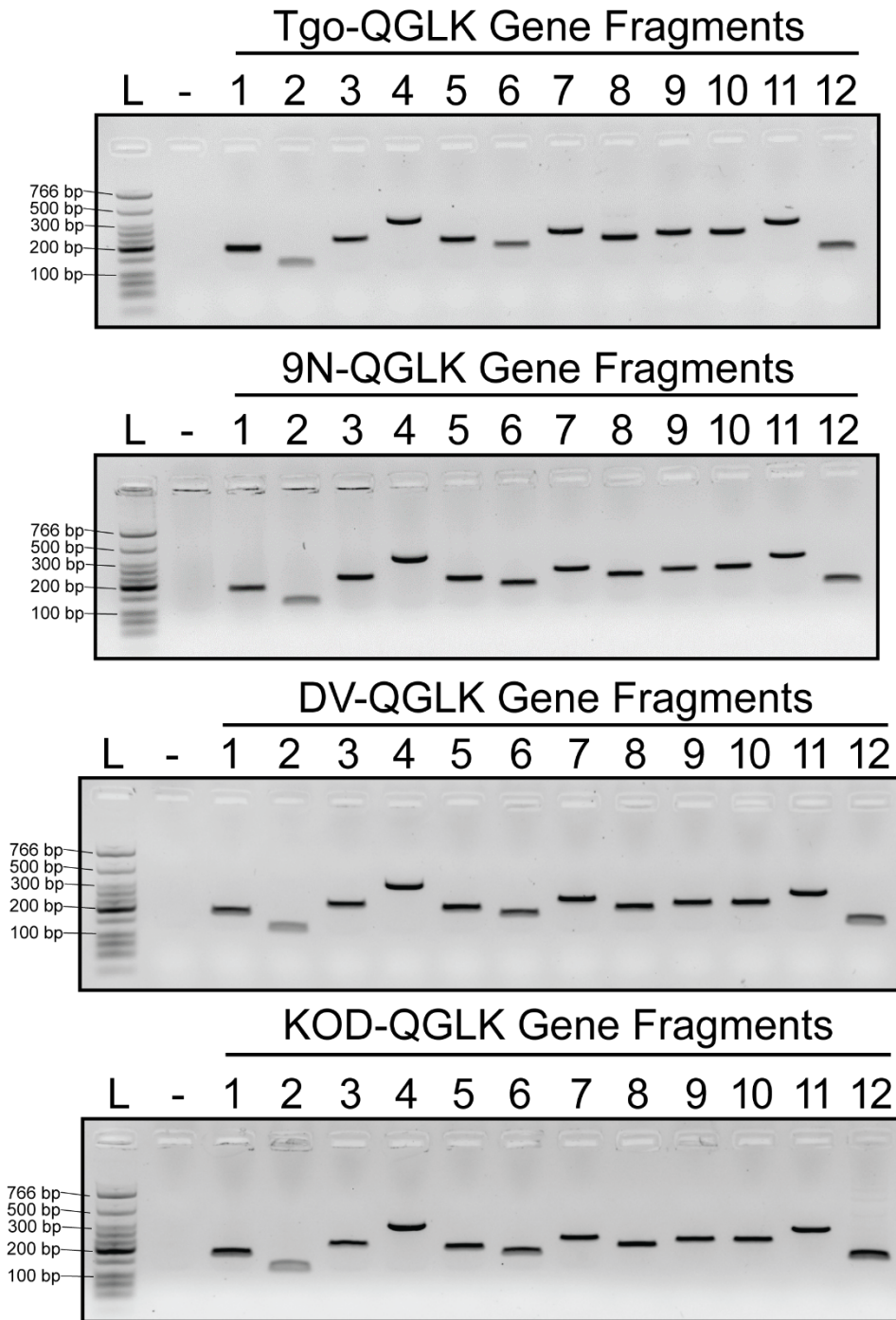
Nucleotide	ν_0	ν_1	ν_2	ν_3	ν_4	P	$\nu_{\max}$	χ	ψ	TYPE
dG2	-32.42	43.51	-37.26	19.69	7.66	150.87	42.65	-82.89	47.37	Beta, C2'-endo
dC3	-33.77	37.83	-27.70	9.09	15.21	137.64	37.48	-111.64	38.12	Beta, C1'-exo
dG4	-20.36	35.42	-36.31	25.47	-3.45	166.49	37.35	-97.67	46.30	Beta, C2'-endo
dA5	-33.10	29.26	-15.17	-3.30	22.78	117.83	32.50	-124.66	48.41	Beta, C1'-exo
dA6	-20.33	31.55	-30.29	19.42	0.33	160.61	32.11	-103.80	36.69	Beta, C2'-endo
dC7	-39.93	29.96	-9.98	-12.45	32.90	104.93	38.75	-118.96	36.11	Beta, O4'-endo
dT8	-32.62	32.58	-20.62	2.56	18.76	127.94	33.53	-118.39	45.84	Beta, C1'-exo
dG9	-31.07	36.50	-27.95	10.95	12.41	141.26	35.84	-104.35	44.47	Beta, C1'-exo
dC10	-38.96	21.97	1.91	-24.35	40.05	87.31	40.76	-143.56	59.09	Beta, O4'-endo
dG11	-5.24	20.33	-26.75	24.45	-12.29	187.73	26.99	-105.20	41.99	Beta, C3'-exo
U12	-31.16	39.41	-32.02	15.01	9.95	146.39	38.45	-102.91	47.90	Beta, C2'-endo

Tgo-QGLK	MILDTDYITE	DGKPVIRIFK	KENGEFKIDY	DRNFEPYIYA	LLKDDSAIED	VKKITAERHG	TTVRVVRAEK	VKKKFLGRPI	EVWKLYFTHP	90
Kod-QGLK			E.....	T.....F.....	E.....		V.T.K.V.....	Q.....V.....		90
9N-QGLK		N.....V.....	E.....	T.....F.....		V.K.K.....	V.K.K.....	Q.....	N.....	90
DV-QGLK	A.....	I.....	VE.....	R.....	Q.DE.....	R.....	K.I.IID.....	R.....	R.E.....	90
	100		120		140		160		180	
Tgo-QGLK	QDQPAIRDKI	KEHPAVVDIY	EYDIPFAKRY	LIDKGLIPME	GDEELKMLAF	AIATLYHEGE	EFAEGPILMI	SYADEEGARV	ITWKNIDLPI	180
Kod-QGLK	R.....	I.....		V.....					V.....	180
9N-QGLK	R.....	RA.....			T.....		GT.....	GSE.....	K.....	180
DV-QGLK	R.....	S.I.F.....		L.....			K.I.....	E.K.....	K.....	180
	200		220		240		260		280	
Tgo-QGLK	VDVSTEKEM	IKRFLKVVKE	KDPDVLITYN	GDNFDFAYLK	KRSEKLGKVF	ILGREGSEPK	IQRMGDRFAV	EVKGRIFHDL	YPVIRRTINL	270
Kod-QGLK	R.....	R.....			C.....IN.....	A.....D.....				270
9N-QGLK		R.....			C.E.I.....	T.....D.....				270
DV-QGLK	S.R.....	IR.....	I.....	S.LP.V.....	A.....I.L.....	P.....D.....	M.L.MT.....	I.....	H.....	270
	280		300		320		340		360	
Tgo-QGLK	PTYTLEAVYE	AIFGQPKKEV	YAEETIAQAW	TGEGLERVAR	YSMEDAKVTY	ELGKEFFPME	AQLSRLVGQS	LWDVSRSSSTG	NLVEWFLLRK	360
Kod-QGLK	V.....	TT.....	N.....			L.....				360
9N-QGLK	V.....	K.....	S.....			R.....	I.....			360
DV-QGLK		K.....	H.E.....	K.....		R.....	P.....		Y.....	360
	380		400		420		440		460	
Tgo-QGLK	AYERNELAPN	KPDERELARR	-RESYAGGYV	KEPERGLWEN	IVYLDFRSLG	PSIIITHNVS	PDTLNREGCE	EYDVAPQVGH	KFKCDFPGFI	449
Kod-QGLK		K.....	Q.E.....					K.....	R.....	449
9N-QGLK	K.....		GG.....	D.....				K.....	E.....	449
DV-QGLK		YE.....	L.....	K.G.....	L.S.....			R.....	E.....	450
	460		480		500		520		540	
Tgo-QGLK	PSLLGDLLEE	RQKVKKMKKA	TIDPIEKKLL	DYRQLIKIL	ANSFYGYGYG	AKARWYCKEC	AESVTAWGRQ	YIETTIREIE	EKFGFKVLYA	539
Kod-QGLK		I.....	R.....		Y.....	R.....		TM.K.....	Y.....	539
9N-QGLK		I.R.....	V.L.....					MV.L.....		539
DV-QGLK	KR.D.....	E.I.R.....	S.K.....	M.....	Y.....			FVRK.L.....	I.....	540
	560		580		600		620		640	
Tgo-QGLK	DTDGGFFATIP	GADAETVKKK	AKEFLDYINA	KLPGLLELEY	EGFYKRGFFV	TKKKYAVIDE	EDKITTRGLE	IVRRDWSEIA	KETQARVLEA	629
Kod-QGLK			M.K.....	A.....			G.....			629
9N-QGLK	LH.....		K.P.....		V.....		G.....			629
DV-QGLK	LY.....	KP.EI.....	L.V.....	V.....		L.....	G.I.....		K.....	630
	640		660		680		700		720	
Tgo-QGLK	ILKHGDVEEA	VRIVKEVTEK	LSKYEVPPEK	LVIIYQITRD	LKDYKATGPH	VAVAKRLAAR	GIKIRPGTVI	SYIVLKGSGR	IGDRAIPFDE	719
Kod-QGLK	D.....K.....			H.....			V.....			719
9N-QGLK				H.....	R.....		V.....		A.....	719
DV-QGLK	N.....	K.....	I.....	P.....	HE.I.....		V.V.M.....	G.R.D.P.....	SK.LAE.....	720
	740		760		780		800		820	
Tgo-QGLK	FDPAKHKYDA	EYYIENQVLP	AVERILRAFG	YRKEDLRYQK	TRQVGLGAWL	KPK-T	773			
Kod-QGLK	T.....				S.....	G.....	774			
9N-QGLK	T.R.....		K.....		K.....	V.GK	774			
DV-QGLK	LR.....		L.E.....	W.....	K.T.T.....	NI.KK	775			

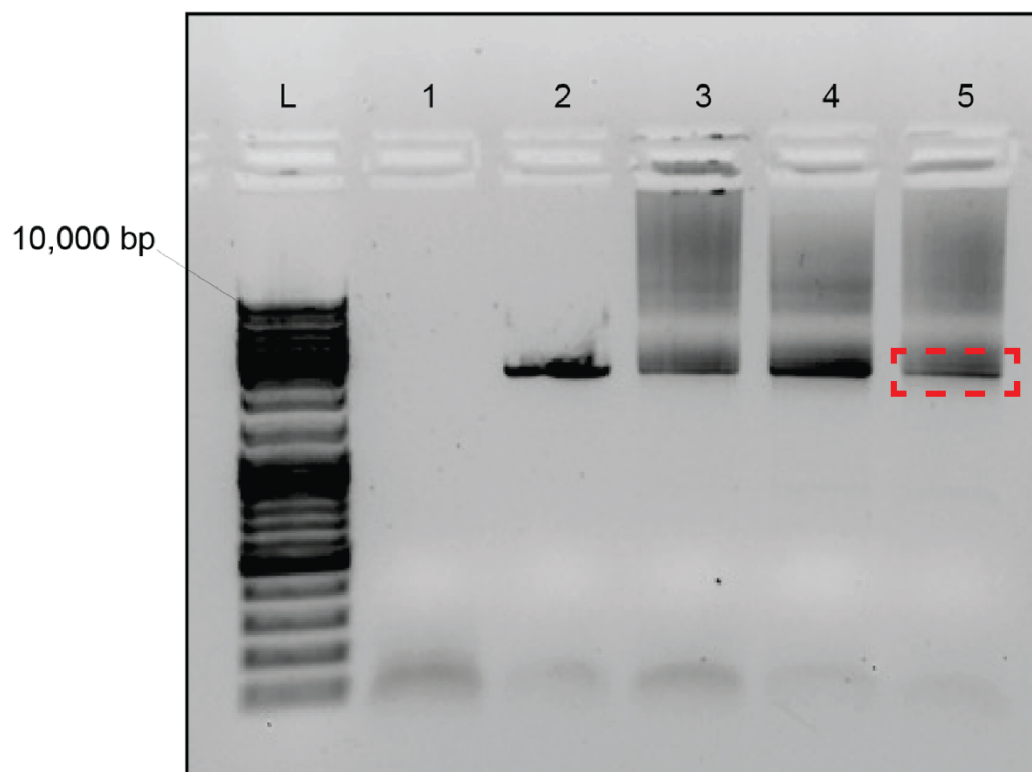
Supplementary Figure 1. Sequence Alignment of B-Family DNA Polymerases. Parent enzymes chosen for homologous recombination all contain four gain-of-function mutations (V93Q, Y409G, A485L, and E664K) and two exonuclease silencing mutations (D141A and E143A). Conserved amino acid residues across all polymerase scaffolds are denoted as (.). Natural amino acid variation from Tgo is provided as single letter amino acid codes at designated positions. Amino acid alignment was generated using CLC Genomics Workbench.



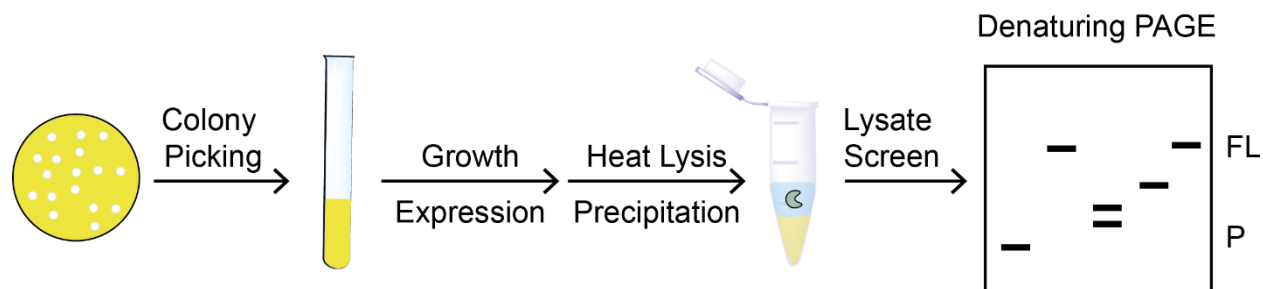
Supplementary Figure 2. C28 Fragment Boundaries Reported at the Amino Acid Level and Mapped to the Domain and Subdomain Regions of the Enzyme. Values inside the boxes correspond to residues that derive from one of the four parent enzymes. Values shown below with dotted lines correspond to the fixed regions used at the DNA level for recombination by PCR.



Supplementary Figure 3. Library Fragmentation by PCR. The polymerase scaffolds (Tgo, 9^N, DV, and Kod) used for recombination were divided into 12 distinct fragments that were separately amplified by PCR, as visualized by ethidium-bromide agarose gel electrophoresis. L: Low molecular weight ladder; -: negative control; 1-12: DNA fragments for homologous recombination.



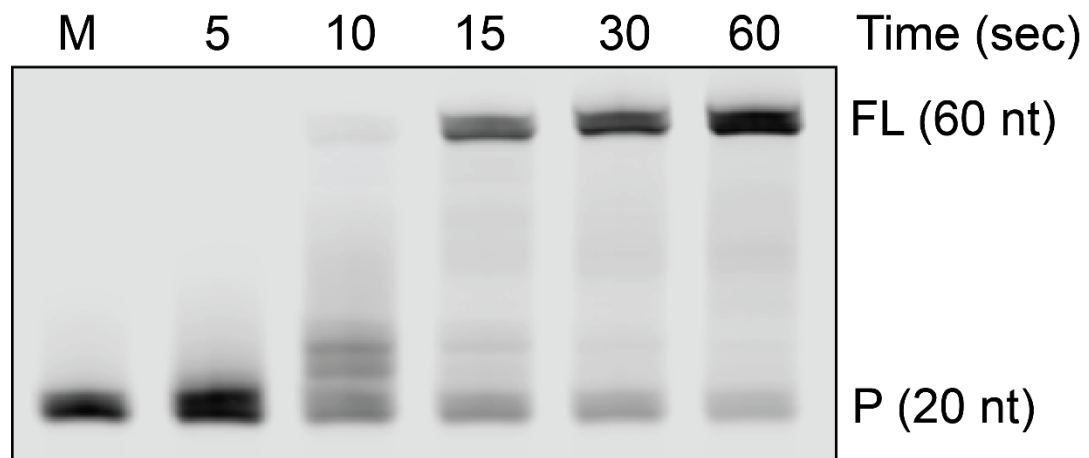
Supplementary Figure 4. Library Assembly by PCR. Agarose gel showing assembly of the homologous recombination library. Lane L: DNA ladder. Lane 1: No polymerase control. Lane 2: Plasmid amplification control. Lane 3: Recombination of Tgo-QGLK and 9°N-QGLK. Lane 4: Recombination of Kod-QGLK, DV-QGLK, and Tgo-QGLK. Lane 5: Recombination of Kod-QGLK, Tgo-QGLK, DV-QGLK and 9°N-QGLK. Red box indicates the band that was gel purified.



Supplementary Figure 5. Schematic of the Lysate Screen. Regenerated libraries are transformed into XL10-Gold *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.

[illegible]

RNA Primed Synthesis

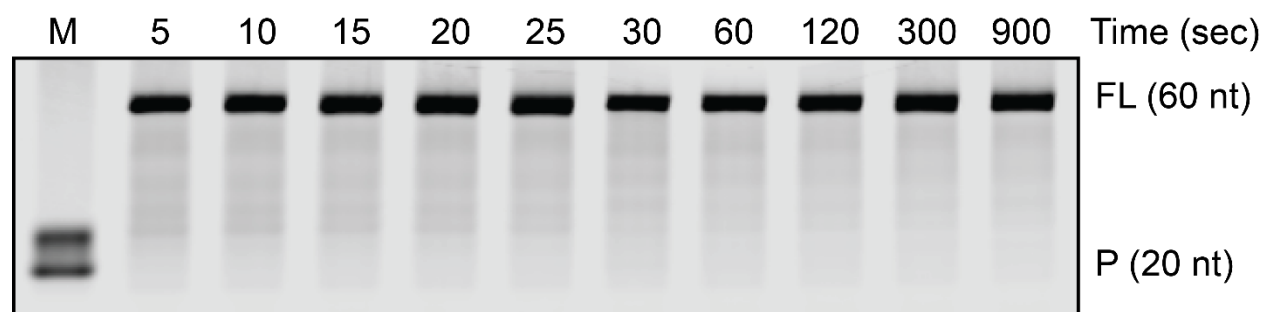


Supplementary Figure 7. RNA-Primed RNA Synthesis. Primer extension time course was performed to evaluate RNA synthesis efficiency from an RNA primer. Reactions were resolved by denaturing polyacrylamide gel electrophoresis and visualized with a LiCOR imager. M: marker lane; FL: full-length product; P: primer. Representative gels of two independent experiments are shown.

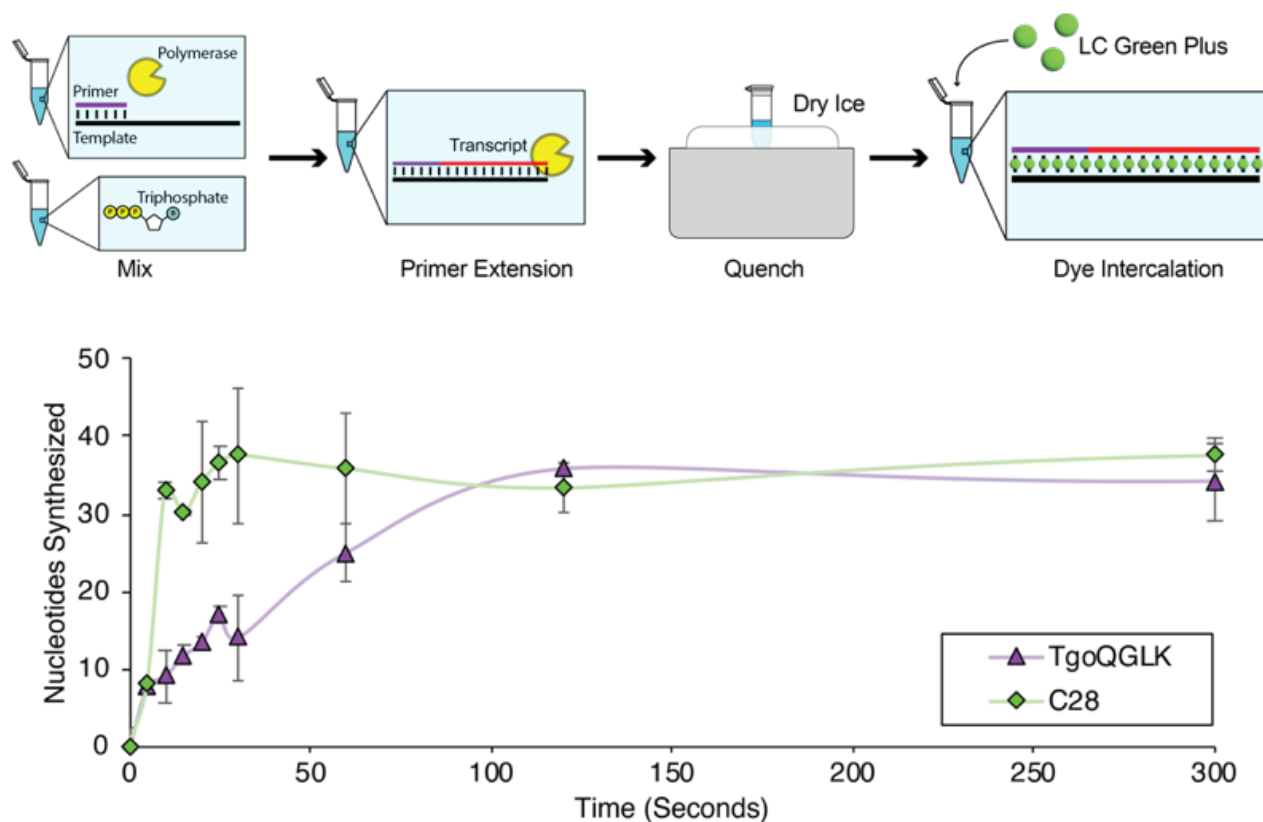


Supplementary Figure 8. Long-Range RNA Synthesis. A linearized 1009 nt template encoding a GFP mRNA containing the T7 promoter sequence and T7 terminator sequence was synthesized by C28 with RNA triphosphates in 2-hours. Two DNA markers were produced using Tgo-WT, a 200 bp DNA sizing marker, and a full DNA transcript of the GFP mRNA sequence as reference for full-length synthesis. Results were visualized on a 3% polyacrylamide gel (2 h, 16 W) on a LiCor imager. M: marker lane; FL: full-length product. Representative gels of two independent experiments are shown.

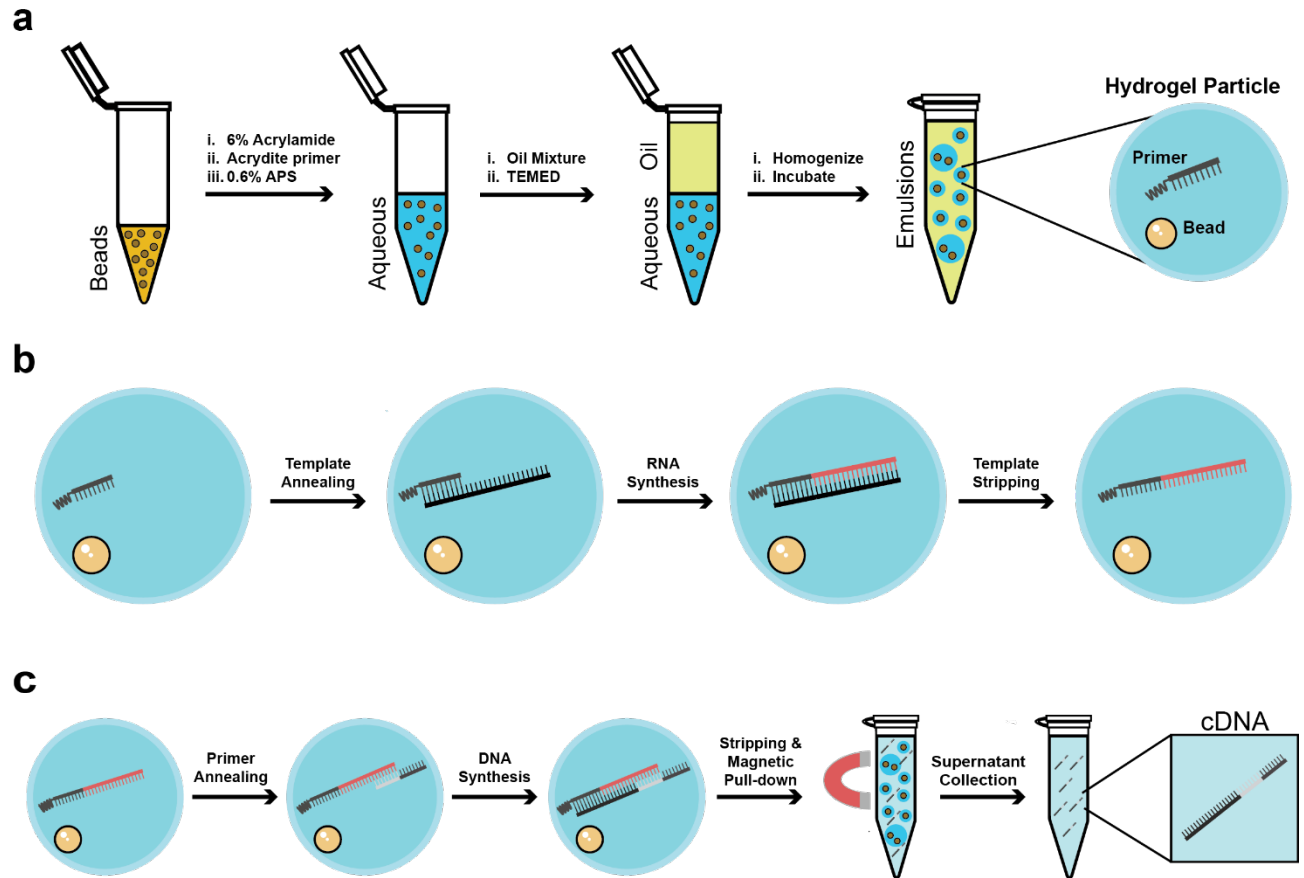
Kinetic Time-Course



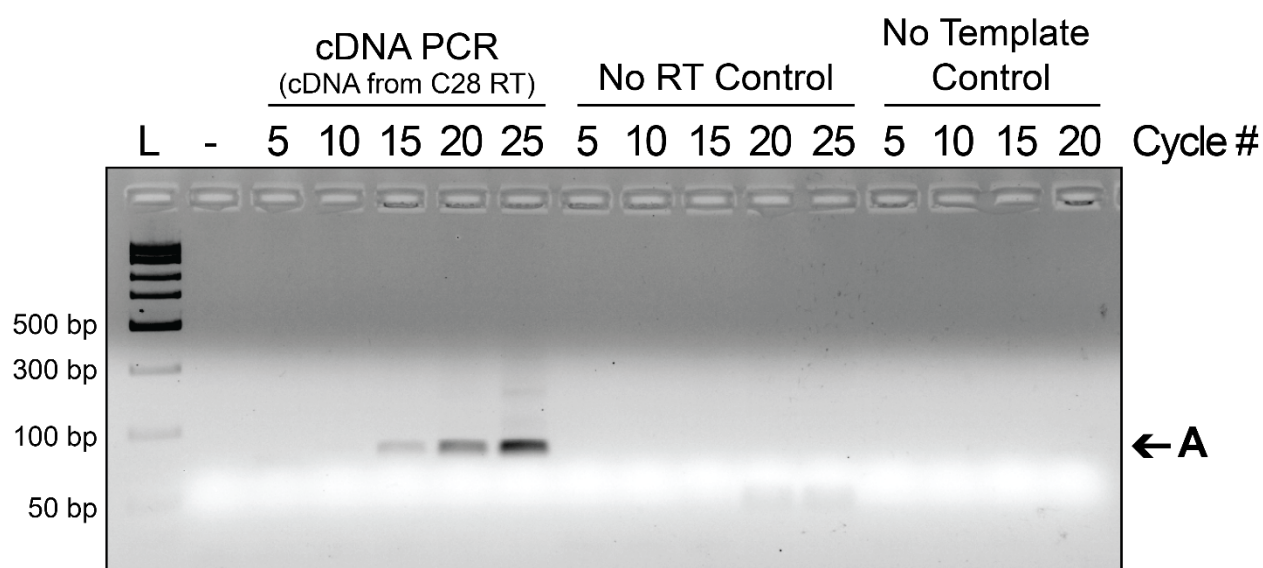
Supplementary Figure 9. Time-course Analysis. Primer extension time course was performed to emulate the kinetic assay with RNA polymerase C28. Reactions were resolved by denaturing polyacrylamide gel electrophoresis and visualized with a LiCOR imager. M: marker lane; FL: full-length product; P: primer. Representative gels of two independent experiments are shown.



Supplementary Figure 10. Kinetics Analysis of RNA Polymerases. Schematic illustration of the kinetic assay used to measure RNA synthesis. RNA polymerases are equilibrated to 55°C in buffer containing the pre-annealed primer-template duplex. Primer extension reaction is initiated with pre-heated triphosphate mix and reactions are quenched at designated time points on powdered dry ice. Samples were thawed on a 4°C cooling block, mixed with intercalating dye, and analyzed on a plate reader. Kinetic curves are shown for each reaction time point using Tgo-QGLK (purple) and RNA pol C28 (green). The slope of the linear range for each kinetic curve corresponds to the rate of RNA synthesis in nucleotides per minute. Error bars denote the standard deviation for 2 independent replicates.



Supplementary Figure 11. Schematic of the Fidelity Assay. **a**, Hydrogel coated particles are prepared by combining magnetic beads (brown) with 6% acrylamide, acrydite primer, and 0.6% APS. An oil mixture [KF6038, mineral oil, and DMF H6CS] is added, forming a biphasic layer that is homogenized after the addition of TEMED. **b**, Primer functionalized hydrogel particles are used to produce RNA coated particles by primer extension. A single-stranded DNA template is hybridized to the primer and extended with an engineered RNA polymerase and NTPs. The DNA template is then stripped to display the RNA strand. **c**, The displayed RNA strand is hybridized to an overhang DNA primer and reverse transcribed with Bst DNA polymerase and dNTPs. cDNA is stripped off the hydrogel particles by NaOH treatment, amplified by PCR, TOPO cloned, and sequenced.



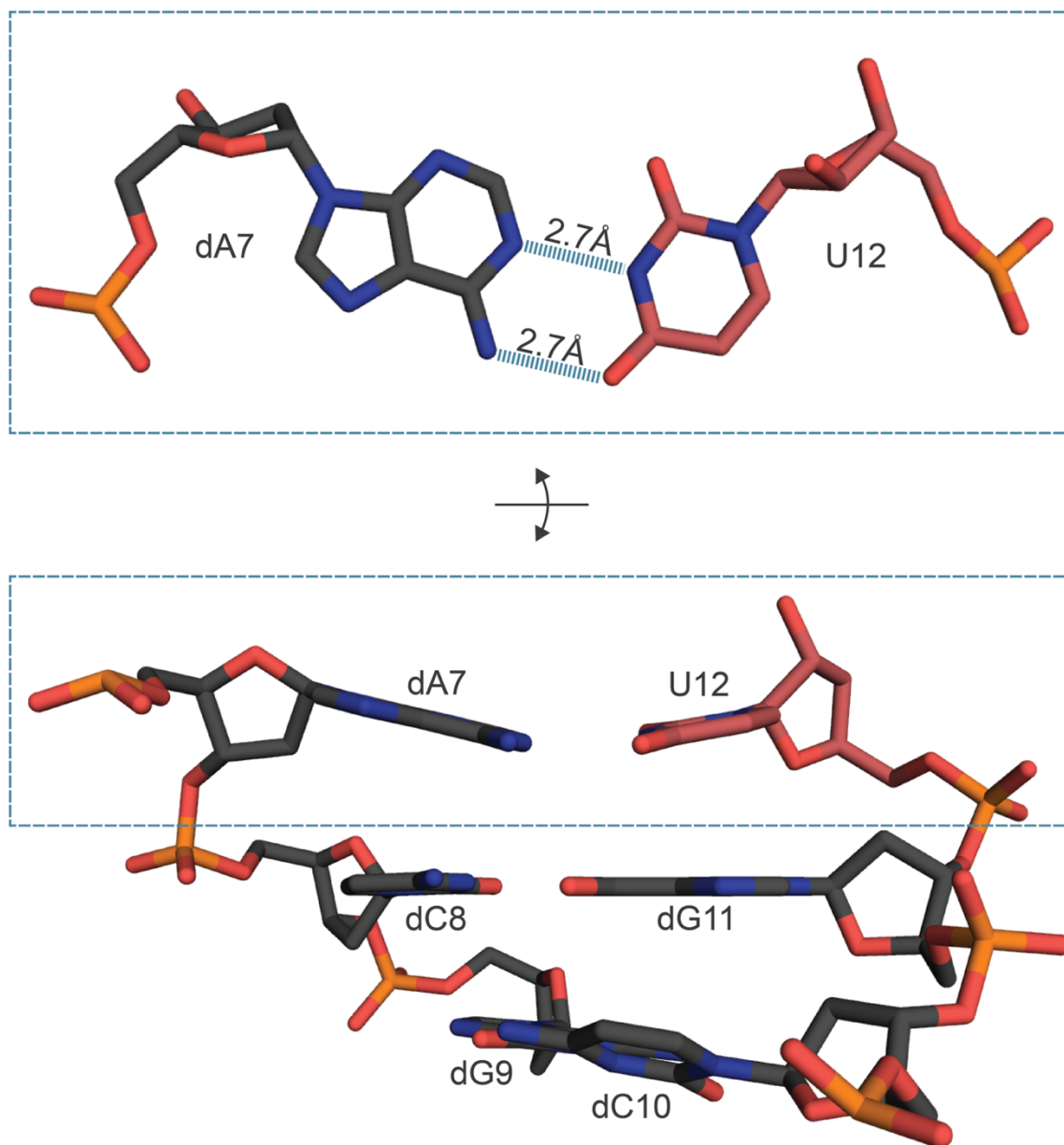
Supplementary Figure 12. cDNA Amplification for the Fidelity Assay. PCR controls are used to assess the background contamination in fidelity assays. The cDNA PCR analyzes the amplification of cDNA resulting from RNA reverse transcription. PCR was performed with collected supernatant from RNA-coated hydrogel particles before reverse transcription. This control is referred to as the ‘no RT control’, to demonstrate that there is no DNA contamination and/or a lack of non-specific amplification of any DNA template contaminant. The second control evaluates the reverse transcription reagents using supernatant from a hydrogel transcription reaction that contained no DNA template. We refer to this control as the ‘no template control’. The above results show that PCR amplification is a result of reverse transcription of the XNA template into cDNA and not the result of DNA contamination. L, DNA ladder; -, negative control; A, amplicon.

4nt9g	T	T	A	A	G	A	C	T	C	G	C	C	A	T	G	T	T	A	C	G	A	T	C	T	G	C	C	A	A	G	T	A	C	A	G	C	C	T	T	G	A	A	T	C	G	T	C	A	C	T	
Colony 1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Colony 2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Colony 3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Colony 4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Colony 5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Colony 6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Colony 7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Colony 8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Colony 9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Colony 10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Colony 11	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Colony 12	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Colony 13	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Colony 14	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Colony 15	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Colony 16	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Colony 17	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Colony 18	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Colony 19	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Colony 20	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Colony 21	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Colony 22	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Colony 23	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Colony 24	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

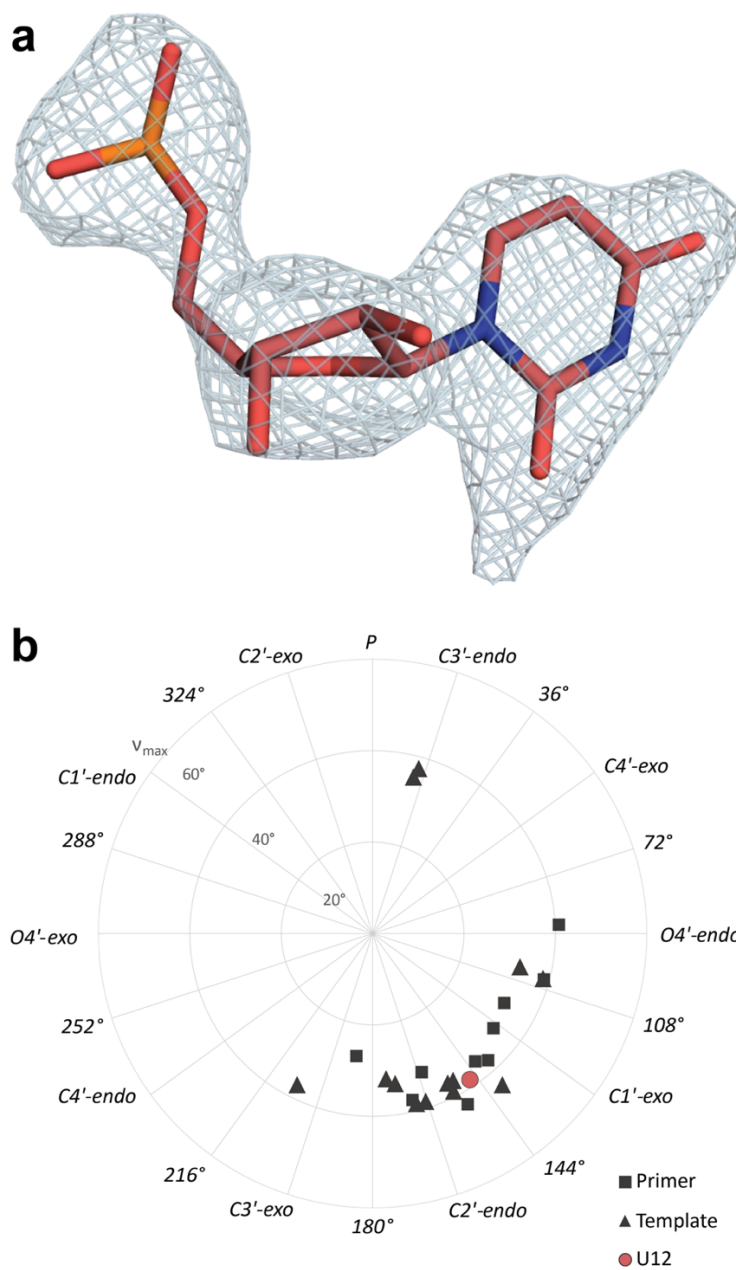
Supplementary Figure 13. Raw Fidelity Data for RNA Replication with C28 and Bst. Sequence alignment of TOPO cloned cDNA from a complete cycle of RNA replication using an engineered RNA polymerase. Raw data was aligned with the expected 4nt9g sequence (colored) displayed in the 5'→3' direction. Correct nucleotide insertions are denoted as (.) and deletions and/or insertions are denoted as (-). Incorrect nucleotide incorporations are shown as the observed base (A, C, G, or T) at specified locations. The calculated aggregate fidelity of C28 is 99.8% (1 misincorporation, 1 deletion) from a total of 1,200 nucleotide incorporations.

4nt9g	T	T	A	G	A	C	T	C	G	C	C	A	T	G	T	T	A	C	G	A	T	C	T	G	C	C	A	A	G	T	A	C	A	G	C	C	T	T	G	A	A	T	C	G	T	C	A	C	T		
Colony 1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Colony 2	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Colony 3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.
Colony 4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Colony 5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Colony 6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Colony 7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Colony 8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Colony 9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Colony 10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Colony 11	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Colony 12	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Colony 13	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Colony 14	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Colony 15	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Colony 16	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Colony 17	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Colony 18	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Colony 19	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Colony 20	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Colony 21	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Colony 22	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Colony 23	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.</																																		

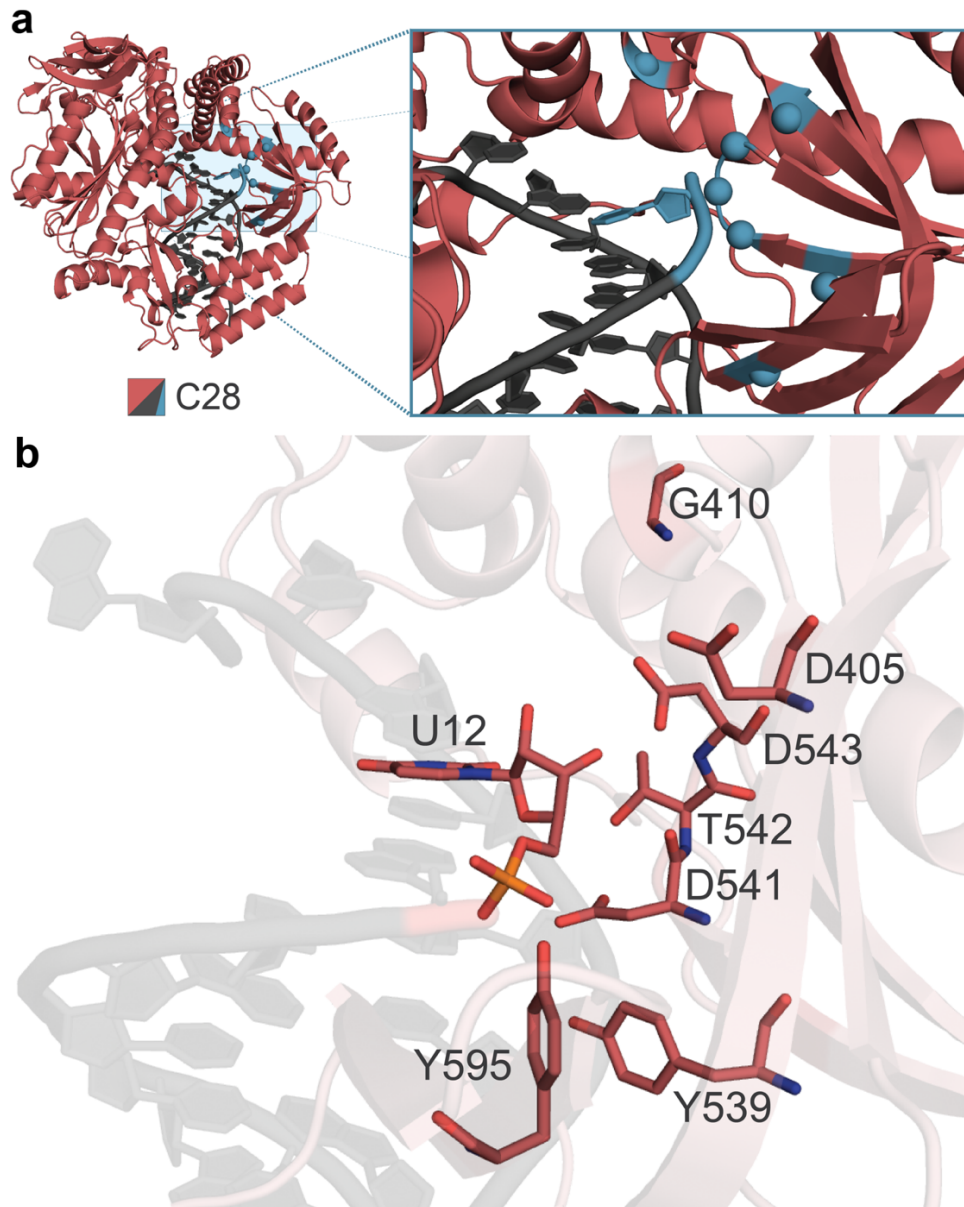
Supplementary Figure 14. Raw Fidelity Data for RNA Replication using only C28. Sequence alignment of TOPO cloned cDNA from a complete cycle of RNA replication using an engineered RNA polymerase for both RNA synthesis and reverse transcription. Raw data was aligned with the expected 4nt9g sequence (colored) displayed in the 5'→3' direction. Correct nucleotide insertions are denoted as (.) and deletions and/or insertions are denoted as (-). Incorrect nucleotide incorporations are shown as the observed base (A, C, G, or T) at specified locations. The calculated aggregate fidelity of C28 is 99.6% (4 misincorporations, 1 deletion) from a total of 1,200 nucleotide incorporations.



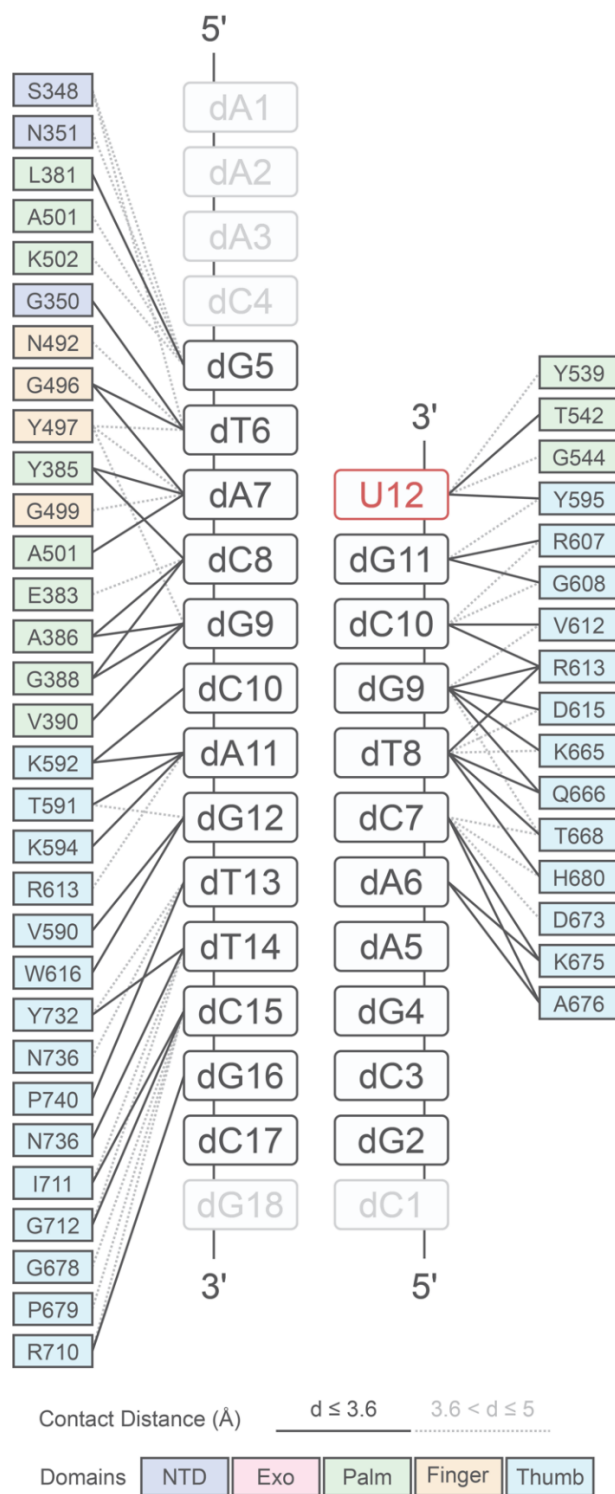
Supplementary Figure 16. U12 Base Pairing. The incorporated U12 exhibits canonical Watson-Crick base pairing with the templating deoxyadenosine dA7 with typical hydrogen bonding distances depicted as dashed lines. A top and side view of the boxed dA7:U12 pair are presented.



Supplementary Figure 17. U12 Sugar Pucker. (a) A stick model of U12 exhibiting a C2'-endo sugar pucker in a $2F_o - F_c$ polder maps contoured at 4.5σ with a carve radius of 2. (b) A pseudorotational wheel for nucleosides built in the resolved structure.



Supplementary Figure 18. Active Site of C28 Accommodates U12. (a) An overview of C28 bound to a duplex (red and dark gray). The aqua color highlights the U12 and the position of select proximal residues in the active site. (b) A stick model of the same residues is overlaid with a transparent cartoon of C28.



Supplementary Figure 19. C28 Duplex Contact Map. A contact map illustrating the interactions between C28 residues, colored by domain, and the primer-template duplex. Solid lines indicate interactions within a 3.6 Å cutoff. Dashed lines indicate interactions within a 3.6 Å to 5 Å.