SUPPLEMENTARY INFORMATION

Programmed Allelic Mutagenesis of a DNA Polymerase with Single Amino Acid Resolution

Ali Nikoomanzar[1], Derek Vallejo[1], Eric J. Yik[1], and John C. Chaput[1,2,3]*

[1]Departments of Pharmaceutical Sciences, University of California, Irvine, CA 92697-3958, [2]Department of Chemistry, University of California, Irvine, CA 92697-3958, and [3]Department of Molecular Biology and Biochemistry, University of California, CA 92697-3958.

Methods and Materials

Reagents

DNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, Iowa). TNA triphosphates were obtained by chemical synthesis as described previously[1,2]. ThermoPol buffer, Q5 site-directed mutagenesis kit, Gibson assembly were purchased from New England Biolabs (Ipswich, MA). DNA clean up kit was purchased from Zymo (Irvine, CA). Chemical reagents including dNTPs, and ammonium persulfate (APS) were purchased from Sigma Aldrich (St. Louis, Missouri). TOPO TA cloning kit, ethylenediaminetetraacetic acid (EDTA), urea, acrylamide, and bis-acrylamide were purchased from Thermofisher Scientific (Waltham, Massachusetts). Tetramethyl-ethylenediamine (TEMED) was purchased from Bio-Rad (Hercules, California). CleanAmp dNTP mix was purchased from TriLink Biotechnologies (San Diego, California). Heparin affinity columns were purchased from GE Healthcare (Little Chalfont, United Kingdom). The scanning mutagenesis library was purchased from Agilent Technologies. Polydimethylsiloxane (PDMS) base and curing agent were purchased from Dow Corning (Midland, MI). SU-8 2025 photoresist was purchased from Fisher Scientific (Hampton, NH). Fluorinated oil HFE-7500 was purchased from 3M Novec (St Paul, MN), and Pico-Surf™ 1 surfactant, Pico-Glide™ 1, and Pico-Break™ 1 were all purchased from Dolomite Microfluidics (UK).

Design and Construction of Scanning Mutagenesis Library (SML)

A custom scanning mutagenesis library (SML) was designed using the eArray software package (Agilent) to span the amino acid region (372-774) in Kod DNA polymerase (GenBank: AP006878.1). The library design was aligned to the Blast ID 48KZ_A (Kod polymerase in binary complex with dsDNA) and found to have a perfect alignment score (403/403 identities matched) with the query sequence. Using the QuikScan19 mutational setting, 31 replicates of 855 oligos were ordered for tiles 1-8 and Tile 9 31 replicates of 798 oligos for printing on a custom DNA microarray. The SML was inserted into a custom pGDR11 polymerase expression plasmid[3] harboring the Kod exo$^-$ polymerase gene using the Agilent QuikChange technology developed for the QuikChange HT Protein Engineering System. Briefly, custom amplification primers (Table S1) were used to PCR amplify the SML (Initial denaturation: 95°C-30 sec followed by 30 cycles: 95°C-30 sec, 56°C-45 sec, 72°C 1 min followed by polishing step of 72°C-2 min) in nine separate PCR reactions. Amplicons were purified using the supplied microspin cups and receptacle tubes. The purified amplicons were used as mega-primers for the linear amplification of the parent pGDR11-Kod exo$^-$ plasmid to incorporate the mutagenic cassettes for each individual sublibrary (Initial denaturation: 95°C-2 min followed by 18 cycles: 95°C-30 sec, 60°C-45 sec, 68°C 4 min followed by polishing step of 68°C-5 min). Subsequently, DpnI digestion (10 min at 37°C) with the supplied enzyme was performed to remove methylated and hemi-methylated plasmid DNA from the pGDR11-Kod exo- parent. This mixture (1 µL) was then transformed into SoloPack Gold Supercompetent cells and recovered for 1 h in 250 uL of NZY$^+$ media with shaking at 225 RPM, plated onto LB-ampicillin (100 µg/mL) agar plates and grown overnight at 37°C. 5 mL of LB-ampicillin (100 µg/mL) was then added to the entire plate and a single, sterile L-shaped spreader (Fisher Scientific) was used to gently scrape off the transformed colonies which were collected and pipetted to homogeneity in the provided liquid. This colony mixture was spun for 10 min at 4,000 RPM and 4°C with the supernatant discarded. The ensuing cell pellet was then purified using the Express Plasmid Miniprep Kit (Biomiga) following the manufacturer's recommended instructions. Plasmid DNA was stored for future transformations into XL1-blue *E. coli* for polymerase expression and DrOPS sorting.

Preparing *E. coli* for Encapsulation in Droplets

A detailed description of photolithography and microfluidic device fabrication for DrOPS sorting has been previously described[6]. Cells expressing the Kod DNA polymerase library were prepared by transforming 250 ng of the plasmid library into XL1-blue cells following the manufacturer's recommended instructions. Recovered cells were then used to inoculate 50 mL of LB-ampicillin (100

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. The culture (1:100 v/v) was used to inoculate 50 mL of LB-ampicillin (100 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 IPTG at a final concentration of 1 mM, 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 discarded. The cells were washed three times with commercial 1x ThermoPol buffer (20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1% TritonX-100, pH 8.8). The bacterial pellet was re-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[5]. Just prior to emulsification, the cells were mixed with the reagents for the polymerase activity assay (PAA) as described previously[5]. The PAA consists of 1 μM of a self-priming hairpin template labeled with Cy3 at the 5' end (ST.1G.HP.44.Cy3 Illumina experiments or 30merHP.V2 PacBio experiments, Table S1), 2 μM of a 3' end labeled Iowa Black quencher sequence (QP08.Iowa, Table S1), and 100 μM of TNA triphosphates (tNTPs) in 1x Thermopol buffer.

All aqueous and oil solutions were sealed in 1.5 mL plastic micro-centrifuge tubes (Sigma-Aldrich) and controlled via pressure driven flow with custom LabVIEW software (National Instruments)[6]. Two lengths of Tygon tubing (EW-06419-01, Cole-Parmer) 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 (Automation Distribution). Another length of tubing was submerged in the reagent solution with the other end connected to the appropriate inlet of the microfluidic device[7]. 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[6]. Single emulsions were formed utilizing a flow focusing geometry[8]. The aqueous phase containing the PAA and *E. coli* cells was sheared by a continuous phase consisting of a low-viscosity fluorinated oil (HFE-7500, 3M Novec) containing 1% (w/w) Pico-Surf surfactant (Dolomite Microfluidics, UK). Pressures were maintained to achieve droplet diameters of 20 μm and production rates of 30-35 kHz, allowing 110-125 x 10^6 droplets to be produced every h. 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 18 h.

FADS Sorting of Single Emulsion Droplets

Following incubation, droplets were injected into a FADS device capable of droplet sorting. Incident light from a 552 nm laser (Coherent OBIS LS, Santa Clara, CA) was focused through a 20x plan apochromatic objective (Motic, Hong Kong) where droplets pass in single file. Emitted light was led through a 405/488/543/635 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 (or AlexaFlour 660) 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, USB-7856R, National Instruments) controlled with custom LabView software. Droplets falling within a user-defined threshold triggered the FPGA to send a square-wave pulse (50 kHz, 50% duty cycle, 60 μs), amplified to 600 V by a high-voltage amplifier (2210, Trek, Lockport New York), 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. A complete summary of all droplet-sorting experiments can be found in SI Table 2.

Recovery of Sorted DNA for Next Generation Sequencing or Gibson Assembly (Colony Screening)

Plasmid DNA was recovered from the population of positively sorted droplets present in the collection tube by extraction with Pico-Break (Dolomite Microfluidics, UK) following the manufacturers recommended protocol. Briefly, DNA samples were recovered from sorted emulsions by extraction with 2 volumes of Pico-Break 1 (Dolomite), which contains 1H,1H,2H,2H-perfluorooctanol (PFO). After addition of Pico-Break 1, the samples were vortexed and centrifuged (15 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 molecular grade water to improve recovery yields. The combined aqueous layers containing the plasmid DNA were concentrated using a spin column (DNA Clean & Concentrator-5, Zymo Research) and eluted with molecular biology grade water (10 µl). Tile specific DNA primers (see Table S1) were used to PCR amplify the appropriate mutagenic region from the plasmid DNA (Initial denaturation: 95°C-30 sec followed by 30 cycles: 95°C-30 sec, 56°C-45 sec, 72°C 1 min followed by polishing step of 72°C-2 min.) The amplicon was then purified using a second spin column (DNA Clean & Concentrator-5, Zymo Research) and size-validated by 2% agarose gel prior to NGS analysis or plasmid reconstruction for colony picking with Gibson assembly.

Illumina Next Generation Sequencing Analysis

DNA from the sorted and unsorted libraries were sequenced at the UCI Genomics High-Throughput Facility. DNA collected from tiles 1-9 were sequenced with an Illumina HiSeq 4000 instrument. After trimming, we obtained a total of 38,934,889 reads for the starting libraries and 58,655,607 reads for the sorted populations that were mapped to amino acids 372-774 of the KOD polymerase gene. An individualized summary of each tile can be found in (SI 3.) The frequency of each mutation at each position was calculated by dividing the number of times the mutation was observed by the total number of mutations observations at that position (these calculations exclude wild-type residues)[9]. Several residues within tiles 1-9 contained less than 10 total observations and were excluded from the analysis (denoted as grey boxes in Figure 2. An enrichment score *E* was assigned to all remaining amino acid substitutions *a* according to the following equation:

$$E = \frac{f_{sort,a}}{f_{unsort,a}}$$

where $f_{sort,a}$ and $f_{unsort,a}$ are the frequencies of amino acid *a* in the sorted and unsorted libraries, respectively.

Enriched DrOPS Library Reconstruction, Colony Picking, Polymerase Activity Assay

Plasmids recovered from positively sorted droplets from libraries Kod-wt-Tile 6, Kod-wt-Tile 8, Kod-RS-Tile 6, and Kod-RS Tile 8 were amplified by PCR (SI Table 1, Tile 6-Fwd/Rvs, Tile 8-Fwd/Rvs), purified, and quantified on a Nanodrop instrument. The pGDR11 polymerase expression vector was reconstructed by Gibson assembly using the purified amplicons from tiles 6 or 8 and the corresponding linearized PCR vector product (Kod-wt or Kod-RS) (SI Table 1, Vector tile-6-Fwd/Rvs, Vector tile-8-Fwd/Rvs) in equimass quantities each (100 ng) following the manufacturer's recommended protocol. Gibson assembly mixes were transformed back into DH5Alpha cells and plated onto LB/ampicillin plates (100 µg /µl) at a dilution of 1:200. Single colonies were picked and grown overnight in 4 ml of LB media at 37°C with shaking (225 rpm). Confluent cultures were diluted 1:100 into fresh LB media (4 mL) and grown at 37°C with shaking (225 rpm) until the cultures reached an OD_{600} of 0.6-0.8 au. Cells were induced with 1 mM IPTG and grown overnight at 25°C with shaking at 225 rpm for an additional 18-24 h. Cells were harvest at 3202 rcf, 4°C, 10 min. Pelleted cells were suspended in 10 mM Tris pH 8.0, 500 mM NaCl, 10% glycerol and subjected to a heat treatment for

1 h at 80°C followed by cooling on ice for 1 h. The lysis mixture was centrifuged at 13,200 rpm for 30 min and the supernatant was used for the polymerase activity assay. Polymerase assays were performed in a 20 µL reaction volume containing a final concentration of 500 nM primer-template complex (Table S1, see 30mer template and short PBS8 or 50/50 template and long PBS8). The primer−template complex was annealed in 1x ThermoPol buffer by heating for 5 min at 95°C and cooling for 10 min at 4°C. Polymerase (2 µL of crude lysate or 0.5 µM purified enzyme) was added to the reaction mixture. The reactions were initiated with the addition of tNTPs (100 µM) and the solutions were incubated for 2 h at 55°C. A 2 µL aliquot was transferred to 38 µL of stop buffer [1x Tris-boric acid buffer, 25 mM EDTA, 50% formamide, pH 8.0] and an aliquot (10 µL) of each reaction was analyzed by denaturing PAGE and visualized using a LI-COR Oddyssey CLx imager.

Construction of Kod DNA Polymerase Variants

Polymerase mutants were constructed using the Q5 site-directed mutagenesis kit (New England Biolabs). Briefly, primers containing the desired mutations were designed using the NEBaseChanger software package (New England Biolabs) to ensure effective annealing during PCR amplification. Mutagenic oligos (Table S1) were first used to conduct whole-plasmid amplification of various permutations of the Kod exo- polymerase gene in the pGDR11 expression vector depending on the particular set of mutations desired (Initial denaturation: 95°C-2 min followed by 25 cycles: 95°C-30 sec, 60°C-45 sec, 72°C 8 min followed by polishing step of 72°C-5 min.) This was followed by a kinase-ligase-DpnI (KLD) treatment (20 min at room temperature) to phosphorylate and ligate the blunt-ended linear PCR product. DpnI was used to digest the parent template background. This mixture (1 µL) was then transformed into DH5-alpha supercompetent cells (New England Biolabs), recovered for 1 h in 250 uL of SOC media with shaking at 225 RPM, plated onto LB-ampicillin (100 µg/mL) agar plates and grown overnight at 37°C. Single colonies were picked and used to inoculate separate 4 mL aliquots of LB-ampicillin (100 µg/mL) liquid medium in 14 mL round-bottom culture tubes (Thermofisher Scientific) with shaking at 225 RPM and 37°C overnight. Overnight cultures were spun for 10 min at 4,000 RPM and 4°C with the supernatant discarded. Cell pellets were purified using the Express Plasmid Miniprep Kit (Biomiga) following the manufacturer's recommended instructions. DNA constructs were sequence validated (Genewiz, San Diego, CA) and analyzed using the CLC Main Workbench (Qiagen) software package. A detailed description of polymerase expression and purification has been previously described[10] and was the basis for all enzymatic characterization experiments performed in this study. Polymerase activity assays from purified enzymes were performed as described in the previous section with the exception of a short 2 min incubation at 95°C prior to extension at 55°C for dNTP (100 µM) activity assays utilizing CleanAmp dNTP to prevent premature extension. For time-course assays, a 2 µL aliquot from each reaction was transferred to 38 µL of stop buffer [1x Tris-boric acid buffer, 25 mM EDTA, 50% formamide, pH 8.0] to quench the reaction at a particular incubation time point.

Construction of the Kod-A485R-N491/R606/T723 Site-Saturation Mutagenesis Library

The Kod-A485R-N491/R606/T723 site-saturation library was designed using CLC Main Workbench. Primers (SI Table 1, see) were designed to produce mutagenic cassettes (SI Fig. 6) of varying size 491 (365 bp), 606 (371 bp), 723 (205 bp) and a longer vector PCR product (6,220 bp) for sealing together the 3 mutagenic cassettes from a pGDR11 Kod-A485R exo- polymerase backbone. NNK degeneracy was used to ensure that all 20 amino acid combinations were present with a limited number of stop codons. PCR conditions for the mutagenic cassettes are as follows (Initial denaturation: 95°C-30 sec followed by 30 cycles: 95°C-30 sec, 60°C-45 sec, 72°C 1 min followed by polishing step of 72°C-2 min) in three separate PCR reactions. PCR conditions for the linearized vector are as follows (Initial denaturation: 95°C-30 sec followed by 30 cycles: 95°C-30 sec, 60°C-45 sec, 72°C 6 min followed by polishing step of 72°C-4 min). The amplicons were purified by silica column and quantified on a NanoDrop spectrophotometer. 100 ng of each cassette and vector (400 ng total) were added to a Gibson assembly Mastermix, incubated at 50°C for 1 hour and transformed

into DH5-alpha super competent cells following the manufacturer's recommended instructions and plated on LB-ampicillin (100 µg/mL) agar plates and grown overnight at 37°C . Recovery and purification of the library was identical to the procedure in the "Design and Construction of Scanning Mutagenesis Library (SML)" section.

DrOPS-based Screening of the Site-Saturation Mutagenesis Library

The Kod-A485R-N491/R606/T723 site-saturation mutagenesis library (250 ng) was transformed into XL1 Blue *E. coli* and grown in 50 ml of LB media with 100 µg/µl ampicillin for 16 h at 37°C. An aliquot (500 µL) of the confluent overnight culture was used to inoculate a new 50 mL culture of LB media containing 100 µg/µl ampicillin. Cells were grown at 37°C and 225 rpm until an OD_{600} of ~ 0.6 au. Cells were induced by adding IPTG to a final concentration of 1mM. Cells were allowed to express recombinant polymerase for 18 h at 25°C. Encapsulation of *E. coli* was performed as described in the "Preparing *E. coli* for Encapsulation in Droplets" section. The PAA was incubated at 55°C for 18 h with the 30merHP.V2 template, which represented a higher stringency from the Illumina sorting experiments. Droplets were sorted as described in the "FADS Sorting of Single Emulsion Droplets" section. We detected 13,840,744 droplets and sorted 5,874 droplets using a 60 RFU gate. The DNA from the sorted droplets was recovered as described in the "Recovery of Sorted DNA for Next Generation Sequencing or Gibson assembly (Colony Screening)" section. PCR amplification was performed (Tile 1-Fwd/Tile 9-Rvs primers, SI Table 1) to amplify the polymerase domain (1260 bp) for PacBio library construction with the SMRTbell template prep kit (Pacific Biosciences, Menlo Park, CA). The amplicons were purified using ampure PB beads followed by ligation of the hairpin loop adapters. Failed ligation products were removed by exonuclease treatment and the SMRTbell templates were purified three times using ampure PB beads. Sequencing was performed on the PacBio Sequel 1M v3 SMRTcell LR using diffusion loading and Sequencing kit 3.0. The loading concentration was 8 pM with an immobilization time of 120 min, pre-extension of 240 min and movie time of 20 h. DNA internal control complex V3 was included in runs for monitoring sequencing quality. Experiments were performed at the UCI Genomics High Throughput Facility (Irvine, CA).

The Kod-A485R-N491/R606/T723 site-saturation library was sequenced on a PacBio Sequel instrument. The raw reads prior to processing for the naïve library was 28,918,205 reads. After trimming and processing, we obtained 60,466 reads from 6,853 possible combinations of mutations at positions 491/606/723 (85.7% of the 8,000 total combination). A filter was applied by setting the minimum number of reads to 10, which reduced the dataset to 41,479 reads and 1,944 combinations (24.3% of the 8,000 total combination). The raw reads prior to processing for the sorted population was 28,699,846 reads. After trimming and processing, we obtained 32,176 reads and 5,418 unique combinations (67.7% of the 8,000 total combination). When filtered with 10 minimum reads, the data was reduced to 17,610 reads with 579 unique combinations (7.2% of the 8,000-total combination). Enrichments scores were calculated with the filtered data set of 10 reads or more as follows: 1) each individual combination read count was divided by the total read count from that set to determine the frequency, 2) frequency of the selected output was divided by the frequency observed in the naïve pool for the specified combination. Enrichment score was transformed by natural log (ln). Upon further analysis only 546 combinations overlapped between the naïve and enriched pool.

Fidelity of TNA Transcription and Reverse Transcription

The fidelity assay was performed as previously described[3]. Briefly, 1 µM primer Extra.PBS8 containing a two-nucleotide mismatch was annealed with 1 µM of the 4NT9G template. For Kod-RSGA, a second fidelity assay was performed using 10 unique 83mer sequences having a G:C content of ~50% in the region of interest. In all cases, the primer-extension reaction was performed for 2 h at 55°C in 1x Thermopol buffer and 100 µM of each tNTP. The Kod-RS and Kod-RSGA polymerases were used at equimolar concentration to the primer-template duplex. The extension

product was purified by denaturing urea PAGE and reverse transcribed into cDNA. The reverse transcription reaction was performed by incubating the TNA template with PBS7 primer (1 μM) with Bst DNA polymerase (3 μM) for 3 h at 50°C in a reaction mixture (20 μl) that contained 500 μM dNTPs, 3 mM $MgCl_2$ and 1x Thermopol buffer. The cDNA product was PCR amplified with Taq polymerase (PCR cycle conditions should be listed here) and the PBS9/PBS7 primer pair, ligated into a TOPO vector and cloned into DH5-alpha competent *E. coli*. Individual colonies were grown in liquid media and sequenced (Genewiz, San Diego, CA). DNA sequences were aligned with the 4NT9G template and analyzed using Molecular Evolutionary Genetics Analysis version 7 and CLC Main (Qiagen, Venlo, Netherlands).

Supplementary Table 1. Oligonucleotides used in this study. Sequences are given in IDT nomenclature. “N” represents a degenerate position with variable proportions of A, T, G, and C. “K” represents a degenerate position with variable proportions of G and T. Underlined and lowercase base are TNA fragments.

Name	DNA Sequence (5’->3’)
Tile 1-Fwd	GAATGAGCTGGCCCCGAACAAG
Tile 1-Rvs	CTTCTCTGTTGAGCGTATCCGGCG
Tile 2-Fwd	CCCCTCAATCATCATCACCCACAAC
Tile 2-Rvs	GTCAATCGTGGCCTTCATCTTCTTC
Tile 3-Fwd	CCTGCTTGGAGACCTCCTAGAGGAGAG
Tile 3-Rvs	CCGTTACGCTCTCTGCACACTC
Tile 4-Fwd	GTTACTACGGCTATGCAAGGGCG
Tile 4-Rvs	CATAGCCTTCTTTTTGACGGTTTCAGC
Tile 5-Fwd	CGGATTTTTTGCCACAATACCTGGAG
Tile 5-Rvs	CGTTGTTATCTTGCCTTCCTCGTC
Tile 6-Fwd	CGTCACGAAGAAGAAGTATGCGG
Tile 6-Rvs	GCTCAGCTTTTCGGTAACTTCTTTGAC
Vector Tile 6-Fwd	GTCAAAGAAGTTACCGAAAAGCTGAGC
Vector Tile 6-Rvs	CCGCATACTTCTTCTTCGTGACG
Tile 7-Fwd	CTAAAGGACGGTGACGTCGAGAAG
Tile 7-Rvs	GGCGTATTTTGACTCCTCTCGC
Tile 8-Fwd	GTTGCCGTTGCCAAGAGGTTG
Tile 8-Rvs	GAATTCTCTCAACGGCTGGGAGAACCTG
Vector Tile 8-Fwd	CAGGTTCTCCCAGCCGTTGAGAGAATTC
Vector Tile 8-Rvs	CAACCTCTTGGCAACGGCAAC
Tile 9-Fwd	GCACAAGTACGACGCCGAGTAC
Tile 9-Rvs	CTAATTAAGCTTGGCTGCAGGTCGACTC
ST.1G.HP.44.Cy3 Template	/5Cy3/ACAACCATACTCTTCCACATCTGCATACGCGAAAGCGTATGCAG
30mer.HP.V2.Cy3 Template	/5Cy3/ACAACCATTATCTAGAGCGATTCGTATAGGTGGTATCCGAAAGGATACCACC
QP08.Iowa	ATGGTTGT/3IABkFQ/
Site-saturation-N491-KOD-Rxxx-Fwd	GGCGTATCAAGATCCTGGCANNKAGCTACTACGGTTACTACGG
Site-saturation-N491-KOD-Rxxx-Rvs	CGTTGTTATCTTGCCTTCCTCGTCTATCACCGCATACTTCTTCTTCG
Site-saturation-R606-KOD-Rxxx-Fwd	AGGAAGGCAAGATAACAACGNNKGGACTTGAGATTGTGAGGCG
Site-saturation-R606-KOD-Rxxx-Rvs	CGGGTCGAACTCGTCGAACGGTATCGCCCTG
Site-saturation-T723-KOD-Rxxx-Fwd	CGTTCGACGAGTTCGACCCGNNKAAGCACAAGTA

Site-saturation-T723-KOD-Rxxx-Rvs	CTAATTAAGCTTGGCTGCAGGTCGACTCAAGTTCCCTTCGGCT
Site-saturation-KOD-Rxxx-Vector-Fwd	CTGCAGCCAAGCTTAATTAGCTGAGCTT
Site-saturation-KOD-Rxxx-Vector-Rvs	TGCCAGGATCTTGATACGCCTCTGC
R606G-KOD-Fwd	GGCAAGATAACAACGGGTGGACTTGAGATT
R606-KOD-Rvs	TTCCTCGTCTATCACCGCATACTTCTTCTTCG
T723A-KOD-Fwd	GACCCGGCGAAGCACAAGTAC
T723-KOD-Rvs	GAACTCGTCGAACGGTATCGCC
R485A-KOD-RSGA-Fwd	CGATTACAGGCAGAGGGCTATCAAGATCC
R485-KOD-RSGA-Rvs	AGGAGCTTCCTCTCGATCGGGTCAAT
S491N-KOD-RSGA-Fwd	CAAGATCCTGGCAAACAGCTACTACG
S491A-KOD-RSGA-Fwd	CAAGATCCTGGCAGCCAGCTACTACG
S491P-KOD-RSGA-Fwd	CAAGATCCTGGCACCAAGCTACTACG
N491-KOD-RSGA-Rvs	ATACGCCTCTGCCTGTAATCGAGG
R606P-KOD-RS-Fwd	GATAACAACGCCTGGACTTGAGATTGTG
R606-KOD-RS-Rvs	TTGCCTTCCTCGTCTATCACCG
30mer HP	TCTCTATAGTGAGTCGTATAGGTGGTATCCGAAAGGATACCACC
Max Standard	TCTCTATAGTGAGTCGTATAGGTGGTATCCGAAAGGATACCACC tatacgactcactatagaga
30mer Template	TCTCTATAGTGAGTCGTATAGGTGGTATCC
Short PBS8	/ 5IRD680 / GGATACCACC
50/50 Template	ATTGAACAACGCAAGGATCGGCGATATAAAGGTGGTATCCCCAAGGGGAC
Long PBS8	/ 5IRD680 / GTCCCCTTGGGGATACCACC
4NT9G template	GGATCGTCAGTGCATTGAGATTAAGACTCGCCATGTTACGATCT GCCAAGTACAGCCTTGAATCGTCACTGGTGGTATCCCCAAGGGGA/3 ddC/
Extra.PBS8	/5' 6-FAM/ CTTTTAAGAACCGGACGAACGTCCCCAAGGGGATACCACC
PBS7	GGATCGTCAGTGCATTGAGA
PSB9	CTTTTAAGAACCGGACGAAC

Seq1	GGATCGTCAGTGCATTGAGAAACACAACCACACCCACAAGGTCGGGTTCGTTCGGTATTTGGTGGTGGTATCCCCTTGGGGAC
Seq2	GGATCGTCAGTGCATTGAGACTCAACCCACACAAACAAACCTCAAACCGGTCCCGCTGTCGCGGGTGGTATCCCCTTGGGGAC
Seq3	GGATCGTCAGTGCATTGAGATTCCCAAACCCACCCACCACTGCGACTCTGACCGTCTTATCGTGGTGGTATCCCCTTGGGGAC
Seq4	GGATCGTCAGTGCATTGAGACAGACAAACCCACCCACAAACGACCAAATTAAATCCTGGCCGCGGTGGTATCCCCTTGGGGAC
Seq5	GGATCGTCAGTGCATTGAGACGTAACCAAACAACCCACCCAGAACCCAGTCATGACTGAGGTTGGTGGTATCCCCTTGGGGAC
Seq6	GGATCGTCAGTGCATTGAGATGGCCCAAACACACAACAAATTGGCCAAAGCAAACATGCGATAGGTGGTATCCCCTTGGGGAC
Seq7	GGATCGTCAGTGCATTGAGAGCAGAACAAACACACAACAAGGTCCGGACGAACACCTTATTTGGGTGGTATCCCCTTGGGGAC
Seq8	GGATCGTCAGTGCATTGAGAAGCTACCCAAACACAACACCTGTCGTCATGCCTAGCGGTCATAGGTGGTATCCCCTTGGGGAC
Seq9	GGATCGTCAGTGCATTGAGAGTATCCAAACAAACCAAACCCTCACCCAATTAAAGCATGCCTTGGTGGTATCCCCTTGGGGAC
Seq10	GGATCGTCAGTGCATTGAGACTTTACAAACCCACCACCAAGGACGCAATGATGGTCTGCAGGGGGTGGTATCCCCTTGGGGAC

Supplementary Table 2. Drops observed and drops detected from different experimental libraries.

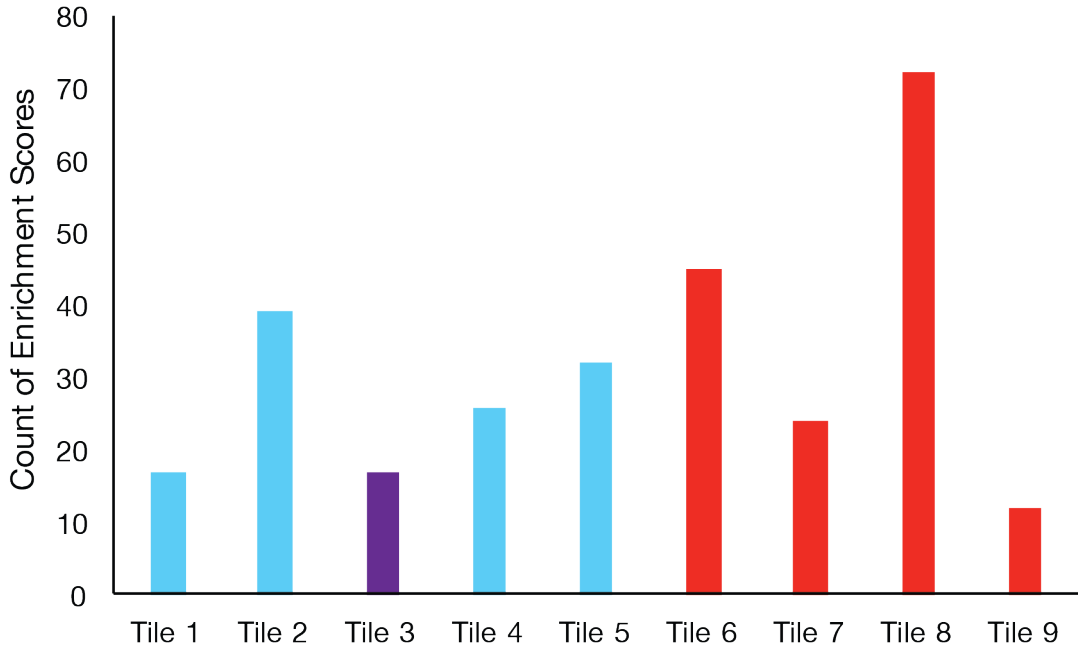
Library	Drops Screened	Drops Sorted
Tile 1	10,338,751	4,252
Tile 2	4,558,463	847
Tile 3	7,359,428	1,172
Tile 4	7,368,352	1,619
Tile 5	9,403,975	3,820
Tile 6	6,064,739	6,435
Tile 7	4,005,398	604
Tile 8	8,338,072	5,259
Tile 9	3,922,286	585
Kod-RS Tile 6	12,430,111	31,583
Kod-RS Tile 8	17,568,008	80,703
Site-Saturation	13,840,744	5,874

Supplementary Table 3. Number of reads observed for the naïve and sorted pools for tiles 1-9.

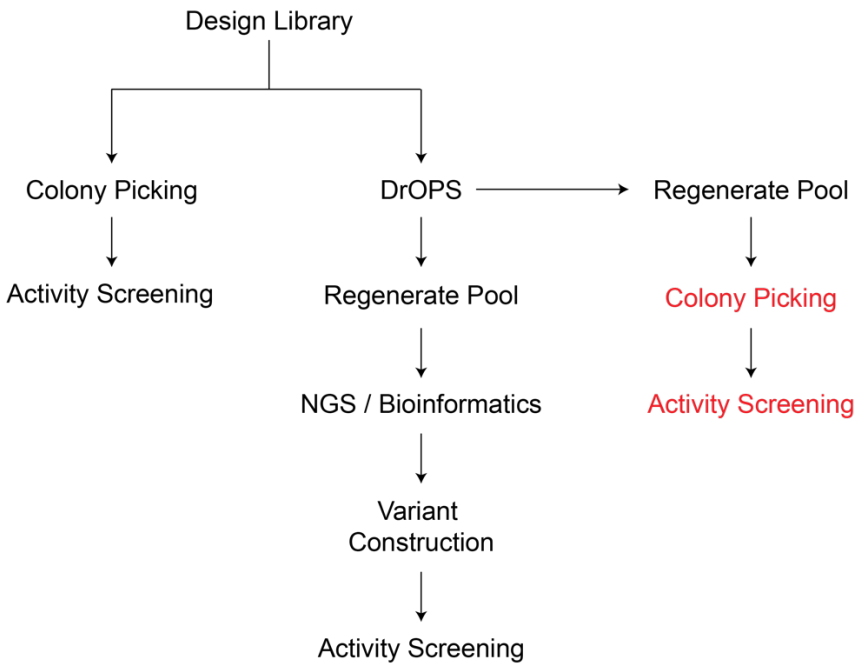
Library	Reads from naïve population	Reads from sorted population
Tile 1	4,988,446	8,990,573
Tile 2	4,026,836	7,941,868
Tile 3	5,703,487	6,757,957
Tile 4	2,432,895	5,212,864
Tile 5	3,501,208	4,944,663
Tile 6	4,664,216	6,945,614
Tile 7	4,566,908	5,793,554
Tile 8	4,161,780	6,934,208
Tile 9	4,889,113	5,134,306

Supplementary Table 4. Assessment of different DNA templates for primer extension with varying composition and GC content. Extension difficulty was evaluated by denaturing PAGE for reactions performed at 30 minutes in 1x ThermoPol, 500 nM primer/template duplex, 100 nM tNTPs and 500 nM of purified Kod-RSGA. Primer binding site is shown in red.

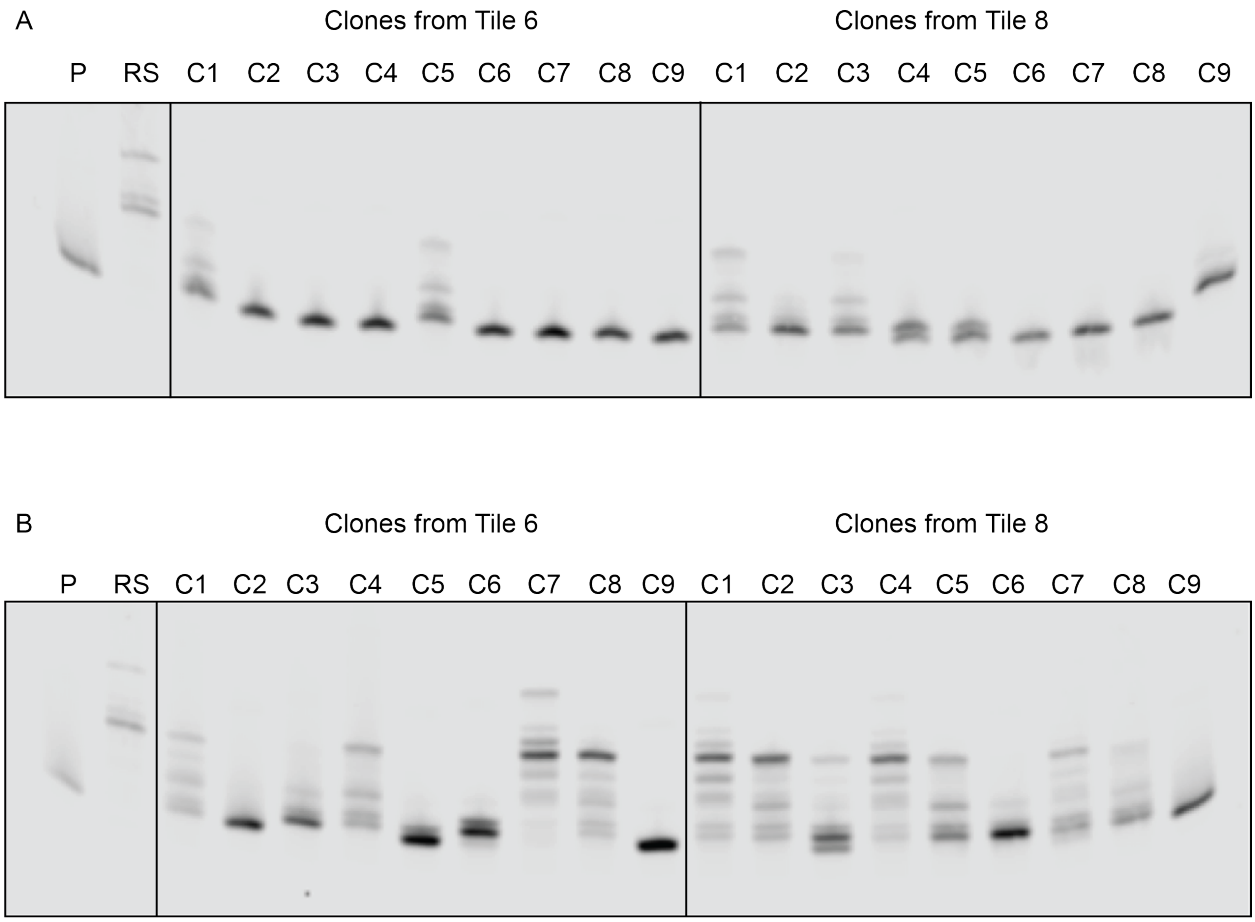
Sequence Name	Nucleotide Sequence (5' → 3')	Difficulty Ranking
50/50 Template 1 (AT:GC)	ATTGAACAACGCAAGGATCGGCGATATAAAGGTGGTATCCCCAAGGGGAC	1
30mer	TCTCTATAGTGAGTCGTATAGGTGGTATCC	2
50/50 Template 2 (AT:GC)	CCGCAAGTTCTAGTGCAATGGCGGTATAGTGGTGGTATCCCCAAGGGGAC	3
60/40 Template 1 (AT:GC)	ATATTTCAGGCGGTACGTCTGCTATCGAGTGGTGGTATCCCCAAGGGGAC	4
60:40 Template 2 (AT:GC)	GAGAAATTACATTCAAGTCTGATTAACCATGGTGGTATCCCCAAGGGGAC	5



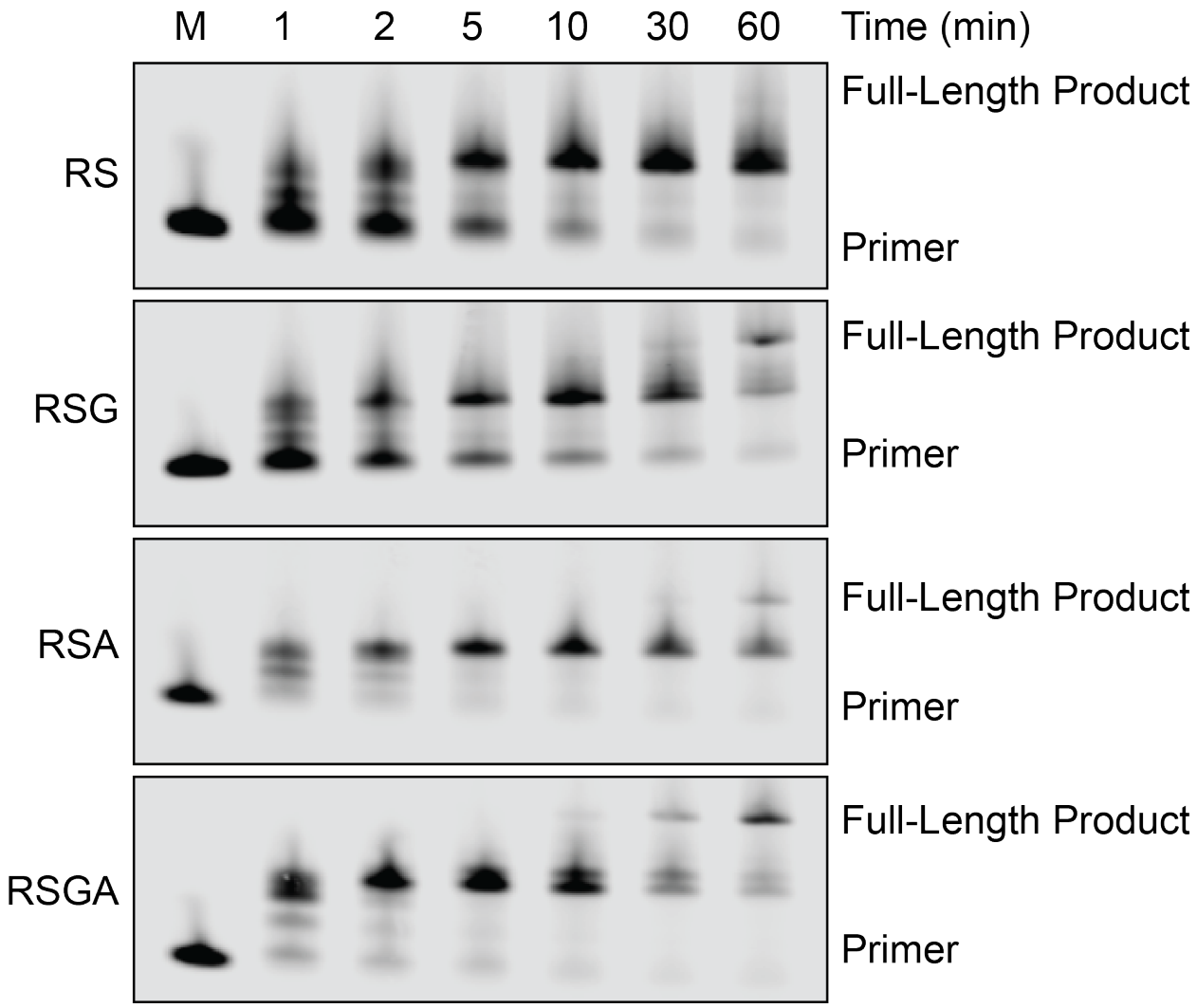
Supplementary Figure 1. Summary of the enriched mutations identified by deep mutational scanning. A summation of the number of enriched mutations (scores > 6) per tile. Tiles 1, 2, 4, 5 correspond to the palm (teal), tile 3 is the finger (purple), and tiles 6-8 correspond to the thumb (red).



Supplementary Figure 2. General strategies for polymerase evolution. Conventional approaches to polymerase engineering involve screening individual library members for activity using enzyme obtained from crude lysate (left). DrOPS provides an ultrahigh throughput assay for screening programmed allelic libraries with high redundancy. Positively sorted variants can be evaluated by NGS analysis, reconstructed, and characterized (center) or the pool could be cloned back into *E. coli* and assayed for activity in a secondary screen (right). Screening the output of a DrOPS selection increases the probability of identifying high activity variants and avoids the need to resynthesize variants before testing. Only variants with high activity in the secondary screen are sequenced, expressed, purified, and validated.



Supplementary Figure 3. Polymerase activity assay comparing randomly sampled clones before and after a high-throughput screening using DrOPS. Eighteen randomly chosen members were sampled from tiles 6 and 8 by picking *E. coli* colonies from the A) naïve and B) enriched pools. The polymerase activity assay (1x ThermoPol, 500 nM primer/template duplex, and 100 nM tNTPs) was performed in a 20 µL reaction volume using 10% (v/v) clarified lysate after thermal denaturation to remove endogenous *E. coli* proteins and incubated at 55°C for 30 min. Reaction was quenched with 1xTBE, 50% Formamide and 25 mM EDTA. Symbols: P, primer; RS, Kod-RS TNA polymerase, C, colony number.



Supplementary Figure 4. Primer-extension assay evaluating Kod TNA polymerase variants for TNA synthesis activity on a DNA primer-template duplex. Mutations R606G and T723A were identified from a secondary screen of tiles 6 and 8, respectively. Primer extension was performed with the following conditions: 1x ThermoPol, 500 nM primer/template duplex, 100 nM tNTPs, 500 nM purified enzyme at 55°C and reaction was quenched (1xTBE, 50% Formamide and 25 mM EDTA) at designated time point.

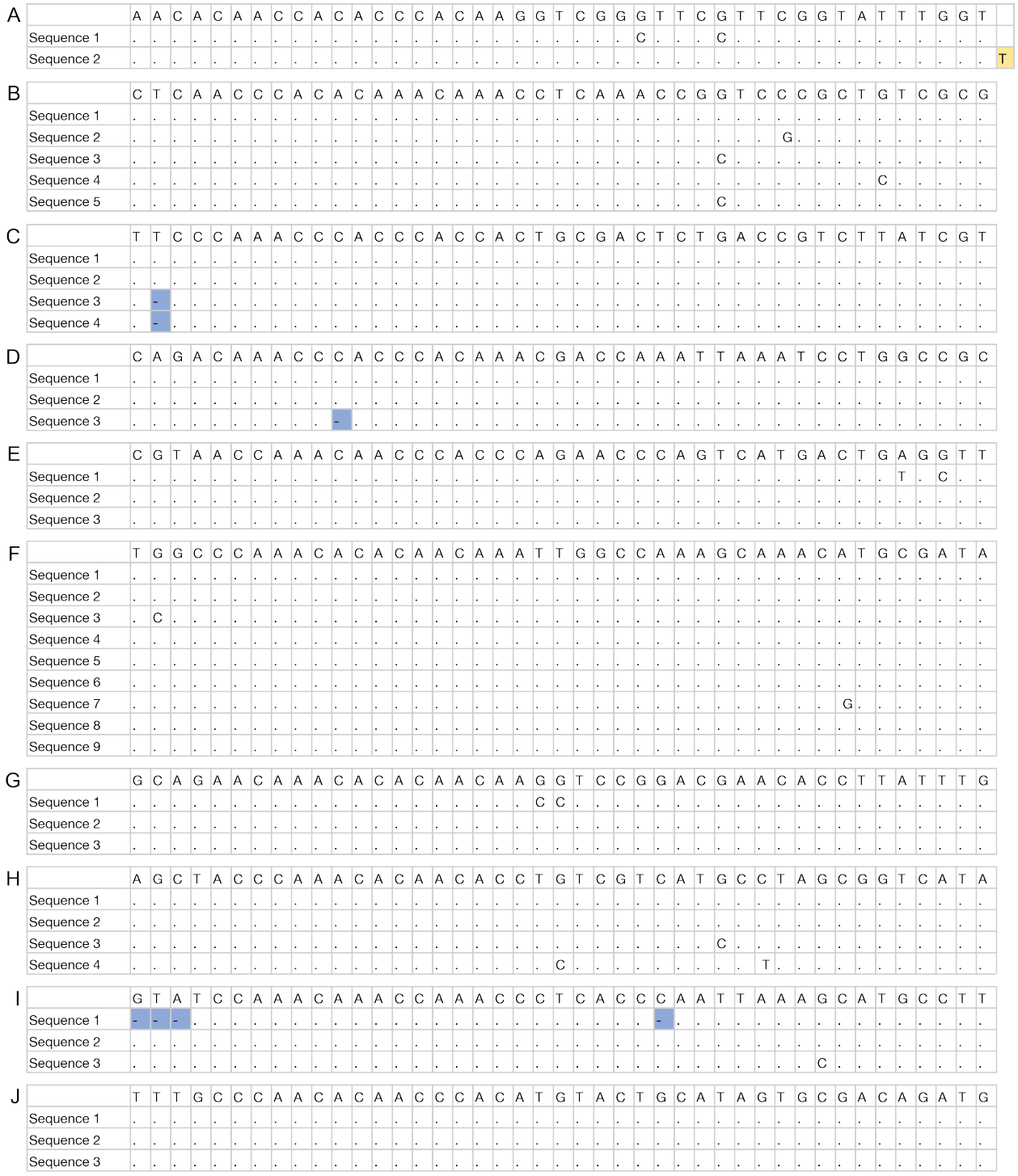
A

	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
Sequence 1	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Sequence 5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Sequence 8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 10	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 11	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 12	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 13	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 14	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 15	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 16	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Sequence 17	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 18	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 19	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 20	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 21	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

B

	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
Sequence 1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
Sequence 2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.
Sequence 4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 11	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 12	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 13	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 14	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 15	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.
Sequence 16	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 17	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 18	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 19	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 20	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sequence 21	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Supplementary Figure 5. Sanger sequencing results obtained for TNA replication experiments performed on the 4NT9G template for Kod-RS (A) and Kod-RSGA (B). Correct bases are denoted as (.) and incorrect bases are denoted by observed base in the specified position. The result signifies that Kod-RSGA is less prone to G→C transversion than Kod-RS.



Supplementary Figure 6. Sanger sequencing results obtained for TNA replication using 10 different templates and Kod-RSGA to mediate TNA synthesis. A-J. The results for templates 1-10 (Seq1-10 from SI Table 1) Correct bases are denoted as (.) and incorrect bases are denoted by observed base in the specified position. Blue boxes represent deletions. Yellow boxes represent insertions.

A

RS

5' End \ 3' End	A	C	T
A	100.00%	80.95%	95.24%
C	100.00%	76.19%	85.71%
T	100.00%	80.95%	85.71%

RSGA

5' End \ 3' End	A	C	T
A	100.00%	100.00%	85.71%
C	95.24%	100.00%	76.19%
T	100.00%	100.00%	100.00%

B

RSGA

5' End \ 3' End	A	C	T	G
A	100.00%	95.00%	100.00%	92.90%
C	100.00%	100.00%	94.10%	100.00%
T	100.00%	95.70%	77.80%	96.20%
G		100.00%	80.80%	100.00%

Supplementary Figure 7. Nearest neighbor analysis. Templating nucleotides on the 5' and 3' side of G residues were analyzed for their contribution to the formation of G:G mismatches during TNA synthesis. (A) Direct comparison of Kod RS and Kod RSGA on the single template sequence used for the fidelity analysis. (B) Analysis of adjacent residues in the 10 unique templates used for fidelity analysis. The results suggest that pyrimidine residues (C's and T's) on the 3' side of templating G residues lead to a higher frequency of G:G mismatches.

References

(1) 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.; Herdewijn, P.; Holliger, P. *Science* 2012, *336*, 341-344.

(2) Yu, H. Y.; Zhang, S.; Chaput, J. C. *Nat Chem* 2012, *4*, 183-187.

(3) Dunn, M. R.; Otto, C.; Fenton, K. E.; Chaput, J. C. *ACS Chem. Biol.* 2016, *11*, 1210-9.

(4) Xia, Y. N.; Whitesides, G. M. *Angew Chem Int Edit* 1998, *37*, 551-575.

(5) Larsen, A. C.; Dunn, M. R.; Hatch, A.; Sau, S. P.; Youngbull, C.; Chaput, J. C. *Nat Commun* 2016, *7*.

(6) Vallejo, D.; Nikoomanzar, A.; Peagel, B. M.; Chaput, J. C. ACS Synth. Biol. 2019, 8, 6, 1430-1440

(7) Vallejo, D.; Lee, S. H.; Lee, A. *Methods Mol Biol* 2017, *1572*, 489-510.

(8) Tan, Y. C.; Cristini, V.; Lee, A. P. *Sensor Actuat B-Chem* 2006, *114*, 350-356.

(9) Romero, P. A.; Tran, T. M.; Abate, A. R. *P Natl Acad Sci USA* 2015, *112*, 7159-7164.

(10) Nikoomanzar, A.; Dunn, M. R.; Chaput, J. C. *Curr Protoc Nucleic Acid Chem* 2017, *69*, 4 75 1-4 75 20.