Enzymatic DNA Synthesis by Engineering Terminal Deoxynucleotidyl Transferase

Xiaoyun Lu[1,2,3†], Jinlong Li[1,2†], Congyu Li[1,4], Qianqian Lou[1,2], Kai Peng[1,2], Bijun Cai[1,2], Ying Liu[1,2], Yonghong Yao[1,2], Lina Lu[1,2], Zhenyang Tian[1,2], Hongwu Ma[1,2], Wen Wang[3], Jian Cheng[1,2*], Xiaoxian Guo[1,2*], Huifeng Jiang[1,2*], Yanhe Ma[1,2]

[1]*National Center of Technology Innovation for Synthetic Biology, Tianjin 300308, China*

[2]*Key Laboratory of Systems Microbial Biotechnology, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, 300308, China*.

[3]*School of Ecology and Environment, Northwestern Polytechnical University, Xi'an, Shanxi 710072, China*

[4]*Tianjin University of Science&Technology, Tianjin, 300457, China*.

[†]These authors contributed equally to this work.

[*]Corresponding authors. E-mail: jiang_hf@tib.cas.cn, guo_xx@tib.cas.cn and cheng_j@tib.cas.cn.

Supplementary Note

Virtual screening and phylogenetic analysis of TdT genes

To obtain an efficient TdT gene for DNA biosynthesis, we screened all potential TdTs in the NCBI database. Firstly, we predicted all potential DNA polymerases by search against the metazoa non-redundant database with the Pfam domain ID PF14791 (hmmscan --cpu 10 --domtblout output.txt -E 1e-4 PF14791.hmm NR_Metazoa.fasta); Secondly, we retrieved all DNA polymerases from KEGG database (https://www.genome.jp/entry/pf:DNA_pol_B_thumb); Thirdly, we made a local blastp search using DNA polymerases from NCBI as the query sequences, and DNA polymerases from KEGG as the BLAST database. After blastp search, 137 potential TdTs (Supplementary Fig. 1) were screened with three standards: the best hit is a DNA nucleotidylexotransferase (DNTT, EC:2.7.7.31, KO: K00977), the identity is more than 40, and the align length is more than 400. Fourthly, 137 TdTs were classified into 13 groups by using OrthoMCL with the identity more than 90 in a group. For each group, we selected a TdT gene which is closest approximation to supposed optimal sequence, consisting of the highest frequency residues in multiple sequences alignment; Finally, we totally synthesized 14 TdT genes (an extra mouse TdT gene was used as reference) for next functional evaluation.

TdT expression and purification

The protein sequences of TdTs from different species were shown in Supplementary Table 5. The codon optimized TdT genes were synthesized by Wuhan GeneCreate Biological Engineering Co., Ltd, and cloned into pET-28a vector between *Nde* I and *Xho* I. All of the TdT variant proteins were expressed in the BL21 (DE3) *E. coli* strain with the addition of IPTG (isopropyl-β-D-thiogalactopyranoside) to a final concentration of 0.5 mM at 16 °C for overnight after the OD_{600} of the cells reached 0.6. The cells were collected by centrifugation at 6,000 g for 10 min and then resuspended in 50 mL lysis buffer (50 mM potassium phosphate buffer, pH 6.8, 100 mM NaCl). The resultant cell pellets were lysed by using a high-pressure homogenizer (JNBIO, China), and removed the cell debris by centrifugation at 10,000 g for 60 min at 4 °C. The

clarified lysate was loaded onto a nickel affinity column (GE Healthcare) under gravity flow. Non-target proteins were removed with 50 mL washing buffer (50 mM potassium phosphate buffer, pH 6.8, 100 mM NaCl and 50 mM imidazole), then the target protein was collected with 20 mL elution buffer (50 mM potassium phosphate buffer, pH 6.8, 100 mM NaCl and 200 mM imidazole). The eluted protein was concentrated and dialyzed against lysis buffer by ultrafiltration with an Amicon Ultra centrifugal filter device (Millipore, USA) and a 30 kDa molecular-weight cutoff. The protein concentration was determined using a BCA Protein Assay Reagent Kit (Pierce, USA) with BSA as the standard.

Incorporation of natural nucleotides by TdT

Catalytic activity of TdTs was determined in 30 μL reaction mixtures. The reaction system comprised 1 μM Oligonucleotides (P1, Supplementary Table 7), 0.1 mM dNTPs, 0.25 mM $CoCl_2$, 100 mM NaCl and 50 mM phosphate buffer (pH 6.8). The catalytic reaction was started by adding 0.5 mg/mL purified protein. After incubation at 30 °C for 5 s, the reaction was stopped by heating at 95 °C. Products were analyzed by Capillary electrophoresis (CE). Samples were assessed by using Qsep-400™ Automatic Nucleic Acid and Protein Analysis System (Bioptic, Taiwan, China). A disposable pen-shaped cartridge was inserted and other run buffers were loaded according to manufacturer's instructions. The run was performed using a high-resolution cartridge with the sample injection protocol at 8 KV for 10 s and separation at 6 KV for 300 s.

Incorporation of modified nucleotides by TdT

The reaction system comprised 1 μM oligonucleotides (P1, Supplementary Table 7), 0.25 mM 3'-ONH_2-dNTPs (Firebird Biomolecular Sciences, LLC, US), 0.25 mM $CoCl_2$, 100 mM NaCl and 50 mM phosphate buffer (pH 6.8). The catalytic reaction was started by adding 0.5 mg/mL purified protein. After incubation at 30 °C for 20 min, the reaction was stopped by heating at 95°C. The products of the reaction were analyzed by gel electrophoresis on a 20% acrylamide gel and 8 M urea. Subsequently, the PAGE gel was stained with SYBR™ Gold nucleicacid gel dye for 20 minutes and scanned

with a fully automatic fluorescence image analysis system (Tanon 5200, China). The result was processed with Gel Image System analysis software (Tanon, China), and the proportion of oligonucleotides with one base added (n+1) to the total oligonucleotides (n+n+1) was calculated.

Molecular-dynamics simulations

Atomistic MD simulations were performed using AMBER18 molecular dynamics package. The bonded and non-bonded description of the interactions between the various atoms have been generated using the AMBER18 force fields, which include the ff14SB force field parameters. The ANTECHAMBER module and GAFF2 with AM1-BCC charges were used to obtain force field parameters for ligands. Initially, we performed a series of energy minimization steps to eliminate any bad contacts in the initially built structures. During the minimization, protein (@CA,O,N,C) were restrained with harmonic force constants of 20 kcal/mol. The minimization process involved 5,000 steps of steepest descent followed by 5,000 steps of conjugate gradient method. After the energy minimization, the system was slowly heated up to 310 K in 100 ps MD using 1 fs integration time step, while restraining the solute with 20 kcal/mol harmonic force constant. After this, we performed 15 ps NPT equilibration of the structures with no harmonic restraints. And then, 5 ns constrained MD simulations were executed, so that the ligand was in a reasonable position to react. Finally, 50 ns NPT production simulations were performed at 310 K and 1 atm pressure with 2 fs integration time step.

We have implemented periodic boundary condition across the system using a TIP3P water box. We used the Particle Mesh Ewald (PME) techniques integrated with the AMBER package to account for the long-range component of the electrostatic interactions. During the dynamics, all the bonds involving hydrogen were restrained using the SHAKE algorithm. Langevin thermostat with collision frequency of 1 ps-1 is used to maintain the constant temperature while the pressure was controlled by anisotropic Monte-Carlo barostat. The accelerated GPU version of PMEMD was performed on NVIDIA GeForce 20 Series cards. We have employed CPPTRAJ

functionality of AMBERTOOLs to perform various analyses on the equilibrium MD simulation trajectories. The images and graphics of the structures shown here were generated using the software packages VMD, and PyMOL.

Protein engineering of ZaTdT

Because a total of 16 residues was selected for high-throughput screening, saturation mutagenesis based on NNN or NNK codons encoding the complete set of standard amino acids requires a tremendous amount of work. In order to strike a balance between the size and diversity of the combinatorial mutant libraries, oligonucleotide primers were designed with degenerate codons encoding a defined subset of amino acids with three standards. For a 95% library coverage, it is necessary to screen strains with more than three times the number of mutations contained in codon degeneracy. Therefore, each double-point combinatorial mutation library screened 270 mutants. Mutation library was generated based on PCR-based QuickChange method. PCR reaction was performed with Fastpfu DNA Polymerase (Transgen, China) under the following conditions: the reaction was started at 94 °C (5 min), followed by 30 cycles 94 °C (20 s), 58 °C (20 s), 72 °C (3.5 min), with a final extension at 72 °C (5 min). The PCR product was digested with *Dpn*I restriction enzyme and transformed into *E. coli* BL21 (DE3) competent cells to create the screening library.

Each of the mutant colonies was picked and incubated 24 hours in 200 μL LB medium with 100 μg mL^{-1} kanamycin while shaking at 37 °C in 96-well microplate, and then scaled up to 1 mL LB medium for TdT protein expression. The cell pellets were harvested by centrifugation at 3,300 g for 1 min and lysed by the re-suspension in 100 μL B-PER Bacterial Protein Extraction Reagent (Thermo Scientific, US), followed by 15 min at 37 °C. Subsequently, add 30 μL of magnetic beads with nickel filler (Beaver, China) to attract protein. And then 40 μL reaction buffer was added directly to the immobilized protein for condensation assay and the plates were further incubated at 30 °C at 750 rpm for 60 min. After removing the precipitate by centrifugation, 30 μL sample was used for PPi detection. PPi was tested by pyrophosphate test kit (AAT Bioquest, US), followed the method of the operation manual.

Enzyme kinetics were determined with 3-ONH_2-dNTPs as substrate. The reaction system comprised 1 μM oligonucleotides (P1-CT, Supplementary Table 7), 0.25 mM 3'-ONH_2-dNTPs, 0.25 mM $CoCl_2$, 100 mM NaCl and 50 mM phosphate buffer (pH 6.8). The catalytic reaction was started by adding 0.5 mg/mL purified protein. For each modified nucleotide, the concentration range was 1 to 200 μM. Time course experiments were used to determine the reaction rate for each concentration. Kinetic parameters were obtained by fitting the data to the Michaelis-Menten equation by nonlinear regression analysis using GraphPad Prism 6 (GraphPad Software, USA). For metal ions, the concentration range was 0.025 to 1 mM. Kinetic parameters were determined as described above.

Functional verification of the reasonably designed ZaTdT variant

Each single-site mutant was generated according to the PCR-based QuickChange method. PCR reaction was performed with Fastpfu DNA Polymerase (Transgen, China) under the following conditions: the reaction was started at 94 °C (5 min), followed by 30 cycles 94 °C (20 s), 58 °C (20 s), 72 °C (3.5 min), with a final extension at 72 °C (5 min). The PCR product was digested with *Dpn* I restriction enzyme and transformed into BL21 (DE3) *E. coli* competent cells. Appropriate single clone was used to express the mutant protein and verify its activity. For modified nucleotides, the gel electrophoresis was used to verify the incorporation efficiency of the mutants, and the detailed functional verification method was consistent with ZaTdT.

Extension of a DNA strand by the sequences 5'- ATGACGTGCT-3'

The 5' biotinylated oligonucleotide (P2, Supplementary Table 7) was immobilized onto the magnetic beads loaded with streptavidin (Dynabeads M-270 Streptavidin, Dynal Biotech). All subsequent reactions were done on the magnetic beads. All nucleotide additions were performed at 30 °C for 10 min. The reaction system comprised 1 μM immobilized oligonucleotides (P2), 0.25 mM $CoCl_2$, 100 mM NaCl and 50 mM phosphate buffer (pH 6.8), added 0.25 mM different modified nucleotides as needed. For the 10-mer synthesis, the following modified nucleotides were used in

the extension steps: 1) 3'-ONH_2-dATP, 2) 3'-ONH_2-dTTP, 3) 3'-ONH_2-dGTP, 4) 3'-ONH_2-dATP, 5) 3'-ONH_2-dCTP, 6) 3'-ONH_2-dGTP, 7) 3'-ONH_2-dTTP, 8) 3'-ONH_2-dGTP, 9) 3'-ONH_2-dCTP, 10) 3'-ONH_2-dTTP.

The first round of primer extension was performed by adding 1 mg/mL TdT-mutant. The mixtures were incubated at 30 °C for 10 min. Then, the products were treated with sodium nitrite buffer (700 mM, pH 5, adjust pH with nitrous acid) to cleave the 3'-ONH_2 group, generating a free 3'-OH group. The cleavage buffer was removed and the beads were washed using lysis buffer (50 mM potassium phosphate buffer, pH 6.8, 100 mM NaCl). The subsequent cycles of reaction, cleavage and washing followed the same steps as described for the first cycle. The synthesized product was purified by HPLC and used for subsequent library construction for next-generation sequencing library preparation.

Sequence library preparation and next-generation sequencing analysis of synthesis products

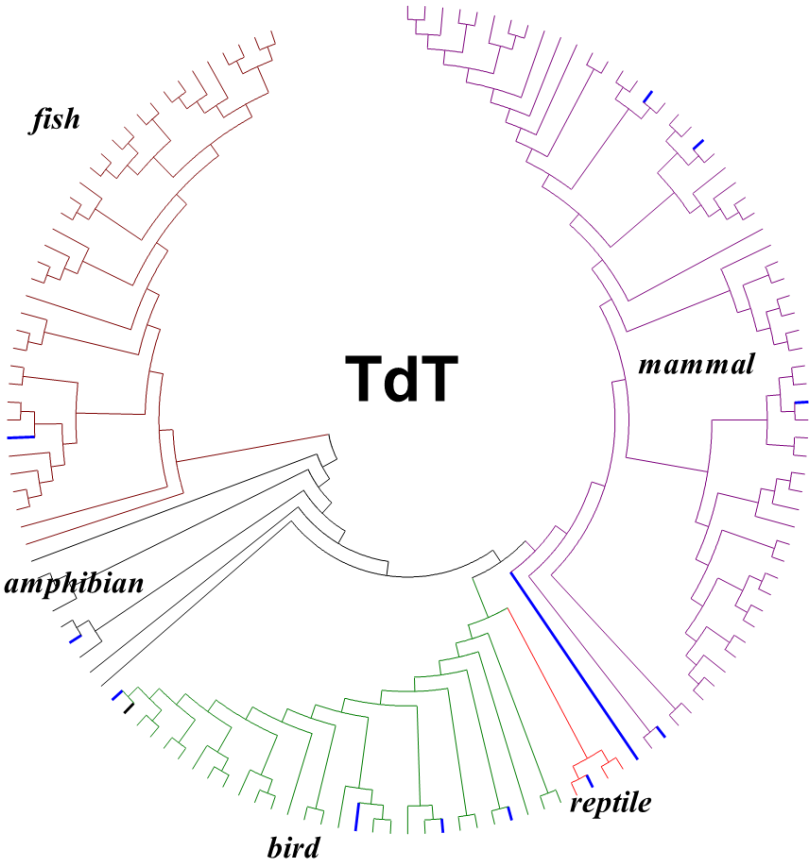
The synthesis products were poly-adenylated on the beads using 1 U/μL TdT (NEB) in combination with 0.25 mM dATP in TdT reaction Buffer (NEB) for 20 min at 30 °C. The beads with immobilized oligonucleotides were then washed and collected. The polyA tailed products were prepared for dual-index paired-end sequencing analysis by two steps. In the first step, the polyA tailed products were amplified using common primers (C1, Supplementary Table 7). PCR conditions were 98 °C for 5 min; 5 cycles of 98 °C for 10 s, 60 °C for 30 s and 72°C for 20 s; and a final extension at 72°C for 30 s. The second step of PCR was used to attach dual indices and Illumina sequencing adapters to the amplicons from the first step PCR reaction. These primers (Supplementary Table 7) were complementary to the sequences on the first PCR primers and included the P5 and P7 adapter sequences to permit binding to the Illumina flow cell and an 8-base pair molecular barcode i5 or i7. PCR reaction conditions were: 98 °C for 2 min; followed by 8 cycles of 98 °C for 20s, 58 °C for 10 s, 72 °C for 20 s; and a 10 min final elongation at 72 °C. PCR reactions were purified by DNA Clean & Concentrator kit ("DCC", Zymo Research) and pooled for a shared 2 × 100 bp HiSeq

(Illumina, San Diego, CA) run.

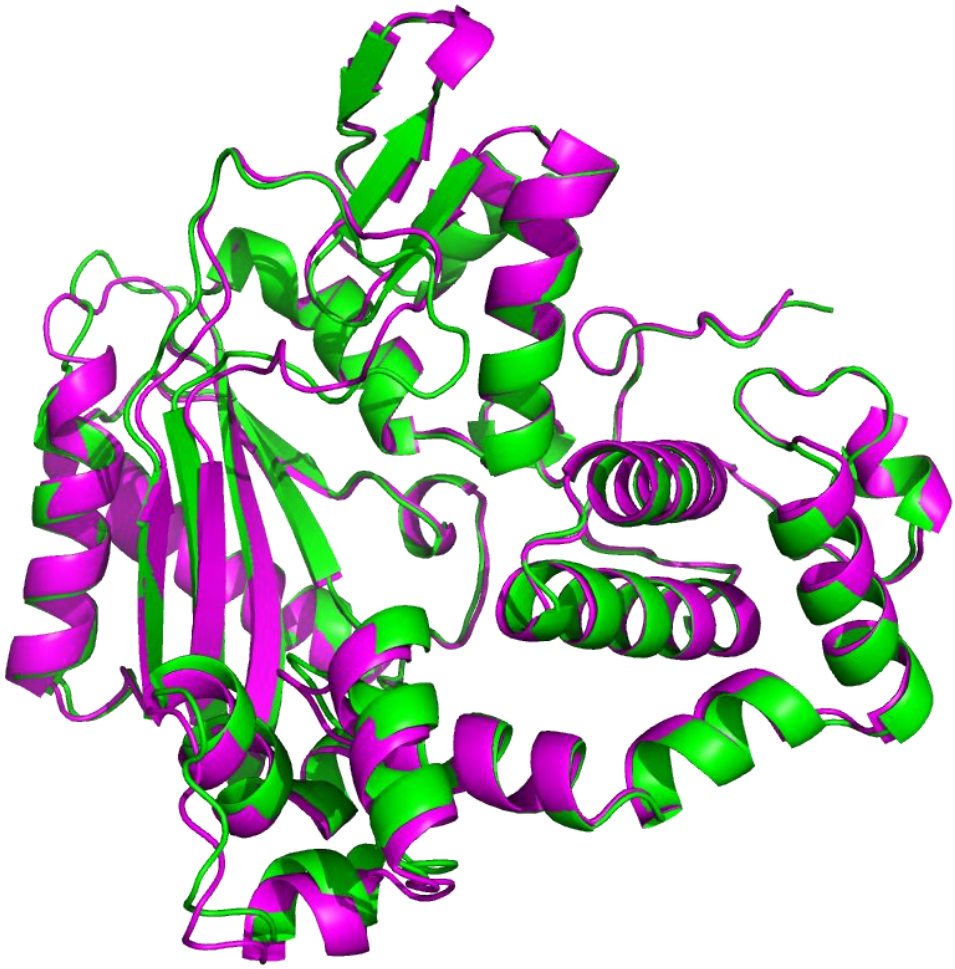
License description

The plasmid sequences of ZaTdT and ZaTdT-R335L-K337G have been deposited in NCBI with the accession number OL989372 and OL989373. When only used for scientific research, the plasmid can be obtained from the author.

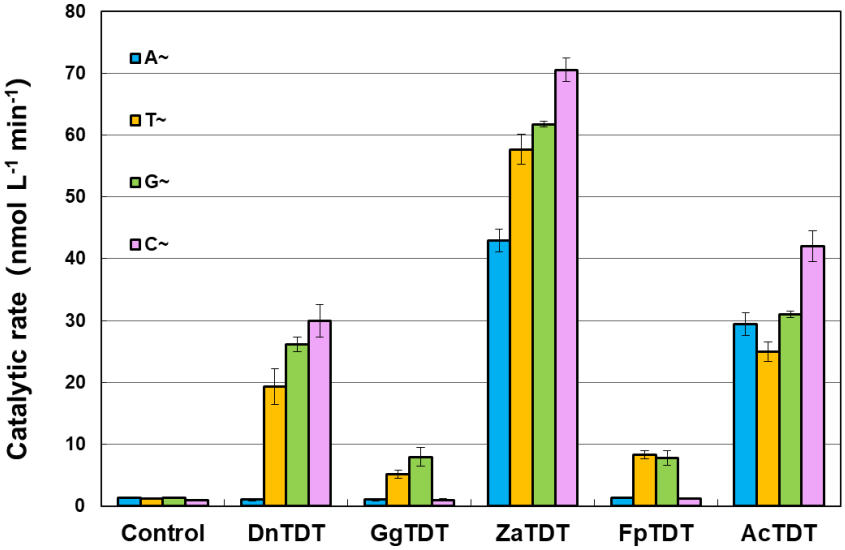
Supplementary Figures



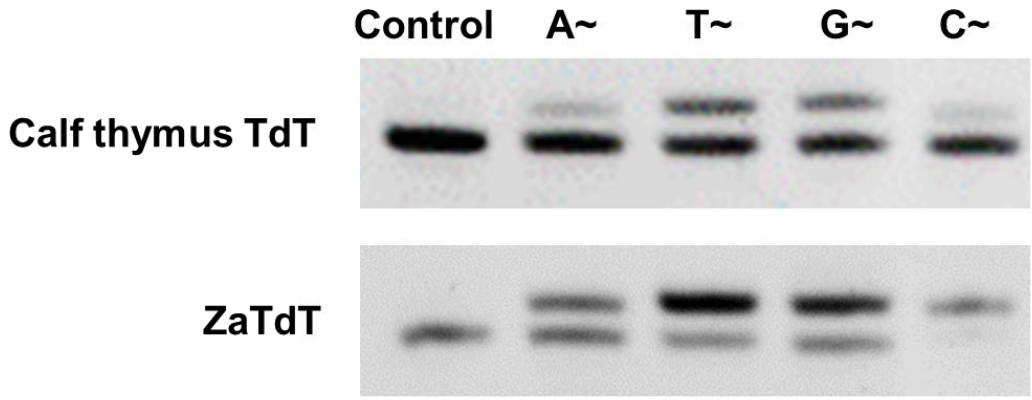
Supplementary Figure 1. Phylogenetic tree of TDTs from 137 vertebrates. The maximum-likelihood tree was constructed by MEGA. Specific sequences of interest are highlighted in blue.



Supplementary Figure 2. Structure model of ZaTdT. The structure of the model used in this experiment (magenta) is not significantly different from the structure predicted by Alphafold2 (green). The RMSD of these two structures is 1.0 Å.



Supplementary Figure 3. Incorporation of four 3′ -ONH$_2$ reversible terminators by avian TdTs. A~: 3'-ONH$_2$-dATP; T~: 3'-ONH$_2$-dTTP; G~: 3'-ONH$_2$-dGTP; C~: 3'-ONH$_2$-dCTP. The catalytic efficiency was calculated by analyzing the percentage of the initiators with nucleotide incorporated in 20 minutes by using the denaturing PAGE. Control means no enzyme was added to the sample. Error bars represent s.d. (standard deviation), *n*=3.



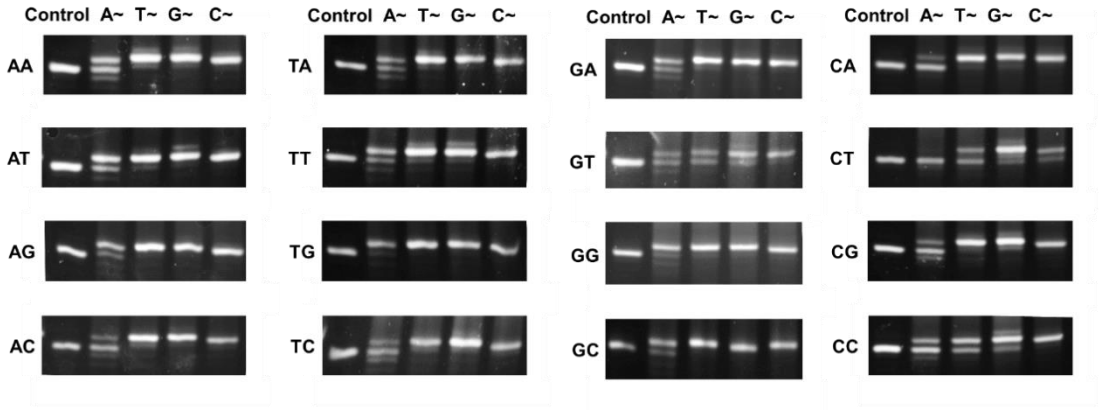
230

231 **Supplementary Figure 4.** Catalytic efficiency of commercial Calf thymus TdT and

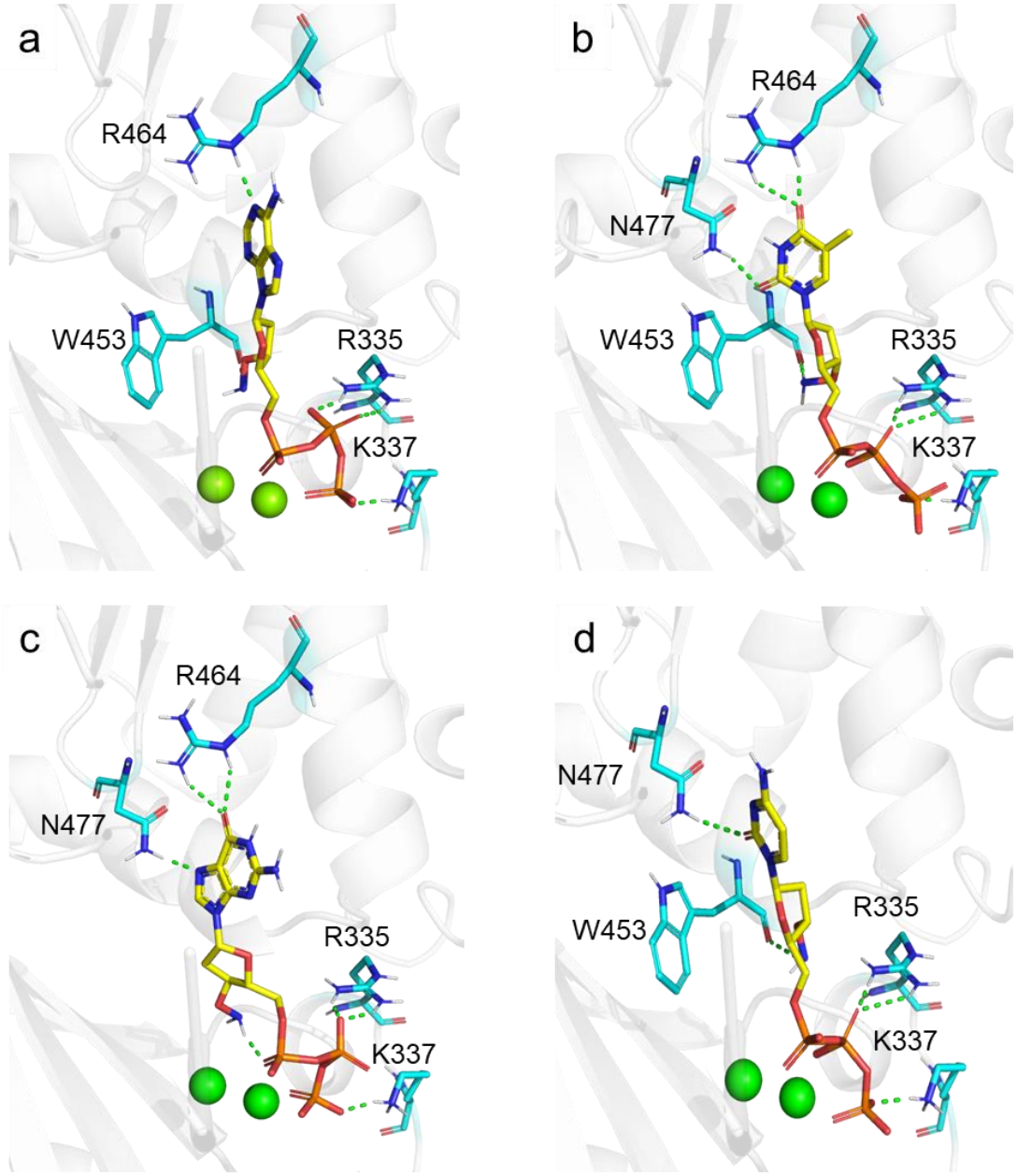
232 ZaTdT. Control means no nucleotide was added to the sample. A~: 3'-ONH_2-dATP; T~:

233 3'-ONH_2-dTTP; G~: 3'-ONH_2-dGTP; C~: 3'-ONH_2-dCTP. The initiator was P1-TA.

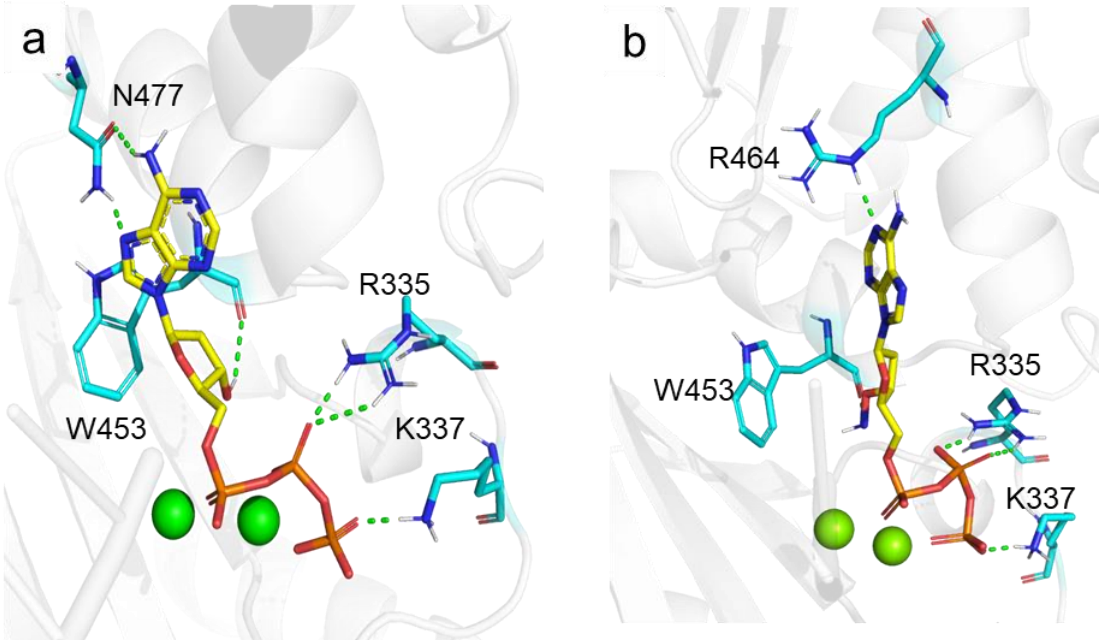
234



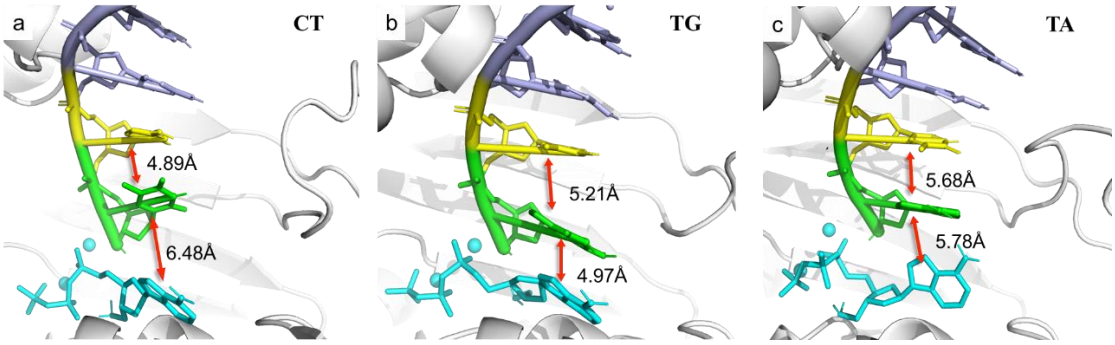
Supplementary Figure 5. The denaturing PAGE analysis of single nucleotide extension assay obtained from ZaTdT. Control means no nucleotide was added to the sample. A~: 3'-ONH_2-dATP; T~: 3'-ONH_2-dTTP; G~: 3'-ONH_2-dGTP; C~: 3'-ONH_2-dCTP. AA to AC respectively represent 16 initiators, which contain all possible dinucleotides at the 3'end. After incorporating a new nucleotide, the initiator had a slower electrophoresis speed due to its larger molecular weight.



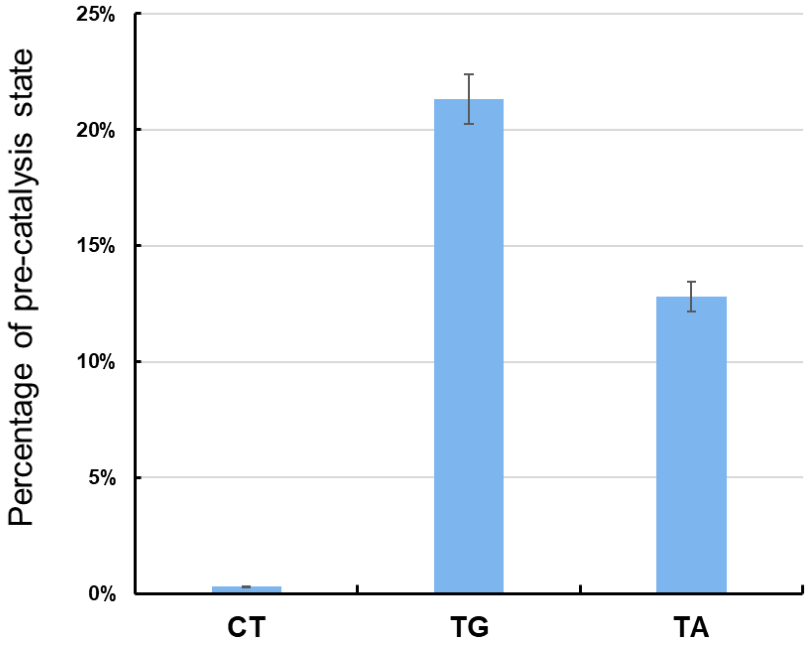
Supplementary Figure 6. Hydrogen bond network between modified nucleotides and ZaTdT. Hydrogen bonding of 3'-ONH$_2$-dATP (a), 3'-ONH$_2$-dTTP (b), 3'-ONH$_2$-dGTP (c), 3'-ONH$_2$-dCTP (d) in ZaTdT, where the triphosphate of each substrate has hydrogen bonding interactions with R335, K337. The binding of the base part of each substrate is different, where 3'-ONH$_2$-dATP (a) has only a weak hydrogen bonding interaction with N-H--N and the other substrates have relatively strong interactions (N-H-O) or have more hydrogen bonds. Residues are in cyan sticks and substrates are in yellow sticks.



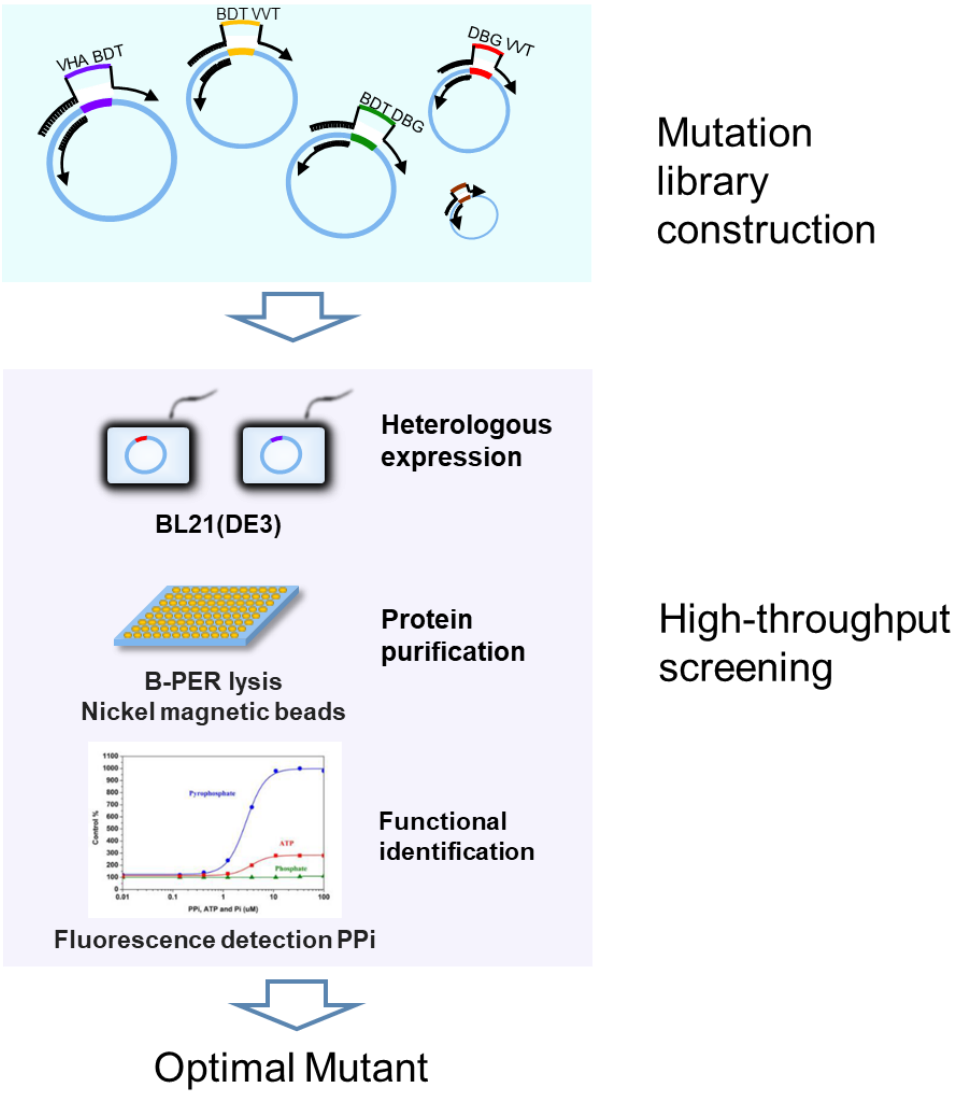
Supplementary Figure 7. Hydrogen bond network. Hydrogen bonding of 3'-dATP (a) and 3'-ONH_2-dATP (b) in ZaTdT, where the triphosphate of each substrate has hydrogen bonding interactions with R335, K337 and the main chain of W453 has hydrogen bonding interaction with OH in 3'-dATP (a) and ONH_2 in 3'-ONH_2-dATP (b). Different hydrogen binding interactions were shown in base part of substrate, there are two hydrogen bonds between dATP with N477 and which has one between 3'-ONH_2-dATP and R464. Residues are shown as cyan sticks and substrates are in yellow sticks.



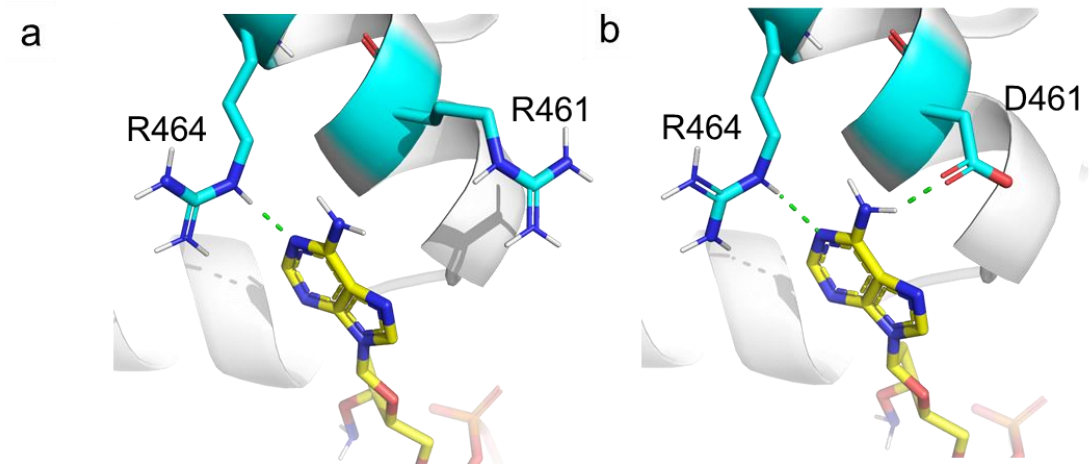
Supplementary Figure 8. The stacking state of different initiators in ZaTdT. 3'-ONH_2-dATP and the last two bases are shown as cyan, green and yellow sticks, respectively. Red lines indicated the corresponding base-base distance (average values calculated during MD trajectories) between two terminal bases, and between the last 3' base and the incoming nucleotide (3'-ONH_2-dATP). CT, TG, TA represent the dinucleotides at the 3'end of the initiator, respectively.



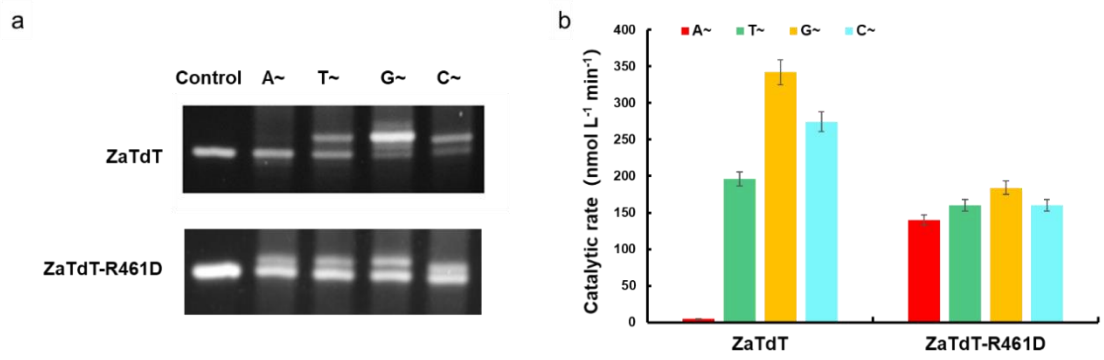
Supplementary Figure 9. Percentage of precatalytic state of different nucleotide chains. The precatalytic state will be counted that satisfy both of the following criteria. The distance between the oxygen atom in OH of the last 3' base and the phosphorus atom in the first phosphate group of triphosphates of the incoming nucleotide (3'-ONH_2-dATP) is smaller than 3.5 Å, and the angle between the oxygen atom in OH of the last 3' base and the phosphorus atom in the first phosphate group and its neighbor oxygen atom which is connect to the second phosphate group of triphosphates of the incoming nucleotide (3'-ONH_2-dATP) should be larger than 150 degrees.



Supplementary Figure 10. Iterative mutation and screening methods. The colors on the different plasmids represent different mutations in the mutant library construction.



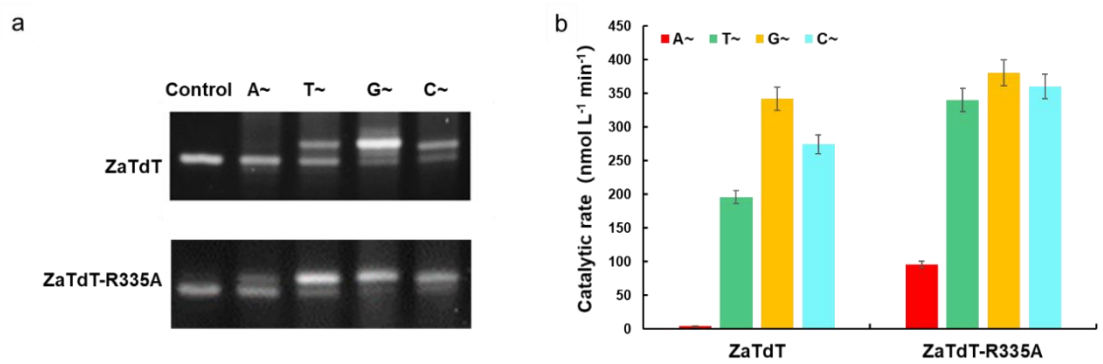
Supplementary Figure 11. Hydrogen bond network. Hydrogen bonding interactions between base part of 3'-ONH_2-dATP and residues R464, R461(a) or D461 (b). The dotted green line indicates hydrogen bonding interaction. Residues are shown as cyan sticks and substrate is in yellow sticks.



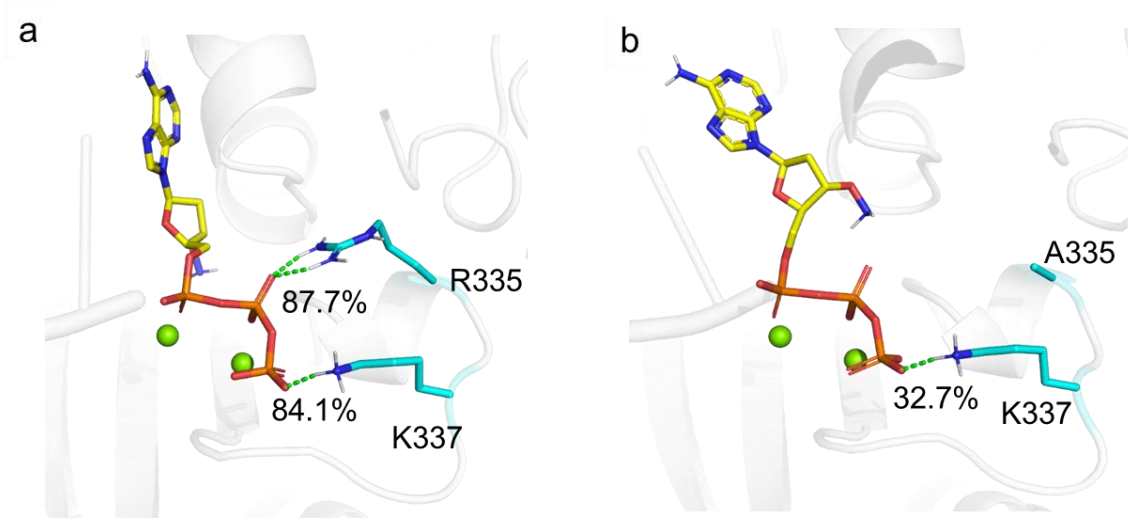
297

298 **Supplementary Figure 12. Effect of R461D mutation on the catalytic efficiency of**
299 **ZaTdT.** a, The denaturing PAGE analysis. Control means no nucleotide was added to
300 the sample. A~: 3'-ONH_2-dATP; T~: 3'-ONH_2-dTTP; G~: 3'-ONH_2-dGTP; C~: 3'-
301 ONH_2-dCTP. After incorporating new nucleotides, the initiator had a slower
302 electrophoresis speed due to its larger molecular weight. b, Percentage of the initiators
303 with nucleotide incorporated. Error bars represent s.d. (standard deviation), *n*=3.

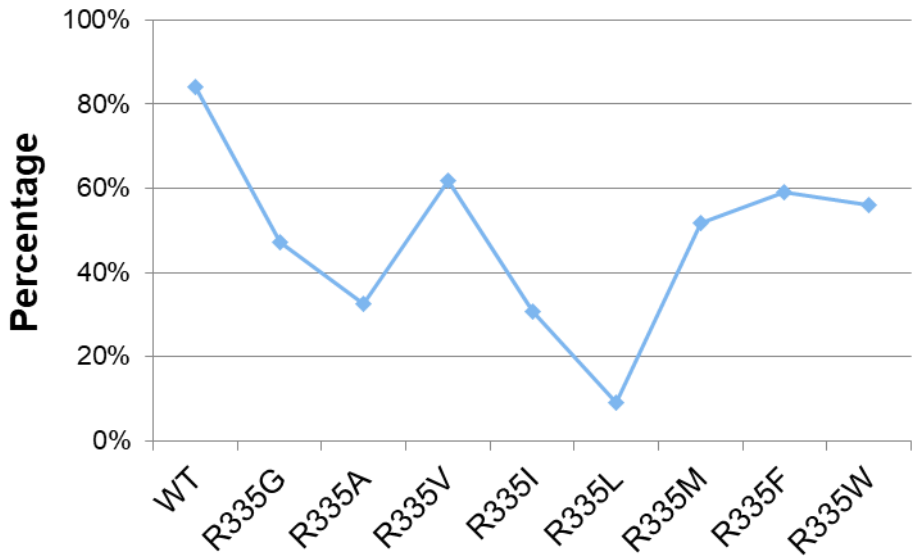
304



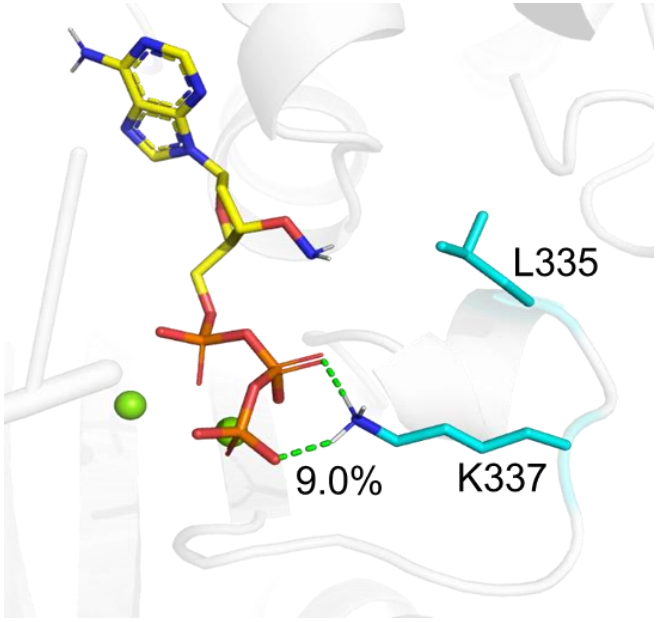
Supplementary Figure 13. Effect of R355A mutation on the catalytic efficiency of ZaTdT. a, The denaturing PAGE analysis. Control means no nucleotide was added to the sample. A~: 3'-ONH_2-dATP; T~: 3'-ONH_2-dTTP; G~: 3'-ONH_2-dGTP; C~: 3'-ONH_2-dCTP. After incorporating new nucleotides, the initiator had a slower electrophoresis speed due to its larger molecular weight. b, Percentage of the initiators with nucleotide incorporated. Error bars represent s.d. (standard deviation), *n*=3.



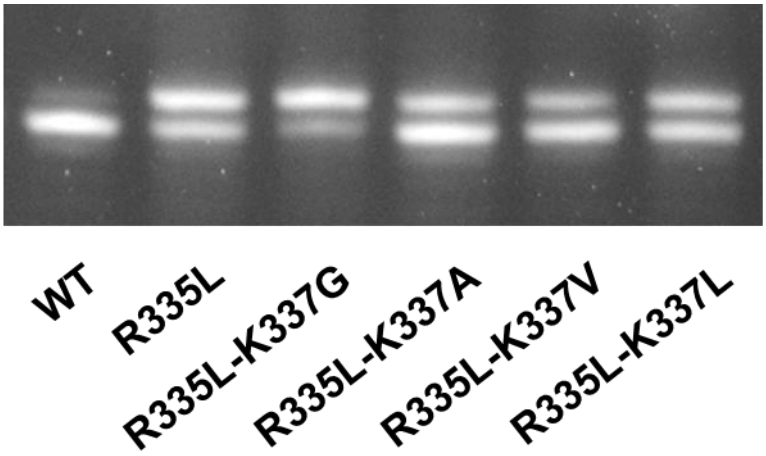
Supplementary Figure 14. The percentage of the hydrogen bonding between triphosphate of 3'-ONH_2-dATP and residues in position 335 and 337 during MD trajectories. a, The hydrogen bonding interaction between residues 335, 337 and 3'-ONH_2-dATP with ZaTdT-WT. b, The hydrogen bonding interaction between residues 335, 337 and 3'-ONH_2-dATP with ZaTdT-R335A. The dotted green line indicates hydrogen bonding interaction. Residues are shown as cyan sticks and substrate is in yellow stick.



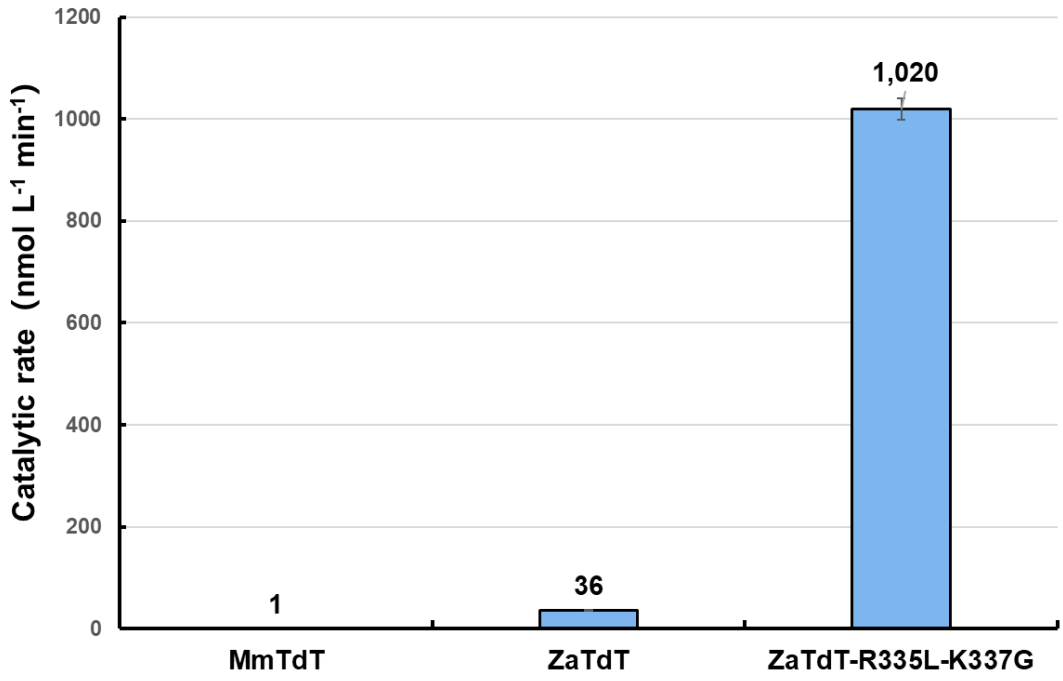
Supplementary Figure 15. The effect of different mutations at position 335 on the proportion of hydrogen bonds at position 337. The vertical coordinate indicates the percentage of the hydrogen bonding interaction between residue 337 and triphosphate of 3'-ONH_2-dATP.



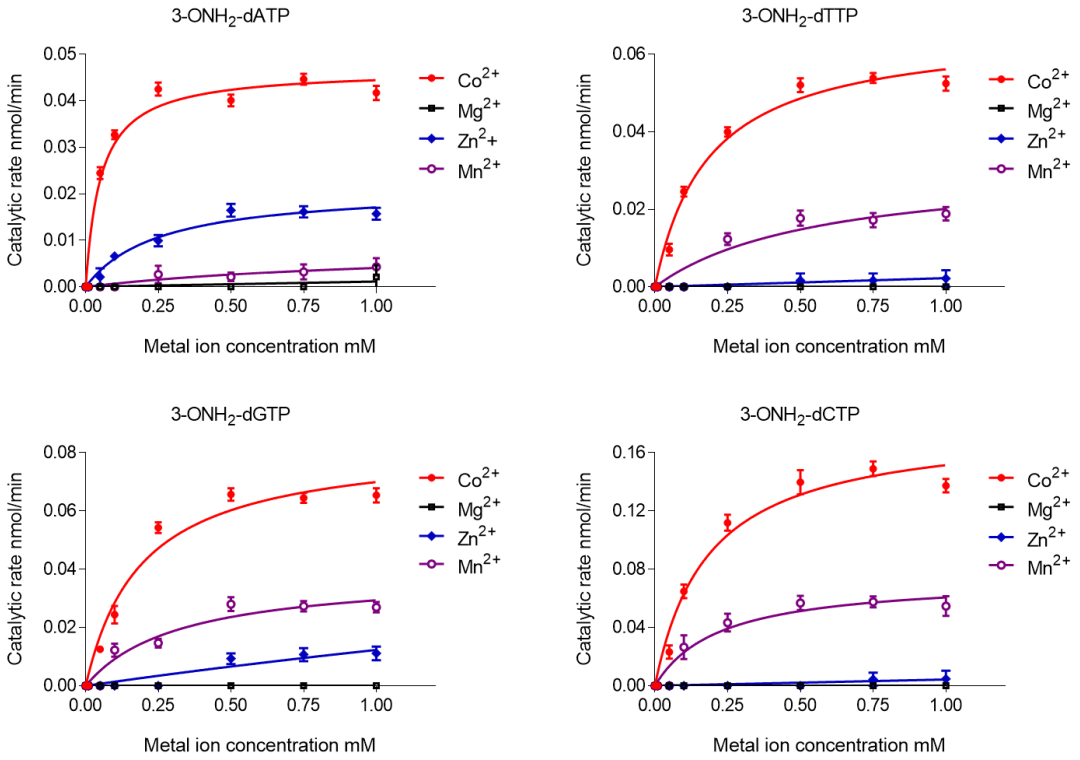
Supplementary Figure 16. The percentage of hydrogen bonding interaction between residues 335, 337 and 3'-ONH_2-dATP with ZaTdT-R335L based on MD analysis. The dotted green line indicates hydrogen bonding interaction. Residues are shown as cyan sticks and substrate is in yellow stick.



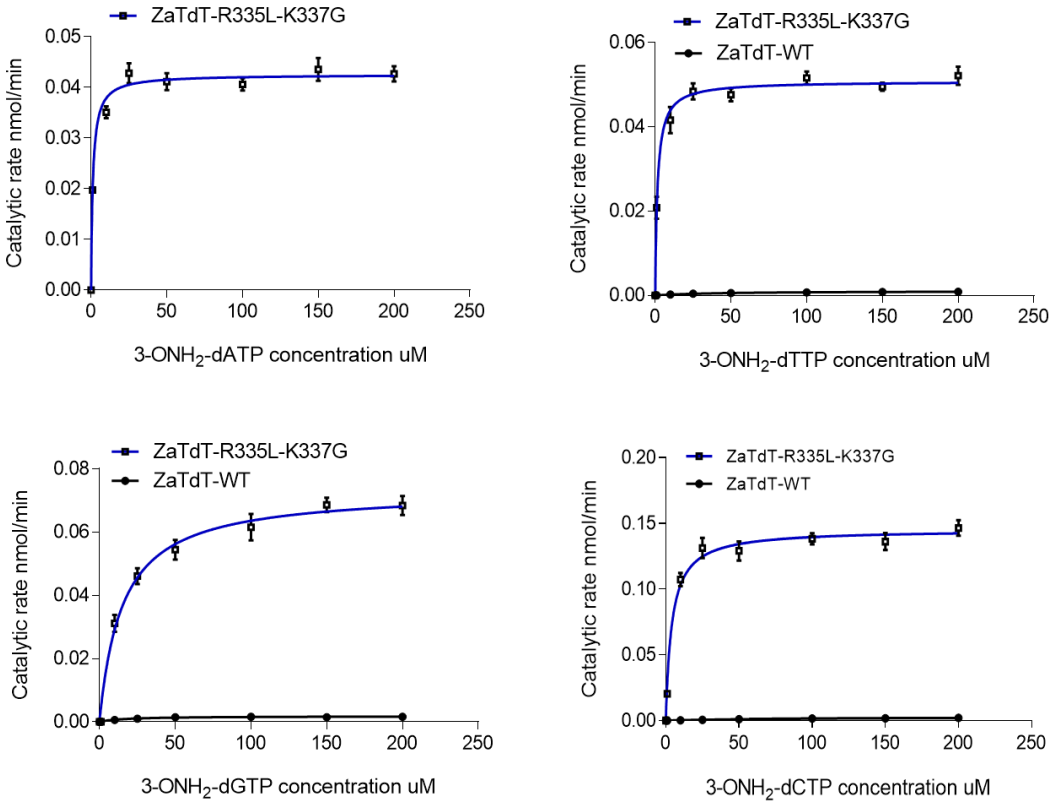
Supplementary Figure 17. PAGE analysis of the activity for different mutants. The end of the initiator was CT, and the inserted nucleotide was 3'-ONH_2-dATP. The initiator with incorporated nucleotides had a higher electrophoresis position than the original initiator.



Supplementary Figure 18. Catalytic efficiency of different TdTs. The catalytic efficiency was determined by testing the incorporation efficiency of 3'-ONH_2-dATP added to the initiator terminated with TA di-nucleotide. Incorporation efficiency was analyzed by NGS sequencing. Error bars represent s.d. (standard deviation), *n*=3.



Supplementary Figure 19. Enzyme kinetics of different metal ions. Kinetics were determined using 0.5 mg mL^{-1} ZaTdT-R335L-K337G. Error bars represent s.d. (standard deviation), *n*=3.



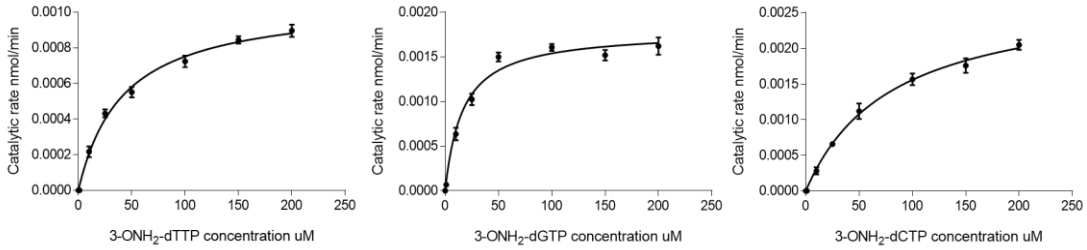
364

365 **Supplementary Figure 20. Comparison of the kinetics for wild-type and mutant**
366 **protein.** Kinetics were determined using 0.5 mg mL^{-1} ZaTdT-WT or ZaTdT-R335L-
367 K337G. ZaTdT-WT refers to wild-type ZaTdT without protein engineering. Error bars
368 represent s.d. (standard deviation), *n*=3.

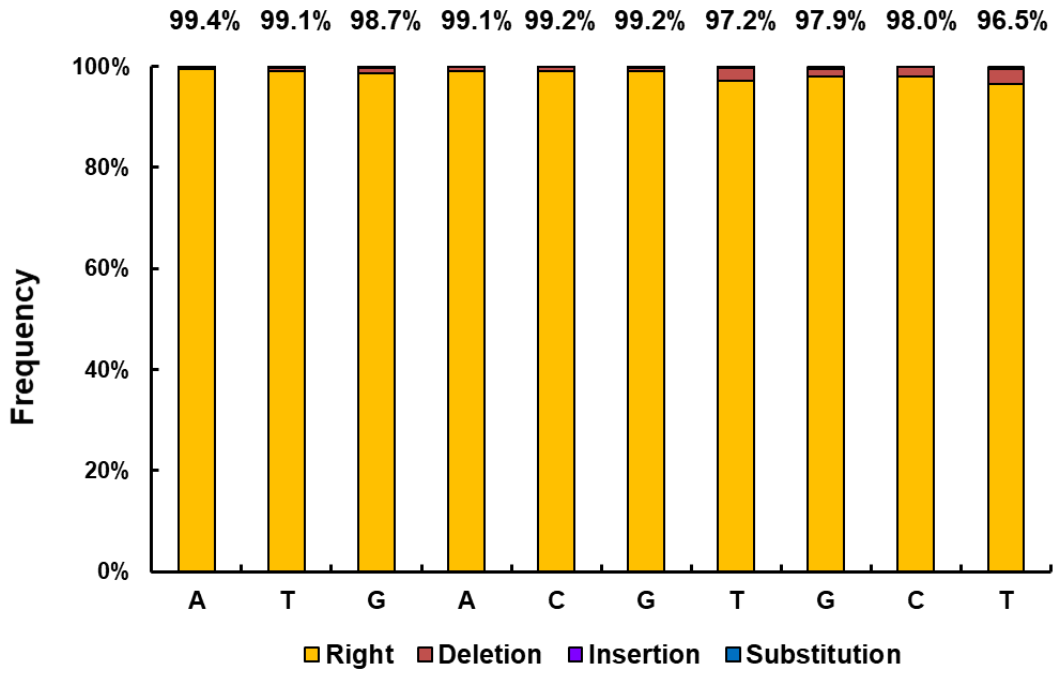
369

370

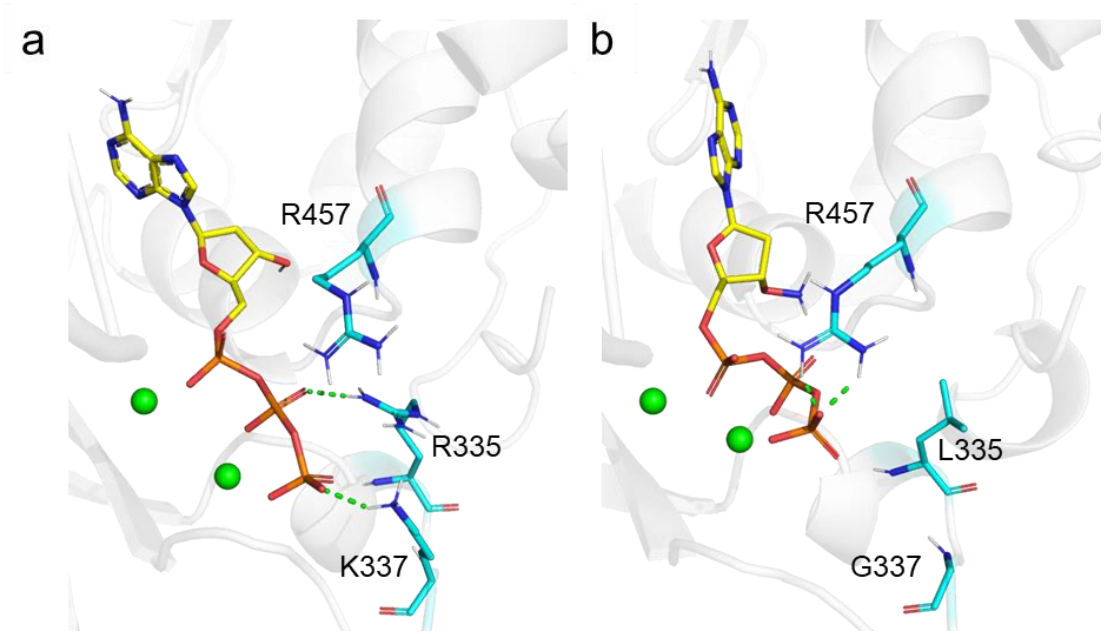
371



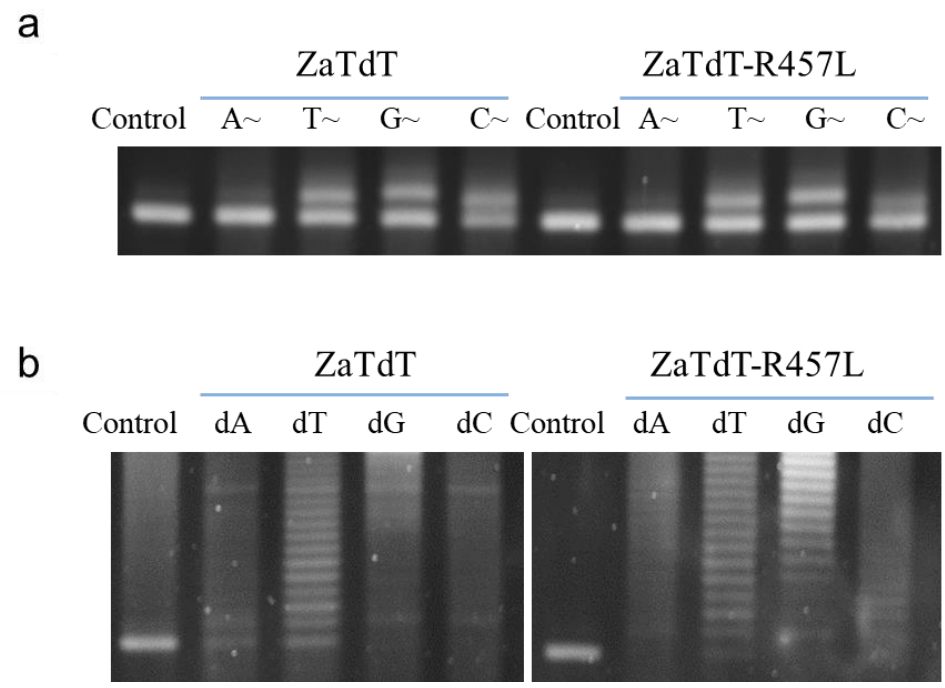
Supplementary Figure 21. Enzyme kinetics of different nucleotides for wild-type ZaTdT. Kinetics were determined using 0.5 mg mL^{-1} protein. Error bars represent s.d. (standard deviation), *n*=3.



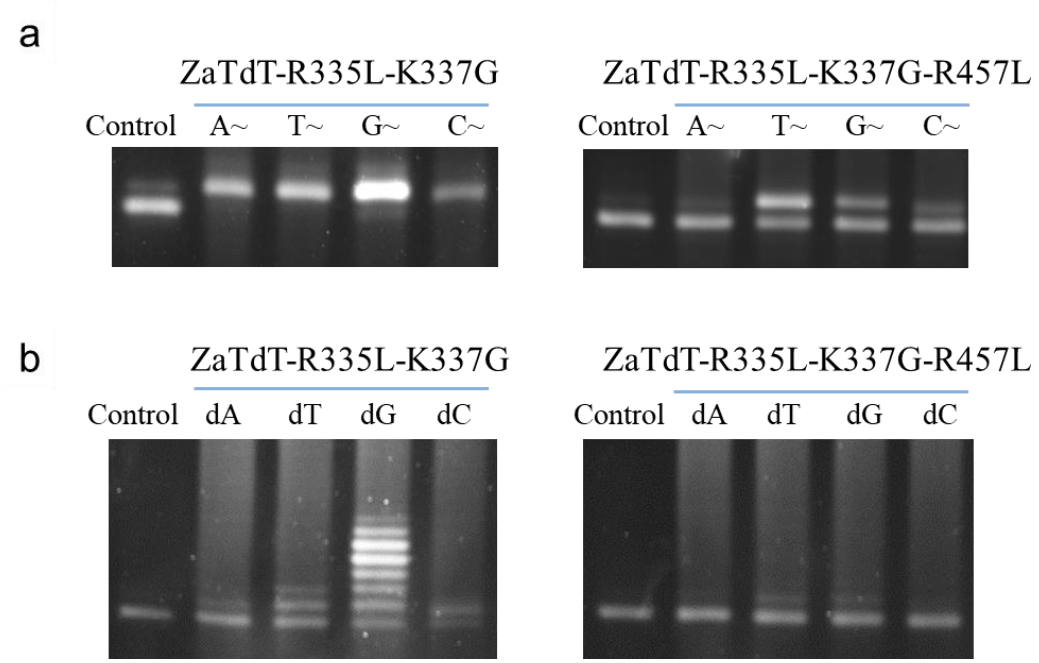
Supplementary Figure 22. Estimates of the stepwise yields of the synthesis based on an alignment against the target sequence. The Y-axis refers to the synthesis accuracy of the single-step reaction, which was calculated by dividing the correct synthesis number of this step by the total.



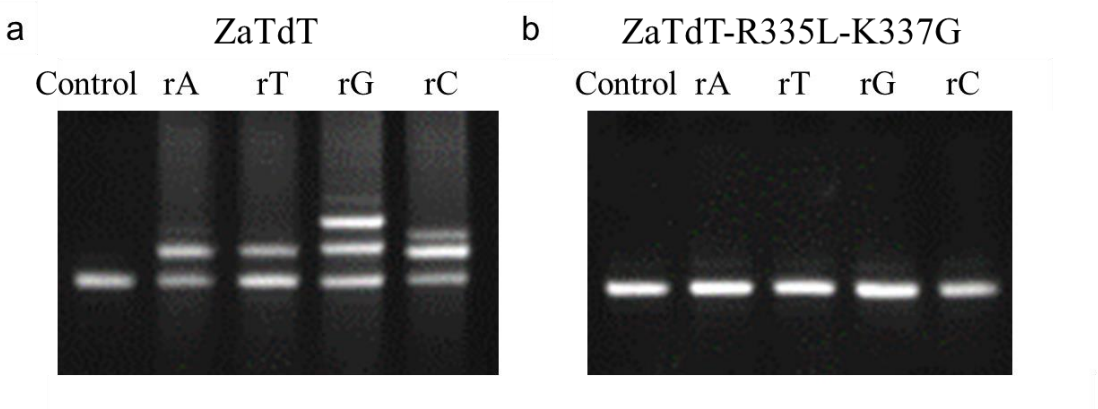
Supplementary Figure 23. The substrate binding conformation. a, The substrate binding conformation in wild ZaTdT. b, The substrate binding conformation in double mutational ZaTdT.



Supplementary Figure 24. The effect of R457L mutation on ZaTdT activity. a, Denaturing PAGE analysis of the products stemming from the reactions with modified nucleotides catalyzed by the R457L variant. b, Denaturing PAGE analysis of the products stemming from the reactions with natural nucleotides catalyzed by the R457L variant. Control means no nucleotide was added to the sample. A~: 3'-ONH_2-dATP; T~: 3'-ONH_2-dTTP; G~: 3'-ONH_2-dGTP; C~: 3'-ONH_2-dCTP. dA: dATP; dT: dTTP; dG: dGTP; dC: dCTP.



Supplementary Figure 25. The effect of R457L mutation on ZaTdT-R335L-K337G activity. a, Denaturing PAGE analysis of the products stemming from the reactions with modified nucleotides catalyzed by the R457L variant. b, Denaturing PAGE analysis of the products stemming from the reactions with natural nucleotides catalyzed by the R457L variant. Control means no nucleotide was added to the sample. A~: 3'-ONH_2-dATP; T~: 3'-ONH_2-dTTP; G~: 3'-ONH_2-dGTP; C~: 3'-ONH_2-dCTP. dA: dATP; dT: dTTP; dG: dGTP; dC: dCTP.



412

413 **Supplementary Figure 26. Catalytic efficiency of wild-type and mutant protein on**
414 **ribonucleotides.** a, Denaturing PAGE analysis of the products stemming from the
415 reactions with ribonucleotides catalyzed by ZaTdT. b, Denaturing PAGE analysis of the
416 products stemming from the reactions with ribonucleotides catalyzed by ZaTdT-
417 R335L-K337G. Control means no nucleotide was added to the sample. rA: rATP; rT:
418 rTTP; rG: rGTP; rC: rCTP.

419

420

Supplementary Tables

Supplementary Table 1. The frequency of pre-catalytic state for the incorporation of natural and modified nucleotides.

Enzyme	Substrate	Percentage of pre-catalysis state
MmTdT	dATP	0.8167
ZaTdT	dATP	0.8920
ZaTdT	3’-ONH_2-dATP	0.1284

Supplementary Table 2. Degenerate codons and their corresponding amino acid residues.

Mutation combination	Degenerate codon	Residues
T330-G331	T330-VHA	A^1 I^1 L^1 V^1 P^3 Q^3 T^3 K^4 E^5
	G331-BDT	G^1 L^1 V^1 F^2 Y^2 C^3 H^4 R^4 D^5
G332-R335	G332-BDT	G^1 L^1 V^1 F^2 Y^2 C^3 H^4 R^4 D^5
	R335-BDT	G^1 L^1 V^1 F^2 Y^2 C^3 H^4 R^4 D^5
G452-W453	G452-BDT	G^1 L^1 V^1 F^2 Y^2 C^3 H^4 R^4 D^5
	W453-DBG	A^1 G^1 L^1 M^1 V^1 W^2 S^3 T^3 R^4
T454-G455	T454-VHA	A^1 I^1 L^1 V^1 P^3 Q^3 T^3 K^4 E^5
	G455-BDT	G^1 L^1 V^1 F^2 Y^2 C^3 H^4 R^4 D^5
S456-R457	S456-DBG	A^1 G^1 L^1 M^1 V^1 W^2 S^3 T^3 R^4
	R457-BDT	G^1 L^1 V^1 F^2 Y^2 C^3 H^4 R^4 D^5
Q458-G460	Q458-VHA	A^1 I^1 L^1 V^1 P^3 Q^3 T^3 K^4 E^5
	G460-BDT	G^1 L^1 V^1 F^2 Y^2 C^3 H^4 R^4 D^5
R461-D462	R461-BDT	G^1 L^1 V^1 F^2 Y^2 C^3 H^4 R^4 D^5
	D462-BDT	G^1 L^1 V^1 F^2 Y^2 C^3 H^4 R^4 D^5
R464-N477	R464-BDT	G^1 L^1 V^1 F^2 Y^2 C^3 H^4 R^4 D^5
	N477-DDT	G^1 I^1 V^1 F^2 Y^2 C^3 N^3 S^3 D^5

The second col indicates the selected degenerate codon libraries for different residues. For example, 'T330-VHA' represents that we used 'VHA' as degenerate codon library for T330, where 'VHA' refers to [ACG][ACT]A. For other degenerate codon libraries, 'BDT' refers to [CGT][AGT]T, 'DBG' refers to [AGT][CGT]G, and 'DDT' refers to [AGT][AGT]T.

[1] represents hydrophobic-aliphatic; [2] represents hydrophobic-aromatic; [3] represents hydrophilic-polar uncharged; [4] represents hydrophilic-acidic; [5] represents hydrophilic-basic.

Supplementary Table 3. The frequency of pre-catalytic state for the mutations in R335 and K337.

Enzyme	Percentage of pre-catalysis state
WT	0.1284
R335A	0.1217
R335L	0.3076
R335L-K337G	0.3360
R335L-K337A	0.2535
R335L-K337V	0.2927
R335L-K337L	0.2740

The substrate was 3'-ONH_2-dATP.

Supplementary Table 4. The kinetic parameters of ZaTdT-R335L-K337G variant influenced by four metal cofactors.

Measures	3'-ONH_2-blocked nucleotides	Co^{2+}	Mg^{2+}	Zn^{2+}	Mn^{2+}
Km [M]	A~	5.14E-05	ND	2.52E-04	1.19E-03
	T~	1.84E-04	ND	ND	5.63E-04
	G~	1.91E-04	ND	6.26E-03	3.21E-04
	C~	1.83E-04	ND	ND	2.22E-04
kcat [min^{-1}]	A~	0.2750	ND	0.1260	0.0528
	T~	0.3910	ND	ND	0.1850
	G~	0.4900	ND	0.5270	0.2280
	C~	1.0500	ND	ND	0.4360
kcat/Km [M^{-1} min^{-1}]	A~	5350	ND	500	44
	T~	2130	ND	ND	329
	G~	2570	ND	84	710
	C~	5740	ND	ND	1960

'ND' represents not detected due to low activity.

443 **Supplementary Table 5. TdT genes used in this study.**

Gene name	Species name	Gene accession in NCBI
DrTdT	Danio rerio	NM_001014817.1
XlTdT	Xenopus laevis	NM_001085782.1
OaTdT	Ornithorhynchus anatinus	XM_029061259.1
ShTdT	Sarcophilus harrisii	XM_003755238.3
MbTdT	Myotis brandtii	XM_005885424.2
HgTdT	Heterocephalus glaber	XM_004838573.2
MmTdT	Mus musculus	NM_001043228.1
PtTdT	Pan troglodytes	XM_521569.6
PmTdT	Podarcis muralis	XM_028728544.1
DnTdT	Dromaius novaehollandiae	XM_026094160.1
GgTdT	Gallus gallus	XM_015288597.2
ZaTdT	Zonotrichia albicollis	XM_026799623.1
FpTdT	Falco peregrinus	XM_027788920.1
AcTdT	Aquila chrysaetos	XM_030029774.1

444

445

Supplementary Table 6. Plasmids used in this study.

Plasmids	Relevant characteristics	Source
pET28a	pBR322 ori with *pT7*; *KanR*	Novagen
pET28a- DrTdT	pET28a vector, *Nde*I-DrTdT-*Xho*I	This study
pET28a- XlTdT	pET28a vector, *Nde*I-XlTdT-*Xho*I	This study
pET28a- OaTdT	pET28a vector, *Nde*I-OaTdT-*Xho*I	This study
pET28a- ShTdT	pET28a vector, *Nde*I-ShTd -*Xho*I	This study
pET28a- MbTdT	pET28a vector, *Nde*I-MbTdT-*Xho*I	This study
pET28a- HgTdT	pET28a vector, *Nde*I-HgTdT-*Xho*I	This study
pET28a- MmTDT	pET28a vector, *Nde*I-MmTDT-*Xho*I	This study
pET28a- PtTdT	pET28a vector, *Nde*I-PtTdT-*Xho*I	This study
pET28a- PmTdT	pET28a vector, *Nde*I-PmTdT-*Xho*I	This study
pET28a- DnTdT	pET28a vector, *Nde*I-DnTdT-*Xho*I	This study
pET28a- GgTdT	pET28a vector, *Nde*I-GgTdT-*Xho*I	This study
pET28a- ZaTdT	pET28a vector, *Nde*I-ZaTdT-*Xho*I	This study
pET28a- FpTdT	pET28a vector, *Nde*I-FpTdT-*Xho*I	This study
pET28a- AcTdT	pET28a vector, *Nde*I-AcTdT-*Xho*I	This study
pET28a- R335A	pET28a vector, *Nde*I-ZaTdT^{R335A}-*Xho*I	This study
pET28a- R335L	pET28a vector, *Nde*I-ZaTdT^{R335L}-*Xho*I	This study
pET28a- R335L-K337G	pET28a vector, *Nde*I-ZaTdT$^{R335L-K337G}$-*Xho*I	This study
pET28a- R335L-K337A	pET28a vector, *Nde*I-ZaTdT$^{R335L-K337A}$-*Xho*I	This study
pET28a- R335L-K337V	pET28a vector, *Nde*I-ZaTdT$^{R335L-K337V}$-*Xho*I	This study
pET28a- R335L-K337L	pET28a vector, *Nde*I-ZaTdT$^{R335L-K337L}$-*Xho*I	This study

Supplementary Table 7. Primers used in this study

Primer Name	Primer Sequence (5'→3')
Primer for functional verification or NGS sequencing	
P1	5'-TAATACGACTCACTA-3'
P1-AA	5'-TAATACGACTCACTAA-3'
P1-AT	5'-TAATACGACTCACTAT-3'
P1-AG	5'-TAATACGACTCACTAG-3'
P1-AC	5'-TAATACGACTCACTAC-3'
P1-TA	5'-TAATACGACTCACTTA-3'
P1-TT	5'-TAATACGACTCACTTT-3'
P1-TG	5'-TAATACGACTCACTTG-3'
P1-TC	5'-TAATACGACTCACTTC-3'
P1-GA	5'-TAATACGACTCACTGA-3'
P1-GT	5'-TAATACGACTCACTGT-3'
P1-GG	5'-TAATACGACTCACTGG-3'
P1-GC	5'-TAATACGACTCACTGC-3'
P1-CA	5'-TAATACGACTCACTCA-3'
P1-CT	5'-TAATACGACTCACTCT-3'
P1-CG	5'-TAATACGACTCACTCG-3'
P1-CC	5'-TAATACGACTCACTCC-3'
P2	5'-biotin-ACTAGGACGACTCGAATT-3'
C1	5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTT-3'
P5	5'-GCTCTTCCGATCTAGTGCTACTAGGACGACTCGAATT-3'
P7	CAAGCAGAAGACGGCATACGAGATCTATCGCTGTGACTGGAGTTCAGACGTG-3'
Cloning construction primers for Protein engineering	
T330-G331-Forward	5'-ATGCACTGGTTACCATTVHABDTGGTTTTCGTCGTGGT-3'
T330-G331-Reverse	5'-AATGGTAACCAGTGCATCCGGCAGAAAGGTACA-3'
G332-R335-Forward	5'-GTTACCATTACCGGTBDTTTTCGTVVTGGTAAAAACATTGGTCA-3'
G332-R335-Reverse	5'-ACCGGTAATGGTAACCAGTGCATCCGGCAGAAAGGT-3'
G452-W453-Forward	5'-CCGTATGCACTGTTABDTDBGACCGGTAGCCGTCAGT-3'
G452-W453-Reverse	5'-TAACAGTGCATACGGATACTGTTCAAACGGGGTAA-3'
T454-G455-Forward	5'-TATGCACTGTTAGGTTGGVHABDTAGCCGTCAGTTTGG-3'
T454-G455-Reverse	5'-CCAACCTAACAGTGCATACGGATACTGTTCAAACGG-3'
S456-R457-Forward	5'-CTGTTAGGTTGGACCGGTDBGVVTCAGTTTGGTCGTGATCT-3'
S456-R457-Reverse	5'-ACCGGTCCAACCTAACAGTGCATACGGATACTGTTCAA-3'
Q458-G460-Forward	5'-TAGGTTGGACCGGTAGCCGTVHATTTBDTCGTGATCTGCGT-3'
Q458-G460-Reverse	5'-ACGGCTACCGGTCCAACCTAACAGTGCATACGGATA-3'
R461-D462-Forward	5'-CCGGTAGCCGTCAGTTTGGTBDTVVTCTGCGTCGTTATGCA-3'
R461-D462-Reverse	5'-ACCAAACTGACGGCTACCGGTCCAACCTAACA-3'
R464-N477-Forward	5'-CAGCCCATGAACGTAAAATGATTCTGGATDDTCATGGCCTGTATGATC-3'
R464-N477-Reverse	5'-CATTTTACGTTCATGGGCTGCATAACGABBCAGATCACGACCAAACTG-3'
Cloning construction primers for Rational design	

R335L-Forward	5’-CCATTACCGGTGGTTTTCGTCTGGGTAAAAACATTGGTC-3’
R335L-Reverse	5’-ACGAAAACCACCGGTAATGGTAACCAGTGCATCCGGCA-3’
R335L-K337G-Forward	5’-ACCGGTGGTTTTCGTCTGGGTCCTAACATTGGTCACGATATC-3’
R335L-K337G-Reverse	5’-ACGAAAACCACCGGTAATGGTAACCAGTGCATCCGGCA-3’
R335L-K337A-Forward	5’-ACCGGTGGTTTTCGTCTGGGTGCAAACATTGGTCACGATATC-3’
R335L-K337A-Reverse	5’-ACGAAAACCACCGGTAATGGTAACCAGTGCATCCGGCA-3’
R335L-K337V-Forward	5’-ACCGGTGGTTTTCGTCTGGGTGTTAACATTGGTCACGATATC-3’
R335L-K337V-Reverse	5’-ACGAAAACCACCGGTAATGGTAACCAGTGCATCCGGCA-3’
R335L-K337L-Forward	5’-ACCGGTGGTTTTCGTCTGGGTCTGAACATTGGTCACGATATC-3’
R335L-K337L-Reverse	5’-ACGAAAACCACCGGTAATGGTAACCAGTGCATCCGGCA-3’

448