Semi-rational evolution of a recombinant DNA polymerase for modified nucleotide incorporation efficiency

Lili Zhai[1,2], Zi Wang[1], Liu Fen[4], Chongjun Xu[3], Jingjing Wang[3], Hongyan Han[1], Qingqing Xie[1,*], Wenwei Zhang[1*], Yue Zheng[1,*], Alexander K. Buell[2,*], Yuliang Dong[1,*]

1. BGI-Shenzhen, Beishan Industrial Zone, Shenzhen 518083, China

2. Department of Biotechnology and Biomedicine, Technical University of Denmark, Kgs. Lyngby 2800, Denmark

3. MGI, BGI-Shenzhen, Shenzhen 518083, China

4. MGI, BGI-Wuhan, Wuhan 430030, China

*Corresponding author

Abstract

Engineering B family DNA polymerases to utilize artificial nucleotides with unique large group labeling is limited by a general lack of understanding about the mechanism of the structural determinants that define substrate specificity. In this study, we report the engineering process of a natural DNA polymerase (KOD pol) from *Thermococcus Kodakaraenis* by using semi-rational design strategies. The essential residues around the active pocket and DNA binding site of KOD pol are characterized by using an established high-throughput microwell-based screening method. We obtained a five-mutant variant by performing site-directed saturation mutagenesis and combinatorial mutations of motif A and position 485, including two previously reported mutation sites (141/143) with inactive exonuclease activity. Then we selected more mutants, which were expected to increase the catalytic activity based on computational simulations, to perform experimental verification. By stepwise combinatorial mutation, we obtained an eleven-mutation variant E10 with over 25 times improvement of kinetic efficiency

which has similar and satisfactory performances on both the BGI NGS platform (SE50) and MGI CoolMPS™ platform (PE100). The results indicated that E10 is a potential candidate for commercialization. The beneficial mutation sites identified in this study might be shared with other archaeal B-family DNA polymerases, which can provide a guide for the mutagenesis of other types of enzymes.

Key Words

DNA polymerases, Enzyme engineering, FRET, Modified nucleotides, MM-GBSA

Introduction

DNA polymerases that are discovered in nature play important roles in molecular biology, synthetic biology, and molecular diagnostics[1]. DNA polymerases can be divided into six families (family A, B, C, D, X, and Y) by phylogenetic analysis based on amino acid sequence similarity[2,3,1,4]. Wilde-type DNA polymerases mostly fail or perform poorly in practical applications; however, these shortcomings can be overcome by enzyme engineering[3]. The application of polymerase engineering to enhance the performance of WT enzymes has been widely reported, such as the engineering of DNA polymerase was used as a research tool for DNA sequencing using acyclic nucleotide analogs in the field off bio-catalysis[4]. Engineered DNA polymerases capable of incorporating modified nucleic acids into DNA support emerging and foundational technologies[5,6].

Many sequencing methods rely on engineered DNA polymerases that efficiently incorporate modified nucleotides[7,8], such as the sequencing-by-synthesis (SBS) method, which uses dye-labeled reversible terminators in an iterative loop[5]; Sanger sequencing methods rely on dideoxynucleoside triphosphates (ddNTPs) and efficient integration of engineered DNA polymerases. In high-throughput sequencing (NGS), the addition of chemically modified nucleotide substrates (such as 3'-O-azidomethyl-dNTPs) relies on

catalysis by archaeal B family DNA polymerases[9,10]. Archaeal B family polymerases accept a broad range of chemically modified nucleotide substrates[11]. Crystal structures were solved showing the two important states of B-family DNA polymerases - the monomeric open state and the ternary ligand-bound closed state, but the mechanism of the transition between these two states remains unclear[12].

KOD DNA polymerase (KOD pol) is an important representative of B family DNA polymerases[24,25]. Since the discovery and isolation of strain KOD1 by Morikawa et al. in 1994[26], an abundant number of related strains and their enzyme products were examined and reported from this archaeon[27,28]. Recombinantly produced KOD DNA polymerases are the key component of numerous biotechnological applications such as PCR[29], RT-qPCR[30], multiplex PCR[31], incorporation of base/sugar modified nucleoside triphosphates[32], and TNA synthesis[33,34,35] etc. Many crystal structures of KOD have been presented, like monomer[36], binary complexe[37], ternary complexe[38], etc. However, there is currently no study reported that demonstrates the directed evolution of a recombinant KOD pol for modified nucleotide incorporation efficiency by deep computational scanning of structure and experimental screening.

Enzyme engineering usually includes computational design components[13], semi-rational design, and large-scale screening by directed evolution[14]. Enzyme screening strategies can include DNA shuffling[15], CSR (Compartmentalized Self-Replication)[16], microfluidic screening technology[17], microplate reader screening technology[18], etc. Most screening techniques will utilize fluorescence as a readout because of its high sensitivity, time efficiency, the potential for high throughput, low cost, and other advantages[19]. In the past 40 years, the application of Förster resonance energy transfer (FRET) has been greatly expanded[20], and the phenomenon is widely exploited in various fields, such as the structural elucidation of biomolecules and their interactions, nucleic acid analysis, and others[21]. FRET is the process in which an excited-state donor (usually a fluorophore) transfers energy to a proximal ground-state acceptor through long-range dipole-dipole interactions[22]. FRET-based sensing systems are designed as

donor-acceptor fluorophore pairs covalently linked through a spacer[23].

In this study, we established a FRET-based polymerase screening platform to screen mutants for improved abilities to incorporate modified nucleotides into DNA using the signal of FRET-cy5 as an indicator. Based on the available structure of KOD pol, we mutated amino acids near the active site pocket and combined mutants of the dominant sites into new polymerase variants. We used all-atom molecular dynamics simulations to predict key amino acid residues based on the most recent crystal structure of KOD pol, all of which are in positions relevant for DNA template strand binding. We also used molecular dynamics simulations and the Molecular Mechanics Generalized Born Surface Area (MM-GBSA) method to calculate binding free energies to identify and select key amino acid residues. We designed and screened variants with mutations in the catalytic pocket and substrate-binding domain of KOD pol and screened for variants with high efficiency. The best-performing variant was evaluated for its potential application on the BGI and MGI sequencing platforms.

Results

Selection of mutation sites for screening

In order to engineer KOD DNA polymerase with an increased incorporation ability for modified nucleotides, we first investigated the structure of KOD pol and selected target residues for mutations. Several crystal structures of KOD pol as binary or ternary complex, bound to primer-template and nucleotide were available, including a closed state (PDB ID:5OMF)[11] and an open state (PDB ID: 4K8Z)[39]. By superimposing the structures of these two states, we identified a significant move of the finger domain from the open state to the closed state by approximately 24°, as shown in **Figure 1 a**. This structural displacement is consistent with Kropp et al.'s studies[11], indicating that the active pocket was formed by the palm and closed finger domain. Especially, residues L408, Y409, and P410, called Motif A, located in the palm domain, face the

incoming dATP as shown in **Figure 1 c**. This Motif A of a DNA polymerase is associated with an increased efficiency of artificial nucleotide incorporation as previously reported[40,60].

Residue A485 located on the α-helix finger domain facing away from the incoming nucleotide in the active site is an important site (**Figure 1 b**). This position was known from the literature to be critical for defining the ability of B family polymerases to handle bases and modified nucleotides[30-33]. For example, Nikoomanzar *et al.* mapped the finger subdomain of KOD pol and found that polymerase specificity may be governed by a small number of highly specific residues such as the 485 position[17]. In addition, during the engineering of B-family polymerases to accept modified nucleotides, exonuclease activity is often inactivated to prevent the excision of the incorporated modified nucleotides[36]. Hoshino *et al*. described that a KOD mutant, exo$^-$ (D141A/E143A)/A485L, had the ability to generate functional nucleic acid molecules by incorporating dN^{am}TP continuously over 50 nt[6]. We were interested in exploring the structural determinants that control the specificity of polymerase so that KOD pol can be engineered in the direction of modified nucleotide incorporation. Structural analysis suggests that motif A and 485 are closely associated with increased efficiency of artificial nucleotide incorporation[11,22,5], which is supported by previous reports[41-42].

These published results provided an ideal starting point for our study, and residues L408, Y409, P410, and A485 were therefore selected for mutagenesis, also including mutational sites 141A/143A with exonuclease deficiency. Although these sites have been reported to associate with catalytic activity, most previous investigations consisted in performing several site-directed mutations and lacked a comprehensive analysis. The present study contains a systematic analysis of these sites by using saturation mutations and combination mutations. We report some new individual and combinatorial mutations that have not been reported previously.

Mutant screening assay establishment

We established a high throughput variant screening assay based on FRET, as shown in **Figure 2**. The screening strategy was inspired by a four-color sequencing with fluorescent nucleotides reported by Seo *et al.*[43] and the reports by Palla *et al.*, of the use of 3'- O- azidomethyl nucleotide reversible terminators in DNA sequencing[44]. The 3'-O-azidomethyl-dATP with dye Cy3 labeled (modified dATP) was used for KOD pol variant screening, according to the molecular structure in **Figure 2 d.** The polymerase activity assay is designed based on the FRET-Cy5 fluorescence signal produced by the incorporation of Cy3-labelled modified nucleotide into the PT-2Cy5, as illustrated in **Figure 2 e**. High-throughput expression of KOD pol variants was carried out in a 96-well plate to give enough proteins for screening. All the cells were lysed, followed by measurement of the protein concentrations using a microplate-based protein quantification assay[45,46]. Then, KOD pol variants proteins were added to the reaction system prepared in a 384 - well plate in advance, and the enzyme activity was measured in a microplate reader. We verified the feasibility of this screening method, and evaluated the quantification of crude protein concentration, as shown in the **SI,** Δ RFUs presents an increased fluorescence signal and Δ RFUs/min shows an increased fluorescence signal in one minute.

Saturation mutagenesis of Motif A and 485

The KOD pol exo$^-$ variant (D141A/E143A) was named as Mut_0. We first constructed the variant Mut_1 (D141A/E143A, L408A/Y409A/P410A), and measured if Mut_1 could incorporate the modified dATP as investigated by the FRET signal. Gardner *et al.* reported that mutations at position 485 likely acted indirectly on nucleotide incorporation by reducing steric barriers for nucleotide analog binding or facilitating a conformational change during the polymerization process[5]. This is also a main consideration for us to design Mut_1 in the first stage. We were encouraged by the

finding that Mut_1 displayed an increase in fluorescence signal presented by 703 Δ RFUs at a rate of 5.9 Δ RFUs/min. Thus, we used Mut_1 as the positive control in the variant screening. We performed saturation mutations at position 408 with mutated 409A/410A, primarily considering that a combinatorial library of alanine substitutions at key positions in proteins allows rapid identification of residues important for protein function, stability, and structure. Each alanine substitution examines the contribution of a single amino acid side chain to protein function[60-47]. Following this, we used alanine scanning mutagenesis to screen motif A and 485 of KOD pol. Thus, we constructed the first library of 60 saturating and combinatorial mutants of 408/409/410 including N/A/A, A/N/A, and A/A/N, and the second library of 20 saturating mutants of position 485 in the case of mutated 408A/409A/410A.

The results of the variant screening are shown **Table 1**, where the inactive mutants are omitted. For position 408, we found 7 mutations (Ile/Tyr/Met/Pro/Val/Ser/Cys) that presented a catalytic ability for modified dATP incorporation, whereby Ile and Tyr showed the best performance. On the other hand, there were only two mutations at position 409, Ala and Gly, that exhibited enzymatic activity. At position 410, there were 9 active mutations identified, Gly/Met/Leu/Pro/Val/Ser/Cys/Ile, of which P and I displayed the best performance, interestingly, Pro was the original (wild type) residue. Brown *et al*. pointed out that dATP adopts the C3' endo conformation and the sugar moiety interacts with Y409, where the C2' atom is closest to the aromatic ring of Y409, which explains the distinction between ribonucleotides[48]. Our results suggested that reducing the size of the side chains at these steric gate positions of the active pocket enables the DNA polymerase to incorporate modified dNTPs. Brown and Suo had previously studied these steric gate amino acids in detail in Family A and B DNA polymerases[48]. In 9 °N DNA polymerase, which belongs to Family B DNA polymerase, the conserved steric gate residue is Y409. Gardner's studies described that when Y409 is mutated to a smaller amino acid such as valine, up to five successive ribonucleotides can be incorporated[49]. Our results suggested the steric gatekeeper amino acid of KOD

pol might also be position 409 where the catalytically active mutations were only alanine and glycine, the amino acids with the smallest side chains.

Table 1. Screening results of the first set of KOD pol variants based on changes in motif A.

Mutants	A*A*A	Δ RFUs	V(ΔRFUs/min)
408	I	653	7.3
	Y	573	7.0
	M	50	1.4
	P	386	3.4
	V	487	6.1
	S	265	3.6
	C	271	4.2
409	A	703	5.9
	G	324	3.2
410	G	662	7.1
	M	196	2.3
	L	786	9.5
	P	1321	17.5
	V	553	5.6
	S	228	2.3
	C	714	7.5
	I	1033	13.2

Δ RFUs: Fluorescence intensity increase in 2 h. V(ΔRFUs/min): Fluorescence intensity increases per min. All experiments were performed in duplicates.

We carried out saturation mutagenesis at position 485 and measured the enzymatic activity of the 20 variants, and the results are shown in **Figure 3 a**. We obtained 18 variants with positive enzyme activity. Five mutants featured a fluorescence signal (Δ RFU) value above 3000, including Phe/Tyr/Leu/lys/Glu, in which Glu showed the best enzymatic activity. The screening results were consistent with the results of previous studies, and 485L is well-established as an important mutation for nucleotide analog discrimination[29-31]. In our study, we found 485E to have a similar and slightly better performance than 485L in modified dATP incorporation. The possible explanation is the KOD mutant (D141A/E143A/L408A/Y409A/A485E) in our study has two more

mutations in motif A in comparison with the previously reported polymerase variants, such as a KOD variant (D141A/E143A/A485L) having ability to generate functional nucleic acid molecules[6] and a 9 °N variant (exo- (D141A/E143A)/A485L) utilizing modified nucleotides[5]. Mutated 408A and 409A generate more space at the active center, but it is not clear whether there is a synergy between 485E and the other two sites for the incorporation of modified nucleotides, thus, the molecular mechanism needs to be explored in further experiments.

Combined mutagenesis of residues in Motif A and of 485

Based on the screening results of motif A and position 485, we finally chose 408 Y/ I, 409 A, 410 P/I, and 485 E/K/L to generate and study combination variants. We constructed and evaluated 12 variants that combined mutations of motif A and position 485, of which 11 variants increased the crude enzyme activity. Mut_C4 (IAP, L) and Mut_C2 (IAP, E) both showed high fluorescence signals with ΔRFU values above 4000 and Mut_C2 presented the highest enzyme activity with a rate of 186.7 ΔRFUs/min. These screening results are shown in **Table 2** and **Figure 3 b**. Interestingly, Mut_C11 showed no catalytic activity for modified dATP. Compared to Mut_C12, there was only one position changed in Mut_C11, residue 408 which had I mutated to Y. The results demonstrated that residue 485 and motif A cooperate with each other in the process of recognizing and polymerizing nucleotides. Apparently, the combination of 408Y and 485K is detrimental to the polymerization reaction, perhaps because of an increase in steric hindrance for finger movement interfering with substrate recognition caused by the benzene ring structure of 408Y.

Table 2. Combination mutation nomenclature and overview

Mutants	Motif A	485
Mut_C1	YAP	E
Mut_C2	IAP	E
Mut_C3	YAP	L
Mut_C4	IAP	L

Mut_C5	YAP	K
Mut_C6	IAP	K
Mut_C7	YAI	E
Mut_C8	IAI	E
Mut_C9	YAI	L
Mut_C10	IAI	L
Mut_C11	YAI	K
Mut_C12	IAI	K

Computational screening and experimental verification

Movements of the KOD subdomains including palm, finger, and thumb depend on the binding of an incoming dNTP. Thus, these domains were identified to contribute to polymerase fidelity because of different conformational dynamics induced by normal dNTPs and modified dNTPs. In addition, the fidelity of nucleotide addition was determined mostly during two general steps of the polymerase kinetic pathway: the binding and the actual incorporation. Previous studies reported that the binding process between enzyme and dsDNA with the correct dNTP pairing took longer than the chemical incorporation process[50]. Raper *et al.* considered that the nucleotide binding and incorporation must be much faster than the binding equilibration of a polymerase and DNA ($E + DNA \rightleftharpoons E{\cdot}DNA$)[51]. They also showed that the binding of a DNA substrate by a polymerase was frequently accompanied by conformational dynamics of the polymerase as well as of the dNTPs. After we selected and evaluated key sites (motif A and 485) affecting the chemical reaction step, we wanted to focus on the binding of polymerase and DNA which might be the overall rate-determining step of the DNA extension process. Identification of the key residues for this binding process would be conducive to the targeted improvement of polymerase catalytic activity.

In order to study the effects of residues interacting with DNA primer/template, we began by computationally designing a scanning mutagenesis library by calculating binding free energies for 93 residues located in the DNA binding region within the distance of 4 Å. We simulated the saturation mutagenesis of 93 residues in five rounds

by the homology modeling method and calculated the change of binding free energy between the mutants and DNA primer/template with the MM-GBSA method, the whole simulated and calculation workflow is presented in **SI Figure 1 and SI Figure 2**. Based on Mut_C2 which was the best-performing variant, we carried out full saturation mutagenesis of the 93 residues by using the Modeler software and then processed all 1860 mutants uniformly. Due to the large number of calculations involved, we designed an automatic processing pipeline in Python, and the whole process included mutant pretreatment, molecular dynamics simulation, and binding energy calculation, as shown in **SI Figure 1**. The average binding energy value of 93 residues, obtained from the MM-GBSA method, is shown in **SI Table 1**. We expressed and experimentally tested several enzyme variants with an average computed binding energy < -250 kcal/mol that was lower than the wild-type value and included also some interesting positions identified in a previous study by Kropp et al.[11], namely residues 349, 383, 389,541,592, and 606. We successfully expressed 52 variants to measure their catalytic efficiency with kinetic data shown in **SI Table 2**. The variants that failed to express are also displayed in SI. The normalized binding free energy of 52 variants containing 26 distinct mutation sites was plotted as a heat map in **Figure 4**. The more negative the free energy values, the stronger the binding affinity. The two variants with the most negative binding free energy were expressed. Expressed proteins were quality controlled by SDS-PAG gel shown in **SI Figure 3**, and the protein concentration was measured by using the Bradford assay.

Mut_C2 was used as the positive control variant in this round of screening. We first tested the enzyme activity (RFUs/min) of these mutants and calculated the relative value of V which was the enzyme activity compared to that of Mut_C2. To further explore the catalytic properties of these enhanced KOD mutants, we measured the Michaelis−Menten kinetics for modified dATP incorporation under increasing concentrations of DNA primer/template. The Michaelis−Menten kinetics towards modified dATP incorporation under increasing concentrations of modified dATP were

also measured, and the data is shown in **SI Table 3**. The variants that had a relative V value > 0.5 were selected for kinetic measurements. K_m and V_{max}/K_m of these combined mutants were measured and calculated and are reported in **Table 3**, and all values were also expressed as relative V/K compared to Mut_C2. The V_{max}/K_m value represented the converted value of K_{cat}/K_m. K_{cat} corresponds to V_{max} divided by the enzyme concentration, as the enzyme concentrations of these mutants were the same in the kinetic assays. Thus, the relative catalytic efficiency of KOD variants is characterized by the ratio of K_{cat}/K_m_(mut) equal to V_{max}/K_m_(mut) compared to that of the Mut_C2, as equation 1:

$$\text{Relative } V/K(mut) = (V_{max}/K_m_(mut))/(V_{max}/K_m_(Mut_C2) \qquad \text{(eq. 1)}$$

The relative V/K(mut) represents the catalytic efficiency of KOD mutants; the higher the value of the relative V/K, the better the performance of modified dATP incorporation.

Table 3. Information on the KOD pol variants generated in the second round of variant screening.

Mutants	Sites	Subdomain	Relative V	K_m (μM)	V_{max} (RFUs/min)	V_{max}/K_m (RFUs/min*μM)	Relative V/K
Mut_C2	-	-	1.0	1.6	596.0	372.5	1.0
Mut_D34	V589H	Palm	2.6	1.0	1943.0	1926.3	5.2
Mut_D35	V589Q	Palm	1.8	1.2	1352.9	1124.9	3.0
Mut_D51	V680M	Thumb	0.5	0.5	350.5	753.9	2.0
Mut_D48	T676K	Thumb	1.2	0.9	536.4	596.0	1.6
Mut_D30	T590K	Palm	1.1	0.9	357.6	411.3	1.1
Mut_D33	V389M	Palm	1.1	0.8	326.5	408.1	1.1
Mut_D36	Y384F	Palm	0.8	2.0	800.6	398.7	1.1
Mut_D40	S383T	Palm	1.1	1.8	676.8	378.7	1.0
Mut_D17	T349I	N-Terminal	1.1	1.4	498.7	368.4	1.0
Mut_D3	Y496I	Finger	0.7	4.4	1579.4	357.2	1.0
Mut_D32	V389I	Palm	0.9	1.3	463.6	357.1	1.0

Notes: Relative V is defined as the relative enzyme activity, and compared with Mut_C2, the value of Mut_C2 is defined as 1.

Relative V/K is defined as the relative enzyme catalytic efficiency, and compared with Mut_C2, the value of Mut_C2 is defined as 1.

Km (μM) presented the affinity between the enzyme and the substrate. Vmax/Km (RFUs/min) values represent the catalytic efficiency of the variants.

The results indicated that if the valine (V) of position 589 located on the palm domain was mutated to histidine (H), the relative V/K value modestly increased about ~5.2-fold. As H is an amino acid with a positively charged side chain, we speculated that the 589H mutation increased the binding affinity of KOD pol to the DNA primer/template, which facilitates the incorporation process, as the double-stranded DNA is strongly negatively charged. 589Q with no charged side chain displayed a lower relative V/K value compared to 589H, but the kinetic activity was still improved about ~3.0-fold compared to V589. There were five mutation sites (589/590/389/385/384) located in the palm domain; two mutation sites (676/680) located in the thumb domain; three mutation sites (496/488/489) located in the finger domain and position 349 located in the N-terminal region. The relative V/K values of 680M and 676K were found to be improved 2 times and 1.6 times compared to Mut_C2. Overall, there were 9 mutants with a relative V/K ≥ 1.0 compared to Mut_C2, and the other values lower than 1.0 are listed in the SI. Kropp *et al*. described that KOD pol's structure displays a long crack with an electro-positive potential, extending from the DNA binding region at the thumb domain upwards along the β-hairpin and the palm domain to the N-terminal domain[11]. Of the mutation sites investigated here, 383T/384F/389M were located at the loop of the KOD pol structure close to the DNA binding region and the nucleobase maintained a stable region, and 589H/590K/676K/680M contributed to the binding between KOD pol and DNA primer/template. These results suggest that crucial sites in these regions were identified that influence enzyme activity and that effective mutations of these sites enhance the affinity of polymerase and DNA for each other, which in turn accelerates the rate of polymerase binding to DNA. We were interested in possible interactions and synergies between these identified sites, and therefore we proceeded to generate

combination variants.

Stepwise combination of effective mutations

Based on the improved performance of the variant Mut_D34, we combined it with other promising mutation sites and expected to achieve further improvements in catalytic ability, shown in Table 4. We first combined 680M (Mut_D51) with the mutations of the mut_D34 variant, and the resulting combination variant Mut_E1 displayed a slight improvement in relative V/K, confirming that the 680M mutation stacked on mut_D34 improved catalytic performance. We further combined the result with 676K (Mut_D48), which added a positive charge located on the thumb domain, and found that these three mutation sites displayed positive interactions for improving polymerase catalytic efficiency, and we obtained an improved variant Mut_E2. We separately combined Mut_E2 with further single site changes, 590K, 389M, 496L, 349I, 385M, and 384F. Four of these mutants (349I, 384F, 496L, and 385M) increased relative V/K values 2 times compared to E1. Especially 384F and 389M showed good catalytic performance, and 383T is in close spatial proximity to these two sites. Thus, we generated two-point and three-point superpositions of 383F, 384T, and 389M, mut_E9, and Mut_E10, which were found to have improved relative V/K. Following that, we tried to combine the 349Y and 496L mutations with Mut_E10, respectively. Unfortunately, we did not obtain an improved variant from E10 in this manner, and the results also showed that the addition of these two mutations had a negative effect on the catalytic activity. These results suggested that the best-performing variant identified in this screen was Mut_E10.

Table 4. Information on the KOD pol variants generated in the final round of variant screening.

Mutant	Mutation Sites	K_m	V_{max}	V_{max}/K_m	Relative V/K
Mut_D34	V589H	1.3	968.9	745.3	1.00
Mut_E1	V589H+680M	1.6	1336.2	835.1	1.1
Mut_E2	V589H+ T676K+ V680M	1.5	1991.6	1327.7	1.8
Mut_E3	V589H+ T349F + T676K+ V680M	0.5	773.9	1547.9	2.1

Mut_E4	V589H+ T676K+ V680M+ Y384F	0.7	1389.1	1984.4	2.7
Mut_E5	V589H+ T676K+ V680M+ Y496L	0.4	709.1	1818.2	2.4
Mut_E6	E385M +V589H+ T676K+ V680M	3.9	1389.1	356.2	0.5
Mut_E7	T590K +V589H+ T676K+ V680M	2.5	664.1	265.7	0.4
Mut_E8	V389I +V589H+ T676K+ V680M	0.4	812.	2030.2	2.7
Mut_E9	V589H+T676K+V680M+S383T+Y384F	0.9	2059.2	2214.1	3.0
Mut_E10	V389I+ V589H+T676K+V680M+S383T+Y384F	0.5	1799.2	3331.9	4.5
Mut_E11	V389I+ V589H+T676K+V680M+S383T+Y384F+Y349F	1.8	1098.1	610.0	0.8
Mut_E12	V389I+ V589H+T676K+V680M+S383T+Y384F+Y496L	1.7	717.1	421.8	0.6

Notes: Relative V/K of mut_D34 was defined as 1 in this round of screening, and the variants' V/K was expressed relative to that of mut_D34.

The whole evolution process that ultimately led to Mut_E10 is illustrated in **Figure 5**. The wild-type KOD pol is not able to add modified dATP. With Alanine screening in 408/409/410, KOD pol mutants could incorporate modified dATP successfully, but the reaction rate was very slow and the observed increase in FRET fluorescence intensity was low. After combining 408I/409A/410P with mutations at site 485E, the ability of the variant Mut_C2 mutant to add modified dATP was significantly enhanced according to the measured reaction rate (186.7 RFUs/min) and the increase in fluorescence (4075.5 RFUs). 589H played an important role in promoting the polymerization of modified dATP by combining it with the mutations of Mut_C2, and the relative kinetic parameter V_{max}/K_m increased by more than 5 times. Based on mut_D34, we carried out four rounds of combinational mutation screening and finally obtained the variant E10 with a relative increase in V/K of more than 20 - fold. These results demonstrate that combinatorial mutations of multiple amino acids lead to efficient interactions with the substrates, which were beneficial to improve the catalytic efficiency of KOD pol towards incorporating the modified nucleotide. We next set out to probe the performance of mut_10 in NGS, and we applied mut_10 on two different sequencing platforms.

Application of the best-performing combination variant in NGS sequencing

Test using the BGISEQ-500 sequencer

Since the launch of the Human Genome Project, next-generation sequencing (NGS) technology has had a huge impact on biological research[52]. The BGISEQ-500 sequencer is a widely used NGS platform[53], so we tested the dominant mutant E10 on the BGISEQ-500 sequence platform and analyzed the sequencing results. The library used for sequencing was E. coli, the number of cycles was SE 50, the temperature was 55°C, and the sequencing process was carried out by the BGI sequencing group. The Commercial Sequencing Enzyme (CSE) in the original system was replaced by Mut_E10, and the sequencing results of CSE and Mut_E10 were compared. The sequence quality of the enzyme-replaced platform was satisfactory and acceptable, and the results are summarized in **Table 5**. The mapped reads and mapping rate of Mut_E10 were similar and slightly higher than the values obtained with CSE. Q30 % of CSE was > 94 although Q30 % of Mut_E10 was slightly lower, the value was still > 93 which was higher than the standard values within the available range in NGS. Lag %, run on %, and ESR% present the polymerization ability of the sequencing enzyme, respectively, and E10 performance was similar to that of CSE. Our results suggest that Mut_E10 was a potential candidate for NGS sequencing, and we also wondered if Mut_E10 could display satisfying performance on the novel autonomous sequencing developed by MGI. Therefore, we evaluated Mut_E10 on an additional sequencing platform, CoolMPS™ sequencing.

Table 5. Sequencing data of Mut_E10 on the BGISEQ-500 sequencer.

BGISEQ-500	L01-CSE	L02- CSE	L01-E10	L02- E10
Reference	Eco.li	Eco.li	Eco.li	Eco.li
Cycles	50	50	50	50
Mapped Reads(M)	198.64	197.78	202.78	200.85
Mapping Rate%	99.5	99.5	99.63	99.53

Q30%	94.41	94.43	93.88	93.76
Lag%	0.11	0.11	0.12	0.12
Runon%	0.05	0.05	0.04	0.04
ESR%	91.45	91.31	92.42	92.11

Test in CoolMPS^{TM} platform using MGISEQ2000

CoolMPS^{TM} is a novel massively parallel sequencing chemistry utilizing cold dNTP (modified) and four natural nucleobase-specific fluorescently labeled antibodies with fast binding, which was developed by the Drmanac group at MGI[54]. We applied Mut_E10 in the CoolMPS^{TM} platform using the MGISEQ-2000 sequencer. The library used for sequencing was E. coli, the number of cycles was PE100, the temperature was 55 °C, and the sequencing process was carried out following the published protocols[54]. In **Figure 6 a**, it can be seen that the change in Q30 was over 95 % in the 100 cycles of the first strand, then a slow decrease in Q30 on the second strand was observed as the complementary template and corresponding sequencing primer was used for extension. During the base-calling of sequencing, Q30 is an important indicator representing the sequencing quality[55]. The higher the quality value (Q), the lower the probability (P) of a base being erroneously detected. The mapping rates of the first strand and second strand were both >99.98, the ESR value was over 78 %, and the average error rate was lower than 0.16. Lag and runon values of the first strand and the second strand were quite low and are plotted in **Figure 6 b-c**, which presented fluorescent bases failing to keep up the rhythm in each copy of rolling circle amplification (RCA). Thus, low lag and runon values indicated that the polymerase exhibited a satisfactory polymerization activity on modified dNTPs. Korostin reported that Illumina HiSeq 2500 and MGISEQ-2000 have similar performance characteristics after comparing both whole-genome sequencing (PE150) in terms of sequencing quality, number of errors and performance[56]. Our sequencing data also presented a similar performance with that including the mapping rate (99.93 %) and Q30 (97.18 %) of HiSeq 2500, although our Q 30 was only about 96 % which can be acceptable. Overall, these results demonstrate that Mut_E10

is ready for practical application in the widespread next generation sequencing (NGS) systems.

Table 6. Sequencing data of E10 measurement on the MGISEQ-2000 sequencer.

MGISEQ-2000	Mut_E10	
Category	Value_1	Value_2
Reference	Ecoli	Ecoli
Cycle Number	210	210
ESR (%)	78,74	81,23
Q30(%)	95,97	96,42
Total Mapping Rate (%)	99,98	99,98
Error Rate 1(%)	0,16	0,12
Error Rate 2(%)	0,16	0,12
Split Rate (%)	99,24	99,23
Runon1(%)	0,04	0,04
Runon2(%)	0,12	0,08
Lag1(%)	0,06	0,06
Lag2(%)	0,13	0,14

Note: Q30: the accuracy of base calling; ESR: estimated error rate; Error Rate: base error rate; Split Rate: barcode split rate. Value_1 and Value_2 were the duplicates sequencing data obtained from Lane1 and Lane2 in the same chip.

Discussion

Mut_E10 performed satisfactorily in both the BGISEQ-500 platform and MGI CoolMPS^{TM} platform. The nucleotides used are fluorescently labeled terminating nucleotides in the BGISEQ500 platform, while in the MGI CoolMPS^{TM} platform, the sequencing was achieved by using reversibly terminated nucleotides and four natural nucleobase- specific fluorescently labeled antibodies with fast binding. The results of the present study illustrate that E10 had similar catalytic efficiency for both modified nucleotides, indicating that E10 has the potential of general adaptability for different types of substrates. However, the direct mechanism by which Mut_E10 recognized

modified nucleotides is not fully understood. Further research will be needed to elucidate how these mutations affect the polymerization efficiency of modified nucleotides by Mut_E10. As a first step towards elucidating the specific mechanism, we constructed the structure of Mut_E10 containing replaced spatial position of 3'-O-azidomethyl-dATP using homology modeling and then analyzed the possible causes and mechanisms. There were eleven mutated positions, including 141A/143A, related to inactivating the endonuclease activity, and the remaining modified sites could be divided into three categories.

First, the sites that affected the binding of polymerase with dsDNA include S383T, Y384F, V389I, V589H, T676K, and V680M. The residues' locations are shown in **Figure 7 a**. MM-GBSA calculation results suggested that modifications at these sites made Mut_E10 polymerase more capable of binding to dsDNA strands, allowing Mut_E10 polymerase to capture DNA strands faster. Another point to note was that 589H and 676K were positively charged and are likely to exhibit higher binding affinities for negatively charged DNA strands when the polymerase bound to the DNA prime/template. 383T, 384F, and 389I are located close to the DNA strands and nucleotides which were the back of the active pocket and might have a positive effect on the formation of a stable active pocket. In addition, 383T, 384F, 389I, and 676K were located on the coil structure. To prevent the mutations from affecting the original coil structure, we selected mutants that were similar to the original residue in the screening. Interestingly, the experimental results also showed that if the properties of the residue side chains at these points were greatly changed, although the binding energy calculated by MM-GBSA may be stronger, the actual test results were rather poor, and these data were not shown here. These results indicated that a stable coil or loop structure of DNA polymerase, often unappreciated, also played an important role in modified nucleotide polymerization activity. Excessive and disruptive changes to these residues on the coil structure are therefore not recommended.

The second category of sites is those where mutations changed the dNTP catalytic

pockets, including L408I and Y409A, as shown in **Figure 7 d**. The modified dATP structure used in mutant screening is shown in **Figure 7 e**. There is an azido methyl terminal group at the 3' end, and the base group is related to a long fluorescent group. When modified dATP with a blocking group entered the catalytic pocket, its blocking group would collide with the surrounding L408/Y409/P410 residue side chains (as shown in **Figure 7 b**). Comparing motif A (L408/Y409/P410) of wild-type KOD pol with motif A (408I, 409A, and 410P) of Mut_E10 revealed more space for the incoming modified dATP in the catalytic pockets as small side chains made it possible to incorporate modified dATP in the DNA primer/template. The results indicated that small side chain residues were recommended to use in the catalytic pockets when the DNA polymerase was transformed to incorporate a modified substrate with a larger group.

The third site category includes the sites that affect the conformational change of the finger domain, including A485E, as shown in **Figure 7 c**. We compared the binary complex structure of KOD polymerase that was only bound to dsDNA with the ternary complex structure of KOD polymerase that was bound to primer/template and dATP in **Figure 7 a**. Two alpha-helices of the finger domain rotated nearly 24° from the binary complex's open state to the ternary complex's close state[11]. This structural transformation might be beneficial for the fidelity of KOD polymerase. In the Mut_E10, the Ala 485 on the rotating alpha helix of the finger domain was mutated to Glu485, which would form a salt bridge (as shown in **Figure 7 c**) with Gln332 on the opposite helix of the exonuclease domain, preventing the rotation of the finger domain. For KOD polymerase, which catalyzed common dNTP, this was evidently a deleterious mutation. However, the modified dNTPs used in NGS have very long fluorescent groups. Locking the finger domain could improve the stability of the dNTP access channel so that the speed of modified dNTP in and out of the catalytic pocket could be improved. The experimental results showed that the mutation of this point significantly improved the enzyme catalytic efficiency.

Based on our knowledge, commercial enzymes currently used in sequencing-by-synthesis mainly include 9°N DNA polymerase mutants[5], Taq DNA polymerase mutants[57,58], Pfu DNA polymerase variants[59], etc. A famous 9 °N DNA polymerase variant containing three mutations (D141A, E143A, A485L) is commercialized with the name "Therminator DNA polymerase" with the ability to incorporate modified nucleotides[5]. E10 contains more mutation sites than Therminator DNA polymerase; in possible follow-up tests in the future, E10 might perform better for incorporating varieties of modified nucleotide classes, as E10 has already exhibited the ability to accept both fluorescently labeled terminating nucleotides and unlabeled nucleotides.

In this study, the semi-rational evolution strategy of KOD pol for modified nucleotide incorporation efficiency displayed a high efficiency compared to some reported methods. For example, Kennedy *et al*. reported that modulation of motif A (408/409/410) greatly shifted both the steady and pre-steady state kinetics of the fidelity of Pfu Pol which was a family B DNA polymerase from *Pyrococcus furiosus*, and their screening strategy employed site-directed mutagenesis constructing mutants and then investigating mutants by using PCR-Based lacZR forward mutation assay and phosphorimaging assays running urea denaturing gel assays[60]. In comparison with Kennedy's strategy, the advantages of our screening methods are higher screening throughput, high sensitivity, and a more straightforward analysis process. Nikoomanzar and his coworkers developed a microfluidic-based deep mutational scanning method to evolve a replicative DNA polymerase(KOD pol) isolated from *Thermococcus kodakarensis* to perform TNA synthesis[17]. They mapped the finger subdomain of KOD pol and discovered two cases of positive sites in the single round of sorting from 912 mutants, which they then combined step-wise to obtain a double mutant variant. In comparison to their screening method, our screening strategies exhibited more efficiency, we obtained Mut_E10 containing nine mutation sites, excluding the mutated 141A/143A, from no more than 150 experimentally validated mutants although the throughput of our screening platform is comparable. However, there are also some

limitations of our detection method, such that, the FRET fluorescence signal reflects the polymerization efficiency of DNA polymerase only at the level of one base extension but cannot reflect directly phosphate bond formation or a pyrophosphate molecule shedding.

In real world sequencing applications, Mut_E10 was found to have similar performance as the CSE, and not to lead to a real improvement. The possible reason for this lack of enhanced sequencing is that Mut_E10 directly was used to replace the enzyme in the original sequencing reagent without buffer optimization. Similarly, the sequencing quality achieved with Mut_E10 in CoolMPS™ chemistry is not as good as that of Illumina HiSeq 2500, but the sequencing data is still within the acceptable range of the "gold standard" of the NGS analysis generated by the Illumina platform[55]. In addition, CoolMPS™ chemistry is still under development[54], thus E10 is possible to continue to be evolved in the future to adapt to new sequencing technologies or related similar application fields.

Conclusion

One of the current trends in enzyme engineering is to design or discover new DNA polymerases with higher performance and wider substrates tolerance by using various protein evolution strategies. This study broadens the limited knowledge of KOD DNA polymerases in their interactions with modified substrates and more understanding of their ability to accept a wide range of modified nucleotides. In this work, we analyzed the crystal structure, catalytic mechanism, active pocket, and substrate-binding region of a KOD DNA polymerase belonging to archaeal B-family DNA polymerases, and generated variant libraries through semi-rational design. The dominant, best-performing variant, containing eleven mutation sites, was found to successfully accept two different types of modified substrates through screening, and these beneficial mutation sites identified could act as inspiration for the engineering and improvement

of other archaeal B-family DNA polymerases. The FRET-based screening methodology which enabled us to rapidly distinguish mutations with beneficial and detrimental effects at high throughput might also guide the screening of other types of enzymes or antibodies.

Method and Materials

Reagents and instruments

The primer and template labeled at the 5'-terminus by Cy5 dye were ordered from Invitrogen. The primer-template (PT - 2Cy5) was prepared by mixing the primer and template at a 1:1 molar ratio and annealing at 80 °C for 10 min and then cooling down to room temperature in a water bath. The sequence information of PT - 2Cy5 is shown in **Figure 8.** The primers designed with mutations were ordered from Genscript Biotech. 3'-O-azidomethyl-dATP-Cy3 (dA') was supplied by the BGI synthetic chemistry group. Potassium phosphate, $MgSO_4$, KCl, $(NH4)_2SO_4$, Tris, NaCl, HCl, EDTA, NaOH, imidazole, glycerol, lysozyme, LB broth, kanamycin, PMSF, and IPTG were purchased from Thermo Fisher. 12% SDS-PAGE gel and gel running buffer were purchased from Bio-Rad. Affinity chromatography columns (HisTrap HP 5ml column) and ion exchange columns (HiTrap Q FF, 5mL and HiTrap SP HP 5ml column) were purchased from GE Healthcare. The ÄKTA Pure Protein Purification system was purchased from GE Healthcare. The Spark Control microplate reader was purchased from TECAN.

Mutation library construction

The starting polymerase of our study was a variant of KOD pol that inactivated the 3'-5' exonuclease activity through the silencing (exo-) mutations D141A/E143A, so that the modified nucleotide could be incorporated without being subsequently removed by the 3'-5' exonuclease proofreading activity. The codon-optimized gene encoding this exonuclease-deficient KOD pol (named Mut_A0) was synthesized by Genscript

Biotech and then cloned into vector pD441 with kanamycin resistance. The corresponding expressed protein features a 6xHis tag followed by TEV (tobacco etch virus) protease cleavage site at the N-terminus. The first library of 60 saturating and combinatorial mutants of 408/409/410 included N/A/A, A/N/A, and A/A/N, and the second library was constructed included 20 mutants at position 485 with 408A/409A/410A, and the primer information was showed in **SI Table 4**. These libraries were constructed by site-directed mutagenesis technology. Site-directed mutagenesis was employed to construct the mutation library using site-directed mutagenesis kits purchased from Thermo Fisher. The PCR reaction conditions were as follows: one cycle at 98 °C for 10 s; 20 cycles at 98 °C for 10 s,72 °C for 2.5 min, followed by elongation at 72 °C for 5 min and hold at 4 °C. The PCR products were

visualized by 1.2 % agarose gel electrophoresis and ultraviolet light. Then, 1 µl of DpnI (NEB) was added to the PCR mix and incubated for 2 h at 37 °C. 5µL of the digested PCR product was transformed into the 50 µL *E. coli* DH5α competent cells (Tiangen). Recombinant plasmids were prepared by using the QIAprep Spin Miniprep Kit (QIAGEN) and confirmed by sequencing analysis. 1 µL of each correct plasmid was transformed into 50 µL *E. coli* BL21(DE3) competent cell (Tiangen) for protein expression.

Polymerase expression and purification

1 mL of LB medium in a 96-deep well plate was used for small-scale expression, and 1L of LB medium in a conical flask was used for large-scale expression. Cells grown in LB medium containing 50 µg/ml kanamycin were induced at OD600 nm = 0.6 - 0.8 by the addition of 0.5 mM IPTG. Protein production was carried out at 25 °C, 220 rpm overnight.

Semi-purification was performed using cell pellets collected from 1 mL culture in a 96-deep well plates. The cell pellets were re-suspended in Buffer 1 (20 mM Tris-HCl, 10

mM KCl, 10 mM $(NH4)_2SO_4$, 0.1% Triton, 4 mM $MgSO_4$, 1.25 mM PMSF, and 1 mg/mL lysozyme, pH 7.6) with a ratio of 0.04 g cell pellets per mL buffer. The resuspended cells were incubated at 37 °C for 10 minutes, followed by thermal denaturation at 80 °C for 30 minutes. After centrifugation at 12,000 X g for 20 minutes, the supernatant containing protein was collected. The protein concentration was measured and estimated by using the Bradford protein assay[45].

We used purified polymerase in NGS sequencing and the purification methods as the following process. For protein purification, cell pellets collected from 1 L cultivation were resuspended in Buffer 2 (500 mM NaCl, 5 % glycerol, 20 mM imidazole, 1.25 mM PMSF, 50 mM potassium phosphate, pH 7.4) and then disrupted with a high-pressure homogenizer AH-1500 purchased from ATS engineering company. After a 30-minute thermal denaturation at 80 °C, the centrifugation at 12,000 rpm 4 °C (Beckman Avanti J-26) for 30 mins was performed to remove cell debris. After centrifugation, the supernatant was filtered through a 0.22 µm membrane filter. The subsequent purification process was performed using AKTA pure 25 (GE Healthcare) with Nickel affinity chromatography (5mL HisTrap HP column, GE Healthcare), anion ion exchange chromatography (5mL HiTrap Q FF column, GE Healthcare), and cation ion exchange chromatography (5mL HiTrap SP HP column, GE Healthcare), sequentially. For the affinity purification process, the supernatant was loaded onto the pre-equilibrated column with a flow rate of 2.5 mL/min so that the retention time will be 2 min. A wash step using 50 mL Buffer 2 with a flow rate of 5mL/min was added before elution. The elution procedure was performed with a linear gradient of Buffer 3 (500 mM NaCl, 5 % glycerol, 500 mM imidazole, 50 mM potassium phosphate, pH 7.4) from 0 to 100 % in 10 CV with a flow rate of 5 mL/min. Fractions with 280 nm UV signal above 100 mAu were collected. The collected samples from Ni purification were diluted 6-fold with Buffer 4 (5% glycerol, 25 mM potassium phosphate, pH 6.6) to decrease the concentration of NaCl to 500 mM. The diluted sample was loaded onto the pre-equilibrated HiTrap Q FF 5 mL column (GE Healthcare) at a flow rate of 2.5

mL/min. The Flowthrough sample was collected and loaded onto the pre-equilibrated HiTrap SP HP 5 mL column (GE Healthcare) with a flow rate of 2.5 mL/min. The HiTrap SP HP 5 mL column was pre-equilibrated with Buffer 5 (50 mM NaCl, 5 % Glycerol, 50 mM Potassium Phosphate, pH 7.4). The elution procedure was performed with a linear gradient of Buffer 6 (1 M NaCl, 5 % Glycerol, 50 mM Potassium Phosphate, pH 6.6) from 0 to 60% in 10 CV. The first peak eluted was collected. Finally, the collected sample was dialyzed with Buffer 7 (20 mM Tris 200 mM KCl, 0.2 mM EDTA, 5 % glycerol, pH 7.4) at 4 °C overnight and then stored at -20 °C with 50 % glycerol. Protein concentration was determined by A_{280} using Spark Control microplate reader (TECAN) and calculated using the extinction coefficient of 1.393 (what are the units?) predicted by the ExPASy server. 12 % SDS-PAGE was used to analyze the purity of the protein.

Polymerase activity screening of KOD variants

The polymerization reaction mixture contained 0.1 µM PT-2Cy5, 0.25 µM Cy3-labelled modified dATP, 2.5 µM native dC/G/TTP each, 2 µg BSA, 1X KOD Screening Buffer (20 mM Tris-HCl, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 0.1% Triton, 4 mM $MgSO_4$, pH 8.8) and 1 µg KOD polymerase. The activity assay was carried out at 40 °C in black 384 - well fluorescence assay plates (Corning®). The reaction was initiated by the addition of each purified KOD polymerase variant. Wilde-type KOD pol exo$^-$ was used as a reference while KOD storage buffer was used as a negative control. The increase of FRET-CY5 fluorescence signal was measured as relative fluorescence units (RFUs) in appropriate time intervals by exciting at 530 nm and detecting the emission at 676 nm. Measurements were performed with setting gain 60 and shaking for 10 seconds before measurement in a Spark Control microplate reader (TECAN). The polymerase activity was defined as the maximum slope of FRET-Cy5 signal, expressed as ΔRFUs/min.

Enzyme Kinetic assays

The reaction mixture contained 8 μM Cy3-labelled modified dATP, 10 μM native dC/G/TTP each, 2 μg BSA, 1X KOD Screening Buffer (20mM Tris-HCl, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 0.1 % Triton, 4 mM $MgSO_4$, pH 8.8) and 0.1 μM KOD polymerase. Various concentrations of PT-2Cy5 were introduced, ranging from 0.4 μM to 4 μM. The maximum slope of Cy5-FRET- signal for each concentration of PT-2Cy5 was calculated and used for determining the kinetic parameters (V_{max} and K_m) based on analysis of the kinetic data with the "Michaelis-Menten equation". Data were processed and analyzed with the Origin software. The change in fluorescence signal, Δ RFUs generated in 1 min was defined as the enzyme activity V.

MMGBSA screening method

The complex structure of KOD pol and dsDNA adopts the 5OMF PDB structure. Additional small molecules in both structures were removed, and the systems were energy minimized in 80 mM sodium chloride aqueous solution, followed by NVT ensemble MD relaxation of 50 ns[61,62]. We used the Modeler 9.19 software to carry out saturation mutagenesis at 93 residues on the polymerase and then processed all 1860 mutants uniformly. Due to this large number of calculations, we designed an automatic processing workflow in Python. In an 80 mM sodium chloride aqueous solution, five rounds of energy minimization were first carried out for the structure generated by homologous modeling. In the five rounds of simulations, the binding coefficient to heavy atoms of protein and DNA was gradually weakened from 80 kcal / (mol · Å) to 0 kcal / (mol · Å). After that, NVT ensemble MD was carried out at 350 K, and finally, NPT ensemble MD of 10-25 ns was carried out at 310 K. We sampled the trajectory of the system that reached equilibrium in the last part and calculated the binding energy by the MM-GBSA method[63].

Evaluation of sequencing performance of the best variant in the NGS platform

BGISEQ-500 is a desktop sequencer developed for DNA or RNA sequencing by BGI. It utilizes DNA NanoBalls (DNBs) technology (DNBSEQ™). The dNTPs used in DNBSEQ™ technology were modified with different fluorescent dyes in the base and reversible blocking group in its 3'-side. To evaluate the sequencing performance of Mut_E10 in the BGISEQ-500 sequencer, a *E. coli* library prepared by following the instructions described in the manufacture's use manual of MGIEasy PCR-Free DNA Library Prep Kit (PN: 1000013452, MGI-tech) was used to make DNB following BGISEQ-500 protocol[29-53]. The DNBs were then loaded into the patterned nanoarrays. The SE50 sequencing kit was supplied by the BGI sequencing group. Mut_E10 was used to replace the Commercial Sequencing Enzyme (CSE) in the sequencing kit, and both Mut_E10 and CSE were used to sequence in different flow cells. Sequencing was performed employing the SE50 mode. Base calling and mapping were performed as previously described[30].

MGISEQ-2000RS is a high-throughput sequencer developed by MGI. The sequencing performance of Mut_E10 was also evaluated in the DNBSEQ-2000RS sequencer using CoolMPS™. CoolMPS™ is a novel DNA sequence technology using labeled antibodies developed by MGI. The dNTPs used in CoolMPS™ are modified with a reversible blocking group in 3'-side, with no modification in its base position. CoolMPS PE100 High-throughput Sequencing kit (PN: 1000016935) was purchased from MGI-tech. The same E. coli library used for BGISEQ-500 was used for CoolMPS on MGISEQ-2000RS. The sequencing process followed the instructions described in the MGISEQ-2000RS High-throughput Sequencing Set User Manual. Mut_E10 was used to replace the Commercial Sequencing Enzyme (CSE) in the CoolMPS PE100 sequencing kit. Both Mut_E10 and CSE were used to sequence the flow cells loaded with DNBs prepared with E.coli library in different runs. Base calling and mapping

steps were performed as reported previously [31-32].

Sequence:

>Seq_1(KOD mut_0)

MILDTDYITEDGKPVIRIFKKENGEFKIEYDRTFEPYFYALLKDDSAIEEVKKITAERHGTVVTVK
RVEKVQKKFLGRPVEVWKLYFTHPQDVPAIRDKIREHPAVIDIYEYDIPFAKRYLIDKGLVPMEGD
EELKMLAFAIATLYHEGEEFAEGPILMISYADEEGARVITWKNVDLPYVDVVSTEREMIKRFLRVV
KEKDPDVLITYNGDNFDFAYLKKRCEKLGINFALGRDGSEPKIQRMGDRFAVEVKGRIHFDLYPVI
RRTINLPTYTLEAVYEAVFGQPKEKVYAEEITTAWETGENLERVARYSMEDAKVTYELGKEFLPME
AQLSRLIGQSLWDVSRSSTGNLVEWFLLRKAYERNELAPNKPDEKELARRRQSYEGGYVKEPERGL
WENIVYLDFRSLYPSIIITHNVSPDTLNREGCKEYDVAPQVGHRFCKDFPGFIPSLLGDLLEERQK
IKKKMKATIDPIERKLLDYRQRAIKILANSYYGYYGYARARWYCKECAESVTAWGREYITMTIKEI
EEKYGFKVIYSDTDGFFATIPGADAETVKKKAMEFLKYINAKLPGALELEYEGFYKRGFFVTKKKY
AVIDEEGKITTRGLEIVRRDWSEIAKETQARVLEALLKDGDVEKAVRIVKEVTEKLSKYEVPPEKL
VIHEQITRDLKDYKATGPHVAVAKRLAARGVKIRPGTVISYIVLKGSGRIGDRAIPFDEFDPTKHK
YDAEYYIENQVLPAVERILRAFGYRKEDLRYQKTRQVGLSAWLKPKGT*

>Seq_5 (Mut_E10)

MILDTDYITEDGKPVIRIFKKENGEFKIEYDRTFEPYFYALLKDDSAIEEVKKITAERHGTVVTVK
RVEKVQKKFLGRPVEVWKLYFTHPQDVPAIRDKIREHPAVIDIYEYDIPFAKRYLIDKGLVPMEGD
EELKMLAFAIATLYHEGEEFAEGPILMISYADEEGARVITWKNVDLPYVDVVSTEREMIKRFLRVV
KEKDPDVLITYNGDNFDFAYLKKRCEKLGINFALGRDGSEPKIQRMGDRFAVEVKGRIHFDLYPVI
RRTINLPTYTLEAVYEAVFGQPKEKVYAEEITTAWETGENLERVARYSMEDAKVTYELGKEFLPME
AQLSRLIGQSLWDVSRSSTGNLVEWFLLRKAYERNELAPNKPDEKELARRRQTFEGGYIKEPERGL
WENIVYLDFRSIAPSIIITHNVSPDTLNREGCKEYDVAPQVGHRFCKDFPGFIPSLLGDLLEERQK
IKKKMKATIDPIERKLLDYRQREIKILANSYYGYYGYARARWYCKECAESVTAWGREYITMTIKEI
EEKYGFKVIYSDTDGFFATIPGADAETVKKKAMEFLKYINAKLPGALELEYEGFYKRGFFHTKKKY
AVIDEEGKITTRGLEIVRRDWSEIAKETQARVLEALLKDGDVEKAVRIVKEVTEKLSKYEVPPEKL
VIHEQITRDLKDYKAKGPHMAVAKRLAARGVKIRPGTVISYIVLKGSGRIGDRAIPFDEFDPTKHK
YDAEYYIENQVLPAVERILRAFGYRKEDLRYQKTRQVGLSAWLKPKGT*

Acknowledgment

This study was supported by Shenzhen Engineering Laboratory Molecular Enzymology (Fa Gaishen [2018] No. 958). W.L. Thanks to the National Natural Science Foundation of China, China Grant No. 21505134. Alexander K. Buell thanks the Novo Nordisk Foundation for support (NNFSA170028392). Thanks to China National GeneBank DataBase & BGI's Sequencing Platform and MGI's CoolMPS™ platform.

Author Contributions

Conceived and designed the experiments: Lili Zhai, Yue Zheng. Performed the experiments: Lili Zhai. Performed the simulation: Zi Wang (main), Lili Zhai. Performed the application test: Liu Fen, Jingjing Wang, Chongjun Xu, Hongyan Han. Analyzed the data: Lili Zhai. Contributed reagents/materials/analysis tools: Qingqing Xie, Yue Zheng, Wenwei Zhang, Yuliang Dong. Wrote the paper and SI: Lili Zhai, Zi Wang. Reviewed and modified the paper: Lili Zhai, Yue Zheng, Qingqing Xie, and Alexander Kai Buell.

Reference

1. Aschenbrenner, J. & Marx, A. DNA polymerases and biotechnological applications. *Curr. Opin. Biotechnol.* **48**, 187–195 (2017).
2. Joyce, C. M. Techniques used to study the DNA polymerase reaction pathway. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **1804**, 1032–1040 (2010).
3. Coulther, T. A., Stern, H. R. & Beuning, P. J. Engineering Polymerases for New Functions. *Trends Biotechnol.* **37**, 1091–1103 (2019).
4. Nikoomanzar, A., Chim, N., Yik, E. J. & Chaput, J. C. Engineering polymerases for applications in synthetic biology. *Q. Rev. Biophys.* **53**, e8 (2020).
5. Gardner, A. F. *et al.* Therminator DNA Polymerase: Modified Nucleotides and Unnatural Substrates. *Front. Mol. Biosci.* **6**, 28 (2019).

6. Hoshino, H. *et al.* Consecutive incorporation of functionalized nucleotides with amphiphilic side chains by novel KOD polymerase mutant. *Bioorg. Med. Chem. Lett.* **26**, 530–533 (2016).

7. Huber, C., von Watzdorf, J. & Marx, A. 5-methylcytosine-sensitive variants of *Thermococcus kodakaraensis* DNA polymerase. *Nucleic Acids Res.* gkw812 (2016) doi:10.1093/nar/gkw812.

8. Fuller, C. W. *et al.* Real-time single-molecule electronic DNA sequencing by synthesis using polymer-tagged nucleotides on a nanopore array. *Proc. Natl. Acad. Sci.* **113**, 5233–5238 (2016).

9. Kazlauskas, D., Krupovic, M., Guglielmini, J., Forterre, P. & Venclovas, Č. Diversity and evolution of B-family DNA polymerases. *Nucleic Acids Res.* **48**, 10142–10156 (2020).

10. Gardner, A. Determinants of nucleotide sugar recognition in an archaeon DNA polymerase. *Nucleic Acids Res.* **27**, 2545–2553 (1999).

11. 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).

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

13. Moreira et al. - 2007 - Computational alanine scanning mutagenesis—An impr.pdf.

14. Kirmizialtin, S., Nguyen, V., Johnson, K. A. & Elber, R. How Conformational Dynamics of DNA Polymerase Select Correct Substrates: Experiments and Simulations. *Structure* **20**, 618–627 (2012).

15. Coco, W. M. *et al.* DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat. Biotechnol.* **19**, 354–359 (2001).

16. Tubeleviciute, A. & Skirgaila, R. Compartmentalized self-replication (CSR) selection of Thermococcus litoralis Sh1B DNA polymerase for diminished uracil binding. *Protein Eng. Des. Sel.* **23**, 589–597 (2010).

17. 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).

18. Zeng, W., Guo, L., Xu, S., Chen, J. & Zhou, J. High-Throughput Screening Technology in Industrial Biotechnology. *Trends Biotechnol.* **38**, 888–906 (2020).

19. A homogeneous, high-throughput fluorescence resonance.pdf.

20. Chapter 1 Förster resonance energy transfer—FRET what is it, why do it, and how it's done.pdf.

21. Pauszek, R. F., Lamichhane, R., Rajkarnikar Singh, A. & Millar, D. P. Single-molecule view of coordination in a multi-functional DNA polymerase. *eLife* **10**, e62046 (2021).

22. Santoso, Y., Torella, J. P. & Kapanidis, A. N. Characterizing Single-Molecule FRET Dynamics with Probability Distribution Analysis. *ChemPhysChem* **11**, 2209–2219 (2010).

23. Stryer, L., Thomas, D. D. & Meares, C. F. Diffusion-Enhanced Fluorescence Energy Transfer. *Annu. Rev. Biophys. Bioeng.* **11**, 203–222 (1982).

24. 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).

25. Peng, Y., Sun, L., Jia, Z., Li, L. & Alexov, E. Predicting protein–DNA binding free energy change upon missense mutations using modified MM/PBSA approach: SAMPDI webserver. *Bioinformatics* **34**, 779–786 (2018).

26. Morikawa, M., Izawa, Y., Rashid, N., Hoaki, T. & Imanaka, T. Purification and characterization of a thermostable thiol protease from a newly isolated hyperthermophilic Pyrococcus sp. *Appl. Environ. Microbiol.* **60**, 4559–4566 (1994).

27. Rashid, N. & Aslam, M. An overview of 25 years of research on Thermococcus kodakarensis, a genetically versatile model organism for archaeal research. *Folia Microbiol. (Praha)* **65**, 67–78 (2020).

28. Choi, W. S. *et al.* How a B family DNA polymerase has been evolved to copy RNA. *Proc. Natl. Acad. Sci.* **117**, 21274–21280 (2020).

29. Terpe, K. Overview of thermostable DNA polymerases for classical PCR applications: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **97**, 10243–10254 (2013).

30. Arezi, B., Xing, W., Sorge, J. A. & Hogrefe, H. H. Amplification efficiency of thermostable DNA polymerases. *Anal. Biochem.* **321**, 226–235 (2003).

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

32. Kuwahara, M. *et al.* Study on Suitability of KOD DNA Polymerase for Enzymatic

Production of Artificial Nucleic Acids Using Base/Sugar Modified Nucleoside Triphosphates. *Molecules* **15**, 8229–8240 (2010).

33. 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).

34. 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).

35. Horhota, A. *et al.* Kinetic Analysis of an Efficient DNA-Dependent TNA Polymerase. *J. Am. Chem. Soc.* **127**, 7427–7434 (2005).

36. Hashimoto, H. *et al.* Crystal structure of DNA polymerase from hyperthermophilic archaeon Pyrococcus kodakaraensis KOD111Edited by R. Huber. *J. Mol. Biol.* **306**, 469–477 (2001).

37. Wynne, S. A., Pinheiro, V. B., Holliger, P. & Leslie, A. G. W. Structures of an Apo and a Binary Complex of an Evolved Archeal B Family DNA Polymerase Capable of Synthesising Highly Cy-Dye Labelled DNA. *PLoS ONE* **8**, e70892 (2013).

38. 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).

39. Hashimoto, H. *et al.* Crystal structure of DNA polymerase from hyperthermophilic archaeon Pyrococcus kodakaraensis KOD111Edited by R. Huber. *J. Mol. Biol.* **306**, 469–477 (2001).

40. Joyce, C. M. Choosing the right sugar: How polymerases select a nucleotide substrate. *Proc. Natl. Acad. Sci.* **94**, 1619–1622 (1997).

41. Kennedy, E. M., Hergott, C., Dewhurst, S. & Kim, B. The Mechanistic Architecture of Thermostable *Pyrococcus furiosus* Family B DNA Polymerase Motif A and Its Interaction with the dNTP Substrate. *Biochemistry* **48**, 11161–11168 (2009).

42. Leconte et al. - 2010 - Directed Evolution of DNA Polymerases for Next-Gen.pdf.

43. Seo, T. S. *et al.* Four-color DNA sequencing by synthesis on a chip using photocleavable fluorescent nucleotides. *Proc. Natl. Acad. Sci.* **102**, 5926–5931 (2005).

44. Palla, M. *et al.* DNA sequencing by synthesis using 3′-O-azidomethyl nucleotide reversible terminators and surface-enhanced Raman spectroscopic detection. *RSC Adv* **4**, 49342–49346 (2014).

45. Kielkopf, C. L., Bauer, W. & Urbatsch, I. L. Bradford Assay for Determining Protein Concentration. *Cold Spring Harb. Protoc.* **2020**, pdb.prot102269 (2020).

46. pour Nouroozi, R. Determination of Protein Concentration Using Bradford Microplate Protein Quantification Assay. *Int. Electron. J. Med.* (2015) doi:10.31661/iejm158.

47. Moreira, I. S., Fernandes, P. A. & Ramos, M. J. Computational alanine scanning mutagenesis—An improved methodological approach. *J. Comput. Chem.* **28**, 644–654 (2007).

48. Brown, J. A. & Suo, Z. Unlocking the Sugar "Steric Gate" of DNA Polymerases. *Biochemistry* **50**, 1135–1142 (2011).

49. Gardner, A. Determinants of nucleotide sugar recognition in an archaeon DNA polymerase. *Nucleic Acids Res.* **27**, 2545–2553 (1999).

50. O'Flaherty, D. K. & Guengerich, F. P. Steady-State Kinetic Analysis of DNA Polymerase Single-Nucleotide Incorporation Products. *Curr. Protoc. Nucleic Acid Chem.* **59**, (2014).

51. Raper, A. T., Reed, A. J. & Suo, Z. Kinetic Mechanism of DNA Polymerases: Contributions of Conformational Dynamics and a Third Divalent Metal Ion. *Chem. Rev.* **118**, 6000–6025 (2018).

52. Huang, J. *et al.* A reference human genome dataset of the BGISEQ-500 sequencer. *GigaScience* **6**, (2017).

53. Huang, J. *et al. BGISEQ-500 Sequencing v1*. https://www.protocols.io/view/bgiseq-500-sequencing-pq7dmzn (2018)

54. Drmanac, S. *et al. CoolMPS ™ : Advanced massively parallel sequencing using antibodies specific to each natural nucleobase*. http://biorxiv.org/lookup/doi/10.1101/2020.02.19.953307 (2020) doi:10.1101/2020.02.19.953307.

55. Lang, J. *et al.* Evaluation of the MGISEQ-2000 Sequencing Platform for Illumina Target Capture Sequencing Libraries. *Front. Genet.* **12**, 730519 (2021).

56. Korostin, D. *et al.* Comparative analysis of novel MGISEQ-2000 sequencing platform vs Illumina HiSeq 2500 for whole-genome sequencing. *PLOS ONE* **15**, e0230301 (2020).

57. Chen, F. *et al.* Reconstructed evolutionary adaptive paths give polymerases accepting reversible terminators for sequencing and SNP detection. *Proc. Natl. Acad. Sci.* **107**, 1948–1953 (2010).

58. Laos, R., Thomson, J. M. & Benner, S. A. DNA polymerases engineered by directed evolution to incorporate non-standard nucleotides. *Front. Microbiol.* **5**, (2014).

59. 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).

60. Kennedy, E. M., Hergott, C., Dewhurst, S. & Kim, B. The Mechanistic Architecture of Thermostable *Pyrococcus furiosus* Family B DNA Polymerase Motif A and Its Interaction with the dNTP Substrate. *Biochemistry* **48**, 11161–11168 (2009).

61. Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **10**, 449–461 (2015).

62. Genheden, S., Kuhn, O., Mikulskis, P., Hoffmann, D. & Ryde, U. The Normal-Mode Entropy in the MM/GBSA Method: Effect of System Truncation, Buffer Region, and Dielectric Constant. *J. Chem. Inf. Model.* **52**, 2079–2088 (2012).

63. Wang, C., Greene, D., Xiao, L., Qi, R. & Luo, R. Recent Developments and Applications of the MMPBSA Method. *Front. Mol. Biosci.* **4**, 87 (2018).

Figures

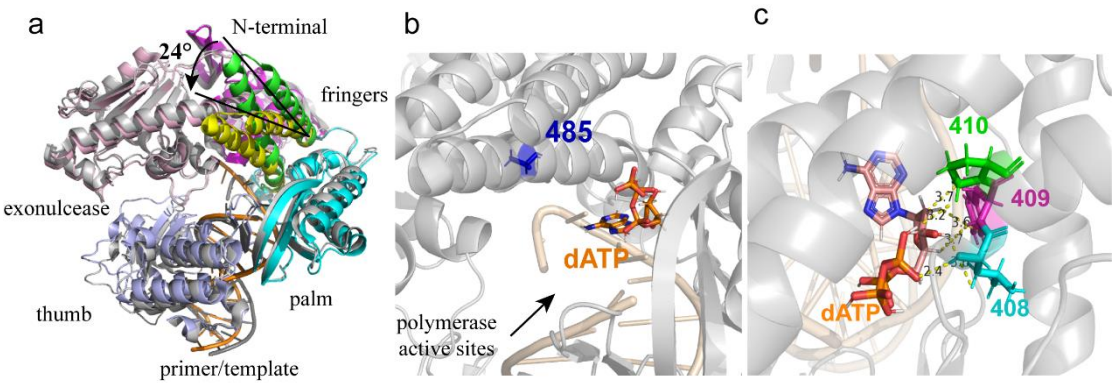


Figure 1. The crystal structure of KOD DNA polymerase. **(a)** shows the superimposed structure of a ternary KOD complex (PDB ID: 5MOF) and a binary KOD complex (PDB ID: 4K8Z) with the finger domain of the ternary complex (yellow) closed by approximately 24˚ in contrast to the finger domain of the binary complex (green), while the other domains of the KOD ternary complex are colored in gray, the exonuclease of

the KOD binary complex in light pink, the N terminal of the KOD binary complex in pink, the thumb domain of the KOD binary complex in light blue, the palm domain of the KOD binary complex is colored in cyan, and the primer/templates of the KOD ternary and binary complex are colored in gray and orange, respectively; **(b)** depicts the active pocket of KOD pol and A485 position colored in blue, dATP was colored by orange elements; **(c)** shows the interaction of motif A (408/409/410) and dATP (orange elements), the atomic distances (yellow) were measured with the PyMOL software, L408 is colored in cyan, Y409 is colored in pink, and P410 is colored in green.

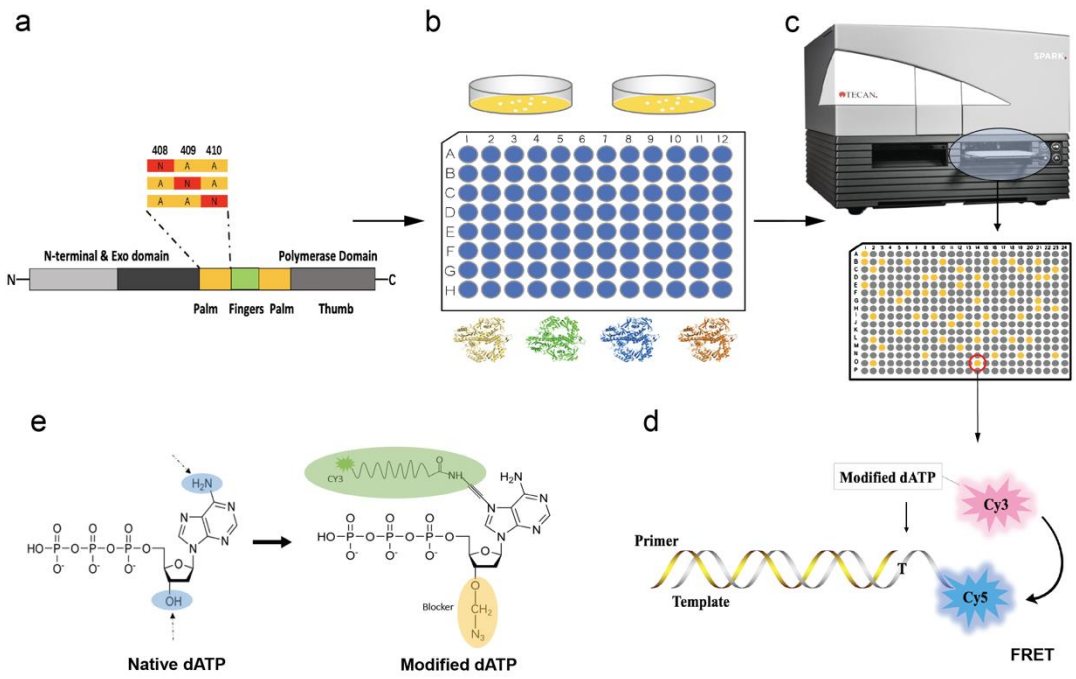


Figure 2. Schematic of KOD DNA pol directed evolution experiments. **(a)** illustrates the library construction containing 60 mutants with independent saturation mutagenesis on positions 408/409/410. **(b)** Illustrates the protein expression. A single colony was selected and cultivated in a 96-deep-well plate to express the polymerase overnight at 25°C, 150 rpm. **(c)** illustrates the variant screening with FRET-Cy5 fluorescence as an

indicator. The crude enzyme from (b) was collected to measure the enzyme activity and kinetics at 40°C for 1-2 h. **(d)** The schematic illustration of the incorporation process. The Cy5 FRET-acceptor fluorescence can be detected when Cy3 dye-labeled modified dATP was added to the primer/template with a Cy3 excitation at 555 nm. **(e)** shows the molecular structures of 3'-O-azidomethyl-deoxyadenosine triphosphate with dye Cy3 labeled (modified dATP) and native dATP.

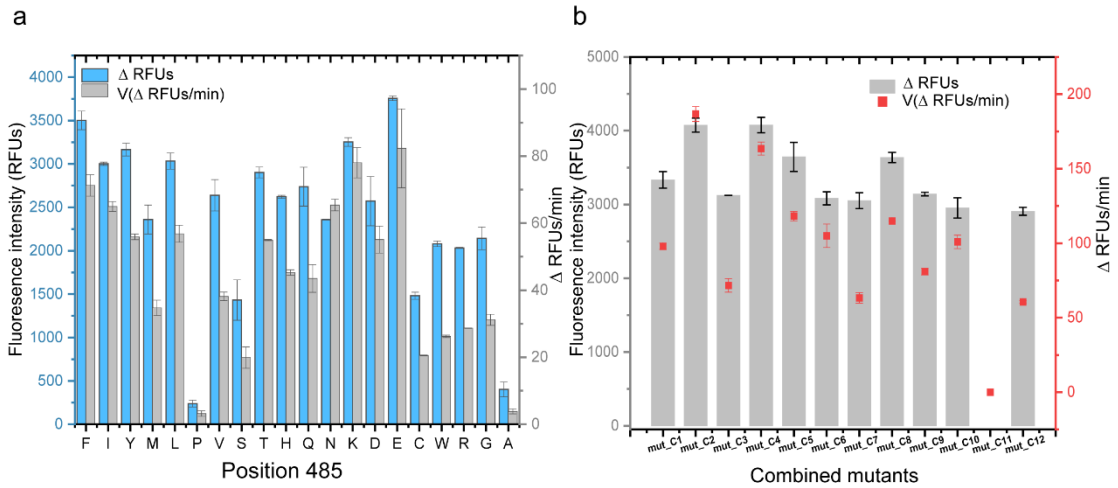


Figure 3. Function analysis of 20 mutant variants of position 485 based on Mut_1 and 12 KOD variants with combined mutations. **(a)** The polymerization activities of 20 variants towards modified dATP were measured for 2 hours by using a microplate reader; the total increase in fluorescence intensity ((Δ RFUs) is shown as blue bars; the enzyme activity V (Δ RFUs/min) is presented as gray bars; **(b)** Enzyme activities of 12 combined mutants of motif A and position 485 were measurement in the same way. Grey bars represent the average relative fluorescence units (RFUs), and the red color showed the average relative fluorescence units (RFUs) per min (RFUs/min).

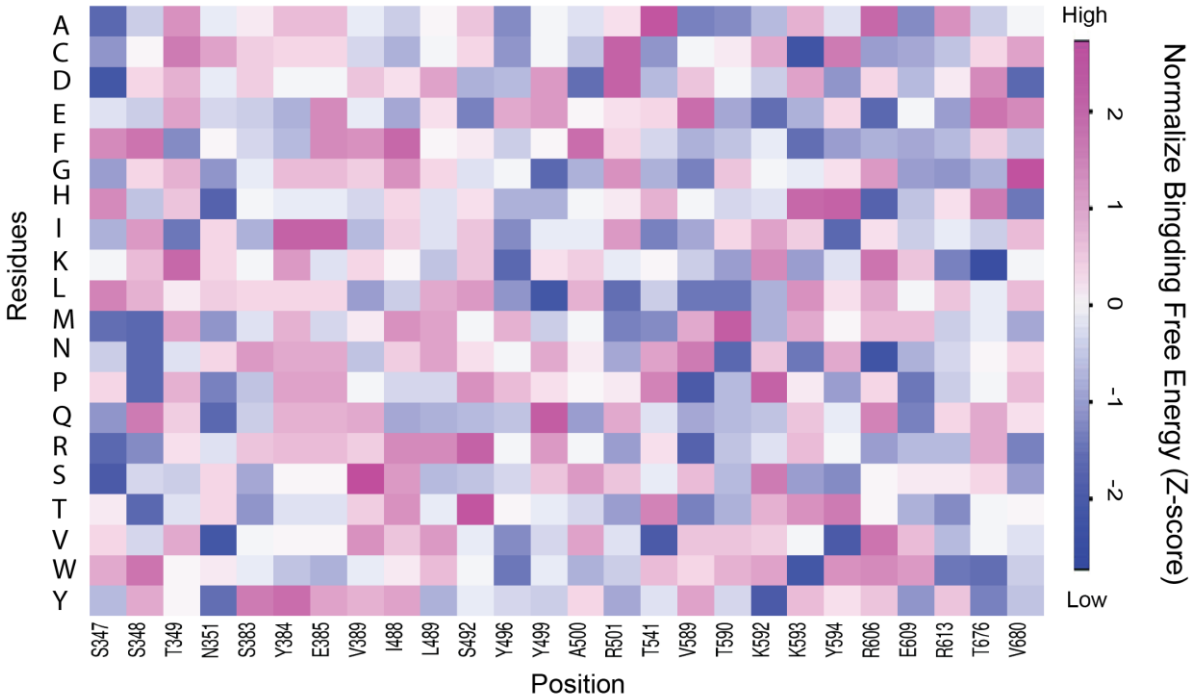


Figure 4. Computational analysis of single mutant variants based on Mut_C2. The heat map illustrates the scaled (calculated by the Z-scored normalization method) of the binding free energy for each single-point mutation. Saturation mutations were performed at one site, and the binding energies of 20 mutants were simulated and calculated. These mutants with the lower binding energy were selected for experimental verification.

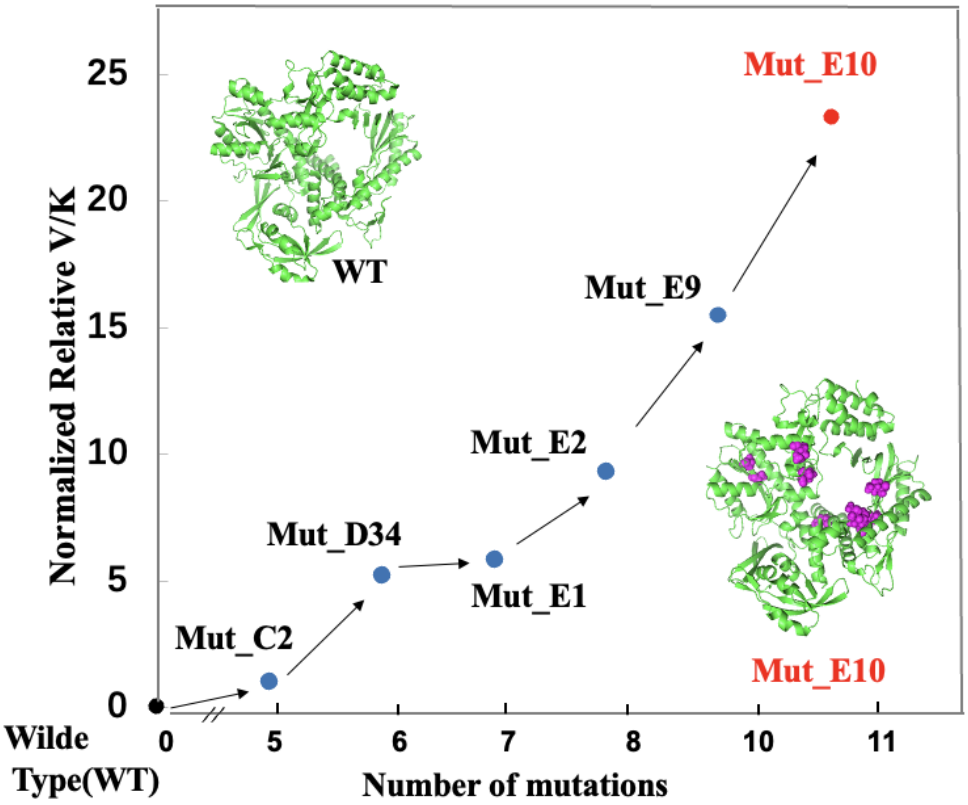


Figure 5. Relative V/K of different mutants was normalized to WT. The normalized relative V/K corresponds to the relative kinetic performance toward the DNA primer/template. Compared to the wild-type KOD pol, Mut_2 contained 5 mutation sites; Mut_D34 contained 6 mutation sites; Mut_E1 contained 7 mutation sites; Mut_E2 contained 8 mutation sites; Mut_E9 contained 10 mutation sites; Mut_E10 contained 11 mutation sites. These mutation sites colored in pink and presented as spheres are illustrated in the structure in the lower right corner of this figure.

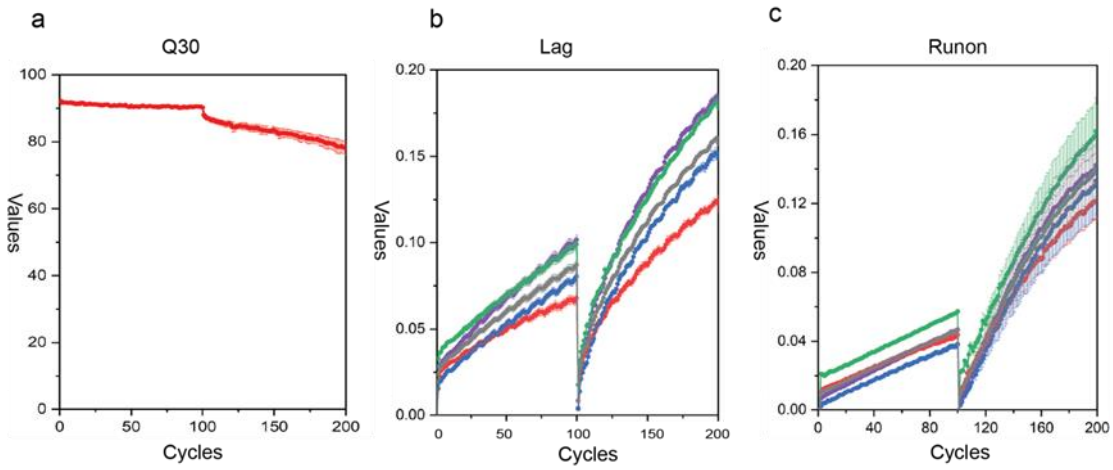


Figure 6. Mut_E10 sequencing data on the MGISEQ-2000 sequencer. **(a)** shows the Q30 curve with PE100. **(b)** presents the lag curve and **(c)** shows the runon curve.

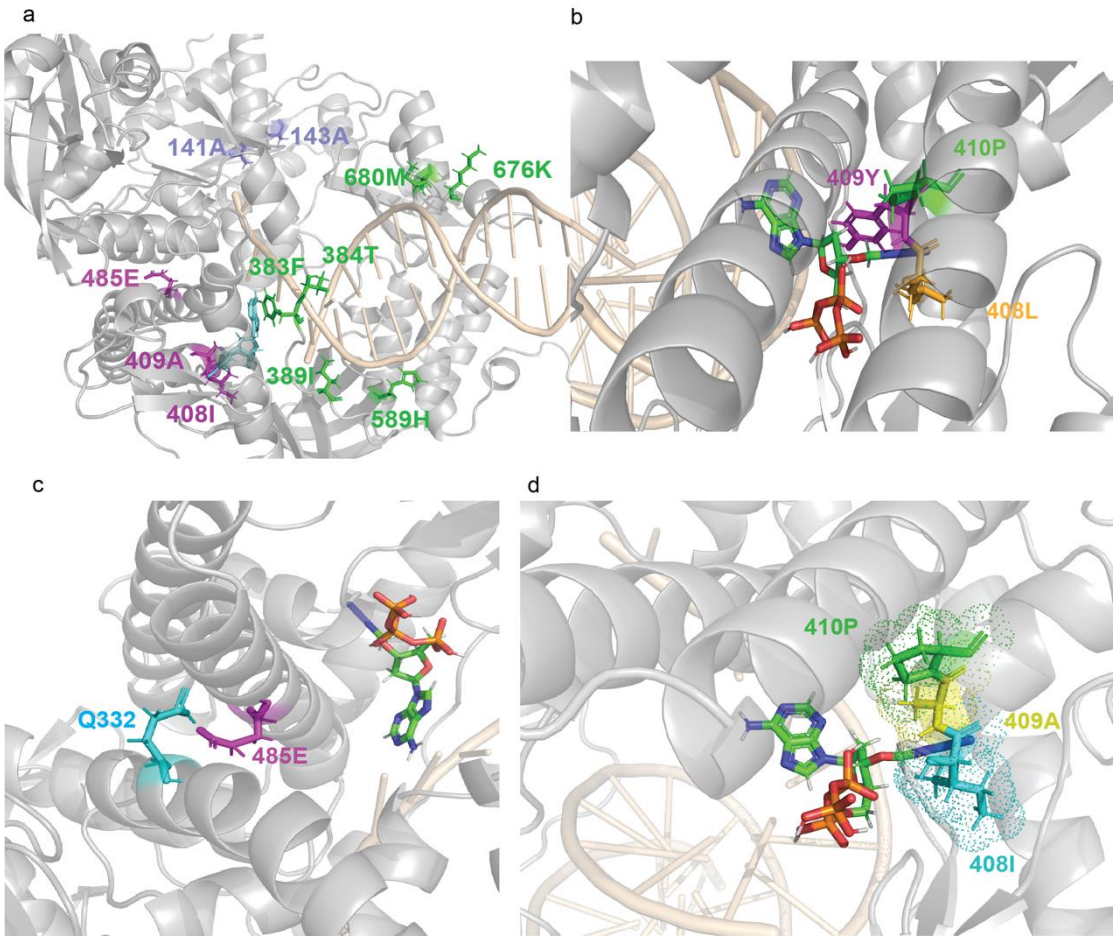


Figure 7． Structural analysis of the designed KOD variant Mut_E10. **(a)** shows the 11 mutation sites of Mut_E10. The exonuclease region inactivation sites 141A and 143A are colored in purple, 408I and 409A located in the palm domain are colored in pink, and 485E located in the finger domain is also colored in pink. Residues contained in the DNA binding region, including 383F, 384T, 389I, 589H, 676K, and 680M, were colored in green. **(b)** depicts the interaction of 408L colored in light orange, 409Y colored in pink, 410P colored in green, and 3'-O-azidomethyl-dATP. **(c)** shows the interaction of Q332 colored in cyan and 485E colored in pink. **(d)** shows Mut_E10's active pocket or catalytic reaction center including 408I colored in cyan, 409A colored in yellow, and 410P colored by atoms. The cloud of dots corresponds to motif A.

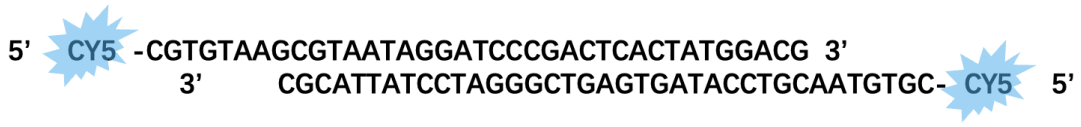


Figure 8. The sequence of the labeled templating primer (PT-2Cy5).

Supplementary Information

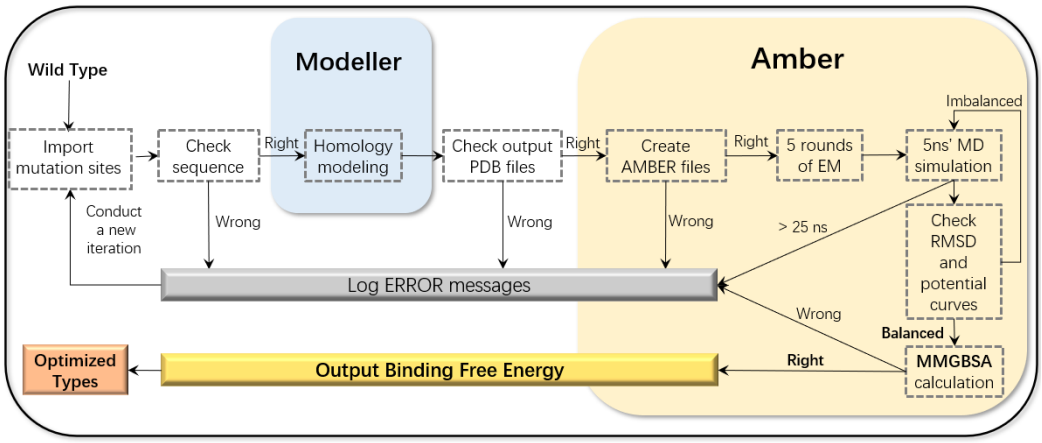
Early studies suggested that the time scale of the binding process between polymerase and dsDNA was longer than that between complex and dNTP. Thus, finding the key residues of this initial binding process was chosen as a strategy for the directional transformation of polymerase and is likely to improve its catalytic activity. In order to study the effects of amino acids interacting with dsDNA of the N-terminal, exonucleases, finger, thumb, and palm domain, we simulated the saturation mutation of 93 residues in five rounds by the homology modeling method and calculated the change of binding free energy between the mutants and dsDNA by MM-GBSA method in Supplementary **Supplementary Figure** 4.

The average binding energy of 93 residues was calculated and is shown in Supplementary **Table 1**. Due to the high standard deviation of MM-GBSA calculation, during data analysis, we first calculate the average value of the calculation results of 20 mutants at a single site and then select some mutants for further experimental testing on the premise of determining their lower average value. We expressed and tested several enzyme mutants with an average binding energy of < 255 kcal/mol and some interesting positions. These successfully constructed and expressed mutants were subjected to measurement of enzymatic activity and polymerization kinetics towards the DNA primer/template, and the results are shown in Supplementary **Table 2**. The kinetics data towards modified dATP of step-wise combined mutants are presented in

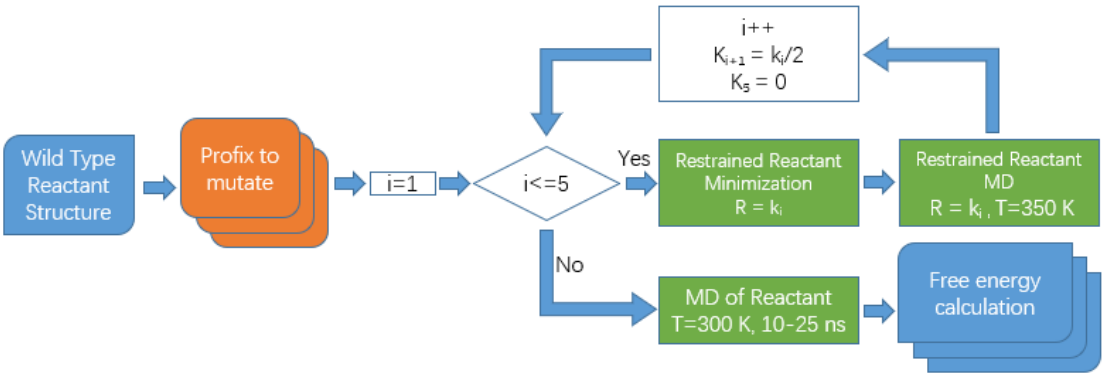
Supplementary **Table 3**.

Supplementary Table 1. Average Binding Free Energy (Kcal/mol) of 93 Mutation Sites in DNA-binding region.

Mutant	Avg. (Kcal/mol)	Mutant	Avg. (Kcal/mol)	Mutant	Avg. (Kcal/mol)	Mutant	Avg. (Kcal/mol)
R265	-236.3	P392	-228.8	W504	-264	G601	-256.3
N269	-231.2	E393	-259.5	Y505	-260	K602	-255.1
L270	-250.9	R394	-258.9	C506	-249.6	I603	-254.6
P271	-220.3	G395	-263.2	K507	-230.4	T604	-242.5
S347	-272.9	Y402	-260.1	A510	-251.3	T605	-270.8
S348	-252.3	L403	-241.4	E511	-243.1	R606	-242.0
T349	-235.1	D404	-258.8	T514	-232.7	G607	-259.8
G350	-260.8	K487	-256.3	R518	-246.8	L608	-256.2
N351	-271.6	I488	-231.9	Y538	-261.6	E609	-273.1
L352	-251.9	L489	-244.9	S539	-266.1	I610	-272.2
V353	-246.3	A490	-277.2	D540	-260.9	V611	-243.5
R379	-278.3	N491	-270.2	T541	-257.5	R612	-253.6
R380	-252.8	S492	-239.6	D542	-268.1	R613	-267.4
R381	-236.6	Y493	-256.4	G543	-258.6	D614	-263.6
Q382	-238.3	Y494	-256.5	F544	-257.5	W615	-259.9
T383	-255.7	G495	-254.6	F545	-268.4	S616	-278.8
F384	-228.5	Y496	-261.7	A546	-245.9	V661	-243.1
E385	-228.5	Y497	-252.4	H589	-265.7	I662	-249.1
G386	-271.4	G498	-259.0	T590	-267.4	H663	-256.2
G387	-241.3	Y499	-254.4	K591	-243.2	K676	-256.2
Y388	-260.7	A500	-250.2	K592	-245.4	M680	-253.1
I389	-238.3	R501	-249.0	K593	-255.3		
K390	-250.8	A502	-248.4	Y594	-256.0		
E391	-258	R503	-253.2	A595	-253.1		



Supplementary Figure 4. Automation process of MM-GBSA calculation. EM signifies Energy Minimization; MD signifies Molecular Dynamics; RMSD signifies the Root-Mean-Square Deviation.



Supplementary Figure 5. Mutants' pretreatment and MD simulation process. i was the cycle number of energy minimization; k_i was the restraint coefficient; T was temperature.

Kinetic results of KOD variants

Supplementary Table 2. Kinetic data towards the DNA primer/template of combined mutants.

Mutants	Sites	Subdomain	Relative V	K_m (μM)	V_{max} (RFUs/min)	V_{max}/K_m (RFUs/min*μM)	Relative V/K
Mut_C2	-	-	1.0	2.3	596.0	259.1	1.0
Mut_D34	V589H	Palm	2.6	1.0	1943.0	1926.3	7.4
Mut_D35	V589Q	Palm	1.8	1.2	1352.9	1124.9	4.3

Mut_D51	V680M	Thumb	0.5	0.5	350.5	753.9	2.9
Mut_D48	T676K	Thumb	1.2	0.9	536.4	596.0	2.3
Mut_D30	T590K	Palm	1.1	0.9	357.6	411.3	1.6
Mut_D33	V389M	Palm	1.1	0.8	326.5	408.1	1.6
Mut_D3	Y496I	Finger	0.7	4.6	1579.4	341.7	1.3
Mut_D6	I488Q	Finger	0.9	0.9	303.1	336.8	1.3
Mut_D32	V389I	Palm	0.9	1.3	363.6	280.1	1.1
Mut_D7	L489Y	Finger	0.9	0.7	181.4	259.1	1.0
Mut_D17	T349I	N-Ter	1.1	1.4	348.7	257.6	1.0
Mut_D20	E385M	Palm	0.9	1.1	283.7	251.0	1.0
Mut_D36	Y384F	Palm	0.8	2.0	500.6	249.3	1.0
Mut_D9	I488L	Finger	1.0	0.8	196.7	245.9	0.9
Mut_D12	S347I	N-Ter	1.1	1.0	238.4	236.8	0.9
Mut_D13	S347M	N-Ter	0.7	2.0	464.9	231.9	0.9
Mut_D40	S383T	Palm	1.1	2.2	476.8	218.0	0.8
Mut_D42	E609Y	Thumb	0.9	1.5	327.8	215.3	0.8
Mut_D8	L489K	Finger	0.8	0.9	196.7	208.0	0.8
Mut_D31	T590L	Palm	0.6	1.3	255.7	204.1	0.8
Mut_D37	Y384W	Palm	0.9	1.7	350.2	201.4	0.8
Mut_D14	S348M	N-Ter	0.9	2.4	453.0	188.4	0.7
Mut_D50	V680D	Thumb	0.5	1.4	256.3	183.1	0.7
Mut_D29	T541w	Palm	0.6	1.1	191.9	174.6	0.7
Mut_D41	S383I	Palm	0.7	1.1	191.9	174.6	0.7
Mut_D19	A500G	Palm	1.4	3.6	619.8	173.9	0.7
Mut_D27	R501M	Palm	0.6	2.0	327.8	163.5	0.6
Mut_D39	Y594V	Palm	0.9	3.1	510.8	162.9	0.6
Mut_D15	S348T	N-Ter	0.9	2.5	389.8	158.1	0.6
Mut_D5	Y499L	Finger	0.7	2.1	333.8	158.0	0.6
Mut_D21	E385W	Palm	1.1	5.2	804.6	154.6	0.6
Mut_D10	N351H	N-Ter	1.4	1.3	190.7	151.1	0.6
Mut_D18	A500D	Palm	1.3	1.3	190.7	150.9	0.6
Mut_D26	R501L	Palm	0.9	1.7	238.4	141.3	0.5
Mut_D11	N351Y	N-Ter	1.2	7.5	822.5	110.2	0.4

Mut_D38	Y594I	Palm	0.0	0.8	71.5	90.7	0.3
Mut_D23	K592Y	Palm	0.6	2.5	222.4	88.0	0.3
Mut_D1	S492E	Finger	0.7	1.7	84.6	50.8	0.2
Mut_D43	E609G	Thumb	0.8	4.4	208.6	47.2	0.2
Mut_D2	S492Y	Finger	0.2	NA	NA	NA	NA
Mut_D4	Y496L	Finger	0.4	NA	NA	NA	NA
Mut_D16	T349F	N-Ter	0.4	NA	NA	NA	NA
Mut_D22	K592L	Palm	0.3	NA	NA	NA	NA
Mut_D24	K593F	Palm	0.1	NA	NA	NA	NA
Mut_D25	K593W	Palm	0.2	NA	NA	NA	NA
Mut_D28	T541I	Palm	0.2	NA	NA	NA	NA
Mut_D44	R606H	Thumb	0.4	NA	NA	NA	NA
Mut_D45	R606N	Thumb	0.3	NA	NA	NA	NA
Mut_D46	R613G	Thumb	0.3	NA	NA	NA	NA
Mut_D47	R613T	Thumb	0.4	NA	NA	NA	NA
Mut_D49	T676Y	Thumb	0.1	NA	NA	NA	NA
Mut_D52	Y499G	Finger	0.2	NA	NA	NA	NA
Mut_D53	S347I	N-Ter	NA	NA	NA	NA	NA
Mut_D54	Y388I	Palm	NA	NA	NA	NA	NA
Mut_D55	A490I	Finger	NA	NA	NA	NA	NA
Mut_D56	A490K	Finger	NA	NA	NA	NA	NA
Mut_D57	N491F	Finger	NA	NA	NA	NA	NA
Mut_D58	S539Q	Palm	NA	NA	NA	NA	NA
Mut_D59	D542E	Palm	NA	NA	NA	NA	NA
Mut_D60	F545K	Palm	NA	NA	NA	NA	NA
Mut_D61	Y505S	Palm	NA	NA	NA	NA	NA
Mut_D62	T605Y	Thumb	NA	NA	NA	NA	NA
Mut_D63	I610Y	Thumb	NA	NA	NA	NA	NA
Mut_D64	R613I	Thumb	NA	NA	NA	NA	NA
Mut_D65	S616R	Thumb	NA	NA	NA	NA	NA
Mut_D66	V611N	Thumb	NA	NA	NA	NA	NA

Mut_D67	Y497V	Finger	NA	NA	NA	NA	NA
Mut_D68	Y497L	Finger	NA	NA	NA	NA	NA
Mut_D69	G543H	Palm	NA	NA	NA	NA	NA
Mut_D70	F544K	Palm	NA	NA	NA	NA	NA

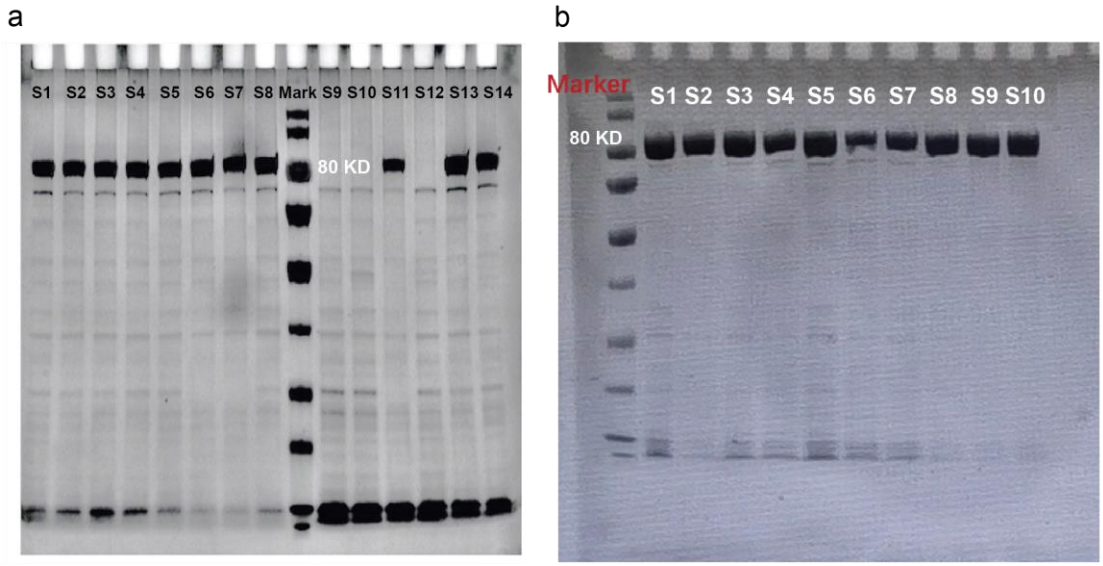
Notes: Relative V was defined as the relative enzyme activity, and compared with Mut_C2, the value of Mut_C2 is defined as 1. N-Ter was the abbreviation of N-Terminal. Relative V/K was defined as the relative enzyme catalytic efficiency, and compared with Mut_C2, the value of Mut_C2 was defined as 1. K_m presents the affinity between the enzyme and the substrate. V_{max}/K_m showed enzyme catalytic efficiency of mutants. From Mut_D53 to Mut_70, the polymerase expression failed; N/A=Not Available.

Supplementary Table 3. Kinetic data towards the modified dATP of these step-wise combined mutants.

Mutant	K_m (μM)	V_{max} (RFUs/min)	V_{max}/K_m (RFUs/min*μM)	Relative V/K
Mut_D34	1.8	894.3	18938.6	1.0
Mut_E1	2.2	1132.5	514.8	1.0
Mut_E2	1.1	1206.0	1096.4	2.2
Mut_E3	3.0	79.8	26.6	0.1
Mut_E4	2.6	1415.6	544.4	1.6
Mut_E5	3.7	172.0	46.5	0.1
Mut_E6	1.8	894.3	496.8	1.4
Mut_E7	2.5	849.3	339.7	1.0
Mut_E8	1.5	1079.5	719.7	2.1
Mut_E9	1.2	1508.2	1256.8	3.7
Mut_E10	2.1	2407.8	1146.6	3.3
Mut_E11	8.8	178.6	20.3	0.1
Mut_E12	2.8	420.7	150.3	0.4

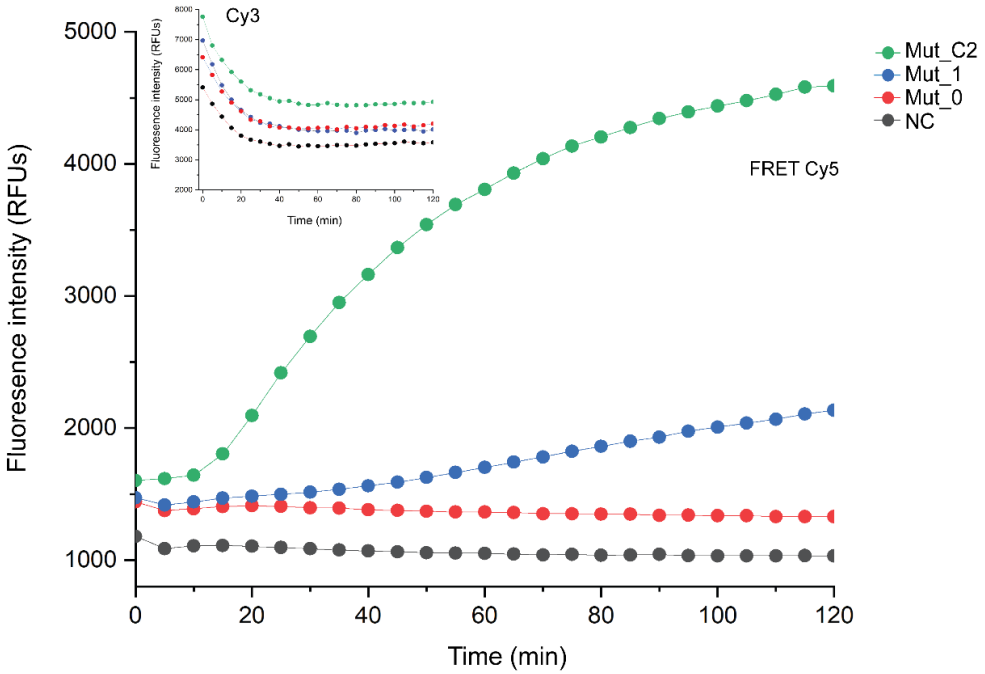
Relative V/K was defined as the relative enzyme catalytic efficiency, and compared with Mut_D34, the value of Mut_D34 was defined as 1.

Enzyme purity verification by SDS-PAGE gel



Supplementary Figure 3. Protein expression of KOD pol and KOD mutants was measured by running 12% SDS-PAGE gel. a) KOD polymerase variant expression is verified without further purification. S1 is Mut_0, S2 is Mut_1, 12 mutants are identified from the first library shown from S3 to S14, and some mutants failed to express polymerases such as S9, S10, and S12. b) Protein purification of KOD variants is verified by SDS-PAGE gel. The KOD polymerase variants obtained according to the crude enzyme expression protocol (S1 to S7) were compared with the pure polymerases (S8 Mut_0, S9 Mut_C2, and S10 Mut_E10) obtained after purification, and the protein concentration was estimated as having a purity of about 80%~85% from SDS-PAGE gels. After calculating the total protein concentration (Conc. $_{Total}$) by the Bradford method, and the real protein concentration = Conc. $_{Total}$ $*$0.8.

Establishment of screening method



Supplementary Figure 4. The establishment that the screening method based on the FRET fluorescence signal reflects the polymerization efficiency of DNA polymerase. The increase of FRET-CY5 fluorescence signal was measured as relative fluorescence units (RFUs) in appropriate time intervals by exciting at 530 nm and measuring the emission at 676 nm. NC was negative control, presenting enzyme was replaced by the buffer during the reaction process; Mut_0 cannot produce a fluorescence signal shown in red color; Mut_1 shows polymerase activity exhibiting FRET-CY5 signal increased; Mut_C2 (IAP, E) is a combined mutant of motif A and 485 that presents efficiency enzyme activity towards modified dATP.

Figures

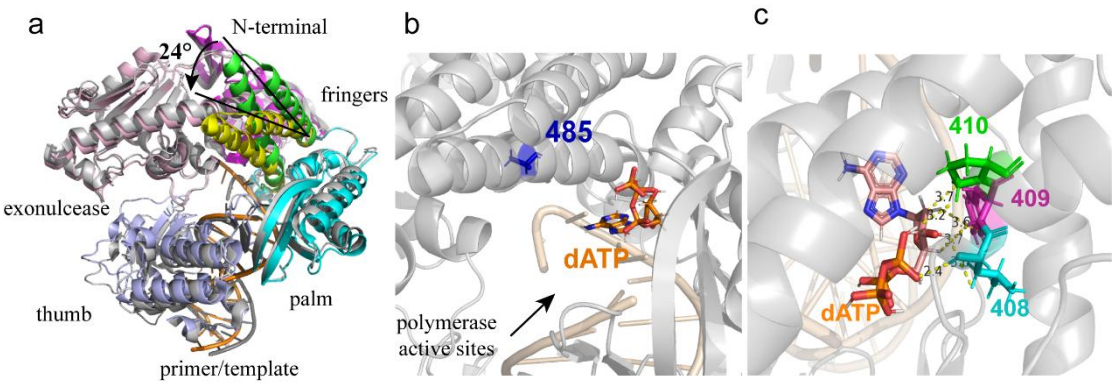


Figure 1. The crystal structure of KOD DNA polymerase. A) shows the superimposed structure of a ternary KOD complex (PDB ID: 5MOF) and a binary KOD complex (PDB ID: 4K8Z) with the finger domain of the ternary complex (yellow) closed by approximately 24˚ in contrast to the finger domain of the binary complex (green), while the other domains of the KOD ternary complex are colored in gray, the exonuclease of the KOD binary complex in light pink, the N terminal of the KOD binary complex in pink, the thumb domain of the KOD binary complex in light blue, the palm domain of the KOD binary complex is colored in cyan, and the primer/templates of the KOD ternary and binary complex are colored in gray and orange, respectively; b) depicts the active pocket of KOD pol and A485 position colored in blue, dATP was colored by orange elements; c) shows the interaction of motif A (408/409/410) and dATP (orange elements), the atomic distances (yellow) were measured with the PyMOL software, L408 is colored in cyan, Y409 is colored in pink, and P410 is colored in green.

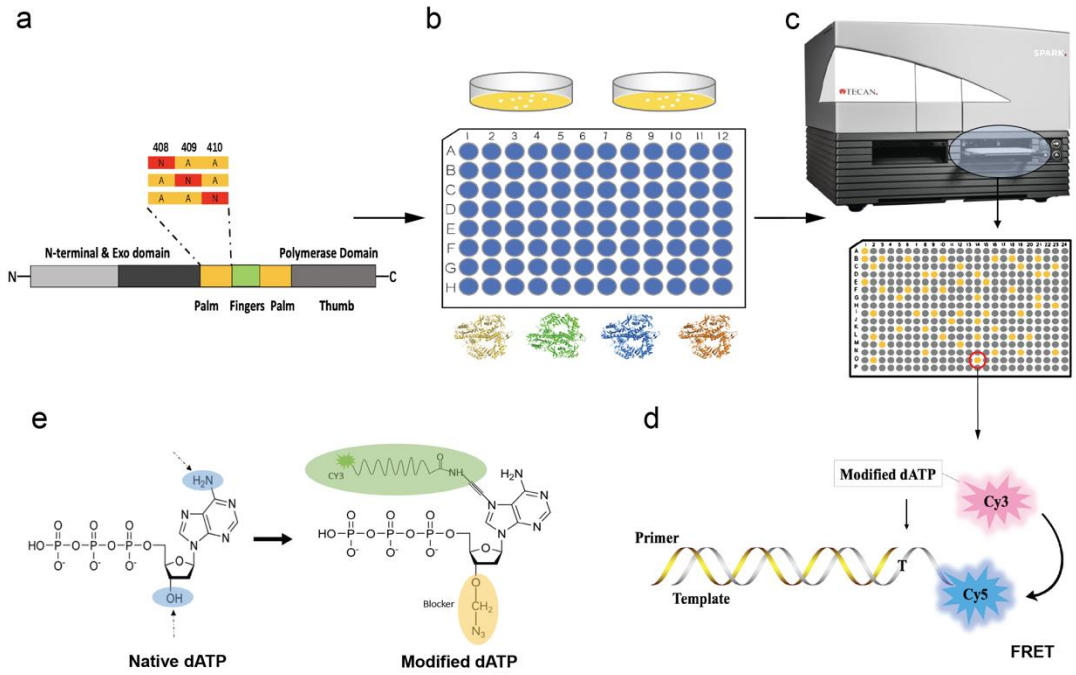


Figure 2. Schematic of KOD DNA pol directed evolution experiments. a) illustrates the library construction containing 60 mutants with independent saturation mutagenesis on positions 408/409/410. b) Illustrates the protein expression. A single colony was selected and cultivated in a 96-deep-well plate to express the polymerase overnight at 25°C, 150 rpm. c) illustrates the variant screening with FRET-Cy5 fluorescence as an indicator. The crude enzyme from b) was collected to measure the enzyme activity and kinetics at 40°C for 1-2 h. d) The schematic illustration of the incorporation process. The Cy5 FRET-acceptor fluorescence can be detected when Cy3 dye-labeled modified dATP was added to the primer/template with a Cy3 excitation at 555 nm. e) shows the molecular structures of 3'-O-azidomethyl-deoxyadenosine triphosphate with dye Cy3 labeled (modified dATP) and native dATP.

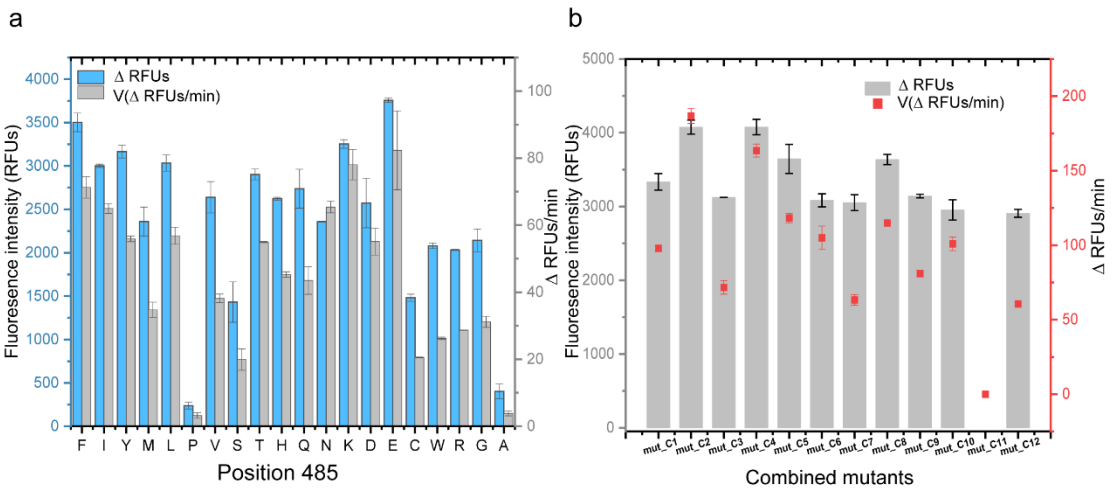


Figure 3. Function analysis of 20 mutant variants of position 485 based on mut_1 and 12 KOD variants with combined mutations. a) The polymerization activities of 20 variants towards modified dATP were measured for 2 hours by using a microplate reader; the total increase in fluorescence intensity ((Δ RFUs) is shown as blue bars; the enzyme activity V (Δ RFUs/min) is presented as gray bars; b) Enzyme activities of 12 combined mutants of motif A and position 485 were measurement in the same way. Grey bars represent the average relative fluorescence units (RFUs), and the red color showed the average relative fluorescence units (RFUs) per min (RFUs/min).

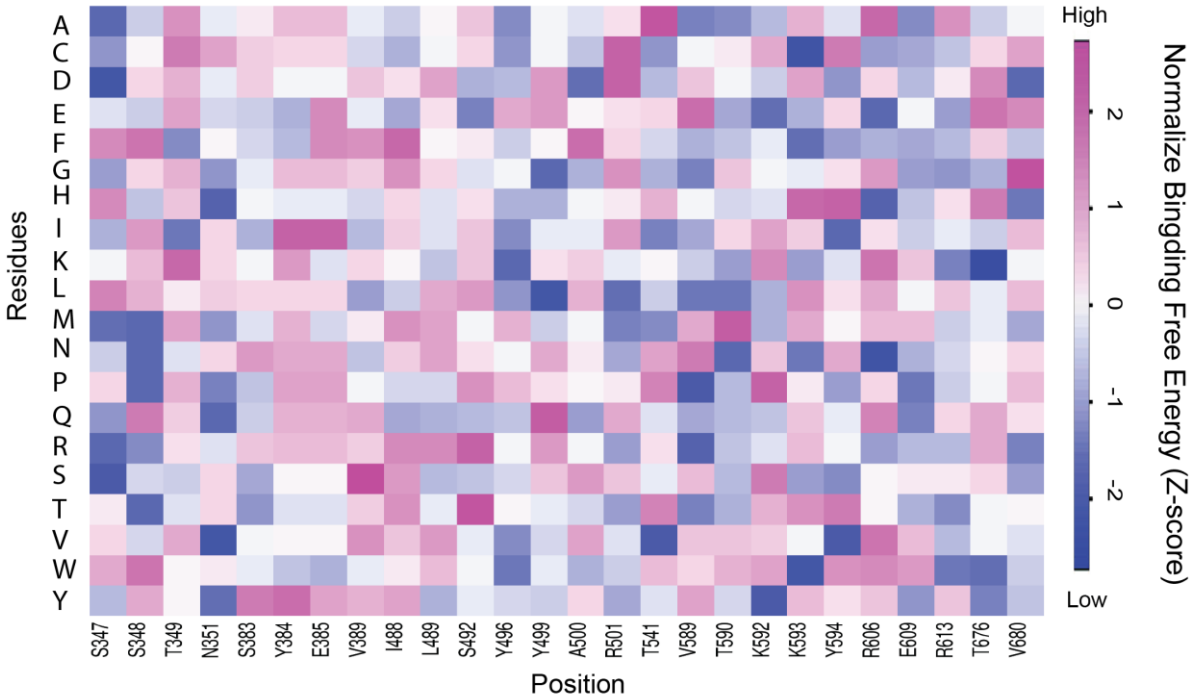


Figure 4. Computational analysis of single mutant variants based on mut_C2. The heat map illustrates the scaled (calculated by the Z-scored normalization method) of the binding free energy for each single-point mutation. Saturation mutations were performed at one site, and the binding energies of 20 mutants were simulated and calculated. These mutants with the lower binding energy were selected for experimental verification.

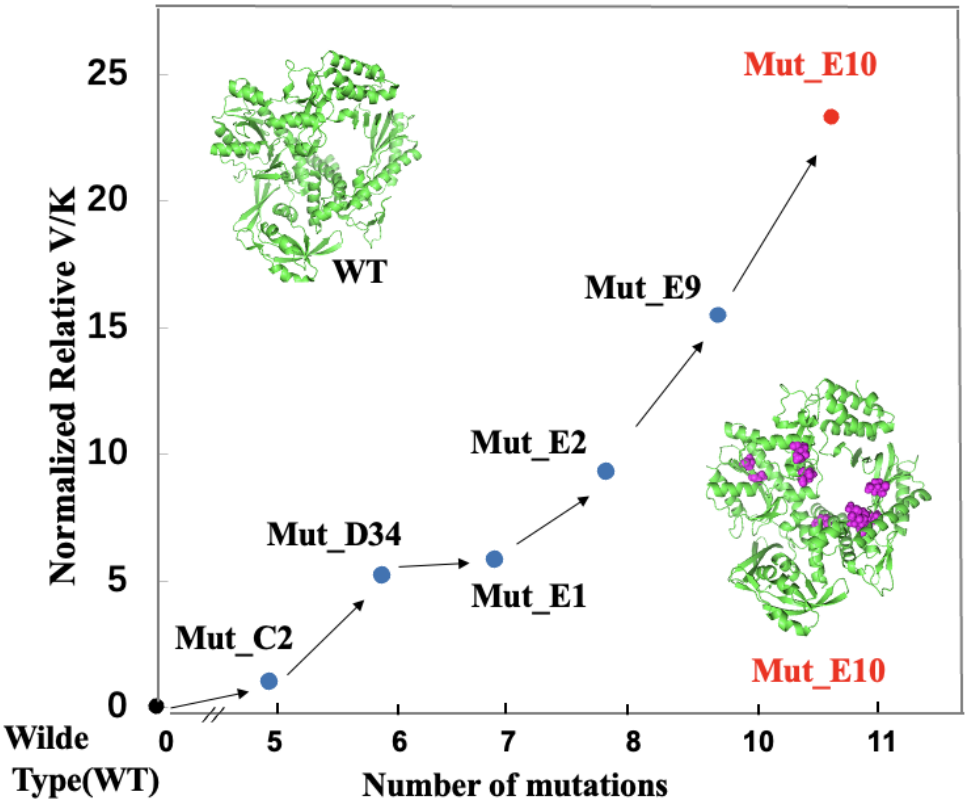


Figure 5. Relative V/K of different mutants was normalized to WT. The normalized relative V/K corresponds to the relative kinetic performance toward the DNA primer/template. Compared to the wild-type KOD pol, mut_2 contained 5 mutation sites; mut_D34 contained 6 mutation sites; mut_E1 contained 7 mutation sites; mut_E2 contained 8 mutation sites; mut_E9 contained 10 mutation sites; mut_E10 contained 11 mutation sites. These mutation sites colored in pink and presented as spheres are illustrated in the structure in the lower right corner of this figure.

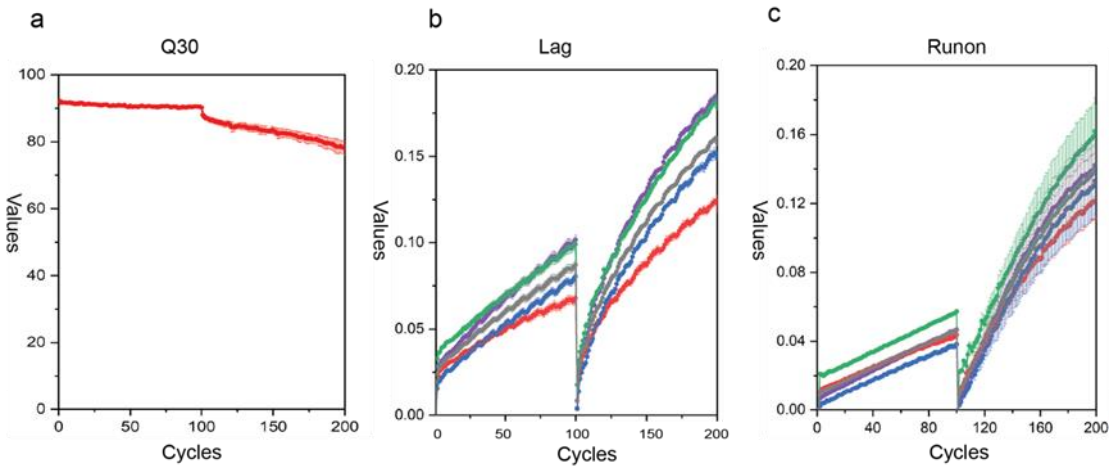


Figure 6. Mut_E10 sequencing data on the MGISEQ-2000 sequencer. A) shows the Q30 curve with PE100. B) presents the lag curve and c) shows the runon curve.

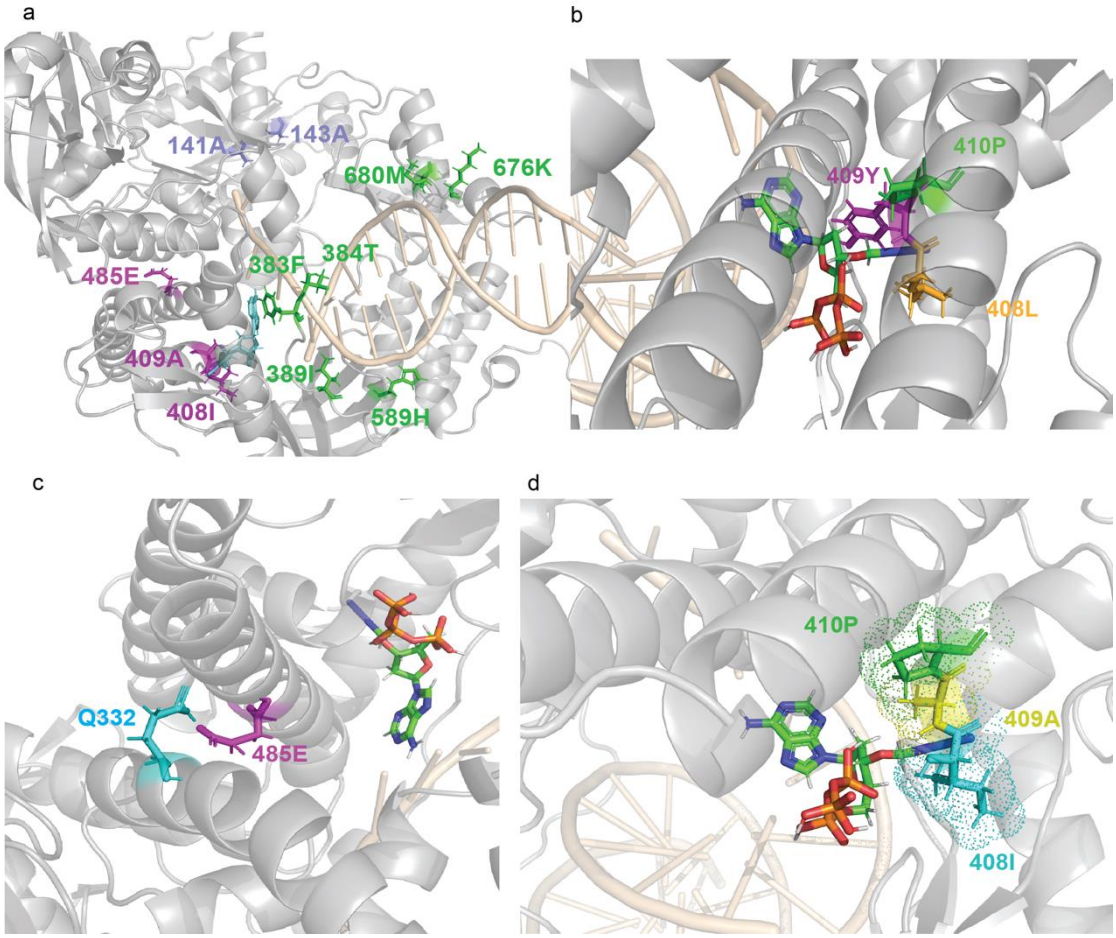


Figure 7. Structural analysis of the designed KOD variant mut_E10. A) shows the 11 mutation sites of mut_E10. The exonuclease region inactivation sites 141A and 143A are colored in purple, 408I and 409A located in the palm domain are colored in pink, and 485E located in the finger domain is also colored in pink. Residues contained in the DNA binding region, including 383F, 384T, 389I, 589H, 676K, and 680M, were colored in green. b) depicts the interaction of 408L colored in light orange, 409Y colored in pink, 410P colored in green, and 3'-O-azidomethyl-dATP. c) shows the interaction of Q332 colored in cyan and 485E colored in pink. d) shows mut_E10's active pocket or catalytic reaction center including 408I colored in cyan, 409A colored in yellow, and 410P colored by atoms. The cloud of dots corresponds to motif A.

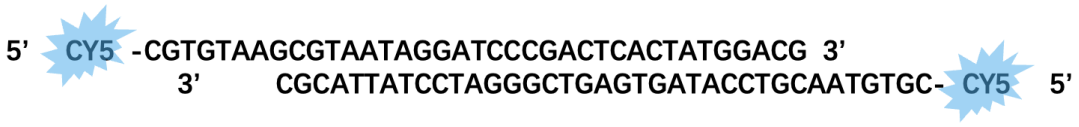


Figure 8. The sequence of the labeled templating primer (PT-2Cy5).