

# SCIENTIFIC REPORTS



OPEN

## DNA binding strength increases the processivity and activity of a Y-Family DNA polymerase

Jing Wu<sup>1,2</sup>, Alexandra de Paz<sup>3</sup>, Bradley M. Zamft<sup>4</sup>, Adam H. Marblestone<sup>5,6</sup>, Edward S. Boyden<sup>7,8,9</sup>, Konrad P. Kording<sup>10,11</sup> & Keith E. J. Tyo<sup>1</sup>

DNA polymerase (pol) processivity, i.e., the bases a polymerase extends before falling off the DNA, and activity are important for copying difficult DNA sequences, including simple repeats. Y-family pols would be appealing for copying difficult DNA and incorporating non-natural dNTPs, due to their low fidelity and loose active site, but are limited by poor processivity and activity. In this study, the binding between Dbh and DNA was investigated to better understand how to rationally design enhanced processivity in a Y-family pol. Guided by structural simulation, a fused pol Sdbh with non-specific dsDNA binding protein Sso7d in the N-terminus was designed. This modification increased *in vitro* processivity 4-fold as compared to the wild-type Dbh. Additionally, bioinformatics was used to identify amino acid mutations that would increase stabilization of Dbh bound to DNA. The variant SdbhM76I further improved the processivity of Dbh by 10 fold. The variant SdbhKSKIP241–245RVRKS showed higher activity than Dbh on the incorporation of dCTP (correct) and dATP (incorrect) opposite the G (normal) or 8-oxoG(damaged) template base. These results demonstrate the capability to rationally design increases in pol processivity and catalytic efficiency through computational DNA binding predictions and the addition of non-specific DNA binding domains.

As we study and engineer organisms, there is a demand for more sophisticated strategies for cloning, mutagenizing, labeling, detecting, and amplifying DNA. To meet this demand, there is an increasing need to modulate a range of DNA polymerase (pol) properties, beyond what exists in nature<sup>1,2</sup>. Processivity, the average number of bases a pol will extend before falling off a template, and catalytic properties, such as  $V_{max}$  and  $K_m$ , are useful properties to modulate. High processivity is important for efficiently copying simple repetitive sequences<sup>3,4</sup>, and high incorporation rates are important to be applicable for biotechnology applications<sup>5</sup>.

Y-family pols are a superfamily of evolutionarily related proteins that exist in cells to bypass DNA damage caused by both radiation and chemicals<sup>6</sup>. Dpo4 from *Sulfolobus solfataricus*, is the most characterized Y-family polymerase. However, Dbh (DinB homologue) from *S. acidocaldarius*, as a close relative of Dpo4, is particularly interesting for pol-based biotechnology applications. Dbh has a relatively high nucleotide misincorporation ratio (the enzyme exhibits misinsertion fidelities in the range of  $8 \times 10^{-3}$  to  $3 \times 10^{-4}$  on undamaged DNA templates, as determined by steady-state kinetic analysis) that could be useful for mutagenesis at mesophilic conditions<sup>7–9</sup>. However, Y-family pols typically have low processivity (only able to polymerize a limited number of nucleotides during a single association-dissociation cycle) and activity/incorporation rate, limiting their usefulness *in vivo*

<sup>1</sup>Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois, United States of America. <sup>2</sup>School of Pharmaceutical Science, Jiangnan University, Wuxi, Jiangsu Province, China. <sup>3</sup>Interdisciplinary Biological Sciences Program, Northwestern University, Evanston, Illinois, United States of America. <sup>4</sup>Department of Genetics, Harvard Medical School, Boston, Massachusetts, United States of America. <sup>5</sup>Biophysics Program, Harvard University, Boston, Massachusetts, United States of America. <sup>6</sup>Wyss Institute, Harvard University, Boston, Massachusetts, United States of America. <sup>7</sup>Media Lab, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America. <sup>8</sup>Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America. <sup>9</sup>McGovern Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America. <sup>10</sup>Department of Physiology, Northwestern University, Chicago, Illinois, United States of America. <sup>11</sup>Department of Applied Mathematics, Northwestern University, Chicago, Illinois, United States of America. Correspondence and requests for materials should be addressed to K.E.J.T. (email: [k-tyo@northwestern.edu](mailto:k-tyo@northwestern.edu))

and *in vitro*<sup>10</sup>. Improving the processivity of Y-family pols, while increasing or maintaining catalytic activity, would enhance the utility of this pol for mutagenesis studies on difficult to replicate DNA.

Y-family pols have important structural differences compared to other pols that contribute to their low fidelity and processivity<sup>11–13</sup>. High processivity is an important attribute of all *replicative* pols, which have a typical overall structure shaped like a right hand, composed of thumb, finger, and palm domains<sup>14</sup>. The catalytic carboxylate-metal ion complex sits in the palm domain while the mobile finger and thumb domains grasp the template and primer to create an active site that aligns the template base and incoming nucleotide<sup>14,15</sup>. Y-family pols have similar structures to replicative pols, however the fingers and thumb domains are smaller than most other polymerases, resulting in virtually no contact with the major-groove side of the nascent base pair by the fingers, and the thumb making fewer contacts with both the DNA substrate and incoming nucleotide. These structural differences contribute to the decreased processivity and fidelity of Y-family pols compared to other pol families<sup>16</sup>. Furthermore, Y-family pols have an additional fourth domain called the “little finger” (LF) or “polymerase associated domain” (PAD), which helps grip on the DNA and has also been associated with the family’s low fidelity and translesion replication capabilities<sup>17–19</sup>.

Thus, compared to the catalytic process of high-processivity pols, the Y-family pols impose relatively few constraints on the nascent base pair and DNA substrates<sup>13,17</sup>. This lack of constraints results in simultaneously reduced fidelity and processivity. However, structure may not be the only determinant of processivity. Increased binding affinity of pols to DNA would likely increase processivity, as the pol is less likely to fall off the DNA. In this scenario, the relaxed active site would still readily accommodate misincorporation, but tighter binding might improve processivity.

Several previous studies have shown that altering DNA binding properties, through adding DNA binding domains and introducing mutations, can affect processivity and catalytic activity<sup>20–23</sup>. For example, the dsDNA binding domain, Sso7d (7 kDa) from *S. solfataricus*, binds to DNA in a sequence-independent manner. The processivity of both A-family and B-family polymerases has been shown to be enhanced by fusing the polymerases with Sso7d<sup>24</sup>.

Certain mutations are also known to affect processivity and catalytic efficiency of Y-family pols<sup>25</sup>. For example, a non-conserved residue Arg332 within the LF domain has been shown to be responsible for DNA binding and DNA translocation after phosphodiester bond formation past damaged DNA(8-oxoG) in Dpo4<sup>26</sup>. The Ala332 and Glu332 mutants were each 2-fold faster at full-length extension opposite unmodified DNA than WT Dpo4<sup>26</sup>. Likewise, increasing glycine content in the linker region between the thumb and LF domain of Dpo4 was shown to decrease DNA binding affinity by 250-fold compared to the wild-type<sup>27</sup>.

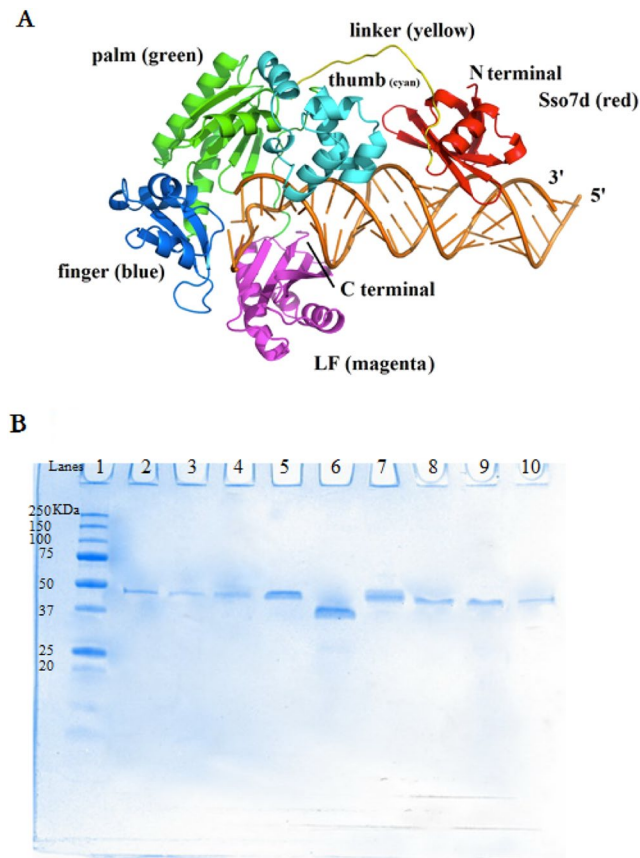
In the present study, we evaluated the effect of pol/DNA binding on processivity and catalytic activity for Dbh by (a) tethering the pol to a non-specific DNA binding domain and (b) introducing point mutations in the pol that increase the binding to DNA. In both cases, we anticipated that increasing DNA binding would increase processivity. We attached Sso7d to the N-terminus of Dbh and introduced mutations that were computationally predicted to strengthen binding free energy ( $\Delta G_{\text{binding}}$ ). Both strategies resulted in increased processivity during extension of a single-stranded DNA (ssDNA) template, M13mp18. The addition of Sso7d did not change kinetic properties, and several computationally guided mutations improved  $k_{\text{cat}}$ . This is a step toward the broader utility of Dbh in biotechnology.

## Results

**Generation of Sso7d-Dbh fusion protein (Sdbh).** Sso7d was covalently linked to Dbh to investigate the role that Sso7d could play on the processivity of Y-family pols. Because steric constraints between Sso7d and Dbh could prohibit Sso7d from interacting with the DNA template, we used a structure-guided approach to design an appropriate linker. Using the published 3D structures of Dbh and Sso7d, we simulated the binary complexes of Sdbh-DNA. A flexible linker SS(GGGGS)<sub>3</sub>GM was found to tether Sso7d to the palm domain of Dbh while maintaining protein-DNA interactions for both Dbh and Sso7d (Fig. 1A). Sso7d was far away from the active site of Dbh on the spatial structure and was predicted to only slightly change the original conformation of Dbh. The flexible linker was predicted to have unconstrained flexibility (due to the small, non-polar amino acid, Gly) and favorable water interactions (due to Ser)<sup>28</sup>. The designed construct was assembled, expressed, and purified (Fig. 1B). A native Dbh was also expressed for comparison.

**Sso7d fusion increased the processivity of Dbh.** Processivity was evaluated with the average extension length of products in the presence of excess herring sperm DNA as a trap which limited the pol from rebinding the DNA. Single-stranded M13mp18 was used as the template to estimate the processivity of native Dbh and fusion Sdbh (Fig. 2). The concentration of primer-annealed template (P-T) was fixed at 12.5 nM, and the amount of enzyme varied from 0 to 250 nM. When the pol was limiting (5 nM) and only one extension per DNA strand was expected, only a small fraction of FAM-labeled primers were extended, and the processivity of Dbh was estimated to be about 10 nt (the length of apparent FAM-labeled primer was 80 nt because the FAM fluorophore alters the electrophoretic properties). As expected, when the pol was not limiting (250 nM enzyme, 20-fold enzyme to P-T), Dbh synthesized replication products of several hundred bases in length, while at an equimolar pol-to-P-T ratio (12.5 nM enzyme), the processivity was about 30 nucleotides (Fig. 2), implying DNA molecules were extended multiple times.

Under the same assay conditions, when the Sso7d-fused pol was limiting (5 nM) and single extension events were favored, processivity of Sdbh increased to approximately 40 nt compared to 10 nt for Dbh (Fig. 2). When Sdbh was in a large molar excess and allowed multiple extension events per DNA strand (20-fold enzyme to P-T) the longest products were comparable to Dbh, but the 200 to 300 nt products were more intense. When the Sdbh was 12.5 nM, the processivity of Sdbh were approximately 2 fold higher than Dbh. Based upon the limiting pol conditions, we estimate that Sdbh incorporates 4-fold more nucleotides per extension than WTDbh. This

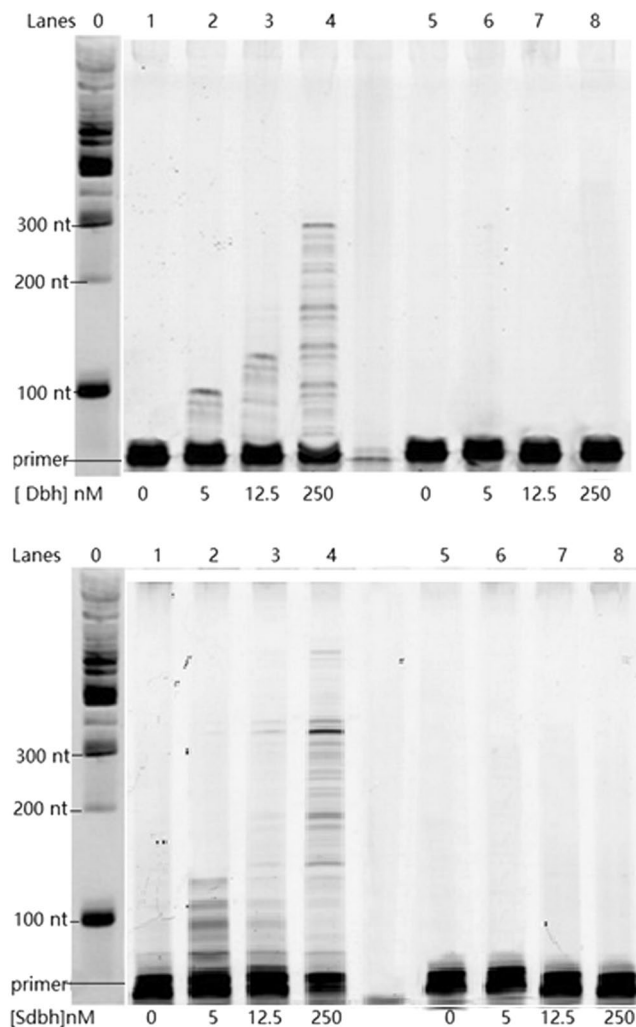


**Figure 1.** (A) A ribbon diagram of Dbh complexed with Sso7d and DNA. The location of the four structural domains in Dbh are color coded as follows: green, palm domain; cyan, thumb domain; blue, finger domain; magenta, little finger (LF) domain; red, Sso7d. Sso7d is located at the N-terminus and is covalently linked to the palm domain. Sso7d binds to DNA by placing a triple-stranded beta-sheet across the DNA minor groove. The amino acid sequence of the flexible linker (yellow) is SS(GGGGS)<sub>3</sub>GM. The DNA substrate is shown in brown. (B) SDS-PAGE of the purified Dbh and the variants. Samples were separated on a 10% Mini-PROTEAN TGXTM precast gel (Bio-Rad) and proteins were visualized after staining with Coomassie Brilliant Blue R-250. Lane 1, protein marker (protein molecular weight (kDa) was showed on the left of the lane 1); Lane2-10, SdbhM76I, SdbhKSKIP241–245RVRKS, SdbhL250V, Sdbh, Dbh, SdbhT37F, SdbhA221S, SdbhI62V and SdbhK337R, respectively.

is consistent with the idea of Sso7d as a processivity enhancer that helps maintain the LF domain and thumb domain around the DNA near the extending 3' terminus (Fig. 1A).

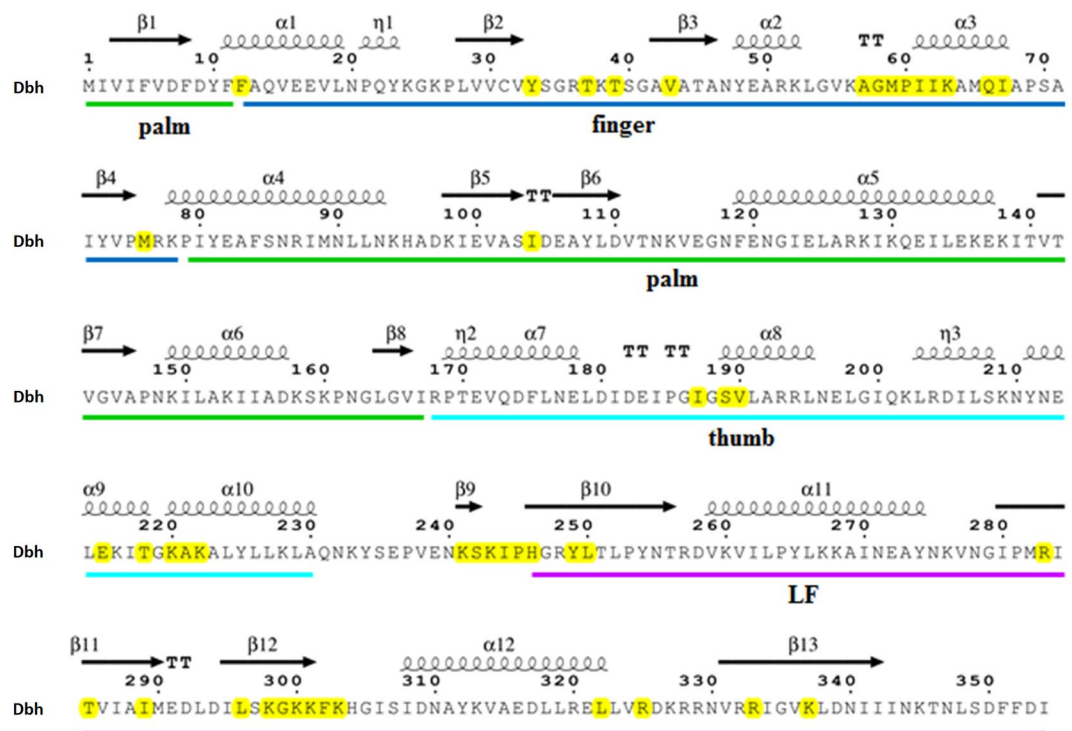
**Identifying mutations in Dbh that increase DNA binding strength computationally.** In order to determine the residues of Dbh that were most amenable to mutations, a multiple sequence alignment of known Y-family pols was used to identify conserved and non-conserved positions. 34 homologous sequences from Archaea, bacteria and eukaryotes with greater than 40% identity were generated. Since the residues located within 6 Å of the DNA duplex are generally considered important for the pol/DNA interaction<sup>29</sup>, the Dbh-DNA complex was analyzed to find the residues located within 6 Å of the double-stranded DNA. As shown in Fig. 3, 72 non-conserved residues in Dbh were found to be potential DNA contacts. Moreover, a library of potential mutation sites with mutation frequency (Table 1) was established based on the multiple sequence alignment and statistical analysis. The potential mutation sites were spread among the finger domain, thumb domain, the β-sheet of the LF domain, and the linker region of the thumb and LF domain. In addition, the contiguous residues K241, S242, K243, I244 and P245 had a high frequency mutation to RVRKS in several species (Table 1, R/24, V/19, R/25, K/25, S/18). These residues are in an interdomain linker that is a key determinant of pol conformation<sup>30</sup>. For each residue, a mutation to the highest frequency amino acid in the multiple sequence alignment was chosen. The mutations T37F, I62V, M76I, A221S, Y249I, L250V, K337R and KSKIP(241–245)RVRKS were made computationally, and alterations to the protein structure were calculated using Modeller 9.11.

Free binding energies of DNA to Dbh or the predicted mutants were estimated using the MM-PBSA method as described in Material and Methods. The predictions are listed in Table 2. All mutations except K337R and Y249I showed more favorable binding to the DNA template strand. With reference to native Dbh, the binding energy of the mutant KSKIP(241–245)RVRKS strengthened by - 532.2 kcal/mol, while for mutants K337R and Y249I, the binding energy weakened by 42.4 and 293.6 kcal/mol, respectively.



**Figure 2.** Processivity of Dbh and Sdbh. The ability of Dbh and Sdbh to extend a FAM-labeled primer (FAM-36 primer) annealed to single-stranded M13mp18 DNA from a single DNA binding event in the presence of a DNA trap. In lanes 1–4, the pol was preincubated with the primer-template DNA (12.5 nM) and the reactions were initiated by the addition of dNTPs (200  $\mu$ M) and excess sperm DNA trap. As a control (lanes 5–8), the pol was preincubated with the primer-template DNA and the herring sperm DNA trap, and reactions were initiated by the addition of dNTPs. The intensity along the length of the lane was quantified and then the marker standards were used to assign a basepair length through the lanes using ImageQuant software. From there the greatest abundant product could be measured as the processivity. The unlabeled marker standards were visualized by SYBR stain independently and showed the length (nt) on the left of the first lane. (Note: the synthesized primer was 36 nt, but the FAM label altered its electrophoretic mobility, such that it ran at about 80 nt). See Supplemental S1 for full-length image.

**Experimental determination of the equilibrium dissociation constant ( $K_d$ ).** To validate the computational prediction of the binding affinity of Dbh and the variants to template DNA, eight Sdbh variants (SdbhT37F, SdbhI62V, SdbhM76I, SdbhA221S, SdbhKSKIP(241–245)RVRKS, SdbhY249I, SdbhL250V, and SdbhK337R) were constructed, expressed and purified, and equilibrium titrations with a 2AP-P/T were performed, using 2AP attached to the primer as the probe. Excitation of 2AP-P/T at 315 nm can minimize protein absorption. The final equilibrium constant was obtained by subtracting the fluorescence of protein. Figure 4 shows a typical titration curve of the 2AP-P/T molecule with Dbh and the variants. The binding isotherms were fit to a hyperbola to calculate  $K_d$  values. As shown in Table 2, the measured  $K_d$  values for Dbh, Sdbh, SdbhM76I and SdbhA221S were  $54.8 \pm 3.8$ ,  $51.2 \pm 2.4$ ,  $42.8 \pm 3.1$  and  $42.5 \pm 3.0$  nM, respectively. The binding of Sdbh, SdbhM76I and SdbhA221S to the 2AP-P/T was a little stronger than that of WTDbh. However, the  $K_d$  value for SdbhKSKIP241–245RVRKS was  $34.8 \pm 1.2$  nM, which indicated a significant increase in the affinity. This result is in agreement with that of the above binding energy calculation. Conversely, SdbhT37F, SdbhI62V and SdbhL250V showed almost identical affinity to that of Sdbh, while SdbhK337R and SdbhY249I showed less affinity than that of Sdbh. Thus variants SdbhKSKIP241–245RVRKS and SdbhM76I were subjected to further processivity analysis.



**Figure 3.** Non-conserved amino acids likely to interact with the DNA substrate. The Dbh amino acid sequence is shown with the domains indicated by color: palm (green), finger (blue), thumb (cyan), and LF (magenta). Non-conserved amino acids key to the interaction of Dbh and DNA duplex are highlighted in yellow. The secondary structures are indicated as coils ( $\alpha$ -helices) and arrows ( $\beta$ -sheets) above the aligned primary amino acid sequence.

**Dbh variants with higher affinity can improve processivity.** The processivity of the variants SdbhKSKIP241–245RVRKS and SdbhM76I were investigated using the same methods previously applied to Dbh and Sdbh. As shown in Fig. 5, when the pol was limiting, favoring single extensions (5 nM), the processivity of SdbhKSKIP(241–245)RVRKS and SdbhM76I were 60 and 100 nt, respectively, compared to 40 nt for Sdbh and 10 nt for Dbh. When the enzymes were 12.5 nM, SdbhKSKIP(241–245)RVRKS and SdbhM76I showed a similar processivity of about 140 nt, higher than that of Sdbh (60 nt). These results show that mutating KSKIP(241–245) and M76 can increase processivity beyond the improvements made by Sso7d alone. These results also demonstrate that increasing DNA binding strength can improve the processivity of Dbh.

**Steady-state kinetic analysis of one-base insertion.** Velocity was studied as the activity of Dbh and the variants on the incorporation of dCTP (correct) and dATP (incorrect) opposite the G (normal) or 8-oxoG (damaged) template base. The polymerase assays and Michaelis–Menten plots were shown in Figure S2, from which  $V_{\max}$  and  $K_m$  values were determined (Table 3). Data in Table 3 showed that the  $V_{\max}$  for correct dCTP incorporation opposite the normal G base by SdbhKSKIP241–245RVRKS was higher than that of WTDbh (~2.3-fold), while the  $V_{\max}$  for correct dCTP incorporation opposite the damaged 8-oxoG base was 2-fold higher than that of WTDbh. The  $V_{\max}$  for incorrect dATP incorporation opposite the normal G base by SdbhKSKIP241–245RVRKS was ~8.4-fold higher than that of WTDbh, and dATP incorporation opposite the damaged 8-oxoG base was also higher (2.8-fold). The  $K_m$  values for correct dCTP incorporation and incorrect dATP incorporation opposite the normal G/damaged 8-oxoG base by SdbhKSKIP241–245RVRKS were all lower than that of WTDbh. Interestingly, the dATP misinsertion opposite 8-oxoG was more rapid than that opposite G by not only WTDbh but also Sdbh, SdbhM76I and SdbhKSKIP241–245RVRKS. These results indicated that increasing affinity can improve the activity of Dbh.

## Discussion

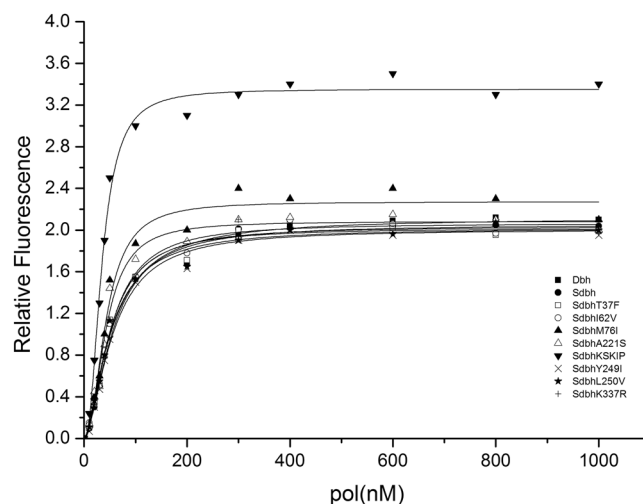
Dbh is an *archaeal* representative of Y-family pols and exhibits a unique structure and set of properties (e.g. low processivity and fidelity)<sup>13,17</sup>. Compared with classical pols, the Dbh active site is wider with fewer contacts to the DNA template and nucleotide substrates, which is useful for incorporating non-natural nucleotides. Our work was based on two observations: pols often require accessory proteins such as proliferating cell nuclear antigen (PCNA) to modify processivity<sup>21</sup>, and the residues present at the non-conserved positions of processive pols like *E. coli* Pol IV result in tighter binding and enhanced processivity<sup>31</sup>. Building on these observations, we show that enhancing interaction by fusing the processivity enhancer Sso7d to Dbh polymerase increases the nonspecific interaction with dsDNA, improving the processivity of the native Dbh. Further, creating bioinformatics-based

Position sites	Original	Mutant	Position sites	Original	Mutant
	AA	AA/frequency <sup>(a)</sup>		AA	AA/frequency <sup>(a)</sup>
12	F	Y/23, A/1	222	K	R/2
33	Y	F/18	241	K	R/24
37	T	F/16, N/3, P2, R1, S/1, E/1	242	S	V/19, T/1, Q/1, I/2, E/5, K/1
39	T	D/20, T/8, P7, K/2, N/4, R/2, S2	243	K	R/25, I/1, V/2
43	V	I/5	244	I	K/25, V/2, T/1, S/1, H/1
57	A	S/19	245	P	S/18, H/6, N/3
58	G	A/11	246	H	I/19, R/2, Q/1, M/1, F/2, Y/1
59	M	I/14	249	Y	I/17
60	P	A/1, S7	250	L	V/20
61	I	A8, C2, S1, L/3	283	R	A/14, K/1, S/1, T/4, E/4
62	I	V/14, P/6, S/2, Y/1, K/1	285	T	H/14, Y/5, S/3, A/3
63	K	R/3, E/11, Q/2	289	I	V/15
66	Q	E/12, K/13, N/1	296	L	V/26, I/2
67	I	L/3	298	K	R/23
76	M	I/7	299	G	S/5, E/3
104	I	V/1	300	K	R/15, Y/2, I/1
187	I	V/12	301	K	T/23, S/4
189	S	N/10, K/5, D/7, E/2, N/10, T/1	302	F	Y/2
190	V	I/15, M/1, S/4	303	K	P/12, T/6, N/2, G/3
215	E	K/20, V/2, S/1, A/1, I/2	322	L	I/25
218	T	I/20, V/5, L/3	325	R	E/19, K/3, S/2, A/1
220	K	R/2, E/16	333	R	E/1, A/1
221	A	S/9, K/1	337	K	R/28

**Table 1.** The mutant residue types and frequencies of the non-conserved amino acids (AA). <sup>(a)</sup>The mutant residue types and frequencies were statistically analyzed using HotSpot Wizard 1.7 software based on the results of the multiple sequence alignment.

Mutation sites	Primer/Template DNA <sup>(a)</sup>	Free binding energy $\Delta G_{\text{bind}}$ (kcal/mol)	$(\Delta G_{\text{bind}})_{\text{mutant}} - (\Delta G_{\text{bind}})_{\text{wt}}$ (kcal/mol)	Variants	Equilibrium dissociation constant $K_d$ (nM)
WTDbh	P	-2251.8	0	WTDbh	54.8 ± 3.8
	T	-3465.7		Sdbh	51.2 ± 2.4
KSKIP(241-245) RVRKS	P	-2579.5	-532.2	SdbhKSKIP(241-245) RVRKS	34.8 ± 1.2
	T	-3670.2			
M76I	P	-2414.7	-380	SdbhM76I	42.8 ± 3.1
	T	-3682.8			
L250V	P	-2422.1	-318.3	SdbhL250V	50.4 ± 3.4
	T	-3603.7			
T37F	P	-2313.6	-159	SdbhT37F	49.7 ± 3.1
	T	-3562.9			
A221S	P	-2344.03	-100.45	SdbhA221S	47.5 ± 3.0
	T	-3473.91			
I62V	P	-2145.6	-60.3	SdbhI62V	49.8 ± 2.6
	T	-3632.2			
K337R	P	-2554.6	42.4	SdbhK337R	55.5 ± 4.1
	T	-3120.5			
Y249I	P	-2305.0	293.6	SdbhY249I	56.2 ± 3.4
	T	-3118.9			

**Table 2.** Binding free energies and equilibrium dissociation constant of Dbh and the mutants to primer/template DNA. <sup>(a)</sup>dsDNA (Primer/Template DNA) referred to the chain primer extended (P(5'-3')) and the corresponding template chain (T(3'-5')).



**Figure 4.** Equilibrium titrations of DNA substrate with Dbh and the variants. A constant amount of 2AP-P/T (100 nM) was titrated with increasing concentration of pol. The fluorescence was excited at 315 nm and observed at 370 nm. Each measurement was repeated four times, and the average value of the fluorescence intensity was recorded. A control experiment was performed with the non-fluorescent promoter DNAs under identical conditions. The fluorescence changes from the control experiments were subtracted from the data obtained with the 2AP-P/T, and the corrected values are plotted against [pol]. The analysis of the data yielded the dissociation constant  $K_d$  for Dbh (■), Sdbh (●), SdbhT37F (□), SdbhI62V (○), SdbhM76I (▲), SdbhA221S (△), SdbhKSKIP (▼), SdbhY249I (×) and SdbhL250V (★), respectively.

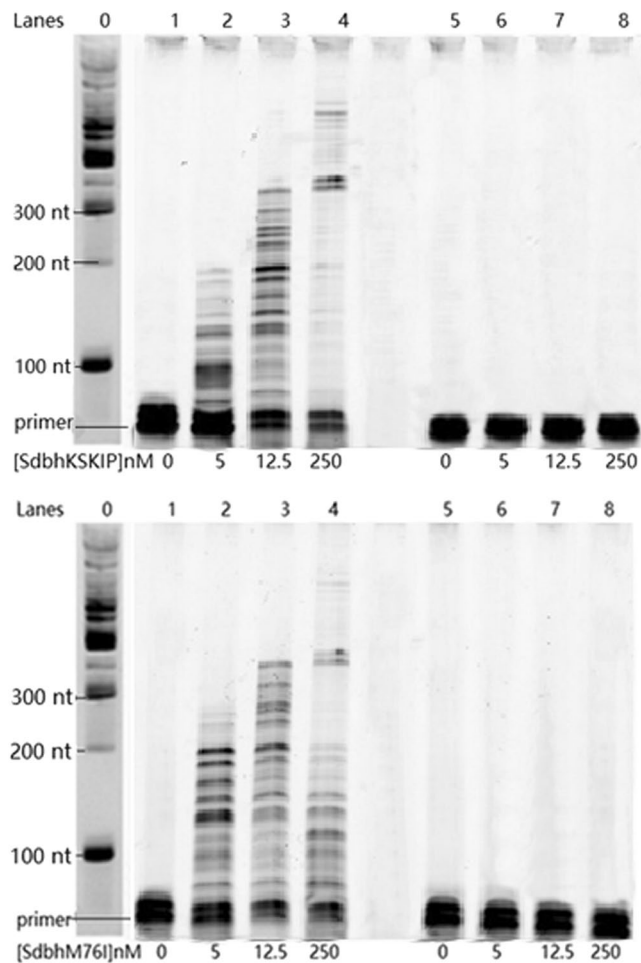
mutations in non-conserved residues increased the DNA template binding ability and processivity of the Sso7d-Dbh fusion. We believe rational modifications and computational predictions such as these could allow for precise control over a range of DNA polymerase properties.

To understand the underlying elements that affect pols binding to their substrates, a large number of structures of pol-DNA complexes have been determined<sup>32,33</sup>. These studies have shown that the pols share special substructures to recognize DNA, including  $\beta$ -sheets,  $\alpha$ -helices, and loops. Forces like hydrophobic, van der Waals, and ionic interactions are attributed with stabilizing the pol-DNA complexes. Hydrogen bonding interactions between the pol and DNA bases in the major groove can mediate the binding specificity of the pol. When the MM-PBSA method was used to calculate the binding energy of the pol-DNA complex in this study, the intramolecular electrostatic and van der Waals, electrostatic and nonpolar contributions in liquid phase were the main concern. Other factors, e.g. non-electrostatic and non-specific components, that also affect DNA binding were ignored. This probably led to an inconsistency between the theoretical and the experimental data, although the calculated affinities in the present study are in agreement with the experimental  $K_d$ . The mutant KSKIP(241–245) RVRKS showed the highest free energy ( $-532$  kcal/mol) computationally, while the variant SdbhKSKIP(241–245) RVRKS showed the highest affinity ( $34.8 \pm 1.2$  nM) experimentally (Table 2).

However, the processivity of SdbhKSKIP(241–245)RVRKS was not the highest, which indicated the affinity is not the only factor that determines the processivity. Processivity of the DNA pols relates to the number of nucleotides added to the nascent strand during one round of binding and dissociation from the primer template. It is essential for processivity that the pols are binding to and sliding along the DNA. Tight binding of a pol to its DNA template is achieved through a large interaction surface, e.g. sequence-specific complexes which display complementarity in shape and polarity. For sliding, however, an enzyme must strike the right energetic balance so as to remain associated with its polymeric substrate, while retaining the ability to move from site to site. SdbhKSKIP(241–245)RVRKS exhibited high affinity, which probably hindered the pol from sliding on the template, and thus decreased the processivity.

Previous steady-state kinetics with Dpo4 showed 90-fold higher incorporation efficiency of dCTP over dATP opposite 8-oxoG, and also faster rates of dCTP incorporation opposite 8-oxoG compared to G<sup>34</sup>. In comparison to Dpo4, Dbh in this study also showed higher reaction velocities during dCTP incorporation vs. dATP opposite both G and 8-oxoG, and so did variants Sdbh, SdbhM76I and SdbhKSKIP 241–245RVRKS (Table 3). Moreover, the variant SdbhKSKIP 241–245RVRKS with the highest binding activity to DNA among Dbh, Sdbh, SdbhM76I and SdbhKSKIP 241–245RVRKS, showed the highest  $V_{max}/K_m$  value both on dCTP incorporation opposite G/8-oxoG and on dATP incorporation opposite G/8-oxoG. These results further demonstrated that increasing binding strength can improve the activity of Dbh. However, the correlation between processivity and activity will be considered in greater detail elsewhere.

As applications requiring DNA polymerases to perform non-natural tasks increase, there will be a need for a molecular toolbox to tailor polymerase properties for different engineering applications. The current study shows two strategies for modifying DNA polymerase processivity. The use of protein fusions in tandem with computationally predicted mutations that improved binding affinity enhanced both processivity and catalytic efficiency in some cases. The approaches used here are likely generalizable to many different pols to increase processivity.



**Figure 5.** Processivity of Sdbh variants. The ability of the variants SdbhKSKIP241–245RVRKS and SdbhM76I to extend a FAM-labeled primer annealed to single-stranded M13mp18 DNA from a single DNA binding event in the presence of a DNA trap. The analysis method employed was the same as that used in Fig. 2. See Supplemental S1 for full-length image.

pol	Template base	dNTP	$K_m$ ( $\mu\text{M}$ ) <sup>(a)</sup>	$V_{\max}$ ( $\text{nMmin}^{-1}$ ) <sup>(a)</sup>	$V_{\max}/K_m$ (min)
Dbh	G	dCTP	$88.2 \pm 12.0$	$1.0 \pm 0.1$	11
		dATP	$215.2 \pm 18.7$	$0.019 \pm 0.001$	0.088
	8-oxoG	dCTP	$67.7 \pm 6.3$	$0.5 \pm 0.02$	7.4
		dATP	$139.8 \pm 3.1$	$0.18 \pm 0.002$	1.3
Sdbh	G	dCTP	$75.2 \pm 12.0$	$1.4 \pm 0.1$	19
		dATP	$128.4 \pm 26.3$	$0.04 \pm 0.003$	0.31
	8-oxoG	dCTP	$59.6 \pm 11.1$	$0.61 \pm 0.07$	10
		dATP	$114.9 \pm 27.8$	$0.23 \pm 0.009$	2.0
SdbhM76I	G	dCTP	$54.9 \pm 8.6$	$1.9 \pm 0.2$	35
		dATP	$76.7 \pm 7.1$	$0.08 \pm 0.01$	0.10
	8-oxoG	dCTP	$55.5 \pm 4.9$	$1.0 \pm 0.05$	18
		dATP	$69.1 \pm 9.3$	$0.4 \pm 0.05$	5.8
SdbhKSKIP 241–245RVRKS	G	dCTP	$40.1 \pm 5.2$	$2.3 \pm 0.1$	57
		dATP	$71.7 \pm 9.6$	$0.16 \pm 0.02$	0.22
	8-oxoG	dCTP	$33.5 \pm 2.7$	$1.0 \pm 0.04$	30
		dATP	$55.1 \pm 2.9$	$0.5 \pm 0.04$	9.1

**Table 3.** Steady-state kinetics of incorporation of dCTP and dATP opposite G and 8-oxoG bases by Dbh and the variants. <sup>(a)</sup>The values of  $V_{\max}$  and  $K_m$  were fit by nonlinear regression of the Michaelis-Menten equation. All reactions were done in duplicate, and the indicated data points are shown as the mean  $\pm$  standard deviation.



Oligo name	Oligo sequence (5'-3')
Fdbh	GGAATTCATATGCATCACCATCACCATCACATGATAGTGATATTCGTTGATT
Rdbh	CGCGGATCCTTAAATGTCGAAGAAATCAGAT
SLD	TCATCCGGTGGGGGAGGCTCTGGTGGTGGTGGTCTGGTGGTGGTGGTCTGGGATG
Fdbh-linker	TCATCCGGTGGGGGAGGCTCTGGTGGTGGTGGTCTGGTGGTGGTGGTCTGGGATGATGATAGTGATATTCGTTGA
Rdbh-dhfr	TAAAGCCTCTGCAGGTTAACCTTACTCGAGAATCCCGGGATCC
	GTGATGGTGGTGGTGGTGGTAAATGTCGAAGAAATCAGAT
Fsso-dhfr	GTCTAGAAATAATTTTGTTAACTTAAAGAAGGAGATATACATATGGCAACCGTAAAGTTCAAGTA
Rsso-linker	CATCCAGAACCCACCACCAGAACCCACCACCAGAGCCTCCCCACCGGATGACTTTTTCTGCTTCCAGC
FAM-36-primer	FAM-ACG CCT GTA GCA TTC CAC AGA CAA CCC TCA TAG TTA

**Table 4.** Oligonucleotides used in this study.

## Materials and Methods

**Construction of the Sso7d-dbh fusion (Sdbh).** The Dbh gene sequence (<https://polbase.neb.com/polymerases/140-dbh>) was synthesized by Integrated DNA Technologies Inc. (IDT) with a C-terminal 6× histidine tag. Previous work has shown that this tag does not significantly alter the polymerization activity of Y-family pols<sup>35</sup>. The *Sso7d* gene was synthesized by IDT based on the published amino acid sequence<sup>24,36</sup>. Gibson assembly was employed to fuse *Sso7d* to the N-terminus of Dbh by a flexible linker (SLD, Table 4)<sup>37</sup>. The overlapping fragments of Dbh and *Sso7d* were PCR-amplified by oligonucleotides Fdbh-linker, Rdbh-dhfr, Fsso-dhfr and Rsso-linker, respectively (Table 4). The high copy plasmid DHFR (dihydrofolatereductase) supplied with PURExpress *In Vitro* Protein Synthesis Kit (New England Biolabs, Inc. (NEB)) was linearized by digestion with *NdeI* and *BamHI* restriction enzymes. The recombinant plasmids DHFR-dbh and DHFR-Sdbh were consequently constructed according to the Gibson assembly protocol (NEB) and verified by DNA sequencing.

**Protein expression and purification.** The complete plasmids DHFR-dbh and DHFR-Sdbh were then transformed into *E. coli* strain BL21 (λDE3) (Invitrogen) for induced expression as described previously<sup>13</sup>. WT Dbh and fused Sdbh contained a 6× histidine tag on the N-terminus and were purified to electrophoretic homogeneity using nickel-nitrilotriacetic acid (Ni-NTA) permeation and Mono-S chromatography. Purified Dbh and Sdbh were stored in small aliquots at −80 °C in 50 mM Tris-HCl buffer (pH 7.7 at 22 °C) containing 50 mM NaCl, 1 mM dithiothreitol, and 50% glycerol (v/v).

**Mutation site identification and 3D-model determination.** Using the Dbh amino acid sequence as a template, all DNA pols with over 40% homology were collected by BLAST from the NCBI database. A multiple sequence alignment was generated using Position Specific Iterated Blast (PSI-BLAST) and conservative sequence analysis was carried out on the alignment. The Dpo4 crystal structure (PDB: 4NLG) which shared a 99% sequence similarity with Dbh was used as a template for the following modeling and computation. A list of potential mutation sites was generated based on residues that were non-conserved and located within 6 Å of the DNA strand. The mutant residue types and frequencies were statistically analyzed using HotSpot Wizard 1.7 software (<http://loschmidt.chemi.muni.cz/hotspotwizard/ActionServlet?action=protein&type=enzyme>) based on the results of the multiple sequence alignment<sup>38</sup>. The 3D structures of mutant enzymes were constructed using the homology modeling program, Modeller 9.11, and then aligned to the selected model structure complex using the PyMol align tool (<http://www.pymol.org>)<sup>39</sup>. The binary complex of the mutant enzyme with the substrate was generated and exported by PyMol.

**Molecular Dynamics (MD) simulation and binding energy calculation.** The Molecular Dynamics (MD) simulations, including energy minimization, system equilibration and production protocols, were performed with the GROMACS 4.5.5 package as described previously<sup>40,41</sup>. Each enzyme-substrate complex was placed in a cubic box after adding GROMOS 9643a1 position, and filled with atomistic TIP3P water. A two-step energy minimization process was performed after system equilibration. The Molecular Mechanics Poisson-Boltzmann surface area (MM-PBSA) method, which has been widely used to predict the binding affinities for a variety of macromolecular complexes and protein-ligand complexes<sup>41-43</sup>, was employed to calculate the binding energy of different pol-DNA template complexes (DNA template in PDB: 4NLG is 5'-GAAGCCGGCGGAA-3'). In this study, the free energy of each molecule is defined as follows:  $\Delta G_{binding} = G_{complex} - (G_{protein} + G_{ligand})$ . Here,  $G_{complex}$ ,  $G_{protein}$ , and  $G_{ligand}$  are the free energy of the polymerase-DNA complex, the free energy of the polymerase, and the free energy of DNA, respectively. The free energy  $G$  can be calculated by the following scheme, based on the MM-PBSA method:  $G = E_{MM} + TS_{MM} + G_{solv}$ , where  $E_{MM}$  is comprised of the intramolecular electrostatic ( $E_{elec}$ ) and van der Waals ( $E_{vdw}$ ) interaction energies. The free energy of solvation,  $G_{solv}$ , was approximated as the sum of electrostatic and nonpolar contributions in liquid phase.  $TS_{MM}$  was ignored as it did not contribute significantly to the binding energy in these conditions. All binding energies were calculated by Gromacs 4.5.5 combined with the *g\_mmpbsa* program developed by Kumari *et al.*<sup>44</sup>.

**Site-directed mutagenesis of Sdbh.** Guided by the computational predictions, the mutations with decreased binding energy were subjected to site-directed mutagenesis. Mutagenesis was performed using a QuikChange site-directed mutagenesis kit by Bio Basic Inc. (Ontario, Canada), and sequences were verified prior to bacterial expression. Mutant proteins were expressed, purified, and stored using the same procedure as for the WT enzyme. These Sdbh variants were then used to verify the residues in Dbh that were most critical for binding to the DNA template.

**Measuring binding of Dbh and the mutants.** Fluorescence titrations were performed to determine the equilibrium dissociation constant ( $K_d$ ) of Dbh and the mutants as previously described<sup>45, 46</sup>. Primer-template DNA with 2-aminopurine (2AP) located at the primer terminus (2AP-P/T) was prepared by Sangon Biotech (Shanghai, China). This 2AP-P/T was excited at 315 nm and the emission was observed at 370 nm. Fluorimetric titration experiments were performed on a Perkin Elmer LS50B Luminescence Spectrometer. A constant amount of 2AP-P/T (100 nM) was titrated against increasing concentration of Dbh and the mutants (0–1000 nM) in the reaction buffer (50 mM Tris acetate, pH 7.5, 50 mM sodium acetate, 10 mM magnesium acetate, 5 mM dithiothreitol, 0.05% Tween 20) at 25 °C. A control experiment was carried out in identical conditions with the presence of the unlabeled P/T (100 nM) and increasing amounts of the pols. The fluorescence changes from the control experiments were subtracted from the data obtained with the 2AP-P/T, and the corrected values were plotted against the corresponding pol concentration. The dissociation constant  $K_d$  was calculated using the following equation,  $F = \frac{F_{max} \times [pol]}{K_d + [pol]}$ , where F is the relative fluorescence intensity, and  $F_{max}$  is its maximum value.

**Primer and template DNA used in polymerase assays.** All DNA substrates used in the extension assays are listed in Table 4 and were synthesized from IDT. The primers used for processivity assays and steady-state kinetic analyses were synthesized with a 5'-6-carboxyfluorescein (FAM) label for fluorescence visualization. Circular single-stranded M13mp18 DNA was used as the template in processivity assays and purchased from Bayou Biolab. Primer was annealed to the template at a 1:1.5 ratio in an annealing buffer containing 10 mM HEPES-NaOH (pH 7.4) and 50 mM NaCl. The annealed complex was heated to 95 °C for 5 min, cooled to 60 °C at 0.1 °C/s, incubated at 60 °C for 10 min, cooled again to 4 °C at 0.1 °C/s, and then stored at –20 °C until use. The annealed primer-template was thawed on ice immediately before assays.

**Processivity assay.** The processivity assay was performed in the presence of a trap to prevent rebinding<sup>19, 35, 47, 48</sup>. The pol at concentrations from 5 nM to 250 nM was preincubated with the FAM-labeled primer-template (12.5 nM) in the reaction buffer (10 mM HEPES-NaOH (pH 7.4), 50 mM NaCl, 10 mM MgCl<sub>2</sub>, 1 mM DTT, 100 µg/mL BSA, and 0.1% Triton X-100). Reactions were initiated by adding 200 µM dNTPs and a 400-fold excess of herring sperm (Promega) as a trap. After incubation for 5 min at 37 °C, reactions were terminated by the addition of 10 µL of 95% formamide, 10 mM EDTA, and the reaction products were denatured at 100 °C for 5 min, then briefly chilled on ice. Reaction mixtures were separated via electrophoresis using 10% TBE-Urea precast polyacrylamide gels (Bio-Rad). The amounts of FAM fluorescence in the unextended and extended primer bands were quantitated using a Typhoon 9400 scanner and ImageQuant software (GE Healthcare). A distribution of DNA lengths were present, and processivity values were reported as approximately 85 percentile of the distribution.

**Steady-state kinetic analysis of one-base insertion.** Steady-state kinetic parameters were analyzed for incorporation of dCTP and dATP opposite the undamaged G or damaged 8-oxoG (7,8-dihydro-8-oxyguanine) template base, and assays were performed using established methods<sup>49–51</sup>. Specifically, each reaction contained 1.0 µM of the annealed primer-template (the primer was 5'-FAM-GGTTGGATGGTAG-3', the template was 5'-CTAACXCTACCATCCAACC-3', X represents oxoG/G). Dbh and the variants (20 nM) were preincubated with the primer-template in the buffer (40 mM Tris-HCl buffer (pH 7.5), 100 mM KCl, 5 mM MgCl<sub>2</sub>, 10 mM DTT, 0.1% Triton X-100, 50 µg/µl BSA). Reactions were initiated by the addition of varying concentrations of a single dCTP or dATP, incubated for 5 min at 37 °C, and then quenched by the addition of 10 µL of 95% formamide, 10 mM EDTA. Substrate and product DNA were then separated by electrophoresis on a 20% polyacrylamide (w/v) containing 7 M urea gel. Fluorescence in the substrate and product primer bands was scanned using the Typhoon 9400 scanner (GE Healthcare) and quantified by ImageQuant software. The reaction rates ( $v$ , nM/min) were plotted as a function of the dNTP concentration, and the data were fit by nonlinear regression of the Michaelis-Menten equation,  $v = \frac{V_{max} \times [dNTP]}{K_m + [dNTP]}$ , to calculate apparent  $K_m$  and  $V_{max}$  steady-state parameters.

## References

- Reha-Krantz, L. J. Recent patents of gene sequences relative to DNA polymerases. *Recent Pat DNA Gene Seq* 2, 145–163, doi:10.2174/187221508786241729 (2008).
- Ishino, S. & Ishino, Y. DNA polymerases as useful reagents for biotechnology - the history of developmental research in the field. *Front Microbiol* 5, 465, doi:10.3389/fmicb.2014.00465 (2014).
- Fazekas, A., Steeves, R. & Newmaster, S. Improving sequencing quality from PCR products containing long mononucleotide repeats. *Biotechniques* 48, 277–285, doi:10.2144/000113369 (2010).
- da Silva, E. F. & Reha-Krantz, L. J. Dinucleotide repeat expansion catalyzed by bacteriophage T4 DNA polymerase *in vitro*. *J Biol Chem* 275, 31528–31535, doi:10.1074/jbc.M004594200 (2000).
- Collins, C. H., Yokobayashi, Y., Umeno, D. & Arnold, F. H. Engineering proteins that bind, move, make and break DNA. *Curr Opin Biotechnol* 14, 371–378, doi:10.1016/S0958-1669(03)00091-0 (2003).
- Lee, C. H., Chandani, S. & Loechler, E. L. Homology modeling of four Y-family, lesion-bypass DNA polymerases: the case that *E. coli* Pol IV and human Pol kappa are orthologs, and *E. coli* Pol V and human Pol eta are orthologs. *J Mol Graph Model* 25, 87–102, doi:10.1016/j.jmglm.2005.10.009 (2006).
- Zamft, B. M. *et al.* Measuring cation dependent DNA polymerase fidelity landscapes by deep sequencing. *PLoS One* 7, e43876, doi:10.1371/journal.pone.0043876 (2012).
- Glaser, J. I. *et al.* Statistical analysis of molecular signal recording. *PLoS Comput Biol* 9, e1003145, doi:10.1371/journal.pcbi.1003145 (2013).
- Schwartz, J. J. & Quake, S. R. Single molecule measurement of the “speed limit” of DNA polymerase. *Proc Natl Acad Sci USA* 106, 20294–20299, doi:10.1073/pnas.0907404106 (2009).
- DeLucia, A. M., Grindley, N. D. & Joyce, C. M. An error-prone family Y DNA polymerase (DinB homolog from *Sulfolobus solfataricus*) uses a 'steric gate' residue for discrimination against ribonucleotides. *Nucleic Acids Res* 31, 4129–4137, doi:10.1093/nar/gkg417 (2003).

11. Pata, J. D. Structural diversity of the Y-family DNA polymerases. *Biochim Biophys Acta* **1804**, 1124–1135, doi:[10.1016/j.bbapap.2010.01.020](https://doi.org/10.1016/j.bbapap.2010.01.020) (2010).
12. Ling, H., Boudsocq, F., Woodgate, R. & Yang, W. Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell* **107**, 91–102, doi:[10.1016/S0092-8674\(01\)00515-3](https://doi.org/10.1016/S0092-8674(01)00515-3) (2001).
13. Silvian, L. F., Toth, E. A., Pham, P., Goodman, M. F. & Ellenberger, T. Crystal structure of a DinB family error-prone DNA polymerase from *Sulfolobus solfataricus*. *Nat Struct Biol* **8**, 984–989, doi:[10.1038/nsb1101-984](https://doi.org/10.1038/nsb1101-984) (2001).
14. Johnson, S. J., Taylor, J. S. & Beese, L. S. Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. *Proc Natl Acad Sci USA* **100**, 3895–3900, doi:[10.1073/pnas.0630532100](https://doi.org/10.1073/pnas.0630532100) (2003).
15. Broyde, S., Wang, L., Rechko, O., Geacintov, N. E. & Patel, D. J. Lesion processing: high-fidelity versus lesion-bypass DNA polymerases. *Trends Biochem Sci* **33**, 209–219, doi:[10.1016/j.tibs.2008.02.004](https://doi.org/10.1016/j.tibs.2008.02.004) (2008).
16. Trincao, J. *et al.* Structure of the catalytic core of *S. cerevisiae* DNA polymerase  $\epsilon$ : Implications for translesion DNA synthesis. *Molecular Cell* **8**, 417–426, doi:[10.1016/S1097-2765\(01\)00306-9](https://doi.org/10.1016/S1097-2765(01)00306-9) (2001).
17. Wilson, R. C. & Pata, J. D. Structural insights into the generation of single-base deletions by the Y family DNA polymerase  $\delta$ . *Mol Cell* **29**, 767–779, doi:[10.1016/j.molcel.2008.01.014](https://doi.org/10.1016/j.molcel.2008.01.014) (2008).
18. Chandani, S. & Loechler, E. L. Structural model of the Y-Family DNA polymerase V/RecA mutasome. *J Mol Graph Model* **39**, 133–144, doi:[10.1016/j.jmgm.2012.09.005](https://doi.org/10.1016/j.jmgm.2012.09.005) (2013).
19. Boudsocq, F. *et al.* Investigating the role of the little finger domain of Y-family DNA polymerases in low fidelity synthesis and translesion replication. *J Biol Chem* **279**, 32932–32940, doi:[10.1074/jbc.M405249200](https://doi.org/10.1074/jbc.M405249200) (2004).
20. Akabayov, B. *et al.* Conformational dynamics of bacteriophage T7 DNA polymerase and its processivity factor, *Escherichia coli* thioredoxin. *Proc Natl Acad Sci USA* **107**, 15033–15038, doi:[10.1073/pnas.1010141107](https://doi.org/10.1073/pnas.1010141107) (2010).
21. Gruz, P. *et al.* Synthetic activity of Sso DNA polymerase Y1, an archaeal DinB-like DNA polymerase, is stimulated by processivity factors proliferating cell nuclear antigen and replication factor C. *J Biol Chem* **276**, 47394–47401, doi:[10.1074/jbc.M107213200](https://doi.org/10.1074/jbc.M107213200) (2001).
22. Pages, S., Kester, H. C., Visser, J. & Benen, J. A. Changing a single amino acid residue switches processive and non-processive behavior of *Aspergillus niger* endopolygalacturonase I and II. *J Biol Chem* **276**, 33652–33656, doi:[10.1074/jbc.M105770200](https://doi.org/10.1074/jbc.M105770200) (2001).
23. Xing, G., Kirouac, K., Shin, Y. J., Bell, S. D. & Ling, H. Structural insight into recruitment of translesion DNA polymerase Dpo4 to sliding clamp PCNA. *Mol Microbiol* **71**, 678–691, doi:[10.1111/mmi.2009.71.issue-3](https://doi.org/10.1111/mmi.2009.71.issue-3) (2009).
24. Wang, Y. *et al.* A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance *in vitro*. *Nucleic Acids Res* **32**, 1197–1207, doi:[10.1093/nar/gkh271](https://doi.org/10.1093/nar/gkh271) (2004).
25. Walsh, J. M. *et al.* Effects of non-catalytic, distal amino acid residues on activity of *E. coli* DinB (DNA polymerase IV). *Environ Mol Mutagen* **53**, 766–776, doi:[10.1002/em.21730](https://doi.org/10.1002/em.21730) (2012).
26. Eoff, R. L., Irimia, A., Angel, K. C., Egli, M. & Guengerich, F. P. Hydrogen bonding of 7,8-dihydro-8-oxodeoxyguanosine with a charged residue in the little finger domain determines miscoding events in *Sulfolobus solfataricus* DNA polymerase Dpo4. *J Biol Chem* **282**, 19831–19843, doi:[10.1074/jbc.M702290200](https://doi.org/10.1074/jbc.M702290200) (2007).
27. Sherrer, S. M. *et al.* Identification of an unfolding intermediate for a DNA lesion bypass polymerase. *Chem Res Toxicol* **25**, 1531–1540, doi:[10.1021/tx3002115](https://doi.org/10.1021/tx3002115) (2012).
28. Chen, X., Zaro, J. L. & Shen, W. C. Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev* **65**, 1357–1369, doi:[10.1016/j.addr.2012.09.039](https://doi.org/10.1016/j.addr.2012.09.039) (2013).
29. Minshull, J., Ness, J. E., Gustafsson, C. & Govindarajan, S. Predicting enzyme function from protein sequence. *Curr Opin Chem Biol* **9**, 202–209, doi:[10.1016/j.cbpa.2005.02.003](https://doi.org/10.1016/j.cbpa.2005.02.003) (2005).
30. Mukherjee P, W. R., Lahiri, I. & Pata, J. D. Three residues of the interdomain linker determine the conformation and single-base deletion fidelity of Y-family translesion polymerases. *J Biol Chem* **289**, 6323–6331, doi:[10.1074/jbc.M113.537860](https://doi.org/10.1074/jbc.M113.537860) (2014).
31. Jarosz, D. F., Beuning, P. J., Cohen, S. E. & Walker, G. C. Y-family DNA polymerases in *Escherichia coli*. *Trends Microbiol* **15**, 70–77, doi:[10.1016/j.tim.2006.12.004](https://doi.org/10.1016/j.tim.2006.12.004) (2007).
32. Privalov, P. L., Dragan, A. I. & Crane-Robinson, C. Interpreting protein/DNA interactions: distinguishing specific from non-specific and electrostatic from non-electrostatic components. *Nucleic Acids Res* **39**, 2483–2491, doi:[10.1093/nar/gkq984](https://doi.org/10.1093/nar/gkq984) (2011).
33. Breyer, W. A. & Matthews, B. W. A structural basis for processivity. *Protein Sci* **10**, 1699–1711, doi:[10.1110/ps.10301](https://doi.org/10.1110/ps.10301) (2001).
34. Zang, H. *et al.* Efficient and high fidelity incorporation of dCTP opposite 7,8-dihydro-8-oxodeoxyguanosine by *Sulfolobus solfataricus* DNA polymerase Dpo4. *J Biol Chem* **281**, 2358–2372, doi:[10.1074/jbc.M510889200](https://doi.org/10.1074/jbc.M510889200) (2006).
35. Wu, Y., Wilson, R. C. & Pata, J. D. The Y-family DNA polymerase Dpo4 uses a template slippage mechanism to create single-base deletions. *J Bacteriol* **193**, 2630–2636, doi:[10.1128/JB.00012-11](https://doi.org/10.1128/JB.00012-11) (2011).
36. Gao, Y. G. *et al.* The crystal structure of the hyperthermophile chromosomal protein Sso7d bound to DNA. *Nat Struct Biol* **5**, 782–786, doi:[10.1038/1822](https://doi.org/10.1038/1822) (1998).
37. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343–345, doi:[10.1038/nmeth.1318](https://doi.org/10.1038/nmeth.1318) (2009).
38. Pavelka, A., Chovancova, E. & Damborsky, J. HotSpot Wizard: a web server for identification of hot spots in protein engineering. *Nucleic Acids Res* **37**, W376–W383, doi:[10.1093/nar/gkp410](https://doi.org/10.1093/nar/gkp410) (2009).
39. Homeyer, N. & Gohlke, H. Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method. *Mol Inform* **31**, 114–122, doi:[10.1002/minf.v31.2](https://doi.org/10.1002/minf.v31.2) (2012).
40. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854, doi:[10.1093/bioinformatics/btt055](https://doi.org/10.1093/bioinformatics/btt055) (2013).
41. Zhang, Y. *et al.* Understanding the molecular mechanism of the broad and potent neutralization of HIV-1 by antibody VRC01 from the perspective of molecular dynamics simulation and binding free energy calculations. *J Mol Model* **18**, 4517–4527, doi:[10.1007/s00894-012-1450-z](https://doi.org/10.1007/s00894-012-1450-z) (2012).
42. Liu, H. & Yao, X. Molecular basis of the interaction for an essential subunit PA-PB1 in influenza virus RNA polymerase: insights from molecular dynamics simulation and free energy calculation. *Mol Pharm* **7**, 75–85, doi:[10.1021/mp900131p](https://doi.org/10.1021/mp900131p) (2010).
43. Nguyen, T. T., Mai, B. K. & Li, M. S. Study of Tamiflu sensitivity to variants of A/H5N1 virus using different force fields. *J Chem Inf Model* **51**, 2266–2276, doi:[10.1021/ci2000743](https://doi.org/10.1021/ci2000743) (2011).
44. Kumari, R., Kumar, R., Open Source Drug Discovery, C. & Lynn, A. g\_mmpbsa—a GROMACS tool for high-throughput MM-PBSA calculations. *J Chem Inf Model* **54**, 1951–1962, doi:[10.1021/ci500020m](https://doi.org/10.1021/ci500020m) (2014).
45. Jia, Y. P., Kumar, A. & Patel, S. S. Equilibrium and stopped-flow kinetic studies of interaction between T7 RNA polymerase and its promoters measured by protein and 2-aminopurine fluorescence changes. *J Biol Chem* **271**, 30451–30458, doi:[10.1074/jbc.271.48.30451](https://doi.org/10.1074/jbc.271.48.30451) (1996).
46. Beechem, J. M. *et al.* Exonuclease-polymerase active site partitioning of primer-template DNA strands and equilibrium Mg<sup>2+</sup>-binding properties of bacteriophage T4 DNA polymerase. *Biochemistry-U S* **37**, 10144–10155, doi:[10.1021/bi980074b](https://doi.org/10.1021/bi980074b) (1998).
47. Boudsocq, F., Iwai, S., Hanaoka, F. & Woodgate, R. *Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4): an archaeal DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic poleta. *Nucleic Acids Res* **29**, 4607–4616, doi:[10.1093/nar/29.22.4607](https://doi.org/10.1093/nar/29.22.4607) (2001).
48. Washington, M. T., Johnson, R. E., Prakash, S. & Prakash, L. Fidelity and processivity of *Saccharomyces cerevisiae* DNA polymerase  $\epsilon$ . *J Biol Chem* **274**, 36835–36838, doi:[10.1074/jbc.274.52.36835](https://doi.org/10.1074/jbc.274.52.36835) (1999).

49. Patra, A. *et al.* Kinetics, structure, and mechanism of 8-Oxo-7,8-dihydro-2'-deoxyguanosine bypass by human DNA polymerase *et al.* *J Biol Chem* **289**, 16867–16882, doi:10.1074/jbc.M114.551820 (2014).
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**, 7.21.1-7.21.13, 10.1002/0471142700.nc0721s59, 7 21 21–7213 (2014).
51. Beckman, J. W., Wang, Q. & Guengerich, F. P. Kinetic analysis of correct nucleotide insertion by a Y-family DNA polymerase reveals conformational changes both prior to and following phosphodiester bond formation as detected by tryptophan fluorescence. *J Biol Chem* **283**, 36711–36723, doi:10.1074/jbc.M806785200 (2008).

## Acknowledgements

We thank Angela Park for experimental assistance, Junqing Wang and Joseph Muldoon for intellectual contribution, and Namita Jayant Bhan for editing. We thank the Recombinant Protein Purification Center (rPPC), Genomics Core, and Keck Biophysics Facility at Northwestern University. This work has been funded by the National Institutes of Health (NIH) Common Fund (5R01MH103910-02); Chicago Biomedical Consortium with support from the Searle Funds at The Chicago Community Trust; the National Nature Science Foundation of China (21576117); Natural Science Foundation of Jiangsu Province of China (BK20131103); and the China Scholarship Council (CSC). The authors declare no financial or commercial conflict of interest.

## Author Contributions

Jing Wu wrote the main manuscript text and prepared all figures and tables. Keith E.J. Tyo guided the research and helped editing. Alexandra de Paz helped to do the processivity assay and helped editing. Bradley M. Zamft gave advice and helped editing. Adam H. Marblestone gave advice. Edward S. Boyden gave advice and helped editing. Konrad P. Kording helped in modeling. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:10.1038/s41598-017-02578-3

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017