THEORETICAL BIOLOGY AND
MEDICAL MODELLING

A nucleotide binding rectification Brownian ratchet model for translocation of Y-family DNA polymerases

Ping Xie

Correspondence: pxie@aphy.iphy.
ac.cn
Key Laboratory of Soft Matter
Physics and Beijing National
Laboratory for Condensed Matter
Physics, Institute of Physics,
Chinese Academy of Sciences,
Beijing 100190, China

Abstract

Y-family DNA polymerases are characterized by low-fidelity synthesis on undamaged DNA and ability to catalyze translesion synthesis over the damaged DNA. Their translocation along the DNA template is an important event during processive DNA synthesis. In this work we present a Brownian ratchet model for this translocation, where the directed translocation is rectified by the nucleotide binding to the polymerase. Using the model, different features of the available structures for Dpo4, Dbh and polymerase ι in binary and ternary forms can be easily explained. Other dynamic properties of the Y-family polymerases such as the fast translocation event upon dNTP binding for Dpo4 and the considerable variations of the processivity among the polymerases can also be well explained by using the model. In addition, some predicted results of the DNA synthesis rate versus the external force acting on Dpo4 and Dbh polymerases are presented. Moreover, we compare the effect of the external force on the DNA synthesis rate of the Y-family polymerase with that of the replicative DNA polymerase.

Introduction

DNA polymerases (Pols) are enzymes to add free nucleotide to the 3' end of the newly-forming DNA strand. They play an essential role in the maintenance of genome integrity. On the basis of sequence similarity, DNA Pols can be broadly classified into A-, B-, C-, D-, X- and Y-families [1-3]. In general, most Pols in A-, B-, C-, and D-families are high-fidelity enzymes primarily involved in faithful DNA replication and in repair of replication mistake. The X-family Pols are involved in a number of DNA repair processes such as base excision repair (BER) and repair of double-strand breaks (DSBs) [4,5]. The Y-family Pols represent a number of recently identified Pols characterized by low-fidelity synthesis on undamaged DNA and the ability to bypass DNA lesions which normally block replication by members of the A-, B-, C-, D-, or X-family Pols [6-12].

The Y-family Pols are ubiquitous and are distributed among the three kingdoms of life. They include *E. coli* Pol IV (also known as DinB) [13] and Pol V (also known as UmuC) [14,15], yeast Pol η [16] and Rev1 [17], human Pols η [18], ι [19,20], κ [21] and Rev1 [22], and archaeal Dbh [23] and Dpo4 [24], etc. Although there is no detectable sequence identity with other family Pols, available crystal structures of some Y-family Pols such as Dbh [25-27], Dpo4 [28], Pol η [29,30], Pol ι [31,32], Pol κ [33,34]

and Rev1 [35] reveal that they retain a catalytic core consisting of fingers, palm and thumb subdomains found in other family Pols. However, the fingers and thumb subdomains of the Y-family Pols are significantly smaller than the corresponding subdomains of the other DNA Pols. In addition to the conserved polymerase core, the Y-family Pols possess also a unique C-terminal domain termed the little finger (LF), wrist or polymerase-associated domain (PAD). In this paper we will use the acronym, LF, to denote this domain. The LF domain is the least conserved of the four domains in the Y-family Pols.

Besides the intensive structural studies of the Y-family Pols, which include the structures in apo, binary and ternary forms as well as the structures complexed with DNA substrates containing different lesions [11,25-42], a variety of biochemical assays have provided insight into the catalytic mechanism, lesion-bypassing property, processivity and fidelity of the Pols [12,43-55]. Both the biochemical and single-molecule assays for Dpo4 indicated that the binding of a nucleotide induces a fast DNA translocation event [55,56], which is consistent with the structural studies showing that, in both of the binary complexes (pre- and post-insertion), the primer terminus occupies the site where the next incoming nucleotide will bind [28,41,42]. However, the structural studies for Dbh showed that, in the pre-insertion binary complex, the templating base and the primer terminus are already positioned so that space is available for the incoming nucleotide to bind and form the ternary complex, while in the post-insertion binary complex, the DNA is located in nearly the same position on the Pol [27]. Similar to Dbh, two pre-insertion binary complexes of Pol ι showed that space is available for the incoming nucleotide to bind [32].

Recently, a Brownian ratchet model has been proposed for the translocation of the high-fidelity replicative DNA Pols, where the translocation depends on the change of the interaction of the fingers subdomain with the single-stranded DNA (ssDNA) template upon a correct incorporation [57,58]. In this work, based on the available structural, biochemical and single-molecule studies for the Y-family Pol, we modify the previous Brownian ratchet model for the replicative Pol to be applicable to the Y-family Pol, where the directed translocation is rectified by the nucleotide binding. Thus, the model can be called nucleotide binding rectification (NBR) Brownian ratchet model, which is abbreviated as the NBR model. Using the model, the observed different features of the structures for Dpo4, Dbh and Pol ι in binary and ternary forms [27,28,32,41,42] can be easily explained. Other dynamic properties for the Y-family Pols such as the considerable variations of the processivity among the Pols and the fast translocation event upon dNTP binding for Dpo4 can also be explained by using the model. In addition, some predicted results of the DNA synthesis rate versus the external force acting on Dpo4 and Dbh Pols are presented. Moreover, we compare the effect of the external force on the DNA synthesis rate of the Y-family Pol with that of the replicative Pol.
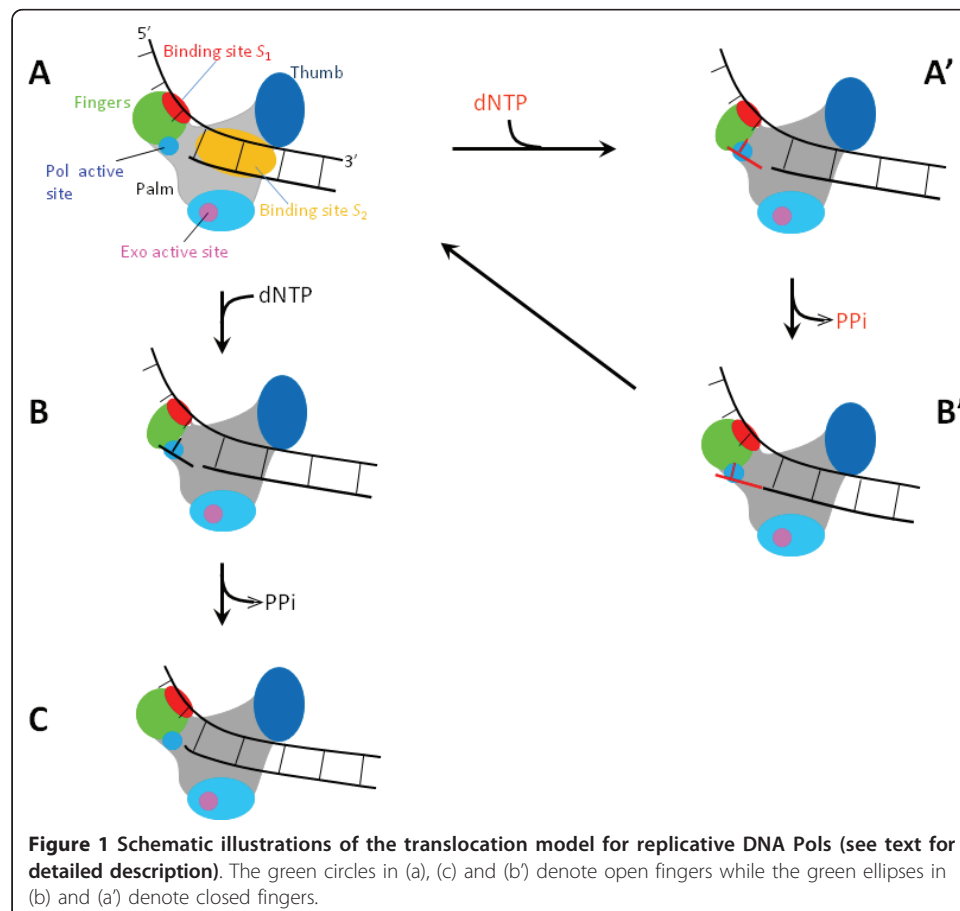
Methods
Brownian ratchet translocation model for replicative DNA Pols
Since the NBR model for the Y-family Pol is modified from the previous model for the replicative Pol [57-59], for convenience of reading, in this section we re-present the latter model. Briefly, the model was based on the Brownian ratchet mechanism

(e.g., see [60,61]) and the directed translocation of the Pol along the template resulted from the potential change induced by dNTP incorporation. The model was built up based on two arguments.
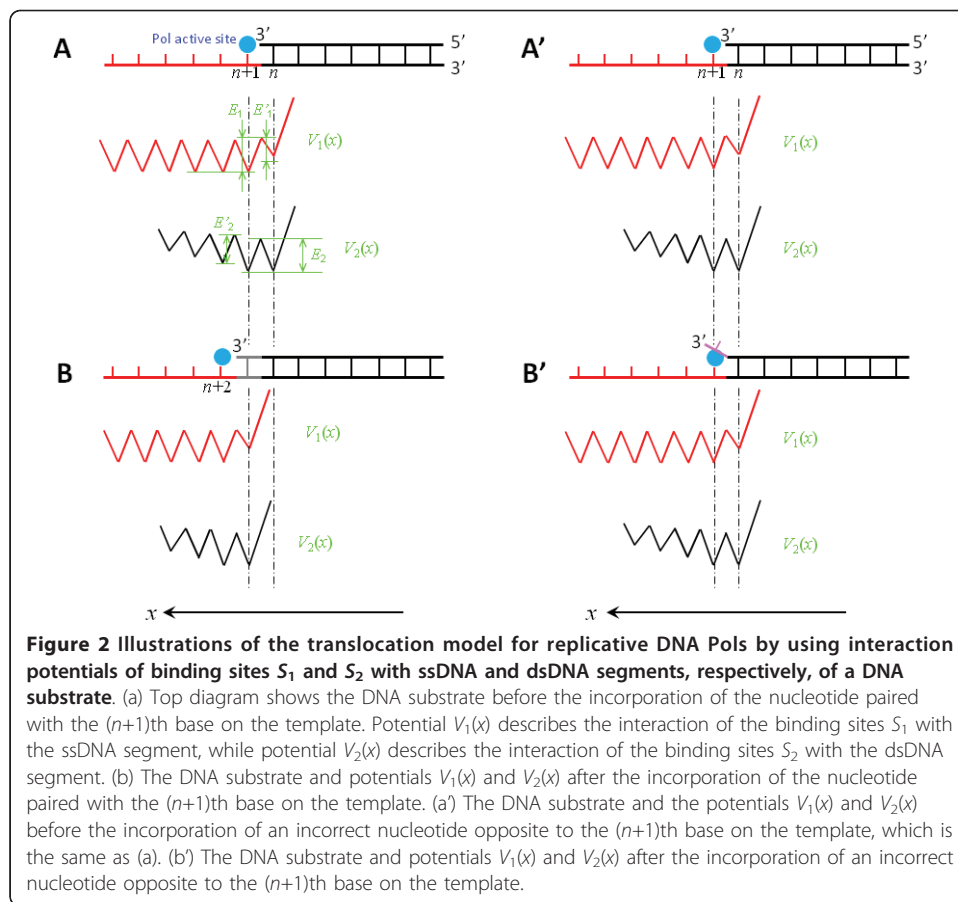
The first argument is on the interaction between the Pol and DNA substrate. The interaction can be characterized by two DNA-binding sites on the Pol. (i) The binding site S_1, which is located in the fingers subdomain (see Figure 1a), shows a high affinity for the unpaired base and/or the sugar-phosphate backbone of the ssDNA template. The presence of binding site S_1 is supported by the experimental data on bacteriophage T4 DNA Pol and Klenow fragment, showing that the fingers subdomain has a high binding affinity for the ssDNA template [62-64]. (ii) The binding site S_2, which is located in the palm and thumb subdomains (see Figure 1a), shows a high affinity for the double-stranded DNA (dsDNA).

The second argument is on the rotation of the fingers subdomain from open (closed) to closed (open) conformation upon the binding (release) of dNTP (pyrophosphate, PPi), which is consistent with the structural studies on bacteriophage T7 DNA Pol [65], *Taq* DNA Pol [66] and HIV-1 reverse transcriptase [67]. The closed conformation of the fingers activates the phosphodiester bond formation (or nucleotide incorporation), while the open conformation of the fingers opens the polymerase active site for nucleotide binding. Moreover, the closed fingers could potentially enhance the interactions of binding sites S_1 and S_2 with the DNA substrate.

Figure 1 Schematic illustrations of the translocation model for replicative DNA Pols (see text for detailed description). The green circles in (a), (c) and (b') denote open fingers while the green ellipses in (b) and (a') denote closed fingers.

Based on the two arguments, the translocation model for the replicative DNA Pol is schematically shown in Figure 1[57-59]. We begin with the binding site S_1 of the Pol binding strongly to the ssDNA at the replication fork, with the binding site S_2 binding to the dsDNA and no nucleotide being in the polymerase active site (Figure 1a). In this nucleotide-free state, either a matched or a mismatched dNTP can bind to the active site, although the matched dNTP has a much larger probability to bind. Thus, we consider the two cases separately. (i) First, consider a correct incorporation. The binding of a matched dNTP induces the fingers to rotate from open to closed conformations (Figure 1b). The closed conformation activates nucleotide incorporation. After the incorporation, the release of PPi induces the fingers to return to the open conformation. At the same time, the binding site S_1 would bind to new nearest unpaired base (i.e., the next unpaired base) of the ssDNA template, because the previous unpaired base where the binding site S_1 has just bound has disappeared due to base pairing (Figure 1c). Then, the next nucleotide-incorporation cycle will proceed. (ii) Second, consider an incorrect incorporation. We still begin with Figure 1a. The binding of a mismatched dNTP also induces the fingers to rotate from open to closed conformations, activating nucleotide incorporation (Figure 1a'). After the incorporation, the release of PPi induces the fingers to return to the open conformation. Now, although the sugar-phosphate backbone of the mismatched dNTP has been connected to the backbone of the already formed dsDNA, the mismatched base is not paired with the sterically corresponding base on the ssDNA template. Thus, the binding site S_1 is still binding strongly to the same unpaired base of the ssDNA template (Figure 1b'). Thus, the polymerization cannot proceed. In other words, the polymerization becomes stalled. In Figure 1b', after the mismatched base is excised, the polymerization will proceed again (Figure 1a).

Using potentials of the two binding sites interacting with the DNA substrate, we describe the model as follows. First, consider potential, $V_1(x)$, of the binding site S_1 interacting with ssDNA, where position, x, of the Pol along the template is represented by that of its active site. Considering that the binding site S_1 covers N_1 bases on the ssDNA template, before the incorporation of nucleotide paired with the $(n+1)$th base (top of Figure 2a), the form of $V_1(x)$ is shown in Figure 2a, where E_1 is the binding affinity for N_1 bases of the ssDNA template while E'_1 is the binding affinity for (N_1-1) bases. Note that the binding affinity E'_1 that corresponds to binding (N_1-1) bases is smaller than E_1 that corresponds to binding N_1 bases. Moreover, it is implicated in the potential that the primer 3' terminus, due to the structural restriction, is not allowed to move forwards relative to the Pol when its active site is located at the primer 3' terminus. Similarly, considering that the binding site S_2 covers N_2 base pairs of dsDNA, before the incorporation of nucleotide paired with the $(n+1)$th base (top of Figure 2a), the potential, $V_2(x)$, of binding site S_2 interacting with dsDNA is shown in Figure 2a. From Figure 2a, it is seen that the deepest well of the total potential, $V(x) = V_1(x) + V_2(x)$, of the Pol interacting with the DNA substrate is located at position of the $(n+1)$th base before the incorporation of the nucleotide paired with the $(n+1)$th base. Thus, the Pol is now located at position of the $(n+1)$th base. After the incorporation (top of Figure 2b), the forms of $V_1(x)$ and $V_2(x)$ are shown in Figure 2b. Now, the deepest well of the total potential, $V(x) = V_1(x) + V_2(x)$, is located at position of the $(n+2)$th base. Thus, the Pol would move from a shallower potential well located at position of the $(n+1)$th base to the deepest well located at position of the $(n+2)$th base.

Figure 2 Illustrations of the translocation model for replicative DNA Pols by using interaction potentials of binding sites S_1 and S_2 with ssDNA and dsDNA segments, respectively, of a DNA substrate. (a) Top diagram shows the DNA substrate before the incorporation of the nucleotide paired with the (n+1)th base on the template. Potential $V_1(x)$ describes the interaction of the binding sites S_1 with the ssDNA segment, while potential $V_2(x)$ describes the interaction of the binding sites S_2 with the dsDNA segment. (b) The DNA substrate and potentials $V_1(x)$ and $V_2(x)$ after the incorporation of the nucleotide paired with the (n+1)th base on the template. (a') The DNA substrate and the potentials $V_1(x)$ and $V_2(x)$ before the incorporation of an incorrect nucleotide opposite to the (n+1)th base on the template, which is the same as (a). (b') The DNA substrate and potentials $V_1(x)$ and $V_2(x)$ after the incorporation of an incorrect nucleotide opposite to the (n+1)th base on the template.

However, after an incorrect incorporation of the nucleotide opposite to the (n+1)th base (see top of Figure 2b'), the forms of $V_1(x)$ and $V_2(x)$ are shown in Figure 2b', which are the same as those before the incorporation. This is because, after the incorrect incorporation, the (n+1)th base has not formed a base pair with the newly incorporated primer base and, thus, the Pol is still located at the position of the (n+1)th base, i.e., the position of the deepest well.

In this model, the translocation step occurs following the incorporation of a correct nucleotide. This is supported by the comparison of the binary (Pol-DNA) with ternary (Pol-DNA-dNTP) structures for the replicative Pol (see, e.g., [66]). Upon an incorrect incorporation, the Pol becomes stalled, which is also consistent with the experimental data [68]. For a lesion such as an abasic lesion having a weak effect on distortion of the DNA structure so that the damaged base still has a high affinity for the binding site S_1, an incorporated base opposite to the lesion, which is equivalent to a mismatched base, also induces the stall of the polymerization. This is consistent with the structural observation [69]. During the stalled period, the mismatched base would be excised. Then another base opposite to the lesion site would be incorporated. Thus, the Pol cannot perform the translesion synthesis. For lesions that severely distort the DNA structure causing damaged DNA substrate not to be tolerated by the replicative Pol, e.g., with the template base being flipped out of the active site, this would preclude closing of the fingers subdomain upon nucleotide binding, as observed by Li et al. [70] for bacteriophage T7 DNA Pol complexed with a DNA template containing a *cis-syn*
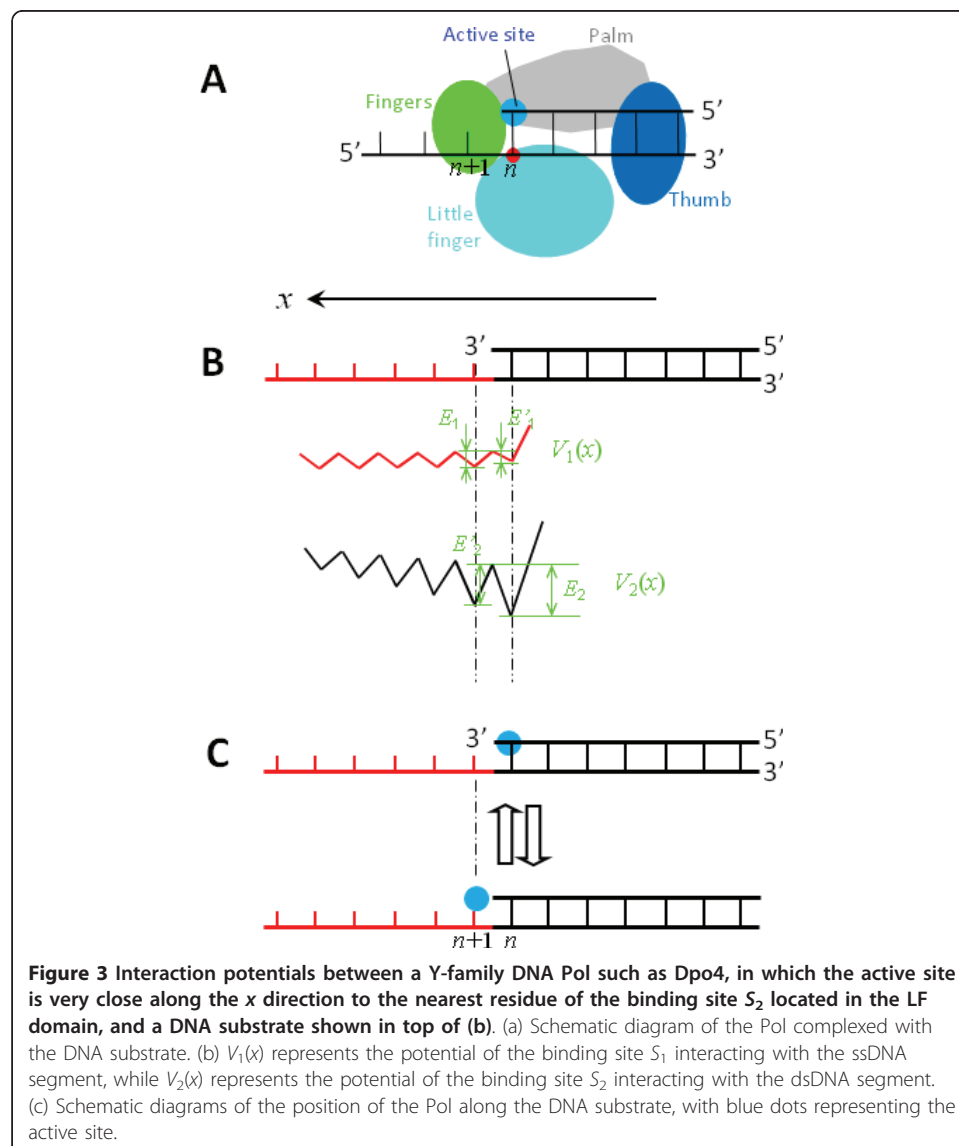
cyclobutane pyrimidine dimer. Without the activation by the closed conformation, the nucleotide incorporation cannot proceed and, thus, the Pol cannot also perform the translesion synthesis.
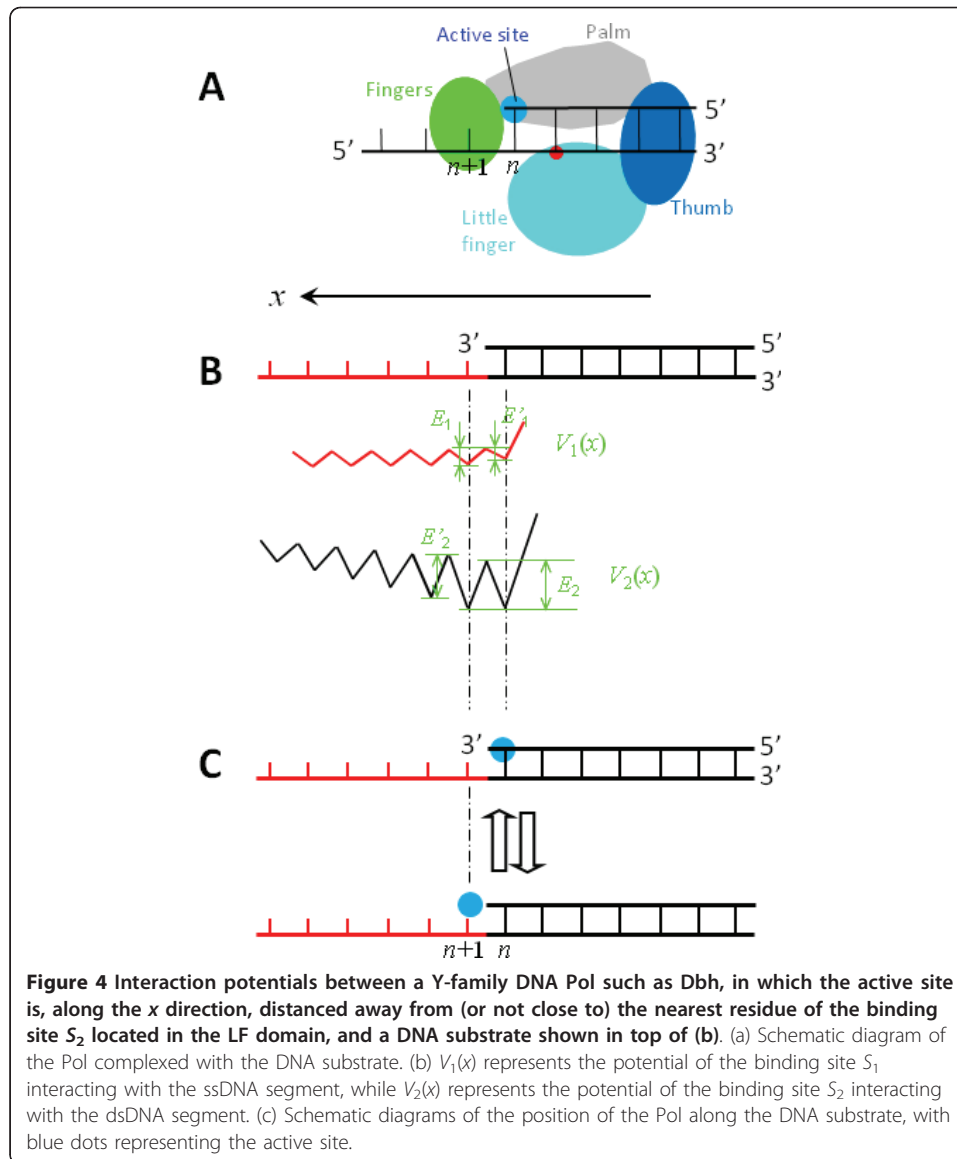
Nucleotide binding rectification Brownian ratchet model for Y-family DNA Pols

The NBR model for the Y-family DNA Pol is modified from the above model for the replicative Pol. The model is also constructed based on two arguments, which are presented in the following two sections.

Interaction of Pol with DNA substrate

As in the replicative Pol (see above), the interaction of the Y-family Pol with the DNA substrate can also be characterized by two DNA-binding sites on the Pol. The binding site S_1 is composed of residues located in the fingers subdomain (see Figure 3a or 4a). However, in contrast to the replicative Pol where the binding site S_1 has a *high* affinity for the unpaired bases and/or the sugar-phosphate backbone of the ssDNA template,

Figure 3 Interaction potentials between a Y-family DNA Pol such as Dpo4, in which the active site is very close along the *x* direction to the nearest residue of the binding site S_2 located in the LF domain, and a DNA substrate shown in top of (b). (a) Schematic diagram of the Pol complexed with the DNA substrate. (b) $V_1(x)$ represents the potential of the binding site S_1 interacting with the ssDNA segment, while $V_2(x)$ represents the potential of the binding site S_2 interacting with the dsDNA segment. (c) Schematic diagrams of the position of the Pol along the DNA substrate, with blue dots representing the active site.

Figure 4 Interaction potentials between a Y-family DNA Pol such as Dbh, in which the active site is, along the x direction, distanced away from (or not close to) the nearest residue of the binding site S_2 located in the LF domain, and a DNA substrate shown in top of (b). (a) Schematic diagram of the Pol complexed with the DNA substrate. (b) $V_1(x)$ represents the potential of the binding site S_1 interacting with the ssDNA segment, while $V_2(x)$ represents the potential of the binding site S_2 interacting with the dsDNA segment. (c) Schematic diagrams of the position of the Pol along the DNA substrate, with blue dots representing the active site.

the binding site S_1 in the Y-family Pol has a *very low or even no* affinity, which is consistent with the available structural studies [27,28,32,41,42]. The binding site S_2, which is composed of residues located in the thumb domain and mainly in the LF domain (see Figure 3a or 4a), has a high affinity for dsDNA, which is also consistent with the available structural studies [27,28,32,41,42].

As in Figure 2, the potential $V_1(x)$ of the binding site S_1 interacting with the ssDNA is shown in Figures 3b and 4b, with E_1 denoting the binding affinity for N_1 bases of the ssDNA template while E'_1 the binding affinity for (N_1-1) bases. However, E'_1 and E_1 have very small or nearly zero values.

Then, consider the potential $V_2(x)$ of the binding site S_2 interacting with the dsDNA. Since the binding site S_2 in the Y-family Pols is composed of residues located in the thumb domain and mainly in the LF domain, the form of potential $V_2(x)$ depends on the distance, L, from the active site to the nearest residue (red dots in Figures 3a and 4a) of the binding site S_2 located in the LF domain along the x direction.

(i) For the case that the active site is very close along the x direction to the nearest residue of the binding site S_2 located in the LF domain (see Figure 3a), as seen from the structure of Dpo4 [28,41,42], the interaction potential $V_2(x)$ can be simply shown in Figure 3b, where $L = 0$. If binding site S_2 is considered to cover N_2 base pairs of the dsDNA, E_2 is the binding affinity for the sugar-phosphate backbones connecting N_2 base pairs on the dsDNA while E'_2 is the binding affinity for the backbones connecting only (N_2-1) base pairs. Moreover, in the potential it is implicated that the primer 3' terminus, due to the structural restriction (see, e.g., [27,28,32,41,42]), is not allowed to move forwards relative to the Pol when its active site is located at the primer 3' terminus. In addition, from the Pol structures complexed with the DNA substrate, it is inferred that that the interaction between the binding site S_2 and the dsDNA is via the hydrogen-bonding, van der Waals and mainly electrostatic forces. On the other hand, the interaction distance of the electrostatic force that is approximately equal to the Debye length (~ 1 nm) in solution is larger than the distance $p = 0.34$ nm between two successive base pairs. Thus, the value at maxima of $V_2(x)$ increases as the binding site S_2 deviates away from the dsDNA segment along the x direction.

(ii) For the case that the active site is, along the x direction, distanced away from (or not close to) the nearest residue of the binding site S_2 located in the LF domain (Figure 4a), as evidently seen from the structure of Dbh [27], the interaction potential $V_2(x)$ can be simply shown in Figure 4b, where we take $L = 1$ bp. From available structures of the binary and ternary complex for Pol ι [31,32], it is also noted that, if the active site is positioned opposite to the first unpaired base on the template, the first unpaired base is distanced by $L = 1$ bp away from the nearest residue of the binding site S_2 located in the LF domain. Thus, the interaction potential $V_2(x)$ for Pol ι also has the form of Figure 4b rather than that of Figure 3b. Similarly, from the available structure of the ternary complex for Pol η [30], we infer that the interaction potential $V_2(x)$ for Pol η also has the form of Figure 4b.

From Figure 3b it is seen that, when the active site is positioned at the nth base pair (top of Figure 3c), the affinity of the Pol for the DNA substrate is $E_n = E'_1 + E_2$; while when the active site is positioned at the $(n+1)$th base (bottom of Figure 3c), the affinity is $E_{n+1} = E_1 + E'_2$. Since E'_1 and E_1 are much smaller than E'_2 and E_2 and $E_2 > E'_2$, it is expected that $E_n > E_{n+1}$. Similarly, from Figure 4b it is seen that, when the active site is positioned at the nth base pair (top of Figure 4c), the affinity of the Pol for the DNA substrate is $E_n = E'_1 + E_2$; while when the active site is positioned at the $(n+1)$th base (bottom of Figure 4c), the affinity is $E_{n+1} = E_1 + E_2$. Since E'_1 and E_1 are much smaller than E_2, it is expected that E_{n+1} is slightly larger than (or nearly equal to) E_n. Moreover, from both Figure 3 and 4 it is noted that, when the active site is positioned at the $(n+1)$th base, the jumping of the Pol from the $(n+1)$th site to the $(n+2)$th site is required to overcome a larger energy barrier than the backward jumping to the nth site. For approximation, we do not consider the jumping to the $(n+2)$th site in this work.

The binding of dNTP induces a slight conformational change, enhancing the interaction of the Pol with DNA substrate

As evidenced from the FRET experimental data [56], it is argued that the dNTP binding involves (at least) two substeps, $E \cdot DNA + dNTP \rightarrow E \cdot DNA \cdot dNTP \rightarrow E^* \cdot DNA \cdot dNTP$, where E represents the DNA Pol. The transition from the unactivated E

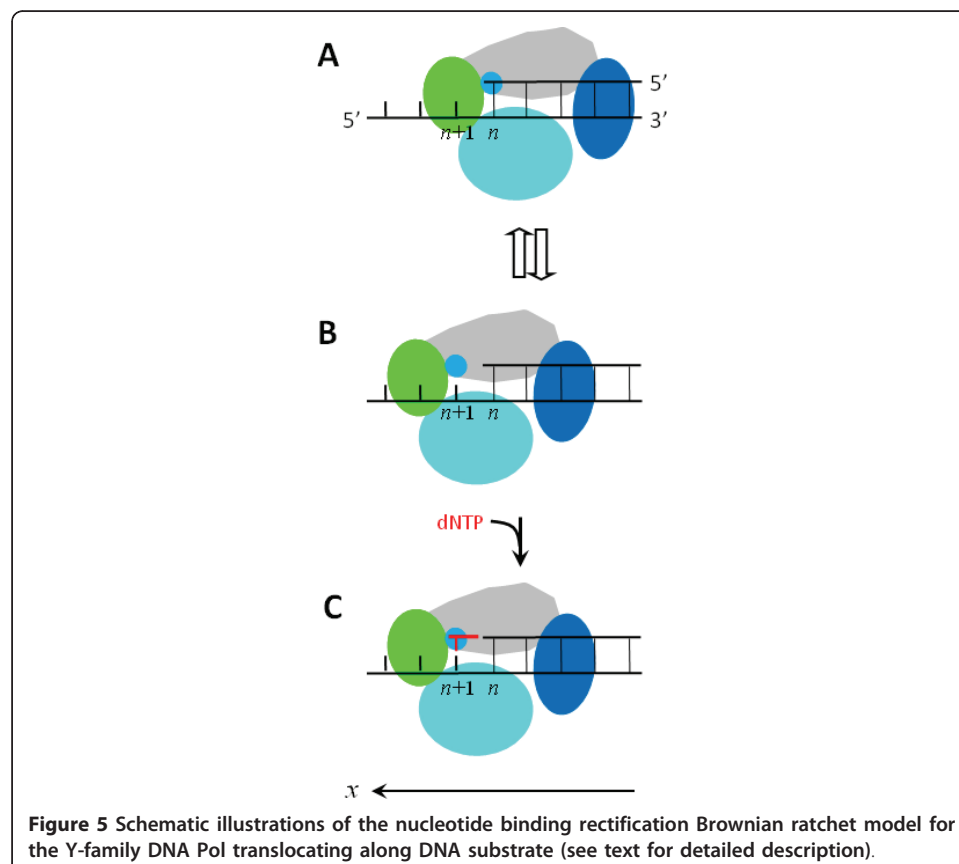· DNA · dNTP ternary complex to activated E* · DNA · dNTP ternary complex induces a slight conformational change of the Pol, enhancing its interactions with both the DNA substrate and dNTP. Similarly, the PPi releasing also involves (at least) two sub-steps, E* · DNA · PPi → E · DNA · PPi → E · DNA + PPi, where the transition from the activated E* · DNA · PPi ternary complex to unactivated E · DNA · PPi ternary complex results in a reverse slight conformational change of the Pol, reducing its interactions with both the DNA substrate and PPi.

Since in the activated E* · DNA · dNTP (or E* · DNA · PPi) complex the Pol has a stronger interaction with DNA substrate and nucleotide than in the unactivated E · DNA · dNTP (E · DNA · PPi) complex, for simplicity of analysis, it is considered that in the activated state the Pol is unable to move relative to the DNA substrate and the dNTP or PPi bound to the active site has a negligible probability to release.

Model for Pol translocation

Using potentials $V_1(x)$ and $V_2(x)$ (Figures 3 and 4), the NBR model for the Y-family Pol translocating along DNA substrate is schematically shown in Figure 5.

We begin with the Pol positioned at the nth site (Figure 5a), just after the incorporation of a nucleotide. In Figure 5a, the active site is occupied by the primer 3'-terminus, which sterically prevents a dNTP from binding to the active site. Due to the thermal noise, the Pol in this nucleotide-free state can jump from the nth site to the $(n+1)$th site (from Figure 5a to 5b) and vice verse (from Figure 5b to 5a). For the case that the active site is very close along the x direction to the nearest residue of the binding site S_2 located in the LF domain (Figure 3a), $E_n > E_{n+1}$ (see above). Thus, the Pol in the

Figure 5 Schematic illustrations of the nucleotide binding rectification Brownian ratchet model for the Y-family DNA Pol translocating along DNA substrate (see text for detailed description).

binary E · DNA state stays most of the time at the nth site (Figure 5a), as will be shown in the Results, which is consistent with the availably resolved binary E · DNA structure for Dpo4 [28,41,42]. For the case that the active site is, along the x direction, distanced away from (or not close to) the nearest residue of the binding site S_2 located in the LF domain (Figure 4a), E_{n+1} is slightly larger than (or nearly equal to) E_n (see above). Thus, the Pol in the binary E · DNA state shows slightly larger (or nearly equal) probability to stay at the $(n+1)$th site (Figure 5b) than (or to) that at the nth site (Figure 5a), implying that the binary E · DNA structure for this case would be observed to be either at the $(n+1)$th site or at the nth site. This is consistent with the observations that the pre-insertion binary E · DNA structures for Dbh [27] and Pol ι [32] showed that their active sites are at the $(n+1)$th site, while the post-insertion binary E · DNA structure for Dbh [27] showed that the active site is at the nth site.

When the Pol jumps to the $(n+1)$th site, since the active site is nucleotide free, a dNTP becomes able to bind to it, as shown in Figure 5b that is equivalent to the state shown at bottom of Figure 3c or Figure 4c. Consider that the dNTP binds to the active site during the period when the Pol stays at the $(n+1)$th site (Figure 5c). Due to the structural restriction (see, e.g., [27,28,32,41,42]), the occupation of the active site by the dNTP sterically prevents the Pol from moving backwards to the nth site unless the dNTP is dissociated, which is consistent with the available structures showing that the active site of the Pols such as Dbh, Dpo4, Pol ι, Pol η, Pol κ and Rev1 in ternary forms is at the $(n+1)$th site [11,27,28,30,32,34,35,41,42]. Then, the transition from the unactivated ternary complex E · DNA dNTP to the activated E* · DNA dNTP complex enhances the interactions of the Pol with the DNA substrate and with the dNTP, thus preventing both the DNA substrate and the dNTP from dissociating from the Pol.

After the phosphodiester bond formation and then the release of PPi, except that the dsDNA segment is elongated by one base pair and the Pol has moved forwards by one base pair, the Pol-DNA complex returns to the state shown in Figure 5a. Correspondingly, the potentials $V_1(x)$ and $V_2(x)$ in Figure 3b and in Figure 4b are shifted by one base pair along the x direction. Then, the next round of the nucleotide incorporation would proceed continuously.

Equations for Pol motion

Consider the movement of Pol relative to the DNA substrate in two dimensions. One is along the DNA, which is represented by the x axis, as shown in Figures 3, 4, and 5. The other one is along the r axis that is perpendicular to the x axis. Then, the movement equations can be written in the following Langevin forms

$$\Gamma\frac{dx}{dt} = -\frac{\partial U(x,r)}{\partial x} + \xi_x(t), \tag{1a}$$

$$\Gamma\frac{dr}{dt} = -\frac{\partial U(x,r)}{\partial r} + \xi_r(t). \tag{1b}$$

Here the potential $U(x,r)$ can be written as $U(x,r) = V(x)[2\exp(-r/r_d) - \exp(-2r/r_d)]$, with $V(x) = V_1(x) + V_2(x) + V_0$, where $V_1(x)$ and $V_2(x)$ have the forms shown in Figures 3b and 4b, and $V_0 \equiv -E_0 < 0$ results from the fact that the electrostatic interaction distance of the Pol with the DNA in solution is larger than the distance between two successive base pairs. The magnitude of $V(x)$ is defined as follows: its minimum

value at the nth site is $- (E_n + E_0)$, while at the $(n+1)$th site the value is $- (E_{n+1} + E_0)$. The term $[2 \exp (-r/r_d) - \exp (-2r/r_d)]$, which has the Morse form, denotes the potential change along the r direction, with $2r_d = 1$ nm (the Debye length) characterizing the interaction distance. The parameter Γ is the frictional drag coefficient on the Pol and $\xi_i (t)$ $(i = x, r)$ is the fluctuating Langevin force with $\langle \xi_i (t) \rangle = 0$ and $\langle \xi_i (t)\xi_j (t') \rangle = 2k_B T\Gamma\delta_{ij} \delta(t - t')$. The drag coefficient is calculated by $\Gamma = 6\pi\eta R$, where η is the viscosity of the aqueous medium and the Pol is approximated as a sphere with radius $R = 5$ nm. As the previous experiment showed that the viscosity of the aqueous cytoplasm does not differ from water [71], we take the viscosity of aqueous to be the same as that of water in the calculation, i.e., $\eta = 0.01$ g cm^{-1} s^{-1}, which gives $\Gamma = 9.4 \times 10^{-11}$ kg s^{-1}. Moreover, the effect of the viscosity variation on the results will be discussed.

Results

Processivity of the Y-family Pol

To study the processivity of the Y-family Pol, we determine the dissociation probability of the Pol from the DNA substrate during one cycle of nucleotide incorporation. To this end, we calculate the mean dissociation time, T_d, of the Pol from the DNA substrate.

First, we consider the motion with the Pol fixed at one potential well (e.g., the potential well at the nth site) along the x direction. Then, the potential $U(x,r)$ in Eq. (1b) becomes: $W(r) = -E_r [2 \exp (-r/r_d) - \exp (-2r/r_d)]$, where the depth of the potential well is $E_r = E_n + E_0$. If it is considered that the Pol is dissociated from its DNA substrate when it moves away from the DNA substrate by a distance of $r = L$, the mean dissociation time T_d, i.e., the mean time for the Pol to move from $r = 0$ to $r = L$, can be obtained by [72]

$$T_d = \frac{1}{D} \int_0^L \exp \left\{ \frac{W(\gamma)}{\Gamma D} \right\} d\gamma \int_0^\gamma \exp \left\{ -\frac{W(z)}{\Gamma D} \right\} dz, \tag{2}$$

where $D = K_B T/\Gamma$. From Eq. (2) we have

$$T_d = \frac{\Gamma}{k_B T} \int_0^L \exp \left\{ \frac{E_r}{k_B T} \left[\exp \left(-\frac{2\gamma}{r_d} \right) - 2 \exp \left(-\frac{\gamma}{r_d} \right) \right] \right\} d\gamma$$

$$\times \int_0^\gamma \exp \left\{ -\frac{E_r}{k_B T} \left[\exp \left(-\frac{2z}{r_d} \right) - 2 \exp \left(-\frac{z}{r_d} \right) \right] \right\} dz, \tag{3}$$

where it is seen that T_d is proportional to the viscosity η.

The dissociation probability per unit time, P_d, of the Pol from the DNA substrate is calculated by

$$P_d = \frac{1}{T_d}. \tag{4}$$

It is noted from Eqs. (3) and (4) that P_d is inversely proportional to the viscosity η.

Based on the model, only during the time period, T_{p1}, after transition to the unactivated E · DNA PPi ternary complex but before the dNTP binding and during the time period, T_{p2}, after the dNTP binding but before transition to the activated E* · DNA

dNTP ternary complex, can the Pol be dissociated from the DNA substrate. During the time period T_{p1}, the Pol can jump between the well at the nth site and the well at the $(n+1)$th site along the x direction. As our results show (see additional file 1), the dissociation probability P_{d1} during this time period T_{p1} is approximately only dependent on the value of $P_d^{(m)}$ ($m = n$ or $n + 1$), with $P_{d1} \approx C \cdot T_{p1} P_d^{(m)}$, where $C = 1 \sim 2$ and $P_d^{(m)}$ represents P_d given by Eqs. (3) and (4) but with $E_r = E_r^{(m)} = E_m + E_0$ ($m = n$ if $E_n > E_{n+1}$, $m = n + 1$ if $E_n < E_{n+1}$). During the time period T_{p2}, the Pol is positioned at the $(n+1)$th site and the dissociation probability P_{d2} is calculated by $P_{d2} = T_{p2} P_d^{(n+1)}$, where $P_d^{(n+1)}$ represents P_d given by Eqs. (3) and (4) but with $E_r^{(n+1)} = E_{n+1} + E_0$.

For the Pol such as Dpo4, in which the active site is very close along the x direction to the nearest residue of the binding site S_2 located in the LF domain, since $E_n > E_{n+1}$, it is noted from Eqs. (3) and (4) that $P_d^{(n)} << P_d^{(n+1)}$. Moreover, it is known that $T_{p1} < T_{p2}$ at saturating concentrations of dNTP for Dpo4 [56]. Thus, we have $P_{d1} \approx C \cdot T_{p1} P_d^{(n)} << P_{d2} = T_{p2} P_d^{(n+1)}$. The mean number of incorporated nucleotides for one binding event of the Pol with the DNA substrate, which characterizes the polymerization processivity, is calculated by

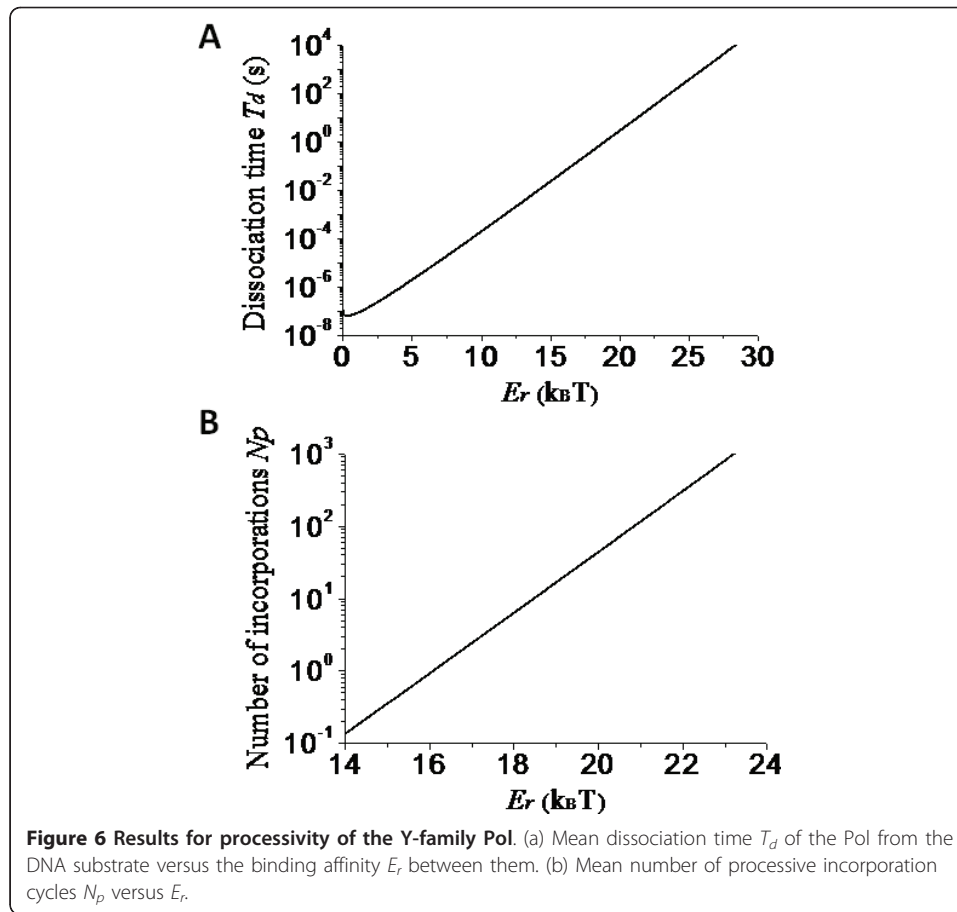$$N_p = \frac{1}{P_{d1} + P_{d2}} \approx \frac{1}{P_{d2}} = \frac{1}{T_{p2} P_d^{(n+1)}}. \tag{5}$$

For the Pol such as Dbh, Pol ι and Pol η, in which the active site is, along the x direction, distanced away from (or not close to) the nearest residue of the binding site S_2 located in the LF domain, since $E_{n+1} \geq E_n$, we have $P_{d1} \approx C \cdot T_{p1} P_d^{(n+1)}$. Thus, the mean number of incorporated nucleotides for one binding event is calculated by

$$N_p = \frac{1}{P_{d1} + P_{d2}} \approx \frac{1}{\left(CT_{p1} + T_{p2}\right) P_d^{(n+1)}}. \tag{6}$$

From Eqs. (5) and (6), it is seen that, whether $E_n > E_{n+1}$ or $E_n \leq E_{n+1}$, the polymerization processivity is mainly determined by the binding affinity, E_{n+1}, of the Pol at the $(n+1)$th site along the DNA substrate. Moreover, it is noted that N_p is proportional to the viscosity η.

Using Eq. (3), the calculated results of the mean dissociation time T_d versus E_r are shown in Figure 6a, where we take $L = 5$ nm that is larger than the interaction distance $2r_d = 1$ nm. It is seen that T_d increases significantly with the increase of E_r. With results of Figure 6a and using Eqs. (4) and (5), the calculated results of N_p versus E_r are shown in Figure 6b, where we take $T_{p2} = 0.065$ s that is consistent with the experimental data of transition rate of 15.3 s^{-1} for Dpo4 [56]. It is seen that, when $N_p = 10 \sim 100$ that is consistent with the experimental data [48], $E_r \approx 18.5k_BT \sim 20.8k_BT$. Now, from this value of $E_r = E_r^{(n+1)} = E_{n+1} + E_0$, we estimate values of E_{n+1} and E_n. Taking a conservative value of $E_0 = E_r/5 = 3.9k_BT$, we estimate that the value of E_{n+1} is at most about $17k_BT$. For reasonable value of $E_n - E_{n+1} = 3k_BT \sim 5k_BT$, we estimate that the value of E_n is at most about $20k_BT \sim 22k_BT$.

In addition, from Figure 6b it is interesting to see that, if E_r is decreased from $20.8k_BT$ to a value below $16k_BT$, N_p is decreased from about 100 to a value smaller than 1. This implies that only an about $5k_BT$ decrease in the binding affinity can

Figure 6 Results for processivity of the Y-family Pol. (a) Mean dissociation time T_d of the Pol from the DNA substrate versus the binding affinity E_r between them. (b) Mean number of processive incorporation cycles N_p versus E_r.

induce the processive polymerization of a hundred nucleotides to a distributive polymerization. From this result, it is also concluded that the considerable variations of the processivity among Y-family Pols result mainly from slight changes in their binding affinities for the DNA substrate. Moreover, since the LF domain is the least conserved of the four domains in the Y-family Pols, the slight differences in the binding affinity of different Pols are mainly due to different interaction strengths of the LF domain with the DNA. For example, comparison of the LF structure of Dpo4 with that of Dbh showed that the DNA-contacting surface in LF domain of Dpo4 is slightly more positively charged than Dbh, and, correspondingly, Dpo4 is much more processive than Dbh [48]. Since the LF domain has a large binding affinity for the DNA, it is expected from Figure 6a that the deletion of the LF domain will significantly reduce the association time of the Pol with the DNA, thus resulting in much less active than the full-length Pol. This is also consistent with the experimental data [28].

Moving time of the Y-family Pol

Now, we study the moving time from the nth site to the $(n+1)$th site and vice verse during the time period after the incorporation of the nth base but before the dNTP binding. To this end, we can consider the motion only along the x direction and the potential $U(x,r)$ in Eq. (1a) becomes $V(x)$. Thus, the mean moving time, $T_{n\rightarrow(n+1)}$, i.e., the mean first-passage time for the Pol to move from the nth site at position $x = 0$ to
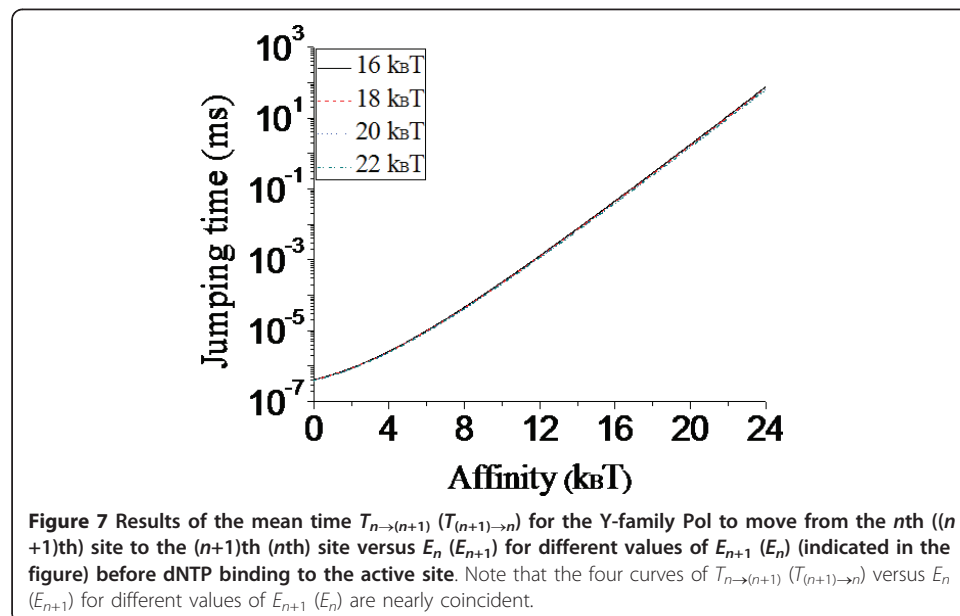
the next $(n+1)$th site at position $x = p = 2l = 0.34$ nm can be approximately calculated by $T_{n\rightarrow(n+1)} = (1/D)\int_0^{2l}\exp[V(y)/(\Gamma D)]\,dy \times \int_0^y \exp[-V(z)/(\Gamma D)]\,dz$, which is similar to Eq. (2). From the integral, we have

$$
\begin{aligned}
T_{n\rightarrow(n+1)} = & \frac{(l\Gamma)^2 D}{E_n}\left[\exp\left(\frac{E_n}{k_B T}\right) - 1\right] - \frac{l^2\Gamma}{E_n} + \frac{l^2\Gamma}{E_{n+1}} \\
& + \frac{(l\Gamma)^2 D}{E_n E_{n+1}}\left[\exp\left(\frac{E_n}{k_B T}\right) - \exp\left(\frac{E_n - E_{n+1}}{k_B T}\right)\right] \\
& \times \left[1 - \exp\left(-\frac{E_n}{k_B T}\right)\left(1 + \frac{E_n}{E_{n+1}}\right)\right].
\end{aligned}
\tag{7}
$$

The mean moving time, $T_{(n+1)\rightarrow n}$, from the $(n+1)$th site to the nth site also has the form of Eq. (7) but with E_n and E_{n+1} being exchanged with each other. From Eq. (7) and $D = K_B T/\Gamma$, it is noted that the mean moving time is proportional to the viscosity η.

Using Eq. (7), the calculated results of the mean moving time $T_{n\rightarrow(n+1)}$ ($T_{(n+1)\rightarrow n}$) versus E_n (E_{n+1}) for different values of E_{n+1} (E_n) are shown in Figure 7. As expected, $T_{n\rightarrow(n+1)}$ increases significantly with the increase of E_n but is insensitive to the variation of E_{n+1}, while $T_{(n+1)\rightarrow n}$ increases significantly with the increase of E_{n+1} but is insensitive to the variation of E_n. It is seen from Figure 7 that, even for the value of $E_n = 20k_B T \sim 22k_B T$ for Dpo4 (see the above section), the mean moving time $T_{n\rightarrow(n+1)}$ is only about $2 \sim 10$ ms. For the value of $E_{n+1} = 17k_B T$ for Dpo4 (see the above section), $T_{(n+1)\rightarrow n}$ is only about 0.1 ms. These results indicated that, after the incorporation of the nth base and before the dNTP binding, Dpo4 would jump between the nth site and the $(n+1)$th site with a high frequency. Thus, within the time resolution used in the FRET experiment [56], this highly frequent jumping between the two positions could not be detected. Moreover, as will be shown in the following section, Dpo4 would stay most probably at the nth site (Figure 5a). Thus, the resolved structure is

Figure 7 Results of the mean time $T_{n\rightarrow(n+1)}$ ($T_{(n+1)\rightarrow n}$) for the Y-family Pol to move from the nth ((n+1)th) site to the (n+1)th (nth) site versus E_n (E_{n+1}) for different values of E_{n+1} (E_n) (indicated in the figure) before dNTP binding to the active site. Note that the four curves of $T_{n\rightarrow(n+1)}$ ($T_{(n+1)\rightarrow n}$) versus E_n (E_{n+1}) for different values of E_{n+1} (E_n) are nearly coincident.

most probably in the state with Dpo4 active site being located at the nth site and the FRET data shows a rapid translocation event for Dpo4 relative to the DNA substrate upon the adding of dNTP, which is consistent with the experimental data [41,42,56].

Effect of external force on DNA-synthesis rate of the Y-family Pol

In the NBR model for the Y-family Pol, after the incorporation of the nth base and before the dNTP binding, the active site jumps between the nth site and the $(n+1)$th site. As noted from Eq. (7), for $E_n >> 1$ and $E_{n+1} >> 1$, the ratio of the time, T_n, for the active site to position at the nth site over the time, T_{n+1}, to position at the $(n+1)$th site approximately has the form

$$\frac{T_n}{T_{n+1}} = \frac{T_{n\to(n+1)}}{T_{(n+1)\to n}} \approx \exp\left(\frac{E_n - E_{n+1}}{k_BT}\right). \tag{8}$$

For Dpo4, since $E_n > E_{n+1}$, it is thus noted from Eq. (8) that the active site has a much larger probability to stay at the nth site than to stay at the $(n+1)$th site. For the value of E_n-$E_{n+1} = 3k_BT \sim 5k_BT$, $T_n/T_{n+1} \approx 20 \sim 100$. For Dbh, Pol ι and Pol η, E_{n+1} is slightly larger than (or nearly equal to) E_n. From Eq. (8) it is noted that the Pols show slightly larger (or nearly equal) probability to stay at the $(n+1)$th site than (or to) that at the nth site.

Consider an external force, F, acting on the Pol bound to a fixed DNA substrate, where F is defined as positive when it points towards the $-x$ direction. The experiment can be realized by using the optical trapping method, with a micro-bead linked to the residues on the palm subdomain or LF domain of the Pol. The linked residues on the Pol should be far away from the active site, thus the external force having no effect on the polymerase activity of the active site. Under the external force F, the depth of potential well at the nth site changes from E_n to $E_n + Fp/2$, while the depth at the $(n+1)$th site changes from E_{n+1} to E_{n+1}-$Fp/2$. Thus, Eq. (8) becomes

$$\frac{T_n}{T_{n+1}} \approx \exp\left(\frac{E_n - E_{n+1}}{k_BT}\right)\exp\left(\frac{Fp}{k_BT}\right). \tag{9}$$

Based on the model, only when the active site is positioned at the $(n+1)$th site can the dNTP bind to the active site, i.e., only during the time period T_{n+1} can the dNTP bind to the active site. Thus, based on Eq. (9), the dNTP-binding rate, $k_b(F)$, versus the external force F has the form

$$k_b(F) = k_b^{(0)} \frac{\exp\left(\dfrac{E_n - E_{n+1}}{k_BT}\right) + 1}{\exp\left(\dfrac{E_n - E_{n+1}}{k_BT}\right)\exp\left(\dfrac{Fp}{k_BT}\right) + 1}. \tag{10}$$

where $k_b^{(0)}$ denotes the dNTP-binding rate under no external force. From Eq. (10), the DNA-synthesis rate, k, is calculated by
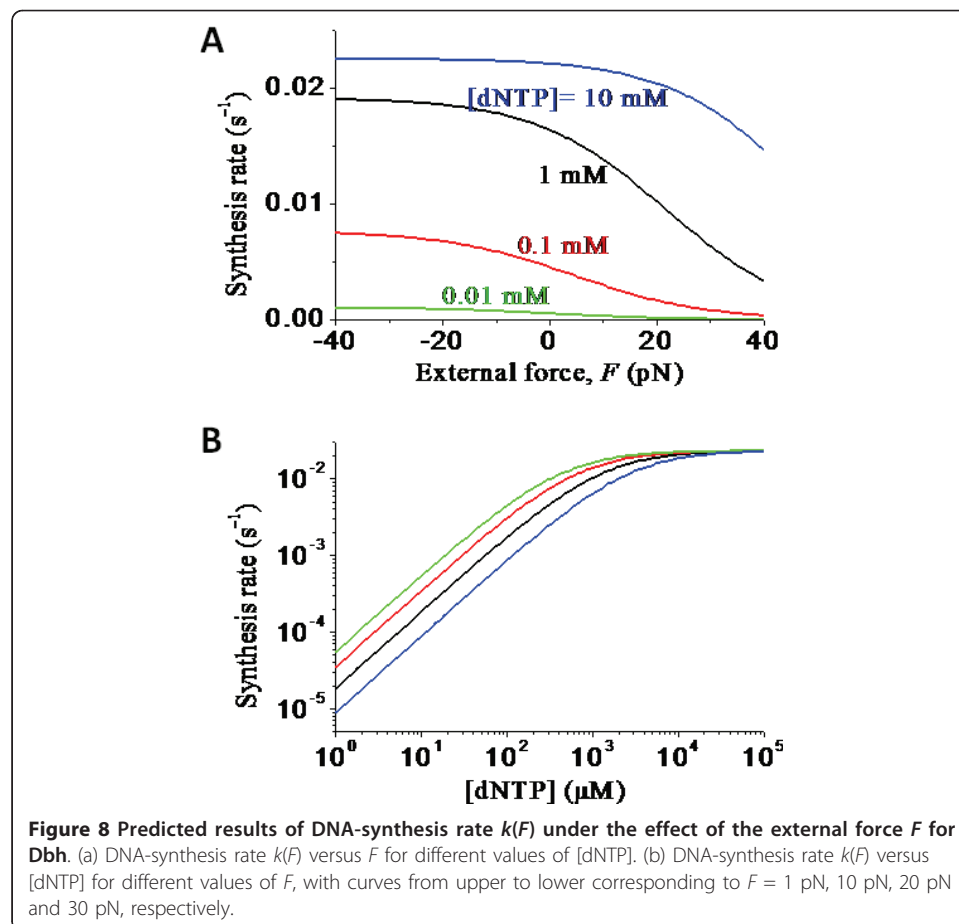
$$k = k_c \frac{[\text{dNTP}]}{[\text{dNTP}] + K_m^{(0)}\dfrac{\left[\exp\left(\dfrac{E_n - E_{n+1}}{k_BT}\right)\exp\left(\dfrac{Fp}{k_BT}\right) + 1\right]}{\left[\exp\left(\dfrac{E_n - E_{n+1}}{k_BT}\right) + 1\right]}}. \tag{11}$$

where k_c is the dNTP-incorporation rate at saturating dNTP concentration and $K_m^{(0)} = k_c / k_b^{(0)}$. It is noted from Eqs. (8) - (11) that the DNA-synthesis rate k is independent of the viscosity η.
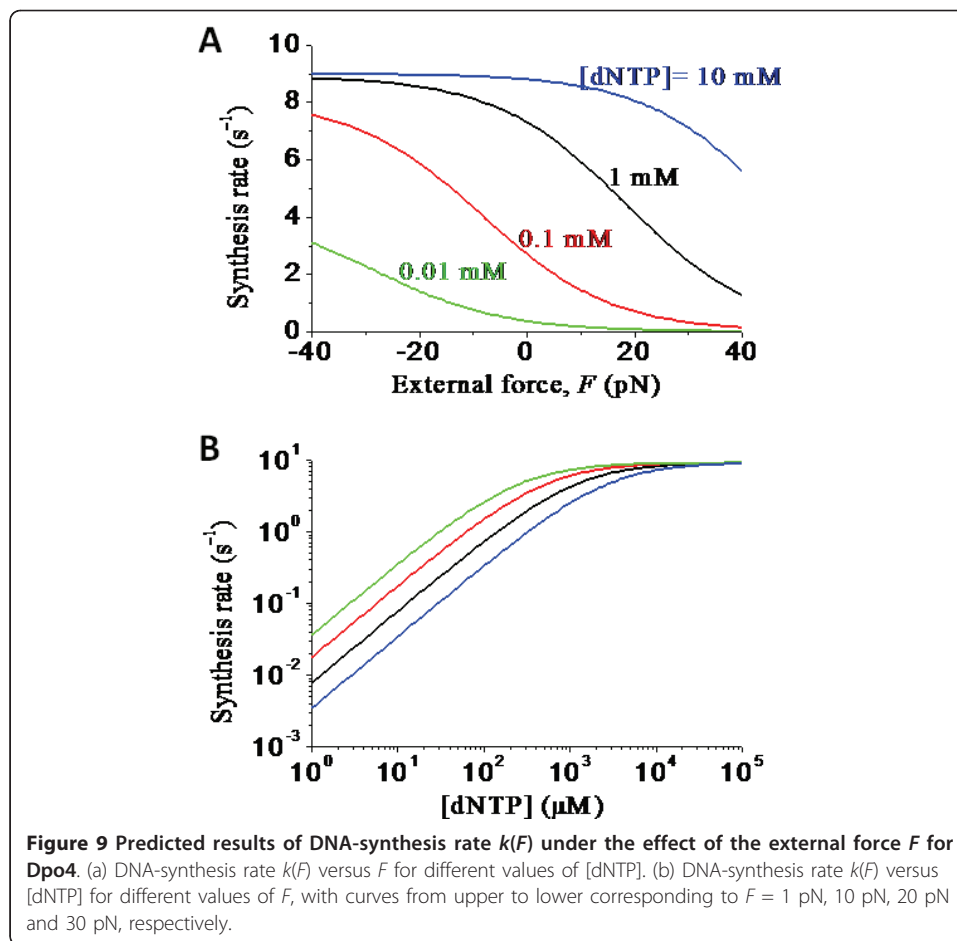
Now, we use Eq. (11) to make some predicted results. From the experimental data, we have $k_c = 2.3 \times 10^{-2}$ s^{-1} and $K_m^{(0)} = 0.4$ mM for Dbh [54]. As discussed before, we take $E_{n+1} \approx E_n$ for Dbh. Using Eq. (11), we calculate the DNA-synthesis rate k versus F for different values of [dNTP], with the results shown in Figure 8a, and 8k versus [dNTP] for different values of F, with the results shown in Figure 8b. For Dpo4, $k_c = 9$ s^{-1} and $K_m^{(0)} = 230$ μM [44]. Moreover, we take $E_n - E_{n+1} = 3k_BT$ for Dpo4. The results of k versus F for different values of [dNTP] and k versus [dNTP] for different values of F are shown in Figures 9a and 9b, respectively. By comparing Figure 8a with Figure 9a, it is seen that, in the range of F = -20 ~ 20 pN and [dNTP] ≤ 1 mM, the external force F has more significant effect on the DNA-synthesis rate k of Dpo4 than on that of Dbh.

Comparison of the effect of external force on dNTP-binding rate of the Y-family Pol with that of the replicative Pol

Based on the model for the replicative Pol (Figure 2) and the modified model for the Y-family Pol (Figures 3, 4, and 5), the dNTP-binding rate $k_b(F)$ versus the external
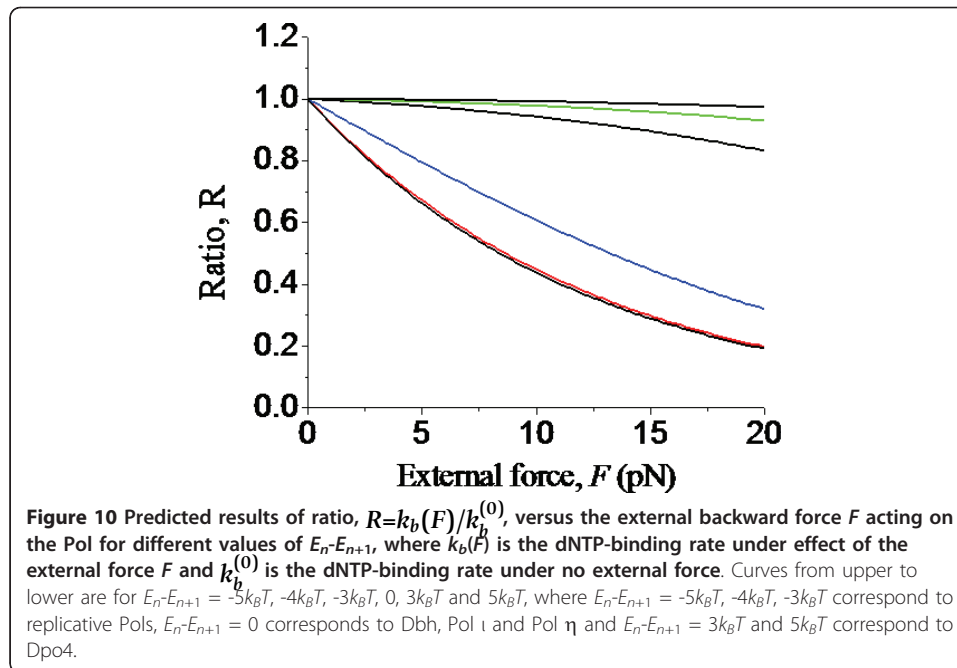

Figure 8 Predicted results of DNA-synthesis rate *k(F)* under the effect of the external force *F* for Dbh. (a) DNA-synthesis rate *k(F)* versus *F* for different values of [dNTP]. (b) DNA-synthesis rate *k(F)* versus [dNTP] for different values of *F*, with curves from upper to lower corresponding to *F* = 1 pN, 10 pN, 20 pN and 30 pN, respectively.

Figure 9 Predicted results of DNA-synthesis rate $k(F)$ under the effect of the external force F for Dpo4. (a) DNA-synthesis rate $k(F)$ versus F for different values of [dNTP]. (b) DNA-synthesis rate $k(F)$ versus [dNTP] for different values of F, with curves from upper to lower corresponding to $F = 1$ pN, 10 pN, 20 pN and 30 pN, respectively.

force F satisfies Eq. (10), where $E_n-E_{n+1} = E'_1-E_1 < 0$ for the replicative Pol (see Figure 2) while $E_n-E_{n+1} \geq 0$ for the Y-family Pol. For the replicative Pol, E'_1-E_1 represents the binding affinity of binding site S_1 for one base of the ssDNA template (or the binding affinity of the binding site S_1 residue that is closest to the palm subdomain for ssDNA), and it is estimated that $E'_1-E_1 = -5k_BT \sim -3k_BT$. As mentioned before, for Dbh, Pol ι and Pol η, $E_n-E_{n+1} \approx 0$, while for Dpo4, $E_n-E_{n+1} = 3k_BT \sim 5k_BT$.

Using Eq. (10), the calculated results of ratio $R = k_b(F)\big/k_b^{(0)}$ versus the external backward force F acting on the Pol for different values of E_n-E_{n+1} are shown in Figure 10. It is seen that, in the range of $F < 20$ pN, with the increase of F the ratio R only decreases slightly for the replicative Pol whereas the ratio R decreases greatly for the Y-family Pol. In other words, the external backward force has much more effect on the the dNTP-binding rate for the Y-family Pol than for the replicative Pol.

The biological implication of these different characteristics between the Y-family and replicative Pols might be imagined as follows. At the replication fork, the replicative Pol generally feels a backward force by the front DNA helicase [73]. Thus, the slight effect of the backward force on the DNA synthesis rate by the replicative Pols is purposed to have little impact on the DNA replication. However, when the replicative Pol becomes stall at the lesion site, since the front helicase is still unwinding the dsDNA, the Pol would not feel a backward force now. Thus, when the relicative Pol is replaced
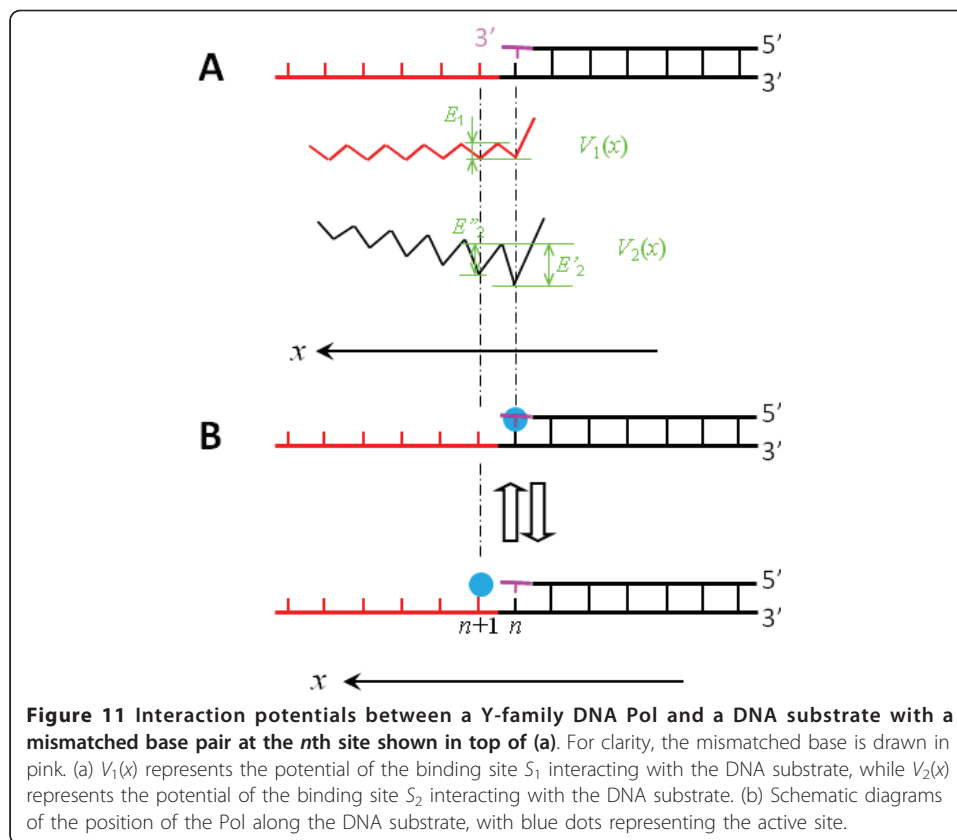
Figure 10 Predicted results of ratio, $R=k_b(F)/k_b^{(0)}$, versus the external backward force F acting on the Pol for different values of E_n-E_{n+1}, where $k_b(F)$ is the dNTP-binding rate under effect of the external force F and $k_b^{(0)}$ is the dNTP-binding rate under no external force. Curves from upper to lower are for E_n-E_{n+1} = -5k_BT, -4k_BT, -3k_BT, 0, 3k_BT and 5k_BT, where E_n-E_{n+1} = -5k_BT, -4k_BT, -3k_BT correspond to replicative Pols, E_n-E_{n+1} = 0 corresponds to Dbh, Pol ι and Pol η and E_n-E_{n+1} = 3k_BT and 5k_BT correspond to Dpo4.

by the Y-family Pol, the latter would also not feel a backward force at the lesion site. After bypass the lesion site, the Y-family Pol would continue to make processive DNA synthesis and, if the Y-family Pol catches up with the helicase, the backward force induced by the front helicase would greatly reduce the DNA synthesis rate, thus enhancing the probability of the Y-family Pol to dissociate from the DNA substrate or the probability of the Y-family Pol to be replaced by the replicative Pol.

The Y-family Pol can easily bypass a mismatched base pair or a lesion site
In this section, we will show how the Pol that uses the NBR mechanism for translocation can easily bypass a mismatched base pair or a lesion site. As an example, we will use Dpo4 to illustrate this bypass ability, in which the active site is very close along the x direction to the nearest residue of the binding site S_2 located in the LF domain.

As shown in Figure 11, consider that a mismatched base is incorporated at the nth site. Then, the interaction potential $V_1(x)$ of the binding site S_1 with the ssDNA template and the potential $V_2(x)$ of the binding site S_2 with the dsDNA segment are shown in Figure 11a. Here, E_1 is the binding affinity of the binding site S_1 for N_1 bases of the ssDNA template that the binding site S_1 can cover, E'_2 is the binding affinity of the binding site S_2 for the backbones connecting (N_1-1) base pairs on the dsDNA, and E''_2 is the binding affinity of the binding site S_2 for the backbones connecting (N_1-2) base pairs. Thus, when the active site is positioned at the nth site (top of Figure 11b), the affinity of the Pol for the DNA substrate is $E_n = E_1 + E'_2$; while when the active site is positioned at the (n+1)th site (bottom of Figure 11b), the affinity is $E_{n+1} = E_1 + E''_2$. Note that the binding affinity E''_2 that corresponds to binding (N_1-2) base pairs is smaller than E'_2 that corresponds to binding (N_1-1) base pairs.

Now we compare the case of a matched incorporation at the nth site (Figure 3) with the case of a mismatched incorporation (Figure 11). Since E'_1 and E_1 are much smaller than E'_2 and E_2, we have $E_n^{(match)} > E_n^{(mismatch)}$, where $E_n^{(match)}$ and $E_n^{(mismatch)}$ represent E_n

Figure 11 Interaction potentials between a Y-family DNA Pol and a DNA substrate with a mismatched base pair at the *n*th site shown in top of (a). For clarity, the mismatched base is drawn in pink. (a) $V_1(x)$ represents the potential of the binding site S_1 interacting with the DNA substrate, while $V_2(x)$ represents the potential of the binding site S_2 interacting with the DNA substrate. (b) Schematic diagrams of the position of the Pol along the DNA substrate, with blue dots representing the active site.

for cases of matched incorporation and mismatched incorporation, respectively. Thus, from the results shown in Figure 7, it is seen that the mean moving time $T_{n\rightarrow(n+1)}$ for the case of mismatched incorporation is shorter than the case of matched incorporation. Moreover, it is noted that the value of $E_n^{(mismatch)} - E_{n+1}^{(mismatch)} = E_2' - E_2''$ is close to that of $E_n^{(match)} - E_{n+1}^{(match)} \approx E_2 - E_2'$, since both value of $E_2'-E_2''$ and that of E_2-E_2' correspond to the affinity of the binding site S_2 for the backbones connecting one base pair on the dsDNA. Thus, from Eq. (8), the ratio $(T_n/T_{n+1})^{(mismatch)}$ is close to $(T_n/T_{n+1})^{(match)}$. As a result, we conclude that the Pol that uses the NBR mechanism for translocation has nearly the same rate to bypass a mismatched base pair as that to bypass a matched base pair. Similarly, for the case of an abasic lesion located at the *n*th site, the interaction potential $V_2(x)$ is the same as shown Figure 11a. Thus, as in the case of a mismatched base, the Pol that uses the NBR mechanism for translocation can also easily bypass the lesion site.

However, since $E_{n+1}^{(mismatch)} \approx E_2'' < E_{n+1}^{(match)} \approx E_2'$, it is noted from Figure 6a that the dissociation probability near the lesion site is larger than that in the absence of the lesion or mismatched base, which is in agreement with the experimental data for Dpo4 [52]. On the hand, it is seen from Figure 6b that, when there is no lesion or no mismatched base, for $N_p = 10$ that corresponds to the case of Dbh [48]$E_r = 18.5k_BT$, while for $N_p = 100$ that corresponds to the case of Dpo4 [48]$E_r = 20.8k_BT$. From the value of $E_1-E_1' = 3k_BT \sim 5k_BT$ (see above), we infer that $E_1'-E_1'' \approx 3k_BT \sim 5k_BT$. Thus, at the lesion site, for the case of Dbh with $E_r = 13.5k_BT \sim 15.5k_BT$, we have $N_p \approx 0.08 \sim 0.6$ from Figure 6b; for the case of Dpo4 with $E_r = 15.8k_BT \sim 17.8k_BT$, we have $N_p \approx 0.8 \sim$

5.2 from Figure 6b. This implies that, at the lesion site, Dbh is prone to dissociate from the DNA substrate while Dpo4 is not easily to dissociate. In other words, Dpo4 can bypass the lesion site, yet Dbh does so with a much lower efficiency. These are consistent with the experimental data [24,48,50,52]. Moreover, as the different lesion-bypassing abilities of the two Pols, Dpo4 and Dbh, are due to the different binding affinities, which in turn result mainly from the different interaction strengths of the LF domain with the DNA, it is inferred that, by interchanging the LF domains, the lesion-bypassing abilities of the two Pols will be exchanged. This is also consistent with the experimental data of Boudsocq et al. [48].

Similarly, we can easily show that the Pols such as Dbh and Pol ι, in which the active site is, along the x direction, not close to the nearest residue of the binding site S_2 located in the LF domain, can also easily bypass the lesion site. Thus, we conclude that, although different values of distance L give different translocation features, all the Y-family Pols that use the NBR mechanism for translocation can easily bypass the lesion site, thus performing the translesion synthesis. By contrast, the replicative Pols that use other Brownian ratchet mechanism for translocation cannot bypass the lesion site and is thus unable to perform the translesion synthesis (see Figures 1 and 2).

Discussion

The explanation of the lesion-bypassing ability by a Y-family Pol in this work is focused only on the translocation activity of the active site from the position opposite to the lesion site to the next position. In fact, the lesion-bypassing ability is dictated by two activities. One is the translocation activity while the other one is the catalytic activity of the phosphodiester bond formation. The latter activity determines that rates of nucleotide incorporation opposite to the lesion site and one position downstream from the lesion site are slower than those at other sites [46]. Moreover, since the latter activity is determined by the structure of the catalytic core, single amino acid substitutions within the active site, palm or fingers subdomains can also have a profound effect on the ability of the enzyme to perform translesion synthesis [74,75].

Distinct translocation features among the Y-family Pols in the NBR model depend on the distance L from the active site to the nearest residue of the binding site S_2 located in the LF domain. It is important to note that this distance L in the model corresponds to the structure of the Pol only after binding to its DNA substrate. Although available structures showed a significant conformational change in the LF domain of the apo-Dpo4 upon binding to the DNA substrate [42], only a little change has been observed between the Pol complexed with the DNA substrate alone and that with both the DNA substrate and the dNTP nucleotide. Thus, the large conformational change in the LF domain of the Pol upon binding to the DNA substrate has no effect on the conclusion of the current work which is involved only with the structures of the Pol either complexed with the DNA substrate alone or complexed with both the DNA substrate and the dNTP nucleotide.

Further comments on the NBR translocation model

In the NBR model (Figure 5), it has been implicitly considered that the Pol has a rigid structure, i.e., different domains have been considered to be linked rigidly. In fact, the residues linking different domains may behave elastically. For example, for Dpo4,

considering an elastic link between the palm and thumb domains. Upon nucleotide binding, after the active site, together with the fingers, palm and LF domains, move from the nth site to the $(n+1)$th site, the thumb domain may not move simultaneously due to the elastic link between the palm and thumb domains, i.e., the thumb contacts with the DNA may fluctuate by the thermal noise between pre- and post-translocation positions with nearly equal probability. This gives an explanation to the available structural data for Dpo4 showing that, upon nucleotide binding, the LF contacts with the DNA shift by one base pair but the thumb contacts do not shift simultaneously [41].

Potential implication of binding site S_1 in the induced-fit mechanism

The strong interaction of the binding site S_1 with the unpaired base on the template induces the conformational change in the residues of the binding site S_1, which in turn results in the conformational change in the active site that is adjacent to the residues of the binding site S_1. This unpaired-base-related conformational change thus results in the active site having a much higher affinity for the structurally compatible nucleotide than structurally incompatible nucleotides. This argument is consistent with the experimental data for high-fidelity DNA Pols showing that the shape of the nascent base pair is important regardless of whether the Watson-Crick hydrogen bonds can be formed [76]. It is also consistent with the recent FRET-based assay on Klenow fragment showing that base discrimination takes place within the open complex rather than occurs during the transition from open to closed fingers conformations [77]. Conversely, the interaction of the structurally incompatible nucleotide may have a negative effect on the conformational change in the active site, which in turn results in the inverse conformational change in the binding site S_1, potentially reducing its binding affinity for the unpaired base on the template. This could account for the increased dissociation caused by the binding of a mismatched nucleotide, as observed by Joyce et al. [77].

Crystal structures of the Y-family DNA Pol complexed with DNA substrate showed weak or no interaction of the Pol with the ssDNA template [27,28,32,41,42]. The weak binding affinity would result in a minor or no conformational change in the active site that is related to the structure of the unpaired base on the template, giving a smaller difference in binding affinity between correct and incorrect nucleotides. This thus results in that the Y-family Pol has a low-fidelity synthesis, which is consistent with the available experimental data [43,45]. Since minor or no conformational change in the active site occurs which is related to the shape of the template base, it is expected that the base discrimination in Y-family Pols should rely mainly on the Watson-Crick hydrogen bonding interaction, in sharp contrast to the high-fidelity replicative DNA Pols where the effect of the Watson-Crick hydrogen bonding interaction can be negligible compared to the dominant effect of the base shape. Indeed, steady state kinetic studies with incoming dNTP and DNA substrates containing difluorotoluene, which has the same shape as thymine but lacks the ability to form Watson-Crick hydrogen bonds, are poor substrates for Y-family pol η [78].

Conclusion

In conclusion, a NBR model is proposed for the translocation of the Y-family DNA Pol along the DNA substrate, which is modified from the translocation model proposed

previously for the replicative DNA Pol. The observed different features of the structures for Dpo4, Dbh and pol ι in binary and ternary forms are consistent with the NBR model. Moreover, since the interaction potential $V_2(x)$ for Pol η has the form of Figure 4b rather than that of Figure 3b, it is predicted that Pol η would show the translocation feature similar to Dbh and Pol ι rather than Dpo4. The obtained theoretical results on dynamic properties of the Y-family Pols by using the NBR model are consistent with the available experimental data. To further verify the model, it is hoped to test the predicted results given in Figures 8, 9, 10.

Additional material

Additional file 1: Dissociation probability P_{d1} during time period T_{p1}.

Authors' contributions
PX is the sole author of this paper and is responsible for developing the concepts and for writing and revising the manuscript.

Competing interests
The author declares that they have no competing interests.

References
1. Braithwaite DK, Ito J: **Compilation, alignment, and phylogenetic relationships of DNA polymerases.** *Nucleic Acids Res* 1993, **21**:787-802.
2. Cann IK, Ishino Y: **Archaeal DNA replication: identifying the pieces to solve a puzzle.** *Genetics* 1999, **152**:1249-1267.
3. Ohmori H, Friedberg EC, Fuchs RP, Goodman MF, Hanaoka F, Hinkle D, Kunkel TA, Lawrence CW, Livneh Z, Nohmi T, Prakash L, Prakash S, Todo T, Walker GC, Wang Z, Woodgate R: **The Y-Family of DNA Polymerases.** *Mol Cell* 2001, **8**:7-8.
4. Moon AF, Garcia-Diaz M, Batra VK, Beard WA, Bebenek K, Kunkel TA, Wilson SH, Pedersen LC: **The X family portrait: structural insights into biological functions of X family polymerases.** *DNA Repair (Amst)* 2007, **6**:1709-1725.
5. Yamtich J, Sweasy JB: **DNA polymerase family X: function, structure, and cellular roles.** *Biochim Biophys Acta* 2010, **1804**:1136-1150.
6. Woodgate R: **A plethora of lesion-replicating DNA polymerases.** *Genes Dev* 1999, **13**:2191-2195.
7. Goodman MF: **Error-prone repair DNA polymerases in prokaryotes and eukaryotes.** *Annu Rev Biochem* 2002, **71**:17-50.
8. Friedberg EC, Wagner R, Radman M: **Specialized DNA polymerases, cellular survival, and the genesis of mutations.** *Science* 2002, **296**:1627-1630.
9. Yang W, Woodgate R: **What a difference a decade makes: Insights into translesion DNA synthesis.** *Proc Natl Acad Sci USA* 2007, **104**:15591-15598.
10. Lehmann AR, Niimi A, Ogi T, Brown S, Sabbioneda S, Wing JF, Kannouche PL, Green CM: **Translesion synthesis: Y-family polymerases and the polymerase switch.** *DNA Repair (Amst)* 2007, **6**:891-899.
11. Pata JD: **Structural diversity of the Y-family DNA polymerases.** *Biochim Biophys Acta* 2010, **1804**:1124-1135.
12. Washington MT, Carlson KD, Freudenthal BD, Pryor JM: **Variations on a theme: Eukaryotic Y-family DNA polymerases.** *Biochim Biophys Acta* 2010, **1804**:1113-1123.
13. Wagner J, Gruz P, Kim S-R, Yamada M, Matsui K, Fuchs RPP, Nohmi T: **The *dinB* gene encodes a novel *E. coli* DNA polymerase, DNA pol IV, involved in mutagenesis.** *Mol Cell* 1999, **4**:281-286.
14. Reuven NB, Arad G, Maor-Shoshani A, Livneh Z: **The mutagenesis protein UmuC is a DNA polymerase activated by UmuD', RecA, and SSB and is specialized for translesion replication.** *J Biol Chem* 1999, **274**:31763-31766.
15. Tang M, Shen X, Frank EG, O'Donnell M, Woodgate R, Goodman MF: **UmuD'$_2$C is an error-prone DNA polymerase, *Escherichia coli* pol V.** *Proc Natl Acad Sci USA* 1999, **96**:8919-8924.
16. Johnson RE, Prakash S, Prakash L: **Efficient bypass of a thymine-thymine dimer by yeast DNA polymerase, pol η.** *Science* 1999, **283**:1001-1004.
17. Nelson JR, Lawrence CW, Hinkle DC: **Deoxycytidyl transferase activity of yeast *REV1* protein.** *Nature* 1996, **382**:729-731.
18. Masutani C, Kusumoto R, Yamada A, Dohmae N, Yokoi M, Yuasa M, Araki M, Iwai S, Takio K, Hanaoka F: **The *XPV* (xeroderma pigmentosum variant) gene encodes human DNA polymerase η.** *Nature* 1999, **399**:700-704.
19. Johnson RE, Washington MT, Haracska L, Prakash S, Prakash L: **Eukaryotic polymerasesι and ζ act sequentially to bypass DNA lesions.** *Nature* 2000, **406**:1015-1019.
20. Tissier A, McDonald JP, Frank EG, Woodgate R: **pol ι, a remarkably error-prone human DNA polymerase.** *Genes Dev* 2000, **14**:1642-1650.

21. Ohashi E, Ogi T, Kusumoto R, Iwai S, Masutani C, Hanaoka F, Ohmori H: **Error-prone bypass of certain DNA lesions by the human DNA polymerase κ.** *Genes Dev* 2000, **14**:1589-1594.

22. Gibbs PEM, Wang X-D, Li Z, McManus TP, McGregor WG, Lawrence CW, Maher VM: **The function of the human homolog of *Saccharomyces cerevisiae* REV1 is required for mutagenesis induced by UV light.** *Proc Natl Acad Sci USA* 2000, **97**:4186-4191.

23. Kulaeva OI, Koonin EV, McDonald JP, Randall SK, Rabinovich N, Connaughton JF, Levine AS, Woodgate R: **Identification of a DinB/UmuC homolog in the archeon *Sulfolobus solfataricus*.** *Mutat Res* 1996, **357**:245-253.

24. Boudsocq F, Iwai S, Hanaoka F, Woodgate R: ***Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4): an archaeal DNA polymerase with lesion-bypass properties akin to eukaryotic pol η.** *Nucleic Acids Res* 2001, **29**:4607-4616.

25. Zhou BL, Pata JD, Steitz TA: **Crystal structure of a DinB lesion bypass DNA polymerase catalytic fragment reveals a classic polymerase catalytic domain.** *Mol Cell* 2001, **8**:427-437.

26. Silvian LF, Toth EA, Pham P, Goodman MF, Ellenberger T: **Crystal structure of a DinB family error-prone DNA polymerase from *Sulfolobus solfataricus*.** *Nat Struct Biol* 2001, **8**:984-989.

27. Wilson RC, Pata JD: **Structural insights into the generation of single-base deletions by the Y family DNA polymerase Dbh.** *Mol Cell* 2008, **29**:767-779.

28. 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* 2001, **107**:91-102.

29. Trincao J, Johnson RE, Escalante CR, Prakash S, Prakash L, Aggarwal AK: **Structure of the catalytic core of *S. cerevisiae* DNA polymerase η: implications for translesion DNA synthesis.** *Mol Cell* 2001, **8**:417-426.

30. Alt A, Lammens K, Chiocchini C, Lammens A, Pieck JC, Kuch D, Hopfner KP, Carell T: **Bypass of DNA lesions generated during anticancer treatment with cisplatin by DNA polymerase eta.** *Science* 2007, **318**:967-970.

31. Nair DT, Johnson RE, Prakash S, Prakash L, Aggarwal AK: **Replication by human DNA polymerase-iota occurs by Hoogsteen base-pairing.** *Nature* 2004, **430**:377-380.

32. Jain R, Nair DT, Johnson RE, Prakash L, Prakash S, Aggarwal AK: **Replication across template T/U by human DNA polymerase-iota.** *Structure* 2009, **17**:974-980.

33. Uljon SN, Johnson RE, Edwards TA, Prakash S, Prakash L, Aggarwal AK: **Crystal structure of the catalytic core of human DNA polymerase kappa.** *Structure* 2004, **12**:1395-1404.

34. Vasquez-Del Carpio R, Silverstein TD, Lone S, Swan MK, Choudhury JR, Johnson R, Prakash S, Prakash L, Aggarwal A: **Structure of human DNA polymerase kappa inserting dATP opposite an 8-OxoG DNA lesion.** *PLoS ONE* 2009, **4**: e5766.

35. Nair D, Johnson R, Prakash L, Prakash S, Aggarwal A: **Rev1 employs a novel mechanism of DNA synthesis using a protein template.** *Science* 2005, **309**:2219-2222.

36. Ling H, Boudsocq F, Plosky BS, Woodgate R, Yang W: **Replication of a *cis-syn* thymine dimer at atomic resolution.** *Nature* 2003, **424**:1083-1087.

37. Ling H, Boudsocq F, Woodgate R, Yang W: **Snapshots of replication through an abasic lesion; structural basis for base substitutions and frameshifts.** *Mol Cell* 2004, **13**:751-762.

38. Ling H, Sayer JM, Plosky BS, Yagi H, Boudsocq F, Woodgate R, Jerina DM, Yang W: **Crystal structure of a benzo[a] pyrene diol epoxide adduct in a ternary complex with a DNA polymerase.** *Proc Natl Acad Sci USA* 2004, **101**:2265-2269.

39. Trincao J, Johnson RE, Wolfle WT, Escalante CR, Prakash S, Prakash L, Aggarwal AK: **Dpo4 is hindered in extending a GT mismatch by a reverse wobble.** *Nat Struct Mol Biol* 2004, **11**:457-462.

40. Vaisman A, Ling H, Woodgate R, Yang W: **Fidelity of Dpo4: effect of metal ions, nucleotide selection and pyrophosphorolysis.** *EMBO J* 2005, **24**:2957-2967.

41. Rechkoblit O, Malinina L, Cheng Y, Kuryavyi V, Broyde S, Geacintov NE, Patel DJ: **Stepwise translocation of Dpo4 polymerase during error-free bypass of an oxoG lesion.** *PLoS Biol* 2006, **4**:e11.

42. Wong JH, Fiala KA, Suo Z, Ling H: **Snapshots of a Y-family DNA polymerase in replication: substrate-induced conformational transitions and implications for fidelity of Dpo4.** *J Mol Biol* 2008, **379**:317-330.

43. Kokoska RJ, Bebenek K, Boudsocq F, Woodgate R, Kunkel TA: **Low fidelity DNA synthesis by a Y family DNA polymerase due to misalignment in the active site.** *J Biol Chem* 2002, **277**:19633-19638.

44. Fiala KA, Suo Z: **Mechanism of DNA polymerization catalyzed by *Sulfolobus solfataricus* P2 DNA polymerase IV.** *Biochemistry* 2004, **43**:2116-2125.

45. Fiala KA, Suo Z: **Pre-steady-state kinetic studies of the fidelity of *Sulfolobus solfataricus* P2 DNA polymerase IV.** *Biochemistry* 2004, **43**:2106-2115.

46. Fiala KA, Hypes CD, Suo Z: **Mechanism of abasic lesion bypass catalyzed by a Y-family DNA polymerase.** *J Biol Chem* 2007, **282**:8188-8198.

47. Washington MT, Prakash L, Prakash S: **Yeast DNA polymerase η utilizes an induced-fit mechanism of nucleotide incorporation.** *Cell* 2001, **107**:917-927.

48. Boudsocq F, Kokoska RJ, Plosky BS, Vaisman A, Ling H, Kunkel TA, Yang W, Woodgate R: **Investigating the role of the little finger domain of Y-family DNA polymerases in low fidelity synthesis and translesion replication.** *J Biol Chem* 2004, **279**:32932-32940.

49. Gruz P, Pisani FM, Shimizu M, Yamada M, Hayashi I, Morikawa K, Nohmi T: **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* 2001, **276**:47394-47401.

50. Potapova O, Grindley ND, Joyce CM: **The mutational specificity of the Dbh lesion bypass polymerase and its implications.** *J Biol Chem* 2002, **277**:28157-28166.

51. McCulloch SD, Kokoska RJ, Masutani C, Iwai S, Hanaoka F, Kunkel TA: **Preferential *cis-syn* thymine dimer bypass by DNA polymerase η occurs with biased fidelity.** *Nature* 2004, **428**:97-100.

52. Kokoska RJ, McCulloch SD, Kunkel TA: **The efficiency and specificity of apurinic/apyrimidinic site bypass by human DNA polymerase η and *Sulfolobus solfataricus* Dpo4.** *J Biol Chem* 2003, **278**:50537-50545.

53. Zang H, Goodenough AK, Choi JY, Irimia A, Loukachevitch LV, Kozekov ID, Angel KC, Rizzo CJ, Egli M, Guengerich FP: **DNA adduct bypass polymerization by *Sulfolobus solfataricus* DNA polymerase Dpo4.** *J Biol Chem* 2005, **280**:29750-29764.

54. DeLucia AM, Grindley ND, Joyce CM: **An error-prone family Y DNA polymerase (DinB homolog from Sulfolobus solfataricus) uses a 'steric gate' residue for discrimination against ribonucleotides.** *Nucleic Acids Res* 2003, **31**:4129-4137.
55. DeLucia AM, Grindley ND, Joyce CM: **Conformational changes during normal and error-prone incorporation of nucleotides by a Y-family DNA polymerase detected by 2-aminopurine fluorescence.** *Biochemistry* 2007, **46**:10790-10803.
56. Xu C, Maxwell BA, Brown JA, Zhang L, Suo Z: **Global conformational dynamics of a Y-family DNA polymerase during catalysis.** *PLoS Biol* 2009, **7**:e1000225.
57. Xie P: **Model for forward polymerization and switching transition between polymerase and exonuclease sites by DNA polymerase molecular motors.** *Arch Biochem Biophys* 2007, **457**:73-84.
58. Xie P: **A possible mechanism for the dynamics of transition between polymerase and exonuclease sites in a high-fidelity DNA polymerase.** *J Theor Biol* 2009, **259**:434-439.
59. Xie P: **A polymerase-site-jumping model for strand transfer during DNA synthesis by reverse transcriptase.** *Virus Res* 2009, **144**:65-73.
60. Magnasco MO, Stolovitzky G: **Feynman's ratchet and pawl.** *J Stat Phys* 1998, **93**:615-632.
61. Kay ER, Leigh DA, Zerbetto F: **Synthetic molecular motors and mechanical machines.** *Agnew Chem Int Ed* 2007, **46**:72-191.
62. Delagoutte E, von Hippel PH: **Function and assembly of the Bacteriophage T4 DNA replication complex.** *J Biol Chem* 2003, **278**:25435-25447.
63. Turner RM, Grindley NDF, Joyce CM: **Interaction of DNA polymerase I (Klenow fragment) with the single-stranded template beyond the site of synthesis.** *Biochemistry* 2003, **42**:2373-2385.
64. Datta K, Wowor AJ, Richard AJ, LiCata VJ: **Temperature dependence and thermodynamics of Klenow polymerase binding to primed-template DNA.** *Biophys J* 2006, **90**:1739-1751.
65. Doublie S, Sawaya MR, Ellenberger T: **An open and closed case for all polymerases.** *Structure* 1999, **7**:R31-R35.
66. Li Y, Korolev S, Waksman G: **Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of thermus aquaticus DNA polymerase I: structural basis for nucleotide incorporation.** *EMBO J* 1998, **17**:7514-7525.
67. Huang H, Chopra R, Verdine GL, Harrison SC: **Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance.** *Science* 1998, **282**:1669-1675.
68. Donlin MJ, Patel SS, Johnson KA: **Kinetic partitioning between the exonuclease and polymerase sites in DNA error correction.** *Biochemistry* 1991, **30**:538-546.
69. Hogg M, Wallace SS, Doublie S: **Crystallographic snapshots of a replicative DNA polymerase encountering an abasic site.** *EMBO J* 2004, **23**:1483-1493.
70. Li Y, Dutta S, Doublie S, Bdour HM, Taylor JS, Ellenberger T: **Nucleotide insertion opposite a cis-syn thymine dimer by a replicative DNA polymerase from bacteriophage T7.** *Nat Struct Mol Biol* 2004, **11**:784-790.
71. Luby-Phelps K, Mujumdar S, Mujumdar RB, Ernst LA, Galbraith W, Waggoner AS: **A novel fluorescence ratiometric method confirms the low solvent viscosity of the cytoplasm.** *Biophys J* 1993, **65**:236-242.
72. Gardiner CW: **Handbook of stochastic methods for physics, chemistry and the natural sciences.** *Springer-Verlag, Berlin* 1983.
73. Stano NM, Jeong Y-J, Donmez I, Tummalapalli P, Levin MK, Patel SS: **DNA synthesis provides the driving force to accelerate DNA unwinding by a helicase.** *Nature* 2005, **435**:370-373.
74. Boudsocq F, Ling H, Yang W, Woodgate R: **Structure-based interpretation of missense mutations in Y-family DNA polymerases and their implications for polymerase function and lesion bypass.** *DNA Repair (Amst)* 2002, **1**:343-358.
75. Glick E, Vigna KL, Loeb LA: **Mutations in human DNA polymerase η motif II alter bypass of DNA lesions.** *EMBO J* 2001, **20**:7303-7312.
76. Moran S, Ren RXF, Kool ET: **A thymidine triphosphate shape analog lacking Watson-Crick pairing ability is replicated with high sequence selectivity.** *Proc Natl Acad Sci USA* 1997, **94**:10506-10511.
77. Joyce CM, Potapova O, DeLucia AM, Huang X, Basu VP, Grindley NDF: **Fingers-closing and other rapid conformational changes in DNA polymerase I (Klenow fragment) and their role in nucleotide selectivity.** *Biochemistry* 2008, **47**:6103-6116.
78. Washington MT, Helquist SA, Kool ET, Prakash L, Prakash S: **Requirement of Watson-Crick hydrogen bonding for DNA synthesis by yeast DNA polymerase eta.** *Mol Cell Biol* 2003, **23**:5107-5112.