Published in final edited form as:
Cell. 2009 December 24; 139(7): 1279–1289. doi:10.1016/j.cell.2009.11.043.

Structural Insight Into Translesion Synthesis By DNA Pol II

Feng Wang and **Wei Yang**
Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, 9000 Rockville Pike, Bldg. 5, Rm B1-03, Bethesda, MD 20892, USA

Summary

E. coli DNA Pol II and eukaryotic Rev3 are B-family polymerases that can extend primers past a damaged or mismatched site when the high-fidelity replicative polymerases in the same family are ineffective. We report here the biochemical and structural properties of DNA Pol II that facilitate this translesion synthesis. DNA Pol II can extend primers past lesions either directly or by template skipping, in which small protein cavities outside of the active site accommodate looped-out template nucleotides one or two basepairs upstream. Because of multiple looping-out alternatives, mutation spectra of bypass synthesis are complicated. Moreover, translesion synthesis is enhanced by altered partitioning of DNA substrate between the polymerase active site and the proofreading exonuclease site. Compared to the replicative B-family polymerases, DNA Pol II has subtle amino acid changes remote from the active site that allow it to replicate normal DNA with high efficiency yet conduct translesion synthesis when needed.

Introduction

DNA polymerases are divided into six families, A, B, C, D, X and Y based on sequence conservation (Bebenek and Kunkel, 2004; Joyce and Benkovic, 2004). Replication is normally carried out by the A, B or C-family polymerases with high fidelity and high processivity. DNA lesions due to loss of bases or chemical modifications, which prevent normal Watson-Crick (WC) pairing and stall replicative polymerases, require specialized polymerases to bypass the road block or DNA recombination to switch template and avoid the impasse (Andersen et al., 2008; Chang and Cimprich, 2009). Most specialized translesion polymerases belong to the Y family and are distinct from replicative polymerases in sequence and structure (Ohmori et al., 2001; Yang and Woodgate, 2007). However, among the most widely spread B-family polymerases, *E. coli* DNA pol II (hereafter referred to as Pol II) and eukaryotic pol ζ are known for translesion and mutagenic DNA synthesis (Al Mamun and Humayun, 2006; Becherel and Fuchs, 2001; Bonner et al., 1990; Gan et al., 2008; Iwasaki et al., 1991; Kroeger et al., 2004; Lawrence and Maher, 2001; Paz-Elizur et al., 1996). How a polymerase possessing the sequence motifs for high-fidelity DNA synthesis and homologous to the mammalian replicases pol δ and ε is capable of TLS presents an enigma.

To date two physical features have been identified that differentiate the replicases in the A, B and C families from repair polymerases in the X- and Y-family. The first is proofreading. All replicative polymeases have a 3′ – 5′ exonuclease that can remove misincorporated nucleotides at the 3′ end of a primer (Bebenek and Kunkel, 2004; Reha-Krantz, 2009). X- and Y-family polymerases have no intrinsic exonuclease activity and cannot proofread

Correspondence and request for material should be addressed to W.Y., Wei.Yang@nih.gov, phone: (301) 402-4645, fax: (301) 496-0201.

(Moon et al., 2007; Yang and Woodgate, 2007). Similarly, DNA pol ζ (Lawrence and Maher, 2001) and human pol ν and pol θ of family? A (Arana et al., 2007; Sharief et al., 1999), which carry out translesion and mutagenic DNA synthesis, also lack the proofreading exonuclease activity. The second feature is the structure and formation of the polymerase active site. The catalytic core of all DNA polymerases consists of palm, finger and thumb domains and uses a two-metal ion mechanism (Moon et al., 2007; Steitz, 1999). Replicative polymerases require a large "closing" movement of the finger domain in the presence of DNA, an incoming nucleotide and metal ions. The "closed" active site is highly complementary to WC base pairs and incompatible with mismatched or damaged bases (Doublie et al., 1999). Similar conformational changes also occur in some but not all X-family polymerases (Moon et al., 2007). In contrast, the active site of Y-family polymerases is usually preformed in the absence of an incoming nucleotide or DNA substrates and has room to accommodate DNA lesions and abnormal base pairs (Yang and Woodgate, 2007).

Pol II retains the sequence motifs of DNA replicases and possesses a 3′– 5′ exonuclease activity (Cai et al., 1995). However, it is induced during SOS responses with the effects of an increased mutation rate and enhanced survival and evolutionary fitness (Al Mamun, 2007; Napolitano et al., 2000; Yeiser et al., 2002). In vitro, Pol II has been shown to bypass abasic (AP) lesions, εC, and AAF-adducts (Al Mamun and Humayun, 2006; Becherel and Fuchs, 2001; Paz-Elizur et al., 1996). In vivo, other than its unequivocal role in two-nucleotide deletion when bypassing an AAF adduct on a NarI site (Fuchs and Fujii, 2007), Pol II leaves no discernible TLS and mutagenic signature.

Crystals of Pol II were reported in 1994 (Anderson et al., 1994) and the structure was determined and deposited at PDB (www.rcsb.org/pdb/, 1Q8I) in the course of a structural genomic project. As expected from sequence conservation, the structure is highly homologous to the B-family replicases of bacterial phages RB69 and ϕ29 (PDB: 1IH7, 1XHX), herpes simplex virus (PDB: 2GV9), and archaeal organisms (PDB: 1QQC, 1WNS, 1TGO, 1D5A, 2JGU, 1QHT). Despite the extensive structural and mechanistic insight into the fidelity of replicative and repair polymerases (Bebenek and Kunkel, 2004; Joyce and Benkovic, 2004; Moon et al., 2007; Yang and Woodgate, 2007), how and why a B-family polymerase be involved in TLS remains an open question.

To solve this puzzle, we have determined and report here eight crystal structures of Pol II, including apo, binary and ternary complexes with normal and AP lesion containing DNAs. When the first five crystal structures yielded no clue to the TLS mechanism of Pol II, we turned to functional assays for the answers. Combining structural and kinetic studies, we show that Pol II is proficient in copying normal DNAs with high fidelity and, in addition it can efficiently extend DNA primers after a variety of lesions. Pol II's TLS ability can be attributed to an altered partitioning between the polymerase and exonuclease active sites and relaxed interactions with the upstream template. Surprisingly, the structural changes that differentiate replicative and repair polymerases in the B family are not located in the polymerase active site.

Results

Structural Basis for DNA synthesis by Pol II

Crystal structures of Pol II alone and in complexes with template-primer DNA and each of four dNTPs (G, A, T and C) as the incoming nucleotide were determined at 1.9–2.4 Å resolutions by molecular replacement (Experimental Procedures) (Fig. 1, Table S1-2). To keep DNA substrates intact, an Exo$^-$ mutant (D335N) Pol II was used for all crystallographic studies. Similar to other B family DNA polymerases, the 783-residue Pol II contains five structural domains: the N (aa 1–146 and 368–388), 3′ – 5′ Exo (aa 147-367), palm (aa 403–

465 and 512–634), finger (aa 466–511) and thumb domains (aa 635–783) (Fig. 1A). The catalytic core assumes the shape of a cupped right hand, and the N and Exo domains extend from the finger toward the thumb giving the polymerase a donut shape (Fig. 1B). Apo Pol II was crystallized in the $p2_12_12_1$ space group, which differs from the reported $P2_12_12$ (PDB: 1Q8I,) by doubling one unit-cell dimension. Two Pol II molecules previously related by crystallographic symmetry now belong to one asymmetric unit and are not identical (silver and copper Cα traces in Fig. 1B). Perhaps as a result, the previously disordered tip of the thumb domain (685–744 aa) becomes traceable. After refinement at 2.2Å resolution, the adjacent thumb and Exo domains show the largest rotational differences (3.7° and 6.5°, respectively) in the two non-crystallographic symmetry-related apo structures.

When associated with a 3′-dideoxy template-primer DNA and a correct incoming dNTP, Pol II structures (G-dC, A-dT, T-dA or C-dG) determined between 1.92 and 2.40 Å are superimposable with an RMSD under 0.2 Å over 750 Cα atoms. The catalytic core tracks the DNA minor groove for ~10 bp (Fig. 1C). As expected, the finger domain undergoes a 15° rotation from the “open” apo structure to the “closed” ternary complexes, where it lies on top of the replicating base pair (Fig. 1A, S1). Notably, thumb and Exo domains also rotate by >10° between the apo and ternary complex structures (Fig. 1B).

Two well-ordered Mg^{2+} ions are found in the active site and are coordinated by the two highly conserved and catalytically essential Asp’s (D419 and D547), the triphosphate of dNTP, and water molecules (Fig. 2A). Metal ion A has five ligands, which would be six if the 3′-OH were present. Two downstream nucleotides 5′ to the templating base are held between the N and Exo domains (Fig. 1D). The upstream 13 bp template-primer duplex is predominantly B-form, but the replicating base pair has the A-like 3′-endo instead of 2′-endo sugar pucker. As a result the minor groove of the replicating base pair is shallow and interacts snugly with the palm and finger domains (Fig. 2B).

Pol II is highly similar to the B-family replicases

The overall structures of Pol II-DNA ternary complexes are highly similar to those of replicative ϕ29 and RB69 polymerases gp43 (hereafter referred to as gp43) (Berman et al., 2007; Franklin et al., 2001). They share the conserved metal-ion ligands, the basic residues and steric gate (Y424) that interact with the triphosphate and deoxyribose of the incoming dNTP, respectively (Fig. 2A). They also share the conserved sequence motif that hugs the minor groove of the replicating base pair (Fig. 2B). Furthermore, after superposition of the palm domains of Pol II and gp43, the DNAs of the two ternary complexes have nearly identical structures (Fig. 2C).

At first glance it is striking that the Pol II finger domain is 60 residues shorter and undergoes a much smaller conformational change upon binding dNTP than that of gp43 (Franklin et al., 2001; Lee et al., 2009) (Fig. S1). But the finger domain of ϕ29 polymerase is as short and undergoes as limited conformational changes as Pol II (Berman et al., 2007) (Fig. S1). Therefore the finger domain alone is unlikely to explain TLS by Pol II.

Pol II extends primer after AP lesions

To compare the abilities of Pol II and gp43 in TLS, we used an AP analogue, tetrahydrofuran (THF), as a test case. When the THF was at the templating position (0) and all four dNTPs were supplied, Pol II extended the primer all the way to the end, but gp43 stopped after incorporating one nucleotide opposite the THF (Fig. 3A). Varying the downstream nucleotide 5′ to the THF had no effect on the behavior of either polymerase. When dA was at the 3′ end of the primer opposite the THF, Pol II succeeded in primer

extension, but gp43 failed (Fig. 3B). The ability to extend a primer after a lesion clearly distinguishes Pol II from the replicative gp43.

In the insertion step when a single dNTP was provided, Pol II and gp43 appeared to be similar in their preference for incorporation of dA opposite the THF regardless of the downstream nucleotide being A, T, C or G (Fig. 3C). This observation is in agreement with the established A rule, which is that in the absence of an instructive template base dA is most frequently incorporated (Strauss, 1985). Close inspection reveals that Pol II and gp43 are noticeably different in their second choice of a nucleotide for insertion opposite the lesion. For gp43, dG was most frequently incorporated after dA in all four sequence contexts (Fig. 3C). But for Pol II, the second choice appeared to vary with the base sequence downstream of the THF (Fig. 3C). We thus set out to determine a crystal structure of nucleotide insertion opposite a THF by Pol II and to uncover whether Pol II could loop out the THF and use the 5′ nucleotide to template nucleotide insertion as observed for the Y-family polymerases (Ling et al., 2004).

Crystal structure of Pol II in TLS nucleotide insertion

Crystals of a Pol II-DNA complex designed to capture the nucleotide insertion opposite a THF were grown, and the structure was determined and refined at 2.7Å resolution (Experimental Procedure). Surprisingly, not only was the AP lesion looped out, but two downstream nucleotides were also looped out. dT_{+3} and not dT_{+1} forms a base pair with the incoming dATP (Fig. 4A). The three looped-out nucleotides were accommodated in the same pocket as the downstream single-stranded template in the normal ternary complex structures (Fig. 4B). This pocket is formed between a β-hairpin loop in the Exo domain (aa 256–266), which is rich in aromatic and basic side chains, and the Exo-N connecting loop (aa 364–369). In the normal ternary complexes, the pocket holds two nucleotides downstream from the active site, and the first base is stacked with F260 and the second is sandwiched between F266 and R365.

This template-binding pocket is enlarged in the TLS nucleotide insertion structure (called $L_t(0,3)$ for looping out 3 nucleotides at the 0 position) by a ~5Å movement of the β-hairpin loop. The β-hairpin loop protects the two flipped-out bases from solvent (Fig. 4A) and appears to keep DNA in the polymerase active site. The $L_t(0, 3)$ structure may not be prevalent in solution in the presence of a long template strand since here the dT_{+3} was at the 5′ end of the unphosphorylated template strand and formed a reverse WC base-pair with the incoming dATP (Fig. 4A). Yet the ability of Pol II to stay in polymerization mode and use a downstream nucleotide as the templating base corroborates the observation of sequence-dependent nucleotide preferences for insertion opposite THF (Fig. 3C).

The β-hairpin loop and altered DNA partitioning

The corresponding β-hairpin loop also exists in gp43 but doesn't interact with the template strand during polymerization (Franklin et al., 2001; Freisinger et al., 2004; Hogg et al., 2004) (Fig. 4C). In the binary complex structure of gp43 with an AP DNA, the tip of the β-hairpin loop is placed between the primer and template strand and appears to direct the 3′ end of primer to the exonuclease active site for degradation (PDB: 2P5O, chain B and D) (Hogg et al., 2004). Nucleotides incorporated opposite an AP lesion by wildtype gp43 are unstable and removed by the proofreading activity. When this β-hairpin loop is truncated by 7 residues (gp43-β^{-}) and could no longer reach the DNA duplex, nucleotides are stably incorporated opposite an AP lesion, and the DNA substrate remains in the polymerase active site (Hogg et al., 2007).

Two structural changes in the N and Exo domains of Pol II may lead to its preference for TLS over primer degradation. Firstly, a deletion in the Exo-N connection in Pol II compared to gp43 removes an obstructive α-helix in the template-binding pocket and allows the template strand to enter and be stabilized for polymerization (Fig. 4C, S2). Secondly, Pol II, which is 121aa shorter than gp43 with deletions in four out of five domains, has a 20-residue insertion in the N domain that replaces a β-sheet in gp43 with a β-barrel (Fig. 4C, S2). This insertion in Pol II is correlated with the shift of the Exo domain by 4–5 Å and the β-hairpin loop by ~ 10Å. As a result, the tip of the β-hairpin loop in Pol II only reaches as far as that of the 7-residue truncated form of gp43 (PDB: 2DTU) (Fig. 4D), although the number of amino acid residues in the β-hairpin loop is comparable between the two proteins.

An altered DNA partitioning between the two active sites is apparent from the 7th crystal structure of Pol II, a protein-DNA binary complex with the AP lesion opposite G at the primer 3′ end (Table S1-2, Fig. 4D). As expected, the protein domains deviate from the apo and the ternary complex structures. Although the single-stranded template is not traceable and DNA is shifted, the primer end remains near the polymerase active site and far away from the exonuclease (Fig. 4D). This is in contrast to the gp43 - AP DNA binary complex structure (PDB: 2P5O), where the DNA primer end moved to the exonuclease active site in two of the four gp43 molecules.

We further compared the exonuclease activities of wildtype Pol II, gp43 and gp43-β$^-$on a mismatched template-primer pair (Fig. 4E) and their ability to degrade versus extend a THF-containing mismatched DNA substrate (Fig. 4F). Both experiments show that Pol II is a weaker exonuclease than gp43 and gp43-β$^-$. The presence of 50 μM dNTP is sufficient for Pol II to extend the lesion-containing DNA substrate and for gp43-β$^-$ to stop primer degradation beyond the lesion site, but it only slows down primer degradation by gp43 (Fig. 4F). Coupled with the structural results, these assays demonstrate that Pol II indeed favors polymerization over proofreading.

Multiple mechanisms for TLS primer extension by Pol II

Primer extension after an AP lesion by Pol II is robust (Fig. 3B). The products can be full-length but are often one, two or even three nucleotides shorter than the template strand. This may be due to looping out 1 or more template nucleotides. To determine the templating mechanism, 13 template strands, each with a THF and varying nucleotides downstream (N_0, N_{+1}, N_{+2} and N_{+3}), and 4 complementary primers with A, T, C or G at the 3′ end opposite the THF (Fig. 5A) were synthesized. Primer extension was assayed for each template-primer pair.

A representative set of template-dependent Pol II primer extension experiments is shown in Fig. 5B. Primer extension after the 3′-A opposite THF is predominantly directed by the nucleotide immediately 5′ to the THF. For example, dTTP was most efficiently incorporated opposite A_0, dGTP opposite C_0, and so on. Surprisingly, when T was at N_0, dTTP was favored second only to dATP for incorporation (Fig. 5B). Since dTTP was not incorporated when G, A or C is at N_0, we suspect that A_{+2} served as the template after looping out 2 nucleotides upstream (THF and T_{-2}) and the last two nucleotides (AA) of the primer strand formed base pairs with T_0 and T_{+1}.

Looping out of the template strand appeared not to be limited to 2 nucleotides. When the N_0 was G and N_{+2} was varied (Fig. 5C), although predominantly dCTP was incorporated opposite G_0 by direct extension, dGTP was also incorporated when N_{+2} was T. This could occur only if C_{+3} served as the template base when three nucleotides (THF, G_0 and T_{+1} or, T_{-2}, THF and G_0) were looped out at 1 or 2 bp upstream from the replicating base pair (Fig. 5C). To determine the location of 3 nt looping out, we paired the primer strand ending with

AT with template strands of varying nucleotides at N_{+1} (Fig. 5D). If looping out occurred 1 bp upstream, dGTP would be incorporated in all four sequence-contexts. But dGTP was incorporated only when N_{+1} was T. We conclude that 3 nucleotides were looped out 2-bp upstream from the replicating base pair.

To verify the 3-nucleotide looping-out mechanism, we co-varied the 3′ end of the primer strand and N_{+2} for base pair formation and showed that C at N_{+3} indeed instructed dGTP incorporation in all four pairing variations (Fig. 5E). In addition, when N_{+3} was varied, different dNTPs were incorporated accordingly (Fig. 5F). Therefore looping out 2 (Fig. 5B) or 3 nucleotides (Fig. 5C–D) both can occur 2-bp upstream (−2 position) from the templating base. They are designated as $L_t(-2,2)$ and $L_t(-2,3)$, respectively, according to the location and size of the bulge in the template strand.

Pol II can also loop out 1 nucleotide 1-bp upstream ($L_t(-1,1)$) (Fig. S3A). When C was at the primer 3′ end, it base paired with G_0 after looping out just THF and resulted in dATP incorporation opposite T_{+1} (Fig. 5E–F, boxed in cyan). The multiple mechanisms for TLS primer extension by Pol II differ from the X or Y family polymerases, which can loop out one nucleotide and at a preferred location only (Garcia-Diaz et al., 2006; Ling et al., 2004; Wilson and Pata, 2008). Remarkably, Pol II can use more than one mechanism on a single template-primer pair if the DNA local sequence permits (Fig. 5, S3B).

Crystal structure of $L_t(-2,2)$ by Pol II and its implications

A ternary complex of $L_t(-2,2)$ was crystallized and diffracted X-rays to 2.04 Å resolution (Table S2). This crystal had different unit cell dimensions from the other five Pol II-DNA ternary complex crystals (Table S2). The refined structure after molecular replacement, however, is rather similar to the normal ternary complexes (Fig. 6A). Two nucleotides including the THF are indeed looped out 2-bp upstream from the replicating base pair. Seven base pairs surrounding the looped-out nucleotides can be superimposed with those in the normal DNA ternary complexes. Significant changes are observed in the DNA duplex further upstream and in the thumb domain of Pol II (Fig. 6A). Of the two looped out nucleotides, the 5′ G interacts with A398, S399 and P400 of Pol II located on the extended linker between the N and palm domains (Fig. 6B–C). The 3′ THF has little interaction with Pol II. This observation may explain why there is little limitation on the size and sequence of the looped-out nucleotides.

Superposition of the $L_t(-2,2)$ and the gp43 ternary-complex structures reveals why gp43 is unable to loop out a damaged template strand and perform TLS primer extension. The length of the N-palm linker is conserved between gp43 and Pol II (Fig. S2). But the Cα trace of the linker in Pol II and gp43 differs significantly due to the insertion in the N domain of Pol II. The resulting β barrel in Pol II, which changes the β-hairpin loop position and the partition between polymerization and proofreading (Fig. 4), in effect shortens the distance between the N and palm domains and thereby relaxes the N-palm linker (Fig. 6B). In gp43 this linker has to travel a longer distance than that in Pol II and is juxtaposed to the template strand at the −2 and −3 positions. Together with the bulky side chains (P and Y replacing A398 and S399 of Pol II, respectively), it prevents the template strand from looping out (Fig. 6C). Other B-family replicases with known structures contain an α-helix in the N-palm linker, which is juxtaposed to the template strand and prevents nucleotides from looping out at the −2 to −4 positions (Fig. S2, S4).

Structural comparison between Pol II and gp43 in this region also reveals a cavity formed by A398 and S399 in Pol II next to the N_{-1} template nucleotide (Fig. 6D). This cavity can potentially accommodate a nucleotide looped out at the −1 position ($L_t(-1,1)$). The cavity may also relax the Pol II - DNA interaction and allow template-primer to adjust relative to

the dNTP in the active site for direct primer extension after a lesion. Substitution of S399 by Tyr in Pol II was thus engineered to study the size effect of this cavity on TLS.

Efficiency of TLS by Pol II

To determine the TLS efficiencies of Pol II, we measured K_M and k_{cat} of nucleotide insertion opposite an AP site (THF) and primer extension by various mechanisms (direct, $L_t(-1,1)$, $L_t(-2,1)$, $L_t(-2,2)$ and $L_t(-2,3)$) and compared them with the efficiencies of nucleotide incorporation on normal DNA. K_M and k_{cat} of S399Y mutant Pol II and gp43 were also measured (Table S2). All kinetic measurements of DNA polymerization were carried out with the Exo^- polymerases (D355N Pol II and D222A/D327A gp43) (Experimental Procedures). Since TLS is not limited to AP sites, K_M and k_{cat} for looping out normal nucleotides and primer extension of mismatched base pair were measured as well (Fig. 4F, S5). Templates and primers were designed to preclude multiple TLS mechanisms occurring in one reaction.

We first determined optimal pH and salt concentrations for the Pol II and gp43 polymerization assays (Fig. S6). Since K_M and k_{cat} may vary depending on the sequence context (Table S3), for each TLS substrate the K_M and k_{cat} of a corresponding normal DNA sequence were measured to normalize the polymerization efficiency ($k_{eff} = k_{cat} / K_M$).

In all, 29 sets of K_M and k_{cat} values were measured (Table S3) and are summarized in Figure 7A. With five normal template-primer pairs, k_{eff} of Pol II varies between 0.98 and 2.36 $s^{-1}\mu M^{-1}$. Pol II and gp43 are comparable during normal DNA synthesis and when incorporating dATP opposite a THF lesion (Table S3). But Pol II surpasses gp43 in mutagenic DNA synthesis and translesion primer extension (Fig. 7A). With Pol II, the efficiency of direct primer extension after THF is comparable to dA insertion opposite the lesion. Among the various looping-out mechanisms for primer extension, Pol II is most efficient in $L_t(-2,1)$, which is only ~40-fold less efficient than regular nucleotide incorporation. This is followed by $L_t(-2,2)$, $L_t(-1,1)$, and $L_t(-2,3)$ with additional 40, 120, and 400 fold reductions in catalytic efficiency (Table S3). Pol II is 20 times more efficient in misincorporation than gp43, and it is 6-fold more efficient in extending a mismatched primer than misincorporation. In contrast, gp43 cannot extend after THF or a mismatch.

As suspected, nucleotide looping out at −1 and −2 positions is not limited to damaged template. Normal bases in the template strand can also be looped out (Fig. S5) albeit at a lower efficiency than for THF (Fig. 7A, Table S3). Substitution by S399Y, which reduces the cavity size at the −1 position (Fig. 6D), clearly hampers but doesn't abolish looping out whether with an AP or normal nucleotide (Table S3). This may be because the peptide backbone of Pol II is not changed and still allows the cavity to form. Interestingly, the S399Y substitution is most detrimental to the direct primer extension after THF and reduces the efficiency by 6 fold.

Discussion

Tuning a B-family polymerase for TLS and mutagenic synthesis

Our structural and kinetic analyses reveal that an insertion in the N domain of Pol II alters enzyme-substrate interactions from a distance (Fig. 1D) and ultimately results in reduced fidelity and TLS. Firstly, the inserted β-barrel loosens the N-palm linker (Fig. 6B), which leaves small cavities one and two base pairs upstream from the active site for looping out DNA lesions in the template. The extra space surrounding the −1 base pair may also allow the template-primer DNA to adjust, thus facilitating direct primer extension after a damaged or mismatched base pair. Template misalignment has long been proposed for mutagenesis and TLS by replicases (Streisinger et al., 1966), but without the special cavities replicases

are not as efficient as Pol II in looping out the template strand (Fig. 7). Secondly, the inserted β-barrel shifts the β-hairpin loop in Pol II and alters substrate partitioning between polymerization and proofreading (Fig. 4). Although Pol II is able to proofread, an increase of DNA residence time in the polymerase active site in the presence of a lesion can give Pol II the opportunity to sample downstream template and carry out TLS extension (Fig. 4F). These two features may also cause misincorporation and efficient extension of mismatched base pairs by Pol II when cells are under stresses (Foster, 2005; McKenzie and Rosenberg, 2001). Such tuning of catalytic specificity at a long distance by nature is in contrast to manmade mutations, which rarely deviate from the site of catalysis (Bebenek et al., 2001; Jarosz et al., 2006).

Function of Pol II *in vivo*

The unusual "acrobatic" looping out of template (Fig. S7) and multiple mechanisms for primer extension by Pol II after a damaged or mismatched base have interesting implications. Pol II is not efficient in nucleotide insertion opposite a lesion, but it may be the second polymerase in mutagenic and translesion synthesis to extend primers after nucleotide insertion by a replicase or specialized Y-family polymerase (Curti et al., 2009; Neeley et al., 2007; Wagner et al., 2002). By looping out DNA lesions, Pol II has no particular preference for a lesion type and thus can have a broad substrate range. Given the many primer extension mechanisms (Fig. S7), the results of TLS by Pol II are complex and have no fixed mutational pattern. The absence of a TLS signature may contribute to the difficulty of identifying in vivo functions of Pol II. A hot spot for 2-nucleotide deletion after AAF modification was found only in the presence of Pol II at a NarI restriction site (GGCGCC), which became GGCC (Fuchs and Fujii, 2007). The proposed looping-out model fully agrees with the $L_t(-2,2)$ structure. Our crystal structures also indicate that the two looped-out nucleotides can remain extra-helical as Pol II continues downstream DNA synthesis. If the sequence is not as GC rich as in the NarI site, it is possible that template and primer realign thus leaving no trace of frameshift.

Pol ζ may share similar structural features with Pol II in TLS

Eukaryotic Pol ζ has been shown to play a critical role in TLS and spontaneous and induced mutagenesis (reviewed recently in (Gan et al., 2008). In addition to a TLS deficiency, reduced Pol ζ expression causes large-scale genomic instability in vertebrate cells (Schenten et al., 2009; Sonoda et al., 2003). Pol ζ is a heterodimer of REV3 and REV7, and REV3 alone contains 1500-3300 aa with regulatory domains in addition to the Pol II-like polymerase domain (Lawrence and Maher, 2001). Partly due to its large size and low yield, Pol ζ has resisted structural characterization, and its mechanism for TLS and mutagenesis is unknown.

Based on sequence alignment, REV3 shares two structural features of Pol II that underlie the mutagenic and TLS capability. Firstly, due to the loss of the catalytically essential carboxylates in the Exo domain (Fig. S2), Pol ζ has no proofreading activity (Lawrence and Maher, 2001). This surpasses the altered DNA partitioning of Pol II and implicitly leads to an increased mutation rate and TLS activity (Reha-Krantz, 2009; Schaaper, 1993). Secondly, Rev3 is predicted to have a flexible N-palm linker devoid of the α-helix (toolkit.tuebingen.mpg.de/hhpred) (Fig. S2, S4). This unstructured linker is even longer than that of Pol II and may produce unique cavities to accommodate looped-out template strand. The potential of misalignment corroborates the sequence-dependent complex mutagenic spectra of pol ζ (Abdulovic et al., 2008; Zhong et al., 2006).

Evolutionary implications

To cope with lesions, nature appears to employ diversely modified DNA polymerases with changes either in the active site as observed with the Y-family polymerases or distant from the active site as observed with Pol II. Y-family polymereases are more efficient in TLS than Pol II (Kroeger et al., 2004; Neeley et al., 2007; Wagner et al., 2002), but on normal DNA Pol II behaves like a replicative polymerase and is more efficient and more accurate than Y-family polymerases. Evolutionarily one can imagine that sudden and drastic changes in the active site of a replicative DNA polymerase could be catastrophic. In accordance the mutagenic Y-family polymerases are modulated by their inefficiency in DNA synthesis (Beard et al., 2002). Gradual changes outside of the active site, however, can be selected when they confer a growth advantage under stress and may give rise to the A- and B-family TLS polymerases in species ranging from *E. coli* to humans. As exemplified by Pol II, tuning catalytic specificity at a distance from the active site and thereby gaining new functions without greatly altering original function may be a general mechanism in evolution.

Experimental Procedures

Protein Expression and Purification

E. coli DNA polymerase II (Pol II) was PCRed (Zeng, 1998) from genomic DNA (ATCC: 10798D) and cloned into PET 28a(+) (pWY2193, with an N-terminal His-tag cleavable by PreScission protease). Pol II exo$^-$ mutants D156N (pWY2194), D229N (pWY2195) and D335N (pWY2196), and D335N/S399Y (PWY2197) were made by QuikChange mutagenesis (Stratagene). Plasmid pCW50 encoding gp43 exo$^-$(D222A/D327A) was a gift of Dr. W. Konigsberg. WT gp43 protein and the plasmid encoding gp43-β$^-$ (I253G and Δ254–260) were gifts of Dr. S. Doublié. All proteins were expressed in BL21(DE3) cells (Novagen) by induction with 0.2–1mM IPTG at OD 0.6~0.8 and further incubation at 16°C for 20 hours (gp43) or 37°C for 3 hours (Pol II). Cells were harvested and lysed by sonication. The proteins were purified to homogeneity after fractionation by a His-trap column, PreScission protease cleavage to remove the His-tag from Pol II, and Mono Q chromatography. Concentrated Pol II in 20 mM Tris-HCl (pH7.5), 150 mM NaCl, 5% glycerol and 2 mM β-mercaptoethanol was stored at −80 °C after flash cooling in liquid Nitrogen. For kinetic assays, Pol II and gp43 were stored in 20 mM Tris-HCl (pH 7.5) 25% glycerol and 1 M NaCl at −20°C.

Oligonucleotides

DNA oligonucleotides (Table S1, S3, and Fig. 5A) were purchased from the Facility for Biotechnology Resources (FBR) or from Integrated DNA Technologies (IDT) and HPLC purified. Template and primer pairs at a 1:1 molar ratio (or otherwise specified) were annealed in pH 8.0 TE buffer by heating for 5 min at 85°C and slow cooling to 25 °C.

Crystallization

Pol II-DNA-dNTP ternary complexes were prepared by mixing ~ 6mg/ml *E.coli* pol II D335N protein, DNA (at a 1.5:1 molar ratio to protein), and 1 mM dNTP in 13.33 mM Tris-HCl (pH 7.5), 5mM $MgCl_2$, 100 mM NaCl, 3.33% glycerol and 1.33 mM β-mercaptoethanol and incubating at the room temperature for 10 min (Table S1). For the L_t(−2,2) complex, dNTP was replaced by ddNTP, and $MgCl_2$ by $CaCl_2$. Crystals were grown at room temperature by the hanging-drop vapor diffusion method after 1:1 mixing of protein complexes and well solutions. Apo-protein and binary complex crystals were grown using similar conditions. After streak seeding, diffraction-quality crystals were obtained in a week.

Structure Determination

All X-ray diffraction data were collected in-house (Rigaku) and processed using HKL2000 (Otwinowski and Minor, 1997). The apo and first ternary complex structure (T-dA) were solved using the PDB entry 1Q8I as a search model by molecular replacement (MOLREP) (CCP4, 1994) (Table S1). The N and palm domains were first located, and then the Exo, thumb and finger domains. The model of protein, DNA and dNTP were completed using Arp/Warp (CCP4, 1994), O and COOT (Emsley and Cowtan, 2004; Jones et al., 1991). Structures were refined using CNS (Brünger et al., 1998). The same diffraction test set was used in R_{free} calculation for the structures of same crystal lattices. The final refinement statistics are summarized in Table S2. Structures were superimposed using CCP4MG (CCP4, 1994).

pH, salt, and $MgCl_2$ concentration optimization

The pH optima were determined using the pH Kit (pH 6.2~9.8) from Hampton Research (Cat No: HR2-241). Salt concentrations of NaCl and KCl were optimized at the optimal pH (8.0). At the optimal NaCl (100 mM) and pH, $MgCl_2$ concentrations (0 – 20 mM) were optimized. Finally, 50 mM Tris–HCl (pH 8.0), 100 mM NaCl, 5 mM $MgCl_2$, 0.1 mg/ml BSA, 1 mM DTT and 5% glycerol (buffer R) were used for all DNA synthesis assays.

Nucleotide incorporation and primer extension assays

Single nucleotide incorporation assays were conducted in 10 μl buffer R with 10 μM DNA and 0.15 μM protein. For primer extension assays, the 5′ 6-carboxyflorescein (6-FAM) labeled primer (purchased from IDT or Syngen) and template DNA were annealed at a 1:1.1 molar ratio. Reactions were initiated by addition of 1 mM (single or four) dNTP-$MgCl_2$, and terminated after 15–60 min incubation at room temperature by addition of 90% formamide in 1X TBE buffer. Samples were incubated at 95°C for 3 min and resolved on a 20% TBE-urea gel. Gels were stained by SYBR-green for 20 min and analyzed by Fluochem imaging system (Alpha Innotech) or directly scanned by a Typhoon Trio Scanner (GE Healthcare). The 6-FAM-labeled products were quantified based on fluorescence intensity using ImageQuant software.

Partition assays

Exonuclease activities of Pol II, gp43-β^- and gp43 (25–200 nM) were assayed using a mismatched DNA (10 μM) (Table S3) in buffer R without dNTPs. Partition between the two active sites were assayed using 150 nM of each protein, 10 μM L_t(μ−2, 1) DNA (Table S3), and 0–5 mM dNTPs. Reactions were carried out at 22°C for 5 min.

K_M and k_{cat} measurements

Time courses were examined first with 1–2 mM dNTP and various enzyme concentrations to insure a steady reaction rate and less than 10% substrate-product conversion. Each 10 μl reaction contained 10 μM unlabeled DNA with 10 nM corresponding 6-FAM-labeled DNA (0.1%), 1–20 nM enzyme and a concentration range of dNTP. The enzyme was freshly diluted from a stock solution, mixed with DNA and pre-incubated for 5 min at room temperature. Reactions were initiated and terminated as described above. After electrophoresis on 20% TBE-urea gels, products were quantified. k_{cat} and K_M were calculated using Graphpad Prism 5.

Accession numbers

The structures reported here have been deposited to Protein Data Bank with accession codes 3K57 (T-dA), 3K58 (A-dT), 3K59 (G-dC), 3K5A (C-dG), 3K5L (L_t(0,3)), 3K5M (L_t(−2,2)), 3K5N (AP-binary complex) and 3K5O (Apo).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. S. Doublié for purified gp43 protein and the gp43-β^{-} clone, Dr. W. Konigsberg for the gp43 exo^{-} clone, our lab mates for technical support, Drs. D. Leahy and R. Craigie for critical reading of the manuscript, and Dr. F. Dyda for maintaining the X-ray equipment. The research is funded by the NIDDK intramural research program at National Institutes of Health.

References

Abdulovic AL, Minesinger BK, Jinks-Robertson S. The effect of sequence context on spontaneous Polzeta-dependent mutagenesis in Saccharomyces cerevisiae. Nucleic Acids Res. 2008; 36:2082–2093. [PubMed: 18276637]

Al Mamun AA. Elevated expression of DNA polymerase II increases spontaneous mutagenesis in Escherichia coli. Mutat Res. 2007; 625:29–39. [PubMed: 17586534]

Al Mamun AA, Humayun MZ. Escherichia coli DNA polymerase II can efficiently bypass 3,N(4)-ethenocytosine lesions in vitro and in vivo. Mutat Res. 2006; 593:164–176. [PubMed: 16171831]

Andersen PL, Xu F, Xiao W. Eukaryotic DNA damage tolerance and translesion synthesis through covalent modifications of PCNA. Cell Res. 2008; 18:162–173. [PubMed: 18157158]

Anderson WF, Prince DB, Yu H, McEntee K, Goodman MF. Crystallization of DNA polymerase II from Escherichia coli. J Mol Biol. 1994; 238:120–122. [PubMed: 8145251]

Arana ME, Takata K, Garcia-Diaz M, Wood RD, Kunkel TA. A unique error signature for human DNA polymerase nu. DNA Repair (Amst). 2007; 6:213–223. [PubMed: 17118716]

Beard WA, Shock DD, Vande Berg BJ, Wilson SH. Efficiency of correct nucleotide insertion governs DNA polymerase fidelity. J Biol Chem. 2002; 277:47393–47398. [PubMed: 12370169]

Bebenek A, Dressman HK, Carver GT, Ng S, Petrov V, Yang G, Konigsberg WH, Karam JD, Drake JW. Interacting fidelity defects in the replicative DNA polymerase of bacteriophage RB69. J Biol Chem. 2001; 276:10387–10397. [PubMed: 11133987]

Bebenek K, Kunkel TA. Functions of DNA polymerases. Adv Protein Chem. 2004; 69:137–165. [PubMed: 15588842]

Becherel OJ, Fuchs RP. Mechanism of DNA polymerase II-mediated frameshift mutagenesis. Proc Natl Acad Sci U S A. 2001; 98:8566–8571. [PubMed: 11447256]

Berman AJ, Kamtekar S, Goodman JL, Lazaro JM, de Vega M, Blanco L, Salas M, Steitz TA. Structures of phi29 DNA polymerase complexed with substrate: the mechanism of translocation in B-family polymerases. Embo J. 2007; 26:3494–3505. [PubMed: 17611604]

Bonner CA, Hays S, McEntee K, Goodman MF. DNA polymerase II is encoded by the DNA damage-inducible dinA gene of Escherichia coli. Proc Natl Acad Sci U S A. 1990; 87:7663–7667. [PubMed: 2217198]

Brünger AT, Adams PD, Clore GM, Delane WL, Gros P, Grosse-Kunstleve RW, Jiang J-S, Kuszewski J, Nilges M, Pannu NS, et al. Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. 1998; D54:905–921.

Cai H, Yu H, McEntee K, Kunkel TA, Goodman MF. Purification and properties of wild-type and exonuclease-deficient DNA polymerase II from Escherichia coli. J Biol Chem. 1995; 270:15327–15335. [PubMed: 7797520]

CCP4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D. 1994; 50:760–763. [PubMed: 15299374]

Chang DJ, Cimprich KA. DNA damage tolerance: when it's OK to make mistakes. Nat Chem Biol. 2009; 5:82–90. [PubMed: 19148176]

Curti E, McDonald JP, Mead S, Woodgate R. DNA polymerase switching: effects on spontaneous mutagenesis in Escherichia coli. Mol Microbiol. 2009; 71:315–331. [PubMed: 19019142]

Doublie S, Sawaya MR, Ellenberger T. An open and closed case for all polymerases. Structure Fold Des. 1999; 7:R31–R35. [PubMed: 10368292]

Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004; 60:2126–2132. [PubMed: 15572765]

Foster PL. Stress responses and genetic variation in bacteria. Mutat Res. 2005; 569:3–11. [PubMed: 15603749]

Franklin MC, Wang J, Steitz TA. Structure of the replicating complex of a pol a family DNA polymerase. Cell. 2001; 105:657–667. [PubMed: 11389835]

Freisinger E, Grollman AP, Miller H, Kisker C. Lesion (in)tolerance reveals insights into DNA replication fidelity. Embo J. 2004; 23:1494–1505. [PubMed: 15057282]

Fuchs RP, Fujii S. Translesion synthesis in Escherichia coli: lessons from the NarI mutation hot spot. DNA Repair (Amst). 2007; 6:1032–1041. [PubMed: 17403618]

Gan GN, Wittschieben JP, Wittschieben BO, Wood RD. DNA polymerase zeta (pol zeta) in higher eukaryotes. Cell Res. 2008; 18:174–183. [PubMed: 18157155]

Garcia-Diaz M, Bebenek K, Krahn JM, Pedersen LC, Kunkel TA. Structural analysis of strand misalignment during DNA synthesis by a human DNA polymerase. Cell. 2006; 124:331–342. [PubMed: 16439207]

Hogg M, Aller P, Konigsberg W, Wallace SS, Doublie S. Structural and biochemical investigation of the role in proofreading of a beta hairpin loop found in the exonuclease domain of a replicative DNA polymerase of the B family. J Biol Chem. 2007; 282:1432–1444. [PubMed: 17098747]

Hogg M, Wallace SS, Doublie S. Crystallographic snapshots of a replicative DNA polymerase encountering an abasic site. Embo J. 2004; 23:1483–1493. [PubMed: 15057283]

Iwasaki H, Ishino Y, Toh H, Nakata A, Shinagawa H. Escherichia coli DNA polymerase II is homologous to alpha-like DNA polymerases. Mol Gen Genet. 1991; 226:24–33. [PubMed: 2034216]

Jarosz DF, Godoy VG, Delaney JC, Essigmann JM, Walker GC. A single amino acid governs enhanced activity of DinB DNA polymerases on damaged templates. Nature. 2006; 439:225–228. [PubMed: 16407906]

Jones TA, Zou J-Y, Cowan SW. Improved methods for building models in electron density maps and the location of errors in these models. Acta Crystallogr A. 1991; 47:110–119. [PubMed: 2025413]

Joyce CM, Benkovic SJ. DNA polymerase fidelity: kinetics, structure, and checkpoints. Biochemistry. 2004; 43:14317–14324. [PubMed: 15533035]

Kroeger KM, Jiang YL, Kow YW, Goodman MF, Greenberg MM. Mutagenic effects of 2-deoxyribonolactone in Escherichia coli. An abasic lesion that disobeys the A-rule. Biochemistry. 2004; 43:6723–6733. [PubMed: 15157106]

Lawrence CW, Maher VM. Mutagenesis in eukaryotes dependent on DNA polymerase zeta and Rev1p. Philos Trans R Soc Lond B Biol Sci. 2001; 356:41–46. [PubMed: 11205328]

Lee HR, Wang M, Konigsberg W. The Reopening Rate of the Fingers Domain Is a Determinant of Base Selectivity for RB69 DNA Polymerase (dagger). Biochemistry. 2009; 48:2087–2098. [PubMed: 19228036]

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. [PubMed: 15023344]

McKenzie GJ, Rosenberg SM. Adaptive mutations, mutator DNA polymerases and genetic change strategies of pathogens. Curr Opin Microbiol. 2001; 4:586–594. [PubMed: 11587937]

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. [PubMed: 17631059]

Napolitano R, Janel-Bintz R, Wagner J, Fuchs RP. All three SOSinducible DNA polymerases (Pol II, Pol IV and Pol V) are involved in induced mutagenesis. Embo J. 2000; 19:6259–6265. [PubMed: 11080171]

Neeley WL, Delaney S, Alekseyev YO, Jarosz DF, Delaney JC, Walker GC, Essigmann JM. DNA polymerase V allows bypass of toxic guanine oxidation products in vivo. J Biol Chem. 2007; 282:12741–12748. [PubMed: 17322566]

Ohmori H, Friedberg EC, Fuchs RPP, Goodman MF, Hanaoka F, Hinkle D, Kunkel TA, Lawrence CW, Livneh Z, Nohmi T, et al. The Y-family of DNA polymerases. Mol Cell. 2001; 8:7–8. [PubMed: 11515498]

Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 1997; 276:307–326.

Paz-Elizur T, Takeshita M, Goodman M, O'Donnell M, Livneh Z. Mechanism of translesion DNA synthesis by DNA polymerase II. Comparison to DNA polymerases I and III core. J Biol Chem. 1996; 271:24662–24669. [PubMed: 8798733]

Reha-Krantz LJ. DNA polymerase proofreading: Multiple roles maintain genome stability. Biochim Biophys Acta. 2009

Schaaper RM. Base selection, proofreading, and mismatch repair during DNA replication in *Escherichia coli* . J Biol Chem. 1993; 268:23762–23765. [PubMed: 8226906]

Schenten D, Kracker S, Esposito G, Franco S, Klein U, Murphy M, Alt FW, Rajewsky K. Pol zeta ablation in B cells impairs the germinal center reaction, class switch recombination, DNA break repair, and genome stability. J Exp Med. 2009; 206:477–490. [PubMed: 19204108]

Sharief FS, Vojta PJ, Ropp PA, Copeland WC. Cloning and chromosomal mapping of the human DNA polymerase theta (POLQ), the eighth human DNA polymerase. Genomics. 1999; 59:90–96. [PubMed: 10395804]

Sonoda E, Okada T, Zhao GY, Tateishi S, Araki K, Yamaizumi M, Yagi T, Verkaik NS, van Gent DC, Takata M, Takeda S. Multiple roles of Rev3, the catalytic subunit of polzeta in maintaining genome stability in vertebrates. Embo J. 2003; 22:3188–3197. [PubMed: 12805232]

Steitz TA. DNA polymerases: structural diversity and common mechanisms. J Biol Chem. 1999; 274:17395–17398. [PubMed: 10364165]

Strauss BS. Translesion DNA synthesis: polymerase response to altered nucleotides. Cancer Surv. 1985; 4:493–516. [PubMed: 2825983]

Streisinger G, Okada Y, Emrich J, Newton J, Tsugita A, Terzaghi E, Inouye M. Frameshift mutations and the genetic code. This paper is dedicated to Professor Theodosius Dobzhansky on the occasion of his 66th birthday. Cold Spring Harb Symp Quant Biol. 1966; 31:77–84. [PubMed: 5237214]

Wagner J, Etienne H, Janel-Bintz R, Fuchs RP. Genetics of mutagenesis in Ecoli: various combinations of translesion polymerases (Pol II, IV and V) deal with lesion/sequence context diversity. DNA Repair (Amst). 2002; 1:159–167. [PubMed: 12509262]

Wilson RC, Pata JD. Structural insights into the generation of singlebase deletions by the Y family DNA polymerase dbh. Mol Cell. 2008; 29:767–779. [PubMed: 18374650]

Yang W, Woodgate R. What a difference a decade makes: insights into translesion DNA synthesis. Proc Natl Acad Sci U S A. 2007; 104:15591–15598. [PubMed: 17898175]

Yeiser B, Pepper ED, Goodman MF, Finkel SE. SOS-induced DNA polymerases enhance long-term survival and evolutionary fitness. Proc Natl Acad Sci U S A. 2002; 99:8737–8741. [PubMed: 12060704]

Zeng G. Sticky-end PCR: new method for subcloning. Biotechniques. 1998; 25:206–208. [PubMed: 9714877]

Zhong X, Garg P, Stith CM, Nick McElhinny SA, Kissling GE, Burgers PM, Kunkel TA. The fidelity of DNA synthesis by yeast DNA polymerase zeta alone and with accessory proteins. Nucleic Acids Res. 2006; 34:4731–4742. [PubMed: 16971464]

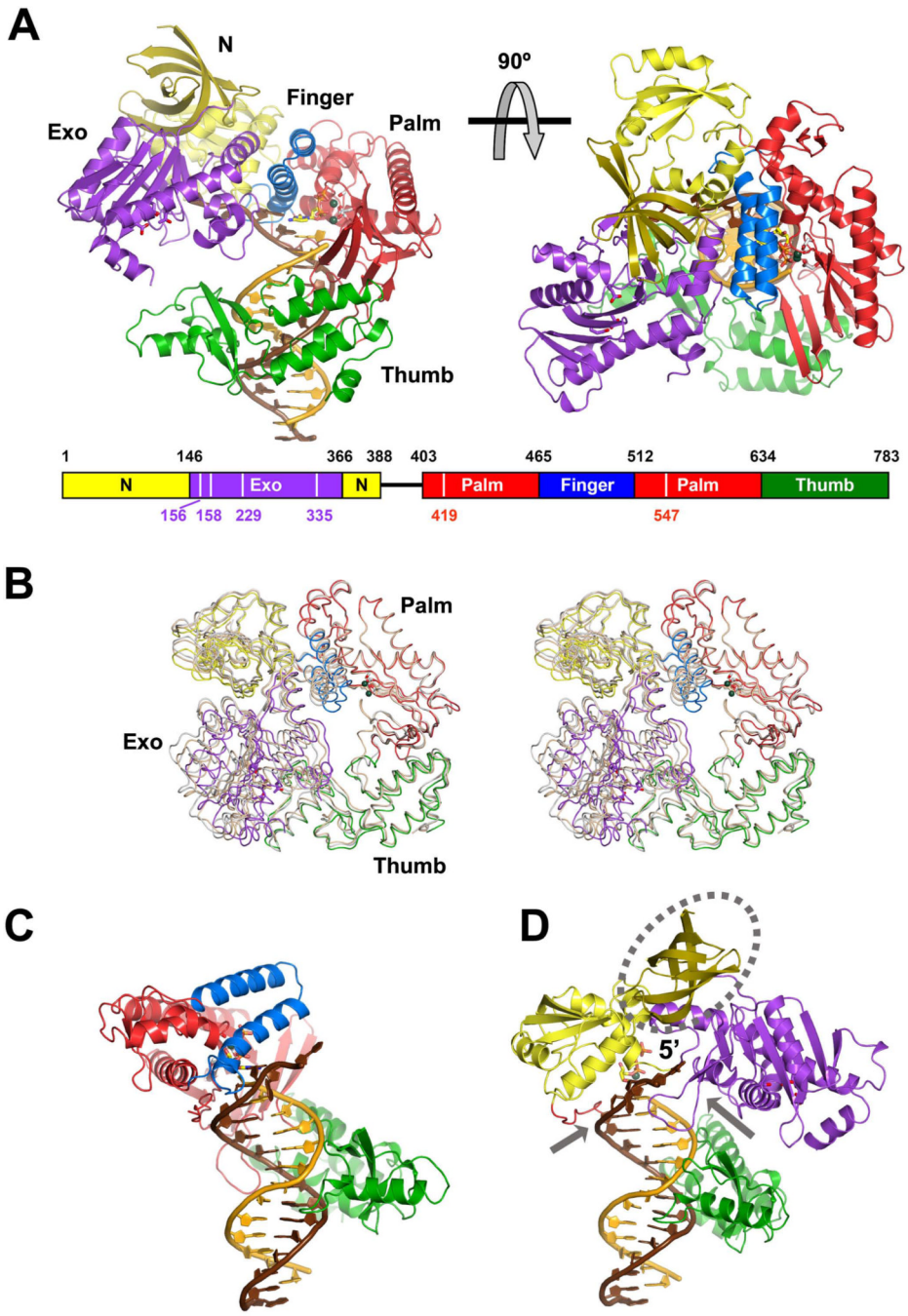


Figure 1.

Crystal structures of Pol II and its complexes with DNA and an incoming dNTP. **A.** Orthogonal views of the Pol II-DNA-dCTP complex. The five domains of Pol II are color-coded as in the one-dimensional (1D) diagramed below. The catalytic residues of the polymerase and exonuclease are indicated in the 1D diagram and shown as grey (Pol) and purple sticks (Exo). The template strand is shown in brown, primer in orange, and the dCTP in yellow. The green spheres represent the two metal ions. **B.** Superimposition of two Pol II Cα traces in the apo structure (silver and copper) and one in the ternary complex (color coded as in A). **C.** Catalytic core of Pol II (thumb, palm and finger) tracks the minor groove

of the DNA substrate. **D.** The downstream template is sandwiched by the N and Exo domains. The β-barrel in the N-domain is encircled in the grey oval, and the β-hairpin and N-palm linker are indicated by the arrowheads.

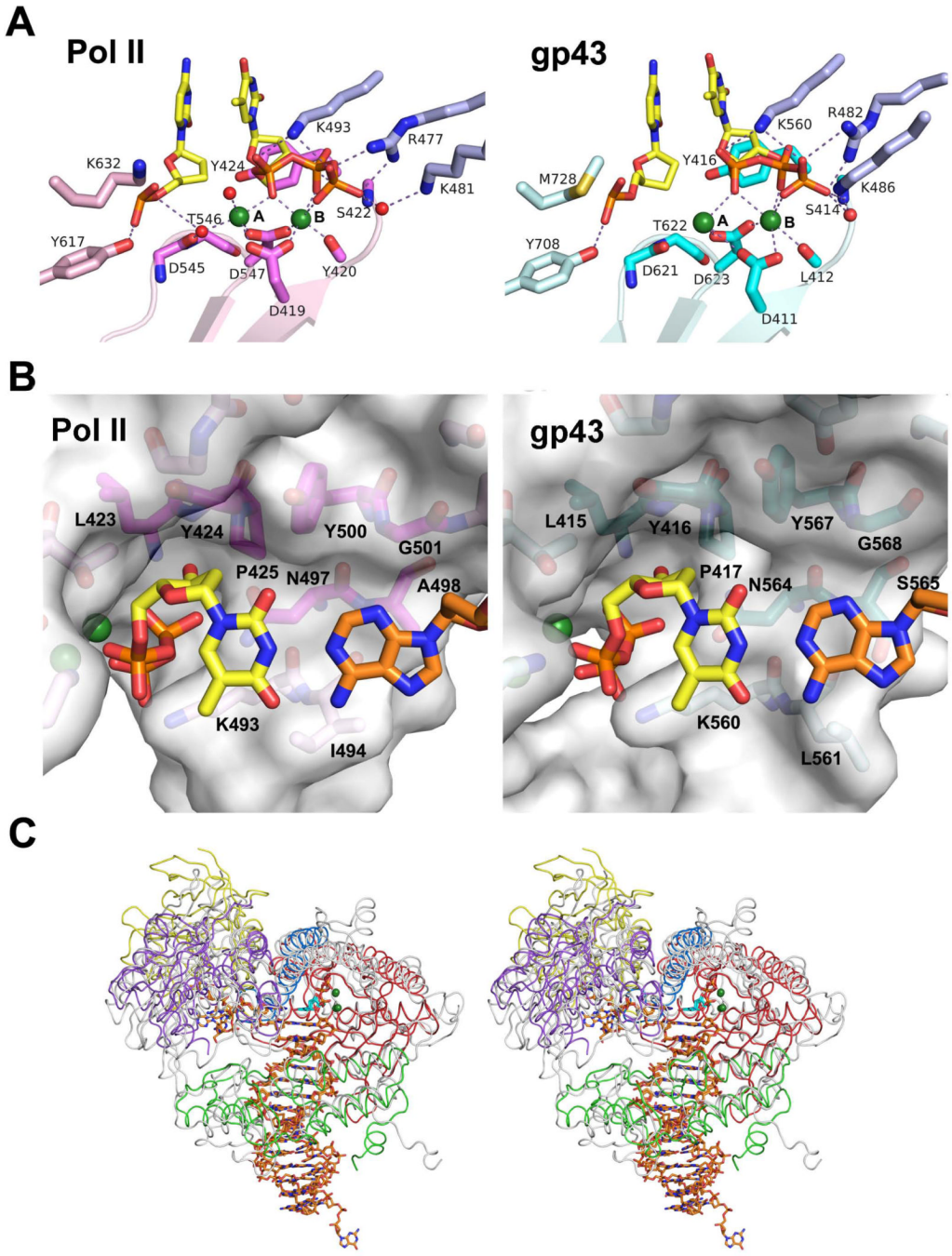


Figure 2.

Comparison of Pol II and gp43. **A.** The polymerase active site. The metal ions are shown as green spheres, the primer end and dNTP as yellow and orange sticks, and basic residues interacting with the triphosphate are shown in blue. The carboxylates and conserved residues around the metal ions are colored in pink (Pol II) or cyan (gp43). **B.** Minor groove interactions. The replicating base pair is shown in orange (template) and yellow (dTTP), and protein residues are shown as pink or cyan sticks enveloped by semi-transparent molecular surface. **C.** Superposition of Pol II (multi-colors) and gp43 (silver) ternary complex

structures in stereo. Although only palm domains are superposed, the DNA substrates (both in orange/blue/red) and dNTP (cyan) are nearly identical.

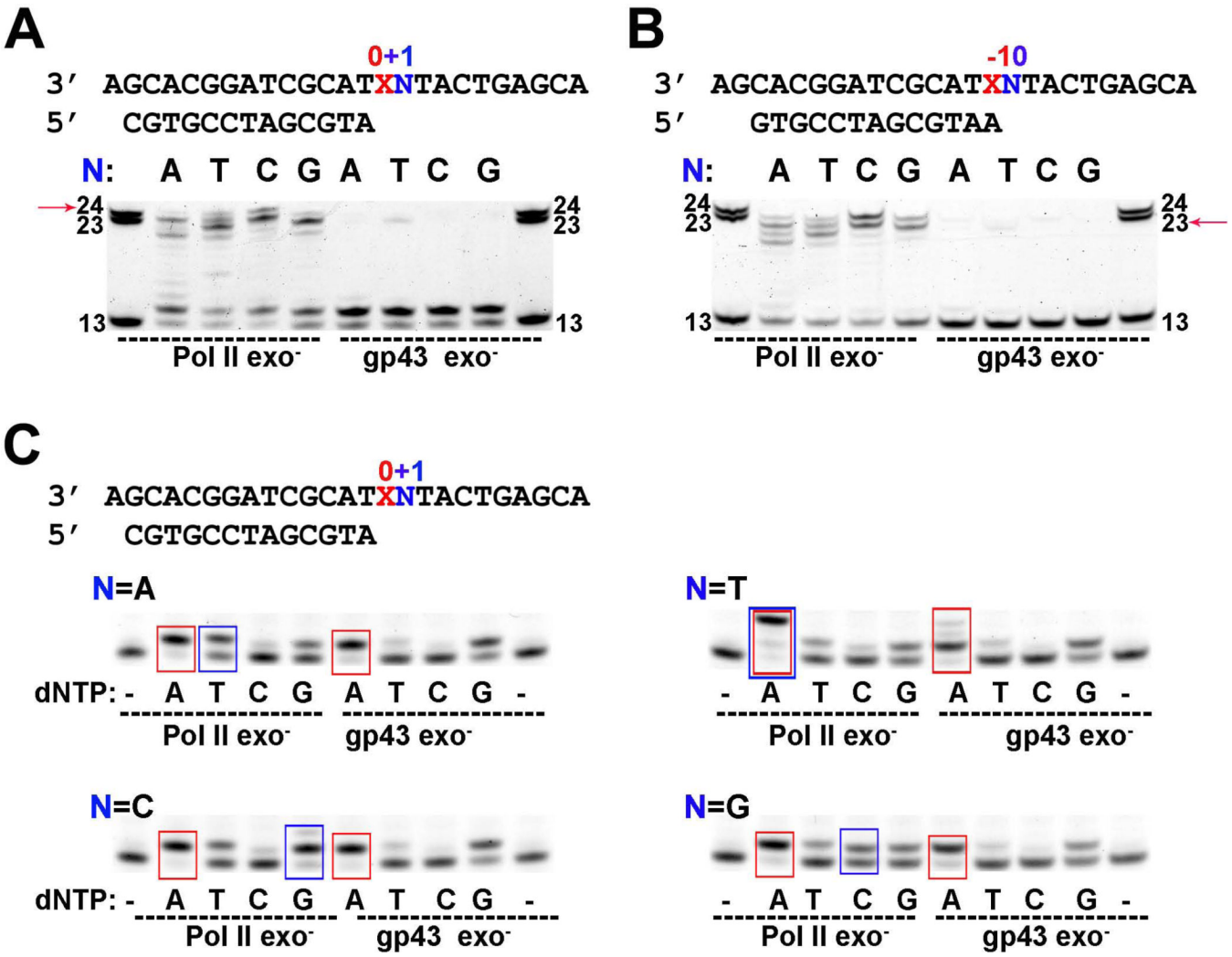


Figure 3.

Lesion bypass by exo⁻ Pol II and gp43. In DNA sequences, X represents the AP analog THF, and the nucleotide 5′ to the lesion was varied as indicated by N. **A.** Nucleotide insertion opposite the lesion. **B.** Primer extension. All four dNTP were present. The red arrows indicate the expected full-length product, 24 nt in A and 23 nt in B. **C.** Nucleotide selection in the insertion step. When only one dNTP was provided, incorporation of dA is most efficient (boxed in red). The second choice for incorporation by Pol II varies with the template sequences (boxed in blue).

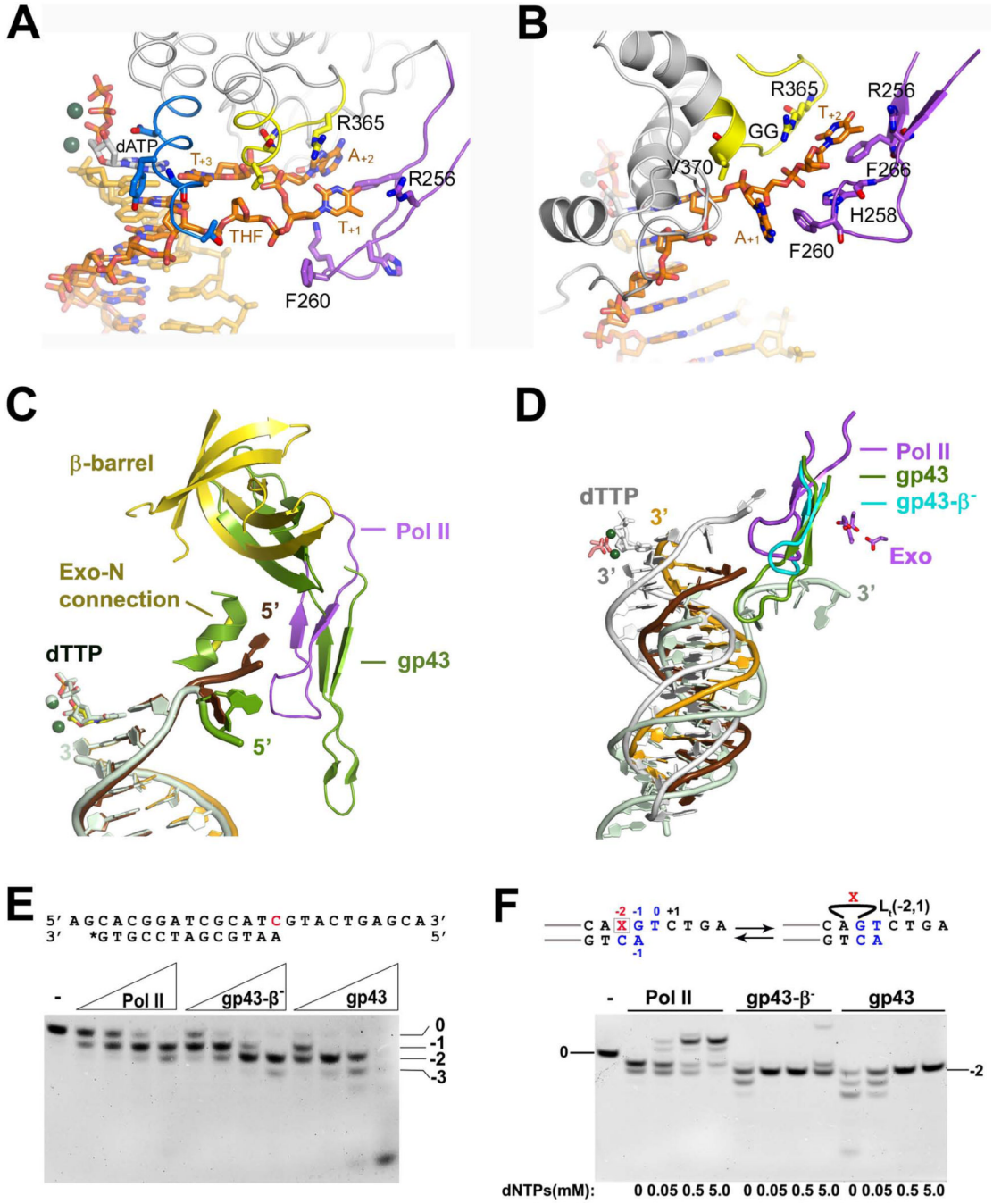


Figure 4.

The Pol II structure of nucleotide insertion opposite an AP lesion. **A.** A zoom-in view of the $L_t(0,3)$ structure. The looped-out THF, T_{+1} and A_{+2} are inserted into the template-binding pocket. T_{+3} forms reverse WC pair with the incoming dATP. **B.** The template-binding pocket in the normal Pol II ternary complex. It is formed between the Exo-N connection (yellow) and the β-hairpin (magenta). **C.** Comparison of the β hairpin loop in Pol II and gp43. The β barrel (gold) in Pol II causes retraction of its β-hairpin loop (magenta), which appears as if shorter than the β-hairpin in gp43 (green). The downstream template in Pol II (brown) and gp43 (green) also differ. **D.** The primer (orange) and template (brown) stay

around the polymerase active site in the Pol II-AP DNA binary complex. The DNA in the Pol II ternary complex structure (silver), and the DNA being proofread by gp43 (PDB: 2P5O) (pale green) are superimposed for comparison. The β-hairpin loop Pol II, gp43) and gp43-β^- are shown in different colors. The active site of Exo is indicated by the magenta catalytic residues, and that of Pol by the green metal ions. **E.** Exonuclease activities of wildtype Pol II, gp43 and gp43-β^- . The enzyme concentrations are 25, 50, 100 and 200 nM, and DNA 10 μM. The first lane (-) is DNA alone. **F**. Partition assay. Primer extension and degradation of L_t(−2, 1) DNA (10 μM) by the three polymerases (150 nM) in the presence of 0–5 mM dNTP. Pol II has the least Exo activity and is the only one that can extend the primer.

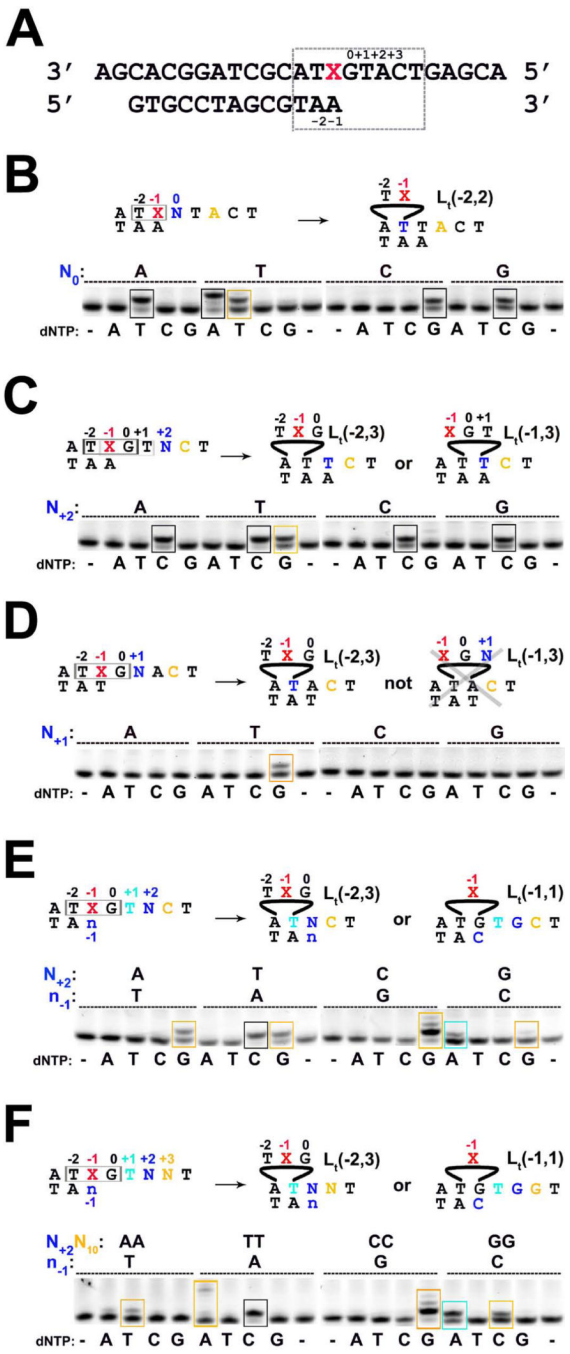


Figure 5.

Multiple mechanisms for primer extension by Pol II after an AP lesion. **A.** A template of the DNA substrates. Nucleotides at 0, +1, +2, and +3 of the template and −1 of the primer are varied. Sequences in the grey box are specified in the following panels. **B.** Direct primer extension. Nucleotide immediately 5′ to the THF instructs dNTP incorporation as highlighted in grey boxes. The exception resulting from $L_t(-2,2)$ is boxed in orange. **C.** Looping out 3 nucleotides. Results of direct primer extension are boxed in grey, and that of looping out 3 nt in orange. **D.** Determining the position of 3-nt looping-out. It occurs only at 2 bp upstream from the replicating base pair. **E** and **F**. Confirmation of the $L_t(-2,3)$

mechanism. Results of direct extension are boxed in grey, of $L_t(-2,3)$ in orange, and $L(-1,1)$ in cyan.

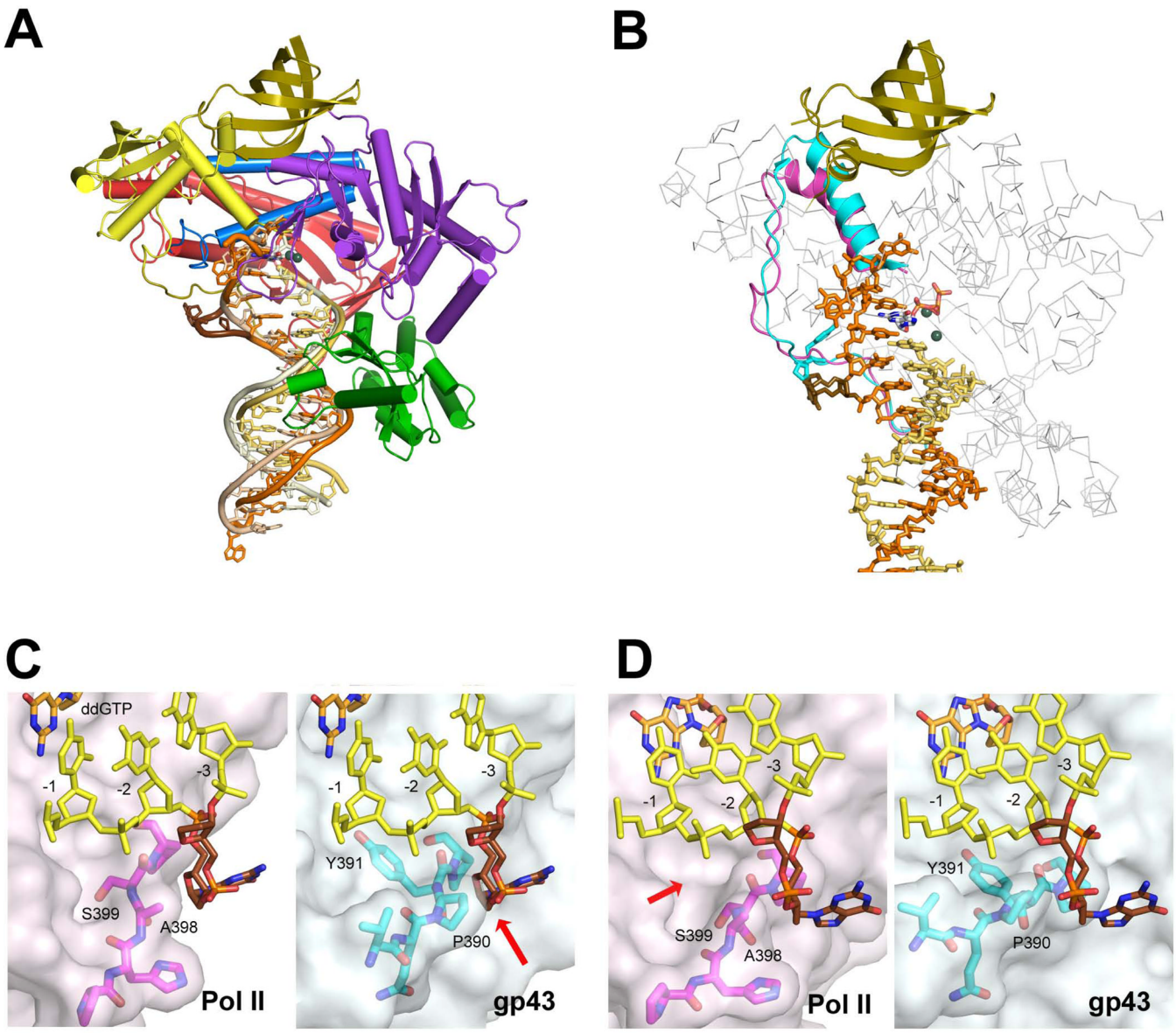
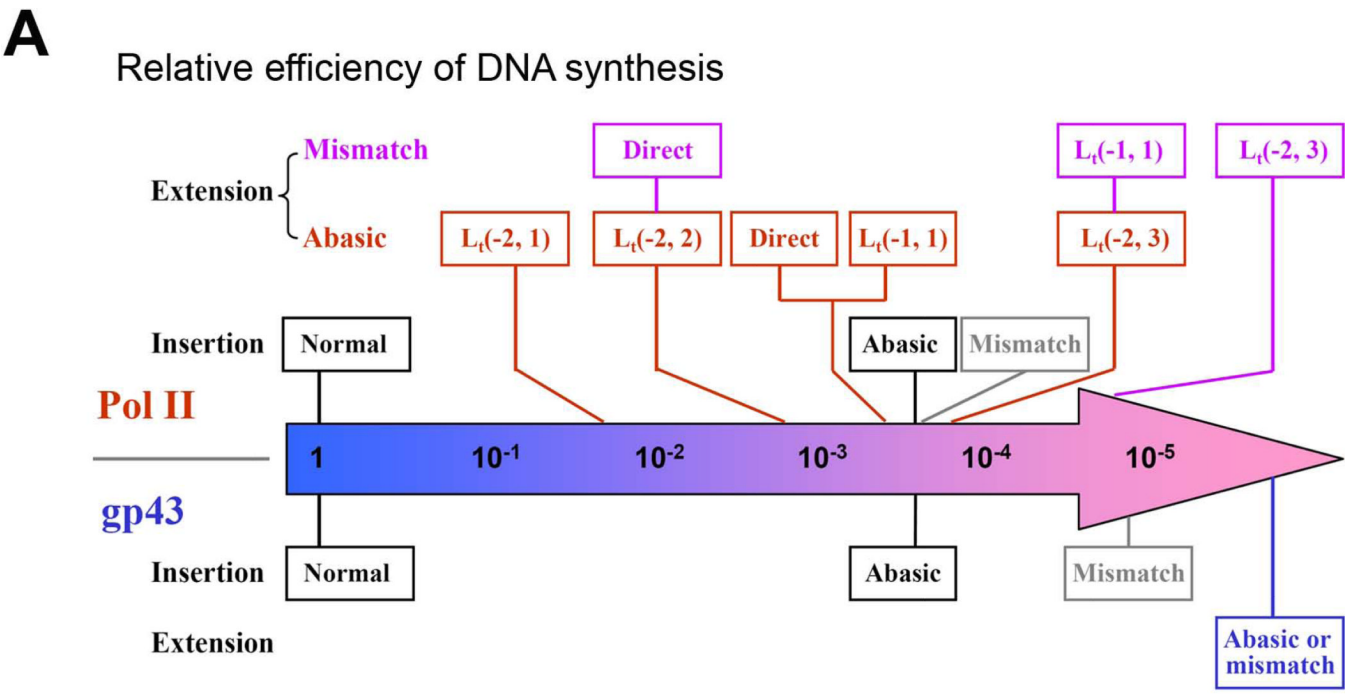


Figure 6.

Crystal structure of primer extension by Pol II using the $L_t(-2,2)$ mechanism. A. Ribbon diagram of the $L_t(-2,2)$ structure. The protein domains are color coded as in Fig. 1A. The primer is colored yellow, and the template brown. The looped out region is highlighted in dark brown. The DNA from a normal ternary complex (pale yellow and orange) is superimposed for comparison. **B.** Superposition of the gp43 ternary complex and $L_t(-2,2)$. The N-palm linker of gp43 and bulky sidechains (Pro and Tyr) (cyan) would clash with the looped-out nucleotides (brown). Due to the β barrel insertion (olive color) in Pol II, the N-palm linker (magenta) is more relaxed than gp43. The Cα trace of the remaining Pol II is shown in silver, the template (orange), primer (yellow), the incoming dGTP (grey and blue) and Ca^{2+} (green) are also shown. **C.** A close-up view of the nucleotides (brown) looped-out from the template (yellow). They are accommodated in Pol II but would clash with gp43 (indicated by the red arrow). **D.** Pol II also has a cavity between the −1 and −2 position (red arrow). Gp43 is devoid of such cavity. In C and D, the molecular surfaces of Pol II and gp43

are shown as semi-transparent, and the DNA-contacting residues are highlighted in magenta (Pol II) and cyan (gp43).



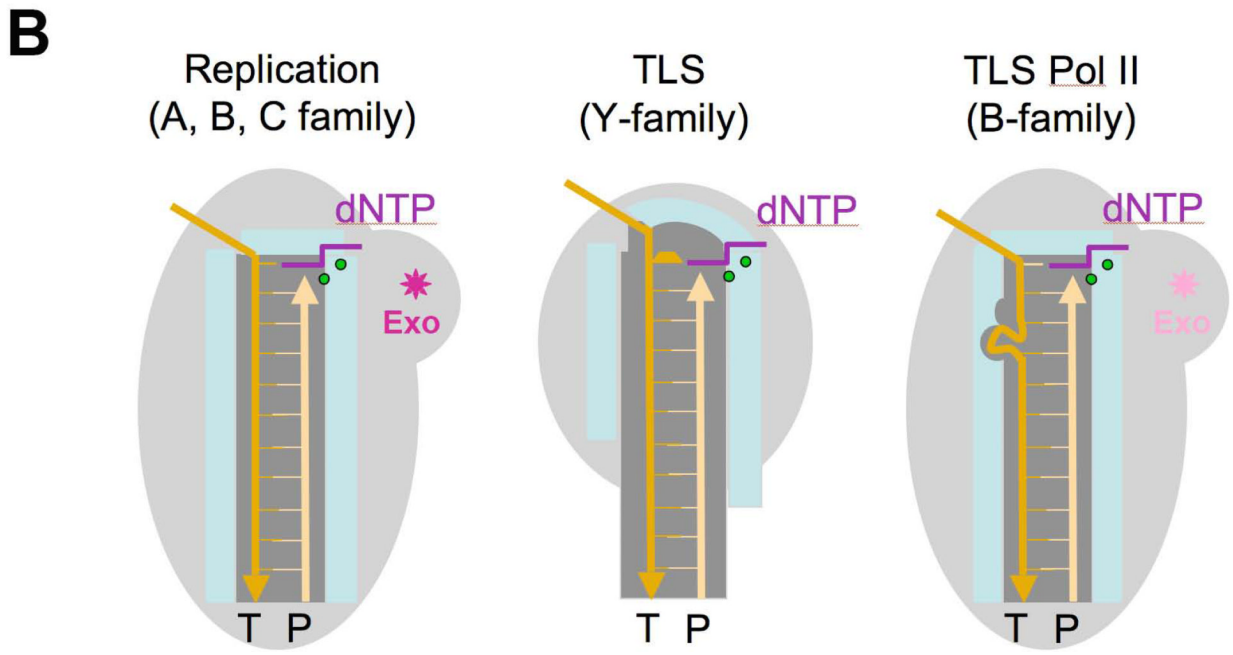


Figure 7.

Differences between repair and replicative polymerases. **A.** Normalized TLS efficiencies (k_{cat}/k_M) of Pol II and gp43 (Table S3) are indicated on the 5-logrithm scale bar. The two polymerases have similar catalytic efficiencies on normal DNA and when inserting dA opposite an abasic lesion. But Pol II is more efficient in all other TLS and mutagenic DNA synthesis. **B.** Features that differentiate repair from replicative polymerases. The left panel shows a high-fidelity polymerase, which has a snug polymerase active site and a proofreading Exo activity. The middle panel shows a characteristic Y-family polymerase, which has a large active site to accommodate DNA lesions and no proofreading activity. The

right panel shows Pol II, which has small pockets along the template strand to loop out DNA lesions and a weakened proofreading activity due to the altered DNA partitioning. The proofreading activity is often absent in other A and B-family TLS polymerases.