6-3-2009

Structural Basis of Error-prone Replication and Stalling at a Thymine Base by Human DNA Polymerase iota

Kevin N. Kirouac
University of Western Ontario, kkiouac@uwo.ca

Hong Ling
University of Western Ontario, hling4@uwo.ca

Structural basis of error-prone replication and stalling at a thymine base by human DNA polymerase ι

Kevin N Kirouac & Hong Ling*

Department of Biochemistry, University of Western Ontario, London, Ontario, Canada, N6A 5C1

***Correspondence should be addressed to H. L**. (hling4@uwo.ca)

Subject Category: Structural Biology; Genome Stability and Dynamics

Characters (with spaces) 48 073

Figures 5

Table 1

Running Title: Polymerase ι misinserts and stalls at T bases

Competing interest statement: The authors declare no competing financial interests

ABSTRACT

Human DNA polymerase ι (polι) is a unique member of Y-family polymerases, which preferentially misincorporates nucleotides opposite thymines (T), and halts replication at T bases. The structural basis of the high error rates remains elusive. We present three crystal structures of polι complexed with DNA containing a thymine base, paired with correct or incorrect incoming nucleotides. A narrowed active site supports a pyrimidine:pyrimidine mismatch and excludes Watson-Crick base pairing in polι. The template thymine remains in an *anti* conformation irrespective of incoming nucleotides. Incoming ddATP adopts a *syn* conformation with reduced base stacking while incorrect dGTP and dTTP maintain *anti* conformations with normal base stacking. Further stabilization of dGTP by H-bonding with Gln59 from the finger domain explains the preferential T:G mismatch. A template 'U-turn' is stabilized by polι and the methyl group of the thymine template, revealing the structural basis of T stalling. Our structural and domain swapping experiments indicate that the finger domain is responsible for polι's high error rates on pyrimidines and determines the incorporation specificity.

Key Words: Y family DNA polymerase/ pol ι/ incorporation specificity/ mutagenesis/ translesion synthesis

INTRODUCTION

A strict adherence to Watson-Crick base pairing is a key feature in determining the high fidelity of replicative DNA polymerases. An induced fit mechanism is employed when a correct incoming nucleotide is optimally paired with the template base (Doublie et al., 1999). A highly restrictive active site with limited solvent accessibility enables this efficient base pair selection (Doublie et al., 1998). In contrast, Y family DNA polymerases, which specialize in traversing DNA lesions have evolved an open and solvent accessible active site allowing for permissive base pairing (Ling et al., 2001). The Y-family polymerases contain a similar catalytic core, consisting of "palm", "finger" and "thumb" domains as in high-fidelity DNA polymerases. The smaller finger and thumb domains in Y-family polymerases generate a solvent-accessible and spacious active site. The finger domain contacts the replicating base pair and is therefore the substrate recognition site. In addition, the Y-family polymerases possess a unique C-terminal domain, called the "little finger" or polymerase associated domain (PAD). The little finger (LF) holds the DNA substrate along with the thumb domain. Around the active site, the major groove of the DNA duplex is fully solvent exposed. Thus, bulky DNA adducts can be accommodated in the active site and multiple nucleotide conformations can be adopted in order for the enzyme to replicate through the lesion. The caveat to performing translesion DNA synthesis is a high error rate of replication on undamaged DNA (Boudsocq et al., 2001; Johnson et al., 2000; Zhang et al., 2000).

Human Y family DNA polymerase ι (polι) is a specialized polymerase that does not utilize Watson-Crick base pairing for DNA replication. Instead, this enzyme functions by inducing a *syn* conformation on template purines, which results in

Hoogsteen base pairing with the correctly matched incoming nucleotide (Johnson et al., 2005; Nair et al., 2005; Nair et al., 2006). The ability to induce a nucleotide *syn* conformation by polι appears to serve as the mechanism for replication opposite damaged template purines. Structural evidence demonstrates that the 1, N^6 ethenodeoxyadenosine and N2 ethyl guanine lesions are presented in the *syn* conformation protruding into the solvent accessible major groove of the DNA helix (Nair et al., 2006; Pence et al., 2008), which allows base pairing with the correct incoming nucleotide.

DNA replication by polι on template pyrimidines displays extremely high error rates, while incorporation opposite template purines is more accurate (Kunkel et al., 2003; Tissier et al., 2001; Tissier et al., 2000; Zhang et al., 2000). Opposite a template thymine (T), polι prefers to incorporate a guanine (G) up to 2.5 fold over the correctly paired adenine (A) in a metal-dependent manner (Frank and Woodgate, 2007). In addition, polι has inefficient replication past a template T base causing a signature T template stall (Zhang et al., 2000). A similar pattern of misincorporation and replication stalling by polι is observed opposite template uracil (U) (Vaisman and Woodgate, 2001). It has been proposed that G misincorporation opposite template U could restore the genomic sequence of cytosines (C) that have undergone deamination. The biological role of this preferred misinsertion of G opposite template T or U within cells remains unknown. However, such an unusual and highly specific property could serve a unique function in DNA maintenance.

The error-prone replication on template T by polι has been implicated in the high rates of DNA mutagenesis presented in patients with the UV-sensitive disorder; *Xeroderma Pigmentosum Variant* (XP-V) syndrome (Wang et al., 2007). When the

human Y family DNA polymerase η (polη) is inactivated by mutations, polι takes over its specialized role of bypassing UV-induced thymine-thymine (T-T) dimers. However, the preference of misinsertion opposite template T by polι results in an increase in mutagenesis and the presentation of the disease. Although polι is responsible for increased DNA mutagenesis when functioning out of context, this enzyme does appear to play a role in tumour suppression. Mice deficient in both polη and polι have an earlier onset on UV-light induced tumors than polη deficiency alone (Dumstorf et al., 2006; Ohkumo et al., 2006), indicating a role for polι in UV-induced lesion bypass. In addition, it has recently been observed that polι plays a significant role in cellular protection from oxidative damage (Petta et al., 2008). Although polι likely facilitates the repair of oxidative DNA lesions, the specificity and mechanism of this repair is unknown. The unique T template misincorporation by polι has been extensively reported, but has remained mechanistically unexplained.

Here, we report three crystal structures of polι in complex with DNA containing a template T base in the active site, which is paired with either correct (A) or incorrect (T or G) incoming nucleotides. Our results reveal, for the first time, the structural basis of preferred G misincorporation and stalling on a template T base by polι.

RESULTS

Polι-DNA-dNTP complexes with template bending back at the T base

In order to position the thymine base at the polι active site, DNA substrates for crystallization were designed containing a thymine 5´ to the template-primer junction. The first (substrate 1) is a 15/9-nt duplex DNA with two thymines 5' to the template-

primer junction (see Methods). The second (substrate 2) is an 18-nt self-complementary duplex (14 base pairs) containing a 2, 3-dideoxy 3′ primer end for trapping ternary complexes (see Methods). Incoming 2, 3-dideoxy ATP (ddATP) was incubated with DNA substrate 1 and co-crystallized with polι, while dGTP and dTTP were incubated with DNA substrate 2 and co-crystallized with polι. Interestingly, the polι-DNA-ddATP complex was trapped at the first T from the template-primer junction without the expected one incorporation, probably due to the replication stalling at T. The resulting three crystal structures are denoted as T:ddADP, T:dGTP and T:dTTP according to the replicating base pair in the active site. The ternary complex crystals are in two different space groups (C2 for T:ddADP; $P6_522$ for T:dGTP and T:dTTP) and diffract to 2.0 Å, 2.0 Å, and 2.2 Å resolutions, respectively (Table I). In T:ddADP, incoming ddATP was hydrolyzed to ddADP. Hydrolysis in this manner has been observed in Dpo4 from *Sulfolobus solfataricus*, the model enzyme of the Y-family DNA polymerase, due to a weak phosphatase activity of the polymerase (Ling et al., 2001). In the T:ddADP structure, the active site metal ions have been refined as Ca^{2+}, due to the presence of 150 mM $CaCl_2$ in the crystallization buffer and the high electron density. In addition, anomalous signal peaks were observed at the metal ion sites which are distinct from surrounding non-metal atoms though weak at the A site (Figure 4S). This is analogous to Dpo4 structures crystallized with 100 mM $Ca(AC)_2$ (Ling et al., 2001; Wong et al., 2008). Such anomalous peaks were not observed for the T:dTTP or T:dGTP structures, which were crystallized in the absence of Ca^{2+} ions (Figure 5S). Thus, the T:dTTP and T:dGTP structures were refined with two active site Mg^{2+} ions. Primer extension assays have been preformed on polι in the presence of 150 mM $CaCl_2$, which demonstrates that Ca^{2+}

ions do not change the nucleotide incorporation specificity of polι (Figure S1). The divalent cation in the B site is positioned identically within all three T template structures, as well as previous polι structures containing template purines (Nair et al., 2006) and Dpo4 ternary structure (Vaisman et al., 2005) (Figure S2). The divalent cation in the A site however, is mobile with variable positions in all three structures (Figure S2) Divalent ion mobility within the A site has also been reported previously for polι (Nair et al., 2005) and Dpo4 (Vaisman et al., 2005).

Polι in all three ternary structures is essentially identical to that of the previously solved, purine-template polι structures (Nair et al., 2005; Nair et al., 2006; Nair et al., 2004) (Figure 1A). The pair-wise comparisons on all Cα atoms produced root mean square deviations (rmsd) within 0.7 Å among our three complex structures. In addition, the Cα rmsd is ~0.8 Å between T:ddADP and a previously solved polι ternary complex (PDB: 2ALZ) containing a purine base at the template position. The close agreement between all of these polι structures indicates that polι, like other Y-family polymerases, does not undergo significant conformational change when replicating through different DNA substrates (Bauer et al., 2007; Ling et al., 2001; Nair et al., 2006).

We use the first complex structure (T:ddADP, Figure 1C) to describe the general features of the three ternary complexes, since these three structures are identical, except for the replicating base pair. All residues of polι have the same side-chain conformations in the current three structures, which are identical to those of the previously reported purine-template polι structures, except for Tyr 61. Tyr 61 flips its side chain conformation 100° from that seen in the purine-template polι structures, and moves its aromatic ring 9 Å closer to the template DNA (Figure 1A). The unique Tyr61 orientation

is observed in all three of our template thymine structures reported here, which are in two different crystal forms. Thus, this conformation is independent of crystal packing and is likely induced by the DNA substrate that contains a template T base (see details below).

A striking structural change is observed in the single-stranded DNA template when comparing our T template base structures with previous polι structures containing template purines and with Y-family polymerase Dpo4 (Figure 1). The previous polι and Dpo4 structures project the single stranded template DNA away from the active site in extended conformations (Figure 1B-D). In our T template base structures, the single stranded DNA is flipped back upon itself in a 'U' shaped conformation, enclosing the replicating base pair from the DNA's major groove and approaching the polymerase thumb domain across the major groove (Figure 1B,C). The DNA backbone is bent ~90^{o} after the template T base towards the major groove, and the +1 nucleotide (5' to T) is oriented 90^{o} to the template T base. This latter difference is completely different from the template/+1 base relationships in the extension template strands in the other Y-family polymerase structures (Figure 1B). In our structures, the +1 base lies perpendicular to the template T base (position 0) due to DNA strand binding. All three of our polι structures display this unique 'U-turn' DNA conformation after the T base, irrespective of the incoming nucleotide or the identity of the bases flanking the template T.

Contributions of polι domains and T template base to template stabilization

The unique 'U-turn' DNA conformation in our polι structures is stabilized by both polι and the unique T template base within the active site, which likely induces replication stalling. The single-stranded template DNA downstream of the T base is held in position by the finger domain, the little finger (LF) domain and the thumb domain of polι (Figure

2A). Tyr 61, Leu 62 and Leu 78 from the finger domain contact the +1 nucleotide and provide strong hydrophobic interactions to the backbone sugar via the aromatic side chain of Tyr 61 and to the +1 nucleotide base from interactions with the two leucine residues (Figure 2B). The Positively charged Arg 347 and polar Ser 307 from the LF domain contact the phosphate backbone of the template at position 0 and +1 nucleotides (Figure 2B). The 5′ end of the template is in contact with the thumb domain at Tyr 244 and Lys 237 (Figure 2C). Tyr 244 stacks with the sugar of the +3 nucleotide, and the positively charged Lys 237 interacts with the negatively charged phosphate to fix the free 5′ end of the template strand in front of the replicating base pair (Figure 2C). Furthermore, the ~90° bend at the T base is stabilized by the interactions between the methyl group of the template T base and the bent single-stranded template DNA (Figure 2D). Although the extensive contacts between this unique methyl group and the +1 nucleotide reinforce the unusual bending template, the methyl group may not be an absolute requirement, due to a similar stalling effect opposite template U (Vaisman and Woodgate, 2001). The unique 'U-turn' conformation is not observed in the presence of template purine bases (Nair et al., 2005; Nair et al., 2006). Two of our template thymine structures (T:dTTP and T:dGTP) have the same sized DNA substrate, single stranded DNA overhang, and crystal form as the previous purine template structures. However, these template thymine structures adopted the 'U-turn' conformation, similar to the T:ddADP structure, which has a different DNA substrate and crystal form (Figure 1A). This excludes any structural variation caused by differences in the single stranded DNA and packing environments of the complexes. It appears that the purine bases A and G are too large to be accommodated in the U-turn conformation observed in our structures, which would

disrupt bending. However, there are likely other unidentified factors also involved in preventing this conformation. Although template C has a similar size to T and U, which may lend itself to a 'U-turn' structure, such a conformation may not be stable when template C is in the active site, as no significant replication stalling has been observed with this template base. The U-turn interactions are not involved with the template bases in the double strand DNA except the template T and are mainly involved with backbone atoms on the downstream single strand DNA. This is consistent with the observations that the stalling only depends on the T base and not the bases flanking it (Zhang et al., 2000). The current T template structures clearly show the structural basis for the signature T template stalling by polι.

Role of finger domain at polι active site

Structural comparison of polι and Dpo4 ternary complexes indicate that the Y-family polymerases are highly structurally conserved in the core area that forms the DNA-binding cleft, except for the finger domains (Figure 3B). The finger domains are structurally conserved with most secondary structural elements aligned between polι and Dpo4 (Figure 3A,B). However, there are two striking differences in these finger domains, which affect the shape of the active sites and their interactions with DNA substrate. These differences are concentrated in the fragment between the β-strands β2 and β3, which is in the non-conserved substrate recognition site that contacts the template DNA and the replicating base pair in the active site (Yang, 2003). First, polι has a much shorter loop between β2 and β3 than that of Dpo4. This loop (L23) forms a structural interface for the finger domains to interact with the LF domains, as well as the-single stranded template DNA in Dpo4 (Figure 3B-E). The shorter loop of polι causes the LF

domain to rotate 12˚ inward to the finger domain and the β9 strand moves ~2 Å towards the template strand, creating a narrowed active site in polι (Figure 3B,C). The narrowed active site limits the C1´-C1´ distance of the replicating base pair to within 9 Å in polι. The C1´-C1´ distances are 8.3 Å, 8.6 Å, and 8.9 Å in the structures of T:ddADP, T:dTTP, and T:dGTP, respectively. In contrast, the replicating base pair in the active site of Dpo4 has a C1´-C1´ distance of ~10.6 Å (Figure 3D), which is a common strand width for B form DNA in all other Y-family polymerase structures (Alt et al., 2007; Lone et al., 2007; Nair et al., 2005).

The second difference between polι and Dpo4 finger domains lies in the fragment that contacts the replicating base pair within the active site. Polι has relatively large amino-acid side chains (Gln 59, Lys 60, Leu 62, Val 64, Leu 78) contacting the replicating base pair at the active site (Figure 3E) when compared to Dpo4, which has relatively small amino acids (Val 32, Gly 41, Ala 42, Ala 44, Gly 58) for the contacts. Although the finger domain is rolled out by ~15˚ relative to Dpo4, the larger side chains of Gln 59, Leu 62, Val 64, and Leu 78 in polι still push the replicating base pair towards the major groove, effectively tilting it off plane relative to that of Dpo4 (Figure 3E). Lys 60 and the residues from strand β9 of the polι LF domain squeeze the template base towards the incoming nucleotide and make the C1´-C1´ distance shorter than 9 Å (Figure 3E). These structures reveal that the finger domain is not only important for contacting the replicating base pair but is also an essential factor for restricting the C1´-C1´ distance in polι. Therefore, the polι finger domain is most likely the functional domain responsible for the nucleotide specificity during replication.

Conformation and stability of replicating base pairs in polι active site

The unique polι active site and the T template base make the replicating base pairs in the ternary complexes different in conformation from those in other Y-family polymerase ternary structures. All three of our polι structures have the template T base in a normal *anti* conformation when it is paired with an incoming nucleotide in the active site (Figure 4). Nucleotide binding does not induce a conformational change in the T template base as observed in the purine-template structures (Nair et al., 2005; Nair et al., 2006). Instead, the incoming dNTPs of the replicating base pairs in our polι structures adopt different conformations, depending on their fit in the enzyme active site. Due to large residues from finger domain, the template T base is pushed out of the stacking area with underneath bases and tilted off plane by the finger domain in all three structures (Figure 4). The tilt (τ) and roll (ρ) angles of the off plane T from the underlying base are around 6˚ and 16˚, respectively.

In the T:ddADP structure, the incoming ddADP adopts a *syn* conformation and forms a Hoogsteen base pair with the template T (Figure 4A). The Hoogsteen base pair in the T:ddADP structure fits the narrowed active site with a C1´-C1´ distance of 8.4 Å, which is similar to other reported polι structures (Nair et al., 2005; Nair et al., 2006; Nair et al., 2006) and is smaller than the required C1´-C1´ distance of ~10.6 Å for proper Watson-Crick base pairing (Ling et al., 2003; Ling et al., 2001; Ling et al., 2004; Ling et al., 2004; Wong et al., 2008). In addition, incoming ddADP is flipped out of the stacking area of the underlying base pair towards the major groove due to its *syn* conformation. The ddADP is tilted ~20° off plane with the underlying base pair and has an elongated

stacking distance of about 4 Å , which weakens the stability of the replicating base pair further (Figure 4A).

In the T:dTTP structure, the mismatched incoming dTTP is in an *anti* conformation (Figure 4B). The narrowed active site holds the pyrimidine- pyrimidine base pair well, due to the pair being smaller than the common pyrimidine-purine base pair in contacting distance. The narrowed active site thus stabilizes the small pyrimidine-pyrimidine mismatched base pair. The C1´-C1´ distance of the T:T base pair is 8.5 Å, which would not be stable in an active site that accommodates a standard Watson-Crick base pair with C1´-C1´ distance of ~10.6 Å. Interestingly, the incoming dGTP is also in the *anti* conformation, which has not been observed in other polι structures that contain template purine bases in the active site (Nair et al., 2005; Nair et al., 2006) (Figure 4C). The C1´-C1´ distance is restricted to 8.9 Å, which causes the template T to tilt an extra 15° off plane in order to accommodate the *anti* conformation of the dGTP nucleotide. Our structural observation shows that a *syn* conformation on purine nucleotides can occur in the template or incoming nucleotide position and is the result of a narrowed active site that constrains the C1´-C1´ distance of the replicating base pair. In contrast to ddADP, the bases of dTTP and dGTP in *anti* conformations remain within the active site, parallel to the underlying base pair (Figure 4B,C) with stacking distances in the normal range of 3.2 - 3.6 Å. Compared to incoming A and T bases, the G base of dGTP has the largest stacking surface due to its purine base and *anti* conformation. Base stacking between the incoming nucleotide and the underlying base pair is critical to the stability and preference of nucleotide incorporation (Yang, 2006). Therefore, G is most

favourable and A is the least favourable opposite template T in terms of its base-stacking properties.

In addition to base stacking, incoming bases are also stabilized by hydrogen bonding with template bases. There are two hydrogen bonds between incoming A (N^6 and N7) and the template T (N3 and O^4) in the Hoogsteen base pair in T:ddADP (Figure 4A). In the T:dTTP structure, there are also two hydrogen bonds formed between the template T (O^2 and N3) and the incoming dTTP (N3 and O^4) (Figure 4B). Accordingly, the hydrogen-bonding forces of the replicating base pairs are comparable in these two complexes. Thus, the loss of base stacking on ddADP makes the mismatched dTTP more favourable for incorporation than the A base. Incoming dGTP also forms two hydrogen bonds to the template T, as well as a unique third hydrogen bond between its N^2 atom and OE1 of Gln59 from the finger domain (Figure 4C). This special hydrogen bonding with Gln59 is the first to be identified in a polι structure and reveals a unique stabilizing force that favours, over other bases, the mis-incorporation of dGTP opposite template T by polι. The Gln 59 is conserved in polι homologues (Figure S3), signifying its functional importance. The structural observation that incoming dGTP is the most stable incoming nucleotide is supported by the observation that dGTP binding affinity opposite template T by polι is greater than dTTP or dATP and is the same for dTTP binding affinity opposite template A (Washington et al., 2004).

Role of polι finger domain in base incorporation specificity and replication stalling

The finger domain of polι contacts the replicating base pair and pulls the LF domain towards the active site, which contributes to the replication specificity. In order to confirm that the polι finger domain determines the replication specificity, we generated

two Dpo4-ι chimeric proteins with functional domains switched between Dpo4 and polι. The LF domain has been implicated in contributing to replication specificity (Boudsocq et al., 2004), thus the finger or LF domains of Dpo4 were replaced with the corresponding counterparts of polι in the chimeras. Four DNA substrates containing either T, A, C, or G at the first replicating position and a T base at the eighth replicating position were used for the functional assays (Figure 5). The chimeric proteins were tested by primer extension assays with wild-type Dpo4 and polι as controls.

Opposite the T template base, polι has a high misincorporation rate of G and T (vertical arrows in Figure 5A) opposite template T, as the primers (bottom bands) are almost fully reacted for dGTP (lane G) and dTTP (lane T). Multiple bands are observed due to the low processivity of these enzymes. In contrast, Dpo4 incorporates the correct A nucleotide preferentially (lane A) with dramatically reduced reactions with dTTP (lane T), dGTP (lane G), and dCTP (lane C) compared to dATP (Figure 5A). Opposite the A template (Figure 5B), both Dpo4 and polι have quite accurate incorporation with preference for the correct incoming nucleotide dTTP (lanes T in Figure 5B). The polymerases against the G template in Figure 5C show similar patterns, with C preferentially inserted. Interestingly, Dpo4 replicates the C template (Figure 5D) accurately, while polι preferentially inserts G with significant misincorporations of T (vertical arrows in Figure 5D). Remarkably, Dpo4-ι finger domain chimera (Dpo4-ι) adopted a high mis-incorporation rate of G and T opposite template T similar to polι, as the primer bands in the lanes T and G are almost fully reacted (vertical arrows in Figure 5A), which is similar to polι and different from Dpo4 (Figure 5A). Accordingly, opposite template A, G, and C, the Dpo4-ι resembles polι and differs from wild type

Dpo4 (Figure 5B,C). The primer extension assays indicate that replication specificity is dominated by the finger domain of the Y-family polymerases as the finger swapping converts Dpo4 into a polι–like protein in terms of nucleotide incorporation. Interestingly, the other Dpo4-ι chimera with the LF domain swapped into Dpo4 (Dpo4-ι-LF) does not show any base incorporation pattern changes from Dpo4 to pol ι (data not shown), in contrast to what we observed in the Dpo4-ι mutant. This LF-replacement chimera is very different from the Dpo4-Dbh chimeric proteins in which the enzymatic properties of the mutants are mainly influenced by their LF domains (Boudsocq et al., 2004). In the latter case, the LF domain is swapped between two very similar Y-family polymerases Dpo4 and Dbh, which have almost identical substrate recognition sites. The Dpo4-Dbh chimeras show functional differences of the LF domains between homologues sharing very similar finger domains (Boudsocq et al., 2004). In our case, the dramatic difference between the finger domains of Dpo4 and polι masks the influence from the LF domains. Overall, the mutagenesis data clearly support the structural observations that the finger domain plays an important role in nucleotide incorporation specificity, particularly for G and T mis-insertion opposite template T and in determining the replication specificity.

Replication stalling is observed for polι at the T template bases (horizontal arrows in Figure 5) but not for Dpo4 and Dpo4-ι. When all four nucleotides (lanes 4) are present in the assay, polι has poor extension beyond the first T template base (Figure 5A) and stops at the downstream 8th T base (labelled with horizontal arrows in Figure 5), while Dpo4 extends the primer to the end of the template DNA (top bands in Figure 5) with better processivity than polι. The finger domain alone does not appear to control the stalling property of the enzymes, as the chimeric protein Dpo4-ι extends replication

beyond the 8th T base. This finding is consistent with our structural observation that three domains, instead of the finger domain alone, contribute to the stabilization of the 'U-turn' DNA that leads to replication stalling at the T base.

DISCUSSION

The polι finger domain creates a unique active site that induces low fidelity opposite pyrimidines

Polι displays a wide diversity on its fidelity between template purines versus template pyrimidines (Kunkel et al., 2003; Tissier et al., 2001). Although this enzyme has low error rates opposite template purines, it has the highest error rate of any known polymerase opposite template thymines (Johnson et al., 2000; Tissier et al., 2000; Zhang et al., 2000). The diversified substrate-recognition site in the finger domain changes in size and residue identity across the Y-family members and is expected to be responsible for the specificity of nucleotide incorporation (Ling et al., 2001; Yang, 2003). The β2-turn-β3 loop (L23) is the only part of the polymerase core that contacts the LF domain in the Y-family ternary complex structures. Our structural analyses indicate that the shorter polι L23 of the finger domain induces a movement of the LF domain towards the template strand, which results in a narrowed active site. It is conceivable that the finger domain causes the LF shifting towards the template in the active site, as the LF is the most flexible domain in the Y-family structures (Wong et al., 2008).

Polι promotes a T:G mismatch by the its unique narrowed active site and specific interactions with the replicating base pair. As a good structural fit, the narrowed active site supports pyrimidine:pyrimidine mismatches when the template base is a pyrimidine.

Higher misincorporations of T:T, T:C, C:T and C:C were observed in our primer extension assays for both polι and Dpo4-ι relative to misincorporations opposite template purines. When the template base is a purine, the smaller, incoming pyrimidine maintains its base stacking in *anti* conformations (Nair et al., 2006), while a larger, incoming purine nucleotide would be difficult to fit in the narrowed active site for a purine:purine mismatch. This mechanism prevents misincorporation against purine template bases and allows for accurate replication. Our structural and biochemical analyses are consistent with the well documented, high error rates against template pyrimidine bases and relatively high fidelity against purine bases by polι (Johnson et al., 2000; Tissier et al., 2000; Zhang et al., 2000).

Interestingly, in the presence of Mn^{2+}, polι has increased fidelity opposite template thymine with a preference of incorporating the correct A nucleotide instead of incorrect G (Frank and Woodgate, 2007). The Mn^{2+} ion has a more relaxed and mobile coordination than Mg^{2+} within the active site of polymerases. This effect likely allows the incoming nucleotide to adopt a variety of conformations that would not be possible with Mg^{2+}. In this manner, Mn^{2+} ion coordination by polι may render a favourable interaction that selects A over G opposite template T.

Replication stalling is stabilized by conserved residues over three domains

Another unique feature of polι is a pronounced stalling of replication in extending a primer strand opposite a T base (Zhang et al., 2000). Our structures of polι reveal a unique template DNA 'U-turn' conformation at the DNA's single-stranded side that may effectively stall replication. The back-bending 'U-turn' conformation is stabilized by specific interactions from the unique methyl group of the template T base and a collection

of interactions from three domains of polι. Three domains are involved in interacting with the "U-turn": the finger domain (Tyr 61, Leu 62, and Leu 78), little finger domain (Ser 307 and Arg 347), and thumb domain (Lys 237 and Tyr 244). These combined interactions stabilize the bent single-stranded DNA and appear to hinder its translation into the active site for primer elongation. In addition, the highly bent DNA may also reduce the catalytic efficiency of polι, due to the observation that dNTP incorporation opposite template T is much slower than opposite template A (Washington et al., 2004). Interestingly, most of the residues contacting the single-stranded template are unique to and conserved in polι from difference species (Figure S3), suggesting conservation of specific functions. Replication stalling by polι may be involved in recruiting another polymerase for primer extension after insertion opposite template T or U.

Conclusions

Polι uniquely replicates DNA with a constrained active site, creating shorter C1´-C1´ strand distances, with the finger domain projecting the template base out towards the solvent-accessible major groove and stabilizing a mismatched G base via H-bonding. The finger domain of polι is responsible for the unique active site, and in turn, its replication specificity. This feature allows polι to maintain a relatively high fidelity on template purines, yet induce high rates of misincorporation on template pyrimidines. The high fidelity on template purines by polι appears to play a role in translesion synthesis by allowing accurate replication through adducted purine bases. The biological role of polι's low fidelity is still unclear; however, it is apparent that when functioning out of context, this unique replication specificity induces high rates of DNA mutagenesis.

METHODS

Dpo4-ι chimeric proteins

To construct the Dpo4-ι finger domain chimera (Dpo4-ι) and the Dpo4-ι little finger domain chimera (Dpo4-ι-LF), plasmid vector pET-22b containing the Dpo4 gene and plasmid vector pHis-parrallel1 containing the polι gene were used as templates for PCR. For Dpo4-ι, the N-terminus of Dpo4 was cloned up to the beginning of the finger domain using primers A

(5'- C GTT ACT GCC ATG GTT GTT CTT TTC GTT G -3') and B

(5'-TTCTACTTGTGCATAAAAGCAGTCAAAATCAACGAAAAGAACAATC-3').

The result was an N-terminus Dpo4 PCR product containing an NcoI cutting site at the N terminus and a C terminal overhang, which was complementary to the beginning of the polι finger domain. The C-terminus of Dpo4 was cloned past the end of the finger domain using primers C

(5'-GTTGGTATTAGTTAATGGAGAAGACAAGGAAGTATATCAGCAAGTTTC-3')

and D (5'-GCTAGTTATTGCTCAGC-3'). The result was a C-terminal Dpo4 PCR product with an N terminal overhang, complementary to the end of the polι finger domain. The finger domain of polι was cloned using primers E

(5'-TGC TTT TAT GCA CAAGTAGAAATG-3') and F

(5'-GTCTTCTCCATTAACTAATACCAAC-3'). The N-terminal Dpo4 product was joined with the polι finger domain product using primers A and F. The resulting N-terminal Dpo4-polι finger domain product was joined with the C-terminal Dpo4 product using primers A and D to produce the final product of a Dpo4 gene containing the finger domain of polι (Dpo4-ι).

For Dpo4-ι-LF, the N-terminus of Dpo4 was cloned up to the beginning of the little finger domain using primers A

(5'- C GTT ACT GCC ATG GTT GTT CTT TTC GTT GAT TTT GAC TAC TTT TAC GCT C -3') and B

(5'- AGT TCT TAT AGG CTC GTT ATA CTC GTC TCT AGC TAG A -3').

The result was an N-terminus Dpo4 PCR product containing an NcoI cutting site at the N terminus. The little finger domain of polι was cloned using primers C

(5'- GCC GTT ACT GCC ATG GTT GTT CTT TTC GTT GAT TTT GAC TAC TTT TAC GCT C -3') and D

(5'- CAT CCT CGA GAC CTA CTT AGC AGT ATT TAG TGC TTT AAG GTT GCA GAA GC-3'). The result was a polι little finger domain with an N-terminal overhang, complimentary to the end of the Dpo4 product. The N-terminal Dpo4 product was joined with the polι little finger domain product using primers A and D to produce the final product of a Dpo4 gene containing the little finger domain of polι (Dpo4-ι-LF).

Both Dpo4-ι and Dpo4-ι-LF genes were cloned into the pHis-parallel1 vector and confirmed by sequencing.

Primer Extension Assays

DNA substrate (10 nM) was incubated with either Dpo4, Dpo4-ι, Dpo4-ι-LF or hpolι (10 nM) and 100 uM of either all four dNTPs or individual dNTPs at 37°C for 2min in reaction buffer containing 40 mM Tris (pH 8.0), 5 mM $MgCl_2$, 250 ug/ml BSA, 10 mM DTT, and 2.5% glycerol. Reactions carried out in the presence of 150 mM $CaCl_2$ were incubated at 37°C for 60min. Reactions were terminated with loading buffer (95%

formamide, 20 mM EDTA, 0.025% xylene, 0.025% bromophenol blue) and resolved on a 20% polyacrylamide gel containing 7 M urea. Gels were visualized using a PhosphorImager.

Protein Preparation

Human DNA pol ι (amino acid 1-420) was cloned into pGST-parrallel1 vector and the subsequent glutathione S-transferase-tagged polι was over expressed in Escherichia coli strain DE3. The polι-GST fusion protein was purified using affinity chromatography and cleaved using a histidine-tagged tobacco etch virus (TEV) protease, which was subsequently removed using nickel affinity chromatography. The cleaved polι containing 2 extra N-terminal residues was further purified using an SP column. Dpo4 used for functional assays was purified as previously described (Ling et al., 2001). The His tagged Dpo4-ι chimeric proteins used for functional assays were overexpressed in Escherichia coli strain DE3 and purified using nickel affinity chromatography followed by an SP column

DNA preparation

Oligonucleotides for crystallization were purchased from Keck Oligo Inc. and gel purified. The 9-nt primer (5'-GTGGATGAG-3') was annealed to a 15-nt template (5'-CTCATTCTCATCCAC-3'), and the self-annealing 18-nt oligonucleotide (5'-TCATGGGTCCTAGGACCC^{dd}-3') was annealed with itself to give a DNA substrate with two replicative ends. Oligonucleotides used for primer extension assays were purchased from Sigma Aldrich and gel purified. A 30-nt template

(5'-GAGCAGTCGCACATGTAGTATCTCTGTGAC-3') was annealed to a 16-nt primer (5'-GTCACAGAGATACTAC-3') resulting in a template T base at the first and eighth position beyond the primer-template junction. The primer was 5'-end labelled using [γ-32P]ATP and T4 polynucleotide kinase. The 5'-labelled primer was mixed with template DNA at a 1.5:1 molar ratio and heated at 95°C, followed by slow cooling to form the annealed DNA substrate.

Crystallization and Structure Determination

Ternary complexes were formed for T:ddADP, T:dGTP, and T:dTTP by incubating protein (0.2 mM) and DNA in a 1:1.2 ratio with either ddNTP or dNTP (1mM), and $MgCl_2$ (5 mM). Crystals of the T:dGTP and T:dTTP complex were obtained in 12% PEG 5000 MME + 0.2M NH_4SO_4 + 5% glycerol + 0.1 M MES, pH 6.5, while crystals of the T:ddADP complex were obtained in 12% PEG 3350 + 0.15 M $CaCl_2$ + 0.01 M DTT + 5% glycerol. All crystals were flash frozen in liquid nitrogen using paratone-N as a cryo protectant. X-ray diffraction data were collected on the beamline 24-ID-C at the Advanced Photon Source in Argonne National Laboratory. The data were processed and scaled using HKL (Otwinowski and Minor, 1997).

All three structures were solved using molecular replacement with a previously solved ternary complex (PDB:2ALZ) as a search model. Rigid body refinement was performed using REFMAC (Murshudov et al., 1997), followed by restrained refinement and then TLS refinement. Electron density was well defined for all structures except for the first 27 residues of the N-terminus, loop regions 332-337, 350-356, and 371-378 and the last 6 residues of the C-terminus. Additionally, the +4 and +5 nucleotide within T:ddADP and the +3, and +2 nucleotides within T:dGTP and T:dTTP were disordered.

All structures have good stereochemistry with over 95% of the residues in the most favoured region of the Ramachandran plot.

Coordinates

The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.rcsb.org, with accession codes 3GV5, 3GV7 and 3GV8 for the structures T:ddADP, T:dGTP, and T:dTTP, respectively.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online

ACKNOWLEDGEMENTS

We thank G. Xing for subcloning the N-terminal (1-420) polymerase domains from the full-length polι, and K. Rajashankar for beamline support at 24-ID of APS in Argonne National Laboratory. This research was funded by the Canadian Institutes of Health Research (H.L.).

References

Alt, A., Lammens, K., Chiocchini, C., Lammens, A., Pieck, J.C., Kuch, D., Hopfner, K.P. and Carell, T. (2007) Bypass of DNA lesions generated during anticancer treatment with cisplatin by DNA polymerase eta. *Science*, **318**, 967-970.

Bauer, J., Xing, G., Yagi, H., Sayer, J.M., Jerina, D.M. and Ling, H. (2007) A structural gap in Dpo4 supports mutagenic bypass of a major benzo[a]pyrene dG adduct in DNA through template misalignment. *Proc Natl Acad Sci U S A*, **104**, 14905-14910.

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

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

Doublie, S., Sawaya, M.R. and Ellenberger, T. (1999) An open and closed case for all polymerases. *Structure Fold Des*, **7**, R31-35.

Doublie, S., Tabor, S., Long, A.M., Richardson, C.C. and Ellenberger, T. (1998) Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. *Nature*, **391**, 251-258.

Dumstorf, C.A., Clark, A.B., Lin, Q., Kissling, G.E., Yuan, T., Kucherlapati, R., McGregor, W.G. and Kunkel, T.A. (2006) Participation of mouse DNA polymerase iota in strand-biased mutagenic bypass of UV photoproducts and suppression of skin cancer. *Proc Natl Acad Sci U S A*, **103**, 18083-18088.

Frank, E.G. and Woodgate, R. (2007) Increased catalytic activity and altered fidelity of human DNA polymerase iota in the presence of manganese. *J Biol Chem*, **282**, 24689-24696.

Johnson, R.E., Prakash, L. and Prakash, S. (2005) Biochemical evidence for the requirement of Hoogsteen base pairing for replication by human DNA polymerase iota. *Proc Natl Acad Sci U S A*, **102**, 10466-10471.

Johnson, R.E., Washington, M.T., Haracska, L., Prakash, S. and Prakash, L. (2000) Eukaryotic polymerases iota and zeta act sequentially to bypass DNA lesions. *Nature*, **406**, 1015-1019.

Johnson, R.E., Washington, M.T., Prakash, S. and Prakash, L. (2000) Fidelity of human DNA polymerase eta. *J Biol Chem*, **275**, 7447-7450.

Kunkel, T.A., Pavlov, Y.I. and Bebenek, K. (2003) Functions of human DNA polymerases eta, kappa and iota suggested by their properties, including fidelity with undamaged DNA templates. *DNA Repair (Amst)*, **2**, 135-149.

Ling, H., Boudsocq, F., Plosky, B.S., Woodgate, R. and Yang, W. (2003) Replication of a *cys--syn* thymine dimer at atomic resolution. *Nature*, **424**, 1083-1087.

Ling, H., Boudsocq, F., Woodgate, R. and Yang, W. (2001) Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell*, **107**, 91-102.

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

Ling, H., Sayer, J.M., Plosky, B.S., Yagi, H., Boudsocq, F., Woodgate, R., Jerina, D.M. and Yang, W. (2004) Crystal structure of a benzo[a]pyrene diol epoxide adduct in a ternary complex with a DNA polymerase. *Proc Natl Acad Sci U S A*, **101**, 2265-2269.

Lone, S., Townson, S.A., Uljon, S.N., Johnson, R.E., Brahma, A., Nair, D.T., Prakash, S., Prakash, L. and Aggarwal, A.K. (2007) Human DNA polymerase kappa encircles DNA: implications for mismatch extension and lesion bypass. *Mol Cell*, **25**, 601-614.

Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*, **53**, 240-255.

Nair, D.T., Johnson, R.E., Prakash, L., Prakash, S. and Aggarwal, A.K. (2005) Human DNA polymerase iota incorporates dCTP opposite template G via a G.C + Hoogsteen base pair. *Structure*, **13**, 1569-1577.

Nair, D.T., Johnson, R.E., Prakash, L., Prakash, S. and Aggarwal, A.K. (2005) Rev1 employs a novel mechanism of DNA synthesis using a protein template. *Science*, **309**, 2219-2222.

Nair, D.T., Johnson, R.E., Prakash, L., Prakash, S. and Aggarwal, A.K. (2006) Hoogsteen base pair formation promotes synthesis opposite the 1,N6-ethenodeoxyadenosine lesion by human DNA polymerase iota. *Nat Struct Mol Biol*, **13**, 619-625.

Nair, D.T., Johnson, R.E., Prakash, L., Prakash, S. and Aggarwal, A.K. (2006) An incoming nucleotide imposes an anti to syn conformational change on the templating purine in the human DNA polymerase-iota active site. *Structure*, **14**, 749-755.

Nair, D.T., Johnson, R.E., Prakash, S., Prakash, L. and Aggarwal, A.K. (2004) Replication by human DNA polymerase-iota occurs by Hoogsteen base-pairing. *Nature*, **430**, 377-380.

Ohkumo, T., Kondo, Y., Yokoi, M., Tsukamoto, T., Yamada, A., Sugimoto, T., Kanao, R., Higashi, Y., Kondoh, H., Tatematsu, M., Masutani, C. and Hanaoka, F. (2006) UV-B radiation induces epithelial tumors in mice lacking DNA polymerase eta and mesenchymal tumors in mice deficient for DNA polymerase iota. *Mol Cell Biol*, **26**, 7696-7706.

Otwinowski, Z. and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol*, **276**, 307-326.

Pence, M.G., Blans, P., Zink, C.N., Hollis, T., Fishbein, J.C. and Perrino, F.W. (2008) Lesion bypass of N2-ethylguanine by human DNA polymerase Iota. *J Biol Chem*.

Petta, T.B., Nakajima, S., Zlatanou, A., Despras, E., Couve-Privat, S., Ishchenko, A., Sarasin, A., Yasui, A. and Kannouche, P. (2008) Human DNA polymerase iota protects cells against oxidative stress. *Embo J*, **27**, 2883-2895.

Tissier, A., Frank, E.G., McDonald, J.P., Vaisman, A., Fernandez de Henestrosa, A.R., Boudsocq, F., McLenigan, M.P. and Woodgate, R. (2001) Biochemical characterization of human DNA polymerase iota provides clues to its biological function. *Biochem Soc Trans*, **29**, 183-187.

Tissier, A., McDonald, J.P., Frank, E.G. and Woodgate, R. (2000) poliota, a remarkably error-prone human DNA polymerase. *Genes Dev*, **14**, 1642-1650.

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

Vaisman, A. and Woodgate, R. (2001) Unique misinsertion specificity of poliota may decrease the mutagenic potential of deaminated cytosines. *Embo J*, **20**, 6520-6529.

Wang, Y., Woodgate, R., McManus, T.P., Mead, S., McCormick, J.J. and Maher, V.M. (2007) Evidence that in xeroderma pigmentosum variant cells, which lack DNA polymerase eta, DNA polymerase iota causes the very high frequency and unique spectrum of UV-induced mutations. *Cancer Res*, **67**, 3018-3026.

Washington, M.T., Johnson, R.E., Prakash, L. and Prakash, S. (2004) Human DNA polymerase iota utilizes different nucleotide incorporation mechanisms dependent upon the template base. *Mol Cell Biol*, **24**, 936-943.

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

Yang, W. (2003) Damage repair DNA polymerases Y. *Curr Opin Struct Biol*, **13**, 23-30.

Yang, W. (2006) Poor base stacking at DNA lesions may initiate recognition by many repair proteins. *DNA Repair (Amst)*, **5**, 654-666.

Zhang, Y., Yuan, F., Wu, X. and Wang, Z. (2000) Preferential incorporation of G opposite template T by the low-fidelity human DNA polymerase iota. *Mol Cell Biol*, **20**, 7099-7108.

Figure 1 Comparison of polι-DNA-nucleotide ternary structures. The colour schemes are shown either as the colour of the appropriate labels or the colour bars in the panels. (**A**) Superposition of T:ddADP (yellow), T:dTTP (cyan), T:dGTP (magenta), and a previously solved polι ternary complex (PDB: 2ALZ, grey). Proteins are shown as Cα traces and Tyr 61 is shown as sticks to highlight its conformational change as a result of the DNA U-turn. Incoming nucleotides were omitted for clarity. (**B**) DNA superposition of T:ddADP (yellow), Dpo4 ternary structure (PDB: 2AGQ, blue), and a previously solved polι ternary complex (PDB: 2ALZ, grey). Top view is also shown with the template T base in orange and the incoming ddADP in red. (**C**)(**D**) Polι (T:ddADP) and Dpo4(PDB: 2ALZ) ternary complexes. DNA template strands are shown in yellow, T bases in orange, and primer strands in grey. The U-turn DNA and position of the 5′ template end are indicated by the appropriate arrows. LF represents the little finger domain.

Figure 2 Structure of T:ddADP showing template DNA 'U-turn' stabilization. Numbers indicate template nucleotide positions relative to T at position 0. Hydrogen bonding is shown as blue dashed lines. (**A**) Overall T:ddADP structure is shown with DNA template strand in yellow, T base in orange, and the primer strand in grey. The finger domain is shown in blue, little finger (LF) domain in purple, thumb domain in green, and palm domain in grey. (**B**) Zoom-in view of the 'U-turn' stabilization by the polι finger (blue) and LF (magenta) domains. (**C**) Zoom-in view of the 'U-turn' stabilization by the polι thumb domain (green). (**D**) Zoom-in view of the 'U-turn' stabilization by the template T base (orange). Hydrophobic interactions are shown as black dashed lines.

View is as seen from underneath and looking up through the red square in panel a; the vertical arrow in panel a points to the side that is the top of panel d.

Figure 3 Polι and Dpo4 active site comparison. (**A**) Structure-based sequence alignment of amino acids for the finger domains of polι (cyan) and Dpo4 (grey). Numbres 2, 3, and 4 indicate the second, third, and fourth β-sheets. Secondary structure is indicated as rectangles for α-helices and arrows for β-sheets. Residues interacting with the replicating base pair are highlighted in magenta. (**B**) Superposition of T:ddADP (cyan, purple) and ternary Dpo4 (type I)-DNA-nucleotide (1JX4, grey). The incoming nucleotides are shown as sticks for Dpo4 (grey) and T:ddADP (yellow). LF represents the little finger domain. (**C**)(**D**) Close-up views of active sites showing finger-LF domain interactions in polι and Dpo4. Finger domains are cyan, LF domains are purple, and DNA is yellow. (**E**) Active site superposition from b of T:ddADP (finger: cyan; little finger: purple; T base: orange; ddADP: yellow) with Dpo4 (grey). Positioning of the replicating base pair by polι side chains is indicated with black arrows.

Figure 4 Base stacking and hydrogen bonding of replicating base pairs with $2F_o$-F_c electron density maps at 1σ contour level. (**A**) T:ddADP structure (**B**) T:dTTP structure (**C**) T:dGTP structure. The template T base is shown in brown, the incoming nucleotide is shown in yellow, and the underlying base pair is shown in grey. Hydrogen bonds represented as blue dashed lines are shown in top views on the left side. Protein side chains involved in hydrogen bonding are shown in cyan. Green spheres represent divalent cations.

Figure 5 The role of the finger domain in incorporation specificity. Primer extension analysis was used to examine nucleotide incorporation opposite (**A**) template **T** base, (**B**) template **A** base, (**C**) template **G** base, and (**D**) template **C** base by polι, Dpo4-ι (finger domain) chimera, and Dpo4. The first replicating template base and the T template base at the 8^{th} position are bolded. Horizontal arrows indicate replication stalling, while vertical arrows indicate misincorporation. Enzymes were incubated with DNA and either no nucleotides (0), all four dNTPs (4), or individual dNTPs (A,T,C,G).

Table I Summary of crystallographic data

Crystal	T:ddADP	T:dTTP	T:dGTP
Space group	C2	$P6_522$	$P6_522$
Complexes per AU[a]	2	1	1
Unit cell			
a, b, c (Å)	140.2, 71.8, 127.4	97.2, 97.2, 201.9	98.1, 98.1, 203.7
β (°)	112.5		
Resolution range (Å)[b]	52.0-2.0 (2.04-2.00)	27.0-2.2 (2.26-2.20)	24.0-2.0 (2.07-2.00)
R_{merge}[b]	7.2 (61.7)	8.44 (83.8)	12.1(57.7)
$I / \sigma I$	24.6 (2.1)	36.8 (2.3)	57.9 (3.0)
Completeness (%)[b]	99.0 (96.3)	99.9 (100)	99.7 (100)
Redundancy[b]	3.6 (3.1)	11.6 (11.7)	13.3 (8.9)
No. reflections (test)	73972 (2%, 1551)	28762 (2%, 604)	38810 (2%, 826)
R_{work} / R_{free}	20.6 / 25.3	20.9 / 25.8	21.8/ 24.9
No. atoms			
Protein	6008	2978	2903
DNA	892	323	323
dNTP	50	28	30
Mg^{2+} ions	-	2	2
Ca^{2+} ions[c]	8	-	-
Waters	757	265	238
B-factors			
Protein	62.5	49.6	43.5
DNA	52.7	46.6	44.5
dNTP	29.0	44.8	48.8
Metal ions	35.0	46.1	47.7
Water	39.2	59.9	46.8
R.m.s.d. bond lengths (Å)	0.016	0.013	0.019
R.m.s.d. bond Angles (°)	1.96	1.56	1.82

[a] AU means asymmetric unit
[b] Data in the highest resolution shell are in parentheses
[c] There are five non-catalytic Ca^{2+} in the structure.