Abstract

DNA polymerase ε (Polε) is a multi-subunit polymerase that contributes to genomic stability via its roles in leading strand replication and the repair of damaged DNA. Here we report the ternary structure of the Polε catalytic subunit (Pol2) bound to a nascent G:C base pair (Pol2G:C). Pol2G:C has a typical B-family polymerase fold and embraces the template-primer duplex with the palm, fingers, thumb and exonuclease domains. The overall arrangement of domains is similar to the structure of Pol2T:A reported recently, but there are notable differences in their polymerase and exonuclease active sites. In particular, we observe Ca²⁺ ions at both positions A and B in the polymerase active site and also observe a Ca²⁺ at position B of the exonuclease active site. We find that the contacts to the nascent G:C base pair in the Pol2G:C structure are maintained in the Pol2T:A structure and reflect the comparable fidelity of Pol2 for nascent purine-pyrimidine and pyrimidine-purine base pairs. We note that unlike that of Pol3, the shape of the nascent base pair binding pocket in Pol2 is modulated from the major groove side by the presence of Tyr431. Together with Pol2T:A, our results provide a framework for understanding the structural basis of high fidelity DNA synthesis by Pol2.

Introduction

The bulk of DNA synthesis in eukaryotes is carried out by three polymerases: Polε, δ, and ε [1,2]. Polε primes the Okazaki fragments on the lagging strand, which are then elongated by Polδ. Polε is believed to be the leading strand polymerase and, like Polδ, achieves fidelity via both accurate DNA polymerization and 3′→5′ proofreading exonuclease (Exo) activities. The DNA polymerization (Pol) activities of Polδ and ε achieve an error rate of ~10⁻⁷, which is then further lowered to ~10⁻⁷ by their proofreading functions. DNA mismatch repair achieves another ~100-fold increase in fidelity, for an error rate of ~10⁻⁹ following DNA synthesis. Accurate DNA replication by Polδ and ε is thus crucial in maintaining genome integrity and mutations that lower the fidelity of these polymerases lead to tumor development. Germline and somatic mutations in the exonuclease domains of Polδ and ε are frequently associated with endometrial and colon cancers [3,4,5,6].

Polε, δ, and ε belong to the B-family of DNA polymerases. Crystal structures of the Polδ [7] and Polε [8] catalytic subunits (Pol1 and Pol3, respectively) have been determined and reveal a characteristic B-family polymerase fold comprised of a palm domain that carries the catalytic residues for dNTP addition, a fingers domain that drapes over the nascent base pair, a thumb domain that makes contacts in the DNA minor groove, and an N-terminal domain (NTD). Pol3 also contains an active exonuclease domain, whereas in Pol1 the exonuclease domain is rendered inactive due to mutations. The ternary structure of the yeast Polε catalytic domain (Pol2) was reported recently by Johansson and colleagues [9]. The structure with a nascent T:A base pair (Pol2T:A) was solved by molecular replacement (MR) using the Pol3 structure as a search model. We present here a crystal structure of yeast Pol2 catalytic domain that differs from the Pol2T:A structure in containing a slightly different protein construct (residues 1–1187 versus 1–1228), a different template-primer, a different incoming nucleotide (dCTP versus dATP), a different nascent base pair (G:C versus T:A), and different metals (Ca²⁺ versus Mg²⁺). Also, the protein construct we used contains wild-type residues compared to the Pol2T:A. Domain with mutated residues (D290A/E292A) which renders it exonuclease deficient. We show here that the two structures are very similar in their overall arrangement but differ in the Pol and Exo active sites. Pol2 is the only eukaryotic B-family polymerase for which structures are now available with both G:C and T:A nascent base pairs. Together with the Pol2T:A structure, our results provide structural insights into the high fidelity of Pol2 for nascent Watson-Crick base pairs.
Results

Structure determination

We crystallized the Pol2 catalytic core (residues 1–1187) in ternary complex with a 12-nt/16-nt primer/template presenting G as the templating base, and with dCTP as the incoming nucleotide. To prevent degradation of the DNA by the Pol2 exonuclease activity, we prepared the protein-DNA complex in the presence of Ca\(^{2+}\) (rather than Mg\(^{2+}\)). The cocrystals diffract to 2.8 Å resolution with synchrotron radiation (Argonne National Laboratory) and belong to space group C2 with unit cell dimensions of \(a = 147.29 \text{ Å}, \quad b = 68.48 \text{ Å}, \quad c = 149.08 \text{ Å}, \quad \text{and} \quad \beta = 109.6^\circ\) (Table 1). The space group is the same as that of Pol2\(_{T:A}\) cocrystals and the unit cell dimensions are very similar. Our attempts to determine the structure by molecular replacement (MR) methods using the Pol3 structure as a search model resulted in a satisfactory MR solution. However, the structure could not be refined to produce a final model. Johansson and colleagues were successful in solving the Pol2\(_{T:A}\) structure by MR using the Pol3 structure, possibly because the X-ray data extended to higher resolution (2.2 Å). Using the Pol2\(_{T:A}\) structure, we obtained an MR solution for the Pol2\(_{G:C}\) complex that readily refined to satisfactory agreement factors. The refined model of Pol2\(_{G:C}\) consists of residues 1–1187 (with missing segments 1–29, 215–219, 225–231, 665–677) of Pol2, nucleotides 1–15 of the template, nucleotides 2–12 of the primer, incoming dCTP, 4 Ca\(^{2+}\) ions, 1 Na\(^{+}\) ion, 147 solvent molecules, and 1 molecule of ethylene glycol.

Overall Arrangement

As in the Pol2\(_{T:A}\) structure, the Pol2\(_{G:C}\) catalytic core surrounds the template-primer with the palm, fingers, thumb and exonuclease domains (Figure 1). The palm interacts with the replicative end of the DNA and carries the active site residues (Asp640 and Asp877) for DNA synthesis. The fingers domain is composed of two long anti-parallel \(\alpha\)-helices that drape over the nascent G:C base pair (Figure 1). These \(\alpha\)-helices are longer by about two turns when compared to the helices in fingers domains of Pol1 [7] and Pol3 [8]. The thumb domain interacts primarily with the duplex portion of the template-primer and can be further divided into two subdomains that pack against the palm and exonuclease domains, respectively (Figure 1). The exonuclease domain lies on the opposite side of the DNA as the thumb domain and contains the catalytic residues (Asp290, Gln292 and Asp477) for proofreading activity. The NTD bridges the exonuclease and fingers domains. The role of the NTD in Pol2 function is unclear at present; in Pol3, the NTD has been suggested to bind both RNA and DNA [8]. The unpaired segment of the DNA template strand kinks sharply out of the Pol2 polymerase active site cleft and tracks a path between the fingers and exonuclease domains. The duplex portion of the template-primer has a B-DNA like conformation with average helical twist and rise values of 34.2\(^\circ\) and 3.2 Å, respectively. Overall, the Pol2\(_{T:A}\) and Pol2\(_{G:C}\) structures are very similar with the enzymes superimposing with an rmsd of \(\sim 0.38\) Å (for 1001 Cαs) (Figure S1).

An Extended Palm Domain

The palm domain is larger and more elaborate than that in Pol1 and Pol3 (Figure 2). In particular, the palm domain contains several insertions that coalesce to define three new subdomains (A, B and C) that extend outward from the \(\alpha/\beta\) core (Figure 2). Subdomain A is delineated by residues 533–555 and 682–760 and has been named the P domain by Johansson and colleagues [9]. The P domain (or subdomain A) consists of a three-stranded \(\beta\)-sheet capped by two \(\alpha\)-helices which drape over the nascent G:C duplex. The palm domain is larger and more elaborate than that in Pol1 and Pol3 (Figure 2). In particular, the palm domain contains several insertions that coalesce to define three new subdomains (A, B and C) that extend outward from the \(\alpha/\beta\) core (Figure 2). Subdomain A is delineated by residues 533–555 and 682–760 and has been named the P domain by Johansson and colleagues [9]. The P domain (or subdomain A) consists of a three-stranded \(\beta\)-sheet capped by two \(\alpha\)-helices which drape over the nascent G:C duplex.

Table 1. Data collection and refinement statistics.

| Data collection statistics | Pol2\(_{T:A}\) | Pol2\(_{G:C}\) |
|---------------------------|-------------|-------------|
| Number of reflections     | 34031       | 127525      |
| Completeness (%)          | 98.9 (83.9) |             |

A single crystal was used or solving this structure. Values for outermost shells are given in parentheses. (doi:10.1371/journal.pone.0094835.t001)
Basis of High Fidelity

Pol2’s fidelity for a nascent Watson-Crick (W-C) base pair is determined primarily by residues Val825, Asn828, Ser829, Tyr831 and Gly832 from the fingers domain, and by Tyr645 from the palm domain (Figure 3a). Most of the contacts to the nascent G:C base pair occur from atop (Asn828 and Ser829) or from the minor groove side (Tyr 645, Tyr831, and Gly832). These contacts are primarily van der Waals (vdW) in nature and are also maintained in the Pol2_{T:A} structure. For example, the vdW contacts to the minor groove acceptors N3 and O2 of the nascent G:C base pair in the Pol2_{G:C} structure are simply switched to O2 and N3 atoms of the T:A base pair Pol2_{T:A} structure (Figure 3a). Conservation of base-pair interactions in the Pol2_{T:A} and Pol2_{G:C} structure reflects the accuracy of Pol2 for pyrimidine-purine and purine-pyrimidine base-pairs.

Unique to Pol2, is the presence of Tyr431 in the major groove of the nascent W-C base pair. Tyr431 lies in a loop (residues 430-439) of the exonuclease domain and approaches the incoming dCTP from the major groove side with its hydroxyl group located ~4.0 Å from the dCTP N4 atom (Figure 3a). In the structure of Pol3, the major grove is devoid of contacts as the corresponding amino acid Lys473 is disordered and lies >7 Å away from the nascent base pair (Figure S2).

Polymerase and Exonuclease Active sites

The Pol2 polymerase and exonuclease active site are separated by ~41 Å in a direction roughly perpendicular to the DNA axis (Figure 1). The polymerase active site is characterized by acidic residues Asp640 and Asp877 and two calcium ions (A and B) (Figure 3b). Ca^{2+} A and B are separated by ~3.6 Å and are analogous to metals “A” and “B” in other DNA polymerases [11,12,13,14]. Ca^{2+} A is more mobile than Ca^{2+} B (B-factor of 68 Å^2 versus 49 Å^2). Although calcium inhibits Pol2 activity, the active site geometry is appropriate for the two-metal mechanism of catalysis [15] with the putative 3’OH located ~3.8 Å from the dCTP γ-phosphate and aligned with respect to the P_{2}’-O_{3} bond (angle of about 148°). Metals A and B are in a position to activate the primer 3’OH for its nucleophilic attack on the dNTP γ-phosphate and to stabilize the pentacovalent transition state.

The exonucleolytic reaction in B-family polymerases is also believed to proceed by a two-metal mechanism [16], with metals occupying sites A and B in the exonuclease domain [17,18,19]. Since the exonuclease catalytic residues (Asp290 and Glu292) were mutated in the Pol2_{T:A} structure there are no bound metal ions [9]. By contrast, we observe Ca^{2+} at site B, coordinated by residue Asp290 (Figure 3c). We also observe strong electron density next to this Ca^{2+}, suggestive of a second Ca^{2+} ion bound at site A (Figure S3). However, the distance between the two putative calcium ions would then be ~3.2 Å, which is shorter than the typical distance of ~3.7–4.0 Å. Given this uncertainty, we have assigned this density at site A as a water molecule, though refining it as such leads to substantial positive electron density in an F_o-F_c map (indicative of a more electron-rich atom) (Figure S3).

Discussion

Pol2_{G:C} and Pol2_{T:A} structures are very similar, including regions of the enzyme and DNA that are visible in the electron density map. Importantly, contacts to the nascent G:C and T:A base pairs are interchangeable and reflect the roughly equal fidelity of Pol2 for Pu:Py or Py:Pu nascent base pairs (Figure S4). A notable difference between the two structures is the number of metals in the polymerase active site. Johansson and colleagues crystallized Pol2_{T:A} in the presence of Mg^{2+}; to prevent
degradation of the DNA by the Pol2 exonuclease activity residues Asp290 and Glu292 were mutated to alanines. The Pol2 T:A structure shows a single Mg$^{2+}$ ion in the polymerase active site at position B (Figure 3b). By contrast, we cocrystallized Pol2 G:C with the wild-type enzyme; to prevent exonucleolytic degradation of DNA we used Ca$^{2+}$ in place of Mg$^{2+}$. The structure reveals Ca$^{2+}$ ions at positions A and B in the polymerase active site. Typically, metal A in DNA polymerases is coordinated by the α-phosphate of the incoming nucleotide, the putative primer 3’OH, the carboxylates of active site residues, and water molecules. It tends to be more mobile than metal B with longer ligation distances and is often not observed in DNA polymerases. For example, the structures of Polk show only a single Mg$^{2+}$ at position B [20]. Also, the structure of Pol with incoming dTTP showed a single Mg$^{2+}$ at position B [21], but a later structure with incoming dCTP showed a second Mg$^{2+}$ at position A [22]. The presence of Ca$^{2+}$ at position A in the Pol2 G:C structure shows that the Pol2 active site is fully capable of binding a second metal. The absence of metal A in the Pol2 T:A structure reflects its intrinsic mobility (compared to metal B) and the lack of a 3’OH ligand at the primer terminus.

A unique feature of Pol2 is the presence of Tyr431 in the major groove of the nascent base pair binding pocket (Figure 3a and S2). In both Pol2 and Pol3, the binding pocket is primarily shaped by residues from the palm and fingers domain which are conserved in all B-family polymerases [8]. The interactions of these residues with the nascent base pair occur from the top or the minor groove.
side and the binding pocket is devoid of interactions in the major groove. The binding pocket of Pol2 is also shaped by Tyr431 from the exonuclease domain that approaches the nascent base pair from the major groove side. Figure S2a shows the structures of Pol2 and Pol3 superimposed by their palm domains. Relative to that of Pol2, the entire exonuclease domain of Pol3 is shifted up and away from the major groove by ≥5 Å. This results in Lys473 of Pol3 (which is equivalent to Tyr431 of Pol2) being positioned >7 Å away from the incoming nucleotide. This may lend to differences in base substitution errors between Pol2 and Pol3 [23,24]. A better understanding of the role of Tyr431 in the fidelity of Pol2 would require structures of wild type and mutant Pol2 with different mismatches. Interestingly, if the exonuclease domain of Pol3 were in the same relative orientation as that of Pol2 (Figure S2b), a ‘β-hairpin’ from its exonuclease domain would collide with the unpaired segment of the template strand. An analogous hairpin in RB69 and T4 Pol has been proposed to facilitate strand separation and the transition of the primer strand between the polymerase and exonuclease sites [25,26,27]. In Pol2, this β-hairpin is much smaller but, surprisingly, does not appear to limit the ability of Pol2 to proofread insertion errors [23,24].

An intriguing feature of Pol2 is the putative metal binding motif in the palm domain, characterized by three conserved cysteines (Cys665, Cys677 and Cys763). Based on spectroscopic and other data, we have shown that these cysteines bind a Fe-S cluster [10]. For example, wild-type Pol2 catalytic core is found to be yellowish-brown in color, but a mutant in which Cys665, Cys677 and Cys763 are mutated is colorless. We also showed that the Cys triple mutant is deficient in DNA polymerase activity but not in the exonuclease activity. This is consistent with the location of the cysteines on palm domain, remote from the exonuclease domain. Considering its functional importance, it is surprising therefore that the cysteine-rich metal binding motif is partially disordered in both the Pol2G:C and Pol2TA structures. Johansson and colleagues positioned a Zn 2+ ion in the Pol2TA structure, coordinated to Cys667 and Cys763 and to partially disordered Cys665 and Cys668. We suspect that the disorder in Cys665 and Cys668 likely reflects the binding of sub-optimal Zn 2+ rather than a Fe-S cluster. Fe-S clusters are labile and can be substituted by Zn 2+ and it is quite possible that the Pol2 form that crystallizes in both Pol2TA and Pol2G:C structures contains Zn 2+ instead of a functional Fe-S cluster. It will be interesting to grow the crystals of Pol2 under anaerobic conditions to see whether a Fe-S cluster will replace Zn 2+ and whether this lead to the ordering of the cysteines.

**Methods**

**Protein and DNA preparation**

The catalytic core of S. cerevisiae Pol2 (residues 1–1187) harboring a N-terminal GST tag was expressed in the protease deficient yeast strain YRP654. The GST tag was engineered to be cleaved with PreScission protease. Protein was purified by affinity chromatography with Glutathione Sepharose 4B beads, removal cleaved with PreScission protease. Protein was purified by affinity chromatography with Glutathione Sepharose 4B beads, removal cleaved with PreScission protease. Protein was purified by anion exchange on a MonoQ column, desalted and lyophilized before crystallization. Purified 12-nt primer harboring Fe-S clusters are labile and can be substituted by Zn 2+. For example, wild-type Pol2 catalytic core is found to be yellowish-brown in color, but a mutant in which Cys665, Cys677 and Cys763 are mutated is colorless. We also showed that the Cys triple mutant is deficient in DNA polymerase activity but not in the exonuclease activity. This is consistent with the location of the cysteines on palm domain, remote from the exonuclease domain. Considering its functional importance, it is surprising therefore that the cysteine-rich metal binding motif is partially disordered in both the Pol2G:C and Pol2TA structures. Johansson and colleagues positioned a Zn 2+ ion in the Pol2TA structure, coordinated to Cys667 and Cys763 and to partially disordered Cys665 and Cys668. We suspect that the disorder in Cys665 and Cys668 likely reflects the binding of sub-optimal Zn 2+ rather than a Fe-S cluster. Fe-S clusters are labile and can be substituted by Zn 2+ and it is quite possible that the Pol2 form that crystallizes in both Pol2TA and Pol2G:C structures contains Zn 2+ instead of a functional Fe-S cluster. It will be interesting to grow the crystals of Pol2 under anaerobic conditions to see whether a Fe-S cluster will replace Zn 2+ and whether this lead to the ordering of the cysteines.

**Cocrystallization**

The Pol2G:C ternary complex was prepared by mixing purified Pol2 and the 12/16 template-primer DNA duplex in the ratio of 1:1, followed by the addition of dCTP and CaCl 2 to final concentrations of 10 mM each. The ternary complex was crystallized from solution containing 10–15% polyethylene glycol 5000 monomethyl ether and 25 mM magnesium acetate, 1% DMSO in 0.1 M Tris-HCl buffer (pH = 7.0). For data collection, crystals were cryoprotected by stepwise soaks in mother liquor solutions containing 3–25% ethylene glycol and then flash frozen in liquid nitrogen. X-ray data on cryocooled crystals were measured at Advanced Photon Source (APS, beamline 23-ID) of Argonne National Laboratory at a wavelength of 1.0332 Å. Data sets were indexed and integrated using the HKL-2000 package [28]. Crystals diffract to 2.8 Å and belong to space group C2 with unit cell dimensions of a = 147.29 Å, b = 68.48 Å, c = 149.08 Å and α = β = 90°, γ = 109.6°. Matthew’s coefficient suggested one protein molecule in the asymmetric unit.

**Structure determination and refinement**

The structure of Pol2G:C was solved by molecular replacement (MR), using the Pol2TA complex as a search model (with the DNA, incoming nucleotide, metal ions and water molecules omitted). The program Phaser [29] gave a unique MR solution. The first round of refinement and map calculation was carried out without the DNA using the program PHENIX [30]. The electron density maps (2Fo-Fc and Fo-Fc) showed unambiguous densities for the DNA and incoming nucleotide, which were then included in the model for subsequent refinement. Iterative rounds of refinement and water picking were performed with PHENIX and model building with program Coot [31]. The final model has good stereochemistry as shown by MolProbity [32] with >99.4% of all residues in allowed regions of the Ramachandran plot and 0.6% in the disallowed regions. Final coordinates have been submitted to the Protein Data Bank with PDB ID 4PTF. Figures were prepared using PyMol [33].

**Structural analysis**

Protein structures were aligned and superimposed using MUS-TANG [34] and LSQMAN [35]. Web 3DNA (w3dna.rutgers.edu) [36] was used for analysis of DNA helical parameters.

**Supporting Information**

**Figure S1** Superimposition of the structures of Pol2G:C (cyan) and Pol2TA (red). Overall, the structures are very similar.

**Figure S2** Major groove interactions in the nascent base pair binding pocket of Pol2G:C (cyan) and Pol3G:C (red).

(a) Superimposition of Pol2G:C and Pol3G:C by their palm domains. Pol2 residues 528:767 and 844:989 were aligned with Pol3 amino acids 577:660 and 713:834. Tyr431 of the Pol2 exonuclease domain approaches the nascent base pair from the major groove side. Compared to Pol2, the Pol3 exonuclease domain is shifted away from the major groove by ≥5 Å. (b) Superimposition of the exonuclease domains. If the Pol3 exonuclease domain (residues 316:531) were in the same relative orientation as the Pol2 exonuclease domain (residues 284:501), a β-hairpin (labeled above) would collide with the unpaired segment of the template strand. This β-hairpin has been implicated in aiding the transition of the primer strand between the polymerase and exonuclease...
active sites. The Pol2 β-hairpin is much smaller and does not interact with the DNA.

(PDF)

Figure S3 Residual F₀-Fᵳ density (green, 3σ) in the Pol2-GcC exonuclease active site with position A modeled as a water molecule. This is suggestive of a more electron rich atom (possible Ca²⁺) in the vicinity of position A.

(PDF)

Figure S4 Schematic of protein-DNA interactions. Amino acids from Pol2 palm, fingers, thumb, exonuclease and N-terminal domains are shown in cyan, yellow, orange, magenta and blue respectively; incoming dCTP is shown in red. A distance cut-off of 3.35 Å was used for protein-DNA interactions. Residues R744, R749 and R751 from subdomain A are in the vicinity of the DNA but at a distance larger than 3.5 Å (not shown). Figure was generated with NUCPLOT (Luscombe N M, Laskowski R A, Thornton J M (1997). NUCPLOT: a program to generate schematic diagrams of protein-DNA interactions. Nucleic Acids Res., 25, 4940-4945) and modified for clarity.

Acknowledgments

We thank the staff at Advanced Photon Source (beamline 23-ID) for facilitating X-ray data collection.

Author Contributions

Conceived and designed the experiments: AKA RJ. Performed the experiments: RJ KRR AB REJ. Analyzed the data: RJ KRR AKA. Contributed reagents/materials/analysis tools: REJ LP SP. Wrote the paper: RJ AKA.