Nucleic Acids Research, 2018 **1**
doi: 10.1093/nar/gky402

NAR Breakthrough Article

Pyrophosphate hydrolysis is an intrinsic and critical step of the DNA synthesis reaction

Jithesh Kottur and Deepak T. Nair*

Regional Centre for Biotechnology, NCR Biotech Science Cluster, 3rd Milestone, Faridabad-Gurgaon Expressway, Faridabad 121 001, India

Received February 26, 2018; Revised April 20, 2018; Editorial Decision April 24, 2018; Accepted May 15, 2018

ABSTRACT

DNA synthesis by DNA polymerases (dPols) is central to duplication and maintenance of the genome in all living organisms. dPols catalyze the formation of a phosphodiester bond between the incoming deoxynucleoside triphosphate and the terminal primer nucleotide with the release of a pyrophosphate (PPi) group. It is believed that formation of the phosphodiester bond is an endergonic reaction and PPi has to be hydrolyzed by accompanying pyrophosphatase enzymes to ensure that the free energy change of the DNA synthesis reaction is negative and it can proceed in the forward direction. The fact that DNA synthesis proceeds *in vitro* in the absence of pyrophosphatases represents a long-standing conundrum regarding the thermodynamics of the DNA synthesis reaction. Using time-resolved crystallography, we show that hydrolysis of PPi is an intrinsic and critical step of the DNA synthesis reaction catalyzed by dPols. The hydrolysis of PPi occurs after the formation of the phosphodiester bond and ensures that the DNA synthesis reaction is energetically favorable without the need for additional enzymes. Also, we observe that DNA synthesis is a two Mg^{2+} ion assisted stepwise associative S_N2 reaction. Overall, this study provides deep temporal insight regarding the primary enzymatic reaction responsible for genome duplication.

INTRODUCTION

In all living organisms, deoxyribonucleic acid (DNA) is synthesized by DNA polymerases (dPols), and these enzymes play a central role in genome duplication. Also, dPols are essential for many applications in research and biotechnology such as DNA sequencing, gene cloning, and Polymerase Chain Reaction-based diagnostic kits. In both natural and artificial settings, dPols catalyze the template-directed addition of deoxynucleoside triphosphates (dNTPs) to the primer strand (1–5). Mg^{2+} ions play a critical role in the polymerization reaction, and dPols extends the primer in 5′ to 3′ direction. The formation of a phosphodiester bond (pB bond) between the α-phosphate of the incoming dNTP and the 3′- hydroxyl group of the terminal primer nucleotide represents the primary chemical reaction catalyzed by dPols. This reaction also involves breakage of the bond between the α-phosphate, and the bridging oxygen (αO bond) between the α- and β- phosphates resulting in the release of the pyrophosphate (PPi) moiety as a by-product of the synthesis reaction.

The reaction involving incorporation of a nucleotide into the primer through the formation of the phosphodiester bond along with the release of PPi is expected to have a low free energy change (1,6,7). To ensure that the reaction moves in the forward direction, it is believed that the PPi moiety is cleaved by an accompanying pyrophosphatase enzyme to render a large negative free energy change (∼7 kcal/mol) to the DNA synthesis reaction (1,8–10). However, DNA synthesis can proceed smoothly *in vitro* in the absence of any pyrophosphatase enzyme, and there is no satisfactory explanation available for this long-standing thermodynamic conundrum.

Based on structural and biochemical studies, two distinct mechanisms have been proposed for the synthesis of the pB bond by dPols. In the classical two-metal mechanism, three conserved acidic residues in the palm subdomain bind two Mg^{2+} ions (11). Structures of dPols in complex with DNA and dNTP showed that one Mg^{2+} (Metal B) coordinates two active-site acidic residues and the triphosphate moiety of the incoming nucleotide (12–17). The other Mg^{2+} (Metal A) coordinates all three active-site acidic residues and 3′OH of the primer terminus (13,18,19) and is thought to facilitate the formation of the attacking oxyanion by lowering the

*To whom correspondence should be addressed. Tel: +91 124 2848844; Email: deepak@rcb.res.in

pK_a of the primer terminus 3′O–. The in-line nucleophilic attack by the 3′O– oxyanion on the α-phosphate should lead to the formation of a penta-coordinated bipyramidal transition state. The metal B may stabilize the developing negative charge on the triphosphate moiety during the nucleophilic attack. According to this model, the reaction is hypothesized to be a concerted S_N2-type reaction wherein there are simultaneous formation and breakage of the pB and αO bonds, respectively.

In the recent past, the appearance of an additional divalent metal ion (Metal C) which is distinct from the catalytic (Metal A) and nucleotide (Metal B) metal ions has been reported (19–21). The metal ion (Metal C) coordinates the non-bridging oxygens on the phosphates of what previously were α- and β-phosphates of the incoming nucleotide and is predicted to assist breakage of the αO bond. According to this model, the binding of metal C is the rate-limiting step of the DNA polymerase reaction (21,22). Both the classical two-metal-ion and the new three-metal-ion mechanisms propose a concerted S_N2-type reaction scheme. However, based on theoretical analysis, it has been suggested earlier that DNA synthesis by dPols may follow either a concerted, stepwise associative or a stepwise dissociative reaction scheme (23,24).

To elucidate the different stages of the DNA synthesis reaction, we have employed DNA polymerase IV (PolIV) from *Escherichia coli*. PolIV is a member of the Y-family of dPols and capable of template-dependent dNTP incorporation (18,25,26). We have conducted time-resolved crystallography on DNA synthesis by PolIV and obtained periodic snapshots of the reaction. Our studies show that hydrolysis of the pyrophosphate moiety is an inherent part of the DNA synthesis reaction catalyzed by dPols. This step ensures that the nucleotide incorporation reaction is energetically favorable without the need for accompanying pyrophosphatases. Also, we observed that the DNA synthesis is a two Mg^{2+} ion assisted stepwise associative S_N2 reaction wherein the formation of the pB bond and dissolution of the αO bond occur sequentially. Overall, the study provides insight regarding the mechanism utilized by dPols to synthesize DNA during DNA replication.

MATERIALS AND METHODS

Protein purification and primer extension assays

PolIV was purified as mentioned before (18,27). Primer extension assays were carried out at different pH and incubation times to unearth conditions wherein the nucleotide addition reaction is considerably slowed down. The primer extension assays were carried out with the following DNA substrate:

1. Template A

```
3′-GCATGAGCATCCGTAATCACACTGGTCGACAAGTCCATC
  CGTGCCATCCT-5′
5′-XCGTACTCGTAGGCAT-3′
```

The concentrations of protein, DNA and incoming nucleotide used were 100 nM, 100 nM and 25 μM, respectively and the reaction products were processed as mentioned previously (28). It was seen that at pH 5.2, the rate of reaction was reduced by >60-fold at 37°C (Supplementary Figure S1).

In crystallo DNA synthesis

An 18mer self-annealing oligonucleotide was used for crystallization and has the following sequence: 5′ TCTAGGGT CCTAGGACCC 3′. The ternary complexes were reconstituted by mixing PolIV (0.3 mM) with dsDNA (0.36 mM) followed by addition of 5 mM dTTP (GE Healthcare) or 5 mM dTMPPnP (Jena Biosciences). Crystallization was carried out in the absence of metal ion and crystals were obtained in 0.1 M acetate (pH 5.2) and 5–12% MPD. The *in crystallo* reaction was initiated by transferring crystals to a cryosolution containing 0.1 M acetate (pH 5.2) and 30% MPD with 5 mM $MgCl_2$. The crystals were incubated for different time periods at 4°C (to further reduce the rate of the nucleotide incorporation reaction) and then flash frozen in liquid nitrogen to stop the reaction. Subsequently, X-ray diffraction data were collected from the frozen crystals. The protocol for preparing crystals utilized here is distinct from that employed in previous time-resolved crystallography studies carried out on DNA polymerases β and η, wherein crystals were grown in the presence of Ca^{2+} first and then incubated with high concentrations (∼100 mM) of Mg^{2+} or with Mn^{2+}.

Structure determination and crystallographic refinement

For the crystals prepared with dTTP, X-ray diffraction data were collected at the BM14 beamline of ESRF. Data were processed using iMosflm and the SCALA program in CCP4 (29,30). Using PHASER, the structures corresponding to different time points were determined using the structure of $PolIV_{dA:dTMPnPP}$ (4IR1) as a search model (18,31). Before molecular replacement, the primer nucleotide, incoming nucleotide, and metal ions were removed from the coordinate file of the search model. The MR solutions were subjected to three cycles of refinement in PHENIX, and $F_o - F_c$ maps were calculated for the resolution range of 2.15–40Å for all datasets and viewed at 3.3σ for comparison (Figure 2). The electron density maps (omit $F_o - F_c$) prepared at the highest resolution possible for each dataset (Supplementary Figure S5) were nearly identical to the previous ones (Figure 2) and in agreement with the all the major conclusions of the analysis. The terminal primer nucleotide and incoming dTTP were positioned at appropriate locations in the electron density maps using Coot, and the structures were refined until convergence using PHENIX (32). For the crystal prepared with dTTP and incubated with 50 mM $MgCl_2$ for 30 min (Supplementary Figure S6), the data were collected at the BM14 beamline of ESRF, and were processed using iMosflm and Scala. Structure solution and refinement were carried out using the same protocol used for structures obtained using 5 mM $MgCl_2$.

For the crystals obtained with dTMPPnP (Figure 5), diffraction data were collected at the ID29 beamline of ESRF and data was processed using XDS, Aimless and Pointless (33–36). Structure solution and refinement were carried out using the same protocol used for the structures with dTTP.

For all the refined structures, the majority of the residues were in the favorable regions of the Ramachandran plot with only 1% of residues in the disallowed regions. All the refined structures show the presence of two complexes of PolIV:DNA:dTTP:Mg^{2+} which, based on the chain IDs of the components, are called ABC and FGH. The electron density maps were clearer for the FGH complex, and hence this complex was selected for analysis. The electron density maps were prepared in PHENIX and viewed using Coot (37). The structures were also analyzed using PyMOL (Schrodinger Inc.), and all the figures were generated using PyMOL.

Primer extension assay with modified nucleotides

The modified deoxynucleoside triphosphates dTMPnPP, dTMPPnP, dGMPnPP and dGMPPnP were obtained from Jena Biosciences. Primer extension assay was conducted with different dPols (PolIV, Dbh, Dpo4, MsDpo4, Taq, Klenow exo-, PfPrex, Pol II and M-MuLV RT) and the substrate DNAs utilized were as follows:

Template A

```
3′-GCATGAGCATCCGTAATCACACTGGTCGACAAGTCCATC
 CGTGCCATCCT-5′
5′-XCGTACTCGTAGGCAT-3′
```

Template C

```
3′-GCATGAGCATCCGTACTCACACTGGTCGACAAGTCCATC
 CGTGCCATCCT-5′
5′-XCGTACTCGTAGGCAT-3′
```

where X = 5′ 6FAM label.

Protein, DNA and incoming nucleotide (normal or modified) were added to a final concentration of 10 nM, 50 nM and 2 μM, respectively. The reaction mixture was incubated for 30 min at 37°C, and the reaction products were processed as mentioned previously (28).

Phosphate and pyrophosphate assay

To quantitate the level of phosphates and pyrophosphates generated during the DNA synthesis reaction, primer extension assays were carried out initially, as mentioned previously. The following DNA substrate was utilized:

Template A

```
3′-GCATGAGCATCCGTAATCACACTGGTCGACAAGTCCATC
 CGTGCCATCCT-5′
5′-GTACTCGTAGGCAT-3′
```

The reaction mixture included 100 nM of PolIV, 15 μM DNA duplex and 2 mM of all four dNTPs. After incubation at 37°C for 1 h, the reaction was terminated by adding EDTA solution to a final concentration of 20 mM. The amount of phosphate in the reaction mix was determined using the Phosphate Assay Kit (Sigma-Aldrich) wherein the sensor reagent is a proprietary formulation of the Malachite green dye. 100 μl of this reagent was added to 50 μl of the reaction mix followed by incubation at room temperature for 30 min, and subsequently, absorbance was recorded at 630 nm on a Spectramax M5 plate reader (Molecular Devices). The amount of pyrophosphate was determined using the Pyrophosphate assay kit (MAK168, Sigma) wherein the sensor reagent is a fluorogenic PPi sensor. 50 μl of the reaction mix was added to 50 μl of sensor reagent followed by 30 min of incubation at room temperature. After incubation, the fluorescence intensity was measured (λ_{ex} = 320 nm and λ_{em} = 456 nm) on a Spectramax M5 plate reader (Molecular Devices). To enable accurate quantitation, initially standard curves were plotted for both Pi and PPi. All measurements were done in triplicate and with appropriate controls.

RESULTS

Entry of Mg^{2+} ions triggers conformational changes

Primer extension assays conducted under different reaction conditions shows that the rate of the polymerization reaction was slowed down considerably at a pH of 5.2 (Supplementary Figure S1). Crystals of the PolIV:DNA:dTTP complex were grown at this pH in the absence of Mg^{2+} ions. The crystals were soaked in Mg^{2+} for 0, 1, 3, 5, 7, 10, 15, 20, 25, 30, 35, 40, 50 and 60 min at 4°C and then frozen. X-ray diffraction data were collected from the frozen crystals, and the data collection and refinement statistics are displayed in Supplementary Table S1.

Alignment of the ground state structure (0 min), where no metal ions are present, with the complex obtained 1 min after addition of the Mg^{2+} ions, showed only marginal differences. The incoming nucleotide, therefore, attains the conformation compatible with productive catalysis in the absence of the co-factor ion (Supplementary Figure S2). The density for both the metal ions appeared 1min after the addition of Mg^{2+} ions and the activation of the 3′-hydroxyl group was observed immediately after the metal ions occupied their respective locations. The sugar pucker of the terminal nucleotide of the primer changed from C2′-endo to C3′-endo (Supplementary Figure S3). Also, the incoming nucleotide moved towards the 3′OH of the primer terminus. As a result of these changes, the distance between the 3′O– of the terminal primer nucleotide and the α-phosphate of the incoming dTTP reduced from 4.6 to 3.5 Å (Figure 1). Furthermore, the angle formed by the 3′O–, the α-phosphate and the bridging oxygen between α- and β- phosphates increased from 150° to ∼170° (Figure 1). The conformational changes triggered by binding of Mg^{2+} ions, therefore facilitate the in-line nucleophilic attack by the 3′O– of the terminal primer nucleotide on the α-phosphate of the incoming dTTP.

Phosphodiester bond formation is a stepwise associative S_N2 reaction

From 5min onwards, electron density developed gradually between the 3′-end of the primer and α-phosphate of the incoming dTTP. By 10 min, the pB bond was completely formed resulting in the formation a pentavalent intermediate. In this intermediate, the α-phosphate is simultaneously bonded to the 3′O– of the primer terminus, the bridging oxygen bonded with the β-phosphate, the bridging oxygen bonded with 5′ carbon of the deoxyribose sugar and two oxygen atoms (Supplementary Figure S4). After 10 min, we observed that there is a gradual decrease in the density of the

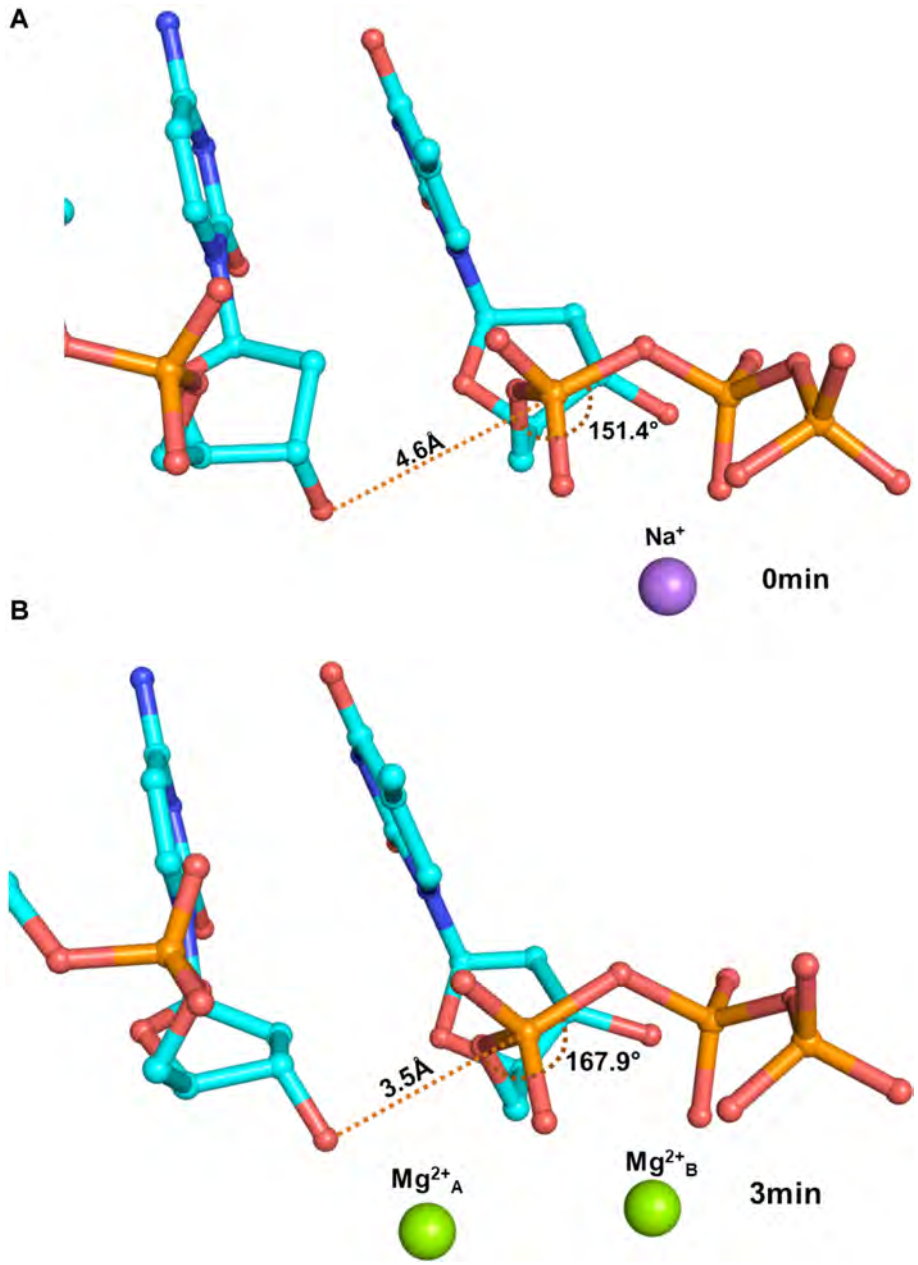

Figure 1. Conformational changes triggered by the Mg^{2+} ions. The terminal primer nucleotide and the incoming dTTP are displayed here in stick representation and coloured by element. Upon entry of two Mg^{2+} ions, the terminal primer nucleotide and the incoming dTTP undergo conformational changes such that the distance between the 3′O– and the α-phosphate decreases from 4.6 Å (0 min) to 3.5 Å (3 min) and the angle between the 3′O–, the β-phosphate and the bridging oxygen between the α- and β- phosphates increases from 150° to 170°. The conformational changes, therefore, facilitate the in-line nucleophilic attack by the 3′O– on the α-phosphate.

αO bond (Figure 2, Supplementary Figure S5). The analysis of the omit $F_o - F_c$ maps showed that the pB bond formation and αO bond dissolution happen sequentially and not concomitantly. The DNA synthesis reaction, therefore, follows the stepwise associative S_N2 scheme (23). Further, the DNA synthesis reaction may also represent an example of the alternate two-step mechanistic model of S_N2(P5) reactions proposed recently by Kolodiazhnyi and Kolodiazhna (38). In this alternate model, the pentavalent intermediate exists in a lower-energy state between two higher-energy transition states corresponding to, in this case, the formation of pB bond and dissolution of the αO bond.

Unlike postulated by the classical two-metal-ion mechanism, the DNA synthesis reaction does not follow a concerted S_N2 reaction scheme as the pB bond formed first followed by dissolution of the αO bond (Movie S1) (11). Additionally, there was no electron density present in the appropriate location for a third Mg^{2+} (Metal C) ion in the maps as observed in the case of DNA polymerases β and η. Hence, our studies show that two Mg^{2+} ions are adequate to enable synthesis of the pB bond (Figure 2 & Supplementary Fig-

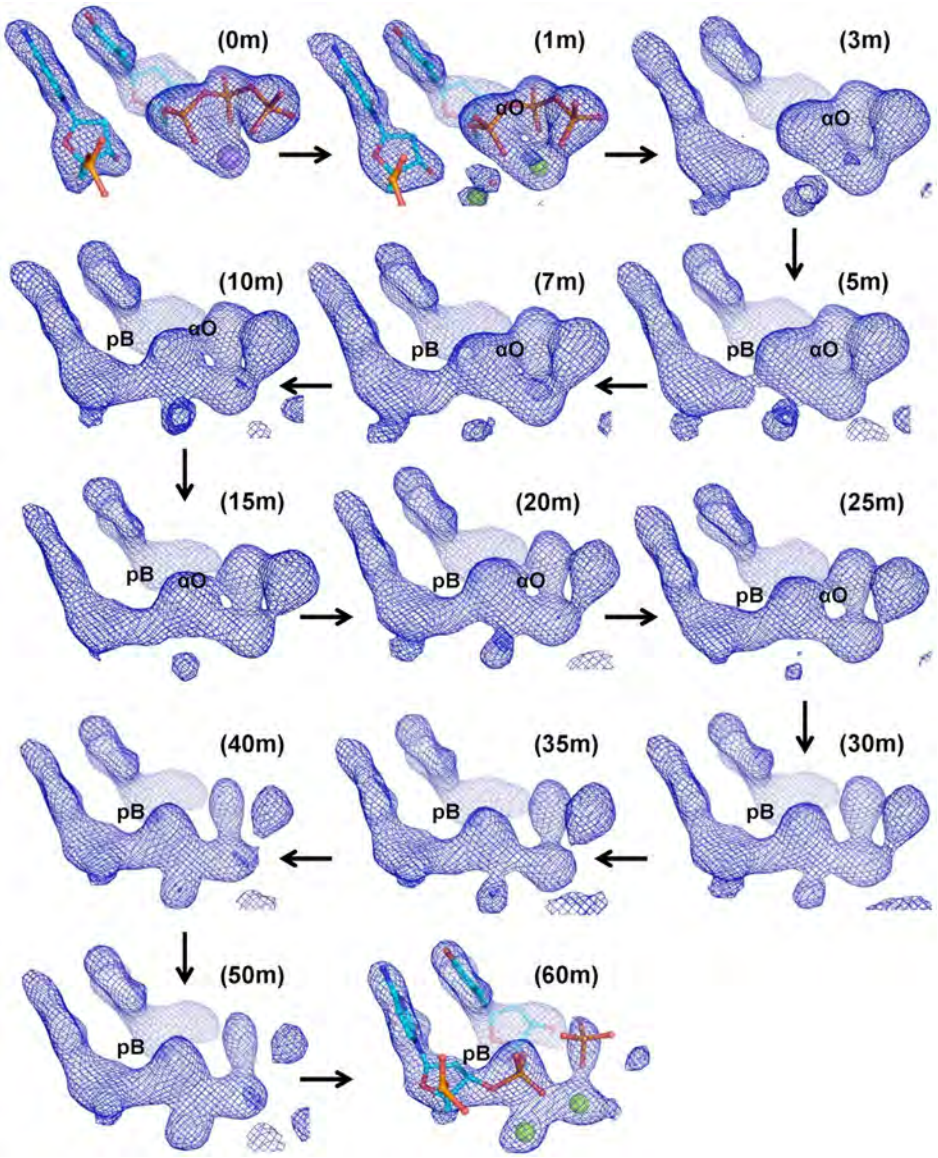

Figure 2. Electron density maps depicting the sequence of events during phosphodiester (pB) bond formation: Omit electron density maps ($F_o - F_c$) calculated to a resolution of 2.15 Å are displayed at a contour of 3.3. These maps show that the electron density corresponding to formation of pB appears first (5, 7 and 10 m) followed by dissolution of αO (15, 20 and 25m) and subsequent hydrolysis of the PPi moiety to phosphate ions (30, 35, 40, 50 and 60 m). The terminal primer nucleotide and the incoming dTTP are displayed here in stick representation and colored by element. Na^+ and Mg^{2+} ions are shown as purple and green spheres respectively.

ure S5). This observation gave rise to the possibility that the third metal ion observed in earlier studies may be due to the use of Mn^{2+} or high concentrations of Mg^{2+}. The electron density maps for a crystal incubated with 50 mM $MgCl_2$ for 30 min showed the presence of density corresponding to the third Mg^{2+} (Metal C) ion (Supplementary Figure S6). Hence, the third metal ion may play no actual part in DNA synthesis.

Also, the time-resolved experiment clearly shows that a pentavalent intermediate is formed and decomposes gradually during the DNA synthesis reaction (Figure 2 & Supplementary Figure S5). The selection of the multivalent phosphorous element to form the backbone of genetic material may in part be due to the need to form this transient pentavalent intermediate during formation of the DNA polymer (39).

Pyrophosphate hydrolysis is critical for DNA synthesis

The products of the nucleotide incorporation reaction are the extended primer strand which has increased in length by one nucleotide, and the PPi moiety from the dNTP. Surprisingly, we observed an additional step in the reaction wherein there is hydrolysis of the PPi moiety to generate two individual phosphate ions- βPi and γPi corresponding to the β- and γ-phosphate of dTTP, respectively (Figure 2 & Movie S1). The Arg49 residue involved in stabilizing the PPi in the active site moved away after breakdown of PPi (Supplementary Figure S7). Consistent with the observed importance of

the R49 residue, the R49→A mutant protein lost the ability to catalyze the DNA synthesis reaction (Supplementary Figure S8). After hydrolysis of the PPi moiety, γPi diffused out first, and the two metal ions were retained in the complex. The mechanism of PPi release by dPols has been the subject of scrutiny in the past, and our study shows that this happens through the hydrolysis of PPi into free phosphates (40,41).

To confirm the importance of breakdown of the PPi in the completion of the polymerization reaction, we conducted primer extension assays using the modified nucleotides dTMPPnP and dGMPPnP, wherein the bond between the β- and γ-phosphates is non-hydrolyzable. If PPi is the final by-product of the reaction, the polymerases should incorporate dTMPPnP and dGMPPnP opposite template dA and dC respectively, as there is no modification in the bond between the α- and β-phosphates. The dPols tested include PolIV (*E. coli*), Dpo4 (*Sulfolobus solfataricus*), Dbh (*Sulfolobus acidocaldarius*), MsDpo4 (*Mycobacterium smegmatis*), the Klenow fragment of DNA polymerase I (*E. coli*), the polymerase module of the Pfprex protein (*Plasmodium falciparum*), DNA polymerase II (*E. coli*) and M-MuLV RT (Moloney Murine Leukemia Virus). PolIV, Dpo4, Dbh and MsDpo4 belong to the Y-family of dPols, the Klenow fragment and Pfprex are representatives of the A-family, DNA Polymerase II is a member of the B-family, and M-MuLV RT belongs to the Reverse Transcriptase family. All the tested dPols failed to incorporate the modified nucleotides (Figures 3 and 4). These studies show that the hydrolysis of PPi for completion of the synthesis reaction may be a conserved feature of all dPols.

Biochemical assays that assess the amount of inorganic phosphate (Pi) and PPi generated during the dNTP incorporation reaction showed that the level of Pi produced is much higher than that of PPi. This observation is in line with the inference that PPi hydrolysis is critical for the completion of the DNA synthesis reaction (Figure 5).

X-ray diffraction data were collected from crystals of $PolIV_{DNA:dTMPPnP}$ complex soaked with Mg^{2+} for 0min and 60min (Supplementary Table S2). The corresponding electron density maps showed that on Mg^{2+} ion entry, the sugar pucker of the terminal primer nucleotide changes and the incoming nucleotide moves closer to the terminal primer nucleotide. The distance between the 3′O- of the terminal primer nucleotide and the α-phosphate of dTMPPnP is 3.3 Å and the angle formed between these two atoms and the bridging oxygen between α- and β- phosphates is 171° (Supplementary Figure S9). However, the reaction does not proceed to completion; incorporation does not occur (Figure 6) and, this observation reinforces the importance of PPi hydrolysis.

Temporal sequence of events during the dNTP incorporation reaction

Based on our studies, the following series of events may occur during the DNA synthesis reaction catalyzed by PolIV (Figure 7 & Movie S2). Initially, binding of dNTP along with Mg^{2+} ions initiates the DNA synthesis reaction in the DNA polymerase active site. (I) Once all the components required for the DNA synthesis are assembled in the correct location and conformation, Mg^{2+} ion activates the 3′-OH of the primer nucleotide, and this results in the change in the sugar pucker from C2′-endo to C3′-endo form, (II-V) phosphodiester bond (pB) formation results in the appearance of a penta-covalent transition state which is followed by αO bond dissolution and release of PPi, (VI) PPi is further hydrolyzed to two phosphate ions followed by reorientation of the Arg49 residue. (VII) γPi diffuses out first (VIII) both metal ions and βPi are released followed by DNA translocation so that the newly incorporated nucleotide moves out of the nucleotide-binding site to make way for the next nucleotide to enter, and a new cycle of DNA synthesis can start. The identity of the incoming nucleotide will be determined by the next unpaired template nucleotide. DNA replication is hypothesized to have existed in DNA viruses even before the emergence of the Last Universal Common Ancestor (42). Hence, it is possible that the proposed scheme may have appeared early in evolution and may be common for all dPols (43,44).

DISCUSSION

Our studies reveal that breakdown of the PPi moiety is critical for completion of the DNA synthesis reaction. The conversion of PPi to Pi ($\Delta G = -7$ kcal/mol) is essential to render a large overall negative free energy change ($\Delta G = -6.5$ kcal/mol) to the DNA synthesis reaction (45,46). It was previously believed that dPols act in tandem with pyrophosphatase enzymes that cleave the PPi, so that the coupling of the two reactions provides an overall negative free energy change (1,8). It is clear from this study that the PPi hydrolysis step is inherently part of the complete dNTP incorporation reaction catalyzed by dPols and therefore, the DNA synthesis reaction carried out by dPols is energetically favorable without the need of any other enzyme activity. The entropic penalty imposed by nucleotide incorporation is expected to be substantial since the mobility of the dNTP reduces drastically on polymerization (24,47). PPi hydrolysis may also serve to reduce the effect of reduction in entropy on the energetics of the DNA synthesis reaction. The energy released due to the breakdown of PPi may also aid in translocation of DNA on the polymerase so that the next unpaired template nucleotide reaches the active site and a new round of DNA synthesis can begin.

PPi is known to participate in pyrophosphorolysis, which is the reverse of the polymerization reaction and involves excision of the terminal nucleotide of the primer strand (48). Pyrophosphorolysis results in restoration of the dNTP and the length of the primer reduces by one nucleotide. The hydrolysis of the PPi moiety ensures that the probability of the pyrophosphorolysis reaction is minimal and DNA synthesis is irreversible.

Recently, a study on DNA polymerase β has shown that utilization of dNTPs that are non-hydrolyzable at the β-γ position leads to the promotion of the pyrophosphorolysis reaction instead of the DNA synthesis reaction (49). Also, the utilization of a PPi analog, with an imido- linkage has been shown to promote pyrophosphorolysis (50). Moreover, modification at the β- and γ- phosphates of dNTPs results in a drastic reduction or complete failure of incorporation of the incoming nucleotide by different dPols (51–

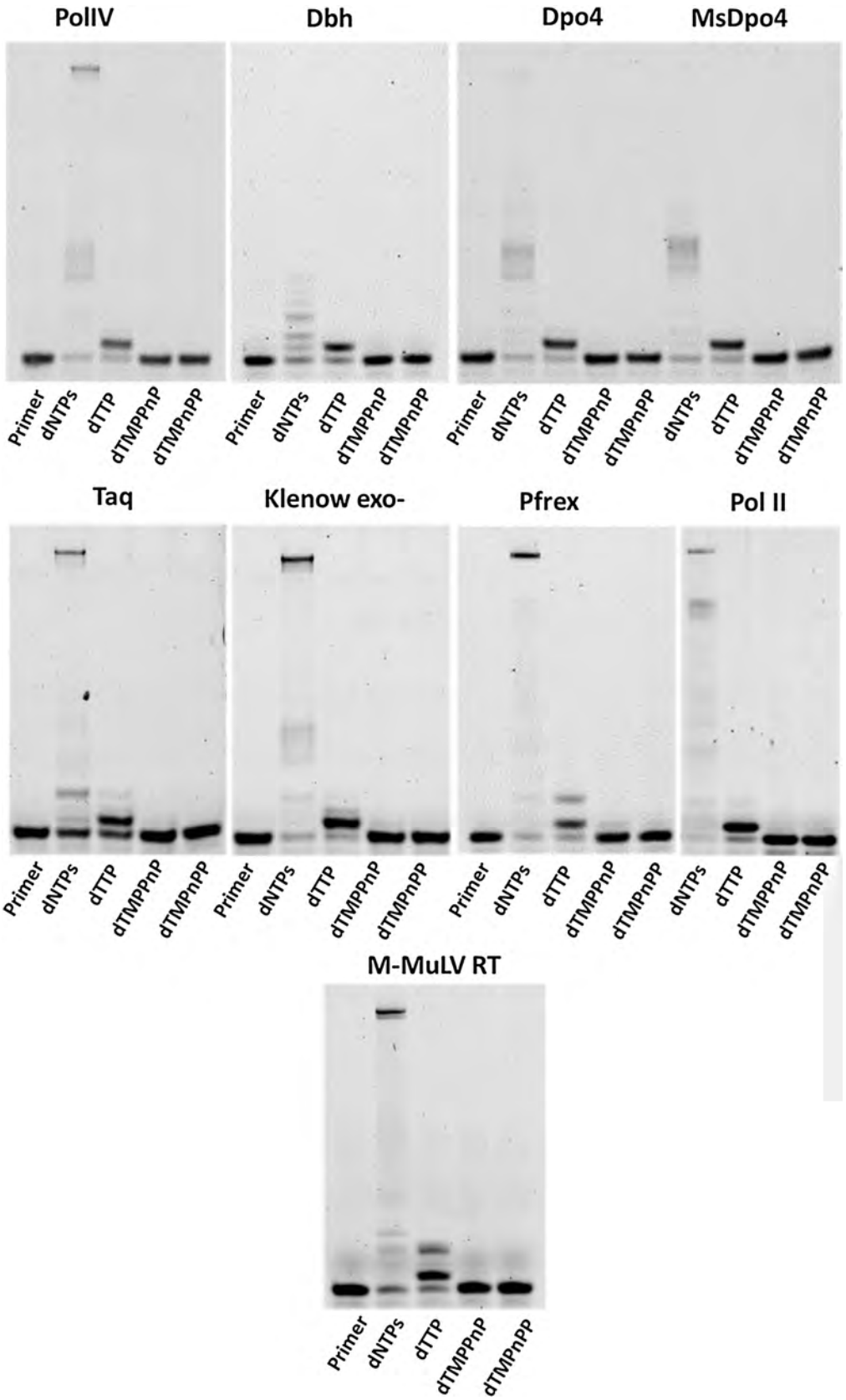

Figure 3. Primer Extension Assays with template dA and incoming dTTP, dTMPnPP and dTMPPnP. Opposite template dA, different dPols added dTTP and could not add the α-β modified dTMPnPP. The β-γ modified dTMPPnP was also not incorporated into the primer by these enzymes and this observation validates the importance of PPi hydrolysis in the DNA synthesis reaction.

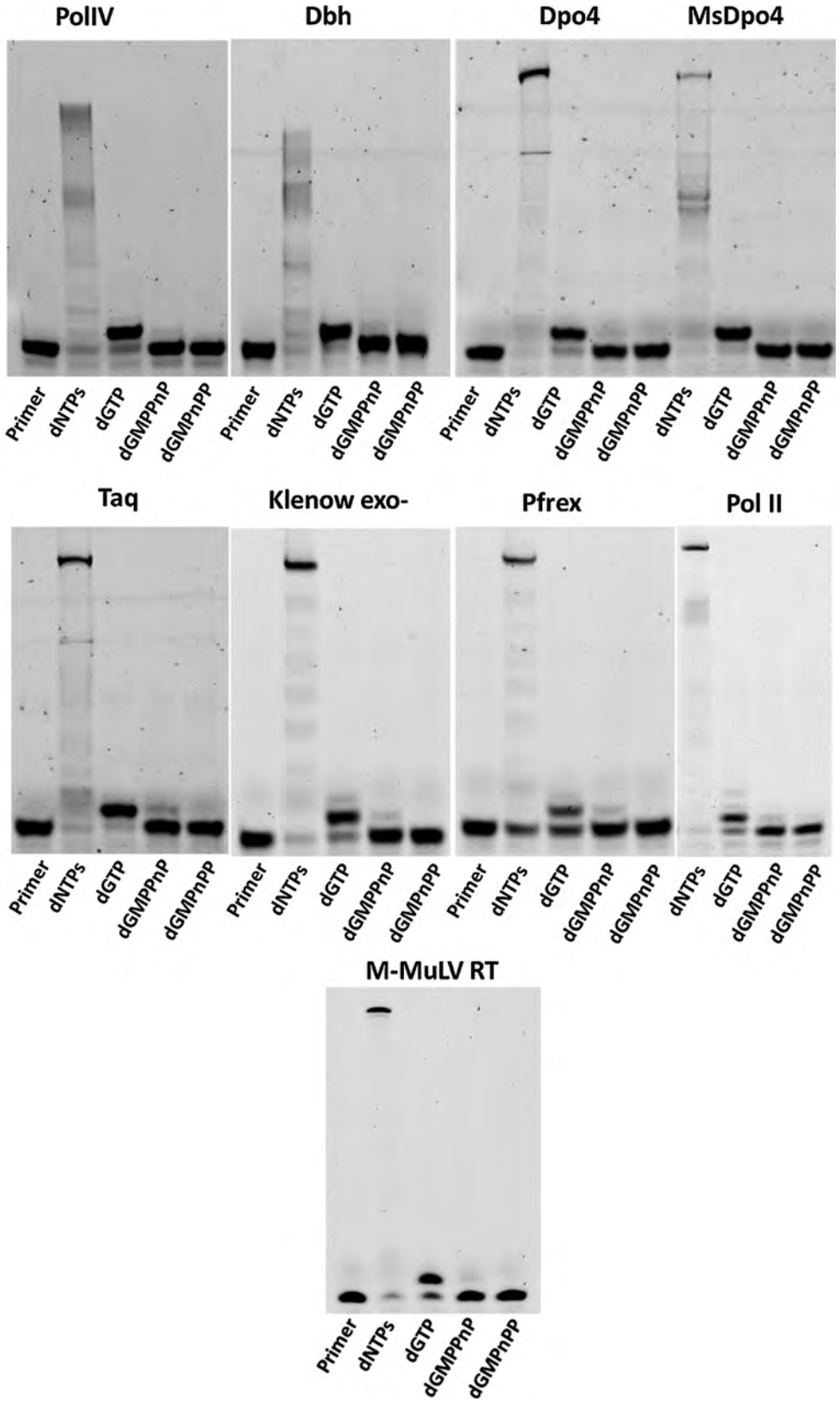

Figure 4. Primer Extension Assays with template dC and incoming dGTP, dGMPnPP and dGMPPnP. The tested dPols added dGTP and could not add the α-β modified dGMPnPP opposite template dC. The β-γ modified dGMPPnP was also not incorporated into the primer and therefore, hydrolysis of PPi is important for completion of the DNA synthesis reaction.

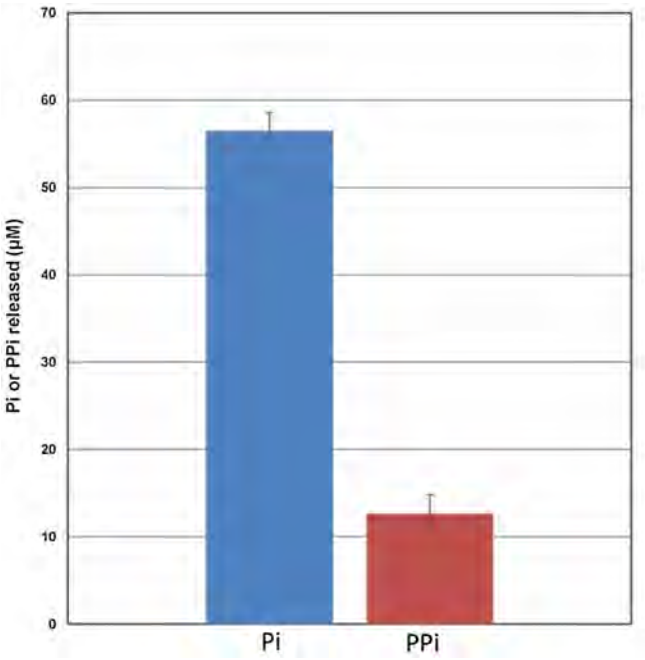

Figure 5. Comparison of PPi and Pi formation during DNA synthesis. The displayed graph shows that, during DNA synthesis, the amount of Pi generated is >5-fold as compared to PPi.

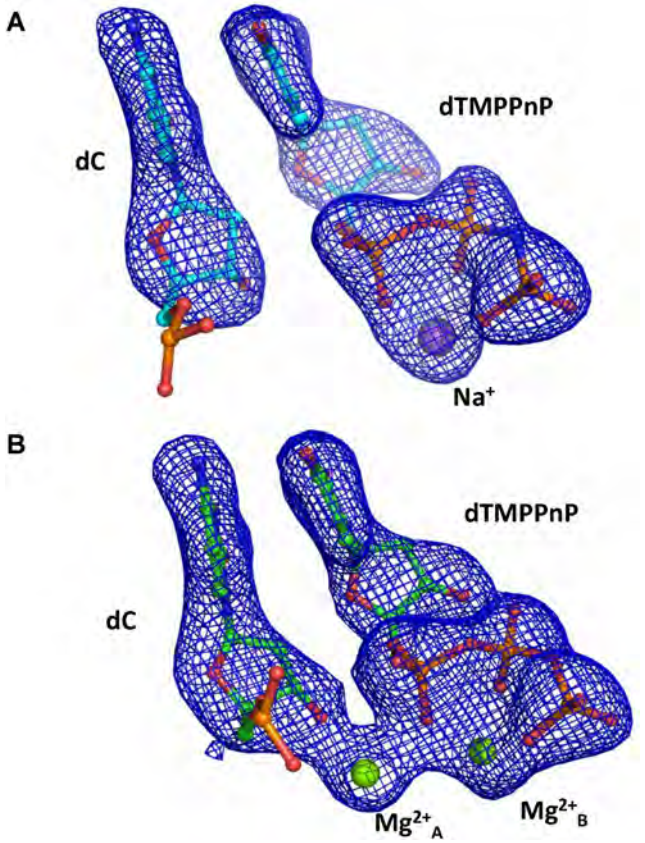

Figure 6. DNA synthesis reaction does not occur in the presence of dTMPPnP. Omit $F_o - F_c$ maps are displayed at a contour of 3.3 for the $PolIV_{DNA:dTMPPnP}$ complex before (**A**) and 60 min after the addition of Mg^{2+} (**B**). The electron density maps show that although the 3′-OH is activated, the DNA synthesis reaction does not proceed further. These observations are in line with the inference that cleavage of the pyrophosphate moiety is critical for the irreversible completion of the dNTP incorporation reaction.

53). These observations are in line with the inference that cleavage of the PPi is critical for completion of the DNA synthesis. It has been suggested previously that dissociation of metal A from the polymerase prevents pyrophosphorolysis (21). However, our studies indicate that the hydrolysis of PPi may be the primary strategy to avoid the reversal of the synthesis reaction. The breakdown of PPi will, therefore allow for smooth progression of the replication fork and ensure that the cell is not subjected to replication stress.

The cleavage of the PPi will contribute towards enforcing fidelity. The presence of a mispair in the active site will lead to distortion of the spatial alignment of the PPi group with respect to the enzyme residues and the Mg^{2+} ion. As a result, the rate at which PPi group is broken down will reduce considerably and thus affect completion of the synthesis reaction (52–55). Although the PPi moiety can potentially move to bind in the correct orientation after the dissolution of the αO bond, pyrophosphorolysis might happen before PPi has attained the right configuration for cleavage and thus prevent incorporation of the wrong nucleotide. It has been shown that the average ΔG° for incorporation of the correct dNTP is −5.2 kcal/mol and that for the addition of the incorrect dNTP is −0.13 kcal/mol (56). These observations suggest that the PPi moiety is cleaved only when the correct dNTP is incorporated and are in line with the inference that cleavage of PPi moiety by dPols will contribute towards ensuring fidelity of the DNA synthesis reaction.

The modification of the γ-phosphate has been shown to affect the ability of viral reverse transcriptase to incorporate nucleotides, and hence, it is possible that PPi cleavage may be an evolutionarily conserved feature of all replicative polymerases (57). Enhancement in pyrophosphorolysis activity has been implicated in the ability of mutant viral polymerases to remove chain terminating inhibitors that are used as drugs such as zidovudine (58,59). It is believed that these mutations enhance the affinity for PPi, and thus promote the reverse reaction. However, our studies raise the possibility that these mutations might prevent/reduce cleavage of PPi by the enzyme and thus promote pyrophosphorolysis.

Overall, we provide concrete experimental evidence that hydrolysis of the pyrophosphate moiety is an intrinsic and critical step of the DNA synthesis reaction. Also, DNA synthesis is a stepwise associative S_N2 reaction assisted by two Mg^{2+} ions. The study brings to light the mechanism of the fundamental reaction responsible for genome duplication and the insight obtained from this study may aid the development of improved PCR-based diagnostic kits and novel therapeutic strategies against retroviruses.

DATA AVAILABILITY

The Structure factors and the refined co-ordinates have been deposited in the PDB with the following codes: 5YUR (0 min), 5YUS (1 min), 5YUT (3 min), 5YUU (5 min), 5YUV (7 min), 5YUW (10 min), 5YV3 (15 min), 5YUX (20 min), 5YUY (25 min), 5YV4 (30 min), 5YUZ (35 min), 5YV0 (40 min), 5YV1 (50 min), 5YV2 (60 min), 5YYD (dTMPPnP: 0 min), 5YYE (dTMPPnP: 60 min) and 5ZLV (50 mM $MgCl_2$; dTTP: 30 min).

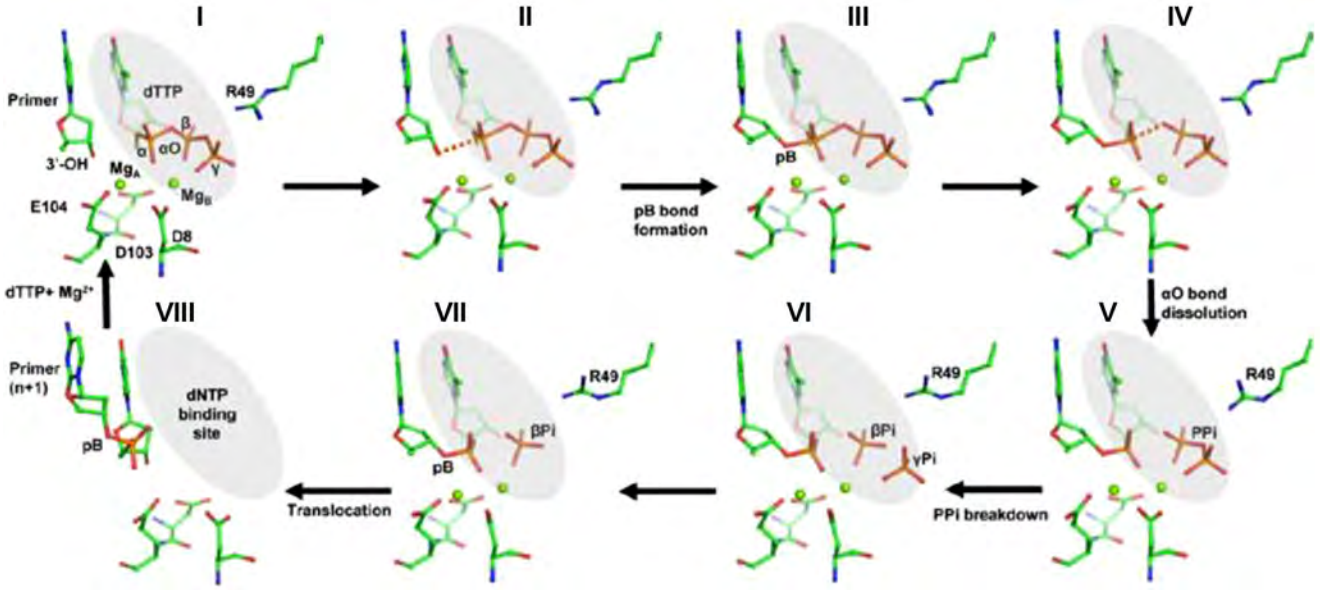

Figure 7. Mechanism of incorporation of dTTP opposite dA by DNA polymerase IV. The different steps associated with the DNA synthesis reaction catalyzed by PolIV are displayed. The terminal primer nucleotide (dC), the incoming nucleotide (dTTP) and active site residues are shown in stick representation and colored according to element. The Mg^{2+} ions are shown in the form of green spheres.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Minakshi Sharma and Mary K. Johnson for purifying Pfprex and MsDpo4, respectively; Dr Dinakar M. Salunke (ICGEB, New Delhi) and Prof. Jayant B. Udgaonkar (NCBS, Bangalore) for critically reading the manuscript; Prof. S. Ramaswamy (inStem, Bangalore) for discussions. We thank the X-ray diffraction facility located at the Regional Centre for Biotechnology. DTN thanks Dr. Hassan Belrhali & Dr Babu Manjashetty (BM14 beamline, ESRF) and Dr Danielle de Sanctis (ID29 beamline, ESRF) for help with X-ray diffraction data collection.

FUNDING

Regional Centre for Biotechnology; Data collection at the BM14 beamline of ESRF (Grenoble, France) was supported by the BM14 project—a collaboration between DBT, EMBL and ESRF; Data collection at ID29 was facilitated by the ESRF Access Program of RCB which is supported by Department of Biotechnology, Government of India [BT/INF/22/SP22660/2017]. Funding for open access charge: Intramural Funding from Regional Centre for Biotechnology.
Conflict of interest statement. None declared.

REFERENCES

1. Watson,J.D., Baker,T.A., Bell,S.P., Gann,A., Levine,M. and Losick,R. (2013) *Molecular Biology of the Gene*. 7th edn. Pearson, (Chapter 3 and Chapter 9).
2. Bessman,M.J., Kornberg,A., Lehman,I.R. and Simms,E.S. (1956) Enzymic synthesis of deoxyribonucleic acid. *Biochim. Biophys. Acta*, **21**, 197–198.
3. Berdis,A.J. (2009) Mechanisms of DNA polymerases. *Chem. Rev.*, **109**, 2862–2879.
4. Johansson,E. and Dixon,N. (2013) Replicative DNA polymerases. *Cold Spring Harb. Perspect. Biol.*, **5**. a012799
5. Aschenbrenner,J. and Marx,A. (2017) DNA polymerases and biotechnological applications. *Curr. Opin. Biotechnol.*, **48**, 187–195.
6. Peller,L. (1976) On the free-energy changes in the synthesis and degradation of nucleic acids. *Biochemistry*, **15**, 141–146.
7. Peller,L. (1966) Thermodynamic factors in the synthesis of two-stranded nucleic acids. *Proc. Natl. Acad. Sci. U.S.A.*, **55**, 1025–1031.
8. Lapenta,F., Monton Silva,A., Brandimarti,R., Lanzi,M., Gratani,F.L., Vellosillo Gonzalez,P., Perticarari,S. and Hochkoeppler,A. (2016) Escherichia coli DnaE polymerase couples pyrophosphatase activity to DNA replication. *PLoS One*, **11**, e0152915.
9. Burke,C.R. and Luptak,A. (2018) DNA synthesis from diphosphate substrates by DNA polymerases. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, 980–985.
10. Aravind,L. and Koonin,E.V. (1998) Phosphoesterase domains associated with DNA polymerases of diverse origins. *Nucleic Acids Res.*, **26**, 3746–3752.
11. Steitz,T.A. (1999) DNA polymerases: structural diversity and common mechanisms. *J. Biol. Chem.*, **274**, 17395–17398.
12. 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.
13. 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.
14. Swan,M.K., Johnson,R.E., Prakash,L., Prakash,S. and Aggarwal,A.K. (2009) Structural basis of high-fidelity DNA synthesis by yeast DNA polymerase delta. *Nat. Struct. Mol. Biol.*, **16**, 979–986.
15. 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.
16. Silverstein,T.D., Johnson,R.E., Jain,R., Prakash,L., Prakash,S. and Aggarwal,A.K. (2010) Structural basis for the suppression of skin cancers by DNA polymerase eta. *Nature*, **465**, 1039–1043.
17. Zahn,K.E., Averill,A.M., Aller,P., Wood,R.D. and Doublie,S. (2015) Human DNA polymerase theta grasps the primer terminus to mediate DNA repair. *Nat. Struct. Mol. Biol.*, **22**, 304–311.

18. Sharma,A., Kottur,J., Narayanan,N. and Nair,D.T. (2013) A strategically located serine residue is critical for the mutator activity of DNA polymerase IV from Escherichia coli. *Nucleic Acids Res.*, **41**, 5104–5114.
19. Nakamura,T., Zhao,Y., Yamagata,Y., Hua,Y.J. and Yang,W. (2012) Watching DNA polymerase eta make a phosphodiester bond. *Nature*, **487**, 196–201.
20. Freudenthal,B.D., Beard,W.A., Shock,D.D. and Wilson,S.H. (2013) Observing a DNA polymerase choose right from wrong. *Cell*, **154**, 157–168.
21. Gao,Y. and Yang,W. (2016) Capture of a third Mg^{2+} is essential for catalyzing DNA synthesis. *Science*, **352**, 1334–1337.
22. Yang,W., Weng,P.J. and Gao,Y. (2016) A new paradigm of DNA synthesis: three-metal-ion catalysis. *Cell Biosci.*, **6**, 51.
23. Kamerlin,S.C., Sharma,P.K., Prasad,R.B. and Warshel,A. (2013) Why nature really chose phosphate. *Q. Rev. Biophys.*, **46**, 1–132.
24. Ram Prasad,B. and Warshel,A. (2011) Prechemistry versus preorganization in DNA replication fidelity. *Proteins*, **79**, 2900–2919.
25. Kottur,J. and Nair,D.T. (2016) Reactive oxygen species play an important role in the bactericidal activity of quinolone antibiotics. *Angew. Chem. Int. Ed. Engl.*, **55**, 2397–2400.
26. Kottur,J., Sharma,A., Gore,K.R., Narayanan,N., Samanta,B., Pradeepkumar,P.I. and Nair,D.T. (2015) Unique structural features in DNA polymerase IV enable efficient bypass of the N2 adduct induced by the nitrofurazone antibiotic. *Structure*, **23**, 56–67.
27. Sharma,A. and Nair,D.T. (2011) Cloning, expression, purification, crystallization and preliminary crystallographic analysis of MsDpo4: a Y-family DNA polymerase from Mycobacterium smegmatis. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, **67**, 812–816.
28. Sharma,A. and Nair,D.T. (2012) MsDpo4-a DinB homolog from mycobacterium smegmatis-Is an Error-Prone DNA polymerase that can promote G:T and T:G mismatches. *J. Nucleic Acids*, **2012**, 285481.
29. Leslie,A.W. and Powell,H. (2007) In: Read,R and Sussman,J (eds). *Evolving Methods for Macromolecular Crystallography*. Springer, Netherlands, Vol. **245**, pp. 41–51.
30. Battye,T.G., Kontogiannis,L., Johnson,O., Powell,H.R. and Leslie,A.G. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 271–281.
31. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L. C. and Read,R. J. (2007) Phase crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
32. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
33. Kabsch,W. (2010) Xds. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 125–132.
34. Evans,P.R. and Murshudov,G.N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.*, **69**, 1204–1214.
35. Evans,P.R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 282–292.
36. Evans,P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 72–82.
37. Emsley,P. and Cowtan,K. (2004) Coot: model building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
38. Kolodiazhnyi,O.I. and Kolodiazhna,A. O. (2017) Nucleophilic substitution at phosphorus: stereochemistry and mechanisms. *Tetrahedron: Assymetry*, **28**, 24.
39. Westheimer,F.H. (1987) Why nature chose phosphates. *Science*, **235**, 1173–1178.
40. Da,L.T., Wang,D. and Huang,X. (2012) Dynamics of pyrophosphate ion release and its coupled trigger loop motion from closed to open state in RNA polymerase II. *J. Am. Chem. Soc.*, **134**, 2399–2406.
41. Genna,V., Gaspari,R., Dal Peraro,M. and De Vivo,M. (2016) Cooperative motion of a key positively charged residue and metal ions for DNA replication catalyzed by human DNA Polymerase-eta. *Nucleic Acids Res.*, **44**, 2827–2836.
42. Forterre,P. (2002) The origin of DNA genomes and DNA replication proteins. *Curr. Opin. Microbiol.*, **5**, 525–532.
43. Cheetham,G.M., Jeruzalmi,D. and Steitz,T.A. (1998) Transcription regulation, initiation, and "DNA scrunching" by T7 RNA polymerase. *Cold Spring Harb. Symp. Quant. Biol.*, **63**, 263–267.
44. Yao,N.Y. and O'Donnell,M.E. (2016) Evolution of replication machines. *Crit. Rev. Biochem. Mol. Biol.*, **51**, 135–149.
45. Fazakerley,G.V., Sowers,L.C., Eritja,R., Kaplan,B.E. and Goodman,M.F. (1987) Structural and dynamic properties of a bromouracil-adenine base pair in DNA studied by proton NMR. *J. Biomol. Struct. Dyn.*, **5**, 639–650.
46. M.Davies,J., J.Poole,R. and Sanders,D. (1993) The computed free energy change of hydrolysis of inorganic pyrophosphate and ATP: apparent significance. for inorganic-pyrophosphate-driven reactions of intermediary metabolism. *Biochim. Biophys. Acta (BBA) - Bioenergetics*, **1141**, 29–36.
47. Minetti,C.A., Remeta,D.P., Miller,H., Gelfand,C.A., Plum,G.E., Grollman,A.P. and Breslauer,K.J. (2003) The thermodynamics of template-directed DNA synthesis: base insertion and extension enthalpies. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 14719–14724.
48. Pandey,M., Patel,S.S. and Gabriel,A. (2008) Kinetic pathway of pyrophosphorolysis by a retrotransposon reverse transcriptase. *PLoS One*, **3**, e1389.
49. Shock,D.D., Freudenthal,B.D., Beard,W.A. and Wilson,S.H. (2017) Modulating the DNA polymerase beta reaction equilibrium to dissect the reverse reaction. *Nat. Chem. Biol.*, **13**, 1074–1080.
50. Rozovskaya,T., Tarussova,N., Minassian,S., Atrazhev,A., Kukhanova,M., Krayevsky,A., Chidgeavadze,Z. and Beabealashvilli,R. (1989) Pyrophosphate analogues in pyrophosphorolysis reaction catalyzed by DNA polymerases. *FEBS Lett.*, **247**, 289–292.
51. Martynov,B.I., Shirokova,E.A., Jasko,M.V., Victorova,L.S. and Krayevsky,A.A. (1997) Effect of triphosphate modifications in 2′-deoxynucleoside 5′-triphosphates on their specificity towards various DNA polymerases. *FEBS Lett.*, **410**, 423–427.
52. Sucato,C.A., Upton,T.G., Kashemirov,B.A., Batra,V.K., Martinek,V., Xiang,Y., Beard,W.A., Pedersen,L.C., Wilson,S.H., McKenna,C.E. *et al.* (2007) Modifying the beta,gamma leaving-group bridging oxygen alters nucleotide incorporation efficiency, fidelity, and the catalytic mechanism of DNA polymerase beta. *Biochemistry*, **46**, 461–471.
53. Sucato,C.A., Upton,T.G., Kashemirov,B.A., Osuna,J., Oertell,K., Beard,W.A., Wilson,S.H., Florian,J., Warshel,A., McKenna,C.E. *et al.* (2008) DNA polymerase beta fidelity: halomethylene-modified leaving groups in pre-steady-state kinetic analysis reveal differences at the chemical transition state. *Biochemistry*, **47**, 870–879.
54. Lecomte,P., Doubleday,O.P. and Radman,M. (1986) Evidence for an intermediate in DNA synthesis involving pyrophosphate exchange. A possible role in fidelity. *J. Mol. Biol.*, **189**, 643–652.
55. 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.
56. Olson,A.C., Patro,J.N., Urban,M. and Kuchta,R.D. (2013) The energetic difference between synthesis of correct and incorrect base pairs accounts for highly accurate DNA replication. *J. Am. Chem. Soc.*, **135**, 1205–1208.
57. Arzumanov,A.A., Semizarov,D.G., Victorova,L.S., Dyatkina,N.B. and Krayevsky,A.A. (1996) Gamma-phosphate-substituted 2′-deoxynucleoside 5′-triphosphates as substrates for DNA polymerases. *J. Biol. Chem.*, **271**, 24389–24394.
58. Arion,D. and Parniak,M.A. (1999) HIV resistance to zidovudine: the role of pyrophosphorolysis. *Drug Resist. Updat.*, **2**, 91–95.
59. Urban,S., Fischer,K.P. and Tyrrell,D.L. (2001) Efficient pyrophosphorolysis by a hepatitis B virus polymerase may be a primer-unblocking mechanism. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 4984–4989.

Published online 23 September 2019 *Nucleic Acids Research, 2019, Vol. 47, No. 20* ***10693–10705***
doi: 10.1093/nar/gkz792

A polar filter in DNA polymerases prevents ribonucleotide incorporation

Mary K. Johnson[1,2], Jithesh Kottur[1] and Deepak T. Nair[1,*]

[1]Regional Centre for Biotechnology, NCR Biotech Science Cluster, 3rd Milestone, Faridabad-Gurgaon Expressway, Faridabad 121001, India and [2]National Centre for Biological Sciences, Tata Institute of Fundamental Research, GKVK Campus, Bangalore 560065, India

Received May 15, 2019; Revised September 02, 2019; Editorial Decision September 04, 2019; Accepted September 09, 2019

ABSTRACT

The presence of ribonucleotides in DNA can lead to genomic instability and cellular lethality. To prevent adventitious rNTP incorporation, the majority of the DNA polymerases (dPols) possess a steric filter. The dPol named MsDpo4 (*Mycobacterium smegmatis*) naturally lacks this steric filter and hence is capable of rNTP addition. The introduction of the steric filter in MsDpo4 did not result in complete abrogation of the ability of this enzyme to incorporate ribonucleotides. In comparison, DNA polymerase IV (PolIV) from *Escherichia coli* exhibited stringent selection for deoxyribonucleotides. A comparison of MsDpo4 and PolIV led to the discovery of an additional polar filter responsible for sugar selectivity. Thr43 represents the filter in PolIV and this residue forms interactions with the incoming nucleotide to draw it closer to the enzyme surface. As a result, the 2'-OH in rNTPs will clash with the enzyme surface, and therefore ribonucleotides cannot be accommodated in the active site in a conformation compatible with productive catalysis. The substitution of the equivalent residue in MsDpo4–Cys47, with Thr led to a drastic reduction in the ability of the mycobacterial enzyme to incorporate rNTPs. Overall, our studies evince that the polar filter serves to prevent ribonucleotide incorporation by dPols.

INTRODUCTION

DNA polymerases (dPols) synthesize DNA through template-dependent addition of dNTPs to the growing primer strand. All dPols share a similar core structure with thumb, fingers, and palm domains. Three acidic residues constitute the catalytic center of these enzymes, and these residues are present in the palm domain. The catalytic residues co-ordinate a divalent metal ion, usually Mg^{2+}, that is critical for DNA synthesis activity (1–3).

In the cellular environment, dPols encounter rNTPs much more frequently than dNTPs. The cellular concentrations of rNTPs are up to 10–100-fold higher than those of dNTPs depending on the type of cell, and stage of cell cycle (4–9). Like dNTPs, rNTPs can form Watson-Crick base pairs with DNA nucleotides, as they are sister molecules of dNTPs with an additional 2′-OH on the sugar moiety (4,10,11). Hence it is possible, that the dPols may inadvertently add rNTPs to the growing primer strand. The presence of ribonucleotides in DNA is detrimental to the integrity of the genome as they are less stable due to the reactive nature of the hydroxyl group on the C2′ atom (7–10,12,13). The presence of ribonucleotides in DNA can lead to the spontaneous appearance of double-strand breaks (5,8,13,14). The misincorporation of rNTPs into DNA is known to slow down the rate of DNA replication by Polδ, Polε and PolIII (5,6,12,15). Ribonucleotides can alter the conformation of DNA from B to A form, and thus adversely affect target sequence recognition by proteins such as transcription factors (10,16,17). Overall, the presence of ribonucleotides in the genomic DNA is detrimental to cellular physiology and, represents a significant challenge for the basic genomic processes such as replication, repair, and transcription (13,17). It is, therefore, imperative to reduce the frequency of incorporation of rNTPs into the genome during DNA replication.

Most dPols employ a steric filter to minimize rNTP incorporation (4,18–20). The steric filter is represented by an amino acid residue with a bulky side chain which will clash with the 2′-OH group on the ribose ring of the incoming nucleotide (Supplementary Figure S1) (9,19,21). As a result, rNTPs cannot bind stably in the active site of the dPol. The steric gate is represented by Glu in A-family polymerases and Tyr or Phe in members of the B, X, Y and RT families (11,18–20,22–36).

MsDpo4 is a Y-family dPol from *Mycobacterium smegmatis* (37,38). MsDpo4 does not possess the steric filter and can incorporate ribonucleotides (28,39). In comparison, DNA polymerase IV (PolIV) from *Escherichia coli,* exhibits stringent sugar selectivity. A rigorous comparison between MsDpo4 and PolIV shows the presence of a strategi-

*To whom correspondence should be addressed. Tel: +91 124 2848844; Email: deepak@rcb.res.in

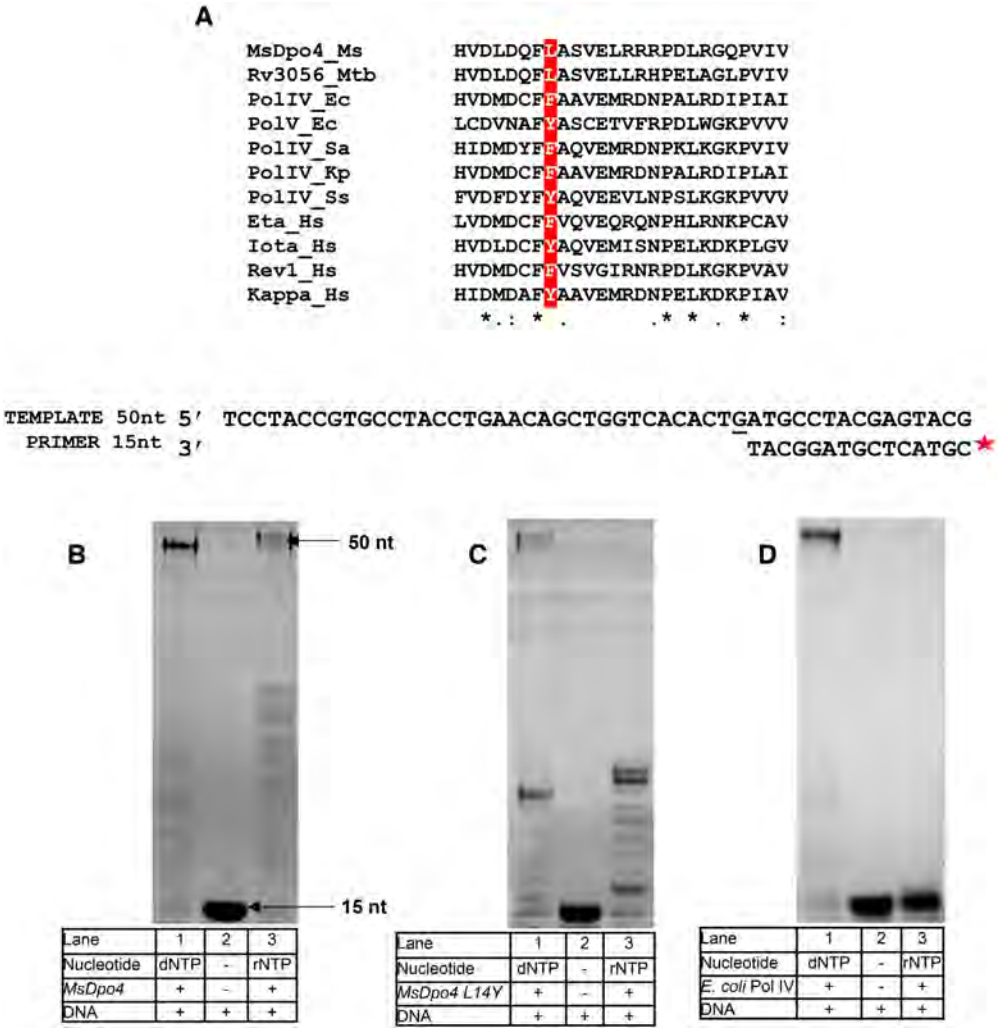

Figure 1. Comparison of rNTP incorporation by MsDpo4-WT, MsDpo4-L14Y, and PolIV. (**A**) Sequence alignment of the N-terminal region of MsDpo4 and homologous proteins from bacteria, archaea, and humans shows that MsDpo4 lacks the steric filter as it has a Leu residue instead of an aromatic residue at the appropriate position (highlighted in red). Primer extension assays show that (**B**) MsDpo4-WT is capable of rNTP and dNTP incorporation. (**C**) MsDpo4-L14Y can incorporate rNTP and dNTP and (**D**) PolIV-WT from *E. coli* is unable to incorporate rNTP and therefore exhibits stringent sugar selectivity.

cally located polar residue that aids sugar selectivity during DNA synthesis. The polar residue, therefore, represents a second filter- termed as a polar filter, which acts in tandem with the steric filter to prevent ribonucleotide incorporation by dPols.

MATERIALS AND METHODS

Protein purification

MsDpo4 (C47T, L14Y, C47T+L14Y) and PolIV (F13A and F13A+T43C) mutants were generated using QuikChange Lightning site-directed mutagenesis kit (Agilent). The MsDpo4 and PolIV proteins were expressed and purified as described previously (37,38,40). For phasing, selenomethionine-labeled MsDpo4 was prepared using B834 strain of *E. coli*. The labeled protein was purified and crystallized using identical protocols as for the native protein.

The PolII gene (*E. coli*) was amplified from *E. coli* genomic DNA and cloned into a modified pET-28b vector (pDJN1) (41). The exonuclease deficient mutant of PolII (D335A) (42) was generated using QuikChange Lightning site-directed mutagenesis kit (Agilent). The protein was expressed in the C41(DE3) strain of *E. coli*. The freshly transformed cells were grown to an $OD_{600} = 0.8$, at 37°C. The expression of PolII was induced with 0.4 mM IPTG followed by incubation at 18°C for 16 h. The cells were then harvested by centrifugation, lysed by sonication and the lysate was clarified by centrifugation. The purification was done using pre-packed Ni-NTA column (GE-Healthcare). The binding buffer (Buffer A) composition was 25 mM Tris pH 8.0, 250 mM NaCl, and 5% glycerol. The elution buffer (Buffer B) composition was 25 mM Tris pH 8.0, 250 mM NaCl, 5% glycerol and 1 M Imidazole. The bound protein was eluted using a step gradient of elution Buffer B. The eluted protein was then dialyzed overnight in Buffer A and then, concentrated and stored in −80°C.

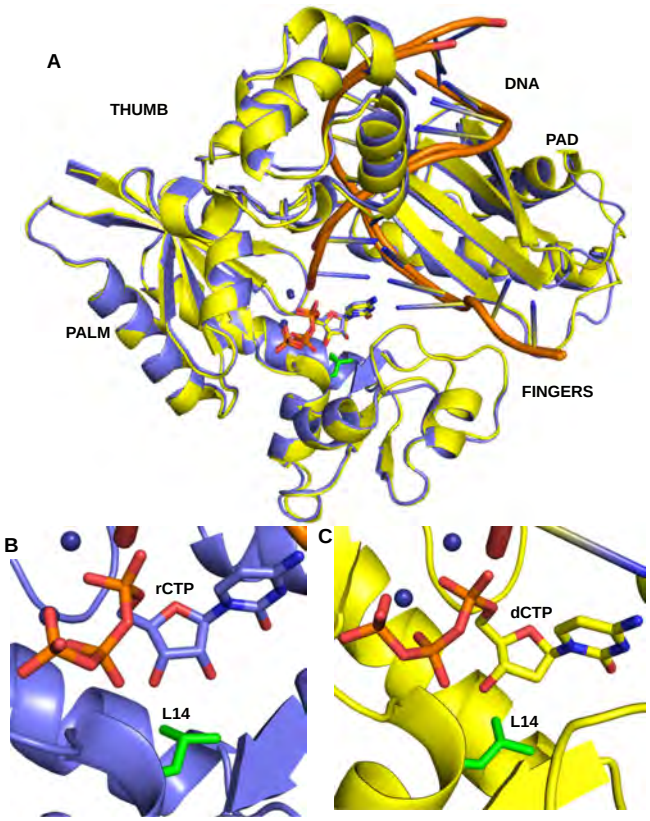

Figure 2. Structural mechanism of rCTP and dCTP incorporation by MsDpo4-WT. (**A**) The structures of MsDpo4-$WT_{DNA(dG):dCTP}$ (colored yellow) and MsDpo4-$WT_{DNA(dG):rCTP}$ (colored violet) superimpose with an RMSD of 0.36 Å. The enzyme structure is shown in cartoon representation, DNA is shown as ribbons, and the incoming nucleotide and the Leu14 residue (in green) are shown in stick representation. A close-up of the region surrounding the incoming nucleotide, is displayed for (**B**) $MsDpo4_{DNA(dG):rCTP}$ and (**C**) $MsDpo4_{DNA(dG):dCTP}$. The comparison shows that the dCTP and rCTP bind the MsDpo4 active site in the same location and conformation with marginal differences in the enzyme structure.

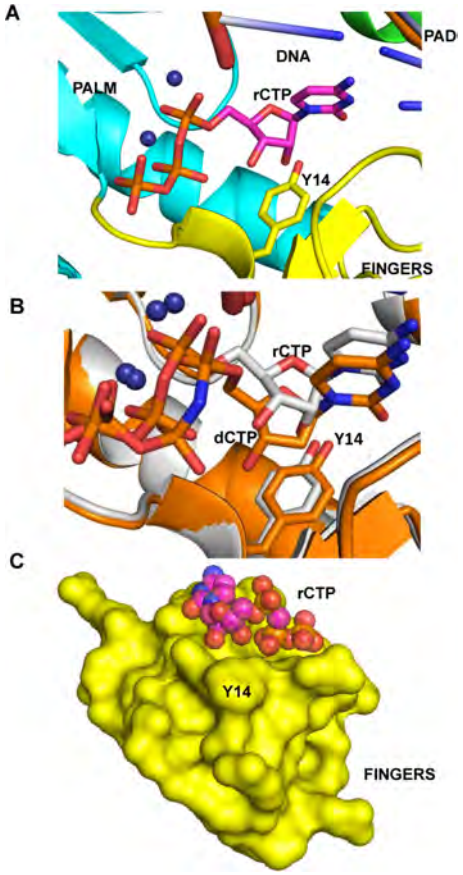

Figure 3. Mechanism of rCTP incorporation by MsDpo4-L14Y. (**A**) A close-up of the region surrounding the incoming nucleotide is displayed for MsDpo4-$L14Y_{DNA(dG):rCTP}$ structure. (**B**) The structures of MsDpo4-$L14Y_{DNA(dG):dCTP}$ (colored orange) and MsDpo4-$L14Y_{DNA(dG):rCTP}$ (colored white), superimpose with an RMSD of 0.37 Å. The comparison shows that the rCTP is repositioned and undergoes a change in the sugar pucker to prevent a steric clash between the 2′-OH and the engineered steric filter. (**C**) The surface of the fingers domain (in yellow) from the MsDpo4-$L14Y_{DNA(dG):rCTP}$ structure is shown along with space filling model of the bound rCTP (colored according to element). The figure shows that there is no steric clashes between the rCTP molecule and the enzyme surface in this structure.

Primer extension assay and steady-state kinetic analysis

To prepare the DNA substrate, a 6FAM labeled 15mer primer (5′-6FAM-CGTACTCGTAGGCAT-3′) was annealed to 50mer template (5′- TCCTACCGTGCCTA CCTGAACAGCTGGTCACACAXATGCCTACGA GTACG-3′). The nucleosides deoxyadenosine, deoxycytidine, deoxythymidine and deoxyguanosine occupied the underlined position for the templates A, C, T and G respectively. The reaction mixture was composed of enzyme (MsDpo4/PolIV/PolII), dsDNA, dNTPs or rNTPs (Jena Bioscience), 25 mM Tris pH 8.0, 2 mM DTT, 0.1 mM $(NH_4)_2SO_4$, 2.5 mM $MnCl_2$ and 0.05 mg/ml BSA in a final volume of 20 μl. The reaction was carried out at 37°C for 60 min. The reaction was terminated by adding 10 μl of stop buffer, composed of 80% formamide, 10% bromophenol blue and 10% 0.5 M EDTA (40,43). The reaction products were denatured by heating at 95°C for 10 min and quick cooling in ice for 10 min. The products were then resolved in a 20% polyacrylamide gel with 8 M urea, in 1× TBE buffer and visualized using Typhoon scanner FLA-7000 (GE Healthcare). The band intensities were quantified using ImageQuant software. Relative percentage incorporation (RPI) in each group was calculated with respect to the highest in the group. RPI was calculated using the following equation

$$\mathrm{RPI} = \left(\mathrm{I}/\mathrm{I}^{'}\right) \times 100,$$

where I is the percentage incorporation by the mutant and I' is the percentage incorporation by the most active version of the protein. For steady-state kinetic analysis, the reaction was carried out for varying time points (30 s to 30 min) at 37°C. The time point at which 20% of the primer had been extended was used for further analysis. Reactions were carried out with varying concentrations of incoming rCTP. After quantification of the resolved products, apparent K_m and V_{max} values were calculated using a Lineweaver–Burk plot. The measurements were carried out in triplicate, and the standard deviation values were calculated.

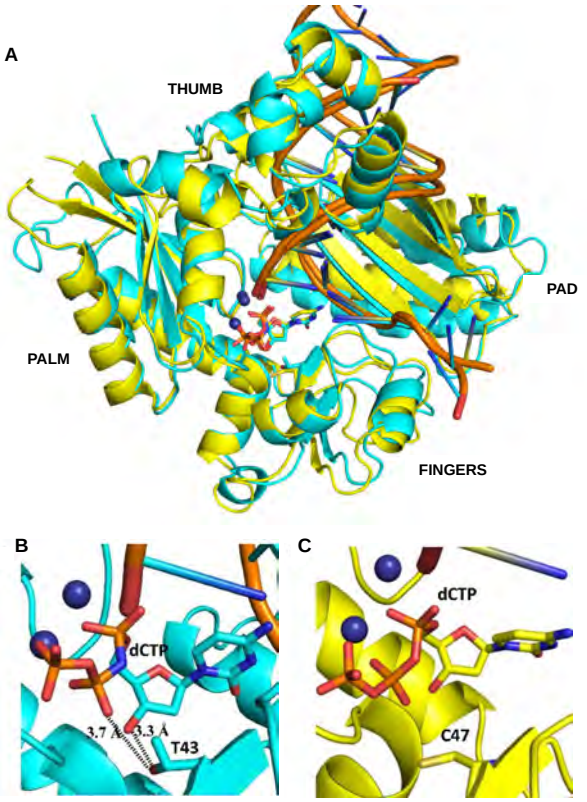

Figure 4. Comparison of the structures of ternary complexes of MsDpo4 and PolIV. (**A**) The two structures, $MsDpo4_{DNA(dG):dCTP}$ (colored yellow) and $PolIV_{DNA(dG):dCTP}$ (colored cyan) superimpose with an RMSD value of 1.47 Å. The comparison shows that, at the position equivalent to Thr43 residue in PolIV, Cys47 is present in MsDpo4. (**B**) In the $PolIV_{DNA(dG):dCTP}$ structure, Thr43 residue forms interactions with the 3′-OH and β-phosphate of the incoming dCTP and may play an important role in the ability of PolIV to exclude ribonucleotides. (**C**) The equivalent residue in the $MsDpo4_{DNA(dG):dCTP}$ structure is Cys47, and the Cys side chain cannot form interactions with the incoming nucleotide.

Crystallization, structure determination & crystallographic refinement

An 18mer self-complementary oligonucleotide (5′TCTGGGGTCCTAGGACCC 3′) was purified by ion-exchange chromatography using a monoQ column. The purified oligonucleotide was desalted, lyophilized, solubilized in water and annealed. The ternary complexes were reconstituted by mixing MsDpo4 (0.3 mM) with annealed dsDNA in a molar ratio of 1:1.2 followed by addition of 5 mM dCpNHpp or CpCpp (Jena Bioscience) and $MgCl_2$ or $MnCl_2$, respectively. Crystallization trials were carried out by the hanging drop method using commercial screens. After screening and optimization, the best crystals were obtained in 0.2 M Bis–Tris Propane (pH 6.0), 15–25% (w/v) PEG 3350 or PEG 2KMME and 0.1–0.7 M NaCl.

The PolIV-F13A+T43C protein in complex with DNA (template dG) and incoming rCpCpp or dCpNHpp was crystallized using the same conditions as that for wt-protein described previously (40,43). The initial characterization of the crystals was carried out using a Metaljet X-ray generator with a PhotonII detector (Bruker Inc.). The X-ray diffraction data for $MsDpo4_{DNA(dG):dCTP}$, $MsDpo4_{DNA(dG):rCTP}$, $MsDpo4\text{-}L14Y_{DNA(dG):dCTP}$ and $MsDpo4\text{-}L14Y_{DNA(dG):rCTP}$ were collected at the BM14 beamline of the European Synchrotron Radiation Facility (ESRF) in Grenoble. Data corresponding to $MsDpo4\text{-}C47T_{DNA(dG):dCTP}$, $PolIV\text{-}F13A+T43C_{DNA(dG):dCTP}$ and $PolIV\text{-}F13A+T43C_{DNA(dG):rCTP}$ were collected at ID30B and ID29 beamline at ESRF, Grenoble, France. The data for $MsDpo4\text{-}C47T+L14Y_{DNA(dG):dCTP}$ was collected at Elettra Sincrotrone in Trieste.

All images were integrated using iMOSFLM of CCP4 (44). The structure of the SeMet–$MsDpo4_{DNA(dG):dCTP}$ complex was determined by single-wavelength anomalous dispersion method using the AUTOSOL program in PHENIX (45). The initial model was built using the AUTOBUILD program in PHENIX, and subsequent manual rebuilding was done using COOT (46). The other structures of ternary complexes of wt- or mutant versions of MsDpo4 were determined by Molecular Replacement using PHASER in CCP4 (47,48). Similarly, the structure of $PolIV\text{-}T43C+F13A_{DNA(dG):rCTP}$ was determined by PHASER using the structure of $PolIV_{DNA(dG):dCTP}$ (4IRC) as a search model. All the structures were refined using PHENIX until convergence after which TLS refinement was carried out using Refmac in CCP4 (49).

Growth assays

The DNA sequence corresponding to the β-clamp binding peptide from *E. coli* PolIV (^{342}QMERQLVLGL^{351}) (50) was added to the 3′ end of the MsDpo4 gene, and this gene construct (*msdpo4β*) was cloned into a modified version of the pET-28b vector (41). The C47T mutation of the *msdpo4β* gene was generated using QuikChange Lightning site-directed mutagenesis kit (Agilent). The presence of the mutation was confirmed by gene sequencing. The growth assays were conducted using the B834 strain of *E. coli*. The spot assay was carried out with a serial dilution of culture harvested at the log phase ($A_{600} = 0.6$). LB-Ampicillin plates with 0.1 mM IPTG were used to induce the expression of the different constructs and assess their effect on cell survival. LB-Ampicillin plates with 1% glucose were used as the control. The plasmids bearing wt- and mutant *msdpo4β* gene constructs, and the empty vector were transformed into fresh competent cells of B834, and the transformed cells were grown overnight. Single colonies were picked and used to inoculate a starter culture which was harvested at the log phase ($A_{600} = 0.6$). The following serial dilutions were prepared- 10^{-1}, 10^{-2} and 10^{-3}. For each dilution, 5 μl was spotted on LB ampicillin plates with 0.1 mM IPTG or 1% glucose. The plates were incubated in 37°C for 12 h after which the observations were recorded.

The standard plate assay described by McDonald *et al.* was performed to quantitate the effect of the C47T mutation on bacterial survival (51). The transformed cells were grown to an $OD_{600} = 0.6$, and each culture was induced with 0.1 mM IPTG. Subsequently, the cultures were grown at 37°C for 5 h, following which the cells were plated and incubated overnight. The colonies obtained were counted, and the experiment was repeated three times. The fold change in the number of colony forming units (CFUs) was then calculated with respect to the wild type. The assays were repeated

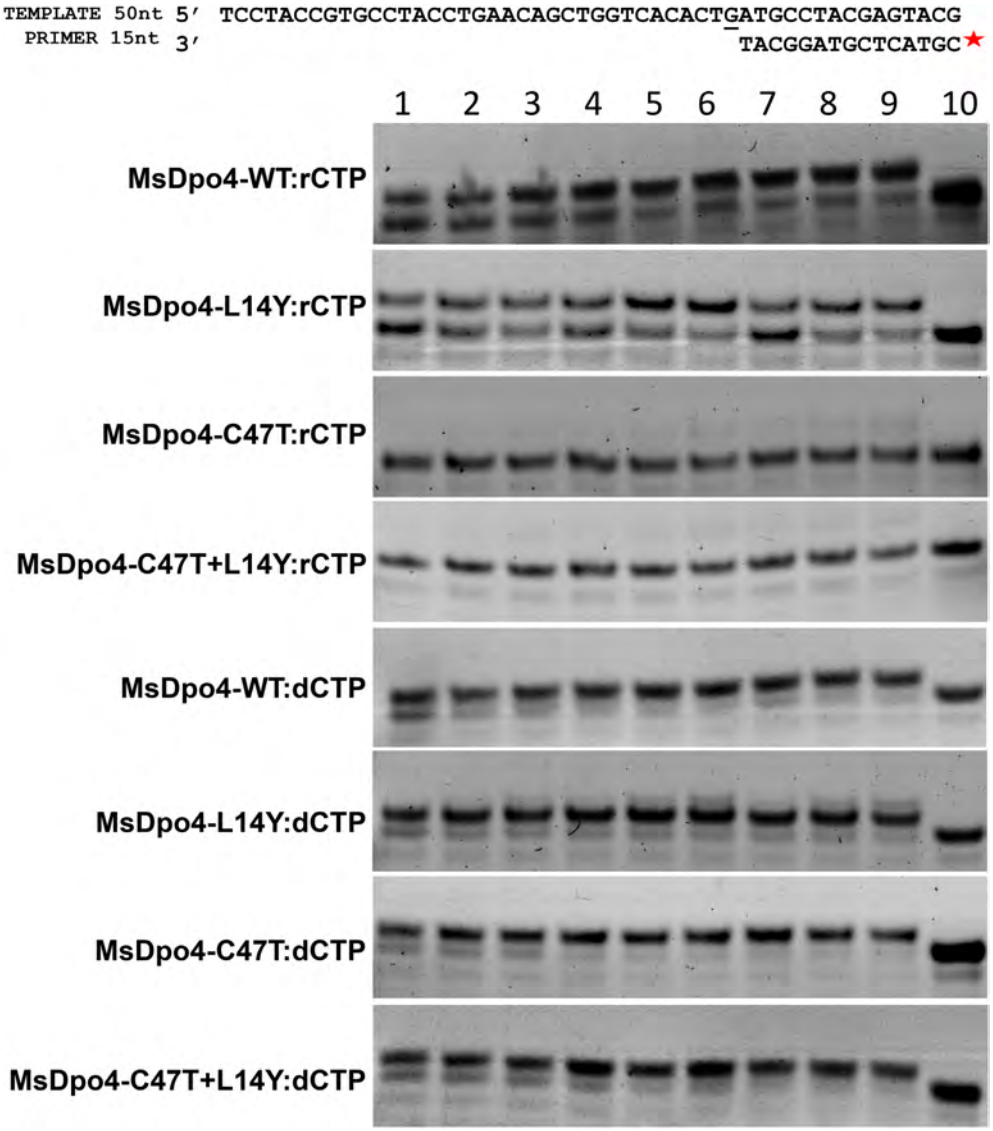

Figure 5. The polar residue reduces rCTP incorporation by MsDpo4. The results of primer extension assays conducted using different concentrations of wt- MsDpo4 and variants (MsDpo4-L14Y, MsDpo4-C47T & MsDpo4-C47T+L14Y) at different concentrations of rCTP or dCTP are shown. The concentrations of enzyme used were 5 nM (Lanes 1, 2 and 3), 10 nM (Lanes 4, 5 and 6) and 15 nM (lanes 7, 8 and 9). The concentrations of incoming nucleotide (rCTP or dCTP) used were 1 μM (lanes 1, 4 and 7), 2 μM (lanes 2, 5, and 8) and 4 μM (lanes 3, 6 and 9). The assays show that mutation of Cys47 to Thr enhances the ability of MsDpo4 to exclude ribonucleotides. The MsDpo4-C47T and MsDpo4-C47T+L14Y proteins show higher selectivity for dCTP than MsDpo4-WT or MsDpo4-L14Y. These experiments shows that engineered residue (Thr47) enhances sugar selectivity.

three times, and the standard deviation values were calculated.

RESULTS

Presence of the steric gate does not abrogate the ability of the MsDpo4 to incorporate ribonucleotides

The steric filter is naturally absent in MsDpo4, and it has a Leu at the 14th position, instead of Tyr or Phe (Figure 1A). Primer extension assays show that MsDpo4 can incorporate both dNTPs and rNTPs (Figure 1B). The aromatic steric gate was engineered in MsDpo4, by mutating the Leu14 residue to Tyr. Surprisingly, primer extension assays showed that the MsDpo4-L14Y enzyme still exhibited substantial ability to incorporate rNTP (Figure 1C). In comparison, PolIV-WT did not show significant incorporation of rNTPs and thus exhibited stringent sugar selectivity (Figure 1D). Steady-state kinetic analysis carried out to compare the ability of MsDpo4-WT and MsDpo4-L14Y to incorporate ribonucleotides showed that addition of the steric gate led to only a six-fold reduction in the catalytic efficiency of rCTP incorporation (Table 1).

To understand the structural basis of ribonucleotide incorporation by MsDpo4-WT and the L14Y mutant, we first determined the structure of the wt- enzyme in complex with DNA bearing dG at the templating position and rCTP or dCTP as the incoming nucleotide to a resolution of 2.5 and 2.3 Å, respectively (Supplementary Table S2A and B). The

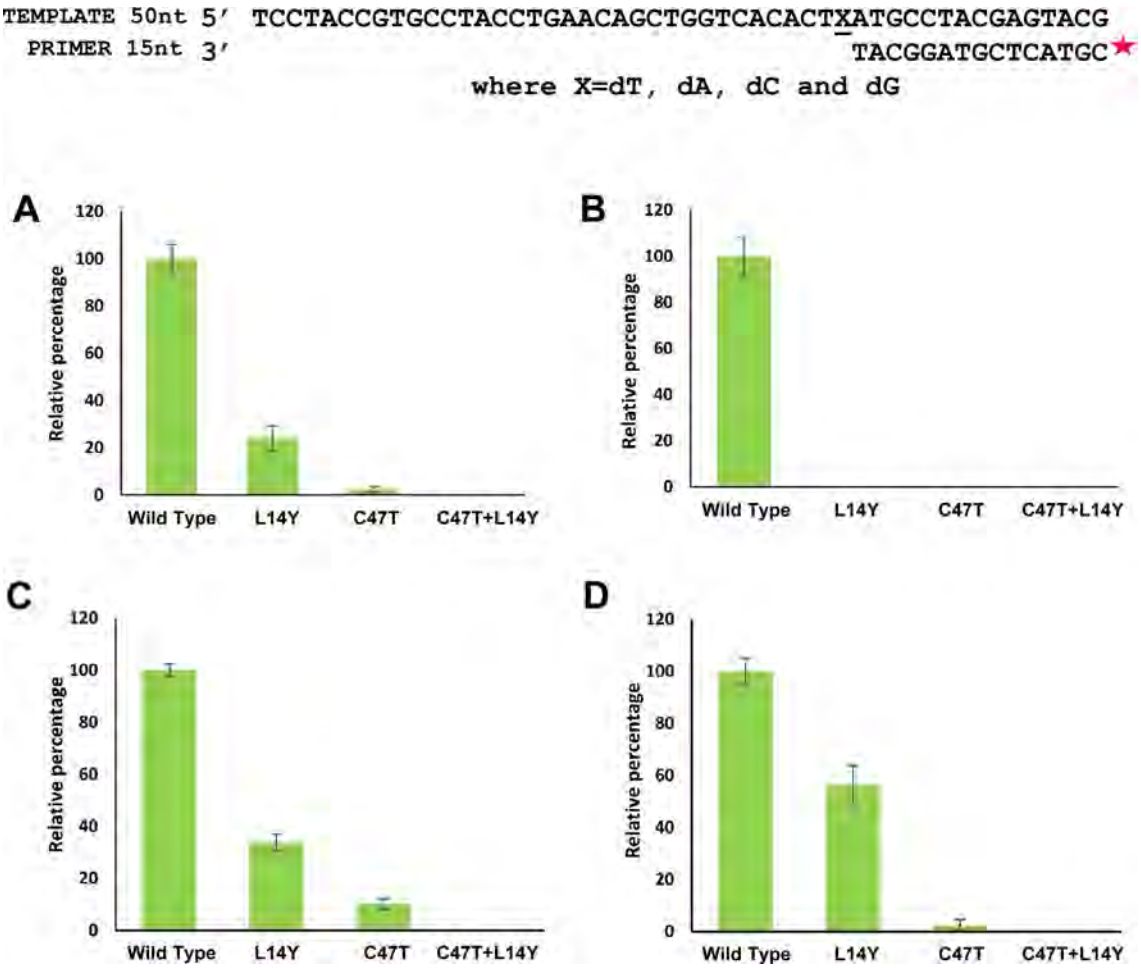

Figure 6. The presence of the polar residue results in drastic reduction in incorporation of all rNTPs. The level of rNTP incorporation by MsDpo4-WT and variants (MsDpo4-C47T, MsDpo4-L14Y & MsDpo4-L14Y+C47T) are displayed. (**A**) The level of incorporation of rATP opposite template dT by MsDpo4-WT and variants is displayed. (**B**) The figure exhibits the level of incorporation of UTP opposite template dA by MsDpo4-WT and variants. (**C**) The level of incorporation of rGTP opposite template dC by MsDpo4-WT and variants is displayed. (**D**) The ability of MsDpo4 and variants to incorporate rCTP opposite dG is displayed. For all graphs, the error bars indicate standard deviation ($n = 3$). For MsDpo4, the presence of the polar filter (C47T) results in a drastic reduction in rNTP incorporation and the presence of both filters in MsDpo4 gives rise to stringent sugar selectivity.

Table 1. Steady-state kinetics of rCTP incorporation by MsDpo4-WT and MsDpo4-L14Y

Protein	k_{cat}[a] (min^{-1})	K_m[a] (μM)	k_{cat}/K_m[a] (μM^{-1} min^{-1})
MsDpo4-WT	0.66 ± 0.07	55.3 ± 4.0	0.01 ± 0.001
MsDpo4-L14Y	0.1 ± 0.01	59.54 ± 13.4	0.0016 ± 0.0003

[a]The numbers after ± denote standard deviation ($n = 3$).

structures for MsDpo4-WT showed that both dCTP and rCTP form Watson-Crick base pairs with the template dG, and are present in the location and conformation compatible with productive catalysis (Figure 2A). The two structures superimposed onto each other with an RMSD of 0.36 Å and showed no difference in the position and conformation of incoming nucleotide . These structures, therefore, show that MsDpo4 can bind rCTP and dCTP without any distortion in enzyme and DNA structure (Figure 2B and C). The structure of MsDpo4-L14Y was also determined with DNA bearing dG at the templating position and dCTP or rCTP as the incoming nucleotide to a resolution of 2.16 and 2.06 Å, respectively (Supplementary Table S2C and D). The structure showed that the rCTP was accommodated in the active site of MsDpo4-L14Y without any steric clash between the aromatic residue and 2′- OH (Figure 3A).

The structures of MsDpo4-L14Y in complex with DNA (template dG) and incoming rCTP or dCTP superimposed onto each other with an RMSD of 0.37 Å (Figure 3B). This comparison showed that, unlike in the case of the wt- enzyme, the incoming rCTP was repositioned. The rCTP is situated slightly upwards such that the C2′ atom of the sugar ring of rCTP is present 1.3 Å above that of the dCTP. Also, the sugar of the rCTP attains a C3′ endo conformation, and due to this, the 3′-OH group is present 2.4 Å above the corresponding location in the case of dCTP (Figure 3B).

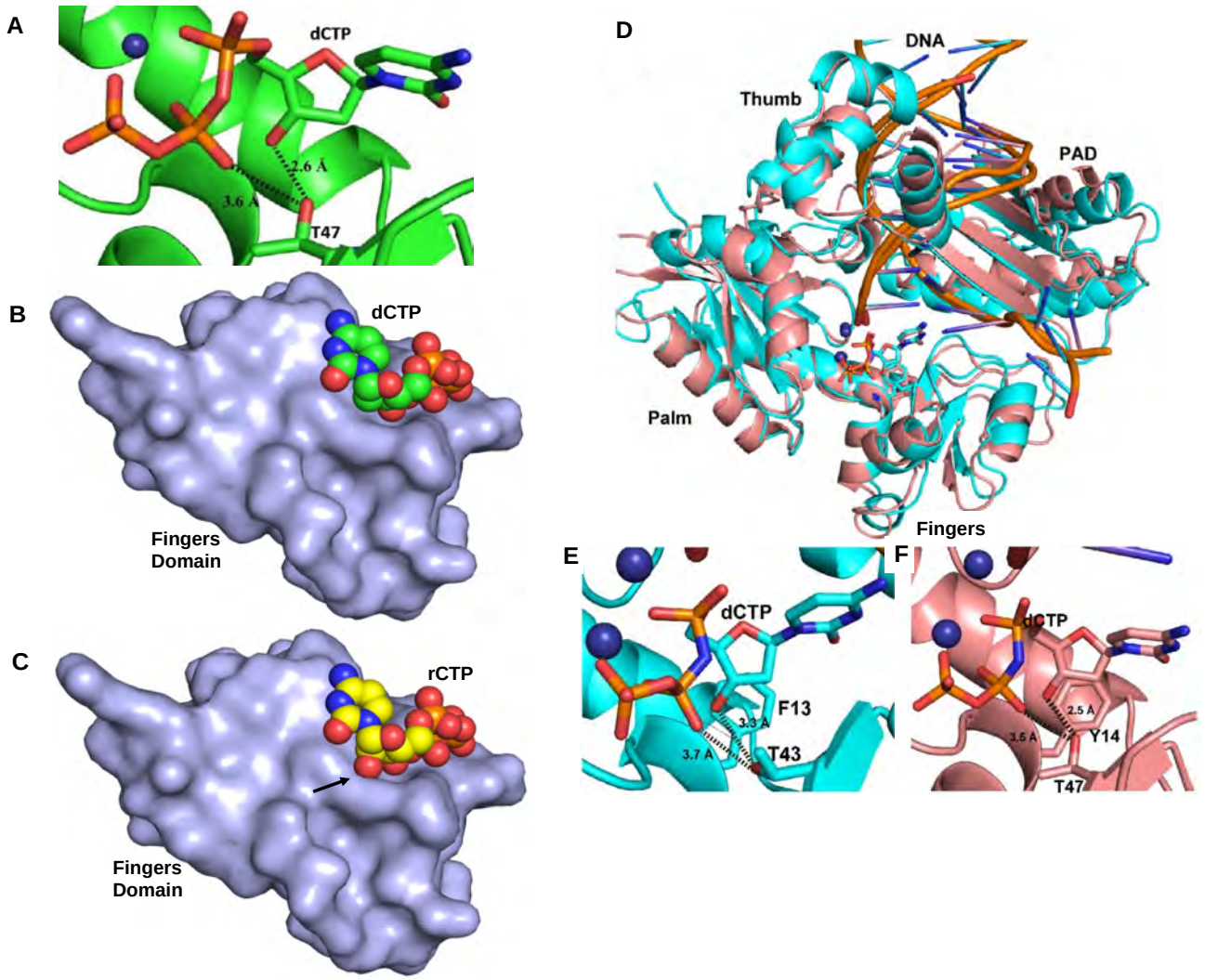

Figure 7. A polar filter aids sugar selectivity. (**A**) The crystal structure of the MsDpo4-C47T$_{DNA(dG):dCTP}$ complex shows that the 3′-OH and the β-phosphate of the incoming nucleotide form interactions with the engineered T47 residue. (**B**) The surface of the fingers domain from the structure of the MsDpo4-C47T$_{DNA(dG):dCTP}$ complex along with the space-filling representation of the incoming dCTP is displayed. The figure shows that the dCTP molecule is accommodated in the catalytic site cavity without any steric clashes. (**C**) The surface of the fingers domain from the structure of the MsDpo4-C47T$_{DNA(dG):dCTP}$ complex along with the space-filling representation of incoming rCTP modeled in the catalytic site is displayed. The 2′-OH of the modeled rCTP forms steric clashes with the surface of the fingers domain (highlighted by an arrow). The engineered residue (Thr47) therefore represents a polar filter that aids sugar selectivity. (**D**) The PolIV$_{DNA(dG):dCTP}$ (Cyan) and MsDpo4-C47T+L14Y$_{DNA(dG):dCTP}$ (peach) structures superimpose with an RMSD of 1.44 Å. The protein structure is shown in cartoon representation, DNA in ribbon form, and ions in the form of spheres. The incoming nucleotide and the steric plus polar filters are shown in stick representation. The dCTP molecules in the two structures are present in identical locations. (**E**) Thr43 forms interactions with the 3′–OH and β-phosphate of incoming dCTP in the PolIV$_{DNA(dG):dCTP}$ structure. (**F**) The engineered T47 residue in the MsDpo4-C47T+L14Y$_{DNA(dG):dCTP}$ structure also forms interactions with the 3′-OH and β-phosphate of the incoming dCTP. Overall, the presence of the steric and polar filters in MsDpo4 lead to similar interactions as that seen in PolIV.

Due to these structural differences, the 2′-OH of the incoming rCTP is present in the same location as the C2′ atom of dCTP and therefore will not clash with the π-electron cloud over the aromatic side chain of the engineered Tyr14 residue (Figure 3C). The location of the triphosphate moiety is similar in the case of rCTP and dCTP and the changes in the position and conformation of the sugar do not adversely affect the ability of the rCTP to form a productive Watson-Crick base pair with the template dG. Overall, due to the repositioning of the rCTP and change in the sugar pucker, rCTP can bind to the MsDpo4-L14Y enzyme in a conformation compatible with catalysis and without any steric clashes with the enzyme surface (Figure 3C). Thus, the comparison shows that the presence of a steric filter is not adequate to achieve stringent sugar selectivity.

The presence of a polar filter prevents ribonucleotide incorporation

Since PolIV exhibits stringent ribonucleotide exclusion, the structures of ternary complexes of MsDpo4-WT and PolIV-WT (PDB ID 4IRC) (40) were compared to identify additional determinants of sugar selectivity. The structures of MsDpo4-WT and PolIV-WT, in complex with DNA (tem-

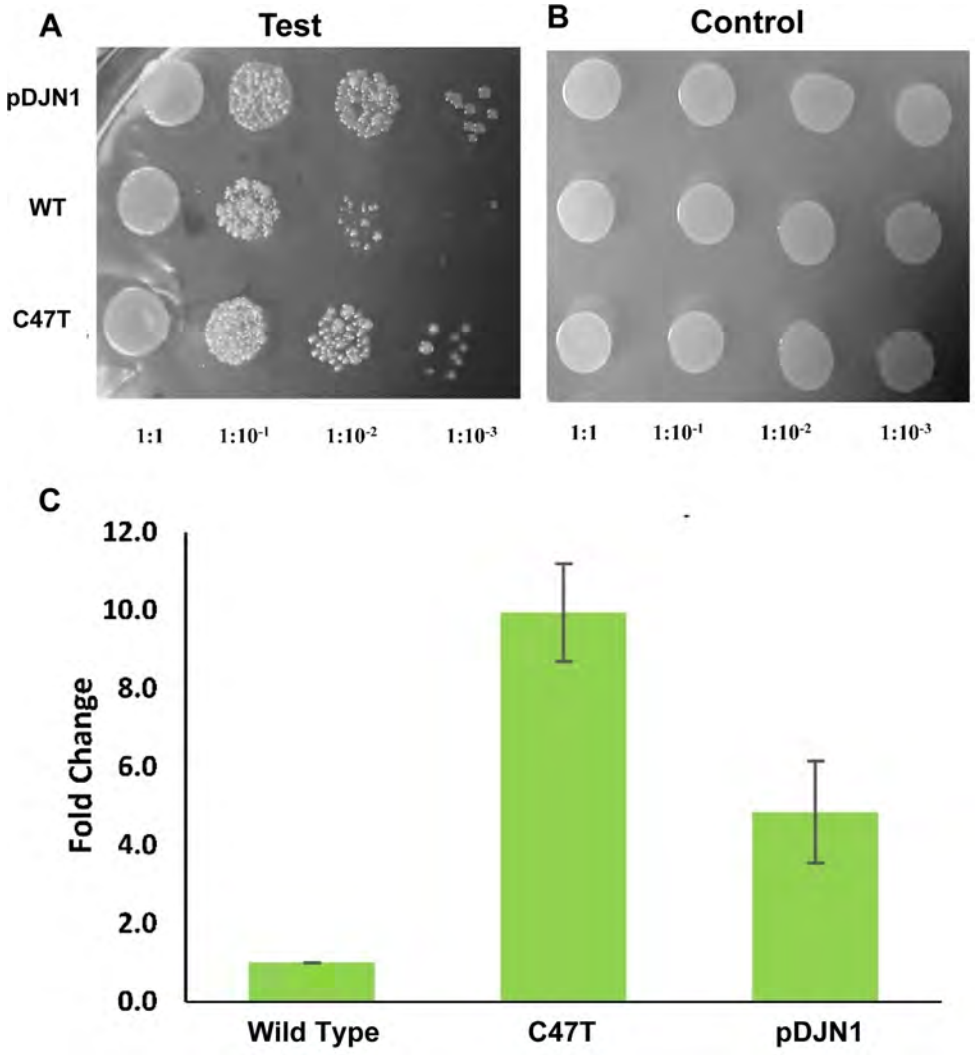

Figure 8. Presence of the polar filter enhances bacterial survival. Spot Assay was conducted to assess the effect of MsDpo4-WT and MsDpo4-C47T on bacterial survival. The pDJN1 (vector alone), MsDpo4-WT and MsDpo4-C47T are shown in rows from top to bottom respectively. (**A**) Serial dilutions of cells expressing MsDpo4-WT and variants were spotted on LB ampicillin plates with 0.1mM IPTG and it was seen that bacteria expressing MsDpo4-C47T exhibit better survival than those expressing MsDpo4-WT.(**B**) Serial dilutions of cells expressing MsDpo4-WT and MsDpo4-C47T were spotted on LB ampicillin plates with 1% glucose and the displayed picture shows that equal amount of cells were spotted for all the samples tested. (**C**) Quantitative growth assay was done to assess the contribution of the polar filter towards bacterial survival. The effect of expression of wt- and mutant versions of MsDpo4 on the survival of the *E. coli* were measured and compared. The cells expressing MsDpo4-C47T exhibited 10-fold higher number of Colony Forming Units than cells expressing MsDpo4-WT. The error bars indicate standard deviation ($n = 3$).

plate dG) and incoming dCTP, were superimposed, and residues located within a distance of 5 Å in the vicinity of dCTP were compared. This comparison showed that at the position equivalent to Cys47 in MsDpo4, there exists a Thr residue in PolIV (Figure 4). This polar residue (Thr43) in PolIV, forms hydrogen bonds with 3′-OH and β-phosphate of the incoming nucleotide (Figure 4B). Since Cys47 in MsDpo4 cannot form equivalent interactions (Figure 4C), it is possible that Thr43 in PolIV may play a role in sugar selectivity.

To test this hypothesis, two additional mutant versions of the MsDpo4 enzyme were prepared- one wherein the Cys47 residue was mutated to Thr, and another with the double mutation C47T+L14Y. It was observed that the MsDpo4-C47T enzyme exhibited substantially higher sugar selectivity than the wt- enzyme. The double mutant of MsDpo4 with both the steric filter and the polar residue showed almost no ribonucleotide incorporation activity at different concentrations of enzyme and nucleotide (Figure 5). The ability of the mutant versions of MsDpo4 to prevent ribonucleotide incorporation was true for all four rNTPs (Figure 6). There was no difference in the ability of the wt- and mutant enzymes to incorporate dNTPs (Supplementary Figure S2).

To ascertain if the engineered polar residue interacts with the incoming dNTP, the structure of MsDpo4-C47T in complex with DNA (with dG in the templating position) and dCTP was determined to a resolution of 1.8 Å (Supple-

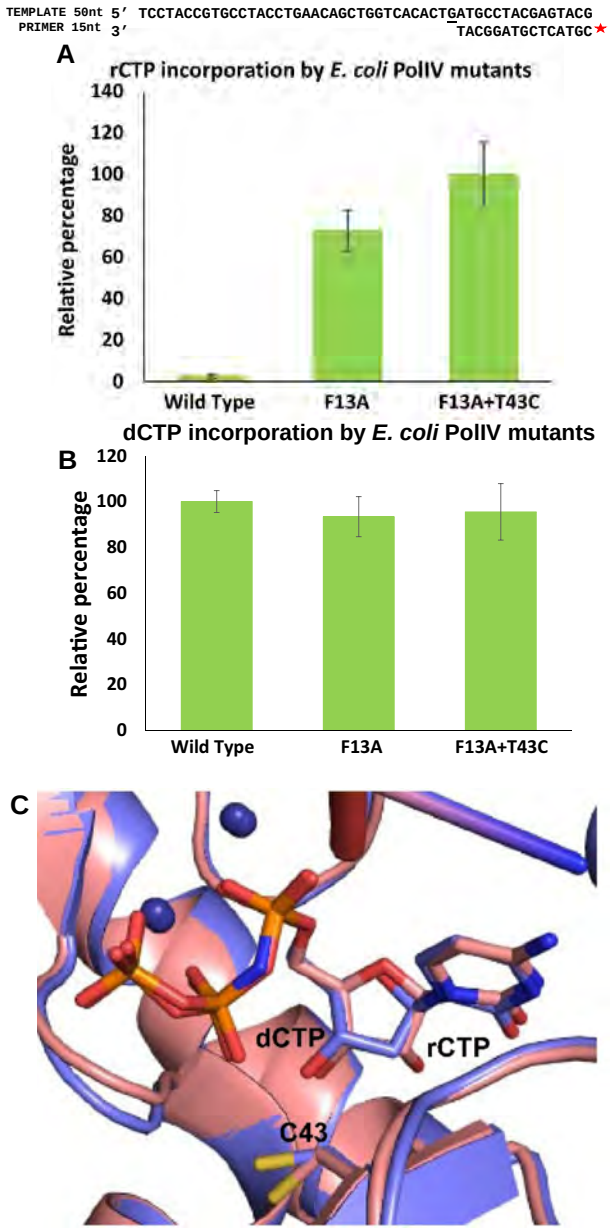

Figure 9. The absence of polar and aromatic steric filter in PolIV results in a drastic reduction in sugar selectivity. (**A**) Primer extension assays showed that the mutant version of PolIV without the steric and polar filter exhibits a 50-fold enhancement in rCTP incorporation. The error bars indicate standard deviation ($n = 3$). (**B**) Primer extension assays showed that the absence of the steric and polar filters did not affect the ability of PolIV to incorporate dCTP. The error bars indicate standard deviation ($n = 3$). (**C**) The PolIV-F13A+T43C$_{DNA(dG):rCTP}$ (peach) structure superimposed with PolIV-F13A+T43C$_{DNA(dG):dCTP}$ (violet) structure with an RMSD of 0.26 Å. The position and conformation of the dCTP and rCTP in the two structures are identical.

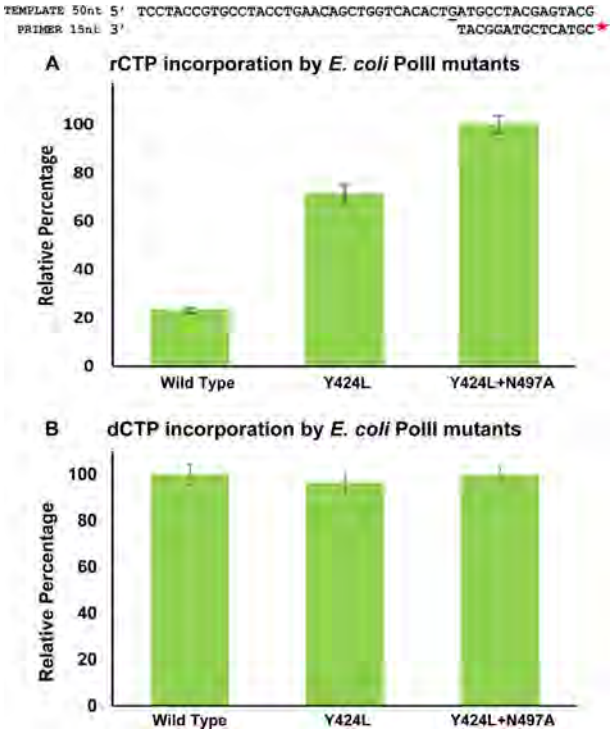

Figure 10. The contribution of polar and steric filters towards sugar selectivity in B-family DNA polymerases. Y424 and N497 represent the steric and polar filters in PolII (*E. coli*), a representative member of the B-family of dPols. (**A**) Primer extension assays were carried out to compare the ability of PolII-WT, PolII-Y424L and PolII-Y424L+N497A to incorporate rCTPs. These assays show that the polar filter contributes substantially towards ribonucleotide exclusion in this family. (**B**) Further, primer extension assays showed that the mutations Y424L and Y424L+N497A did not have a significant effect on the ability of PolII to incorporate dCTPs. For both the graphs, the error bars indicate standard deviation ($n = 3$).

mentary Table S2F). The structure showed that, as in the case of PolIV, the Thr47 residue forms polar interactions with 3′-OH and the β-phosphate of the incoming dCTP (Figure 7A).

The structures of the ternary complexes of PolIV and MsDpo4-C47T show that the interaction with the polar residue holds the incoming dCTP closer to the polymerase surface (Figure 7B). As a result, it is not possible to reposition the incoming nucleotide, such that the 2′-OH of rNTP will not clash with the enzyme surface as was seen in the ternary complex of MsDpo4-L14Y with DNA and rCTP. Thus, the residue represented by Thr43 in PolIV constitutes a polar filter that mediates a closer association of the enzyme with the incoming nucleotide and thus prevents the repositioning required to accommodate rNTP in the active site (Figure 7C).

Further, the structure of MsDpo4-L14Y+C47T in complex with DNA (template dG) and incoming dCTP was determined to a resolution of 2.5 Å (Supplementary Table S2E) and compared with that of the similar complex of PolIV (PDB ID: 4IRC (40)). This comparison showed that the orientation and conformation of incoming nucleotide in the two structures were nearly identical (Figure 7D–F). The presence of the steric and polar filters in MsDpo4, therefore, led to a similar interactions as that for PolIV-WT and complete loss of ribonucleotide incorporation activity.

Presence of polar filter reduces cellular lethality due to ectopic expression of MsDpo4

It has been observed previously that, incorporation of rNTPs into genomic DNA, leads to cellular lethality in *E.*

coli and higher organisms (7,10,13,33,51,52). To assess the importance of the polar filter in ribonucleotide incorporation, growth assays were performed using *E. coli* as a surrogate host. To enable the recruitment of the MsDpo4 protein to the replication fork in *E. coli*, the β-clamp binding motif of PolIV (50) (^{342}QMERQLVLGL^{351}) was added to C-terminus of MsDpo4. Wt- and mutant versions (C47T) of this construct were prepared.

It was seen that when the expression of these constructs was induced in *E. coli*, fewer colonies were obtained in case of wt- protein as compared to the construct bearing the C47T mutation or empty vector (Figure 8A-B). In the quantitative plate assay, cells expressing the C47T mutant exhibited 10-fold higher survival with respect to wild type (Figure 8C). The only difference between the wt- and mutant constructs is the presence of the polar filter in the latter. Hence, the low level of survival observed in the case of wt- enzyme may be due to the higher incorporation of rNTP into the genome.

PolIV lacking polar and steric filter can incorporate ribonucleotides

The variant of MsDpo4 with the polar filter showed substantially enhanced sugar selectivity. To observe the effect of removal of the polar filter in PolIV on ribonucleotide incorporation, the polar residue Thr43 was replaced with Cys (corresponding to Cys47 in MsDpo4) and Phe13 (steric filter) was replaced by Ala. The ability of the mutant proteins (PolIV-F13A and PolIV-F13A+T43C) to incorporate rCTP and dCTP was compared to that of PolIV-WT using primer extension assays. The level of rCTP incorporation by PolIV-F13A+T43C was 50-fold higher than the wt- enzyme; while that of the F13A mutant was ∼35-fold higher (Figure 9A). However, the mutations did not show a significant difference in the ability of PolIV to incorporate dCTP (Figure 9B). The structure of the complex of PolIV-F13A+T43C with DNA (template G) and incoming rCTP or dCTP was determined to a resolution of 2.77 and 2.44 Å, respectively (Supplementary Table S2G and H). The comparison of the two structures shows that the dCTP and rCTP superimpose well on each other (Figure 9C). These structures show that the loss of polar residue helps accommodation of rCTP in a catalytically competent conformation without any steric clash between the 2′–OH and the enzyme surface. Thus, the mutant version of PolIV without the polar and steric filters exhibits a considerably heightened ability to incorporate ribonucleotides as seen in the case of MsDpo4-WT. These experiments also show that the polar and aromatic filters act in tandem to achieve stringent dNTP selectivity in PolIV-WT.

The polar filter is conserved in B and Y family of DNA Polymerases

The structures of ternary complexes of different dPols (enzyme:DNA:dNTP) available in PDB were analyzed to ascertain if the presence of the polar filter is a general feature of dPols. It was observed that the polar and aromatic residues were conserved in the majority of the dPols belonging to Y- and B- families.

For the Y- family, it was seen that the polar filter is represented by a Ser or Thr residue, and a Tyr or Phe residue is present as the steric filter (Supplementary Figure S3A) (53–56) with both filters present in the fingers domain. In the B-family, Asn (fingers domain) represents the polar filter, and Tyr (palm domain) is present as the steric filter (Supplementary Figure S3B) (29,42,57–60).

To validate the role of the polar filter in ribonucleotide exclusion by B-family dPols, DNA polymerase II (PolII) from *E. coli* was utilized as a model enzyme. In this enzyme, Tyr424 is present as the steric filter, and Asn497 is predicted to represent the polar filter (Supplementary Figure S3B3). Primer extension assays showed that, compared to PolII-WT, the mutant proteins PolII-Y424L and PolII-Y424L+N497A exhibit considerably enhanced ability to incorporate ribonucleotides (Figure 10A). The levels of rCTP incorporation by the single and double mutants were 3 and 4.5 times higher than the wild type. In comparison, the ability of the mutant proteins to incorporate dNTPs was unaffected (Figure 10B). The results show that polar residue Asn497 contributes significantly towards ribonucleotide exclusion by PolII. In summary, the studies on MsDpo4, PolIV, and PolII described here point to the presence of a polar filter in dPols that acts in tandem with the steric filter to exclude rNTPs during DNA synthesis.

DISCUSSION

Our results evince that, to achieve stringent sugar selectivity, dPols possess two filters. The first one, known as the steric filter, is composed of a bulky aliphatic or aromatic side chain that is present underneath the ribose sugar and will clash with the 2′-OH of the rNTP molecule to reduce the chances of it occupying the active site in a conformation compatible with productive catalysis. The present study shows the existence of a second polar filter, which interacts with the 3′-OH and triphosphate moiety of the incoming nucleotide and orients it closer to the surface of the fingers domain. As a result, the presence of the polar filter ensures that the 2′-OH of an rNTP molecule will clash with the surface of the fingers domain and consequently there is a limited possibility of the rNTP binding in a catalytically competent conformation. The steric filter is present below the sugar, and the polar filter is present close to the 3′-OH and the triphosphate moiety of the incoming nucleotide. Hence, the filters fall on nearly perpendicular planes to each other, and synergize to ensure high sugar selectivity.

The presence of the steric and polar filters that adopt chemically distinct strategies to prevent ribonucleotide incorporation appears to be a conserved feature of the Y- and B-family of dPols. Polη is a Y-family dPol and exhibits a heightened ability to incorporate ribonucleotides (61). It is believed that this is because it has an extraordinarily spacious active site (30). However, the ability of the enzyme to incorporate ribonucleotides may also be due to the fact that it does not possess a polar filter. An alignment of hPolη with PolIV (RMSD = 2.36 Å) shows that an Ala residue (Ala49) is present at the position equivalent to the Thr43 residue (Supplementary Figure S4). The alignment of Polη incorporating rCTP and dCTP (RMSD = 0.19 Å) shows that the rCTP is repositioned away from the aromatic fil-

ter (Supplementary Figure S4D and E) (30). The presence of a polar filter that interacts with the rNTP will prevent such a repositioning. It has been reported before Polη has a heightened ability to incorporate ribonucleotides, and these observations are in line with the present study (61).

The strategy utilized by Y- and B- family dPols to exclude ribonucleotides is distinct from that observed in the case of A, X and C- family members. In A-family, it is suggested that dual steric filters exist in the form of Glu residue below the sugar of the incoming nucleotide and an aromatic residue lining the fingers domain (19,62–64). The two steric filters are positioned such that the 2′-OH of an rNTP molecule will clash with these residues, thus preventing its incorporation (65,66). Based on the structures of DNA polymerase III (PolIII) from *E. coli and Geobacillus kaustophilus*, C-family dPols may utilize a histidine residue to perform the function of the steric filter. It was seen that the side chains of His760 and His 1275 in PolIII from *E. coli and G. kaustophilus*, respectively (67,68) stack against the ribose sugar of the incoming nucleotide. A number of X-family members exhibit lower selectivity for dNTPs and even though, some members show the presence of an aromatic residue at the position equivalent to the steric filter, the structures show that it does not stack with the ribose sugar (18,69). Instead, it appears that the polypeptide backbone in the region which is present close to the C2′ atom of the sugar acts as a steric filter to provide some level of sugar selectivity (18). For the PrimPol family, it is suggested that the main chain of residues 289–291 in human PrimPol may act as a steric filter to prevent ribonucleotide incorporation (32). Also, Y100 present in the fingers domain, which does not stack with the sugar of the incoming nucleotide is known to play an important role in nucleotide selectivity (70,71). The members of the RT family of dPols do possess the steric filter to reduce ribonucleotide incorporation (72–74). The D-family of dPols represents the least studied family of dPols and strategy utilized by these enzymes to prevent ribonucleotide incorporation is unclear (75).

B-family dPols are highly accurate and responsible for DNA synthesis during replication and repair, especially in eukaryotes (76–80). In humans, members of the B-family include Polα, Polδ, Polε, and Polζ. A thorough examination of different databases such as NIH-NCI GDC data portal, NCBI ClinVar and BIH-DB SNP Short Genetic Variation, revealed a correlation between mutations at the polar filter site and disease in human dPols δ and ζ (Supplementary Table S1). Thermostable members of B-family dPols have been used as enzymes for PCR-based applications (81,82). The results presented in the present study may be useful in the development of engineered enzymes that can directly amplify RNA without the need for cDNA formation. Such engineered dPols may be a valuable tool for a number of applications in biotechnology (83).

DATA AVAILABILITY

The structure factors and the refined co-ordinates have been deposited in the PDB with the following codes: 6JUL MsDpo4-$WT_{DNA(dG):dCTP}$, 6JUM MsDpo4-$C47T_{DNA(dG):dCTP}$, 6JUN MsDpo4-$C47T+L14Y_{DNA(dG):dCTP}$, 6JUO MsDpo4-$L14Y_{DNA(dG):dCTP}$, 6JUP PolIV-$F13A+T43C_{DNA(dG):dCTP}$, 6JUQ PolIV-$F13A+T43C_{DNA(dG):rCTP}$, 6JUR MsDpo4-$L14Y_{DNA(dG):rCTP}$ and 6JUS MsDpo4-$WT_{DNA(dG):rCTP}$.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank the X-ray diffraction facilities located at the Regional Centre for Biotechnology, Faridabad and NII, New Delhi. We thank Dr Hassan Belrhali and Dr Babu Manjashetty (BM14 beamline, ESRF), Dr Nicolas Foos (ID 30B Beamline, ESRF) and Dr Danielle de Sanctis (ID29 beamline, ESRF) for help with X-ray diffraction data collection. We thank Dr Babu Manjashetty and Dr Annie Heroux at for help with data collection at the XRD2 beamline of the Elettra Sincrotrone in Trieste, Italy. We thank Prof. S. Ramaswamy (inStem, Banglore) for critically reading the manuscript.

FUNDING

Science & Engineering Research Board of the Department of Science & Technology, Government of India [EMR/2016/000986]. Data collection at the BM14 beamline of ESRF (Grenoble, France) was supported by the BM14 project—a collaboration between DBT, EMBL and ESRF; Data collection at ID29 and ID30B was facilitated by the ESRF Access Program of RCB which is supported by Department of Biotechnology, Government of India [BT/INF/22/SP22660/2017]. Funding for open access charge: Intramural Funding from Regional Centre for Biotechnology. Access to the XRD2 beamline at Elettra was made possible through grant-in-aid from the Department of Science and Technology, India [DSTO-1668].

Conflict of interest statement. None declared.

REFERENCES

1. Jain,R., Aggarwal,A.K. and Rechkoblit,O. (2018) Eukaryotic DNA polymerases. *Curr. Opin. Struct. Biol.*, **53**, 77–87.
2. Steitz,T.A. (1998) A mechanism for all polymerases. *Nature*, **391**, 231–232.
3. Kottur,J. and Nair,D.T. (2018) Pyrophosphate hydrolysis is an intrinsic and critical step of the DNA synthesis reaction. *Nucleic Acids Res.*, **46**, 5875–5885.
4. Yao,N.Y., Schroeder,J.W., Yurieva,O., Simmons,L.A. and O'Donnell,M.E. (2013) Cost of rNTP/dNTP pool imbalance at the replication fork. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 12942–12947.
5. Nick McElhinny,S.A., Watts,B.E., Kumar,D., Watt,D.L., Lundstrom,E.-B., Burgers,P.M.J., Johansson,E., Chabes,A. and Kunkel,T.A. (2010) Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 4949–4954.
6. Schroeder,J.W., Randall,J.R., Hirst,W.G., O'Donnell,M.E. and Simmons,L.A. (2017) Mutagenic cost of ribonucleotides in bacterial DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 11733–11738.
7. Reijns,M.A., Rabe,B., Rigby,R.E., Mill,P., Astell,K.R., Lettice,L.A., Boyle,S., Letich,A., Keighren,M., Kilanowski,F. *et al.* (2012) Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. *Cell*, **149**, 1008–1022.
8. Mcelhinny,S.A.N., Kumar,D., Clark,A.B., Watt,D.L., Watts,B.E., Lundström,E., Johansson,E., Chabes,A. and Kunkel,T.A. (2010) Genome instability due to ribonucleotide incorporation into DNA. *Nat. Chem. Biolology*, **6**, 774–781.

9. Joyce,C.M. (1997) Choosing the right sugar: How polymerases select a nucleotide substrate. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 1619–1622.
10. Caldecott,K.W. (2014) Ribose-An internal threat to DNA. *Science*, **343**, 260–261.
11. Dahl,J.M., Wang,H., Lazaro,J.M., Salas,M. and Lieberman,K.R. (2014) Kinetic mechanisms governing stable ribonucleotide incorporation in individual DNA polymerase complexes. *Biochemistry*, **53**, 8061–8076.
12. Kennedy,E.M., Amie,S.M., Bambara,R.A. and Kim,B. (2012) Frequent incorporation of ribonucleotides during HIV-1 reverse transcription and their attenuated repair in macrophages. *J. Biol. Chem.*, **287**, 14280–14288.
13. Potenski,C.J. and Klein,H.L. (2014) How the misincorporation of ribonucleotides into genomic DNA can be both harmful and helpful to cells. *Nucleic Acids Res.*, **42**, 10226–10235.
14. Williams,J.S., Lujan,S.A. and Kunkel,T.A. (2016) Processing ribonucleotides incorporated during eukaryotic DNA replication. *Nat. Rev. Mol. cell Biol.*, **17**, 351–363.
15. Watt,D.L., Johansson,E., Burgers,P.M. and Kunkel,T.A. (2011) Replication of ribonucleotide-containing DNA templates by yeast replicative polymerases. *DNA Repair (Amst).*, **10**, 897–902.
16. Egli,M., Usman,N. and Rich,A. (1993) Conformational influence of the ribose 2′-Hydroxyl group: Crystal structures of DNA-RNA chimeric duplexes. *Biochemistry*, **32**, 3221–3237.
17. Mcginness,K.E. and Joyce,G.F. (2002) Substitution of ribonucleotides in the T7 RNA polymerase promoter element *. *J. Biol. Chem.*, **277**, 2987–2991.
18. Brown,J.A., Fiala,K.A., Fowler,J.D., Sherrer,S.M., Newmister,S.A., Duym,W.W. and Suo,Z. (2010) A novel mechanism of sugar selection utilized by a human X-family DNA polymerase. *J. Mol. Biol.*, **395**, 282–290.
19. Astatke,M., Ng,K., Grindley,N.D.F. and Joyce,C.M. (1998) A single side chain prevents Escherichia coli DNA polymerase I (Klenow fragment) from incorporating ribonucleotides. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 3402–3407.
20. Vaisman,A. and Woodgate,R. (2018) Ribonucleotide discrimination by translesion synthesis DNA polymerases. *Crit. Rev. Biochem. Mol. Biol.*, **53**, 382–402.
21. Brown,J.A. and Suo,Z. (2011) Unlocking the sugar 'steric gate' of DNA polymerases. *Biochemistry*, **50**, 1135–1142.
22. Cavanaugh,N.A., Beard,W.A. and Wilson,S.H. (2010) DNA polymerase B ribonucleotide discrimination insertion, misinsertion, extension, and coding. *J. Biol. Chem.*, **285**, 24457–24465.
23. Yang,G., Franklin,M., Li,J., Lin,T. and Konigsberg,W. (2002) A Conserved tyr residue is required for sugar selectivity in a pol alpha DNA polymerase. *Biochemistry*, **41**, 10256–10261.
24. Sherrer,S.M., Beyer,D.C., Xia,C.X., Fowler,J.D. and Suo,Z. (2010) Kinetic basis of sugar selection by a Y-Family DNA polymerase from sulfolobus solfataricus P2 †. *Biochemistry*, **49**, 10179–10186.
25. Ogawa,M., Tosaka,A., Ito,Y., Yoshida,S. and Suzuki,M. (2000) Enhanced ribonucleotide incorporation by an O-helix mutant of Thermus aquaticus DNA polymerase I. *DNA Repair (Amst).*, **485**, 197–207.
26. Randrianjatovo-gbalou,I., Rosario,S., Sismeiro,O., Varet,H., Legendre,R., Coppée,J., Huteau,V., Pochet,S. and Delarue,M. (2018) Enzymatic synthesis of random sequences of RNA and RNA analogues by DNA polymerase theta mutants for the generation of aptamer libraries. *Nucleic Acids Res.*, **46**, 6271–6284.
27. Chen,T., Hongdilokkul,N., Liu,Z., Adhikary,R., Tsuen,S.S. and Romesberg,F.E. (2016) Evolution of thermophilic DNA polymerases for the recognition and amplification of C2′-modified DNA. *Nat. Chem.*, **8**, 556–562.
28. Ordonez,H., Uson,M.L. and Shuman,S. (2014) Characterization of three mycobacterial DinB (DNA polymerase IV) paralogs highlights DinB2 as naturally adept at ribonucleotide incorporation. *Nucleic Acids Res.*, **42**, 11056–11070.
29. Swan,M.K., Johnson,R.E., Prakash,L., Prakash,S. and Aggarwal,A.K. (2009) Structural basis of high fidelity DNA synthesis by yeast DNA polymerase delta. *Nat. Struct. Mol. Biol.*, **16**, 979–986.
30. Su,Y., Egli,M. and Guengerich,F.P. (2016) Mechanism of ribonucleotide incorporation by human DNA polymerase η. *J. Biol. Chem.*, **291**, 3747–3756.
31. Nevin,P., Engen,J.R. and Beuning,P.J. (2015) Steric gate residues of Y-family DNA polymerases DinB and pol kappa are crucial for dNTP-induced conformational change. *DNA Repair (Amst).*, **344**, 1173–1178.
32. Rechkoblit,O., Gupta,Y.K., Malik,R., Rajashankar,K.R., Johnson,R.E., Prakash,L., Prakash,S. and Aggarwal,A.K. (2016) Structure and mechanism of human PrimPol, a DNA polymerase with primase activity. *Sci. Adv.*, **2**, e1601317.
33. Crespan,E., Furrer,A., Rösinger,M., Bertoletti,F., Mentegari,E., Chiapparini,G., Imhof,R., Ziegler,N., Sturla,S.J., Hübscher,U. *et al.* (2016) Impact of ribonucleotide incorporation by DNA polymerases β and lambda on oxidative base excision repair. *Nat. Commun.*, **7**, 10805.
34. Bonnin,A., Lázaro,J.M., Blanco,L. and Salas,M. (1999) A single tyrosine prevents insertion of ribonucleotides in the eukaryotic-type φ29 DNA polymerase. *J. Mol. Biol.*, **290**, 241–251.
35. Moon,A.F., Pryor,J.M., Ramsden,D.A., Kunkel,T.A., Bebenek,K. and Pedersen,L.C. (2017) Structural accommodation of ribonucleotide incorporation by the DNA repair enzyme polymerase Mu. *Nucleic Acids Res.*, **45**, 9138–9148.
36. Delucia,A.M., Grindley,N.D.F. and Joyce,C.M. (2003) An error-prone family Y DNA polymerase (DinB homolog from Sulfolobus solfataricus) uses a 'steric gate' residue for discrimination against ribonucleotides. *Nucleic Acids Res.*, **31**, 4129–4137.
37. Sharma,A., Subramanian,V. and Nair,D.T. (2012) The PAD region in the mycobacterial DinB homologue MsPolIV exhibits positional heterogeneity. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **68**, 960–967.
38. Sharma,A. and Nair,D.T. (2011) Cloning, expression, purification, crystallization and preliminary crystallographic analysis of MsDpo4: a Y-family DNA polymerase from Mycobacterium smegmatis. *Acta Crystallogr. Sect. F*, **67**, 812–816.
39. Ordonez,H. and Shuman,S. (2014) Mycobacterium smegmatis DinB2 misincorporates deoxyribonucleotides and ribonucleotides during templated synthesis and lesion bypass. *Nucleic Acids Res.*, **42**, 12722–12734.
40. Sharma,A., Kottur,J., Narayanan,N. and Nair,D.T. (2013) A strategically located serine residue is critical for the mutator activity of DNA polymerase IV from Escherichia coli. *Nucleic Acids Res.*, **41**, 5104–5114.
41. Jain,D. and Nair,D.T. (2013) Spacing between core recognition motifs determines relative orientation of AraR monomers on bipartite operators. *Nucleic Acids Res.*, **41**, 639–647.
42. Wang,F. and Yang,W. (2009) Structural insight into translesion synthesis By DNA Pol II. *Cell*, **139**, 1279–1289.
43. Kottur,J., Sharma,A., Gore,K.R., Narayanan,N., Samanta,B., Pradeepkumar,P.I. and Nair,D.T. (2015) Unique structural features in DNA polymerase IV enable efficient bypass of the N2 adduct induced by the Nitrofurazone Antibiotic. *Structure*, **23**, 56–67.
44. Battye,T.G.G., Johnson,O., Powell,H.R. and Leslie,A.G.W. (2011) iMOSFLM: a new graphical interface for diffraction- image processing with MOSFLM. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **D67**, 271–281.
45. Adams,P.D., Pavel,V., Chen,V.B., Ian,W., Echols,N., Moriarty,N.W., Read,R.J., Richardson,D.C., Jane,S. and Thomas,C. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **D66**, 213–221.
46. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **D60**, 2126–2132.
47. Mccoy,A.J., Grosse-kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
48. Winn,M.D., Charles,C., Cowtan,K.D., Dodson,E.J., Leslie,A.G.W., Mccoy,A., Stuart,J., Garib,N., Powell,H.R. and Randy,J. (2011) Overview of the CCP 4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **4449**, 235–242.
49. Murshudov,G.N. and Nicholls,R.A. (2011) REFMAC 5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **D67**, 355–367.
50. Bunting,K.A., Roe,S.M. and Pearl,L.H. (2003) Structural basis for recruitment of translesion DNA polymerase Pol IV / DinB to the b -clamp. *EMBO J.*, **22**, 5883–5892.
51. McDonald,J.P., Vaisman,A., Kuban,W., Goodman,M.F. and Woodgate,R. (2012) Mechanisms employed by escherichia coli to

prevent ribonucleotide incorporation into genomic DNA by Pol V. *PLoS Genet.*, **8**, 1–12.
52. Moss,C.F., Rosa,I.D., Hunt,L.E., Yasukawa,T., Young,R., Jones,A.W.E., Reddy,K., Desai,R., Virtue,S., Elgar,G. *et al.* (2017) Aberrant ribonucleotide incorporation and multiple deletions in mitochondrial DNA of the murine MPV17 disease model. *Nucleic Acids Res.*, **45**, 12808–12815.
53. 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 κ encircles DNA: implications for mismatch extension and lesion bypass. *Mol. Cell*, **25**, 601–614.
54. 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.
55. 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.
56. Gaur,V., Vyas,R., Fowler,J.D., Efthimiopoulos,G., Feng,J.Y. and Suo,Z. (2014) Structural and kinetic insights into binding and incorporation of L-nucleotide analogs by a Y-family DNA polymerase. *Nucleic Acids Res.*, **42**, 9984–9995.
57. Berman,A.J., Kamtekar,S., Goodman,J.L., Lázaro,J.M., De Vega,M., Blanco,L., Salas,M. and Steitz,T.A. (2007) Structures of phi29 DNA polymerase complexed with substrate: The mechanism of translocation in B-family polymerases. *EMBO J.*, **26**, 3494–3505.
58. Xia,S., Wang,J. and Konigsberg,W.H. (2013) DNA mismatch synthesis complexes provide insights into base selectivity of a B family DNA polymerase. *J. Am. Chem. Soc.*, **135**, 193–202.
59. Hogg,M., Osterman,P., Bylund,G.O., Ganai,R.A., Lundström,E.B., Sauer-Eriksson,A.E. and Johansson,E. (2014) Structural basis for processive DNA synthesis by yeast DNA polymerase ε. *Nat. Struct. Mol. Biol.*, **21**, 49–55.
60. Baranovskiy,A.G., Duong,V.N., Babayeva,N.D., Zhang,Y., Pavlov,Y.I., Anderson,K.S. and Tahirov,T.H. (2018) Activity and fidelity of human DNA polymerase α depend on primer structure. *J. Biol. Chem.*, **293**, 6824–6843.
61. Mentegari,E., Crespan,E., Bavagnoli,L., Kissova,M., Bertoletti,F., Sabbioneda,S., Imhof,R., Sturla,S.J., Nilforoushan,A., Hübscher,U. *et al.* (2017) Ribonucleotide incorporation by human DNA polymerase eta impacts translesion synthesis and RNase H2 activity. *Nucleic Acids Res.*, **45**, 2600–2614.
62. Joyce,C.M., Potapova,O., Delucia,A.M., Huang,X., Basu,V.P. and Grindley,N.D.F. (2008) Fingers-Closing and other rapid conformational changes in DNA Polymerase I (Klenow Fragment) and their role in nucleotide selectivity †. *Biochemistry*, **47**, 6103–6116.
63. Astatke,M., Grindley,N.D.F. and Joyce,C.M. (1998) How E. coli DNA polymerase I (Klenow fragment) distinguishes between Deoxy- and dideoxynucleotides. *J. Mol. Biol.*, **278**, 147–165.
64. Patel,P.H., Suzuki,M., Adman,E., Shinkai,A. and Loeb,L.A. (2001) Prokaryotic DNA polymerase I: Evolution, structure, and "base flipping" mechanism for nucleotide selection. *J. Mol. Biol.*, **308**, 823–837.
65. Kropp,H.M., Diederichs,K. and Marx,A. (2019) Structural biology the structure of an archaeal B-Family DNA polymerase in complex with a chemically modified nucleotide. *Angew. Chem.*, **58**, 5457–5461.
66. Zahn,K.E., Averill,A.M., Aller,P., Wood,R.D. and Doublié,S. (2015) Human DNA polymerase θ grasps the primer terminus to mediate DNA repair. *Nat. Struct. Mol. Biol.*, **22**, 304–311.
67. Fernandez-leiro,R., Conrad,J., Scheres,S.H.W. and Lamers,M.H. (2015) cryo-EM structures of the E. coli replicative DNA polymerase reveal its dynamic interactions with the DNA sliding clamp, exonuclease and tau. *Elife*, **4**, e11134.
68. Evans,R.J., Davies,D.R., Bullard,J.M., Christensen,J., Green,L.S., Guiles,J.W., Pata,J.D., Ribble,W.K., Janjic,N. and Jarvis,T.C. (2008) Structure of PolC reveals unique DNA binding and fidelity determinants. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 20695–20700.
69. Gosavi,R.A., Moon,A.F., Kunkel,T.A., Pedersen,L.C. and Bebenek,K. (2012) The catalytic cycle for ribonucleotide incorporation by human DNA Pol Lambda. *Nucleic Acids Res.*, **40**, 7518–7527.
70. Díaz-talavera,A., Calvo,P.A., González-acosta,D., Díaz,M., Sastre-moreno,G., Blanco-franco,L., Guerra,S., Martínez-jiménez,M.I., Méndez,J. and Blanco,L. (2019) A cancer-associated point mutation disables the steric gate of human PrimPol. *Sci. Rep.*, **9**, 1–13.
71. Bianchi,J., Rudd,S.G., Jozwiakowski,S.K., Bailey,L.J., Soura,V., Taylor,E., Stevanovic,I., Green,A.J., Stracker,T.H., Lindsay,H.D. *et al.* (2013) PrimPol bypasses UV photoproducts during eukaryotic chromosomal DNA replication. *Mol. Cell*, **52**, 566–573.
72. Gao,G., Orlova,M., Georgadis,M.M., Hendrickson,W.A. and Goff,S.P. (1997) Conferring RNA polymerase Activity to a DNA polymerase: A single residue in reverse transcriptase controls substrate selection. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 407–411.
73. Cases-gonza,C.E. and Mene,L. (2000) Coupling ribose selection to fidelity of DNA synthesis. *J. Biol. Chem.*, **275**, 19759–19767.
74. Beck,J., Vogel,M. and Nassal,M. (2002) dNTP versus NTP discrimination by phenylalanine 451 in duck hepatitis B virus P protein indicates a common structure of the dNTP-binding pocket with other reverse transcriptases. *Nucleic Acids Res.*, **30**, 1679–1687.
75. Sauguet,L., Raia,P., Henneke,G. and Delarue,M. (2016) Shared active site architecture between archaeal PolD and multi-subunit RNA polymerases revealed by X-ray crystallography. *Nat. Commun.*, **7**, 12227–12238.
76. Gómez-Llorente,Y., Malik,R., Jain,R., Choudhury,J.R., Johnson,R.E., Prakash,L., Prakash,S., Ubarretxena-belandia,I. and Aggarwal,A.K. (2013) The architecture of yeast DNA polymerase ζ. *Cell Rep.*, **5**, 1–16.
77. Coloma,J., Johnson,R.E., Prakash,L., Prakash,S. and Aggarwal,A.K. (2016) Human DNA polymerase α in binary complex with a DNA: DNA template - primer. *Sci. Rep.-Nat.*, **6**, 1–10.
78. Johnson,R.E., Klassen,R., Prakash,L. and Prakash,S. (2015) A Major Role of DNA polymerase Delta in replicaton of Both the laeading and Lagging DNA strands. *Mol. Cell*, **59**, 163–175.
79. Jain,R., Rajashankar,K.R., Buku,A., Johnson,R.E., Prakash,L., Prakash,S. and Aggarwal,A.K. (2014) Crystal structure of yeast DNA polymerase epsilon catalytic domain. *PLoS One*, **9**, e94835.
80. Jozwiakowski,S.K., Keith,B.J., Gilroy,L., Doherty,A.J. and Connolly,B.A. (2014) An archaeal family-B DNA polymerase variant able to replicate past DNA damage: occurrence of replicative and translesion synthesis polymerases within the B family. *Nucleic Acids Res.*, **42**, 17–18.
81. Biles,B.D. and Connolly,B.A. (2004) Low-fidelity Pyrococcus furiosus DNA polymerase mutants useful in error-prone PCR. *Nucleic Acids Res.*, **32**, 1–7.
82. Pavlov,A.R., Pavlova,N.V., Kozyavkin,S.A. and Slesarev,A.I. (2004) Recent developments in the optimization of thermostable DNA polymerases for efficient applications. *Trends Biotechnol.*, **22**, 253–260.
83. Blatter,N., Bergen,K., Nolte,O., Welte,W., Diederichs,K., Mayer,J., Wieland,M. and Marx,A. (2013) Structure and function of an RNA-reading thermostable dna polymerase. *Angew. Chem.*, **52**, 11935–11939.

256–266 *Nucleic Acids Research, 2018, Vol. 46, No. 1* *Published online 22 November 2017*
doi: 10.1093/nar/gkx1149

Mechanism of formation of a toroid around DNA by the mismatch sensor protein

Shivlee Nirwal[1,2], Dhananjaya S. Kulkarni[3], Amit Sharma[4], Desirazu N. Rao[3] and Deepak T. Nair[1,*]

[1]Regional Centre for Biotechnology, NCR Biotech Science Cluster, 3rd Milestone, Faridabad-Gurgaon Expressway, Faridabad, 121001, Haryana, India, [2]Manipal University, Manipal, 576104, Karnataka, India, [3]Department of Biochemistry, Indian Institute of Science, Bangalore, 560012, Karnataka, India and [4]TERI-DIAKEN Nanobiotechnolgy Centre, TERI-Gram, Gual Pahari, Gurgaon-Faridabad Road, Gurgaon, 122001, Haryana, India

Received April 03, 2017; Revised October 26, 2017; Editorial Decision October 27, 2017; Accepted November 01, 2017

ABSTRACT

The DNA mismatch repair (MMR) pathway removes errors that appear during genome replication. MutS is the primary mismatch sensor and forms an asymmetric dimer that encircles DNA to bend it to scan for mismatches. The mechanism utilized to load DNA into the central tunnel was unknown and the origin of the force required to bend DNA was unclear. We show that, in absence of DNA, MutS forms a symmetric dimer wherein a gap exists between the monomers through which DNA can enter the central tunnel. The comparison with structures of MutS–DNA complexes suggests that the mismatch scanning monomer (B_m) will move by nearly 50 Å to associate with the other monomer (A_m). Consequently, the N-terminal domains of both monomers will press onto DNA to bend it. The proposed mechanism of toroid formation evinces that the force required to bend DNA arises primarily due to the movement of B_m and hence, the MutS dimer acts like a pair of pliers to bend DNA. We also shed light on the allosteric mechanism that influences the expulsion of adenosine triphosphate from A_m on DNA binding. Overall, this study provides mechanistic insight regarding the primary event in MMR i.e. the assembly of the MutS–DNA complex.

INTRODUCTION

The mismatch repair (MMR) pathway removes errors that appear during replication (1–4). MMR pathway initiates correction of mispairs, insertions or deletions that have escaped proofreading during DNA replication. Perturbation of this pathway results in a 100- to 1000-fold enhancement in the frequency at which mutations appear in *Escherichia coli* (2). The fact that the onset of Hereditary Non Polyposis Colorectal Cancer, Oral Squamous Cell Carcinoma and gliomas in humans are strongly co-related with mutations in MMR genes, highlights the importance of this pathway (5–8). Among specific proteins of the MMR system, MutS is responsible for detecting non-Watson–Crick base pairs in newly synthesized DNA (9). Structural and biophysical studies have shown that the functional form of MutS and orthologues is an oval, disc-shaped, asymmetric dimer (10–14). Two monomers associate to form an oval disc with a central channel into which DNA is loaded to be scanned for mismatches. The structures of MutS from *Thermus aquaticus* (TqMutS) and *E. coli* (EcMutS) in complex with DNA bearing a mismatch show that bound DNA is encircled and bent, and these attributes are vital for recognition of mismatches (11–13). MutS is an asymmetric dimer in the DNA-bound state and only one of the monomer makes base-specific contacts through the minor groove of DNA (monomer B) while the corresponding residues from the other monomer are involved in non-pecific contacts with the DNA phosphate backbone (monomer A). Single molecule microscopy studies on TqMutS suggest that the MutS monomers exhibit vibrational motion in the absence of DNA, which reduces considerably on adenosine triphosphate (ATP) binding (15,16) and that MutS translocates on DNA to scan for mismatches (17,18). However, it is still not known how the MutS–DNA complex is assembled. In the functional state of MutS, DNA is encircled by MutS dimer and it is unclear how DNA is loaded into the scanning tunnel of the MutS clamp. The processivity factor involved in DNA replication (dimeric β-clamp in prokaryotes and trimeric PCNA in eukaryotes) and the Ku protein (involved in double-stranded break repair) represent other toroidal proteins that encircle DNA (19,20). In the case of the processivity factor, a five subunit clamp loader enzyme opens the clamp dimer/trimer to load DNA into the central channel (21). The Ku protein is thought to slide onto DNA at the end generated by a double-stranded break (22). There is no known enzymatic activity that aids

*To whom correspondence should be addressed. Tel: +91 129 284 8844; Fax: +91 129 284 8803; Email: deepak@rcb.res.in

loading of the MutS clamp on DNA and therefore unlike β-clamp/PCNA, this enzyme may not require any *trans* factors. To understand the mechanism of assembly of the MutS–DNA complex, we have utilized the MutS homologue from *Neisseria gonorrhoeae*- named NgoS- as a model enzyme. NgoS has been shown to participate in MMR in *Neisseria* through detection of mismatches and small insertions or deletions (23).

In the present study, we have obtained structural information regarding the NgoS dimer in complex with Adenosine Diphosphate (ADP) and in the absence and presence of DNA using Macromolecular Crystallography (MX) and small angle X-ray scattering (SAXS), respectively. The structural analysis shows that in the absence of DNA, there exists a gap between the monomers through which DNA can enter the central tunnel. On entry, the DNA probably associates non-specifically with one monomer (A_m) and this designates the other monomer as the mismatch scanning monomer (B_m), which moves by nearly 50 Å to associate with A_m. Consequently the N-terminal domains of both monomers press onto DNA to bend it. In line with this mechanism, Dynamic Light Scattering (DLS) studies show that NgoS undergoes a compaction in the presence of DNA and that ATP binding, and not hydrolysis, is critical for this transition. Structural analysis and fluorescence anisotropy studies show that DNA binding involves expulsion of ATP from the A_m monomer. Overall, the study provides insight regarding the structural transitions in MutS on DNA binding that enable formation of the dimer clamp around DNA. The study also sheds light on the allosteric mechanism that influences the dynamics of ATP binding in the presence of DNA.

MATERIALS AND METHODS

Cloning, expression and purification of WT and mutant NgoS

The gene segment corresponding to the *mutS* gene from *N. gonorrhoeae was* amplified by polymerase chain reaction using primer sequences derived from the 5′- and 3′- ends of the open reading frames obtained from the genomic DNA. The amplified product (*ngoS*) was cloned in pGEX_6P1, using the BamHI/XhoI sites to produce a fusion polypeptide with N-terminal Glutathione S-transferase (GST) tag attached to the NgoS by a linker with the *PreScission* protease site. The authenticity of the cloned gene was confirmed by DNA sequencing. Then a truncated construct NgoSΔC814 (1–814) was generated by mutating codon corresponding to 815th residue to stop codon, referred as GST-NgoS in the paper. The GST-NgoS fusion protein was over expressed in the *Escherichia coli* C41(DE3) strain. Five litres of LB medium (100 μg/ml ampicillin) were induced with 0.1 mM of Isopropyl β-D-1-thiogalactopyranoside for 18 h at 18°C at 150 rpm. Cells were harvested by centrifugation and resuspended in buffer A [25 mM Tris–Cl (pH 8.0 at 4°C), 500 mM NaCl, 5% glycerol, 2 mM dithiothreitol, 0.01% IGEPAL CA-630 and 1 mM phenylmethylsulfonyl fluoride]. Cells were lysed by sonication followed by centrifugation at 17 000 rpm to remove cell debris. The recombinant protein was purified from the supernatant primarily using GST-sepharose (GE Healthcare Inc.). The column was pre-equilibrated with buffer B [Tris–Cl (pH 8.0 at 4°C), 250 mM NaCl, 5% glycerol, 2 mM dithiothreitol (DTT) and 0.01% IGEPAL CA-630]. Lysate was loaded on to column followed by washing with buffer C [Tris–Cl (pH 8.0 at 4°C), 1M NaCl, 5% glycerol, 2 mM DTT and 0.01% IGEPAL CA-630]. Protein was eluted using 15 mM reduced glutathione in buffer B and incubated with *PreScission* protease to cleave the tag and release the native protein (NgoS). The cleaved protein was analyzed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE), concentrated and further purified by size exclusion chromatography. Superdex-200 column (GE Healthcare Inc.) was used with buffer containing 25 mM Tris–Cl (pH 8.0 at 4°C), 500 mM NaCl, 5% glycerol and 2 mM dithiothreitol. Eluted fractions were analyzed for purity by SDS-PAGE and fractions corresponding to pure protein were pooled and concentrated to 15 mg/ml, aliquoted, flash frozen and stored at −80°C.

The catalytic mutant for the ATPase activity of NgoS was generated by site directed mutagenesis of the catalytic residue E687, in the Walker B motif, to alanine. The resulting mutant protein was purified as mentioned above for wild-type (WT) NgoS.

Estimation of ATPase activity

The ATPase activity of NgoS WT and its mutant NgoS (E687A) was measured using an EnzChek Phosphate Assay Kit (Invitrogen) in 96 well clear bottom Costar® plates (Corning). NgoS WT and mutant (1 μM) was added to the reaction buffer with purine nucleoside phosphorylase enzyme and 2-amino-6-mercapto-7-methylpurine riboside substrate (as described in the kit), followed by an incubation for a period of 10 min at 22°C. The reaction was then started by adding ATP (2 mM) and absorbance was measured continuously at 360 nm at each 3-min interval. A standard curve was generated using K_2PO_4, as the source of inorganic phosphate, provided with the kit. The P_i released during the reaction was calculated using the standard curve thereby calculating the rate of reaction and initial velocity. The experiment was done in triplicate for both WT and the mutant.

Binding of WT and E687A mutant to heteroduplex DNA

Fluorescence anisotropy was measured for both NgoS WT and NgoS(E687A) using 31mer heteroduplex DNA with G:T mismatch. The 3′ end of one of the oligonucleotide in the duplex was labelled with 6-fluorescein amidite (6-FAM), procured from Keck Biotechnology Resource Laboratory (Yale University).

- **5′ GCGATGTTGCTGACGCTGGTGCCTGGCAGCT -6FAM 3′**
- **3′ CGCTACAACGACTGTGACCACGGACCGTCGA 5′**

The protein in varying concentration (0–100 nM) was incubated with 10 nM of respective DNA in a binding buffer with 25 mM Tris pH 8.0, 5% glycerol, 1 mM DTT, 150 mM NaCl at room temperature for 45 min. All the reactions were carried out in 96 well black bottom Costar® plates

(Corning), in triplicates. SpectraMax M5 micro plate reader (Molecular Devices) was used for measuring fluorescence anisotropy. Fluorescence anisotropy was monitored at the excitation wavelength of 492nm and emission wavelength of 517 nm and calculated by the Spectramax software using the following equation

$$FA = [(I//) - (GxI\perp)]/[(I//) + (2GxI\perp)]$$

wherein I// = Fluorescence intensity parallel; I⊥ = Fluorescence intensity perpendicular; G = G factor, (= 1.000 for the given calculations)

Average anisotropy was calculated from three independent measurements, for each protein concentration. Reduced anisotropy values for each protein concentration were calculated by subtracting the averaged anisotropy value for without protein reaction from the respective averaged anisotropy value. The reduced anisotropy values was then plotted (on *Y*-axis) against protein concentration (on *X*-axis). The reduced anisotropy was fitted to calculate the K_d values using the following logistic equation.

$$y = A2 + [(A1 - A2) / (1 + (x/x0)^p]$$

where, A2 is maximum reduced anisotropy, A1 is minimum reduced anisotropy, x0 is the *x*-value of the point of inflection, corresponding to the K_d and p is the Hill's slope of the curve.

Crystallization

NgoS (0.1 mM) was incubated with ADP (0.2 mM) or AMPPNP (0.2 mM) in presence of $MgCl_2$ (1 mM) for 2 min at 25°C followed by incubation for 30 min at 4°C for crystallization trials. Extensive screening and optimization was carried out and finally large single crystals were obtained in 10–14% (w/v) PEG 4000, 0.1M MES buffer (pH 6.5), 10–14% (v/v) glycerol and 0.1–0.2 M $MgCl_2$. Crystals were cryoprotected by soaking into reservoir solution with 5,10,15 and 20% ethylene glycol for 1 min followed by flash freezing in liquid nitrogen.

Structure determination

X-ray diffraction data was collected at BM14 beamline of the European Synchrotron Radiation Facility. NgoS-ADP and NgoS-AMPPNP crystals diffracted to a resolution of 2.97 Å and 3.3 Å, respectively. The collected data were processed and merged using data processing programs like HKL3000 (24). The structure was determined by the molecular replacement (MR) method using a monomer from the available $EcMutS_{DNA:ADP}$ complex structure (PDB code: 1E3M) as a search model. MR with Phaser provided a solution with two molecules in the asymmetric unit (25). Iterative rounds of crystallographic refinement using the PHENIX program and model building using Coot was carried out till convergence of the R-factors (26,27). Finally, Translation-Libration-Screw (TLS) refinement was carried out with the different domains defined as discrete rigid units. The final R_{free} and R_{work} values are 26.8 and 23.4%, respectively for the NgoS–ADP complex. The final R_{free} and R_{work} values are 25.9 and 21.2%, respectively for the NgoS–AMPPNP complex. All structural comparisons were done using the Superpose tool in Coot and the interactions between the clamp regions were mapped using CONTACT in CCP4 (27,28). All figures were prepared using PyMol (Schrödinger Corp.) and the movies were prepared in PyMol (Schrödinger Corp.) and Chimera (29).

Small angle X-ray scattering

Small angle X-ray scattering data were collected at BM29 BioSAXS beamline, ESRF, Grenoble. SAXS measurements were carried out for the $NgoS_{DNA:ADP}$ complex in buffer containing 25 mM TRIS pH 8.0, 150 mM NaCl, 5% Glycerol and 1 mM DTT, with ADP (0.15 mM) and MgCl2 (0.15 mM). The data were acquired at three protein concentrations (2, 3.5 and 5 mg/ml) for the $NgoS_{DNA:ADP}$ complex (DNA concentration being 1.2 molar excess of NgoS dimer). The samples and buffer blanks were centrifuged prior to the data collection at 20 000 *g* for 30 min at 4°C to remove any possible aggregates in the sample. Data collection was carried out at 10°C and a set of 10 measurements were made per sample. Buffer scattering was also collected prior to each sample measurement and was used to generate buffer subtracted intensity profiles. The data was processed using the Primus software (30) available in the data analysis suite ATSAS2.8.1 (31). The software helps to calculate two SAXS invariants, the Radius of gyration (R_g) and the Intensity at zero scattering angle ($I(0)$) using Guinier analysis. The low q range data limited by $q.R_g < 1.3$ was fitted using the straight line plot, ln $[I(q)]$ versus q^2, where q represents the scattering vector ($q = 4\pi \sin\theta/\lambda$) at scattering angle 2θ and wavelength λ of the X-ray. R_g was estimated from the slope ($R_g^{2/3}$) of the fit and $I(0)$ is the Intensity estimation where q is zero. R_g and $I(0)$ were also estimated using the GNOM program (32) that help to evaluate molecular size by plotting the pair distance distribution function (PDDF), $P(r)$, for the scattering data. $P(r)$ gives information about the shape of molecule in real-space, and approaches zero at its maximum dimension, D_{max}.

Results from GNOM were then used as inputs to generate 12 independent *ab initio* bead models of $NgoS_{DNA-ADP}$ complexes through DAMMIF software (33). These models were subsequently superimposed and averaged using DAMAVER (34). For the docking of $NgoS_{DNA:ADP}$ in to the SAXS-derived model, the structure was modelled in Coot (35) using PDB 1E3M, followed by energy minimization in Discovery Studio Client 3.5 software tool. SITUS2.8.1 was used to fit the energy minimized homology model of $NgoS_{DNA:ADP}$ into the SAXS envelope derived from the averaged bead model (36,37). Figures were generated using VMD (38).

Dynamic light scattering

DLS experiments were performed using a Zetasizer Nano ZS90 unit (Malvern Instruments Ltd). All measurements were carried out using a 12 μl cuvette at 298K. NgoS was diluted from the stock solution in buffer with 25 mM Tris–Cl (pH 8.0), 150 mM NaCl, 5% glycerol and 2 mM DTT to a concentration of 0.75 mg/ml. ADP or AMPPNP was added to a final concentration of 0.15 mM. To probe the

effect of DNA, NgoS was mixed with the following 30mer dsDNA heteroduplex (containing a G:T mismatch)

- **5′AGCTGCCAGGCACCAGTGTCAGCGTCCTAT 3′**
- **3′TCGACGGTTCGTGGTCACAGTCGCAGGATA 5′**

and homoduplex (without mismatch).

- **5′AGCTGCCAGGCACCAGTGTCAGCGTCCTAT 3′**
- **3′TCGACGGTCCGTGGTCACAGTCGCAGGATA 5′**

Protein and DNA was mixed in the ratio of 1:1.2 in the presence of 0.15 mM ADP and AMPPNP and incubated for 2 min at 25°C followed by incubation for 30 min at 4°C. The buffer, apo- protein or protein–DNA solutions were subjected to centrifugation at 14 000 rpm for 45 min immediately before measurements. The data were acquired, processed and analysed using Zetasizer software associated with the instrument to calculate hydrodynamic radius R_h values (using the Stokes Einstein equation).

Fluorescence anisotropy measurement to probe expulsion of ATP

The protein in varying concentration (0–500 nM) was incubated with 10 nM of N^6-(6-Amino)hexyl-ATP-6-FAM (Jena Bioscience GmbH) in a binding buffer (25 mM Tris–Cl pH 8.0, 5% glycerol, 500 mM NaCl and 2 mM DTT) to determine the concentration of protein NgoS where all the FAM-labelled ATP is saturated with NgoS, that turned out to be 100 nM of NgoS. Further measurements were carried out with 100 nM of NgoS.

All the reactions were carried out in 96 well black bottom Costar® plates (Corning), in triplicates. SpectraMax M5 micro plate reader (Molecular Devices) was used for measuring fluorescence anisotropy. Fluorescence anisotropy was monitored at the excitation wavelength of 492nm and emission wavelength of 517 nm and calculated by the Spectramax software from the following equation

$$FA = [(I//) - (GxI\perp)]/[(I//) + (2GxI\perp)]$$

wherein I// = Fluorescence intensity parallel, I⊥ = Fluorescence intensity perpendicular, G = G factor, (= 1.000 for the given calculations).

For monitoring the change in anisotropy in absence and presence of DNA, 100 nM of NgoS was mixed with 10 nM FAM labelled ATP alone or in presence of a 60mer dsDNA containing a G:T mismatch and a homoduplex dsDNA respectively (NgoS: DNA = 1:1.2), in the binding buffer and incubated for 90 min on ice.

- **5′CCCCTGACTCAGTTATCGTTCGTCAAGTCGAG CTGCCAGGCACCAGTGTCAGCGTCCTAT3′**
- **3′GGGGACTGAGTCAATAGCAAGCAGTTCAGCTC GACGGTTCGTGGTCACAGTCGCAGGATA5′**
- **5′CCCCTGACTCAGTTATCGTTCGTCAAGTCGAG CTGCCAGGCACCAGTGTCAGCGTCCTAT3′**
- **3′GGGGACTGAGTCAATAGCAAGCAGTTCAGCTC GACGGTCCGTGGTCACAGTCGCAGGATA5′**

The raw average anisotropy was calculated in absence and presence of DNA. The average anisotropy of free FAM-labelled ATP was also determined. Then the absolute FA in presence of DNA and absence of DNA was calculated by subtracting the FA for FAM-labelled ATP alone from the corresponding raw values

To obtain the percentage drop in anisotropy on DNA binding, the following equation was utilized,

$$\text{Percentage of anisotropy} = \frac{\text{FA in the presence of DNA}}{\text{FA in the absence of DNA}} \times 100$$

Table 1. Crystallographic data and refinement statistics for $NgoS_{ADP}$ complex

Data collection statistics	
Wavelength (Å)	0.97926
Space Group	$P2_122_1$
Cell Constants (Å)	89.1 101.6 235.2
Resolution (Å)	2.97 Å (3.047–2.97)
R_{merge}	6.4 (56.1)
I/σI	8.6 (2.0)
Completeness (%)	98.7 (99.2)
Redundancy	3.5 (3.6)
Refinement	
Resolution (Å)	50–2.97 Å
No. of Reflections	43285
R_{work}/R_{free}	23.4/26.8
No. of atoms	
Protein	11810
ADP	54
R. M. S. deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.449
Average B-factors	
Protein	70.1 (Am); 108.0 (Bm)
ADP	55.4 (Am); 83.6 (Bm)

[a]Values in parentheses are for the highest resolution shell.
[b]$R_{merge} = \Sigma|I–\langle I\rangle|/\Sigma I$, where, I is the integrated intensity of a given reflection.
[c]$R_{work} = \Sigma||F_{obs}| - |F_{calc}||/\Sigma\ |F_{obs}|$. R_{free} was calculated using 5% of data excluded from refinement.

RESULTS

Structural analysis of the $NgoS_{ADP}$ and $NgoS_{DNA:ADP}$ complexes

The E687A mutant showed very low ATPase activity compared to WT protein (Supplementary Figure S1A). The purified NgoS and the E687A mutant proteins were able to bind to heteroduplex DNA bearing a G:T mismatch (Supplementary Figure S1B). The structure of NgoS was determined bound to ADP and in the absence of DNA (Table 1). The structure shows that NgoS exists as a dimer and each monomer is composed of different domains (Figure 1A) namely N-Terminal Domain (NTD- residues 1–118), Central Domain (residues 119–403 and 534–559), Clamp region (residues 404–533) and C-Terminal domain (CTD- residues 560–785). There was clear electron density for the bound ADP molecule (Supplementary Figure S2) and is held in place due to stabilizing interactions formed with residues from the CTD (Figure 1B) including those belonging to the Walker A motif (residues 607–615: GPMGGKST). The previously determined structures of MutS–DNA complexes show that Clamp regions and CTDs from each monomer are in contact with each other to form the large central channel in which substrate DNA is loaded (11,13).

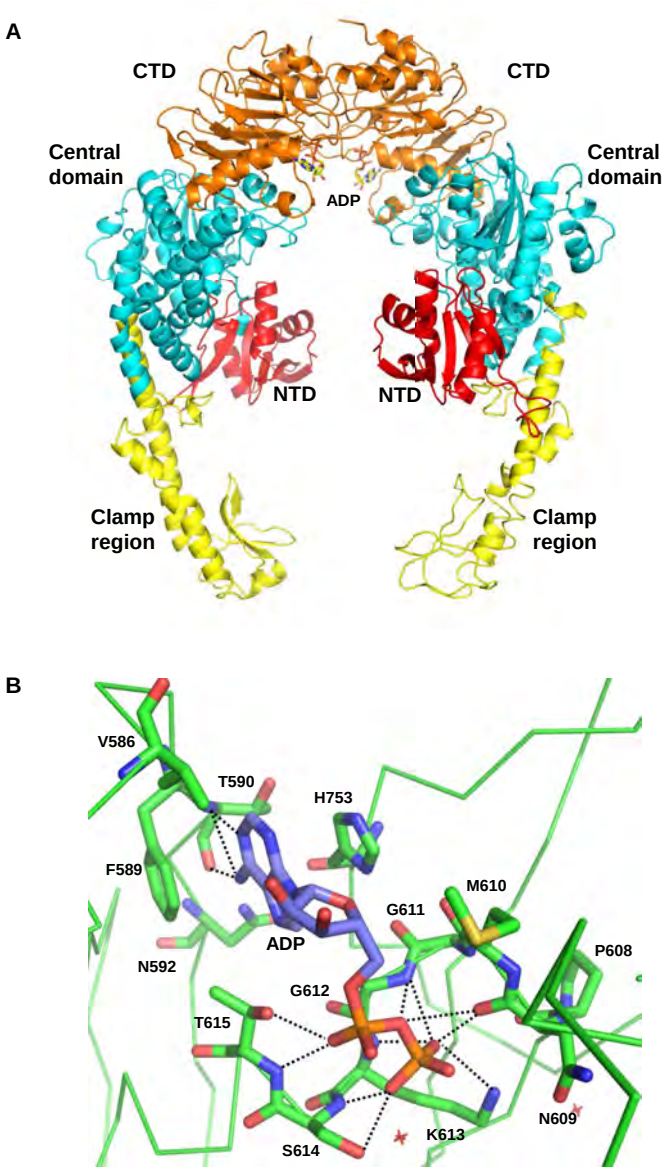

Figure 1. Structure of $NgoS_{ADP}$ complex: (**A**) Structure of $NgoS_{ADP}$ complex: The overall structure of the $NgoS_{ADP}$ complex is displayed. The N-terminal domain (NTD), Central Domain, Clamp region and the C-terminal Domain (CTD) are coloured red, cyan, yellow and brown, respectively. The ADP molecule bound to each monomer is shown in stick representation and coloured according to atom. (**B**) Interactions of ADP with NgoS: The residues of NgoS that form Van der Waals and hydrogen bonding interactions with the ADP molecule are displayed. The backbone of NgoS is displayed as ribbon and the interacting residues are shown in stick representation with C- atoms of NgoS coloured in green and that of ADP coloured in purple. The hydrogen bonds are displayed as dashed lines. The stretch 607–615 corresponds to the Walker A motif or the P-loop of NgoS.

SAXS was also used to obtain structural information regarding the $NgoS_{DNA:ADP}$ complex (Table 2). Based on the $EcMutS_{DNA:ADP}$ structure, an energy-minimized homology model of the NgoS protein in complex with DNA and ADP was used for fitting the SAXS data corresponding to the $NgoS_{DNA:ADP}$ complex. The modelled $NgoS_{DNA:ADP}$ structure fitted the SAXS envelope derived from the average bead

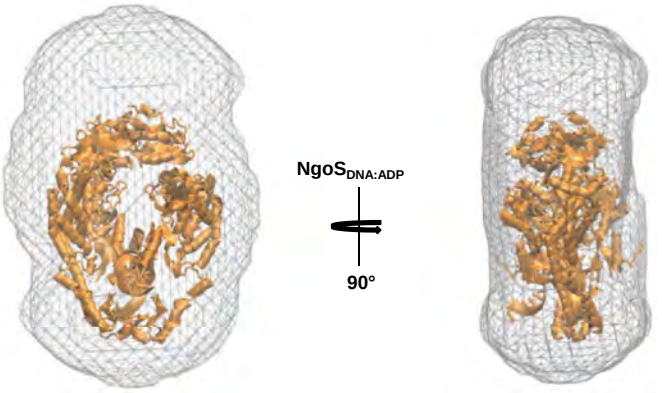

Figure 2. SAXS envelopes of the NgoS–ADP–DNA complex: Homology model of $NgoS_{DNA:ADP}$ (in orange ribbon) is superimposed with the corresponding SAXS derived *ab initio* model. The SAX envelop obtained for the complex shows that $NgoS_{DNA:ADP}$ complex adopts a shape similar to that seen for the $EcMutS_{DNA:ADP}$ and the $TqMutS_{DNA}$ complexes.

Table 2. SAX data collection and statistics

	$NgoS_{DNA:ADP}$
Data Collection parameters	
Beam line	BM29
Detector	Pilatus 1M
Wavelength (Å)	0.991
s range (nm^{-1})	0.035–4.94
Exposure time per frame (s)	0.5
No. of frames collected	10
Concentration (mg/ml)	2, 3.5, 5
Measurement Temperature (K)	283
Structural parameters	
I(0) [from P(r)]	113.30 ± 0.049
Rg (nm) [from P(r)]	6.08 ± 0.002
I(0) [from Guinier]	113.98 ± 0.96
Rg (nm) [from Guinier]	6.10 ± 0.05
Dmax (nm)	17.37

model with a correlation coefficient of ∼0.7 (Figure 2). The SAXS experiments show that the NgoS molecule adopts a conformation on binding DNA that is similar to that seen for the $EcMutS_{DNA:ADP}$ and the $TqMutS_{DNA}$ complexes.

Comparison of $NgoS_{ADP}$ structure with that of $EcMutS_{DNA:ADP}$ and $TqMutS_{DNA}$

The structure of the $NgoS_{ADP}$ complex shows drastic differences with that of MutS–DNA complex structures especially in the relative orientation of the two monomers (Figure 3 and Supplementary Figure S3). Monomer A (A_m) of $TqMutS_{DNA}$ and $EcMutS_{DNA:ADP}$ superimpose well on A_m of $NgoS_{ADP}$ complex with rmsd values of 2.3 Å (770 Cα pairs) and 2.0 Å (770 Cα pairs), respectively. This superimposition shows that the entire Monomer B (B_m) is pivoted in a different orientation in the $NgoS_{ADP}$ structure. In the DNA bound structures, the two monomers are in the same plane and form an oval disc with contacts between CTD and Clamp domains from each monomer. In the $NgoS_{ADP}$ structure, A_m and B_m are associated through CTD but are at an angle (∼45°) with respect to each other and there is no contact between residues of clamp region from each monomer (Figure 3). The comparison also shows that the

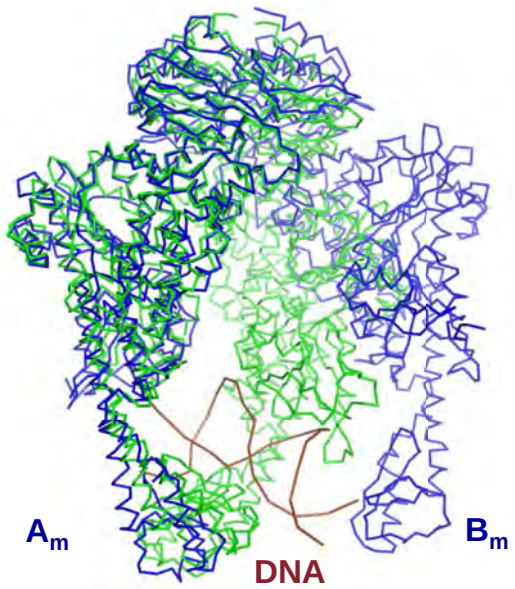

Figure 3. Comparison of $NgoS_{ADP}$ and $EcMutS_{DNA:ADP}$: a superimposition of the $NgoS_{ADP}$ and $EcMutS_{DNA:ADP}$ structures is displayed here. The $NgoS_{ADP}$ molecule is coloured blue and the $EcMutS_{DNA:ADP}$ structure is coloured green with DNA in brown. The comparison shows that there is a gap between the monomers through which DNA can enter and that monomer B_m so that the clamp regions of both monomers can interact and encircle DNA.

clamp region of A_m will have to move inwards slightly to encircle the DNA (Supplementary Figure S4).

In the $TqMutS_{DNA}$ and $EcMutS_{DNA:ADP}$ structures, it was seen that the A_m is involved in forming non-specific contacts with DNA and B_m is involved in detecting the mismatch. This partition of function leads to formation of an asymmetric dimer which is characteristic of the MutS functional complex. In line with this, A_m of $EcMutS_{DNA:ADP}$ superimposes on B_m with an rmsd of 1.4 Å. However in the case of the $NgoS_{ADP}$, A_m superimposes on B_m with an rmsd of 0.8 Å, with most of the structural differences localized to the mismatch recognizing NTD. This comparison also shows that MutS undergoes a transition from a symmetric to asymmetric dimer on DNA binding.

In $TqMutS_{DNA}$ complex, T468 of the clamp in A_m interacts with the same residue in B_m through the backbone atoms (3.1 Å). In the $NgoS_{ADP}$ structure, the corresponding residue is T488 and the distance between equivalent atoms is 48.5 Å (Figure 4). Thus, the comparison shows that in the absence of DNA, the two monomers are at angle to each other and a large gap is formed between the clamp regions from each monomer.

Structure of the $NgoS_{AMPPNP}$ complex

We were able to determine the structure of NgoS in complex with AMPPNP to a resolution of 3.3 Å (Supplementary Figure S5) (Table 3). The structure shows clear density for the AMPPNP molecule (Supplementary Figure S6) and both monomers showed the presence of one bound AMPPNP. The bound AMPPNP molecule is held in place due to stabilizing interactions formed with residues

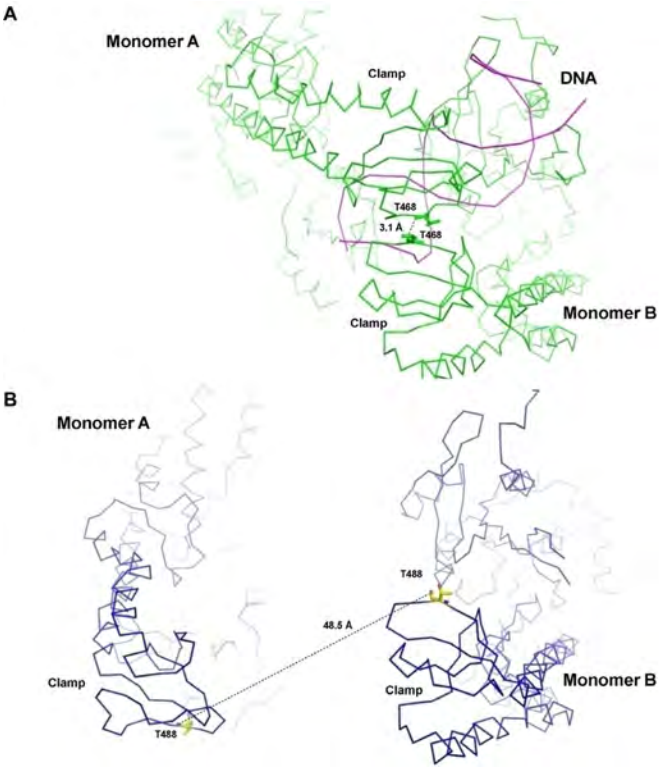

Figure 4. Comparison of the clamp regions in $NgoS_{ADP}$ and $TqMutS_{DNA}$ (**A**) The backbone atoms of the residue T468 in the clamp region of each monomer interact in $TqMutS_{DNA}$ (B) The equivalent residue T488 in monomer A of $NgoS_{ADP}$ is located 48.5 Å from the same residue in Monomer B.

from the CTD (Supplementary Figure S7). In addition to the residues that interact with ADP in the $NgoS_{ADP}$ complex, the residues D686 and E687 of the Walker B motif form interactions with the γ-phosphate of AMPPNP in the $NgoS_{AMPPNP}$ complex. The $NgoS_{AMPPNP}$ structure superimposed onto the $NgoS_{ADP}$ structure with an rmsd of 0.338 Å (1560 Cα pairs) and the structures of the two complexes are nearly identical (Supplementary Figure S8). The relative orientation of the monomers in $NgoS_{AMPPNP}$ is identical to that seen in the case of $NgoS_{ADP}$ structure and the $NgoS_{AMPPNP}$ structure also shows the presence of the gap between the clamp regions through which DNA can enter the central channel. The structure of TqMutS in its apo-state is also available and the clamp regions are disordered in this structure with no symmetry contacts between rows of molecules (Supplementary Figure S9). The comparison of TqMutS and $TqMutS_{DNA}$ structures suggests that there is only marginal difference in the orientation of the two monomers in the absence and presence of DNA. The superimposition of the two structures suggests that B_m exhibits a slight outward movement away from A_m on DNA binding and the difference in relative orientation of the two monomers that is observed in the case of $NgoS_{ADP}$ is not seen in the TqMutS structure (Supplementary Figure S10). The comparison of the TqMutS in the apo- state with the $NgoS_{ADP}$ structure suggests that on ATP binding, the two monomers attain the relative orientation seen in the case of $NgoS_{ADP}$ and the clamp regions undergo a disorder to or-

Table 3. Crystallographic data and refinement statistics for $NgoS_{AMPPNP}$ complex

Data collection statistics	
Wavelength (Å)	0.97864
Space Group	$P2_122_1$
Cell Constants (Å)	89.1 102.1 235.7
Resolution (Å)	3.3 Å (3.48–3.30)
R_{merge}	12.7 (77.5)
I/σI	8.5 (2.5)
Completeness (%)	98.1 (98.9)
Redundancy	5.3 (5.3)
Refinement	
Resolution (Å)	50–3.3 Å
No. of Reflections	32442
R_{work}/R_{free}	21.2/25.9
No. of atoms	
Protein	11820
AMPPNP	62
R.M.S. deviations	
Bond lengths (Å)	0.0011
Bond angles (°)	1.457
Average B-factors	
Protein	72.6 (Am); 108.2 (Bm)
AMPPNP	75.3 (Am); 99.5 (Bm)

[a]Values in parentheses are for the highest resolution shell.
[b]$R_{merge} = \Sigma|I-\langle I\rangle|/\Sigma I$, where I is the integrated intensity of a given reflection.
[c]$R_{work} = \Sigma||F_{obs}| - |F_{calc}||/\Sigma\,|F_{obs}|$. R_{free} was calculated using 5% of data excluded from refinement.

der transition. The overall effect of these transitions is that there is a gap formed between the two monomers and the MutS dimer now attains a state that is capable of threading DNA without the need for trans factors.

$NgoS_{ADP}$ exhibits compaction on DNA binding

Based on the crystal structures and the SAXS experiments, it appears that the NgoS transitions from an open to a closed complex on DNA entry. The presence of the gap allows DNA to enter the mismatch scanning channel. After DNA entry, the B_m will have to move by about 48 Å to be in the same flat plane as the A_m. Due to this movement the NTD from each monomer press onto the substrate DNA, and this will result in the bend observed in the DNA in MutS–DNA complexes. The bending of DNA destabilizes the helix and is vital for mismatch recognition. The proposed mechanism implies that DNA binding will involve a reduction in the hydrodynamic volume. In line with this, DLS studies showed that the addition of DNA leads to significant reduction of the hydrodynamic radius of NgoS (Figure 5). The compaction was observed in the presence of both heteroduplex and homoduplex DNA suggesting that the DNA loading process is same for mismatch scanning as well as detection.

The similarity in the structures of NgoS in complex with ADP and AMPPNP suggests that ATP hydrolysis may not be important for the DNA loading process. In agreement with this observation, the compaction was observed both in the presence of ADP and AMPPNP, implying that ATP binding, but not hydrolysis, is critical for assembly of the MutS–DNA complex (Figure 5). Mutation of the catalytic residue E687 in the Walker B motif to Ala was seen to have low ATPase activity (Supplementary Figure S1A). DLS

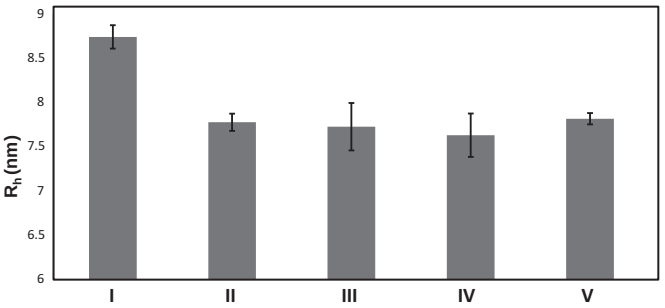

Figure 5. NgoS exhibits compaction on DNA binding. The graph displays the average R_h values obtained using DLS for (I) NgoS in the presence of ADP and absence of DNA, (II) NgoS in the presence of ADP and heteroduplex DNA bearing a G:T mismatch, (III) NgoS in the presence of AMPPNP and heteroduplex DNA bearing a G:T mismatch, (IV) NgoS in the presence of ADP and homoduplex DNA without any mismatch & (V) NgoS in the presence of AMPPNP and homoduplex DNA without any mismatch. The NgoS-ADP/AMPPNP complex undergoes significant compaction on binding DNA.

experiments showed that this mutant also exhibits compaction in the presence of DNA, further confirming that ATP hydrolysis is not critical for the DNA loading process (Supplementary Figure S11).

ATP is expelled from A_m on DNA binding

In the $NgoS_{ADP}$ structure, the ADP molecule is bound to both monomers as compared to the $EcMutS_{DNA:ADP}$ structure wherein it is only seen bound to the B_m. The ATP binding cavity in NgoS is formed by the loops 583–591, 607–615 (P-loop) and 740–755. Superimposition of the A_m of $NgoS_{ADP}$ on that of $EcMutS_{DNA:ADP}$ shows that the conformation of the loop 583–591 remains unchanged, and the loop 740–755 moves inwards towards the ATP binding cavity (Figure 6A). The P-loop exhibits maximal difference and in the conformation seen in $EcMutS_{DNA:ADP}$, the residues of this loop will clash with the phosphate moiety of the bound ADP. It is therefore possible, that the movement of the B_m towards A_m results in a conformational change in two of the loops in A_m that form the ADP binding cavity and expulsion of the bound nucleotide from the A_m.

The comparison of the $NgoS_{ADP}$ and the $EcMutS_{DNA:ADP}$ structures shows that for the mismatch binding monomer (B_m), the C-terminal helix-turn-helix (residues 760–785 in NgoS) is the only region that shows some degree of overlap in the two structures (Supplementary Figure S12). Conversely, while the rest of the monomer structure shows good overlap, the C-terminal region (CTR) made up of a hairpin and helix-turn-helix (738–779 in NgoS) of A_m does not overlap with corresponding region in $EcMutS_{DNA:ADP}$ This region of A_m appears to move downward on DNA binding and this movement forces the P-loop (residues 607–615) to enter the cavity in which the triphosphate moiety of ATP is present (Figure 6B). The residue H753 from the C-terminal hairpin and helix of NgoS forms an interaction with the backbone atom of M610 and this interaction holds the CTR and P-loop together and couples their downward movement. The residues H753 and M610 are conserved in EcMutS as

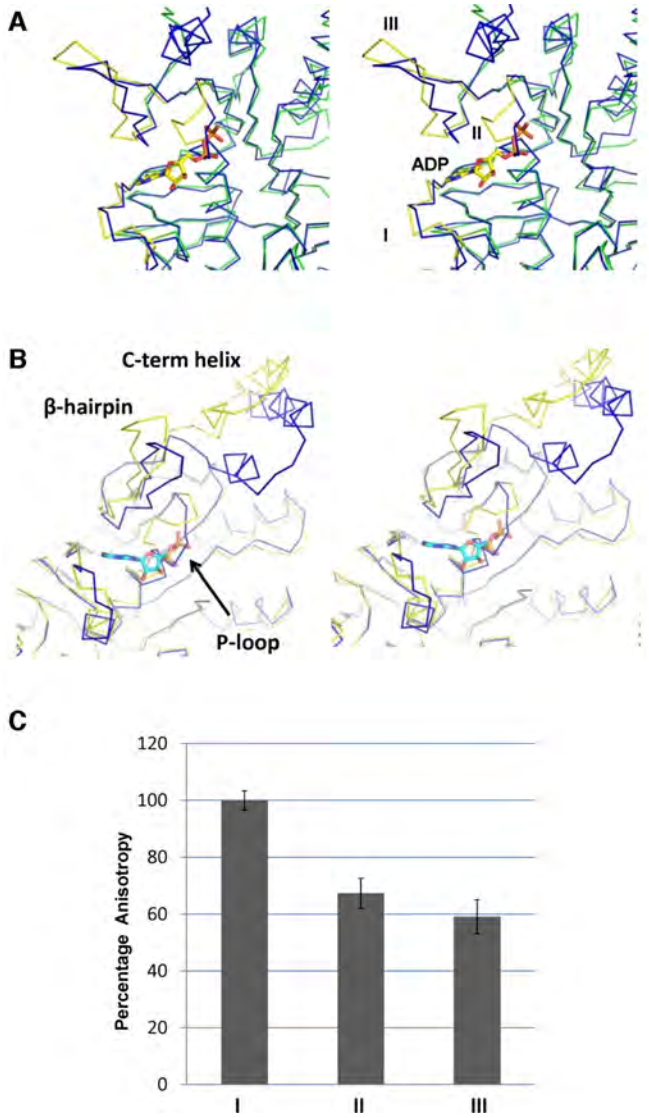

Figure 6. ATP expulsion on DNA binding: Stereo Figures showing comparison of the ADP binding site in Monomer A of $EcMutS_{DNA:ADP}$ (in blue) and $NgoS_{ADP}$ (in green). (**A**) The ATP binding loops in $NgoS_{ADP}$ are displayed in yellow with the residues 583–591, 607–615 (P-loop) and 740–755 labelled I, II and III, respectively. The P-loop appears to move inward on DNA binding and will clash with the phosphate moiety of the ADP molecule. (**B**) The repositioning of B_m (not shown here) is concomitant with a downward movement of the C-terminal helix and this ultimately results in the inward movement of the P-loop. Movie 1: the movie shows the changes in the C-terminal region (residues 740–755) and P-loop of A_m on movement of B_m. The CTR is coloured blue and the P-loop is coloured red. The final conformation attained by the P-loop in A_m will clash with the bound ATP resulting in its expulsion. (**C**) The graph displays the change in fluorescence anisotropy of 6-FAM labelled ATP (I) in the presence of NgoS and absence of DNA (II) in the presence of NgoS and presence of heteroduplex DNA bearing a G:T mismatch and (III) in the presence of NgoS and presence of homoduplex DNA without any mismatches.

H760 and M617, respectively. The comparison therefore provides insight into why the ADP is not bound to A_m in the $EcMutS_{DNA:ADP}$ structure. Overall, it appears that the CTR acts like a hinge around which B_m pivots so that the clamp regions of the two monomers can associate after DNA binding (Movie 1). Deletions in this region have been shown to adversely affect function of TqMutS (39). Regions of different lengths starting from the C-terminus were deleted and it was seen that if the TqMutS protein was truncated at or before residue 757 then the ability to dimerize was lost and consequently mismatch recognition was affected. However, TqMutS truncated till residue 760 could dimerize but transformation with this construct in a MutS deleted strain led to a mutation frequency that was higher than a MutS null strain. Our studies suggest that truncation till 760 may permit dimerization but perturb the ability of the C-terminal region to act as a hinge to enable movement of the B_m monomer and thus adversely impact mismatch recognition. Overall, the movement of B_m is coupled to the movement of C-terminal hairpin & helix in A_m which ultimately results in the expulsion of ATP from A_m.

The observed fluorescence anisotropy due to binding of 6FAM- labelled ATP to NgoS or E687A mutant dropped by about 40% on the addition of heteroduplex DNA bearing a G:T mismatch (Figure 6C and Supplementary Figure S13). This observation is in line with the prediction that binding of DNA will result in expulsion of ATP/ADP molecule bound to the A_m. A similar drop for WT and mutant protein was observed in the presence of homoduplex DNA also and this suggests that ATP expulsion is part of the DNA scanning process. The $TqMutS_{DNA}$ complex structure was initially determined in the absence of ATP/ADP and soaking these crystals with ADP showed that both monomers are bound by ADP, unlike $EcMutS_{DNA:ADP}$ (14,40). It is possible that the two monomers are occupied in $TqMutS_{DNA}$ due to high concentrations of ATP/ADP used or that there are significant differences regarding the dynamics of ATP binding during MutS–DNA assembly in the case of mesophilic and thermophilic bacteria.

DISCUSSION

Based on single molecule microscopy studies, it was suggested that the MutS molecules exhibits vibrational scissoring motion with high frequency association and dissociation of the clamp regions (15). Our studies suggest that binding of ATP may reduce the frequency of these motions considerably and give rise to a more rigid structure wherein the two monomers are held at an angle to each other and not present in the same flat plane. This configuration leads to the formation of a gap between the two monomers through which DNA can be loaded onto the central channel. The MutS molecule probably loads onto DNA such that one monomer forms non-specific contacts with DNA. The other monomer- which now represents the mismatch binding monomer then has to undergo a large movement in which the clamp region of this monomer moves by as much as 50 Å to come in contact with the clamp region of the other monomer and encircle DNA. The contacts formed between DNA and the clamp regions and the NTD will ensure that the mobility in the clamp regions is abolished to form a stable toroid around DNA. FRET-based experiments showed that the distance between NTD varies and this variation reduces on DNA binding (16). It was inferred from these studies that the NTD moves inwards and outwards in the absence of DNA and in the presence

of DNA, NTD from both monomers move towards each other. In contrast to this inference, our studies suggest that the change in distance between the NTDs is due to the movement of the entire monomer as a rigid unit towards the other monomer.

In addition to the conjecture that the NTDs flip between open and closed states, it was also believed that the clamp regions are mobile in the absence of DNA. Earlier models of DNA assembly suggested that DNA slips in through the gap created due to the movement of only these clamp regions. The clamp regions were believed to exhibit an inward movement in the presence of DNA in order to associate with each other and encircle DNA (13). However, it is difficult to envision that the movement of these mobile regions can generate the force required to bend DNA and destabilize the helix to scan for mismatches. The DNA double helix is known to be rigid even over short length scales and therefore a considerable amount of force would be required to bend DNA (41). Our studies show that the entire mismatch recognizing monomer moves towards the other monomer and this will result in the NTD pressing down on DNA to bend it. The proposed mechanism suggests that the two monomers may act like the jaws of a pair of wire bending pliers and the force generated due to movement of the entire monomer should be adequate to bend DNA and destabilize the helix to scan for mismatches.

The complexes presented here show the presence of ADP/AMPPNP in the binding site of both monomers. However, in the DNA bound complexes of EcMutS, the mismatch binding monomer shows the presence of ADP but the other monomer does not (11). The studies presented here also show that the MutS–ATP complex undergoes a transition from symmetric to asymmetric complex on binding DNA. This is in contrast to earlier beliefs that the asymmetry exists in the MutS even prior to DNA binding (42). These observations imply that choice of which monomer will recognize the mismatch and which monomer will bind DNA non-specifically is arbitrary and happens stochastically in the presence of DNA. Such a situation will ensure that initially there is no polarity in the MutS dimer vis-a-vis DNA binding. This will enable MutS to load onto DNA in either orientation of the dimer (related by 180 degree rotation around the long axis) and diffuse in either direction to scan for mismatches. These attributes will ensure that the probability of finding a mismatch is independent of the location of MutS loading and dimer orientation and therefore the chances of non-detection of a mismatch are minimal. The observed asymmetry in the MutS–DNA complex manifests itself only on binding DNA and not before to optimize the chances of finding mismatches during scanning.

Overall, the structural analysis and biophysical studies shown here along with previous observations suggest the following sequence of events (Figure 7 and Movie 2): (i) MutS binds nucleotide resulting in reduced vibrational motion of the two monomers. The nucleotide binding results in the formation of a stable structure as observed in the case of $NgoS_{ADP}$. (ii) DNA enters the scanning tunnel through the gap present between the clamp regions of the two monomers. (iii) B_m moves to be in the same plane as A_m and the clamp region associate to thread the substrate DNA through the scanning tunnel. This movement also results in the release of the nucleotide bound to the A_m. (iv) Due to the movement of B_m, the NTDs from each monomer press down on DNA at both faces of the oval disc and this result in bending and destabilization of the DNA double helix. Residues of the NTD from B_m can then form base-specific contacts to detect the mismatch. Overall, the movement of the B_m appears to be the primary event that ensures encirclement and bending of DNA and therefore the MutS acts like a pair of pliers to bend DNA.

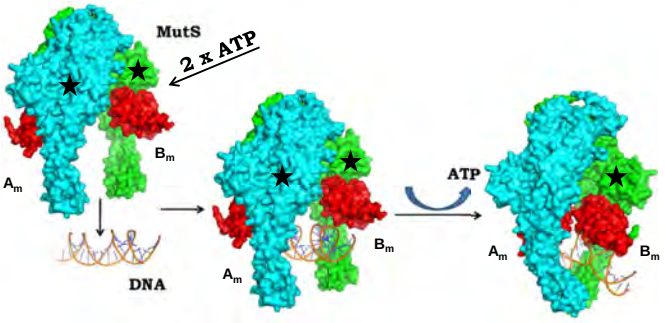

Figure 7. Mechanism of assembly of the MutS–DNA complex: The different stages in the assembly of MutS–DNA complex are displayed in. The two monomers are shown in surface representation, labelled A_m and B_m and are coloured cyan and green, respectively. The N-terminal domain of each monomer is coloured red. In addition, a movie that displays the mechanism of DNA loading onto MutS is also provided (Movie 2)**.** Movie 2: The sequence of events that occur during the assembly of the MutS–DNA complex are displayed here. The A_m and B_m monomers are shown in cyan and green colour, respectively. The P-loop of monomer Am is coloured red for the DNA is shown in ribbon representation and the bound ATP molecules are shown in stick representation.

The proposed mechanism suggests that the C-terminal region of the MutS molecule acts like a hinge to facilitate movement of the B_m towards A_m. The INSIGHT database lists all mutations in human MMR genes present and many of them co-segregate with diseases such as HNPCC. There are two MutS orthologues in eukaryotes and these molecules are heterodimers. MutSα is formed by the gene products of the *msh2* and *msh6* genes. MutSβ is formed by the gene products of msh2 and msh3 genes. The INSIGHT database lists at least 20 mutations in the msh6 gene and two mutations in the msh2 genes that will affect the region corresponding to the C-terminal helix-turn-helix of NgoS (Supplementary Table S1). This region maps to residues 1294–1335 and 798–855 in MSH6 and MSH2, respectively. It is possible that the model of MutS–DNA assembly proposed for mesophilic bacteria is conserved in eukaryotes. The mutations in the C-terminal region probably perturb the ability of MutSα to form a toroid around DNA and thus affect the efficiency of MMR in humans. In line with this hypothesis, the mutations in msh6 that lead to truncated protein that is missing the C-terminal hairpin and helix have been implicated in recurrent glioma, endometrial cancer and rectal tumour (Supplementary Table S1). Further studies on MutSα are required to validate this hypothesis and these efforts will provide deeper insight regarding mechanism of assembly of a functional MutSα–DNA complex and the effect of pathogenic mutations in the C-terminus on MMR.

DATA AVAILABILITY

The coordinates and structure factors for the $NgoS_{ADP}$ and $NgoS_{AMPPNP}$ have been deposited in the PDB with coordinates 5YK4 and 5X9W.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEGEMENTS

We thank the X-ray diffraction facilities located in the Molecular Biophysics Unit of the Indian Institute of Science and at the Regional Centre for Biotechnology. D.T.N. thanks Dr Hassan Belrhali and Dr Babu Manjashetty (BM14 beamline, ESRF) for help with X-ray diffraction data collection. D.T.N. also thanks Dr Bart Van Laer (BM29 beamline, ESRF) for help with the collection of Small angle X-ray scattering data and the ESRF Access Program of RCB (supported by the Grant No. BT/INF/22/SP22660/2017 of the Department of Biotechnology, Ministry of Science and Technology).
Author contributions: Conceived and designed the experiments: D.T.N., D.N.R. Performed the experiments: S.N., D.S.K. Analyzed the data: S.N., A.S., D.T.N., D.N.R. Wrote the manuscript: S.N., D.T.N., D.N.R.

FUNDING

CSIR (Govt. of India) Senior Research Fellowship (to S.N.); DBT Research Associateship (to D.S.K.); Science & Engineering Research Board of the Department of Science & Technology, JC Bose Fellowship (to D.N.R.); Regional Centre for Biotechnology (to D.T.N.); Department of Biotechnology, Ministry of Science and Technology ['BT/PR882/BRB/10/935/2011' to D.T.N., D.N.R.]. Funding for open access charge: Regional Centre for Biotechnology, Intra-mural Grant.
Conflict of interest statement. None declared.

REFERENCES

1. Joseph,N., Duppatla,V. and Rao,D.N. (2006) Prokaryotic DNA mismatch repair. *Prog. Nucleic Acid Res. Mol. Biol.*, **81**, 1–49.
2. Watson,J.D., Baker,T.A., Bell,S.P., Gann,A., Levine,M. and Losick,R. (2013) *Molecular Biology of the Gene*. 7th edn. CSHL Press, NY.
3. Iyer,R.R., Pluciennik,A., Burdett,V. and Modrich,P.L. (2006) DNA mismatch repair: functions and mechanisms. *Chem. Rev.*, **106**, 302–323.
4. Fishel,R. (2015) Mismatch repair. *J. Biol. Chem.*, **290**, 26395–26403.
5. Xie,C., Sheng,H., Zhang,N., Li,S., Wei,X. and Zheng,X. (2016) Association of MSH6 mutation with glioma susceptibility, drug resistance and progression. *Mol. Clin. Oncol.*, **5**, 236–240.
6. Peltomaki,P. (2001) DNA mismatch repair and cancer. *Mutat. Res.*, **488**, 77–85.
7. Leach,F.S., Nicolaides,N.C., Papadopoulos,N., Liu,B., Jen,J., Parsons,R., Peltomaki,P., Sistonen,P., Aaltonen,L.A., Nystrom-Lahti,M. *et al.* (1993) Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell*, **75**, 1215–1225.
8. Amaral-Silva,G.K., Martins,M.D., Pontes,H.A., Fregnani,E.R., Lopes,M.A., Fonseca,F.P. and Vargas,P.A. (2016) Mismatch repair system proteins in oral benign and malignant lesions. *J. Oral Pathol. Med.*, **46**, 241–245.
9. Su,S.S. and Modrich,P. (1986) Escherichia coli mutS-encoded protein binds to mismatched DNA base pairs. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 5057–5061.
10. Sixma,T.K. (2001) DNA mismatch repair: MutS structures bound to mismatches. *Curr. Opin. Struct. Biol.*, **11**, 47–52.
11. Lamers,M.H., Perrakis,A., Enzlin,J.H., Winterwerp,H.H., de Wind,N. and Sixma,T.K. (2000) The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch. *Nature*, **407**, 711–717.
12. Natrajan,G., Lamers,M.H., Enzlin,J.H., Winterwerp,H.H., Perrakis,A. and Sixma,T.K. (2003) Structures of Escherichia coli DNA mismatch repair enzyme MutS in complex with different mismatches: a common recognition mode for diverse substrates. *Nucleic Acids Res.*, **31**, 4814–4821.
13. Obmolova,G., Ban,C., Hsieh,P. and Yang,W. (2000) Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature*, **407**, 703–710.
14. Junop,M.S., Obmolova,G., Rausch,K., Hsieh,P. and Yang,W. (2001) Composite active site of an ABC ATPase: MutS uses ATP to verify mismatch recognition and authorize DNA repair. *Mol. Cell*, **7**, 1–12.
15. Qiu,R., DeRocco,V.C., Harris,C., Sharma,A., Hingorani,M.M., Erie,D.A. and Weninger,K.R. (2012) Large conformational changes in MutS during DNA scanning, mismatch recognition and repair signalling. *EMBO J.*, **31**, 2528–2540.
16. Hingorani,M.M. (2016) Mismatch binding, ADP-ATP exchange and intramolecular signaling during mismatch repair. *DNA Repair (Amst)*, **38**, 24–31.
17. Cho,W.K., Jeong,C., Kim,D., Chang,M., Song,K.M., Hanne,J., Ban,C., Fishel,R. and Lee,J.B. (2012) ATP alters the diffusion mechanics of MutS on mismatched DNA. *Structure*, **20**, 1264–1274.
18. Jeong,C., Cho,W.K., Song,K.M., Cook,C., Yoon,T.Y., Ban,C., Fishel,R. and Lee,J.B. (2011) MutS switches between two fundamentally distinct clamps during mismatch repair. *Nat. Struct. Mol. Biol.*, **18**, 379–385.
19. Kuriyan,J. and O'Donnell,M. (1993) Sliding clamps of DNA polymerases. *J. Mol. Biol.*, **234**, 915–925.
20. Fell,V.L. and Schild-Poulter,C. (2015) The Ku heterodimer: function in DNA repair and beyond. *Mutat. Res. Rev. Mutat. Res.*, **763**, 15–29.
21. Kelch,B.A., Makino,D.L., O'Donnell,M. and Kuriyan,J. (2011) How a DNA polymerase clamp loader opens a sliding clamp. *Science*, **334**, 1675–1680.
22. Roberts,S.A. and Ramsden,D.A. (2007) Loading of the nonhomologous end joining factor, Ku, on protein-occluded DNA ends. *J. Biol. Chem.*, **282**, 10605–10613.
23. Criss,A.K., Bonney,K.M., Chang,R.A., Duffin,P.M., LeCuyer,B.E. and Seifert,H.S. (2010) Mismatch correction modulates mutation frequency and pilus phase and antigenic variation in Neisseria gonorrhoeae. *J. Bacteriol.*, **192**, 316–325.
24. Minor,W., Cymborowski,M., Otwinowski,Z. and Chruszcz,M. (2006) HKL-3000: the integration of data reduction and structure solution–from diffraction images to an initial model in minutes. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 859–866.
25. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
26. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
27. Emsley,P.A.C.K. (2004) Coot: model building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
28. Winn,M.D., Ballard,C.C., Cowtan,K.D., Dodson,E.J., Emsley,P., Evans,P.R., Keegan,R.M., Krissinel,E.B., Leslie,A.G., McCoy,A. *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 235–242.
29. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
30. Konarev,P.V., Volkov,V.V., Sokolova,A.V., Koch,M.H.J. and Svergun,D.I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.*, **36**, 1277–1282.

31. Franke,D., Petoukhov,M.V., Konarev,P.V., Panjkovich,A., Tuukkanen,A., Mertens,H.D.T., Kikhney,A.G., Hajizadeh,N.R., Franklin,J.M., Jeffries,C.M. *et al.* (2017) ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.*, **50**, 1212–1225.
32. Svergun,D. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.*, **25**, 495–503.
33. Franke,D. and Svergun,D.I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.*, **42**, 342–346.
34. Volkov,V.V. and Svergun,D.I. (2003) Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.*, **36**, 860–864.
35. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
36. Wriggers,W. (2010) Using Situs for the integration of multi-resolution structures. *Biophys. Rev.*, **2**, 21–27.
37. Wriggers,W. (2012) Conventions and workflows for using Situs. *Acta Crystallogr. D Biol. Crystallogr.*, **68**, 344–351.
38. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph*, **14**, 33–38.
39. Biswas,I., Obmolova,G., Takahashi,M., Herr,A., Newman,M.A., Yang,W. and Hsieh,P. (2001) Disruption of the helix-u-turn-helix motif of MutS protein: loss of subunit dimerization, mismatch binding and ATP hydrolysis. *J. Mol. Biol.*, **305**, 805–816.
40. Alani,E., Lee,J.Y., Schofield,M.J., Kijas,A.W., Hsieh,P. and Yang,W. (2003) Crystal structure and biochemical analysis of the MutS.ADP.beryllium fluoride complex suggests a conserved mechanism for ATP interactions in mismatch repair. *J. Biol. Chem.*, **278**, 16088–16094.
41. Mazur,A.K. and Maaloum,M. (2014) Atomic force microscopy study of DNA flexibility on short length scales: smooth bending versus kinking. *Nucleic Acids Res.*, **42**, 14006–14012.
42. Lamers,M.H., Winterwerp,H.H. and Sixma,T.K. (2003) The alternating ATPase domains of MutS control DNA mismatch repair. *EMBO J.*, **22**, 746–756.