Published online 8 February 2020

Nucleic Acids Research, 2020, Vol. 48, No. 6 ***3165–3180***
doi: 10.1093/nar/gkaa075

Mycobacterial DNA polymerase I: activities and crystal structures of the POL domain as apoenzyme and in complex with a DNA primer-template and of the full-length FEN/EXO–POL enzyme

Shreya Ghosh[1], Yehuda Goldgur[2] and Stewart Shuman[1,*]

[1]Molecular Biology Program, Sloan-Kettering Institute, New York, NY 10065, USA and [2]Structural Biology Program, Sloan-Kettering Institute, New York, NY 10065, USA

Received January 07, 2020; Revised January 23, 2020; Editorial Decision January 25, 2020; Accepted February 06, 2020

ABSTRACT

Mycobacterial Pol1 is a bifunctional enzyme composed of an N-terminal DNA flap endonuclease/5′ exonuclease domain (FEN/EXO) and a C-terminal DNA polymerase domain (POL). Here we document additional functions of Pol1: FEN activity on the flap RNA strand of an RNA:DNA hybrid and reverse transcriptase activity on a DNA-primed RNA template. We report crystal structures of the POL domain, as apoenzyme and as ternary complex with 3′-dideoxy-terminated DNA primer-template and dNTP. The thumb, palm, and fingers subdomains of POL form an extensive interface with the primer-template and the triphosphate of the incoming dNTP. Progression from an open conformation of the apoenzyme to a nearly closed conformation of the ternary complex entails a disordered-to-ordered transition of several segments of the thumb and fingers modules and an inward motion of the fingers subdomain—especially the O helix—to engage the primer-template and dNTP triphosphate. Distinctive structural features of mycobacterial Pol1 POL include a manganese binding site in the vestigial 3′ exonuclease subdomain and a non-catalytic water-bridged magnesium complex at the protein-DNA interface. We report a crystal structure of the bifunctional FEN/EXO–POL apoenzyme that reveals the positions of two active site metals in the FEN/EXO domain.

INTRODUCTION

The roster of DNA polymerases in the human pathogen *Mycobacterium tuberculosis* and its avirulent relative *Mycobacterium smegmatis* comprises nine enzymes (1). DnaE1 is the essential replicative DNA polymerase with intrinsic proofreading 3′-5′ exonuclease activity (2,3). DnaE2, though inessential, is induced by DNA damage and is involved in adaptive mutagenesis (4,5). The polymerase component of DNA ligase D (LigD-POL) is dispensable for growth but necessary for mutagenic non-homologous end joining (NHEJ) (6,7). DNA polymerases PolD1 and PolD2 (also inessential for growth) are stand-alone paralogs of LigD-POL (8). PolD1 (also known as Prim-PolC) plays a gap-filling role in base excision repair (9). Among the three mycobacterial DinB polymerase paralogs (all inessential for growth), DinB2 is notable as being error-prone and for its ability to incorporate ribonucleotides owing to its lack of a steric gate in the polymerase active site (10–12).

DNA polymerase I (Pol1) was the first mycobacterial DNA polymerase for which a gene (*polA*) was assigned (13); it was also the first to be produced in recombinant form and characterized biochemically as a bifunctional enzyme with 5′-3′ exonuclease and primer-dependent templated DNA polymerase activities (14). The 5′-3′ exonuclease activity relies on an ensemble of putative metal-binding acidic amino acids located in the N-terminal portion of the protein (Figure 1A). The C-terminal segment of mycobacterial Pol1 is homologous to the Klenow polymerase domain of *Escherichia coli*, *Thermus aquaticus* and *Bacillus stearothermophilus* Pol1 (Figure 1A and Supplementary Figure S6). Whereas mycobacterial Pol1 has a central segment that is a putative homolog of the *E. coli* Klenow 3′-5′ exonuclease proofreading domain, it has no 3′-5′ exonuclease activity because it is missing several key metal-binding residues of the 3′-5′ exonuclease module (13,14).

Genetic analysis of *M. smegmatis* Pol1 by the Mizrahi lab had shown that a *polA* strain with targeted insertion of a kanamycin-resistance cassette at the position specifying amino acid 755 of the *polA* ORF was viable, albeit sensitized to UV irradiation and hydrogen peroxide (15). Subsequent genome-wide transposon mutagenesis studies indicated that viable *M. tuberculosis* strains could be recovered with transposon inserts in the 3′ segment of the *polA*

*To whom correspondence should be addressed. Tel: 212 639 7145; Email: s-shuman@ski.mskcc.org

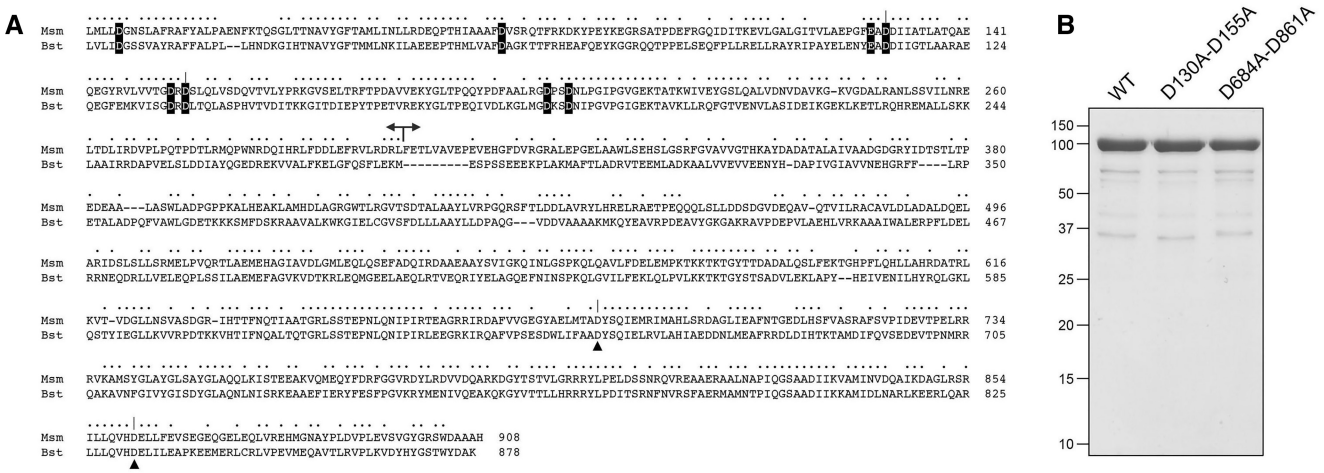

Figure 1. Primary structure and purification of *Mycobacterium smegmatis* Pol1. (**A**) The amino acid sequence of *M. smegmatis* Pol1 is aligned to that of *B. stearothermophilus* Pol1 (NCBI accession WP_053413883.1). Positions of side chain identity/similarity are denoted by dots above the alignment. Gaps in the alignment are denoted by dashes. The border between the catalytically active N-terminal FEN/EXO domain (aa 1–303) (20) and the C-terminal POL domain (aa 304–908) (this study) is indicated by the double-arrowhead line. Amino acids in the N-terminal domain of *M. smegmatis* Pol1 that were shown to be essential for Pol1 5′ exonuclease activity (14) are shown in white font on black background. Two of the metal-binding amino acids in the POL domain are denoted by black triangles below the alignment. The pairs of amino acids mutated in the present study (Asp130A and Asp155, in the FEN/EXO domain and Asp684 and Asp861 in the POL domain) are indicated by |. (**B**) Aliquots (5 μg) of purified recombinant wild-type Pol1 and the indicated double-alanine mutants were analyzed by SDS-PAGE. The Coomassie Blue-stained gel is shown. The positions and sizes (kDa) of marker polypeptides are indicated on the left.

gene (16). Such inserts truncate the C-terminus of the Pol1 protein and presumably ablate the Pol1 polymerase activity. In principle, such *polA* insertion mutants would still produce the N-terminal exonuclease domain of Pol1 fused to variable lengths of the polymerase domain. The observation that transposon inserts were excluded from the proximal portion of the *polA* gene (16) suggested that the 5′-3′ exonuclease activity of Pol1 might be essential for viability.

Though not addressed when mycobacterial Pol1 was initially characterized, contemporaneous studies showed that the N-terminal domains of *E. coli* and *Thermus aquaticus* (Taq) Pol1 are structure-specific endonucleases that incise 5′ flap junctions (17–19). It is thought that the flap endonuclease (FEN) activity of the Pol1 N domain plays a role in processing of Okazaki fragments during DNA replication. We reported recently that the isolated N-terminal 303-amino acid segment of *M. smegmatis* Pol1 (Pol1-N) is indeed an autonomous DNA 5′ flap endonuclease (20). Pol1-N catalyzes magnesium- or manganese-dependent incision of a nicked 5′ flap DNA between the first and second nucleotides of the duplex segment (20). The mycobacterial proteome includes a second stand-alone flap endonuclease, FenA, that has been characterized biochemically and structurally (20,21). FenA is inessential for mycobacterial growth, by the criterion of transposon insertion (16), raising the prospect that its function is redundant to that of the Pol1-N FEN activity.

In the present study, we address the activity and structure of mycobacterial Pol1 and its component catalytic domains. We report that *M. smegmatis* Pol1 has FEN activity on the flap RNA strand of an RNA:DNA hybrid duplex as well as reverse transcriptase activity on a DNA-primed RNA template. By targeting the predicted metal-binding sites of the FEN and polymerase domains, we obtained separation-of-function mutants. We report crystal structures of the polymerase (POL) domain, as the apoenzyme and in complex with a primer-template and dNTP. A structure of the full-length mycobacterial Pol1 affirms the high homology of the Pol1 FEN/EXO domain to mycobacterial FenA and reveals the position of two catalytic metal ions in the FEN/EXO domain that are coordinated by essential carboxylate side chains. Our results prompt a discussion of whether and how the polymerase and flap endonuclease activities of Pol1 might be coordinated during strand displacement synthesis.

MATERIALS AND METHODS

Recombinant *M. smegmatis* proteins

The open reading frames of the full-length 908-amino acid *M. smegmatis* Pol1 (MSMEG_3839) and the C-terminal POL domain (aa 304–908) were PCR amplified from genomic DNA with primers that introduced a BglII site immediately flanking the first codon of the Pol1 or POL ORF and a XhoI site downstream of the stop codon. The PCR products were digested with BglII and XhoI and ligated into pET28B-His$_{10}$Smt3 vector that had been digested with BamHI and XhoI. The resulting pET28B-His$_{10}$Smt3-Pol1 and pET28B-His$_{10}$Smt3-POL expression plasmids encode Pol1 and POL, respectively, fused to an N-terminal His$_{10}$Smt3 tag under the transcriptional control of a T7 RNA polymerase promoter. Plasmids encoding double-alanine mutants of Pol1 (D130A–D155A and D684A–D861A) were obtained by site-directed mutagenesis using the quick-change method with Agilent Pfu Turbo DNA polymerase. The plasmid inserts were sequenced to verify that no unintended coding changes were acquired during amplification and cloning.

The Pol1 and POL expression plasmids were introduced into *E. coli* BL21 (DE3) cells. Cultures (1 l) amplified from single kanamycin-resistant transformants were

grown at 37°C in Luria-Bertani broth containing 60 μg/ml kanamycin until the A_{600} reached 0.6. The cultures were then adjusted to 2% (vol/vol) ethanol and 0.4 mM isopropyl-β-D-thiogalactopyranoside and incubated for 3 h at 30°C with constant shaking at 225 rpm. Cells were harvested by centrifugation, and the pellets were stored at −80°C. All subsequent steps were performed at 4°C. Cells were harvested by centrifugation and resuspended in 25 ml of buffer A (50 mM Tris–HCl, pH 8.0, 500 mM NaCl, 10% sucrose, 20 mM imidazole) containing 1 EDTA-free protease inhibitor cocktail tablet (Roche). Lysozyme was added to a concentration of 1 mg/ml. After incubation for 1 h, the lysate was sonicated to reduce viscosity and the insoluble material was removed by centrifugation at 38 000 × g for 45 min. The supernatant was mixed for 1 h with 5 ml of Ni-nitriloacetic acid (NTA)-agarose resin (Qiagen) that had been equilibrated with buffer A. The resin was recovered by centrifugation and was washed twice with 50 ml of buffer A, then with 20 ml of 50 mM Tris–HCl, pH 8.0, 3 M KCl, followed by 100 ml of buffer B (50 mM Tris–HCl, pH 8.0, 500 mM NaCl, 10% glycerol) containing 20 mM imidazole. The resin was then poured into a column, and the bound material was serially step-eluted with 50, 100 and 200 mM imidazole in buffer B. The polypeptide compositions of the fractions were monitored by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE). The 100 and 200 mM imidazole eluates containing the His_{10}Smt3-tagged Pol1 or POL were pooled, supplemented with Smt3-specific protease Ulp1 (60 μg) and then dialyzed overnight against 3 liters of buffer C (20 mM Tris–HCl, pH 8.0, 250 mM NaCl, 1 mM DTT, 20 mM imidazole, 1 mM EDTA, 10% glycerol), during which time the His_{10}Smt3 tag was cleaved. The dialysates were mixed with 5 ml of Ni-NTA agarose equilibrated with buffer C. The mixtures were nutated for 30 min and then poured into a column. The tag free proteins were recovered in the flow-through fractions. The cleaved His_{10}-Smt3 tag was bound to the resin and recovered in the 500 mM imidazole eluate. The respective protein preparations were purified further by gel filtration through a Superdex-200 column equilibrated in buffer D (20 mM Tris–HCl, pH 8.0, 150 mM NaCl, 1 mM DTT, 10% glycerol). All the proteins eluted as a single discrete peak at an elution volume consistent with them being monomers in solution. The ratio of A_{260} to A_{280} of the resulting peak for each of the purified proteins was 0.54–0.58. Peak fractions for each protein were pooled and stored at −80°C. Protein concentrations were determined by using the Bio-Rad dye reagent with bovine serum albumin as the standard. The yields per liter of bacterial culture were as follows: wild-type Pol1 (11 mg); Pol1-(D130A-D155A) (12 mg); Pol1-(D684A-D861A) (1.9 mg); POL domain (7.5 mg).

Nucleic acid substrates

5′ ^{32}P-labeled DNA or RNA strands were prepared by reaction of synthetic oligonucleotides with T4 polynucleotide kinase and [γ^{32}P]ATP. The labeled DNA or RNA strands were separated from free ATP by electrophoresis through a nondenaturing 18% polyacrylamide gel and then eluted from an excised gel slice. The flap nick and nicked duplex substrates for assays of FEN and EXO activity were formed by annealing a 5′ ^{32}P-labeled 28-mer or 18-mer strand and an unlabeled 18-mer DNA_{OH} strand to a 36-mer template strand at 1:5:2 molar ratio in buffer containing 10 mM Tris–HCl, pH 7.5, 200 mM NaCl, 1 mM EDTA. The primer-template for assay of DNA polymerase was formed by annealing a 5′ ^{32}P-labeled 18-mer pDNA strand to a 36-mer template strand at 1:3 molar ratio. The DNA mixtures were incubated serially at 65, 37 and 22°C to promote strand annealing.

Nuclease assay

Reaction mixtures (10 μl) containing 20 mM Tris–HCl, pH 8.0, 5 mM $MgCl_2$ or 1 mM $MnCl_2$, 1 mM DTT, ^{32}P-labeled FEN (flap nick DNA) or EXO (nicked DNA) substrate, and enzyme as specified in the figure legends were incubated at 37°C for 30 min. The reactions were quenched by adding 10 μl of 90% formamide, 50 mM EDTA, 0.01% bromophenol blue-xylene cyanol. The samples were analyzed by electrophoresis through a 40-cm 18% polyacrylamide gel containing 7.5 M urea in 44.5 mM Tris–borate, pH 8.3, 1 mM EDTA. The products were visualized by autoradiography.

Polymerase assay

Reaction mixtures (10 μl) containing 50 mM Tris–HCl, pH 8.0, 8 mM $MgCl_2$, 5 mM DTT, 0.125 mM each of dATP, dGTP, dCTP and dTTP, ^{32}P-labeled primer-template DNA and Pol1 as specified were incubated at 37°C for 20 min. The reactions were quenched by adding 10 μl of 90% formamide, 50 mM EDTA, 0.01% bromophenol blue-xylene cyanol. The samples were heated at 95°C for 5 min and then analyzed by electrophoresis through a 40-cm 18% polyacrylamide gel containing 7.5 M urea in 44.5 mM Tris–borate, pH 8.3, 1 mM EDTA. The products were visualized by autoradiography.

Crystallization

Crystals of POL and POL in complex with DNA were grown by sitting drop vapor diffusion in 96-well plates at room temperature. For POL apoenzyme, a 1 μl aliquot of a solution of 15 mg/ml (225 μM) POL and 10 mM $MnCl_2$ in buffer D was mixed with 1 μl of precipitant/reservoir solution containing 0.2 M $MgCl_2$, 20% (w/v) PEG3350 and then equilibrated against 150 μl of the same reservoir solution. POL crystals were cryoprotected by transfer to a solution of 0.2 M $MgCl_2$, 20% PEG3350, 20 mM $MnCl_2$, 25% (v/v) glycerol before being flash-frozen in liquid nitrogen. For POL•DNA complex, a solution of 15 mg/ml (225 μM) POL, 225 μM each of 10-mer primer (5′-GCGATCACGT) and 16-mer template (5′-GACGTACGTGATCGCA-3′) strands, 10 mM $MgCl_2$, 1 mM ddATP and 1 mM dCTP was incubated for 20 min at 4°C. A 1 μl aliquot of the POL•DNA solution was mixed with 1 μl of precipitant/reservoir solution containing 0.2 M $MgCl_2$, 0.1 M MES, pH 6.5, 25% (w/v) PEG4000 and then equilibrated against 150 μl of the same reservoir solution. POL•DNA crystals were cryoprotected by transfer to a solution of 0.2 M $MgCl_2$, 0.1 M MES, pH 6.5, 25% PEG4000, 20% glycerol before being flash-frozen in liquid

nitrogen. Crystals of full-length Pol1 were grown by mixing a 1 μl aliquot of 30 mg/ml Pol1 and 10 mM $MnCl_2$ in buffer D with 1 μl of precipitant/reservoir solution containing 1 M LiCl, 0.1 M MES, pH 5.5, 20% (w/v) PEG6000, 3% (w/v) dextran sulfate sodium salt (Mr 5000). Pol1 crystals were cryoprotected by transfer to a solution of 1 M LiCl, 0.1 M MES, pH 5.5, 20% PEG6000, 3% dextran sulfate, 20 mM $MnCl_2$, 25% glycerol before being flash-frozen in liquid nitrogen.

Diffraction data collection and structure determination

X-ray diffraction data were collected at the Advanced Photon Source beamline 24ID-C. Data for the POL apoenzyme and POL•DNA complex were collected from single crystals. Data from two full-length Pol1 crystals were merged to improve completeness. The structures were solved by molecular replacement implemented in Phenix (22). A model of *M. smegmatis* POL that was generated in Phyre 2 (23) based on the structure of BstPOL (pdb 4UQG) was used as a search model to solve the *M. smegmatis* POL apoenzyme structure. The refined *M. smegmatis* POL structure was then employed as a search model for the POL-DNA complex and full-length Pol1. Iterative model building into electron density was performed in O (24). Refinements were executed in Phenix. Data collection and refinement statistics are compiled in Supplementary Table S1.

RESULTS

Active site mutations separate FEN/EXO and POL activities of *M. smegmatis* Pol1

We produced *M. smegmatis* Pol1 in *E. coli* as a His_{10}Smt3 fusion and isolated the recombinant protein from a soluble extract by nickel-agarose chromatography. After removing the His_{10}Smt3 tag with the Smt3-specific protease Ulp1, the native Pol1 was separated from the tag by Ni-agarose chromatography, and then purified further by gel filtration. In parallel, we produced and purified two double-alanine mutants of *M. smegmatis* Pol1: D130A-D155A, in which two of the putative metal ligands in the N-terminal flap endonuclease/5′-exonuclease (FEN/EXO) domain were changed to alanine; and D684A-D861A, in which two of the predicted metal ligands in the C-terminal polymerase (POL) domain were replaced by alanine (the mutated positions are indicated by | in Figure 1A). SDS-PAGE revealed prominent ∼100 kDa polypeptides corresponding to the wild-type or mutant Pol1 polypeptide in each case (Figure 1B). The recombinant FEN/EXO domain of *M. smegmatis* Pol1 (aa 1–303, or Pol1-N) was produced in *E. coli* and purified as described previously (20).

Full-length wild-type Pol1 and Pol1-N were assayed for flap endonuclease activity by reaction with a 'flap nick' substrate formed by annealing a 5′ ^{32}P-labeled 28-mer strand and an unlabeled 18-mer DNA_{OH} strand to a 36-mer template DNA strand to form a singly nicked duplex with a 10-nucleotide 5′ flap (Figure 2B). In the presence of 5 mM magnesium as the metal cofactor (the optimum concentration determined previously; 20), Pol1-N incised the flap strand between the first and second nucleotides of the duplex segment to liberate a 5′ ^{32}P-labeled 11-mer product (Figure 2A). With 1 mM manganese as the metal cofactor (the optimal concentration observed previously; 20), Pol1-N again cleaved predominantly between the first and second nucleotides of the duplex segment, but also generated a minor 5′ ^{32}P-labeled 12-mer product by incising between the second and third nucleotides of the duplex segment (Figure 2A). The full-length Pol1 enzyme displayed a different spectrum of cleavage products, as follows. With magnesium as the metal cofactor, Pol1 cleaved between the first and second nucleotides of the duplex segment, but also generated a minor 5′ ^{32}P-labeled 10-mer product by hydrolyzing the last phosphodiester of the single stranded segment of the flap strand (Figure 2A). The noteworthy finding was that manganese prompted full-length Pol1 to cleave the flap strand at the third, fourth, and fifth phosphodiesters of the duplex portion of the substrate (Figure 2A); these cleavage products were not formed by the isolated FEN domain Pol1-N. We speculate that the interaction of the POL domain of full-length Pol1 with the primer terminus at the 3′-OH strand of the nick in the presence of manganese elicited local melting of the flanking duplex, thereby creating additional flap junctions at which the FEN domain could cleave.

Having shown that full-length Pol1 has flap endonuclease activity, we proceeded to assay the two Pol1 mutants and found that the D130A-D155A enzyme was inert as a flap endonuclease, whereas the D684A-D861A enzyme retained FEN activity and incised at the same sites as wild-type Pol1 (Figure 2C). We also assayed the 5′ exonuclease activity of Pol1 by reaction with a nicked duplex substrate formed by annealing a 5′ ^{32}P-labeled 18-mer strand and an unlabeled 18-mer DNA_{OH} strand to a 36-mer template DNA strand (Figure 2B). Wild-type Pol1 resected the nick 5′ terminus, predominantly by cleavage between the first and second nucleotides to liberate a ^{32}P-labeled dNMP product (Figure 2C). The D130A-D155A mutation eliminated the EXO activity. By contrast, the D684A-D861A mutant retained EXO activity and generated the same cleavage products as wild-type Pol1 (Figure 2C).

In order to assay polymerase activity, we employed a 5′ ^{32}P-labeled substrate consisting of an 18 bp duplex with an 18-nucleotide 5′-tail (Supplementary Figure S1). Reaction of wild-type Pol1 with the primer-template in the presence of 500 μM deoxyribonucleoside triphosphates (dNTPs) resulted in extension of the labeled primer strand to the end of the template strand to generate a ^{32}P-labeled 36-mer product (Supplementary Figure S1). The efficiency of primer utilization was high, i.e. the majority of the primer strand was extended at a Pol1:primer ratio of 1:100 to form a ladder of elongation products (Supplementary Figure S1). The extent of complete fill-in synthesis increased at Pol1:primer ratios of 1:40 and 1:20. The D684A-D861A mutation in the POL domain drastically reduced polymerase activity, to <1% of the wild-type level (as gauged by the fact that 1 pmol of D684A-D861A enzyme catalyzed less primer extension than 0.01 pmol of wild-type Pol1; Supplementary Figure S1). The FEN/EXO-dead D130A-D155A enzyme retained polymerase activity similar to wild-type Pol1 (not shown).

These experiments show or affirm that: (i) native Pol1 has both flap endonuclease and DNA polymerase activities; (ii)

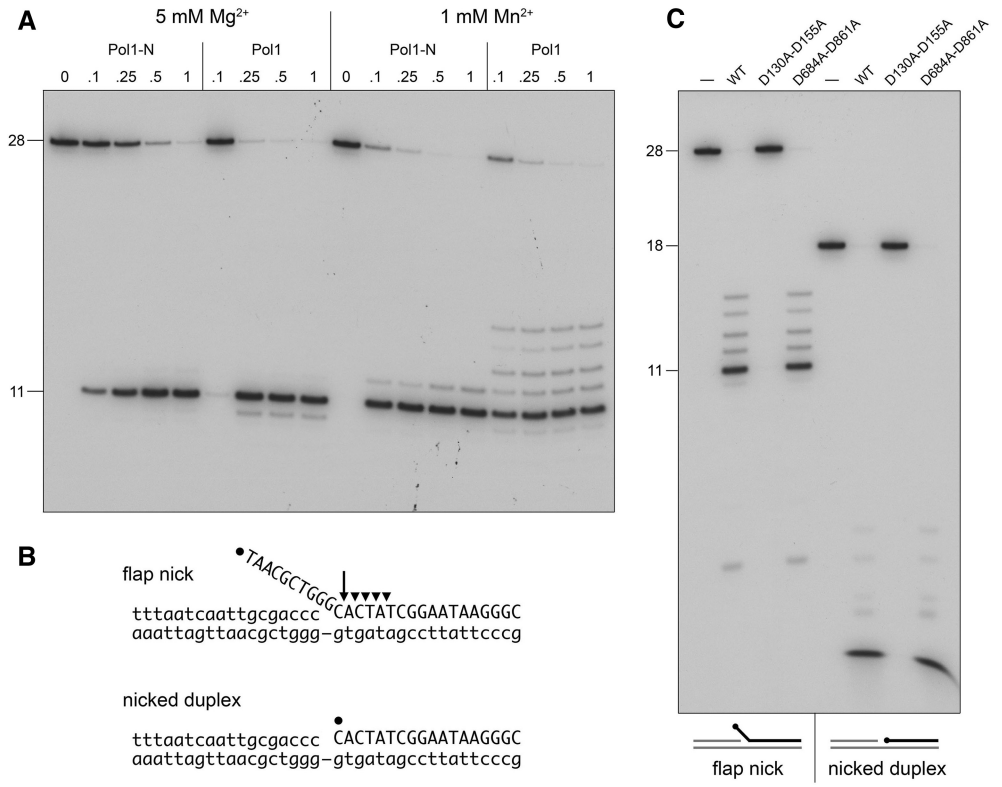

Figure 2. FEN and EXO activities of Pol1 and Pol1-N. (**A**) Flap endonuclease. Reaction mixtures (10 μl) containing 20 mM Tris–HCl, pH 8.0, 1 mM DTT, 1 pmol (100 nM) of ^{32}P-labeled flap nick DNA (depicted in panel B), 5 mM $MgCl_2$ or 1 mM $MnCl_2$ as specified, and 0. 0.1, 0.25, 0.5 or 1 pmol Pol1 or Pol1-N as specified were incubated at 37°C for 30 min. The products were analyzed by urea-PAGE and visualized by autoradiography. (**B**) Flap nick and nicked duplex substrates for assay of FEN and EXO activity. The 5′ ^{32}P label on the scissile DNA strand (nucleobases in upper case letters) is denoted by •. The unlabeled DNA strands are shown with nucleobases in lower case letters. The principal site of scission by the Pol1-N and Pol1 FEN activities in the presence of magnesium is indicated by the arrow. Additional sites of cleavage within the duplex segment by Pol1 in the presence of manganese are denoted by arrowheads. (**C**) Reaction mixtures (10 μl) containing 20 mM Tris–HCl, pH 8.0, 1 mM DTT, 1 mM $MnCl_2$, 1 pmol (100 nM) of ^{32}P-labeled flap nick DNA or nicked duplex DNA as indicated, and either no enzyme (lanes –) or 1 pmol of wild-type Pol1 or the indicated mutants were incubated at 37°C for 30 min. The products were analyzed by urea-PAGE and visualized by autoradiography.

FEN/EXO and POL functions are performed by distinct and mutually independent active sites.

Pol1 FEN activity on an RNA flap-nick substrate

Bacterial Pol1 is imputed to play a role in the removal of RNA primers laid down during lagging strand DNA replication and in the subsequent filling-in of the gaps between tandem Okazaki fragments. In some species, the primer removal function can be executed either by Pol1 or RNase H (25). Because *M. smegmatis* viability requires the activity of either one of the two type I RNase H enzymes in the mycobacterial proteome (26), it was of interest to see if *M. smegmatis* Pol1 has the capacity to cleave a flap structure in which the flap strand is RNA. To answer this question, we prepared an RNA flap-nick substrate (Figure 3) that was otherwise identical to the flap-nick DNA substrate (Figure 2B), except that the 5′ ^{32}P-labeled 28-mer flap strand was all RNA. Pol1 incised the RNA flap strand in the presence of 5 mM magnesium to generate a predominant 5′ ^{32}P-labeled 11-mer product (Figure 3) that comigrated with a 5′ ^{32}P-labeled synthetic RNA strand, pUAACGCUGGGC, corresponding to the terminal 11-nucleotide segment of the RNA flap substrate strand (not shown). Thus, the preferred site of RNA flap nick cleavage was at the first duplex phosphodiester bond (akin to what was seen with a DNA flap nick substrate). The extent of RNA flap cleavage increased with input Pol1 and was nearly quantitative at 1 pmol of Pol1. Whereas there was no evidence of cleavage at other sites within the RNA:DNA hybrid duplex in the presence of magnesium, we detected five sites of cleavage within the single stranded segment of the RNA flap strand (Figure 3). Control assays showed there was no RNA flap cleavage by 1 pmol of Pol1 in the absence of added divalent cation and that reaction of the RNA flap substrate with 1 pmol of the Pol1-D130A-D155A mutant in the presence of 5 mM magnesium elicited no flap cleavage (not shown). We conclude that the observed RNA flap incision activity was inherent to Pol1. Replacing magnesium with 1 mM manganese did not significantly affect the enzyme-dependence of substrate consumption (Figure 3), but did change the distribution of RNA cleavage products, such that: (i) cleavages within the single-strand segment of the RNA flap strand now predominated (generating a mixture of 5′-labeled mononucleotide, dinucleotide, trinucleotide, pentanucleotide and hexanucleotide products); (ii) the 11-mer product of incision at the first duplex phosphodiester was still evident, albeit not the major product and (iii) there

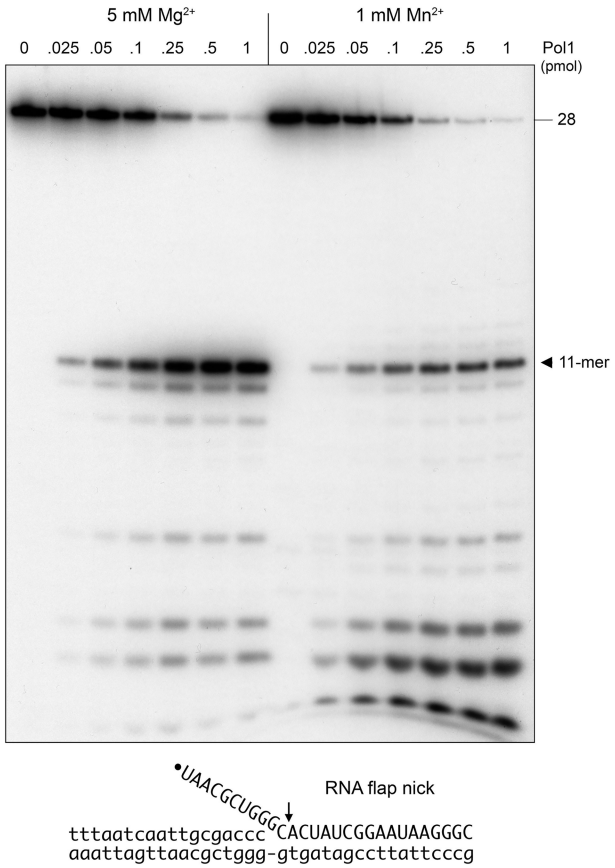

Figure 3. Pol1 FEN activity on an RNA flap-nick substrate. Reaction mixtures (10 μl) containing 20 mM Tris–HCl, pH 8.0, 1 mM DTT, 1 pmol of ^{32}P-labeled RNA flap nick substrate (depicted at bottom), either 5 mM $MgCl_2$ or 1 mM $MnCl_2$ as indicated, and 0, 0.025, 0.05, 0.1, 0.25, 0.5 or 1 pmol Pol1 were incubated at 37°C for 30 min. The products were analyzed by urea–PAGE and visualized by autoradiography. The '11-mer' label indicates the position of a 5′ ^{32}P-labeled RNA oligonucleotide, pUAACGCUGGGC, corresponding to the terminal 11-nucleotide segment of the RNA flap substrate strand.

was a low level of cleavage at the second and third phosphodiesters of the duplex segment of the RNA flap nick (reminiscent of the manganese effect on the DNA flap cleavage activity of Pol1 seen in Figure 2).

Reverse transcriptase activity of *M. smegmatis* Pol1

Recent studies have highlighted the capacity of RNA to serve as a template for homology-directed DNA break repair in vivo (27,28). Such a mechanism requires that a cellular polymerase act as a reverse transcriptase. *Escherichia coli* DNA Pol1 has an RNA-templated DNA polymerase activity *in vitro*, albeit one that is significantly less vigorous than its DNA-dependent DNA polymerase activity (29). The weak reverse transcriptase activity of *E. coli* Pol1 was attributed, in part, to a much higher K_m for dNTP substrates with an RNA template versus a DNA template (29). Whereas several mycobacterial DNA polymerases have been characterized as being adept at incorporating ribonucleotides during DNA-templated synthesis (8,11,12), relatively little is known about their ability to use

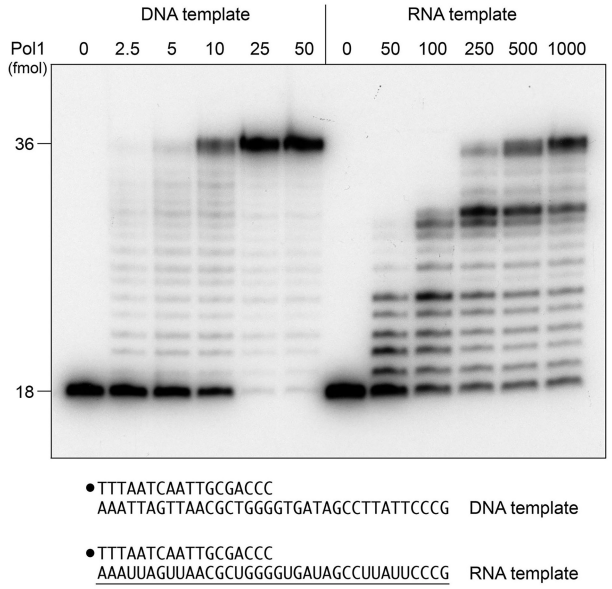

Figure 4. Reverse transcriptase activity of *M. smegmatis* Pol1. Reaction mixtures (10 μl) containing 50 mM Tris-HCl, pH 8.0, 8 mM $MgCl_2$, 5 mM DTT, 0.125 mM each of dATP, dGTP, dCTP and dTTP, 1 pmol ^{32}P-labeled 18mer-DNA/36mer-DNA or 18mer-DNA/36mer-RNA primer-template (shown at bottom with the 5′ ^{32}P label denoted by a filled circle and the RNA strand underlined underlined), and increasing amounts of Pol1 as specified were incubated at 37°C for 20 min. The products were analyzed by urea–PAGE and visualized by autoradiography.

RNA as a template, with the exception that mycobacterial DinB2 was reported to have reverse transcriptase activity with manganese as the metal cofactor (11). Here we tested *M. smegmatis* Pol1 in parallel for RNA-templated DNA synthesis and DNA-templated DNA synthesis using substrates composed of a 5′ ^{32}P-labeled 18-mer DNA primer annealed to a 36-mer template that was either all-RNA or all-DNA (Figure 4). Reaction mixtures contained magnesium as the metal cofactor and 125 μM of each of the four dNTPs, this concentration being 6-fold higher than the observed K_m for the *E. coli* Pol1 reverse transcriptase activity (29). We found that *M. smegmatis* Pol1 has reverse transcriptase activity, but one that is significantly weaker, on a per enzyme basis, than its DNA polymerase activity (Figure 4). To wit: (i) complete fill-in of the 18-nucleotides of the RNA template was first evident at 250 fmol of input Pol1, versus at 10 fmol of Pol1 for the DNA template; and (ii) whereas nearly all primers were fully extended on the DNA template by 25–50 fmol of Pol1, most extensions on the RNA template were incomplete even at 1000 fmol of Pol1 (Figure 4). These results suggest that mycobacterial Pol1, though possessed of reverse transcriptase activity, is not well suited *per se* to synthesis of long stretches of DNA on an RNA template.

Crystal structure of the POL domain as apoenzyme

We produced recombinant POL domain (aa 304–908), verified that it has DNA polymerase activity (not shown), and grew crystals of the POL domain that diffracted X-rays to 2.4 Å resolution. The crystals were in space group $C222_1$ and contained one POL protomer in the asymmetric unit.

The structure was solved by molecular replacement using a search model of *M. smegmatis* POL that was generated in Phyre 2 based on the structure of BstPOL. The refined *M. smegmatis* POL structure ($R_{work}/R_{free} = 21.9/26.5$; see Supplementary Table S1) is shown in Figure 5A colored according to the secondary structure elements, which include 14 β strands, 17 α helices, and one 3_{10} helix. One of the long α helices (aa 812–848) is punctuated by a central π-helix segment (aa 826–829; colored green in Figure 5A). A DALI (30) search of the PDB recovered a structure of *B. stearothermophilus* (Bst) POL (pdb 4DQQ, protomer A) as the top hit (*Z* score of 48.3, rmsd of 2.0 Å at 550 Cα positions, and 41% amino acid identity). The organization of the POL domain of mycobacterial Pol1 echoes that of BstPOL and other family A polymerases that consist of a central palm subdomain at the base of a cleft formed by fingers and thumb subdomains (Figure 5B) that are so named because they collectively resemble a right hand. The palm subdomain rests atop an N-terminal vestigial 3′ exonuclease subdomain (Figure 5B). It is well established that mycobacterial Pol1 lacks a proofreading 3′ exonuclease activity (14). As with BstPOL (31), the structure of *M. smegmatis* POL rationalizes this deficiency, insofar as the vestigial exonuclease module lacks several of the carboxylate-containing amino acids that coordinate the essential metal ions in the 3′ exonuclease active site of *E. coli* Pol1 Klenow fragment (32).

A notable finding was the presence of a single manganese ion engaged at the base of the vestigial exo subdomain (Figure 5), where it is coordinated octahedrally to His337, Asp367, Asp369 and three waters (Figure 6). Two of the metal-bound waters are coordinated in turn to Glu336. Assignment of the metal ion as manganese is supported by the anomalous difference density around the metal atom (Figure 6). The apoenzyme crystals were grown from a solution of POL containing 10 mM manganese, using a precipitant solution containing 200 mM magnesium. Retention of the POL-bound manganese in the face of excess magnesium suggests that this novel metal site is specific for manganese engagement. The functional significance of the manganese site is unclear. The ensemble of manganese-binding residues in the vestigial exo module of MsmPOL is not conserved in the Bst, Eco or Taq Pol1 proteins (Supplementary Figure S6).

The POL apoenzyme structure is punctuated by several gaps in the polypeptide chain (indicated in Figure 5) where electron density was insufficient or absent. The disordered segments included aa 453–468 in the vestigial exonuclease subdomain, aa 546–601 in the thumb subdomain, and aa 725–726 and 745–760 in the fingers subdomain. As shown below, some of these segments became ordered when POL is engaged with its DNA substrate.

Structure of POL in complex with DNA primer-template and dNTP

Purified POL was pre-incubated with a DNA primer-template (consisting of a 10-bp duplex with a 1-nucleotide 3′ overhang and a 5-nucleotide 5′ single-strand tail on the template strand; 33) in the presence of 10 mM magnesium chloride, 1 mM ddATP (dideoxy ATP) and 1 mM dCTP. The rationale was to allow the DNA polymerase to engage the primer terminus and catalyze the incorporation of ddAMP from ddATP opposite thymidine in the template, thereby rendering the elongation complex unable to undergo a second cycle of dNMP addition, in which case we hoped to capture the next incoming nucleotide substrate dCTP in the active site opposite the template deoxyguanosine (see Figure 7C). Crystals grown from the pre-incubated POL•DNA solution were in space group $C222_1$ and diffracted X-rays to 1.9 Å resolution. The asymmetric unit included two POL•DNA complexes. The structure was solved by molecular replacement using the *M. smegmatis* POL apoenzyme as the search model ($R_{work}/R_{free} = 19.1/23.7$; Supplementary Table S1). Discussion of the structure will focus on the A protomer•DNA complex. The DNA model in the structure (shown in stereo in Supplementary Figure S2, with its overlying electron density map) includes all 10 of the nucleotides of the primer strand and 14/16 nucleotides of the template strand; two nucleotides at the 5′ single-strand terminus of the template strand were disordered (these are colored gray in the cartoon in Figure 7C). The electron density showed that ddAMP was incorporated at the 3′ end of the input primer strand across from the first available thymidine deoxynucleotide in the template strand (Supplementary Figure S2). The triphosphate moiety of the next incoming dNTP could be modeled into the electron density; we could not confidently model the sugar or the nucleobase (we suspect the nucleoside adopts more than one conformation in the crystal).

A stereo view of the POL•DNA structure is shown in Figure 7A. The POL interface with the DNA primer-template is predominantly via the fingers and thumb subdomains, both of which become fully ordered in the POL•DNA complex. Figure 7B shows a stereo view of the superimposed structure of POL apoenzyme (colored gold) and DNA-bound POL (colored according to subdomain). The DNA-bound POL, which was crystallized in the presence of 200 mM magnesium, had no metal ion visible in the manganese site observed in the apoenzyme structure. Whereas there was little or no change in the palm and vestigial exo subdomains upon DNA binding, the thumb segment from aa 546–601 that was disordered in the apoenzyme became structured, forming a longer α helix (aa 528–553) than that seen in the apoenzyme plus two new α helices (aa 562–569 and 589–599) and a hairpin loop (aa 579–585) (Figure 7B). The newly ordered thumb region engages both strands of the primer-template duplex across the minor groove (Figure 7A). In particular, the thumb subdomain makes multiple contacts to three successive backbone phosphates of the primer strand—via the Thr581, Thr583, Thr586, Thr587 and Arg611 side chains and the Lys582, Thr583, Thr587 and Asp589 main chain amides (Figures 7C and 8A)—then contacts the last two phosphates at the primer terminus, via the Arg660 side chain and the Ile659 main chain amide. The thumb domain also contacts two successive phosphates of the template strand via the Gln564, Ser561 and Asn558 side chains (Figures 7C and 8A).

The fingers subdomain segment that was disordered in the apoenzyme (aa 745–760) became ordered as an α helix (aa 749–756) when POL was bound to DNA (Figure 7B). This newly ordered helix includes the Tyr750 side chain

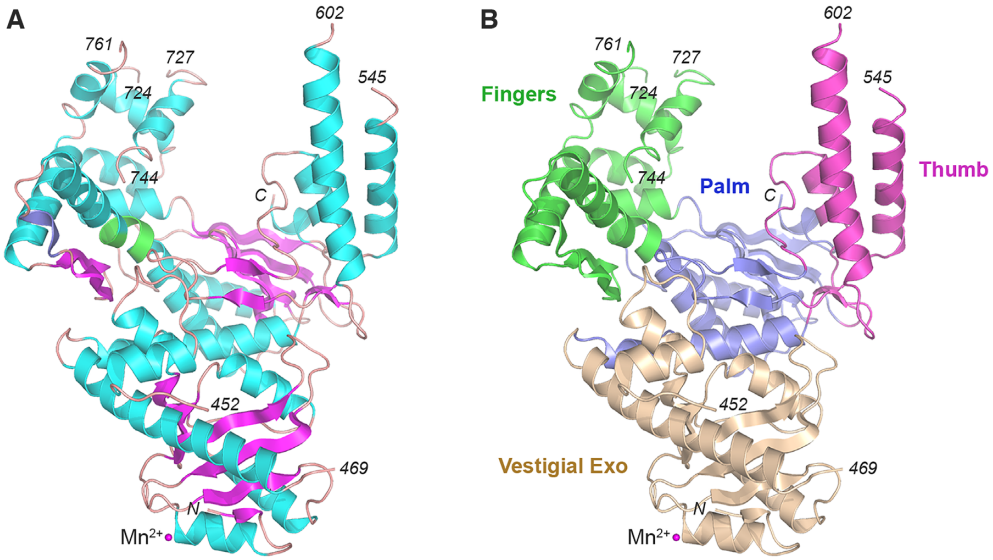

Figure 5. Structure of the POL domain of mycobacterial Pol1 as apoenzyme. (**A**) The tertiary structure of the POL domain is depicted as a cartoon model colored according to the secondary structure elements, with β strands in magenta, α helices in cyan, a π helix in green, and a 3_{10} helix in blue. (**B**) The POL structure is colored with the fingers subdomain in green, the thumb subdomain in magenta, the palm subdomain in blue, and the vestigial exonuclease domain in beige. A manganese ion engaged by the vestigial exo domain is shown as a magenta sphere.

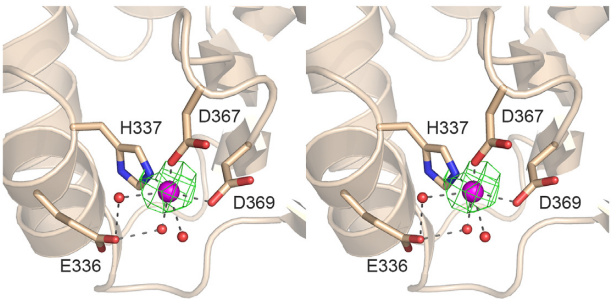

Figure 6. A manganese binding site in the vestigial 3′ exonuclease subdomain. Close up stereo view of the manganese ion (magenta sphere) bound to the vestigial 3′ exonuclease subdomain (beige cartoon trace). Amino acid side chains are rendered as stick models and waters as red spheres. Atomic contacts are indicated by dashed lines. Anomalous difference density for the manganese atom, contoured at 4σ, is shown in green mesh.

that makes a π stack on the first unpaired cytosine base of the template DNA strand (denoted by a green bar in Figure 7C). The fingers module engages three sequential backbone phosphates at the 5′ end of the template strand, via the Arg820, Arg802, and Ser748 side chains (Figures 7C and 8B) and it makes extensive contacts to the triphosphate of the incoming dNTP (discussed below). DNA binding elicits an inward movement of the fingers subdomain (Figure 7B), by 5.6 Å at the Ser810 Cα atom. The palm subdomain contacts three successive phosphates on the template strand (immediately preceding those engaged by the fingers subdomain), via the Glu651 main chain amide, the Ser648 side chain, and the Ala643 main chain amide (Figures 7C and 8B).

The electron density map revealed a non-catalytic magnesium ion as a constituent of the POL•DNA interface (Figure 8B). The octahedral magnesium coordination complex is filled by five waters and the Ile641 main chain carbonyl. The metal-bound waters make bridging contacts to the carbonyls of Pro432 and Gln424 (in the vestigial exo subdomain) and to a phosphate of the template DNA strand (Figure 8B).

In addition to the extensive phosphate backbone interface, POL interacts in the minor groove with several of the primer strand nucleobases near the primer terminus CGTA-3′. To wit: Arg615 donates a hydrogen bond to the cytosine-O2; Asn656 makes a hydrogen bond to guanine-N2; and Arg646 makes hydrogen bonds to adenine-N3 (Figure 7C). These minor groove contacts near the primer terminus, which are considered sequence-independent, are analogous to those made by BstPOL and TaqPOL (34,35).

The POL active site

A stereo view of the active site is shown in Figure 9, highlighting the added ddAMP at the 3′ terminus of the primer strand, the triphosphate moiety of the incoming dNTP, and two magnesium ions. Although the nucleoside moiety was not modeled, we see that the acceptor template dG nucleotide is positioned in the 'insertion site' (33). The dNTP triphosphate was modeled into the electron density as a complex with a magnesium ion that is coordinated octahedrally to non-bridging oxygen atoms of the α, β, and γ phosphates and to three waters (Figure 9). This magnesium is engaged to the enzyme via water-mediated contacts to Asp684 of the palm subdomain. The triphosphate is engaged in an extensive network of electrostatic and hydrogen bonding interactions with amino acids of the fingers subdomain, three of which (Arg733, Lys737 and Tyr741) emanate from the 'O helix' (Figure 9). Arg733 makes a bidentate contact from its terminal guanidium nitrogens to the γ phosphate. Lys737 makes a bifurcated contact to the α and γ phosphates. Tyr741 donates a hydrogen bond to the

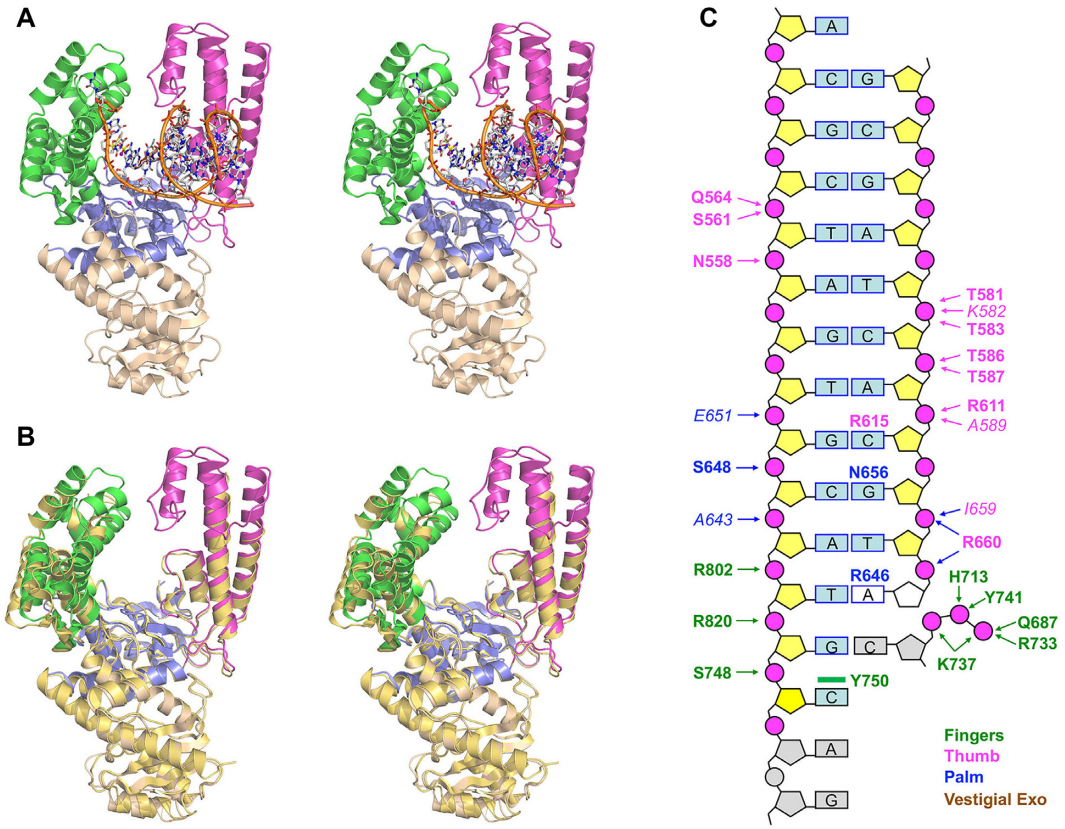

Figure 7. Structure of the POL domain in complex with primer-template DNA. (**A**) Stereo view of the POL•DNA complex with the protein depicted as a cartoon model colored by subdomains as in Figure 5 and as indicated in panel C. The primer-template DNA is depicted as a stick model with a gold cartoon trace through the phosphodiester backbone. The triphosphate moiety of the incoming dNTP is shown with phosphorus atoms colored yellow. Magnesium ions are rendered as magenta spheres. (**B**) Stereo view of the superimposed structures of the POL domain of the POL•DNA complex (colored as in panel A) and the POL apoenzyme (as a cartoon model colored gold). (**C**) Cartoon diagram summarizing POL–DNA contacts. The primer-template DNA is depicted as a 2D base pair ladder with the primer strand on the right and the template strand on the left. The input primer-template visualized in the structural model is colored with deoxyribose sugars in yellow, backbone phosphates in magenta and nucleobases in cyan. The ddAMP incorporated at the primer terminus during pre-incubation is colored white with respect to the sugar and nucleobase. The triphosphate moiety of the incoming dNTP was visualized and is depicted as three magenta spheres. The sugar and nucleobase of the incoming dNTP and two deoxynucleotides at the template strand 5′ terminus that were disordered are colored in gray. Amino acids that contact particular backbone phosphates and the triphosphate of the incoming dNTP are labeled to the left and right of the phosphates in bold font for side chain phosphate contacts and italic font for main chain phosphate contacts, as indicated by the arrows pointing at the phosphates. Amino acids R615, N656 and R646 that contact the nucleobases of the primer strand in the minor groove are shown above the relevant bases. Tyr750 makes a π stack on the first unpaired cytosine template base, depicted as a green bar. Amino acid labels are colored according to their subdomain location, as indicated.

β phosphate. His713 makes a bifurcated hydrogen bond to the β phosphate and the β–γ bridging oxygen. Gln687 coordinates the γ phosphate. A second magnesium ion is coordinated to the terminal ddAMP phosphate of the primer strand and is engaged to the enzyme via water-bridged contacts to Asp861, Glu862, and Asp684 of the palm subdomain and to the α phosphate of the incoming dNTP (Figure 9). This magnesium corresponds to the 'catalytic' metal that, in a canonical DNA polymerase Michaelis complex, coordinates the primer 3′-OH and the dNTP α phosphate (36). A primer 3′-OH is not present in the ddAMP-terminated primer strand in our mycobacterial POL•DNA structure, and this likely explains the displacement of the catalytic magnesium. Note that the 5.4 Å distance between the two magnesium ions is greater that what is expected for a polymerase active site poised for catalysis. This inter-magnesium distance is 3.4 Å in the DNA polymerase β Michaelis complex (36). We envision that were the primer 3′-OH present, the catalytic magnesium would occupy the position of the water that presently bridges to the α phosphate (placing it 3.7 Å from the other magnesium), thereby establishing proximity of the catalytic magnesium to the 3′-OH and its direct coordination to Asp861.

Comparison of the POL•DNA complex to a closed BstPOL•DNA complex with incoming dNTP

Figure 10 shows the structure of the mycobacterial POL•DNA complex superimposed on the 1.7 Å structure of BstPOL in complex with a ddCMP-terminated primer-template and an incoming dCTP nucleotide (37). The BstPOL structure exemplifies the fully closed conformation of a Pol1 active site, in which the dNTP is paired to the template nucleotide and a magnesium is coordinated octahedrally to the dNTP α, β and γ phosphates, the two metal-binding aspartates, and a main chain carbonyl oxygen. Al-

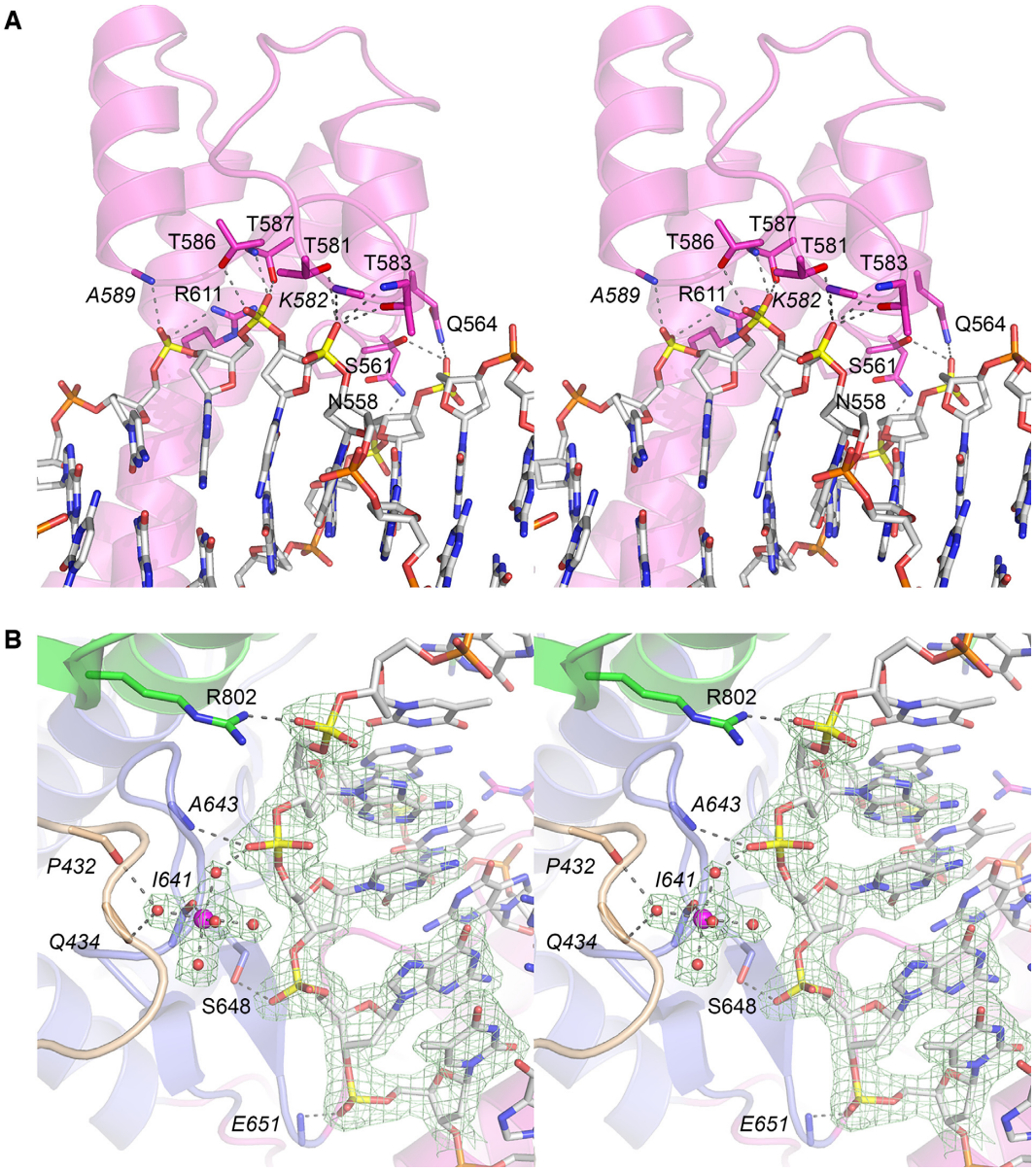

Figure 8. POL•DNA interface. (**A**) Stereo view of the interface of the thumb subdomain (magenta cartoon) with the DNA minor groove. DNA is depicted as a stick model with gray carbons. Thumb amino acids that contact specific DNA phosphates (shown with phosphorus atoms highlighted in yellow) are depicted as stick models with magenta carbons. Atomic contacts are denoted by dashed lines. (**B**) Stereo view of the POL interface with template strand phosphates, with phosphorus atoms of the contacted phosphates colored yellow. An octahedrally coordinated magnesium ion at the POL•DNA interface is engaged by water-bridged contacts to main chain atoms of the vestigial exo subdomain (Pro432 and Gln434) and the fingers subdomain (Ile641) and to a DNA phosphate. Other POL contacts to DNA are denoted by dashed lines. The $2F_o - F_c$ electron density map of the magnesium complex and the contacted template DNA nucleotides is shown as a green mesh contoured at 1.5σ.

though the catalytic magnesium is missing from the BstPOL structure (there being no primer 3′-OH), the dNTP α phosphorus atom is 3.8 Å from the C3′ atom of the dideoxy primer terminus and the PP_i leaving group is oriented for in-line displacement, were a 3′-OH and catalytic metal present. The amino acid contacts to the dNTP triphosphate in the BstPOL structure are similar to those seen in the mycobacterial POL•DNA complex. However, the triphosphate moiety and the associated magnesium are shifted away from the dideoxy C3′ atom in the MsmPOL structure (distance of 5.2 Å from the α phosphorus atom to the dideoxy C3′ atom) compared to the closed BstPOL complex. This is attributable to a lateral displacement of the O helix in the MsmPOL structure versus the closed BstPOL complex, by 4.3 Å at the Arg773 Cα position (Figure 10). The many available crystal structures of DNA-bound BstPOL indicate that the enzyme can sample a continuum of conformational states during the transition from an open pre-insertion complex to the fully closed Michaelis complex owing to the dynamics of the fingers subdomain, particularly the O helix and the dNTP•Mg complex that it engages (38,39). Thus, we envision that the mycobacterial POL•DNA complex

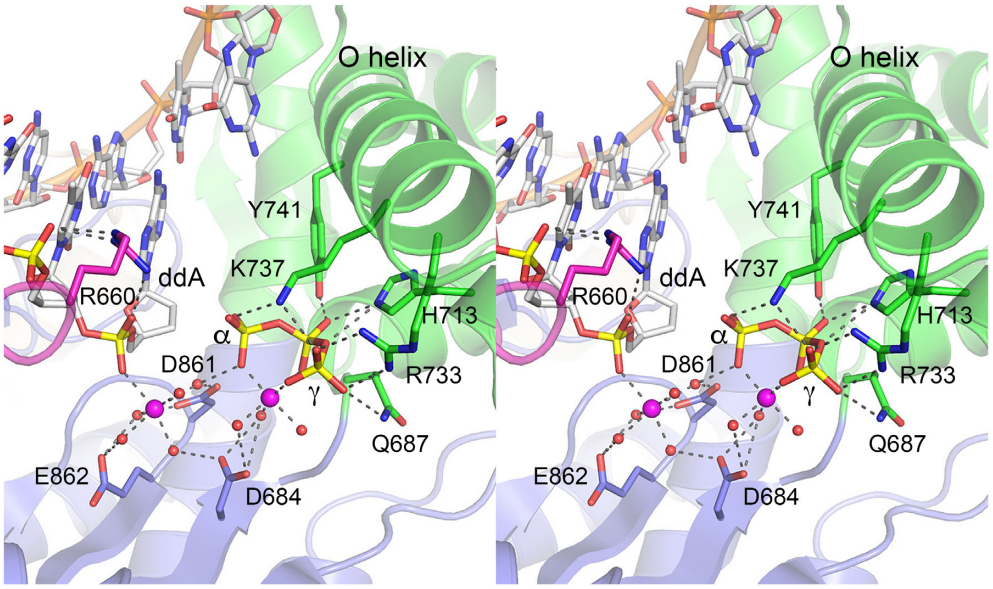

Figure 9. Polymerase active site in the POL•DNA complex. Stereo view of the primer-template terminus (stick model) engaged in the POL active site. The two terminal phosphates of the primer strand and the incoming dNTP triphosphate are colored yellow with respect to their phosphorus atoms. The POL cartoon model and amino acid side chains are colored by subdomain as in Figure 5. Magnesium ions and waters are rendered as magenta and red spheres, respectively. Atomic contacts are indicated by dashed lines.

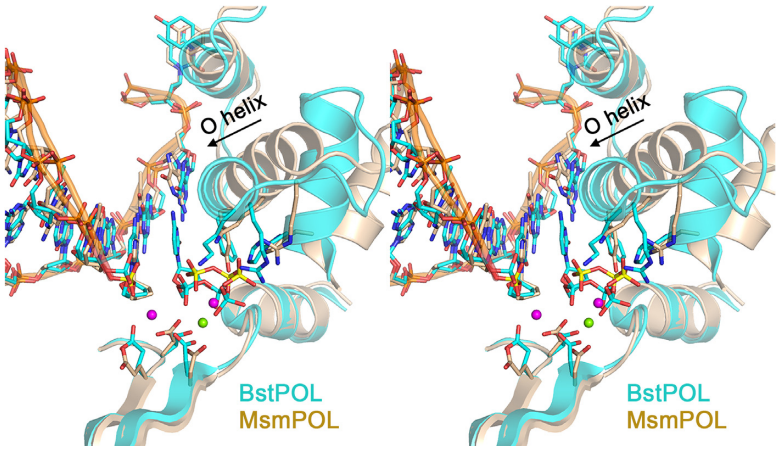

Figure 10. Comparison to the closed structure of BstPOL at a dideoxy primer-template with an incoming dNTP. The mycobacterial POL•DNA complex (MsmPOL) was aligned to a 1.7 Å structure of BstPOL in complex with a ddCMP-terminated primer-template and an incoming dCTP nucleotide (stick model with gray carbons and cyan phosphorus atoms) opposite a dG nucleotide in the template strand (pdb 4DQI). A stereo view is shown focusing on the elements around the primer-terminus, the magnesium ions (magenta spheres for MsmPOL and a green sphere for BstPOL), and the triphosphate moiety of the dNTP (colored with yellow phosphorus atoms for MsmPOL). Secondary structure elements of the palm (metal-binding) and fingers (dNTP binding) modules are shown as cartoon traces in cyan for BstPOL and beige for MsmPOL. The primer-template termini are shown as stick models (with cyan carbons for BstPOL and beige carbons for MsmPOL) with a cartoon trace through the phosphate backbone. Equivalent amino acids in the structures are shown as stick models with cyan carbons for BstPOL and beige carbons for MsmPOL. The lateral movement of the O helix in transition from the partly closed MsmPOL conformation to the fully closed BstPOL conformation is indicated by the arrow.

seen here captures an intermediate conformation *en route* to the closed state.

Structure of full-length *M. smegmatis* Pol1

The two-domain architecture of bacterial Pol1 enzymes raises interesting questions of whether and how the FEN and POL activities might be coordinated temporally and spatially when Pol1 is engaged in Okazaki fragment processing and/or gap repair. Whereas there are numerous insightful structures of the isolated POL domains of *Bacillus*, *Thermus* and *E. coli* Pol1 at various steps along the polymerase reaction pathway, only in the case of TaqPol1 is there a structure of the full-length Pol1 protein (40, 41; pdb 1TAU and 1TAQ). The 293-aa FEN/EXO domain of TaqPol1, which is structurally homologous to the stand-alone mycobacterial flap endonuclease FenA (rmsd of 3.1 Å at 254 Cα positions; ref 21), makes scant contact in *cis* with the POL domain; indeed, the EXO and POL active sites are >70 Å apart (40) (Supplementary Figure S3A). Thus, it is not

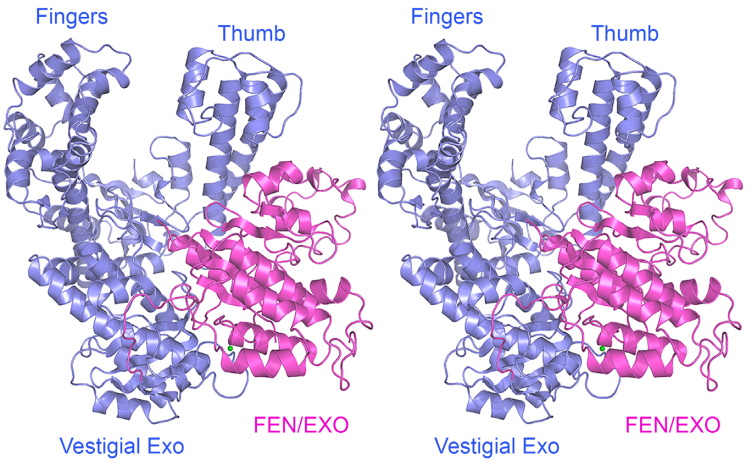

Figure 11. Structure of full-length mycobacterial Pol1. Stereo view of the Pol1 structure, depicted as a cartoon model with the POL domain colored blue and the FEN/EXO domain colored magenta. A manganese ion at the interface of the FEN/EXO domain and vestigial 3′ exo subdomain is depicted as a green sphere.

clear from the TaqPol1 crystal structure how the FEN/EXO domain could access and incise a flap concomitant with its formation.

In an attempt to clarify this issue for *M. smegmatis* Pol1, we crystallized the full-length recombinant protein after preincubation with 10 mM manganese. The crystals diffracted to 2.7 Å, were in space group P1, and had two MsmPol1 protomers in the asymmetric unit. The MsmPol1 model was refined to R_{work}/R_{free} of 23.4/30.4 (Supplementary Table S1). The Pol1 model comprises three continuous polypeptide segments: aa 18–79, aa 98–203, and aa 218–908. The FEN/EXO fold is interrupted by three disordered segments—from amino acids 1 to 17, 80 to 97, and 204 to 217—where electron density was absent or insufficient for modeling. Discussion of the Pol1 structure will focus on the B protomer, which is shown in Figure 11 with the POL domain colored blue and the FEN/EXO domain colored magenta. The FEN/EXO domain packs against the vestigial exo and thumb modules of the POL domain. Interdomain atomic contacts between FEN/EXO and the POL vestigial exo module include: Arg278–Asp440 and Arg278–Asp441 salt bridges; main chain–side chain hydrogen bond pairs Val121(O)–Gln473, Pro271(O)–Arg481, Glu313(N)–Glu450 and Gly317(O)–Arg449; and main chain–main chain hydrogen bond pairs Val121(N)–Leu463(O) and His316(O)–Glu450(N) (Supplementary Figure S4). A manganese ion is bound at the interdomain interface (green sphere in Figure 11 and Supplementary Figure S4) where it is surrounded by the Asp465, Asp471 and Gln473 side chains in the vestigial exo domain and by the Ala123 main chain carbonyl and Glu124 side chain of the FEN/EXO domain (Supplementary Figure S4). The amino acid side chains at the interface of the FEN/EXO and POL domains in the MsmPol1 structure are not conserved in the Bst, Eco or Taq Pol1 proteins (Supplementary Figure S6).

The compact arrangement of the FEN/EXO and POL domains in MsmPol1 contrasts with the splayed apart domain configuration in TaqPol1, as shown in Figures S3A and B in which the two full-length Pol1 structures are superimposed with respect to their POL domains and then offset laterally. There is a very large rigid-body movement of the respective FEN/EXO domains entailing rotation about the interdomain linker peptide that results in an 88 Å displacement in the position of a conserved aspartate: Asp222 in TaqPol1, Asp238 in MsmPol1 (Supplementary Figure S3). We surmise that: (i) there is potential for significant movement of the POL and FEN/EXO domains relative to one another in solution; and (ii) the differences in the positions of the FEN/EXO domains in MsmPol1 versus TaqPol1 likely reflect crystal packing effects.

Superposition of the MsmPOL•DNA structure on that of full-length MsmPol1shows that there is a steric clash between the duplex segment of the DNA primer-template and the FEN/EXO domain (Supplementary Figure S5), signifying that the position of the FEN/EXO domain seen in the crystal structure cannot pertain during DNA synthesis by MsmPol1. The superposition indicates that the POL active site and the FEN/EXO active site are far apart in the Pol1 structure: at a distance of 50 Å between the divalent metal ions that demarcate the active sites.

The Pol1 FEN/EXO domain

A DALI search with the FEN/EXO domain returned three proteins as the top hits with Z scores >20. These are: Taq polymerase (pdb 1TAQ; Z score 26.1; rmsd of 2.1 Å at 245 Cα positions; 36% amino acid identity); mycobacterial FenA (pdb 6C33; Z score 25.3; rmsd of 2.8 Å at 256 Cα positions; 32% amino acid identity); and *E. coli* ExoIX (pdb 6C33; Z score 21.8; rmsd of 2.4 Å at 216 Cα positions; 28% amino acid identity). Figure 12A shows a superposition of the FEN/EXO structure on that of FenA. The Pol1 FEN/EXO tertiary structure comprises a six-strand β-sheet flanked by helices on both sides. The FEN/EXO model includes two Mn^{2+} ions in the active site (magenta spheres in Figure 12A and B) that correspond to the M1 and M2 manganese ions present in the FenA structure (cyan spheres in Figure 12A and B). There was no equivalent of the FenA M3 active site manganese in the Pol1 FEN/EXO structure,

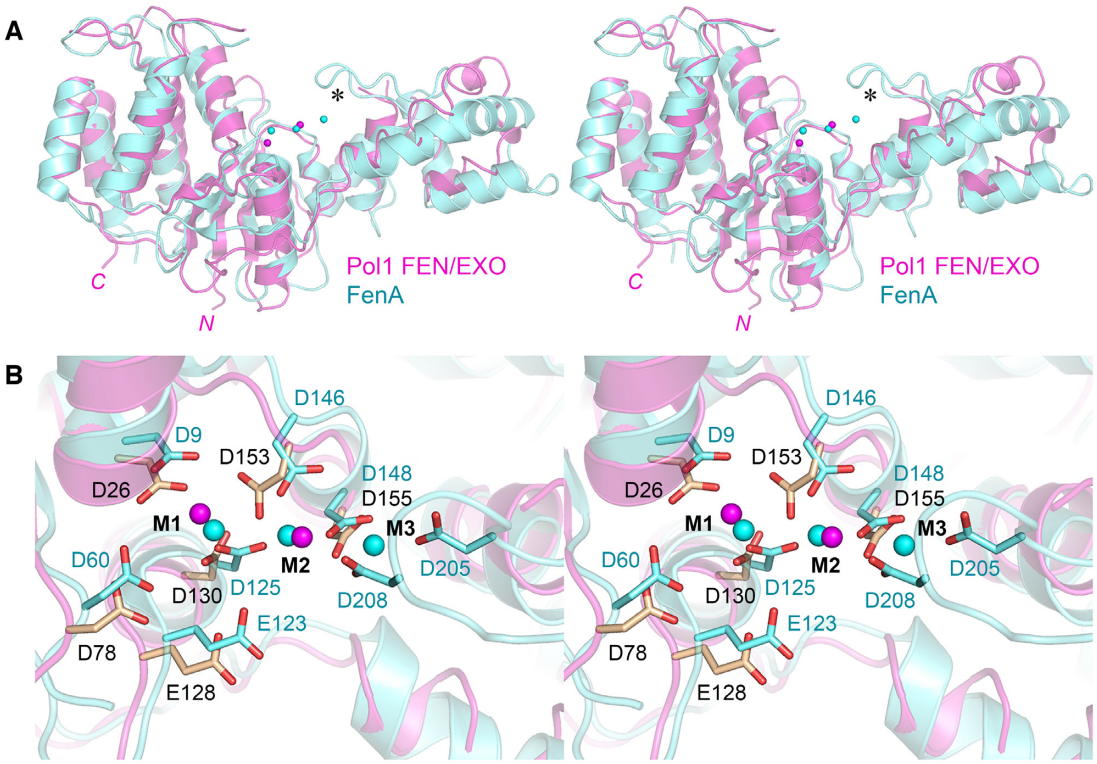

Figure 12. The Pol1 FEN/EXO domain. (**A**) Stereo view of the mycobacterial Pol1 FEN/EXO domain structure (colored magenta) superimposed on the structure of mycobacterial FenA (pdb 6C33, colored cyan). The N and C termini of the FEN/EXO domain are indicated. Active site manganese ions (two in the case of FEN/EXO and three in FenA) are shown as magenta or cyan spheres. The asterisk denotes a meta-binding loop in FenA that is disordered in the FEN/EXO structure. (**B**) Stereo view of the superimposed Pol1 FEN/EXO and FenA active sites, with the polypeptide and metal ions colored as in panel A. The M1, M2, and M3 metals are numbered according to their nomenclature in FenA. Eight conserved carboxylate side chains that coordinate the three metals in FenA are depicted as stick models with cyan carbons. Six equivalent side chains in FEN/EXO are shown as stick models with beige carbons; the other two carboxylates (equivalent to FenA Asp205 and Asp208) are disordered in FEN/EXO, thereby rationalizing the absence of an M3 metal in the FEN/EXO active site.

presumably because the equivalent of the FenA loop that coordinates the M3 metal (indicated by the asterisk in Figure 12A) is disordered in the Pol1 FEN/EXO domain structure.

The three active site manganese ions in FenA are coordinated by a rich network of direct and water-mediated contacts to the carboxylate oxygens of conserved acidic side chains (Figure 12B). The eight metal-binding residues in FenA that comprise the metal-binding sites are: Asp9, Asp60, Glu123 and Asp125 (site M1); Asp125, Asp146, Asp148 and Asp208 (site M2); Asp148, Asp205 and Asp208 (site M3). The eight counterparts in MsmPol1 FEN/EXO are Asp26, Asp78, Asp128, Asp130, Asp153 and Asp155 (located near the M1 and M2 metals in the Pol1 structure) as well as Asp204 and Asp207, which are in the disordered segment implicated in M3-binding and thus not present in the model. Mutational analysis of *M. tuberculosis* (Mtu) Pol1 showed that all eight metal-binding carboxylates are essential for MtuPol1′s 5′ exonuclease activity (14). The available structural and functional data suggest that the FEN/EXO domain of mycobacterial Pol1 adheres to a three-metal phosphodiesterase mechanism. All eight metal-binding carboxylates are conserved in the FEN/EXO domains of BstPol1, EcoPol1 and TaqPol1 (shaded in gold in Supplementary Figure S6).

DISCUSSION

This study extends previous biochemical characterizations of mycobacterial Pol1 (14,42) and provides the first structural snapshots of its POL domain in action. Thereby, Pol1 joins DnaE1, LigD POL and PolD1/Prim-PolC as the subset of mycobacterial DNA polymerases for which crystal structures have been solved (3,9,40). Only in the cases of LigD POL (44) and Pol1 (present study) have mycobacterial polymerases been crystallized in complexes with DNA.

As might be expected from primary structure conservation (Figure 1A and Supplementary Figure S6), the tertiary structure of the mycobacterial POL domain is organized along the same lines as BstPOL and TaqPOL with respect to its constituent palm, thumb, fingers, and vestigial exo subdomains. A distinctive feature of the mycobacterial POL revealed by the apoenzyme structure was the presence of a manganese-binding site in the vestigial exo subdomain. This site apparently does not bind magnesium even when it is present at high concentrations during crystallization. Several segments of the thumb and fingers modules that were disordered in the mycobacterial POL apoenzyme became fully ordered in the structure of POL that had catalyzed ddAMP addition to a DNA primer-template and engaged the next dNMP and metal cofactors in the POL active site. Indeed, the segments undergoing this disordered-

to-ordered transition from the apo to the DNA-bound state are ones that contact the DNA directly. The POL•DNA complex structure was of sufficient quality and resolution to delineate an extensive interface of the thumb, palm, and fingers subdomains with the primer-template DNA and the triphosphate moiety of the incoming dNTP (Figure 7C). A unique feature of the mycobacterial POL•DNA structure was the presence of a 'non-catalytic' octahedral magnesium complex engaged at the protein-DNA interface via water-bridged contacts to amino acids in the vestigial exo subdomain, the palm subdomain, and a backbone phosphate of the template DNA strand.

A key aspect of the POL transition from an 'open' conformation of the apoenzyme to a 'nearly closed' conformation of the ternary complex is the inward motion of the fingers subdomain to engage the primer-template and the dNTP triphosphate. Yet, by comparison to an analogous BstPOL structure, we see that the mycobacterial POL•DNA structure has not closed completely. The state of the POL active site in the present structure is not that of a true Michaelis complex, insofar as the catalytic magnesium imputed to bridge the primer 3′-OH and the dNTP α phosphate is displaced and instead contacts the phosphate of the 3′ terminal nucleotide of the primer strand. This is likely due to the absence of a 3′-OH at the ddAMP-terminated primer strand in the structure. (Note that the analogous dideoxy-terminated BstPOL structure lacked this magnesium altogether.) In addition, the O helix and its associated dNTP triphosphate•Mg^{2+} complex have not completed their movement toward the primer terminus in the mycobacterial POL•DNA structure. These caveats notwithstanding, the structure does provide a valuable picture of the ternary complex in an intermediate conformation along the mycobacterial POL reaction pathway.

The structures also allow for interpretation of the few mutational results reported for the polymerase component of mycobacterial Pol1. The Modak lab (42) had made alanine mutations at three residues in the fingers subdomain of *M. tuberculosis* Pol1: Gln683, Glu706 and Asp707, corresponding to *M. smegmatis* Pol1 amino acids Gln687 (which contacts the dNTP phosphate; Figure 9), Glu710 and Asp711 (located in the loop between the 'M' and 'N' helices of the fingers module). The DNA polymerase specific activities of the recombinant Gln-Ala, Glu-Ala, and Asp-Ala mutants were 76%, 45% and 5% of the wild-type specific activity (42). The benign effect of the Gln-Ala change suggests that its contact to the γ phosphate is functionally redundant to those of the O helix Lys and Arg side chains (Figure 9). The mutated Glu side chain is located on the surface of the Pol domain and makes no atomic contacts; thus, the reason for the 2-fold activity decrement is unclear. By contrast, the severe effect of the Asp-Ala mutation is most likely reflective of the fact that its *M. smegmatis* equivalent Asp711 forms a bidentate salt bridge with Arg733, an active site constituent in the O helix that makes a bidentate hydrogen bond to the dNTP γ phosphate. We infer that the Asp-Arg salt bridge helps tether the N and O helices and positions the Arg side chain for its interaction with the dNTP, whereby it (along with the triphosphate-bound magnesium complex) orients the pyrophosphate leaving group during the polymerization step. The Mizrahi lab (14) reported that changing *M. tuberculosis* Pol1 Tyr737 to phenylalanine resulted in a 6-fold decrement in DNA polymerase activity; the equivalent residue in *M. smegmatis* Pol1 is Tyr741 in the O helix, which donates a hydrogen bond from the tyrosine side chain hydroxyl to the dNTP β phosphate (Figure 9). We surmise that the loss of this hydrogen bond accounts for the activity decrement.

Two previous studies of the role of Pol1 in mycobacterial DNA metabolism were conducted by assessing clastogen sensitivity of *M. smegmatis* mutant strains with a *polA* gene encoding a C-terminally truncated Pol1 protein (aa 1–775). The Pol1-CΔ strains were sensitized to killing by UV irradiation and oxidizing agents hydrogen peroxide and cumene hydroxyperoxide (9,15). We can discern from the POL•DNA structure that this truncation eliminates the two palm subdomain β strands that contain the active site magnesium-binding residues; it also deletes secondary structure elements of the fingers subdomain that comprise part of the DNA interface. Thus, whereas it is certain that Pol1-(1–775) does not have DNA polymerase catalytic activity, it is not a 'pure' polymerase-defective mutant insofar as the truncation is likely to compromise DNA binding and, perhaps, protein-protein interactions of Pol1. It is also unclear whether the C-terminal truncation might affect the folding or activity of the N-terminal FEN/EXO domain. The separation-of-function Ala-Ala mutations in the FEN and POL active sites characterized in the present study should prove useful in assessing genetically the contributions of the FEN and POL enzymatic activities of Pol1 to *M. smegmatis* physiology.

The crystal structure of full-length MsmPol1, and a comparison to the TaqPol1 crystal structure reported previously (40), reveals that the relative positions of the POL and FEN/EXO domains can vary greatly by virtue of rigid-body mobility of FEN/EXO about an interdomain linker. The position of FEN/EXO in the MsmPol1 structure is incompatible with DNA synthesis by POL because it overlaps with the DNA duplex of the primer-template. Clearly, FEN/EXO must move out of the way during DNA synthesis. The key question is whether it does so, or can do so, in a manner that places the FEN/EXO active site in a position where it can access the 5′ end of a DNA duplex segment downstream of the POL active site during gap repair or access a 5′ flap downstream of the POL active site during strand displacement synthesis.

Because there are no structures available of any Pol1 POL enzyme in complex with a gapped primer-template that has a downstream duplex or 5′ flap-duplex nucleic acid component, we cannot judge exactly where FEN/EXO needs to be to gain such access. A recent study using single-molecule FRET and molecular modeling to probe the interaction of *E. coli* Pol1 POL with a 1-nucleotide gapped DNA duplex suggested that DNA is bent by 120° about the gap upon formation of a POL•DNA binary complex and that POL can also form a ternary complex with two POL protomers bound to the bent gapped DNA (45). Molecular dynamics simulations of a FRET-based structural model of a POL•gapped DNA binary complex suggested that the downstream duplex was melted out to form a 4- to 5-nucleotide flap (45). Their model is consistent with our inferences here, based on comparison of the flap endonuclease

cleavage products generated by full-length MsmPol1 versus the isolated Pol1-N FEN/EXO domain (Figure 2), that binding of POL to the flap nick DNA elicits local melting of up to 4 bp of the duplex downstream of the nick.

It has long been known that native bacterial Pol1 can perform nick translation, a process in which the polymerase activity resident in the POL domain adds dNMPs to the 3′-OH at a duplex nick and the 5′ exonuclease or flap endonuclease activity of the FEN/EXO domain degrades the non-template strand of the duplex downstream of the mobile nick. This mode of strand displacement synthesis by Pol1 has been an invaluable tool for the preparation of radiolabeled DNA probes (46). The salient mechanistic question is whether a single Pol1 enzyme protomer can execute both reactions in a single DNA-binding event (47). We envision two ways in which Pol1 could do this in *cis*: (i) by a large movement of the FEN/EXO domain in the context of an actively elongating POL domain that allows its access to the downstream flap; or (ii) an exchange of domains at the flap-nick so that the POL domain is displaced from the 3′-OH primer end, freeing up the flap nick for recognition by the FEN/EXO domain. The latter exchange could occur without Pol1 dissociation if the POL domain slides away from the nick to afford access to FEN/EXO and then slides back to re-engage the primer terminus in the POL active site.

An alternative model is that the polymerization and 5′ processing reactions are not performed by a single Pol1 protomer during a single DNA binding event. Plausible scenarios include: (i) simultaneous binding of two Pol1 enzyme protomers, one of which executes DNA synthesis by its POL domain and the other protomer catalyzes 5′ processing by its FEN/EXO domain; (ii) dissociation of Pol1 after polymerization and rebinding from solution to effect 5′ processing. To our knowledge, biochemical experiments to definitively discriminate among these various modes of action are not available. Obtaining a structure of Pol1 bound to a flap-nick DNA will go a long way toward clarifying the mechanism of Pol1 action during gap filling and Okazaki fragment processing.

DATA AVAILABILITY

Structural coordinates have been deposited in Protein Data Bank under accession codes 6VDC, 6VDD and 6VDE.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

U.S. National Institutes of Health [AI64693, GM126945]; the MSKCC structural biology core laboratory is supported by National Cancer Institute [P30-CA008748]; X-ray diffraction data were collected at synchrotron facilities supported by grants and contracts from the National Institutes of Health [P41GM103403, HEI-S10RR029205]; U.S. Department of Energy [DE-AC02-06CH11357]. Funding for open access charge: U.S. National Institutes of Health [AI64693].

Conflict of interest statement. None declared.

REFERENCES

1 Ditse,Z., Lamers,M.H. and Warner,D.Y. (2017) DNA replication in *Mycobacterium tuberculosis*. *Microbiol. Spectrum*, **5**, doi:10.1128/microbiolspec.
2 Rock,J.M., Lang,U.F., Chase,M.R., Ford,C.B., Gerrick,E.R., Gawande,R., Coscolla,M., Gagneux,S., Fortune,S.M. and Lamers,M.H. (2015) DNA replication fidelity in *Mycobacterium tuberculosis* is mediated by an ancestral prokaryotic proofreader. *Nat. Genet.*, **47**, 677–681.
3 Baños-Mateos,S., van Roon,A.M., Lang,U.F., Maslen,S.L., Skehel,J.M. and Lamers,M.H. (2017) High-fidelity DNA replication in *Mycobacterium tuberculosis* relies on a trinuclear zinc center. *Nat. Commun.*, **8**, 855.
4 Boshoff,H.I., Reed,M.B., Barry,C.E. and Mizrahi,V. (2003) DnaE2 polymerase contributes to *in vivo* survival and the emergence of drug resistance in *Mycobacterium tuberculosis*. *Cell*, **113**, 183–193.
5 Warner,D.F., Ndwandwe,D.E., Abrahams,G.L., Kana,B.D., Machowski,E.E., Venclovas,C. and Mizrahi,V. (2010) Essential roles for imuA'- and imuB-encoded accessory factors in DnaE2-dependent mutagenesis in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 13093–13098.
6 Zhu,H., Nandakumar,J., Aniukwu,J., Wang,L.K., Glickman,M.S., Lima,C.D. and Shuman,S. (2006) Atomic structure and NHEJ function of the polymerase component of bacterial DNA ligase D. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 1711–1716.
7 Aniukwu,J., Glickman,M.S. and Shuman,S. (2008) The pathways and outcomes of mycobacterial NHEJ depend on the structure of the broken DNA ends. *Genes Dev.*, **22**, 512–527.
8 Zhu,H., Bhattarai,H., Yan,H., Shuman,S. and Glickman,M. (2012) Characterization of *Mycobacterium smegmatis* PolD2 and PolD1 as RNA/DNA polymerases homologous to the POL domain of bacterial DNA ligase D. *Biochemistry*, **51**, 10147–10158.
9 Płociński,P., Brissett,N.C., Bianchi,J., Brzostek,A., Korycka-Machała,M., Dziembowski,A, Dziadek,J. and Doherty,A.J. (2017) DNA Ligase C and Prim-PolC participate in base excision repair in mycobacteria. *Nat. Commun.*, **8**, 1251.
10 Kana,B.D., Abrahams,G.L., Sung,N., Warner,D.F., Gordhan,B.G., Machowski,E.E., Tsenova,L., Sacchettini,J.C., Stoker,N.G., Kaplan,G. *et al.* (2010) Role of the DinB homologs Rv1537 and Rv3056 in *Mycobacterium tuberculosis*. *J. Bacteriol.*, **192**, 2220–2227.
11 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.
12 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.
13 Huberts,P. and Mizrahi,V. (1995) Cloning and sequence analysis of the gene encoding the DNA polymerase I from *Mycobacterium tuberculosis*. *Gene*, **164**, 133–136.
14 Mizrahi,V. and Huberts,P. (1996) Deoxy- and dideoxynucleotide discrimination and identification of critical 5′ nuclease domain residues of the DNA polymerase I from *Mycobacterium tuberculosis*. *Nucleic Acids Res.*, **24**, 4845–4852.
15 Gordhan,B.G., Andersen,S.J., De Meyer,A.R. and Mizrahi,V. (1996) Construction by homologous recombination and phenotypic characterization of a DNA polymerase domain *polA* mutant of *Mycobacterium smegmatis*. *Gene*, **178**, 125–130.
16 DeJesus,M.A., Gerrick,E.R., Xu,W., Park,S.W., Long,J.E., Boutte,C.C., Rubin,E.J., Schnappinger,D., Ehrt,S., Fortune,S.M. *et al.* (2017) Comprehensive essentiality analysis of the *Mycobacterium tuberculosis* genome via saturating transposon mutagenesis. *mBio*, **18**, e02133-16.
17 Lyamichev,V., Brow,M.A. and Dahlberg,J.E. (1993) Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. *Science*, **260**, 778–783.
18 Xu,Y., Derbyshire,V., Ng,K., Sun,X.C., Grindley,N.D. and Joyce,C.M. (1997) Biochemical and mutational studies of the 5′-3′ exonuclease of DNA polymerase I of *Escherichia coli*. *J. Mol. Biol.*, **268**, 284–302.

19 Lyamichev,V., Brow,M.A., Varvel,V.E. and Dahlberg,J.E. (1999) Comparison of the 5′ nuclease activities of Taq DNA polymerase and its isolated nuclease domain. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 6143–6148.
20 Uson,M.L., Ghosh,S. and Shuman,S. (2017) The DNA repair repertoire of *Mycobacterium smegmatis* FenA includes the incision of DNA 5′ flaps and the removal of 5′ adenylylated products of aborted nick ligation. *J. Bacteriol.*, **199**, e00304-17.
21 Uson,M.L., Carl,A., Goldgur,Y. and Shuman,S. (2018) Crystal structure and mutational analysis of *Mycobacterium smegmatis* FenA highlight active site amino acids and three metal ions essential for flap endonuclease and 5′ exonuclease activities. *Nucleic Acids Res.*, **46**, 4164–4175.
22 Adams,P.D., Afonine,P.V., Bunkóczi,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 Cryst.*, **D66**, 213–221.
23 Kelley,L.A., Mezulis,S., Yates,C.M., Wass,M.N. and Sternberg,M.J (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, **10**, 845–858.
24 Jones,T.A., Zou,J.Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst.*, **A47**, 110–119.
25 Fukushima,S., Itaya,M., Kato,H., Ogasawara,N. and Yoshikawa,H. (2007) Reassessment of the in vivo function of DNA polymerase I and RNase H in bacterial cell growth. *J. Bacteriol.*, **189**, 8575–8583.
26 Gupta,R., Chatterjee,D., Glickman,M.S. and Shuman,S. (2017) Division of labor among *Mycobacterium smegmatis* RNase H enzymes: RNase H1 activity of RnhA or RnhC is essential for growth whereas RnhB and RnhA guard against killing by hydrogen peroxide in stationary phase. *Nucleic Acids Res.*, **45**, 1–14.
27 Keskin,H., Shen,Y., Patel,M., Yang,T., Ashley,K., Mazin,A.V. and Storici,F. (2014) Transcript-RNA-templated DNA recombination and repair. *Nature*, **515**, 426–439.
28 Shen,Y., Nandi,P., Taylor,M.B., Stuckey,S., Bhadsavle,H.P., Weiss,B. and Storici,F. (2011) RNA-driven genetic changes in bacteria and human cells. *Mutation Res.*, **717**, 91–98.
29 Ricchetti,M. and Buc,H. (1993) *E. coli* DNA polymerase I as a reverse transcriptase. *EMBO J.*, **12**, 387–396.
30 Holm,L., Kaariainen,S., Rosenstrom,P. and Schenkel,A. (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics*, **24**, 1780–2781.
31 Kiefer,J.R., Mao,C., Hansen,C.J., Basehore,S.L., Hogrefe,H.H., Braman,J.C. and Beese,L.S. (1997) Crystal structure of a thermostable *Bacillus* DNA polymerase I large fragment at 2.1 Å resolution. *Structure*, **5**, 95–108.
32 Derbyshire,V., Pinsonneault,J.K. and Joyce,C.M. (1995) Structure-function analysis of 3′-to-5′ exonuclease of DNA polymerases. *Meth. Enzymol.*, **262**, 363–385.
33 Johnson,S.J., Taylor,J.S. and Beese,L.S. (2003) Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 3895–3900.
34 Kiefer,J.R., Mao,C., Braman,J.C. and Beese,L.S. (1998) Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. *Nature*, **391**, 304–307.
35 Li,Y., Korolev,S. and Waksman,G. (1998) Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.*, **17**, 7514–7525.
36 Batra,V.K., Beard,W.A., Schock,D.D., Krahn,J.M., Pedersen,L.C. and Wilson,S.H. (2006) Magnesium-induced assembly of a complete DNA polymerase catalytic complex. *Structure*, **14**, 757–766.
37 Wang,W., Wu,E.Y., Hellinga,H.W. and Beese,L.S. (2012) Structural factors that determine selectivity of a high fidelity DNA polymerase for deoxy-, dideoxy-, and ribonucleotides. *J. Biol. Chem.*, **287**, 28215–28226.
38 Miller,B.R., Beese,L.S., Parish,C.A. and Wu,E.Y. (2015) The closing mechanism of DNA polymerase I at atomic resolution. *Structure*, **23**, 1609–1620.
39 Chim,N., Jackson,L.N., Trinh,A.M. and Chaput,J.C. (2018) Crystal structures of DNA polymerase I capture novel intermediates in the DNA synthesis pathway. *eLife*, **7**, e40444.
40 Kim,Y., Eom,S.H., Wang,J., Lee,D.D, Suh,S.W. and Steitz,T.A. (1995) Crystal structure of *Thermus aquaticus* DNA polymerase. *Nature*, **376**, 612–616.
41 Eom,S.H., Wang,J. and Steitz,T.A. (1996) Structure of *Taq* polymerase with DNA at the polymerase active site. *Nature*, **382**, 278–281.
42 Arrigo,C.J., Singh,K. and Modak,M.J. (2002) DNA polymerase I of *Mycobacterium tuberculosis*: functional role of a conserved aspartate in the hinge joining the M and N helices. *J. Biol. Chem.*, **277**, 1653–1661.
43 Pitcher,R.S., Brissett,N.C., Picher,A.J., Andrade,P., Juarez,R., Thompson,D., Fox,G.C., Blanco,L. and Doherty,A.J. (2007) Structure and function of a mycobacterial NHEJ DNA repair polymerase. *J. Mol. Biol.*, **366**, 391–405.
44 Brissett,N.C., Martin,M.J., Pitcher,R.S., Bianchi,J., Juarez,R., Green,A.J., Fox,G.C., Blanco,L. and Doherty,A.J. (2011) Structure of a preternary complex involving a prokaryotic NHEJ DNA polymerase. *Mol. Cell*, **41**, 221–231.
45 Craggs,T.D., Sustarsic,M., Plochowietz,A., Mosayebi,M., Kaju,H., Cuthbert,A., Hohlbein,J., Domicevica,L., Biggin,P.C., Doye,J.P. *et al.* (2019) Substrate conformational dynamics facilitate structure-specific recognition of gapped DNA by DNA polymerase. *Nucleic Acids Res.*, **47**, 10788–10800.
46 Rigby,P.W., Dieckmann,M., Rhodes,C. and Berg,P. (1977) Labeling of deoxyribonucleic acid to high specific activity by in vitro nick translation with DNA polymerase I. *J. Mol. Biol.*, **113**, 237–251.
47 Xu,Y., Grindley,N.F. and Joyce,C.M. (2000) Coordination between the polymerase and 5′-nuclease components of DNA polymerase I of *Escherichia coli*. *J. Biol. Chem.*, **275**, 20949–20955.