Structure and Function of a Mycobacterial NHEJ DNA Repair Polymerase

R S Pitcher, Nigel Brissett, A J Picher, P Andrade, R Juarez, D Thompson, G C Fox, L Blanco, Aidan Doherty

Publication date
01-02-2007

Published in
Journal of Molecular Biology

Link to external publisher version
https://doi.org/10.1016/j.jmb.2006.10.046

Structure and Function of a Mycobacterial NHEJ DNA Repair Polymerase

Article (Unspecified)

Structure and Function of a Mycobacterial NHEJ DNA repair polymerase

Robert S. Pitcher[1], Nigel C. Brissett[1], Angel J. Picher[2], Paula Andrade[2], Raquel Juarez[2], Darren Thompson[3], Gavin C. Fox[4], Luis Blanco[2] and Aidan J. Doherty[1]*

[1]Genome Damage and Stability Centre, University of Sussex, Brighton BN1 9RQ, UK.

[2]Centro de Biología Molecular "Severo Ochoa", CSIC-UAM, Cantoblanco, Madrid, Spain.

[3]Department of Biochemistry, University of Sussex, Brighton BN1 9RQ, UK.

[4]European Synchrotron Radiation Facility, BP-220, F-38043, Grenoble CEDEX 9, France

*Corresponding author:

Email: ajd21@sussex.ac.uk

Subject categories-: Structural biology; DNA; Microbiology; Repair

Key words: NHEJ; DNA breaks; Ligase D; primase; polymerase

Abstract

Non homologous end-joining (NHEJ)-mediated repair of DNA double-strand breaks in prokaryotes requires Ku and a specific multidomain DNA ligase (LigD). Here we present crystal structures of the primase/polymerization domain (PolDom) of *Mycobacterium tuberculosis* LigD, either alone or complexed with nucleotides. The PolDom structure combines the general fold of the archaeo-eukaryotic primase (AEP) superfamily with additional loops and domains that together form a deep cleft on the surface, likely used for DNA binding. Enzymatic analysis indicates that the PolDom of LigD, even in the absence of accessory domains and Ku proteins, has the potential to recognize DNA end-joining intermediates. Strikingly, one of the main signals for the specific and efficient binding of PolDom to DNA is the presence of a 5′-phosphate, located at the single/double-stranded junction at both gapped and 3′-protruding DNA molecules. Although structurally unrelated, Pol λ and Pol μ, the two eukaryotic DNA polymerases involved in NHEJ, are endowed with a similar capacity to bind a 5′-phosphate. Other properties that are beneficial for NHEJ, such as the ability to generate template distortions and realignments of the primer, displayed by Pol λ and Pol μ, are also shared by the PolDom of bacterial LigD. In addition, PolDom can also perform non-mutagenic translesion synthesis on termini containing modified bases. Significantly, ribonucleotide insertion also appears to be a recurrent theme associated with NHEJ, maximized in this case by the deployment of a dedicated primase, although its *in vivo* relevance is presently unknown.

Introduction

DNA double-strand breaks (DSB) are particularly dangerous lesions because of the loss of integrity of both strands of the DNA duplex. These lesions are repaired by two fundamentally different processes, homologous recombination (HR) and non-homologous end-joining (NHEJ) that can be distinguished based on the involvement of a DNA sequence that is homologous to the break site.[1] In HR, the repair mechanism utilizes an intact duplex that is homologous to the DSB site to guide the repair event and accurately restore the DNA structure and sequence at the break site. By contrast, in NHEJ-mediated DSB repair DNA ends are brought together, processed and then directly joined together. This process has the potential to be error-prone, leading to genetic alterations ranging from the loss or the addition of a few nucleotides at the break site, or even the joining of previously unlinked DNA molecules.[1]

Unusually, key factors in NHEJ, the Ku70/Ku80 heterodimer, XRCC4 and DNA ligase IV, were first identified in mammals. Subsequent studies have revealed the presence of functionally homologous factors in the lower eukaryote *Saccharomyces cerevisiae*,[2] and more recently, in prokaryotes[3], indicating that the mechanism of NHEJ has been conserved throughout evolution. The DNA end-binding Ku heterodimer and DNA-PKcs are assembled together on a DNA end to constitute the DNA-dependent protein kinase (DNA-PK) in higher eukaryotes. Both have been shown to have end-bridging activity,[4,5] as has the yeast Rad50/Mre11/Xrs2 complex.[6] Furthermore, each of these factors functionally interacts with the DNA ligase complex, DNA ligase IV/XRCC4 in mammalian cells and Dnl4/Lif1 in yeast, that completes the repair pathway. [6-8]

The majority of DSBs generated *in vivo* have ends that are non-complementary and may also have damaged *termini*. Such DNA ends require processing by factors including nucleases and DNA polymerases to generate ligatable termini. Many DNA polymerases, including Pol μ Pol λ and TdT in mammals and pol4 in yeast,[9-11] and the nucleases Artemis, Mre11 and FEN-1,[11-14] have been implicated in eukaryotic NHEJ. In contrast, the prokaryotic NHEJ pathway appears to consist of only two repair factors (Ku and Ligase D; LigD) that form a fully functional two-component DSB repair complex.[3,15] Many of the Ku-associated ligases (LigD) possess modular domains with end-processing activities,[15-21] including polymerases that belong to the Archaeo-Eukaryotic Primase (AEP) family and DNA exonuclease domains capable of removing single-stranded 3' overhangs. These domains play a direct role in resecting the *termini* of DSBs prior to ligation[15] and can function independently of one another.[20,21] These additional activities appear to compensate for the lack of the various processing factors implicated in eukaryotic NHEJ. Although bacterial Ku is not directly involved in end-processing, it is essential for recruiting LigD, *via* a direct interaction with the polymerase domain, to DSBs and may also co-ordinate the order and extent of resection.[20] Thus, bacterial NHEJ-mediated DSB repair can be reconstituted, both *in vitro* and *in vivo*, simply by the addition of Ku and LigD.[15-20] Interestingly, LigD and Ku are genetically associated in many diverse bacterial species suggesting that NHEJ is functionally relevant to many prokaryotes.[19] It has been proposed that bacteria in stationary phase rely on NHEJ, in an analogous manner to non-cycling eukaryotic cells.[3,15]

LigD from *Mycobacterium tuberculosis* (*Mt*-LigD) and *Mycobacterium smegmatis* (*Ms*-LigD) is composed of polymerase (*Mt*-PolDom/*Ms*-PolDom) and nuclease domains that reside as N-

terminal distal and proximal extensions of the ligase domain, respectively. *Mt*-PolDom possesses a remarkable variety of nucleotidyltransferase activities including DNA-dependent RNA primase, terminal transferase and DNA-dependent DNA/RNA gap-filling polymerase activities.[3,15,16,20] Previously, we reported that the *Mt*-PolDom (amino acids 1-300), when expressed in isolation, retains both DNA end-filling and primase activities.[20] Here we present the crystal structure of the polymerase domain from *M. tuberculosis*, and the co-crystal structures with either GTP or dGTP bound in the active site. We also describe a detailed functional characterisation of the interaction between either *Mt*-PolDom or *Ms*-PolDom and various NHEJ-relevant DNA substrates that illuminates the basic *modus operandi* of these versatile DSB repair polymerases.

Results

Structure of the NHEJ polymerase domain of *Mt* DNA ligase D

Many bacterial NHEJ DNA ligases contain additional catalytic domains that encode specific polymerase and nuclease activities.[15] These domains play a direct role in resecting the *termini* of DSBs prior to ligation and each of the domains (ligase, polymerase and nuclease) can function independently of one another. Previously, we reported that the polymerase domain (amino acids 1-300) of the *M. tuberculosis* NHEJ ligase D can be readily over-expressed in *E. coli* and purified to near homogeneity using nickel affinity and size-exclusion chromatography steps.[20] The recombinant polymerase domain (PolDom) retains both DNA end-filling polymerase activities and also possesses an intrinsic primase activity.[20]

To understand the molecular basis of these, and other polymerase activities ascribed to this domain, we have crystallized both native and the selenomethionine-substituted PolDom mutant (F64L) and also co-crystallized PolDom with either dGTP or GTP and divalent metals (see Materials & Methods). The native crystals diffracted to beyond 1.6Å, whilst the co-crystals diffracted to ~1.8 Å, and belong to the monoclinic space group $P2_1$ or orthorhombic space group $P2_12_12_1$ with typical unit cell dimensions of a = 41.0, b = 75.8, c = 96.1 Å and β = 92.7º. There are two molecules in the asymmetric unit, giving a V_m of 2.13 $Å^3$ Da^{-1}, corresponding to 41.8% solvent content. The structure of selenomethionine-substituted PolDom (F64L) was determined to 1.8 Å using MAD phasing methods (see Materials and Methods) and the refinement statistics are shown in Table 1. The native structure and co-crystal structures with bound nucleotides were determined to resolutions of 1.65 Å, 1.8 Å and 1.78 Å by molecular replacement using the selenomethionine-substituted structure (see Materials and Methods). The structure comprises residues 2 to 288 of the Mt-PolDom

sequence with no density observed for 12 amino acids at the C-terminal end. Over 90% of the residues in the structure are in the most favorable region of the Ramachandran plot (Table I).

The overall architecture of the N-terminal polymerase domain of *Mt*-ligase D is shown in Figure 1(a). PolDom is a globular monomeric protein with a mixed α+β structure and is composed of three sub-domains (two mixed α+β and one β domain) with 18 β-stands and 6 α-helices on the periphery of the structure. The central core of the PolDom molecule is composed of two β-sheets, comprising of a four-stranded anti-parallel sheet (β4, 3, 7 and 15) and a six-stranded anti-parallel β-sheet (β13, 14, 10, 12, 11 and 17). The single β-domain is made up of two distinct β-sheets (β5,6 and β8,9) which protrude from the globular structure, forming the upper edge of the active site cleft.

Searches of the protein structural data base (PDB) using the SSM server (http://www.ebi.ac.uk/msd-srv/ssm) revealed that PolDom shares significant structural homology with the catalytic domains of the DNA primases of the archaea *Sulfolobus solfataricus,*[22] *Pyrococcus furiosus* [23] and *Pyrococcus horikishii*[24] (Figure 1(b)). For example, *Mt*-PolDom and *S. solfataricus* primase structures align with a *Z* score of 4.4 and a root-mean-square (rms) deviation of 2.4 Å over 167 Cα positions. PolDom also shares significant structural homology with the pRN1 DNA primase-polymerase domain from *Sulfolobus islandicus*[25] (*Z* score of 2.7 and rms deviation of 3.0 over 116 Cα positions). Although these primases share a common catalytic core, many of the proteins contain additional loops and domains (Figure 1(b)) that are likely to determine the substrate specificity, and hence the specific role of these primases in DNA metabolism. These findings confirm previous predictions that the prokaryotic NHEJ polymerases do not belong to the eukaryotic polymerase

X family (that includes the NHEJ repair polymerases, Pol μ, Pol λ and TdT) but rather are members of the archaeo-eukaryotic primase (AEP) superfamily,[15,26] although unlike their close relatives they have evolved to have a role in DSB repair processes rather than in priming DNA replication.[27]

Structures of the polymerase active site with and without bound nucleotide

To understand the molecular basis of nucleotide binding by PolDom, we elucidated a number of crystal structures of the polymerase bound to either dGTP or GTP, in the presence of manganese. In both structures, the nucleoside triphosphates substrates are bound in a large positively charged central cleft on the surface of the protein (Figure 2(a); see also Supplementary Figure1 online), although the relative orientation of the respective guanine base differs significantly (see below). A closer view of the structure and molecular interactions in the active site of PolDom are shown in Figure 2. In the apo-enzyme structure a large number of ordered water molecules are observed in the active site (Figure 2(b)). Many of these hydrogen-bonding interactions between conserved active site residues are lost upon binding of the nucleotide (Figure 2(c),(d)) and replaced with nucleotide/metal-specific contacts. We can group the atomic interactions within the active site into three overall categories: (1) the metal binding sites, (2) triphosphate contacts and, (3) nucleoside contacts. Overall, the conformation of the triphosphate tail and the positioning and co-ordination of the metal ion (manganese) is nearly identical in both the GTP (Figure 2(c)) and dGTP (Figure 2(d)) co-crystal structures. In the co-crystal structures, a Mn^{2+} ion is co-ordinated in an octahedral complex with contacts being made with the δ−oxygens of the essential catalytic aspartate residues (Asp137 and Asp139), a couple of ordered water molecules and with non-bridging oxygens of the β− and γ−phosphates of the triphosphate tail. Interestingly, Asp139 is pointing away from the Mn binding site in the apo-enzyme but undergoes a conformational change upon addition of GTP,

causing it to rotate towards the other conserved aspartates. The utilization of a two-metal catalysis mechanism is highly conserved among DNA polymerases.[28] Although a single metal ion is present in the native co-crystal structures, we did observe a second metal binding site in the active site of co-crystal structures of a mutant form of PolDom (F64L; Supplementary Figure 2 online). The second metal (magnesium) binding site, is partially coordinated in a distorted tetrahedral complex by δ–oxygens of Asp137 and Asp139, as well as to an ordered water molecule shared with the Mn^{2+} site. In addition, the side-chain of Asp227 also coordinates the Mg^{2+} ion. Based on active-site models described for other polymerases, we predict that the Mg^{2+} ion is most likely positioned to activate the 3'-OH of the primer for an in-line attack on the α-phosphate of the bound NTP, whilst the location of Mn^{2+} makes it likely that it acts to stabilize the optimal conformation of the pyrophosphate (PPi) leaving group. Although two different metal ion species are present in the structures, it remains to be determined which is the divalent metal of choice *in vivo*.

The triphosphate tail of the bound GTP is stabilized by a number of specific interactions with the manganese ion, specific amino acid residues and with ordered water molecules. Arg244 (a His residue in archaea and eukarya) co-ordinates to the non-bridging oxygens of the α– and β–phosphates, His178 (invariant in all groups of eukaryotic-type primases) co-ordinates the β-phosphate non-bridging oxygen via Nε and the γ-phosphate hydrogen-bonds to a number of conserved residues (Ser174, Lys175 and Gly176), as well as making a number of water and metal-mediated contacts in the active site. Ser172, which is invariant in all groups of AEPs, also interacts with the γ-phosphate.

The most profound difference between the two nucleotide-bound structures (GTP vs dGTP) is the position adopted by the nucleosides in the active site (Figure 2(c),(d)). In the

PolDom-GTP structure, the guanosine moiety is positioned away from the triphosphate tail with the guanine base stacking against the ring of a conserved phenylalanine (Phe64) and it also interacts with Lys52 (via N7 of the purine). This docking exposes the guanosine base to the solvent but positions the ribose hydroxyl groups towards the active site. The nucleoside is held in this orientation via direct interactions with the ribose moiety. These contacts include hydrogen bonding between the ribose hydroxyls O2' and O3' with the oxygen atoms of Thr236 (invariant in bacterial AEPs; invariant Arg in archaea; invariant His in eukarya) and His111 (Nδ) (invariant in bacterial AEPs). In contrast, the deoxyguanosine of dGTP appeared to bind less specifically, with the nucleoside adopting a number of orientations in the active site of crystal structure. In the co-crystal structure (PolDom:dGTP), refined with the best occupancy for the dGTP nucleotide, the nucleoside adopts a conformation that positions the deoxyribose moiety in a more solvent-exposed position than is observed in the GTP complex. The deoxyribose O3' hydrogen bonds to the ε−oxygen atom of side-chain of Gln230 (invariant in *Mt*-like AEPs; substituted by an Arg in *Pae*-LigD). The base still stacks with Phe64 (invariant in bacterial AEPs), but instead the base interacts the γ−oxygen of Thr236 (via N7 of the purine). In the dGTP complex and the apo-enzyme structures, an ordered water molecule occupies the same position as the ribose O2' in the GTP complex and engages in similar interactions with residues in the active site. These two distinct molecular "snap-shots" of different nucleotides bound to the polymerase suggest that the observed binding modes may be in equilibrium between a productive mode (GTP) and an unproductive one (dGTP). It also suggests that the enzyme has a stronger preference for binding NTPs over dNTPs.

Terminal transferase activity of PolDom preferentially inserts ribonucleotides

The terminal transferase activity of the PolDom of *Mt*- and *Ms*-LigD was assessed using homopolymeric ssDNA substrates, conditions that are unambiguous in comparison with others

that use heteropolymeric ssDNA, the latter prone to self-priming (and therefore templated) events. Thus, in these strict conditions, Pol μ, like TdT, displays significant terminal transferase activity, preferentially activated by Mn^{+2} ions and using dNTPs as substrates[29] (Figure 3). When using NTPs, Pol μ showed a more restricted pattern of terminal transferase extension, being limited to the addition of either a single or a few nucleotides. In these conditions, in the presence of magnesium or manganese ions and with either dNTPs or NTPs, Pol λ was inactive (data not shown). Mt-PolDom displayed an intrinsic terminal transferase activity also on a homopolymeric substrate, preferentially inserting ribonucleotides versus deoxynucleotides. As in the case with Pol μ, terminal transferase extension with NTPs was restricted to a few nucleotide units, and preferentially activated by manganese versus magnesium ions (Figure 3). Very similar results were obtained with the PolDom of Ms-LigD (data not shown). When the terminal transferase activity was assayed on a double-stranded and blunt ended DNA substrate (Figure 3) the activity had the same specificity for NTPs and Mn^{+2} ions, but it was more prominent, perhaps as a consequence of an improved binding to dsDNA versus ssDNA substrates. Therefore, the capacity of PolDom to carry out the preferential insertion of rNTPs in the absence of template information (terminal transferase) is in agreement with the structural information provided here, in which an enzyme/NTP complex appears to be stable in the absence of a template, and in an active conformation compared to the enzyme/dNTP complex. It is also consistent with previous reports on LigD that showed proficient single nucleotide non-templated additions on DSBs with a marked preference for NTP over dNTP incorporation.[15,21] The larger degradation of the primer observed in the presence of manganese appears to be due to traces of a contaminant 3´-5- exonuclease accompanying the MtPolDom fraction.

Preferences for DNA binding parallel optimal polymerization in a gap

In Pol ß, a structural domain named "8 kDa" is very important for DNA binding at a gap, and it also contains a dRP lyase activity required for its role in base excision repair (BER). [30,31] A key feature of this domain is a cluster of positively charged residues that strongly bind the 5´recessive end, particularly if it is endowed with a terminal phosphate.[30] In gap filling studies, it was demonstrated that the presence of a 5´-phosphate makes processive the polymerization reactions required to fill-in the gap[32].

Using electromobility gel-shift assays (EMSA), we observed that the PolDom of *Mt*-LigD has a strong preference for binding gapped DNA molecules having a 5´-phosphate at the gap (Figure 4 (a)), giving rise to a single and sharp retardation band, using as little as 5 ng of protein, that represents a stable enzyme/DNA complex. Similar results were also obtained by using the PolDom of *Ms*-LigD (data not shown). In support of the functional significance of the enzyme/DNA complex observed, Figure 4(b) shows that the templated insertion of C occurred much more efficiently on the 5' P-containing gapped molecule, in agreement with the preferred binding of PolDom to this DNA substrate. Insertion of dC also occurred efficiently on the 5´ P-gapped substrate, although at a lower rate in comparison with the insertion of C. Similarly to template-independent reactions (terminal transferase), manganese ions (1 mM) rather than magnesium ions (3 mM) stimulated the templated insertion of both C and dC nucleotides. The magnesium-activated insertion of NTPs versus dNTPs, assayed in a 1nt-gapped DNA substrate containing a 5´ P, was estimated to be about 20-70 fold higher, depending on each particular base pair (see Supplementary Figure 3 online).

An important matter is the insertion fidelity of PolDom on its preferred DNA substrates.

It is very likely that the optimal interaction with DNA could affect selectivity of the incoming nucleotides, and a more efficient use of the dNTP substrates. This is particularly important because, in the absence of DNA, dNTPs can bind to the enzyme in a "sub-optimal", perhaps inactive conformation (as shown here, terminal transferase activity was very poor with these substrates). For a qualitative analysis, we selected 1nt-gapped DNA substrates having the 4 possible template bases (Figure 5), and tested +1 extension of the labelled primer by addition of each independent deoxy or ribonucleotide at different concentrations in the presence of magnesium ions. In all cases, the template nucleotide directed the preferential addition of the complementary nucleotide (note the different concentrations used for complementary versus non-complementary nucleotides, and also the different concentrations used for ribo versus deoxynucleotides). As previously reported on different DNA substrates, there is a significant level of misincorporation, irrespective of using either deoxy or ribonucleotides as substrates. This suggests that misinsertion is occurring after template stabilization of an active conformation with both nucleotides. In summary, the PolDom of *Mt*- and *Ms*-*Lig*D strongly bind gapped DNA molecules containing a 5'-phosphate, which represents an important NHEJ intermediate, mimicking the bridging step on an end-joining reaction. This strong interaction is functional, as it improves polymerization and nucleotide selection on these substrates. As the structure shows, *Mt*-PolDom lacks an equivalent "8 kDa" domain and showed no detectable dRP lyase activity (data not shown). These enzymes must therefore utilize a novel structural element to facilitate interaction with the 5'-phosphate moiety.

Efficient binding to 3´-protruding DNA ends supports a role for PolDom in NHEJ

As recently demonstrated, Pol μ and Pol λ are capable of bridging two 3´ protruding DNA ends, depending on the level of existing microhomology.[33-35] Pol λ requires some complementarity between the 3´-protruding ends, whereas Pol μ is capable of "aligning"

ends even without any investment in base pairing, thus supporting a non-redundant role for these two DNA polymerases in NHEJ. Interestingly, both enzymes have a similar preference for binding DNA with a phosphate at the 5´ recessive end of a gap (our unpublished data), in agreement with the conservation of the "8 kDa" domain. Therefore, and especially for those polymerases dealing with broken ends (and with 3´-protruding ends in particular), a tight interaction with the 5'P-containing recessive end could be generally required as a first step of the end-joining reaction.[33-35] A second step would imply the connection between the two 3´ protruding DNA ends. Strikingly, as shown in Figure 6, the PolDom of LigD can form a similarly stable enzyme/DNA binary complex even in the absence of a primer strand (P), provided that the 5´-phosphate at the downstream (D) oligonucleotide is maintained. Identical results were obtained with the PolDom of either *Mt*-LigD or *Ms*-LigD. The capacity to bind T/D DNA substrates, containing a 5´P in the downstream oligo, would likely be very important to improve any further reactions occurring at the 3´ end of the template (via end-joining or self annealing), as shown in Figure 6. In agreement with this conclusion, the insertion reaction was much less efficient using a T/D molecule without the 5´P in the downstream oligo (data not shown).

Dislocation and realignment capacity of PolDom

Pol μ is the paradigm of a DNA polymerase with a very potent template dislocation and realignment capacity.[36,37] When assayed on template/primer molecules, or gaps longer than 1 nucleotide, Pol μ extends the primer by inserting nucleotides that are complementary to templating bases that are distal to the primer-terminus. Such behaviour is based on Pol μ´s intrinsic capacity to dislocate (distort) one or more proximal templating bases, thus generating

frameshift deletions and base substitutions.[36,37] This dislocation capacity is instrumental for end-joining of DNA ends with limited complementarity. To analyze if the PolDom of *Mt*- and *Ms*-LigD also have dislocation capacity, we analyzed nucleotide incorporation in a 2nt gapped DNA. As shown in Figure 7(a), in addition to the preferred incorporation of either dC or C, complementary to the first position of the gap, insertion of dG or G, complementary to the second base of the gap also occurred very efficiently, suggesting that it is templated by the second base (dC) of the gap after dislocation of the first templating base (dG). This is particularly obvious when using the *Ms*-PolDom, which has a higher specific activity. In this case, when adding either dG or G, 2 nucleotide units are inserted, filling in the gap. This result suggests that in addition to the dislocation capacity, PolDom is capable of extending both a dG:dGMP and a dG:GMP mismatch.

The capacity to accept distorting nucleotides is very important in allowing the connection of two protruding 3´-ends with imperfect complementarity. An experiment that substantiates this capacity, beneficial for NHEJ, is shown in Figure 7b. In this case, the DNA substrate is a template/primer molecule whose 3´-terminus is not properly paired. As a control, it can be seen that Pol μ uses its capacity to accept distorted nucleotides in the template strand (2 nt in this case) to induce a primer realignment, thus forming a correct pair between the 3´-terminal base of the primer (dC) and a complementary base (dG) available in the template (see scheme in Figure 7(b)). Thus the mismatched end can be efficiently extended by Pol μ only with a dGTP residue, complementary to the template after primer relocation. *Ms*-PolDom was also capable of inducing realignment of the mismatched end, as deduced from the insertion of either dGTP or GTP (the latter not shown). Moreover, the alternative insertion of dTTP suggests that after primer realignment (TC dislocated), the second available template base (dA) can also direct the

extension reaction, by accepting a second dislocation (C dislocated) as described above (Figure 7(a)). Similar results were obtained with *Mt*-PolDom (data not shown). In conclusion, the suggested parallels with Pol μ are strengthened by these results that show that PolDom has an intrinsic capacity to dislocate (distort) one or more proximal templating bases and can induce realignment of a mismatched end, two important properties that have been implicated in microhomology-mediated end joining.

Polymerization at 8-oxoG containing templates by PolDom is preferentially non-mutagenic

Translesion synthesis at damaged templates requires an insertion step, usually carried out by a specialized DNA polymerase, followed by an elongation step than can require an additional specialized DNA polymerase before resuming long-term elongation by a processive replicase. However, there are other situations in which a damaged template base must be accepted either for base-pairing or "to instruct" a DNA repair reaction uncoupled to replication. A relevant situation concerns the repair of a double-strand break by the NHEJ pathway. If the break is associated with DNA damage, the bridging step and filling-in reactions required to seal the break could incidentally require base-pairing with damaged bases in the template strand. Thus, we first analyzed the capacity of the *Mt*-PolDom and *Ms*-PolDom to synthesize in front of a 8-oxoG template base. The substrate selected was a 1 nt-gapped DNA, having a 5´-phosphate, in which 8-oxoG was the templating base (see Figure 8(a), left panels). Such a substrate could serve to mimic the gap-filling reaction that follows an end-joining/bridging reaction. Interestingly, as shown in Figure 8(b) (left panels), insertion of dCTP (the correct nucleotide) in front of 8-oxoG occurred very efficiently, even greater than the insertion rate of dC in front of a non-damaged dG templating base. Conversely, and in contrast to most DNA-dependent

DNA polymerases with the exception of human Pol η, insertion of dA in front of 8-oxoG (a mutagenic event) by PolDom was very inefficient. The dC/dA ratio of insertion in front of 8-oxoG was 12-fold for *Mt*-PolDom and 27-fold for *Ms*-PolDom. When the elongation step was studied (Figure 8(a) & (b), right panels), there was a clear preference to extend the 8-oxoG:dCMP pair versus a 8-oxoG:dAMP pair. Strikingly, the efficiency of extension of the "correct" 8-oxoG:dCMP pair was even better than that corresponding to a "bona fide" dG:dCMP pair. Again, it is worth noting that other DNA polymerases have a stronger preference (~27-fold in the case of Polδ) for extending the 8-oxoG:dAMP pair.

Interestingly, 8-oxoG was also taken as an acceptable template for the insertion of ribonucleotides (the preferred nucleotide substrates for the PolDom of *Mt*- and *Ms*-LigD). Again, as shown in Figure 8(c) (left panels) there was a preferred selection of the complementary base (C in this case) to be inserted in front of 8-oxoG (the C/A ratio of insertion in front of 8-oxoG was 55-fold for *Mt*-PolDom and 5-fold for *Ms*-PolDom), and the efficiency was comparable to the insertion of CTP in front of a non-damaged dG templating base. Moreover, the extension of a 8-oxoG:CMP pair occurred with the same efficiency than the extension of a dG-CMP pair, and the extension of a 8-oxoG:AMP was very inefficient (Figure 8(c), right panels). Therefore, the PolDom's of *Mt*- and *Ms*-LigD are able to carry out efficient polymerization at 8-oxoG containing templates. Both reactions, insertion in front of the damaged base, and extension of a damaged base pair, are preferentially non-mutagenic.

Discussion

There are two major non-related DNA primase families in prokaryotes, the DnaG-like proteins, which function as the bacterial replicative primases, and the AEPs that are closely related to replicative primases in archaea and eukaryotes. DNA primases have long been regarded as polymerases with a limited cellular role, namely the synthesis of short oligoribonucleotide primers at replication forks, which are then extended by the replicative polymerases. However, this restricted role for primases has recently been called into question with the discovery of a wide variety of bacterial and archaeal primases that possess a remarkable variety of nucleotidyl transferase activities including DNA-dependent RNA primase, terminal transferase and DNA-dependent DNA/RNA gap-filling polymerase activities.[18,19,27] DNA primases may therefore have additional roles in DNA metabolism, although the role of novel archaeal primase activities remains to be established.[27] In contrast, the discovery of novel primase domains in prokaryotic DNA repair ligases[3] and the subsequent demonstration that these domains are capable of catalyzing a wide variety of nucleotidyltransferase activities during NHEJ processes,[15,20,21] establishes that AEPs are highly adaptable polymerases that can be utilized for a variety of cellular functions, including DNA repair, in addition to ascribed roles in priming of replication.

Here, we have presented the crystal structure of the primase/polymerase domain of ligase D from *M. tuberculosis*. In agreement with previous *in silico* studies, the fold of *Mt*-PolDom shares a structural core with members of the AEP family, that also includes the classical eukaryotic replicative primases. AEPs contain a highly derived version of the RNA recognition motif (RMM) fold that is found in the catalytic palm module of other DNA and

RNA polymerases.[26] It appears that AEPs were originally recruited for primer synthesis at the base of the archaeo-eukaryotic lineage and were subsequently acquired by prokaryotes via horizontal gene transfer.[26] It has been suggested that the AEPs may be related to the PolX superfamily[38], which includes Pol β, μ and λ. However, a comparison of the structures of AEPs (including PolDom) and PolX (Pol β, Pol λ and TdT) reveals that the PolX members possess a fold unrelated to the RRM fold, although both families possess a triad of aspartate metal-chelating residues that are structurally super-imposable, which is unexpected given the considerable degree of difference in the secondary structures of the protein folds. The conserved geometry of important active site resides suggests that members of the AEP and PolX families may have evolved by convergent evolution to have similar catalytic mechanisms as is discussed below. The structure of the PolDom of *Pseudomonas* LigD possesses a very similar AEP-like architecture.[55]

Unexpectedly, in the co-crystal structures with bound nucleotide (ribo or deoxyribo), the base of GTP and dGTP adopted distinct orientations in the active site of PolDom suggesting a preferred binding mode within the enzyme for these nucleotides. In-line with this observation, we observed a 20-70 fold difference, depending of the templating base, in PolDom's preference for incorporation of rNTPs versus dNTPs. Considering that the *in vivo* concentration of ribonucleotides is much higher (at least 10-fold higher) than that of dNTPs, the probability of using NTP for a templated repair reaction by PolDom could be >700-fold. PolDom's preference for incorporating NTPs *in vitro* and its lack of 2' specificity reflect a catalytic plasticity that has also been maintained during the evolution of other unrelated NHEJ polymerases, such as Pol μ. [15,18,39] It has been postulated that both eukaryotic and prokaryotic NHEJ polymerases may incorporate RNA to repair DSBs when intracellular pools of dNTPs are depleted, as is the case during stationary phase, when break repair cannot be accomplished

using homologous recombination.[15,18,39] The majority of prokaryotic species that possess NHEJ-associated PolDom's can subsist for extended periods in stationary phase, suggesting that the repair of DSBs in these organisms by NHEJ would result in the incorporation of a significant number of short stretches of RNA into the genomic DNA. Although it remains to be resolved how RNA patches introduced during DSB repair processing are tolerated by the cell. One possibility is that the replicative polymerases may be capable of synthesizing through short stretches of hybrid RNA/DNA. Alternatively, incorporated RNA may be excised from DNA by short-patch repair processes, such as the RNase HII pathway.[40] The incorporation of RNA could potentially act as a signal to the cell, signalling the presence of excessive sites of DNA damage that must be processed, for example to remove modified bases, prior to the restart of replication. Whatever the exact nature of the cellular response to RNA incorporation, our findings, and those of others, strongly imply that there is a direct connection between RNA synthesis and the repair of DSBs by NHEJ, that is conserved even though the end-joining polymerases that perform this task are specific to each kingdom.

Our findings directly connect several paradigms among Pol X and AEP proteins that have been implicated in NHEJ-mediated DSB repair. In mammalian cells, three PolX members (Polμ, Polλ and TdT) with overlapping activities partake in the repair of DSBs by NHEJ.[33-35] In contrast, PolDom appears to be the sole polymerase involved in NHEJ processing in prokaryotes. As we have shown, PolDom possesses not only a remarkable variety of polymerase/primase activities but also has many of the essential features required to participate in unusual nucleotidyltranseferase activities required during end-processing. Remarkably, PolDom appears to have evolved into an enzyme that possesses many of the combined

activities present in Pol μ, Pol λ and TdT in a single active site. In common with Polμ, and TdT, PolDom exhibits a reduced dependence on template as evidenced by its terminal transferase activity. Either gapped-DNA molecules or 3′-protruding DNA, containing a 5′ phosphate, strongly interact with PolDom. PolDom's versatility as an end-joining enzyme extends from its capacity to combine the template dislocation properties and ability to realign mismatched ends of Pol μ,[29] with some capacity to perform direct extension of mismatched ends, an activity associated with Pol λ.[41] PolDom also exhibits the desirable facility to perform non-mutagenic translesion synthesis on termini containing modified bases, such as 8-oxoG. It therefore appears that, unlike replicative polymerases, it is beneficial if the active site of end-joining polymerases can accommodate a variety of unstable DNA alignments that are encountered during NHEJ.

Considering the biochemical similarities between PolDom, Pol μ, Pol β and Pol λ, including the shared requirement for a 5' phosphate, it will be intriguing to discover how DNA recognition is accomplished by PolDom, which lacks the 8 kDa domain responsible for phosphate binding by the eukaryote polymerases. Pol μ and TdT both possess template-independent polymerase activities, that have been attributed, at least in part, to the presence of a surface loop (loop 1).[29,34] Pol λ contains an attenuated loop 1, which does not confer template-independent activity when presented with homopolymeric DNA substrates, although it does allow it to act upon DNA ends with limited microhomology. In contrast, Pol β does not possess this loop and is unable to act on DSBs with limited or no homology. PolDom clearly possesses terminal transferase activity and it will be of great interesting to discover if convergent evolution has provided it with an equivalent to the loop 1 of the Pol X polymerase family members.

Acknowledgements

AJD is a Royal Society University Research Fellow and work in the AJD laboratory is supported by grants from BBSRC, Cancer Research UK, Association for International Cancer Research, and the Royal Society. Work in the LB laboratory was supported by Ministerio de Ciencia y Tecnología Grant BMC 2003-00186, and by an institutional grant to Centro de Biología Molecular "Severo Ochoa" from Fundación Ramón Areces. AJP is a recipient of a fellowship from the Spanish Ministry of Science and Technology. We acknowledge the European Synchrotron Radiation Facility for provision of synchrotron radiation facilities and would like to thank the staff of ID14 for assistance in using the beamline.

Materials and Methods

Crystallization and data-collection of PolDom

Mt-PolDom and *Ms*-PolDom were expressed and purified as previously described.[20] Native *Mt*-PolDom and *Mt*-PolDom (F64L) crystals were grown by the hanging drop, vapour diffusion method at 12°C by mixing equal volumes of protein solution (433 μM) and reservoir buffer: 20% (w/v) PEG 3350, 0.2 M Mg-acetate. Crystals formed overnight and typically grew to a size of 160 μm x 160 μm x 270 μm in 6-9 days. The *Mt*-PolDom (F64L) selenomethionine (SeMet) derivative protein was expressed in *E. coli* B834(DE3). SeMet PolDom (F64L) was purified in an identical manner to the native protein, while crystals of SeMet PolDom were obtained by microseeding from native crystals. Co-crystals were grown in 0.1 M Na Hepes pH 7.5, 10 mM $MnCl_2$, 10% (v/v) iso-Propanol, 20% (w/v) PEG 4000, plus 10 mM (d)GTP by microseeding from native crystals. Multiple wavelength anomalous diffraction (MAD) data of the apo- *Mt*-PolDom were collected at beam-line ID14-1 of the ESRF (Grenoble, France). The crystals belong to space group $P2_1$ with unit cell dimensions: a = 48.8Å, b = 76.0Å, c = 95.8Å. Data sets for the native crystals and co-crystals were collected in-house on a Raxis IV++ with a rotating anode X-ray generator RU-H3R. The crystals of apo- *Mt*-PolDom belong to space group $P2_1$ with unit cell dimensions: a = 41.0Å, b = 75.8Å, c = 96.1Å. The co-crystals of *Mt*-PolDom with GTP belong to spacegroup $P2_12_12_1$ with unit cell dimensions: a = 42.6Å, b = 75.5Å, c = 89.5Å. The co-crystals of *Mt*-PolDom with dGTP belong to space group $P2_1$ with unit cell dimensions: a = 41.2Å, b = 75.9Å, c = 95.8Å. All crystals were harvested and cryoprotected with Paratone-N before freezing in liquid nitrogen, while all data sets were collected at 100K. The diffraction data were

processed with IPMOSFLM[42] and were merged, scaled and reduced with programs from the CCP4 suite.[43] The statistics for data processing are summarized in Table I.

Structure solution and refinement of *Mt*-PolDom

The structure of SeMet (apo) PolDom was determined by direct methods using the autoSHARP program.[44] Heavy atom position refinement and initial phasing was carried out using SHARP.[45] The crystallographic model derived from the MAD data was used as a molecular replacement search model against higher resolution native data within the program PHASER.[46] Initial refinement was carried out against 95% of the data with REFMAC5.[47] The remaining 5%, which were randomly excluded from the full data set, was used for cross-validation by calculating the R_{free} to follow the progress of the refinement. The same subset of reflections was used throughout the refinement. A final refined model at 1.65Å resolution, with an R_{cryst} of 19.0% and R_{free} of 23.9%, was used for isotropic refinement using TLS parameters in REFMAC5.[48] Each cycle of refinement was accompanied by manual rebuilding using the programs O[49] and COOT.[50] For the determination of the GTP/dGTP structures, the models were built/refined using COOT/REFMAC5 after an initial molecular replacement search using the apo structure in PHASER. The structure images were prepared with CCP4mg.[51,52]

DNA Substrates. Synthetic oligonucleotides, purified by PAGE, were obtained from Invitrogen or Eurogentec. Homopolymeric poly-dA (15-mer) or blunt-ended DNA (generated by aligning oligonucleotide P1 (5´ GATCACAGTGAGTACAATA 3´) to its complement) were used as substrates for the terminal transferase activity assay shown in

Fig.3. Template-primer (T/P) and 1-nucleotide gapped (Gap-1OH; Gap-1P) molecules used in the electrophoretic mobility shift assay (EMSA) and primer extension reactions shown in figure 4, were generated by annealing either the primer oligonucleotide P2 (5´ TCTGTGCAGGTTCTT 3´) alone, or together with the downstream oligonucleotide D2 (5´ GTCGAGAGGGACTTC 3´), the latter having either a 5´ OH group or a 5´ phosphate group, to the template oligonucleotide T2 (5´-TGAAGTCCCTCTCGACGAAGAACCTGCACAGA-3´). The 1-nucleotide gapped molecule with a 5´ phosphate group shown in figure 6 was obtained by annealing the P1 primer and downstream oligonucleotide D1 (5´-AGATACACTTCT-3´, having a 5´ phosphate group) to the template oligonucleotide T1: (5´-AGAAGTGTATCTCGTACTCACTGTGATC-3´). Template/downstream (T/D) molecule used in figure 6 were obtained by annealing different template oligonucleotides T1-X (5´-AGAAGTGTATCTXGTACTCACTGTGATC-3', where X is A, C, G or T) to D1 oligonucleotide. Four different 1-nucleotide gapped molecules with a 5´ phosphate group used in figure 5 and Supplementary Fig. 3 were obtained by annealing the four T1-X template oligonucleotides to P1 and D1 oligonucleotides. The 2-nucleotide gapped molecule used in figure 7a was obtained by annealing oligonucleotides P2 and D2* (5´-GTCGAGAGGGACTTC-3´, having a 5´-phosphate), to oligonucleotide T2. The mismatched T/P molecule used in figure 7b was obtained by annealing the primer P3 (5'-GATCACAGTGAGTAC-3´) to the template oligonucleotide T3 (5´-TCTACGTCTACTCACTGTGATC-3´). The 1nt-gapped DNA selected to analyze insertion in front of 8-oxoG (Fig. 8) was obtained by annealing the primer P4 (5´-CTGCAGCTGATGCGC-3´) and the downstream oligonucleotide D4 (5´-GTACGGATCCCCGGGTAC-3´, having a 5´phosphate) to either the template

oligonucleotide T4 (5´ GTACCCGGGGATCCGTACGGCGCATCAGCTGCAG-3´) or T4-oxoG (5´-GTACCCGGGGATCCGTACG8-oxoGCGCATCAGCTGCAG-3´). To analyze extension of a 8-oxoG/dCMP versus 8-oxoG/dAMP, either oligonucleotide P4-C (5´-CTGCAGCTGATGCGCC-3´) or P4-A (5´-CTGCAGCTGATGCGCA-3´) were used as primers hybridized to the template T4-oxoG. As a control, primer P4-C was annealed to template T4. Selected oligonucleotides were labelled at the 5´-end with [γ-^{32}P] ATP (3000 Ci/mmol, Amersham) and T4 polynucleotide kinase (New England Biolabs). These labelled oligonucleotides were then hybridized to one or two oligonucleotides to generate different DNA substrates in the presence of 50 mM Tris-HCl, pH 7.5, and 0.3 M NaCl, and heating to 80 ºC for 10 min before slowly cooling to room temperature overnight.

Preparation of PolX enzymes. Purified human Pol μ and Pol λ were obtained as originally described.[53,54]

Electrophoretic mobility shift assay (EMSA). EMSAs were carried out using different DNA substrates, obtained as indicated in the previous section, to analyze the interaction of either *Ms*- or *Mt*-PolDom and DNA. Incubation was performed in a final volume of 12.5 μl, containing 50 mM Tris–HCl, pH 7.5, 0.1 mg/ml BSA, 1 mM DTT, 4% glycerol, 5 nM labeled DNA and different concentrations of *Ms*- or *Mt*-PolDom. After incubation for 10 min at 30°C to allow the formation of enzyme–DNA complexes, samples were mixed with 3 μl of 30% glycerol and resolved by native gel electrophoresis on a 4% polyacrylamide gel (80:1 monomer/bis). After autoradiography, stable PolDom/DNA complexes are detected as slow migration bands compared to the mobility of free DNA.

Polymerization assays. The incubation mixture (20 μl) contained 50 mM Tris-HCl pH 7.5, 3 mM $MgCl_2$ or 1 mM $MnCl_2$, 1 mM DTT, 4 % glycerol, 0.1 mg/ml BSA, 5 nM DNA, the indicated concentration of *Mt- or Ms*-PolDom, and dNTPs or NTPs. After 30 min of incubation at 30ºC, reactions were stopped by adding loading buffer (10 mM EDTA, 95% formamide, 0.03% bromophenol blue, and 0.3% cyanol blue) and subjected to electrophoresis in 8 M urea-containing 20% polyacrilamide sequencing gels. After electrophoresis, the unextended and extended DNA primers were detected by autoradiography.

FIGURES LEGENDS

Figure 1. Structure of polymerase/primase domain of *Mt*-Ligase D

(a) The crystal structure of *Mt*-PolDom is depicted in ribbon form. α-helices are coloured light blue and β-strands coloured yellow with intervening loop regions coloured grey. The secondary structural elements are numbered according to their assignment in DDSP (Dictionary of protein secondary structure).

(**b**) Structural comparison of *Mt*-PolDom, with the primase small subunits of *Sso* (PDB code:1ZT2) and *Pfu* (PDB code:1G71). Structurally conserved regions are coloured red with structurally variant regions coloured green, yellow and blue for *Mt*-PolDom, *Sso* and *Pfu,* respectively.

Figure 2. Detailed view of the active site conformation of *Mt*-PolDom.

(**a**) Electrostatic surface representation of *Mt*-PolDom with GTP bound in the active site**.** A solvent accessible surface representation of *Mt*-PolDom with electrostatic potential mapped is shown for the structure with bound GTP and cations. Negatively charged regions are coloured red and positively charged regions are coloured blue. Manganese is coloured pink.

(b-d) Close-up views of the active-site pocket of *Mt*-PolDom, apoenzyme (b), GTP-bound (c) and, dGTP-bound (d) co-crystals. Potential hydrogen bond interactions are shown as dashed lines. Manganese ions and water molecules are represented as pink and red spheres respectively. Amino acids and nucleotides are labelled and shown as stick representations.

Figure 3. Terminal transferase activity of Mt-PolDom. Homopolymeric poly-dA (**a**) or blunt-ended DNA (**b**) were used as substrate to analyze the terminal transferase activity of the *Mt*-PolDom. Polymerization assays were carried out as described (see Materials and Methods), in the presence of 380 nM of either human Polμ or the *Mt*-PolDom, and activated by either magnesium ions (3 mM $MgCl_2$; **a**, left and right; **b**, left) or by manganese ions (1 mM $MnCl_2$; **b**, right). After 30 min of incubation at 30 ºC in the presence of each indivcated dNTP or NTP (100 μM), reactions were stopped by adding loading buffer. Terminal transferase activity is detected by the appearance of extension products of the labelled primer used in each case.

Figure 4. PolDom prefers DNA gapped substrates containing a 5´-P. (a) DNA-binding capacity of *Mt*-PolDom. Electrophoretic mobility shift assays (EMSA) were performed on the different DNA substrates (T/P; Gap 1-OH; Gap 1-P) shown in the scheme (asterisk indicates the 5´-labelled oligonucleotide), as described in Materials and Methods, in the presence of the indicated amounts of *Mt*-PolDom. Mobility of free DNA and enzyme/DNA complexes was detected by autoradiography. **(b)** Gap-filling synthesis and substrate preferences of *Mt*-PolDom. Primer extension reactions on the indicated DNA substrates (T/P; Gap 1-OH; Gap 1-P) was carried out as indicated in the Material and Methods, using 380 nM *Mt*-PolDom and the indicated concentrations of metal (Mg^{2+} or Mn^{2+}) and the complementary nucleotide (dCTP or CTP). After incubation for 30 min at 30°C, primer extension products were analyzed by 8M urea-PAGE and autoradiography.

Figure 5. Nucleotide insertion fidelity by *Mt*-PolDom at 1nt-gapped DNA. (a) Four 1-nt gapped DNA substrates, differing in the templating base (X), and having a phosphate at the 5´-side of the gap (GAP1P), were used. The primer (P) and downstream (D) oligonucleotides were annealed to four oligonucleotide templates (T), where X is A, C, G or T. Labeling of the oligonucleotide primer is indicated by an asterisk. (b) Gap-filling DNA synthesis reactions were carried out as described (Materials and Methods), using the four variant gapped substrates shown in a, in the presence of 380 nM *Mt*-PolDom and 3 mM Mg_2Cl. Extension of the labeled primer in the presence of either the correct or the incorrect dNTP or rNTP (as indicated in the figures) was analyzed by 8 M urea-20% PAGE and autoradiography.

Figure 6. PolDom efficiently binds 3´-protruding ends. DNA binding capacity of *Ms*- and *Mt*-PolDom for NHEJ intermediates. The two types of DNA molecules used (1-nucleotide gapped and template/downstream molecules with a 5´-phosphate) are shown in the scheme. Labelled oligonucleotides are indicated with an asterisk. EMSAs were performed as described (Materials and Methods) using the indicated amounts of each PolDom (Pol) corresponding to either *Mt*-LigD or *Ms*-LigD. Mobility of free DNA and enzyme/DNA complexes are indicated on the autoradiograph.

Figure 7. Characterization of PolDom template dislocation and primer realignment capacities. Schemes representing template sequence contexts that are appropriate to evaluate dNTP selection-mediated dislocation (**a**), and mismatched-primer realignment (**b**) are shown (see text for details). Labelled primer (asterisk) and 5'-end phosphate group (P) are indicated. Transiently misaligned or mispaired nucleotides are indicated in

bold letters. Polymerization assays were carried out as described (Materials and Methods) using 380 nM of either human Pol μ, or *Mt*-PolDom (Mt), or *Ms*-PolDom (Ms), in the presence of each individual or combined dNTP or NTP (1 μM; part A), or in the presence of the indicated concentration of each dNTP (part B). After incubation for 30 min at 30°C, extension of the labelled primer was analyzed by 8M urea-20% PAGE and autoradiography.

Figure 8. Polymerization at 8-oxoG containing templates is preferentially non-mutagenic. (**a**) Schemes representing the template sequence contexts used to evaluate insertion in front of a 8-oxoG versus a normal dG base (1nt-gap containing a 5´P), and extension of either a 8-oxoG/dC or 8-oxoG/dA primer terminus versus extension of a control dG/dC pair, are shown at the left and right, respectively. Labelled primer (asterisk) and 5'-end phosphate group (P) are indicated. Polymerization assays in the presence of the indicated deoxy (**b**) or ribonucleotides (**c**), were carried out as described (Materials and Methods) using different nucleotide concentrations as a function of the PolDom (*Mt* or *Ms*, provided at 400 nM) and the DNA substrate used. To analyze insertion (**b**,**c**; left) dNTPs were provided either at 100 μM (*Mt*) or at 1 μM (*Ms*), and NTPs were provided either at 0.5 μM (*Mt*) or at 0.1 μM (*Ms*). To analyze extension (**b,c**; right) dNTPs were provided either at 1 mM (*Mt*) or at 10 μM (*Ms*), and NTPs were provided either at 20 μM (*Mt*) or at 0.5 μM (*Ms*). After incubation for 30 min at 30°C, in the presence of 2.5 mM $MgCl_2$, extension of the labelled primer was analyzed by 8 M urea-20% PAGE and autoradiography.

References

1. Krejci, L., Chen, L., Van Komen, S., Sung, P. & Tomkinson, A. E. (2003). Mending the break: two repair machines in eukaryotes. *Prog. Nucl. Acid Res. Mol. Biol.* **74**, 159-201.

2. Critchlow, S. E. & Jackson, S. P. (1998). DNA end-joining: from yeast to man. *Trends Biochem. Sci.* **23**, 394-398.

3. Weller, G. R., Kysela, B., Roy, R., Tonkin, L. M., Scanlan, E., Della, M. *et al.* (2002). Identification of a DNA non homologous end-joining complex in bacteria. *Science* **297**, 1686-1689.

4. DeFazio, L. G., Stansel, R. M., Griffith, J. D. & Chu, G. (2002). Synapsis of DNA ends by DNA-dependent protein kinase. *EMBO J.* **21**, 3192-3200.

5. Ramsden, D. & Gellert, M. (1998). Ku protein stimulates end joining by mammalian DNA ligases: a direct role for Ku in repair of DNA double-strand breaks. *EMBO J.* **17**, 609-614.

6. Chen, L., Trujillo, K., Ramos, W., Sung, P. & Tomkinson, A. E. (2001). Promotion of Dnl4-catalyzed DNA end-joining by the Rad50/Mre11/Xrs2 and Hdf1/Hdf2 complexes. *Mol. Cell* **8**, 1105-1115.

7. Chen, L., Trujillo, K., Sung, P. & Tomkinson, A. E. (2000). Interactions of the DNA ligase IV-XRCC4 complex with DNA ends and the DNA-dependent Protein Kinase. *J. Biol. Chem*. **275**, 26196-26205.

8. Nick McElhinny, S. A., Snowden, C. M., McCarville, J. & Ramsden, D. (2000). Ku recruits the XRCC4-ligase IV complex to DNA ends. *Mol. Cell. Biol*. **20**, 2996-3003.

9. Mahajan, K. N., Nick McElhinny S.A., Mitchell, B. S. & Ramsden, D. A. (2002). Association of DNA polymerase mu (pol μ) with Ku and ligase IV: Role for pol μ in End-Joining Double-Strand Break Repair. *Mol. Cell. Biol*. **22**, 5194-5202.

10. Tseng, H.-M., & Tomkinson, A. E. (2002). A physical and functional interaction between Yeast Pol4 and Dnl4-Lif1 links DNA synthesis and ligation in nonhomologous end joining. *J. Biol. Chem.* **277**, 45630-45637.

11. Wilson, T. E., & Lieber, M. R. (1999). Efficient processing of DNA ends during yeast nonhomologous end joining. *J. Biol. Chem.* ***274***, 23599-23609.

12. Ma, Y., Pannicke, U., Schwarz, K. & Lieber, M. R. (2002). Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* **108**, 781-794.

13. Paull, T. T. & Gellert, M. (2000). A mechanistic basis for Mre11-directed DNA joining at microhomologies. *Proc. Natl. Acad. Sci. U.S.A*. **97**, 6409-6414.

14. Wu, X., Wilson, T. E. & Lieber, M. R. (1999). A role for FEN-1 in nonhomologous DNA end joining: the order of strand annealing and nucleolytic processing events. *Proc. Natl. Acad. Sci. U.S.A*., **96**, 1303-1308.

15. Della, M., Palmbos, P. L., Tseng, H. M., Tonkin, L. M., Daley, J. M., Topper, L. M., *et al*., (2004). Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. *Science* **306**, 683-685.

16. Gong, C., Martins, A., Bongiorno, P., Glickman, M. & Shuman, S. (2004). Biochemical and genetic analysis of the four DNA ligases of mycobacteria. *J. Biol. Chem.* **279**, 20594-20606.

17. Gong, C., Bongiorno, P., Martins, A., Stephanou, N. C., Zhu, H., Shuman, S. & Glickman, M. S. (2005). Mechanism of nonhomologous end-joining in mycobacteria: A low-fidelity repair system driven by Ku, ligase D and ligase C. *Nat. Struct. Mol. Biol.* **12**, 304-312.

18. Pitcher, R. S., Wilson, T. E. & Doherty, A. J. (2005). New Insights into NHEJ Repair Processes in Prokaryotes. *Cell Cycle* **4**, 675-678.

19. Bowater, R. & Doherty, A.J. (2006). Making Ends Meet: Repairing Breaks in Bacterial DNA by Non-Homologous End-Joining. *PLOS Genetics* **2**, 0093.

20. Pitcher, R. S., Tonkin, L. M., Green, A. J. & Doherty, A. J. (2005). Domain Structure of a NHEJ DNA Repair Ligase from Mycobacterium tuberculosis. *J. Mol. Biol.* **351**, 531-544.

21. Zhu, H. & Shuman, S. (2005). A primer-dependent polymerase function of pseudomonas aeruginosa ATP-dependent DNA ligase (LigD). *J. Biol. Chem.* **280**, 418-427.

22. Lao-Sirieix, S. H., Nookala, R. K., Roversi, P., Bell, S. D. & Pellegrini, L. (2005). Structure of the heterodimeric core primase. *Nat. Struct. Mol. Biol*. **12,** 1137-1144.

23. Augustin, M. A., Huber, R. & Kaiser, J. T. (2001). Crystal structure of a DNA-dependent RNA polymerase (DNA primase). *Nat. Struct. Biol.,* **1**, 57-61.

24. Ito, N., Nureki, O., Shirouzu, M., Yokoyama, S. & Hanaoka, F. (2003). Crystal structure of the Pyrococcus horikoshii DNA primase-UTP complex: implications for the mechanism of primer synthesis. *Genes Cells*. **8**, 913-923.

25. Lipps, G., Weinzierl, A. O., von Scheven, G., Buchen, C. & Cramer, P. (2004). Structure of a bifunctional DNA primase-polymerase. *Nat. Struct. Mol. Biol*. **11**, 157-162.

26. Iyer, L. M., Koonin, E. V., Leipe, D. D. & Aravind L. (2005). Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. *Nucleic Acids Res*. **33**, 3875-3896.

27. Lao-Sirieix, S. H, Pellegrini, L. & Bell, S. D. (2005). The promiscuous primase. *Trends Genet.* **21**, 568-572.

28. Steitz, T.A. (1999). DNA polymerases: structural diversity and common mechanisms. *J. Biol. Chem.* **274**, 17395-17398.

29. Juárez, R., Ruiz, J. F., Nick McElhinny, S., Ramsden, D. A. & Blanco, L. (2006). A specific loop in human DNA polymerase mu allows switching between creative and DNA-instructed synthesis. *Nucleic Acids Res., in press.*

30. Prasad, R., Beard, W. A. & Wilson S. H. (1994) Studies of gapped DNA substrate binding by mammalian DNA polymerase beta. Dependence on 5'-phosphate group. *J. Biol. Chem*. **269**, 18096-18101.

31. Matsumoto, Y. & Kim, K. (1995) Excision of deoxyribose phosphate residues by DNA polymerase beta during DNA repair. *Science* **269**, 699-702.

32. Singhal, R. K. & Wilson, S. H. (1993). Short gap-filling synthesis by DNA polymerase beta is processive. *J. Biol. Chem.* **268**, 15906-15911.

33. Ma, Y., Lu, H., Tippin, B., Goodman, M. F., Shimazaki, N., Koiwai, O., Hsieh, C. L., Schwarz, K. & Lieber, M. R. (2004). A biochemically defined system for mammalian nonhomologous DNA end joining. *Mol. Cell* **16**, 701-713.

34. Nick McElhinny, S. A., Havener, J. M., Garcia-Diaz, M., Juarez, R., Bebenek, K., Kee, B. L., Blanco, L., Kunkel, T. A. and Ramsden D. A. (2005). A gradient of template dependence defines distinct biological roles for family X polymerases in nonhomologous end joining. *Mol. Cell* **19**, 1-10.

35. Daley, J. M., Laan, R. L., Suresh, A. & Wilson, T. E. (2005). DNA joint dependence of pol X family polymerase action in non homologous end-joining. *J. Biol. Chem.* **280**, 29030-2937.

36. Zhang, Y., Wu, X., Yuan, F., Xie, Z. & Wang, Z. (2001). Highly frequent frameshift DNA synthesis by human DNA polymerase mu. *Mol, Cell Biol.* **21**, 7995-8006.

37. Ruiz, J. F., Lucas, D., Garcia-Palomero, E., Saez, A. I., Gonzalez, M. A., Piris, M. A., Bernad, A. & Blanco, L. (2004). Over-expression of human DNA polymerase mu (Pol mu) in a Burkitt's lymphoma cell line affects the somatic hypermutation rate. *Nucleic Acids Res.,* **32**, 5861-5873.

38. Arezi, B. & Kuchta, R. (2000). Eukaryotic DNA primase. *Trends Biochem. Sci*. **11**, 572-576.

39. Nick McElhinney, S. A. & Ramsden, D. A. (2003). Polymerase Mu is a DNA-directed DNA/RNA polymerase. *Mol. Cell. Biol.,* **23**, 2309-2315.

40. Rydberg, B. & Game, J. (2002). Excision of misincorporated ribonucleotides in DNA by RNase H (type 2) and FEN-1 in cell-free extracts. *Proc. Natl. Acad. Sci. USA,* **99**, 16654-16659.

41. Picher, A.J., Garcia-Diaz, M., Bebenek, K., Pedersen, L.C., Kunkel, T.A. & Blanco, L. (2006). Promiscuous mismatch extension by human DNA polymerase lambda. *Nucleic Acids Res.* **34**, 3259-3266.

42. Powell, H. R. (1999). The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallogr. D* **55**,1690-1695.

43. CCP4 (1994). Collaborative Computational Project No. 4. The CCP4 suite: Programs for Protein Crystallography. *Acta Crystallogr. D* **50**, 760-763

44. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2005) Automated structure solution with autoSHARP. In *Crystallograpic Methods* (S. Doublie, ed.) Humana Press, NJ, *submitted*.

45. La Fortelle, E. de & Bricogne, G. (1997). *Methods in Enzymology, Macromolecular Crystallography* (R. M. Sweet, and C. W. Carter, Jr, eds.) Academic Press, New York **276**, 472-494.

46. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). Likelihood-enhanced fast translation functions. *Acta Crystallogr. D* **61**, 458-464.

47. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr. D* **53**, 240.

48. Winn, M., Isupov, M. & Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D* **57**, 122-133.

49. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110-119.

50. Emsley, P. & Cowtan, K. (2004). Coot: Model-Building Tools for Molecular Graphics, *Acta Crystallogr. D* **60**, 2126-2132.

51. Potterton, E., McNicholas, S., Krissinel, E., Cowtan, K. & Noble, M. (2002). The CCP4 molecular-graphics project. *Acta Crystallogr. D* **58**, 1955-1957.

52. Potterton, L., McNicholas, S., Krissinel, E., Gruber, J., Cowtan, K., Emsley, P., Murshudov, G. N., Cohen, S., Perrakis, A. & Noble, M. (2004). Developments in the CCP4 molecular-graphics project *Acta Crystallogr. D* **60**, 2288-2294.

53. Dominguez, O., Ruiz, J. F., Lain de Lera, T., Garcia-Diaz, M., Gonzalez, M.A., Kirchhoff, T., Martinez-A, C., Bernad, A. & Blanco, L. (2000). DNA polymerase mu (Pol mu), homologous to TdT, could act as a DNA mutator in eukaryotic cells. *EMBO J.* **19**, 1731-1742.

54. Garcia-Diaz, M., Bebenek, K., Sabariegos, R., Dominguez, O., Rodriguez, J., Kirchhoff, T., Garcia-Palomero, E., Picher, A.J., Juarez, R., Ruiz, J.F., Kunkel, T.A. & Blanco, L. (2002). DNA polymerase lambda, a novel DNA repair enzyme in human cells. *J. Biol. Chem*. **277**, 13184-13191.

55. Zhu, H., Nandakumar, J., Aniukwu, J., Wang, L. K., Glickman, M. S., Lima, C. D. & Shuman, S. (2006) Atomic structure and nonhomologous end-joining function of the

polymerase component of bacterial DNA ligase D. *Proc Natl Acad Sci U S A.,* **103**, 1711-1716.

Table 1. X-ray diffraction data and refinement statistics

Crystal		SeMet		Native	GTP	dGTP
Data collection						
Source		ESRF ID14-1			Raxis IV++	
Space group		$P2_1$		$P2_1$	$P2_12_12_1$	$P2_1$
Unit Cell Dimensions (Å)						
a /*b*/*c*		48.8 / 76.0 / 95.8		41.0 / 75.8 / 96.1	42.6 / 75.5 / 89.5	41.2 / 75.9 / 95.8
a/b/g		90.0 / 92.5 / 90.0		90.0 / 92.7 / 90.0	90.0 / 90.0 / 90.0	90.0 / 92.6 / 90.0
Wavelength (Å)	0.97966 peak	0.97982 inflection	0.90790 remote		1.54	
Resolution (Å)	35.92-1.80	35.92-1.80	35.92-1.80	20.37-1.65	34.79-1.80	40.49-1.78
Total Number of observations[a]	195708 (27655)	194939 (27630)	194460 (28252)	406247 (27876)	242961 (18410)	207094 (29287)
Number of unique reflections[a]	54008 (7707)	54090 (7720)	54189 (7852)	63571 (7663)	27079 (3490)	56549 (8240)
Average I/(I)[a]	20.4 (5.6)	17.9 (3.8)	20.5 (5.6)	23.7 (3.0)	26.5 (2.9)	15.5 (4.3)
Overall completeness (%)[a]	99.6 (98.1)	99.7 (98.3)	99.9 (100.0)	90.1 (74.8)	98.4 (89.1)	100.0 (99.9)
R_{merge} (%)[a,b]	4.3 (18.9)	5.1 (26.9)	4.3 (19.5)	7.0 (23.5)	5.9 (40.8)	7.0 (17.5)
Multiplicity[a]	3.6 (3.6)	3.6 (3.6)	3.6 (3.6)	6.4 (3.6)	9.0 (5.3)	3.7 (3.6)
$\boldsymbol{FOM_{Centric/Acentric}}$		***0.5144 / 0.6376***				
Refinement						
Resolution (Å)				20.37-1.65	34.79-1.80	40.49-1.78
No. of reflections				60313	25666	53660
R_{fac} /R_{free}[c,d]				0.190 / 0.239	0.164 / 0.209	0.182 / 0.230
Contents of asymmetric unit				2 mols.	1 mol.	2 mols.
No. atoms						
Protein				4344	2164	4371
Manganese				2	1	2
Nucleotide molecules				-	1 (GTP)	2 (dGTP)
Water molecules				665	327	659
Mean B value (Å^2)				22.2	23.9	23.2

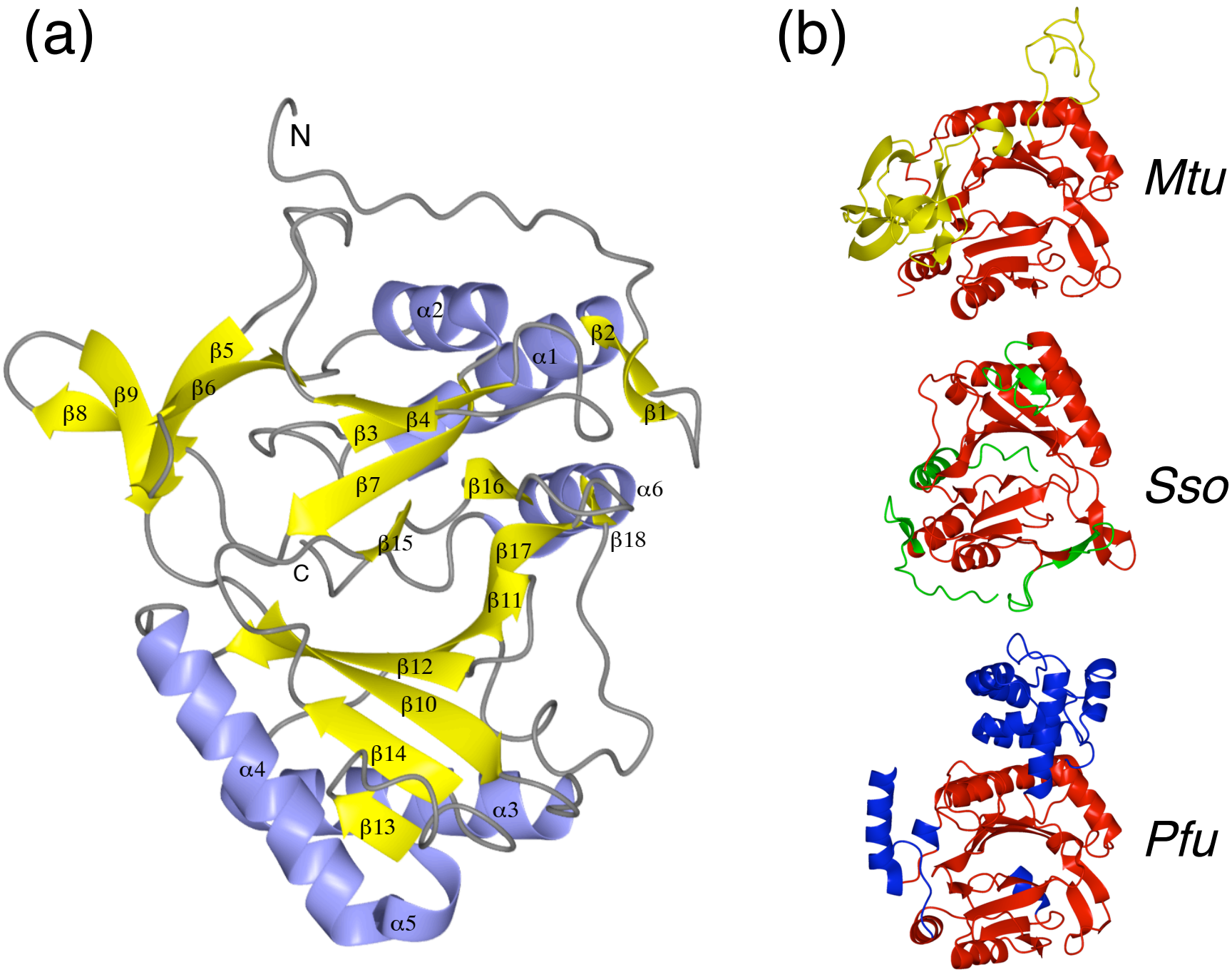
R.m.s. deviations			
Bonds (Å)	0.015	0.014	0.015
Angles (degrees)	1.478	1.429	1.414
Ramachandran Statistics			
Most favored regions (%)	93.6	95.0	93.8
Additionally allowed regions (%)	6.2	5.0	6.0
Generously allowed regions (%)	0.2	0.0	0.2
Disallowed regions (%)	0.0	0.0	0.0
PDB accession codes	xxxx	xxxx	xxxx

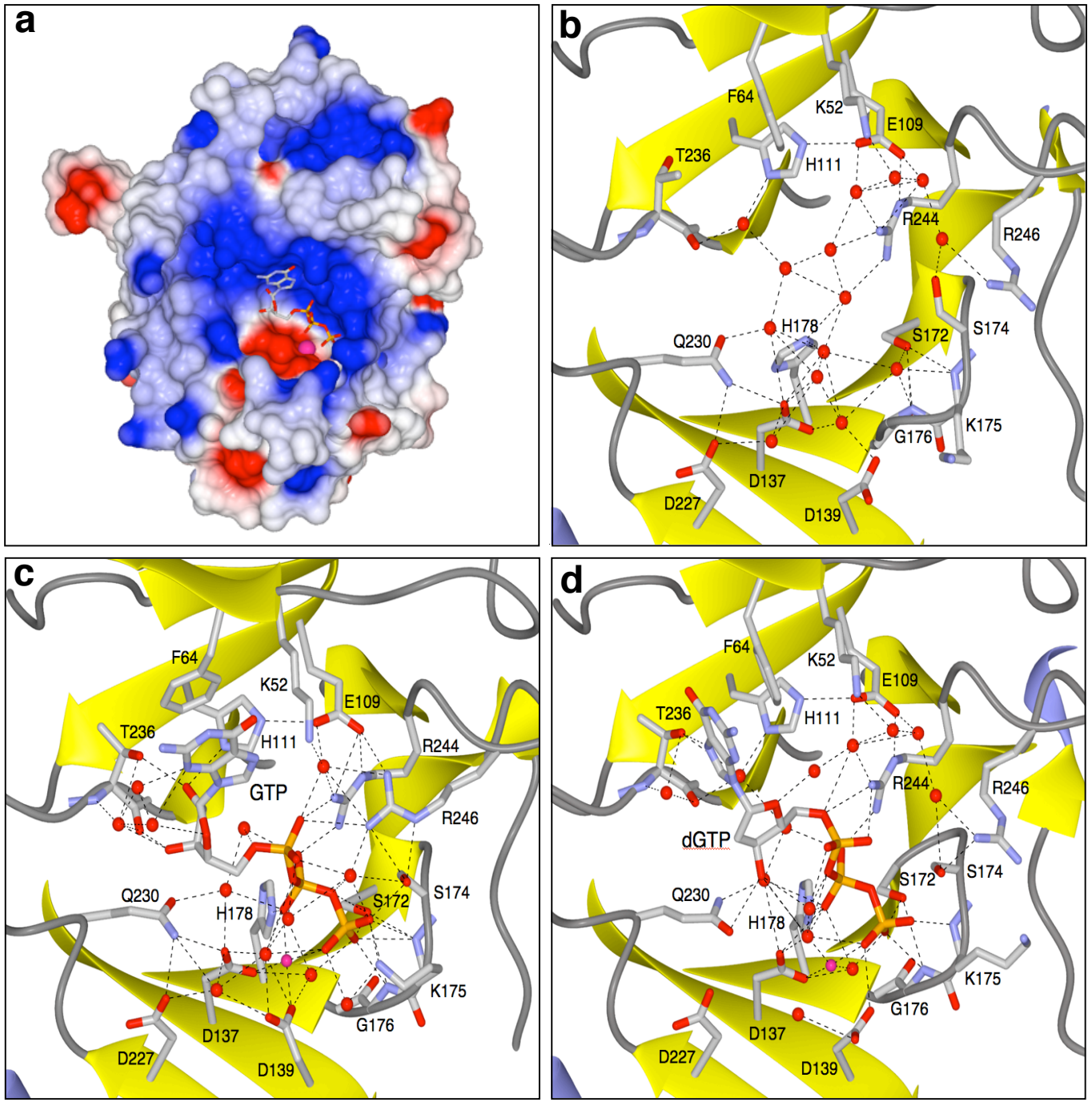
[a]Outermost shell in brackets.

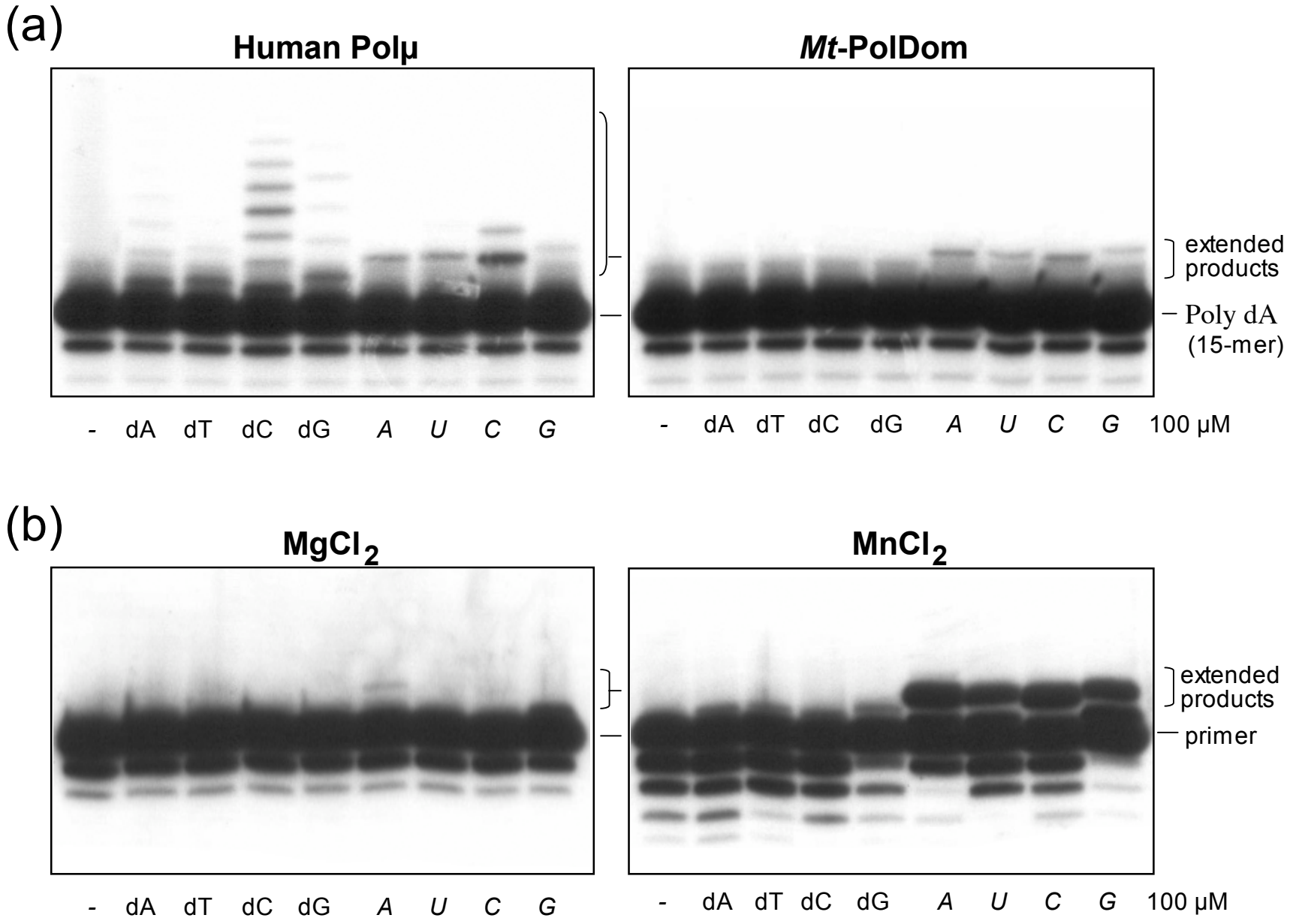
[b] $R_{merge} = \Sigma_{hkl}\Sigma_i|I_i - \langle I\rangle| / \Sigma_{hkl}\Sigma_i\langle I\rangle$, where I_i is the intensity of the i^{th} measurement of a reflection with indices hkl and $\langle I\rangle$ is the weighted mean of the reflection intensity.

[c] $R_{fac} = \Sigma||F_o| - |F_c||/\Sigma|F_o|$, where F_o and F_c are the observed and calculated structure factor, respectively.

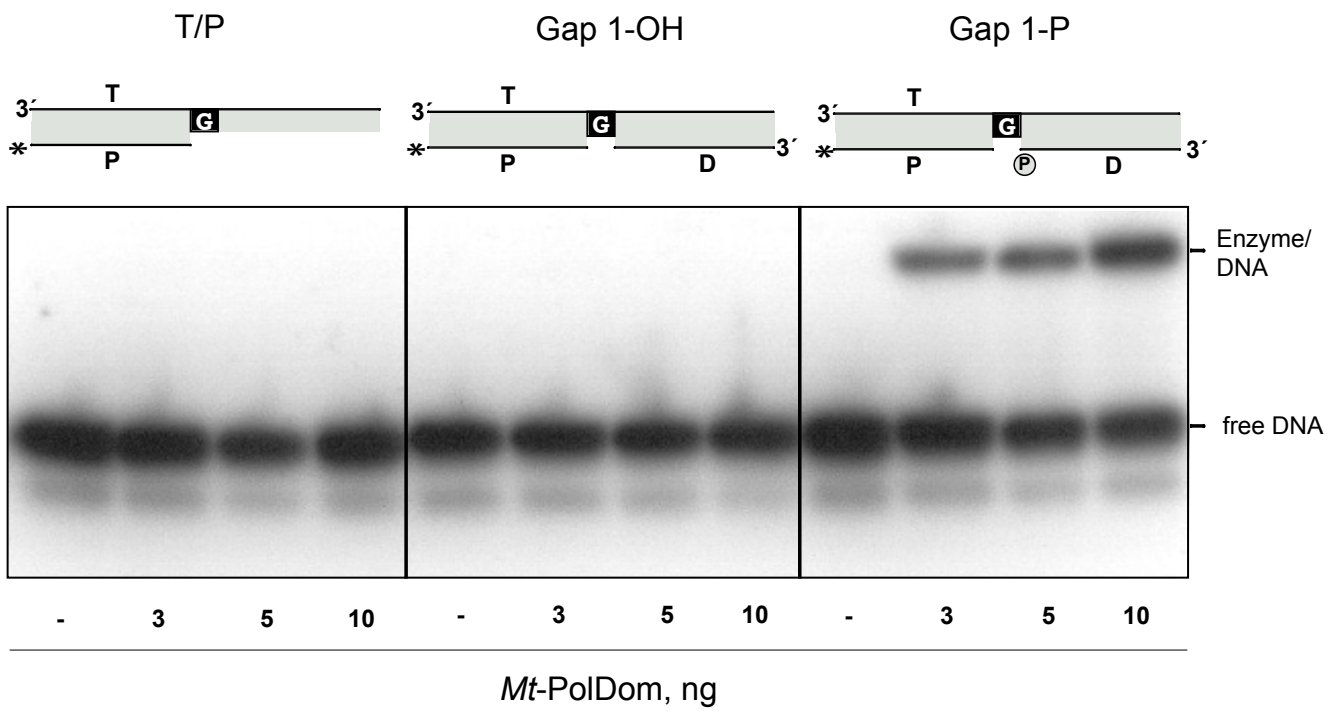
[d] R_{free} is equal to R_{fac} for a randomly selected 5% subset of reflections not used in the refinement.


(a)
N
α2
β2
α1
β5
β6
β9
β8
β4
β3
β1
β7
β16
α6
β15
β17
β18
C
β11
β12
β10
β14
α4
α3
β13
α5
(b)
Mtu
Sso
Pfu

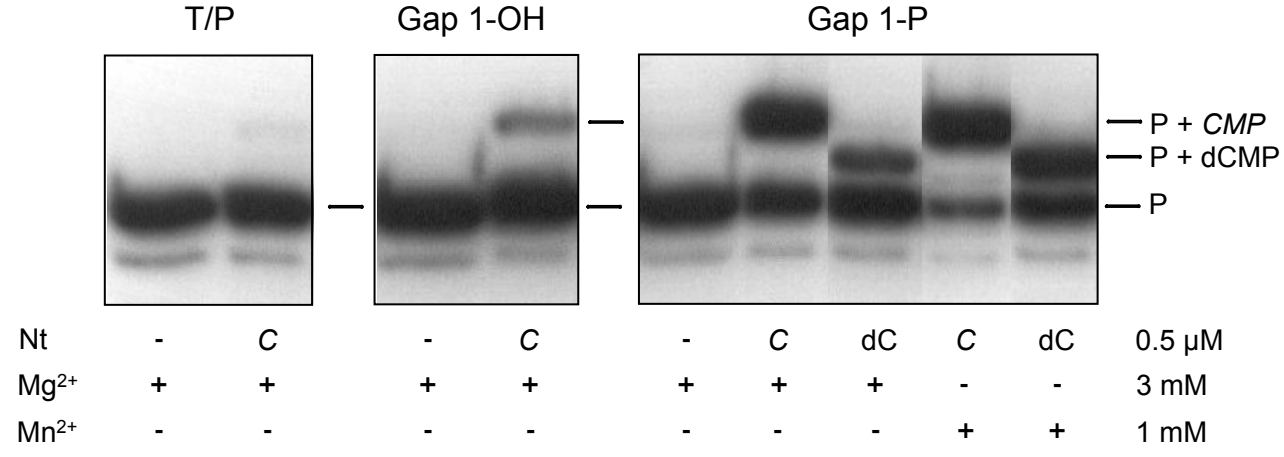

a
b
F64
K52
E109
T236
H111
R244
R246
H178
Q230
S172
S174
K175
G176
D137
D227
D139
c
F64
K52
E109
T236
H111
R244
GTP
R246
Q230
H178
S172
S174
K175
G176
D137
D227
D139
d
F64
K52
E109
T236
H111
R244
R246
dGTP
S172
S174
Q230
H178
K175
G176
D137
D227
D139


(a)
Human Polμ
Mt-PolDom
extended products
Poly dA (15-mer)
- dA dT dC dG A U C G
- dA dT dC dG A U C G 100 μM
(b)
MgCl2
MnCl2
extended products
primer
- dA dT dC dG A U C G
- dA dT dC dG A U C G 100 μM

a


T/P
Gap 1-OH
Gap 1-P
3´
T
G
P
D
Enzyme/
DNA
free DNA
- 3 5 10 - 3 5 10 - 3 5 10
Mt-PolDom, ng

b

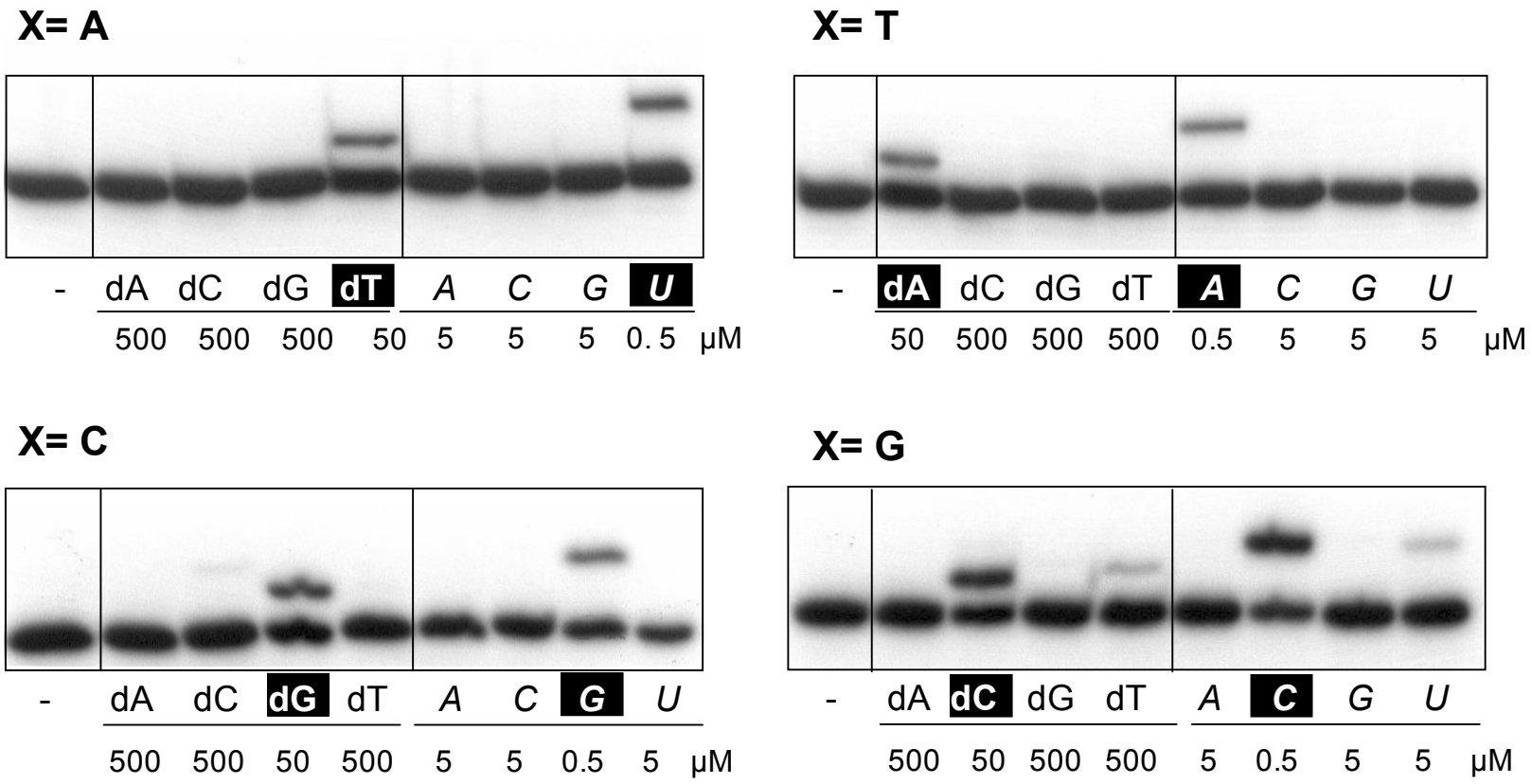

T/P
Gap 1-OH
Gap 1-P
P + *CMP*
P + dCMP
P
Nt - *C* - *C* - *C* dC *C* dC 0.5 µM
Mg^{2+} + + + + + + + - - 3 mM
Mn^{2+} - - - - - - - + + 1 mM

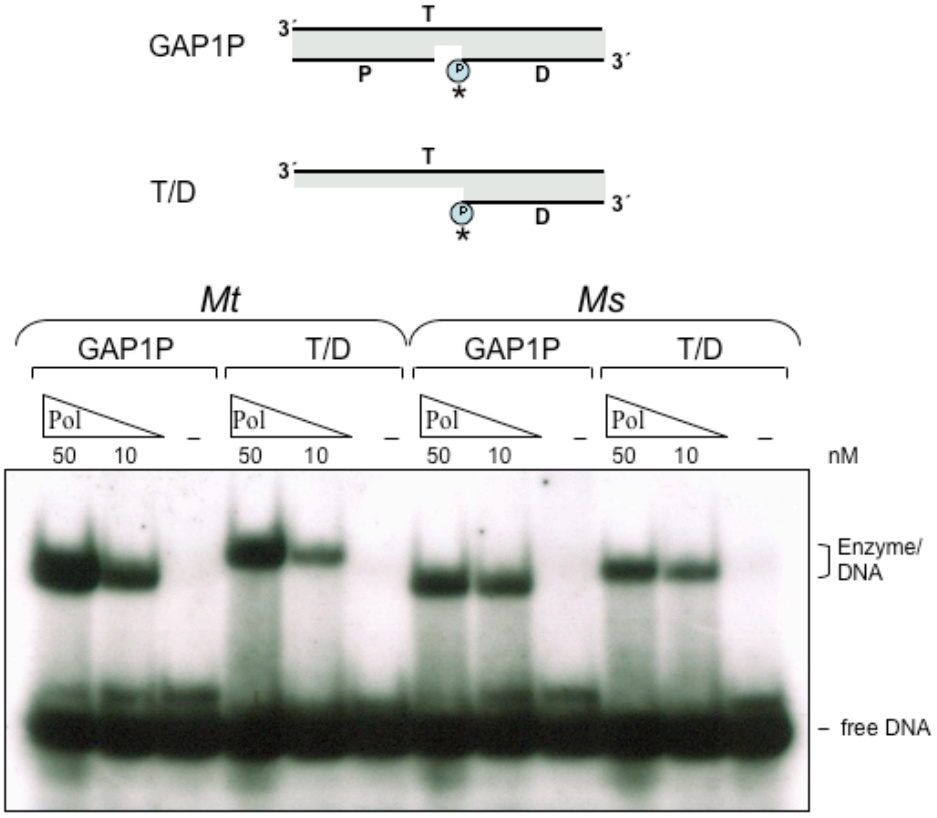
(a)

GAP1P :

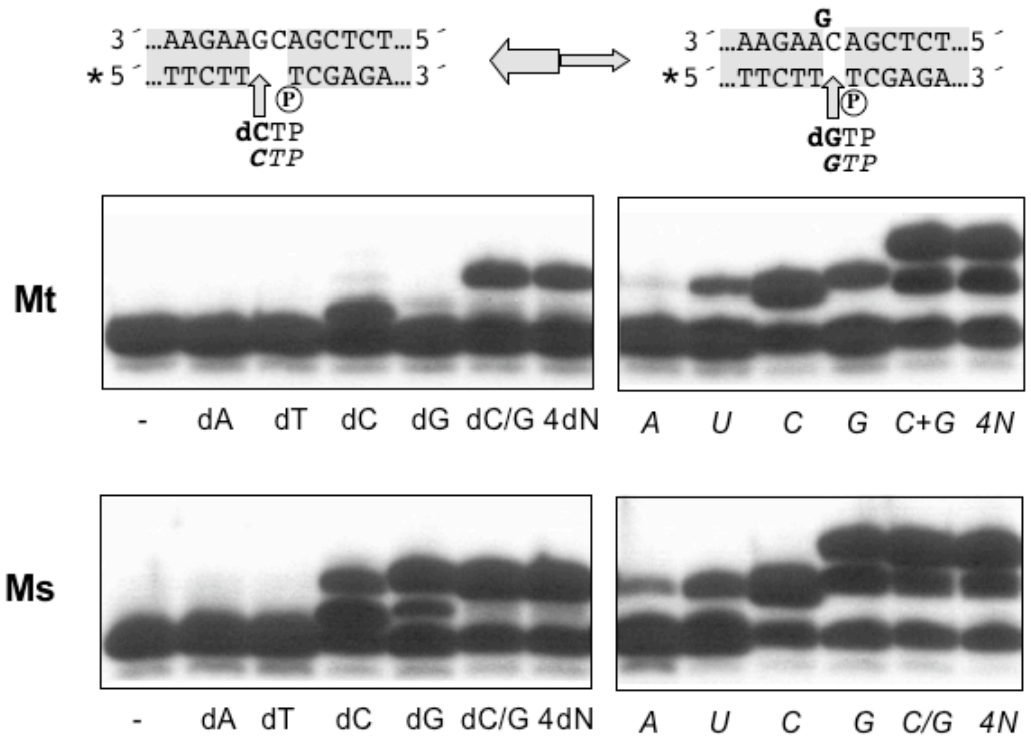
```
T    3´CTAGTGTCACTCATGXTCTATGTGAAGA 5´
P  * 5´GATCACAGTGAGTAC AGATACACTTCT 3´  D
                      Ⓟ
```

(b)

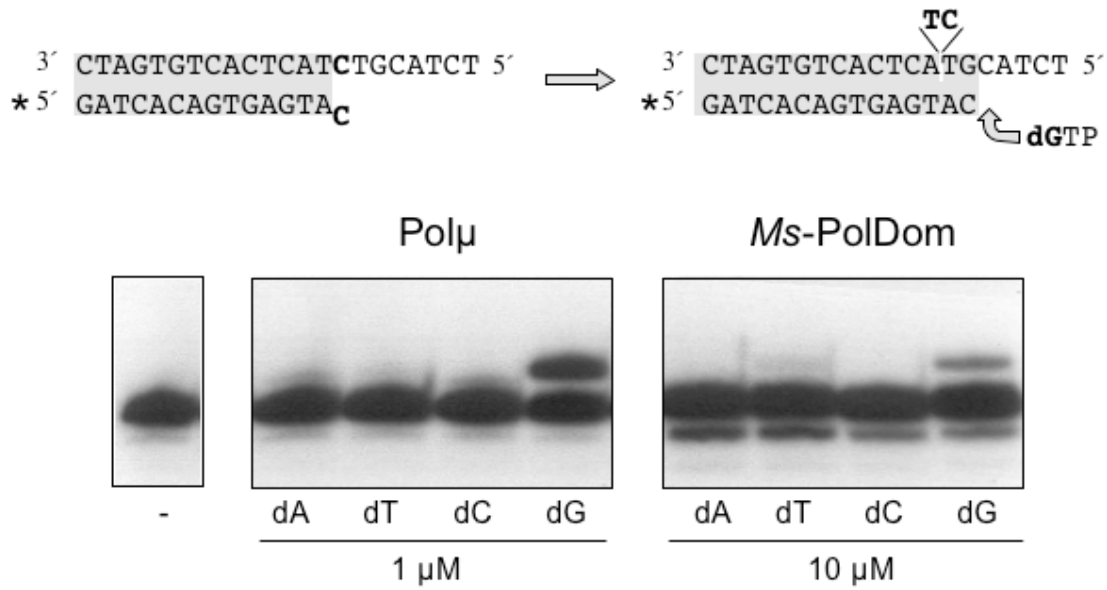


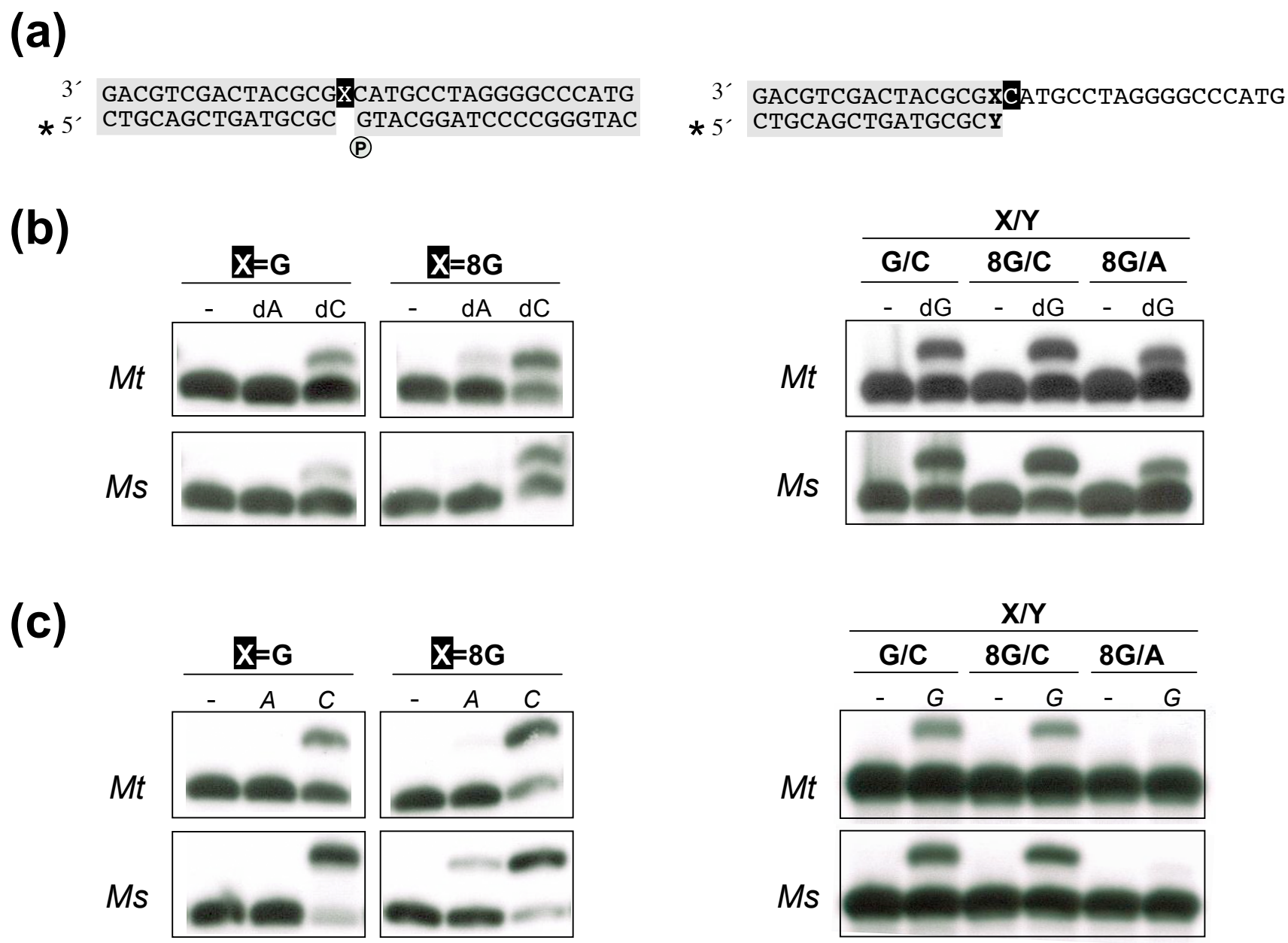

GAP1P
3´
T
P
D
3´
T/D
3´
T
D
3´
Mt
Ms
GAP1P
T/D
GAP1P
T/D
Pol
50
10
–
nM
Enzyme/
DNA
free DNA

a


3´…AAGAAGCAGCTCT…5´
*5´…TTCTT TCGAGA…3´
dCTP
CTP
G
3´…AAGAACAGCTCT…5´
*5´…TTCTT TCGAGA…3´
dGTP
GTP
Mt
- dA dT dC dG dC/G 4dN
A U C G C+G 4N
Ms
- dA dT dC dG dC/G 4dN
A U C G C/G 4N

b


3´ CTAGTGTCACTCATCTGCATCT 5´
*5´ GATCACAGTGAGTAC
TC
3´ CTAGTGTCACTCATGCATCT 5´
*5´ GATCACAGTGAGTAC
dGTP
Polμ
Ms-PolDom
-
dA dT dC dG
dA dT dC dG
1 μM
10 μM


(a)
3′ GACGTCGACTACGCGXCATGCCTAGGGGCCCATG
* 5′ CTGCAGCTGATGCGC GTACGGATCCCCGGGTAC
P
3′ GACGTCGACTACGCGXCATGCCTAGGGGCCCATG
* 5′ CTGCAGCTGATGCGCY
(b)
X=G
X=8G
- dA dC
- dA dC
Mt
Ms
X/Y
G/C
8G/C
8G/A
- dG - dG - dG
Mt
Ms
(c)
X=G
X=8G
- A C
- A C
Mt
Ms
X/Y
G/C
8G/C
8G/A
- G - G - G
Mt
Ms