






# Analysis of diverse double-strand break synapsis with Pol $\lambda$ reveals basis for unique substrate specificity in nonhomologous end-joining

Andrea M. Kaminski <sup>1</sup>, Kishore K. Chiruvella<sup>2</sup>, Dale A. Ramsden <sup>2</sup>, Katarzyna Bebenek <sup>1</sup>✉, Thomas A. Kunkel <sup>1</sup> & Lars C. Pedersen <sup>1</sup>

DNA double-strand breaks (DSBs) threaten genomic stability, since their persistence can lead to loss of critical genetic information, chromosomal translocations or rearrangements, and cell death. DSBs can be repaired through the nonhomologous end-joining pathway (NHEJ), which processes and ligates DNA ends efficiently to prevent or minimize sequence loss. Polymerase  $\lambda$  (Pol $\lambda$ ), one of the Family X polymerases, fills sequence gaps of DSB substrates with a strict specificity for a base-paired primer terminus. There is little information regarding Pol $\lambda$ 's approach to engaging such substrates. We used *in vitro* polymerization and cell-based NHEJ assays to explore the contributions of conserved loop regions toward DSB substrate specificity and utilization. In addition, we present multiple crystal structures of Pol $\lambda$  in synapsis with varying biologically relevant DSB end configurations, revealing how key structural features and hydrogen bonding networks work in concert to stabilize these tenuous, potentially cytotoxic DNA lesions during NHEJ.

<sup>1</sup>Genome Integrity and Structural Biology Laboratory, National Institute of Environmental Health Sciences, National Institutes of Health, 111 TW Alexander Dr., Bldg. 101, Research Triangle Park, NC 27709, USA. <sup>2</sup>Department of Biochemistry and Biophysics, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. ✉email: [bebenek@niehs.nih.gov](mailto:bebenek@niehs.nih.gov)

Genomic integrity exists under imminent threat of breakage and damage due to endogenous exposures to reactive oxygen species generated by normal cellular metabolism, and from exogenous environmental exposures to ionizing radiation<sup>1</sup>. The most toxic form of genomic damage occurs when breaks in the phosphodiester backbone on opposing DNA strands cluster, forming DNA double-strand breaks (DSBs). Such breaks must be repaired quickly and efficiently, since their persistence within the human genome can lead to cancers and other diseases<sup>2</sup>. DSBs can be repaired through a variety of pathways, but nonhomologous end-joining (NHEJ) is favored in non-replicating cells, or those that have yet to proceed through S phase of the cell cycle<sup>3</sup>. Though DSBs can lead to disastrous consequences, they can also prove beneficial when they occur in a controlled, programmed manner—through a specialized form of NHEJ known as V(D)J recombination, which is responsible for immunoglobulin gene maturation<sup>4</sup>.

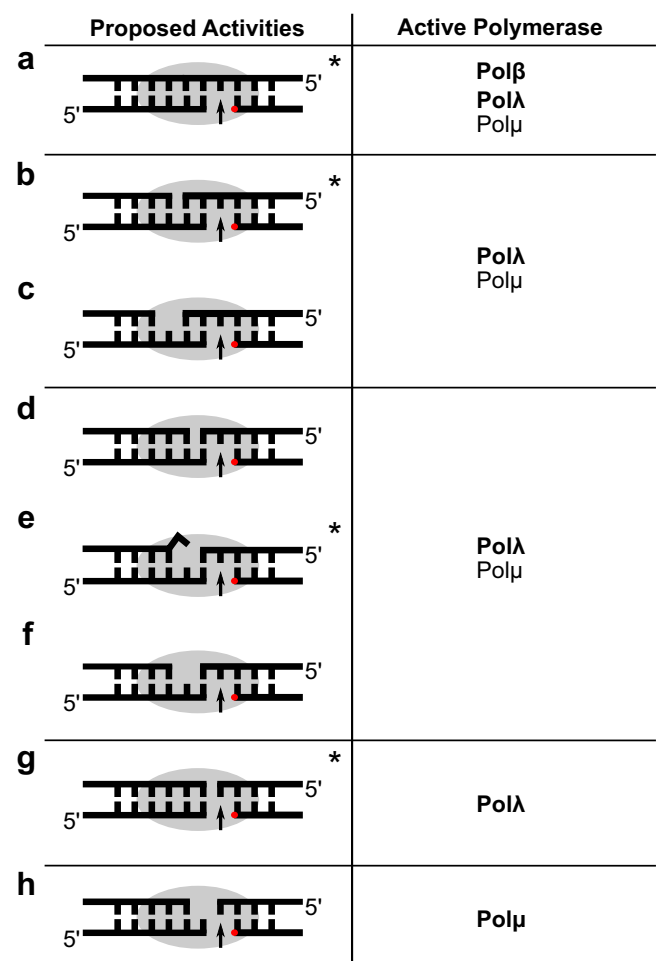
NHEJ is carried out by a multiprotein complex capable of a variety of DNA sequence editing processes. In this pathway, broken DSB ends are bound by the Ku70/80 heterodimer, in conjunction with DNA-PK. Other processing factors are recruited to the complex (Artemis, DNA ligase IV, XRCC4, XLF, and DNA polymerases) to realign and trim the broken/damaged ends, fill in sequence gaps, and reseal the phosphodiester backbone<sup>5</sup>. The Family X polymerases (Pols), Polλ (gene name *Poll*), Polμ (gene name *Polm*), and terminal deoxynucleotidyl transferase (TdT), all participate in V(D)J recombination, each with a specific role<sup>6–8</sup>. Polλ and Polμ are widely expressed, template-dependent enzymes capable of repairing damage-induced DSBs in classical NHEJ, in addition to their roles in V(D)J recombination. Polμ has the unique ability to polymerize on entirely incompatible ends lacking any sequence homology at the break site, where the primer terminus is unpaired<sup>9</sup> (Fig. 1). In contrast, Polλ is preferentially active on DNA substrates with a base-paired primer terminus. Polλ thus acts on DSB substrates containing 3'-overhangs with one or more complementary nucleotides that aid in repair of broken upstream/downstream ends—a context where Polμ is only modestly active. Polλ is also predominantly responsible for initiating synthesis from a blunt-ended upstream end, using as template the 3'-overhang from a downstream end in *trans*<sup>10</sup>. To better understand the structural determinants of how Polλ engages its preferred DSB substrates, we determined the X-ray crystal structures of its catalytic domain mediating end synapsis with a myriad of single- and double-strand break configurations that have been shown to be physiologically relevant substrates<sup>9,10</sup> (Fig. 1). These structures demonstrate how Polλ engages DSB substrates, employing conserved hydrogen-bonding networks to synergistically stabilize both sides of the break simultaneously. A long, flexible loop region in the Polλ thumb subdomain was found to be specifically essential for blunt-end substrate utilization.

## Results

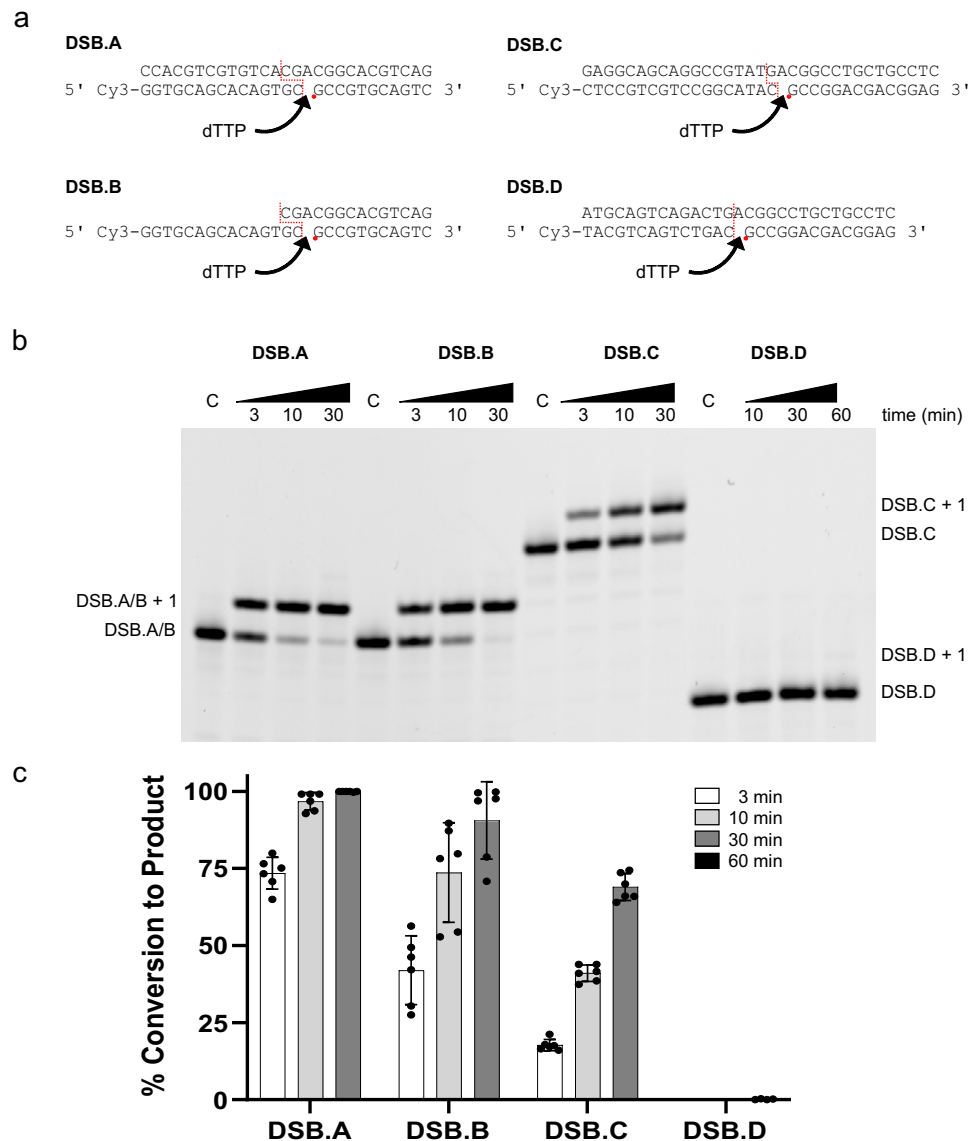
**Structural determinants of Polλ DSB substrate specificity.** In this study, we focused on how Polλ mediates synapsis on physiologically relevant DSB substrates containing either two nucleotides (2nt—DSB.A and DSB.B, similar to Fig. 1b, c), one nucleotide (1nt—DSB.C, similar to Fig. 1d–f), or no (blunt-end—DSB.D, similar to Fig. 1g) overlap between the upstream and downstream duplexes. We first employed a polymerization assay to determine how efficiently synapsis and subsequent gap-filling are accomplished on these substrates *in vitro* (Fig. 2a), in the absence of other NHEJ protein binding partners. These assays reveal that Polλ, alone, binds and fills gaps on substrates DSB.A, DSB.B, and DSB.C containing microhomology at the break site,

which aids in bridging the upstream and downstream duplexes (Fig. 2b). The extent of microhomology also contributes to more efficient synapsis and gap-filling, as indicated by the increase in product formation in substrates containing two-nucleotide (DSB.A and DSB.B) rather than single-nucleotide microhomology (DSB.C). Without the aid of either NHEJ protein binding partners or sequence microhomology to bridge the upstream and downstream ends, Polλ struggles to efficiently fill the gap on the blunt-end DSB.D, as evidenced by the minimal amount of observed incorporation.

Since repair of the blunt-end DSB substrates is primarily Polλ-specific<sup>10</sup>, we attempted to define the structural determinants bestowing Polλ, rather than Polμ, with this capability. The structural superposition of pre-catalytic SSB ternary complexes of these polymerases (Polλ PDB ID code 1XSN<sup>11</sup> and Polμ PDB ID code 4M04<sup>12</sup>) reveal that their global structures are similar (RMSD of 1.55 Å over 257 Ca atoms), with differences primarily



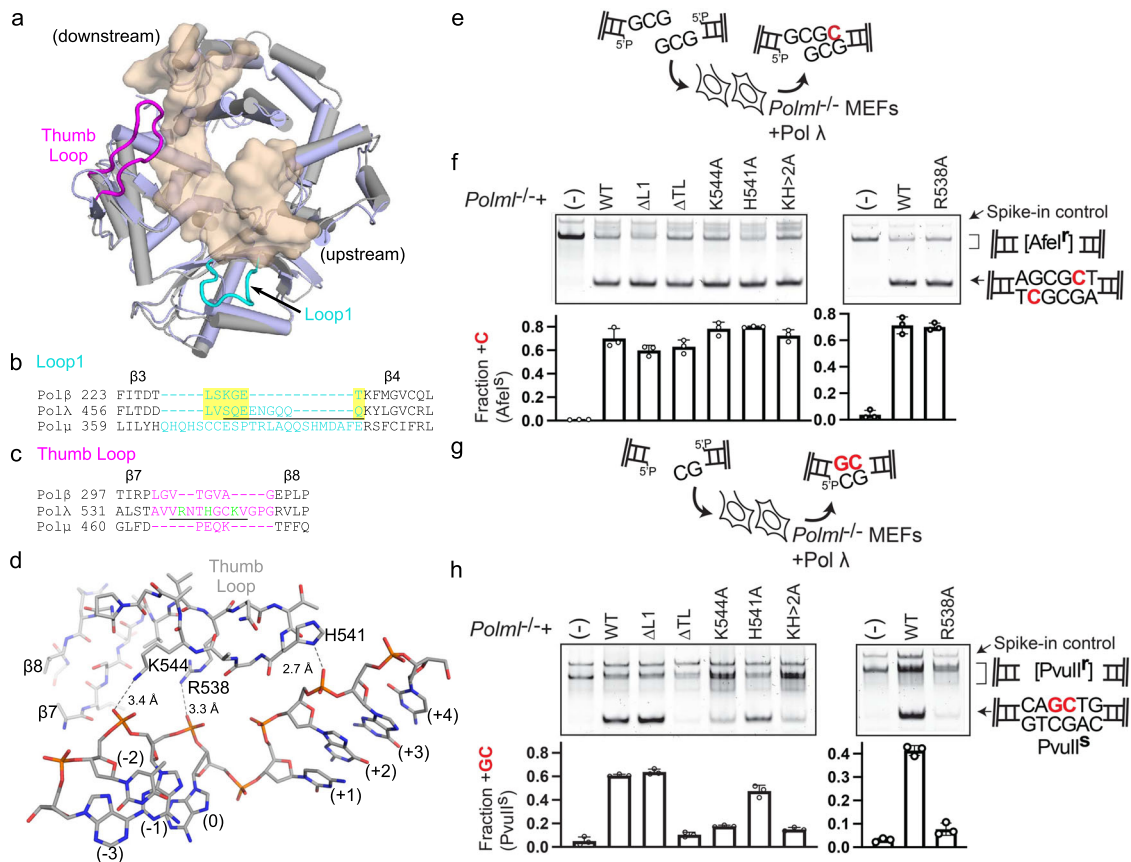
**Fig. 1 Schematic diagram of DNA substrates utilized by Family X polymerases.** Single- and double-strand break DNA substrates are shown in stick on the left, with the rough footprint of the bound polymerase indicated by a gray oval. Locations of the 5'-phosphate (if present) and the incoming nucleotide to fill the gap are given by a red dot and an upward black arrow, respectively. Polymerases that have been found to utilize each type of substrate are listed on the right, with the identities of the polymerase shown to predominantly repair this configuration shown in bold. Types of substrates include single-strand breaks (a) and double-strand breaks with 2nt (b, c) or 1nt (d–f) complementarity, blunt-ended breaks (g), or noncomplementary breaks (h). Crystal structures of substrates with asterisks will be presented in this study.



**Fig. 2** Pol $\lambda$  synapses and fills a variety of DSB end-configurations. **a** DSB substrates utilized for in vitro gap-filling assays. All substrates include a 5'-Cy3-labeled upstream primer and a 5'-phosphorylated (indicated by red dot) downstream primer. The break site location is indicated by a dashed red line. **b** Gap-filling activity of full-length wildtype Pol $\lambda$  on the different DSB end configurations examined in this study. **c** Quantification of gap-filling proficiency. The mean conversion from substrate to product is shown for each substrate, and the error bars correspond to the standard deviation. Reactions were prepared as biological replicates, with  $n = 6$  for DSB substrates A-C, and  $n = 4$  for DSB substrate D. The % conversion for the DSB.D substrate was less than 0.5% for all replicates, after a 60 min incubation.

in the regions of Loop1 and the 'thumb loop' (Fig. 3a). Loop 1 influences polymerase fidelity in Pol $\lambda$ <sup>13,14</sup>, allows Pol $\mu$  to repair DSB ends with unpaired primer termini<sup>9</sup>, and endows Pol $\mu$  and TdT with template-independent synthesis behavior<sup>15,16</sup>. Loop1 varies in length between Family X members, with the longest loops observed in TdT or Pol $\mu$ . A smaller loop exists in Pol $\lambda$  and dwindles to a mere turn in Pol $\beta$  (Fig. 3b). However, Loop1 is disordered in all DNA-bound Pol $\mu$  structures<sup>12,17</sup>. The sequence of Loop1 varies between the Family X members, but is well conserved between orthologs of Pol $\lambda$ <sup>13</sup>. The 'thumb loop' (TL), connecting  $\beta$ -strands 7 and 8 in the thumb subdomain, varies in length between the Family X polymerases, with the longest loop in Pol $\lambda$  (Fig. 3c) exhibiting conservation of both length and sequence. The TL is considerably shorter in Pol $\beta$  and exists as a turn in Pol $\mu$ . This loop is often disordered in Pol $\lambda$ <sup>18</sup> structures, but has been previously observed in an ordered conformation, juxtaposed with the template strand (PDB ID code 1XSN<sup>11</sup>,

Fig. 3d). When ordered, residues in the TL (Lys544, Arg538, and His541) hydrogen bond with nonbridging oxygens along the template backbone. In order to determine whether Loop1 or the TL contribute to Pol $\lambda$ 's utilization of blunt-end DSB substrates during NHEJ, loop deletions or point mutations were created (Fig. 3b, c). Due to the relative inefficiency of synopsis and gap-filling on this substrate in vitro (Fig. 2b), a cell-based assay utilizing the endogenous cellular NHEJ machinery was instead used to explore the consequences of the loop deletions. All variants exhibited similar levels of end-joining on a 1nt gapped complementary DSB substrate with 2nt microhomology at the break site (Fig. 3e, f), suggesting these alterations do not broadly affect polymerase activity. The Pol $\lambda$  Loop1 variant<sup>13</sup> ( $\Delta$ L1) also exhibited end-joining similar to that of the wildtype on a blunt-end substrate with a 2nt 3'-overhang (Fig. 3g, h). In contrast, the  $\Delta$ TL deletion variant was markedly deficient in correct repair of this substrate, suggesting a significant role for the 'thumb loop'



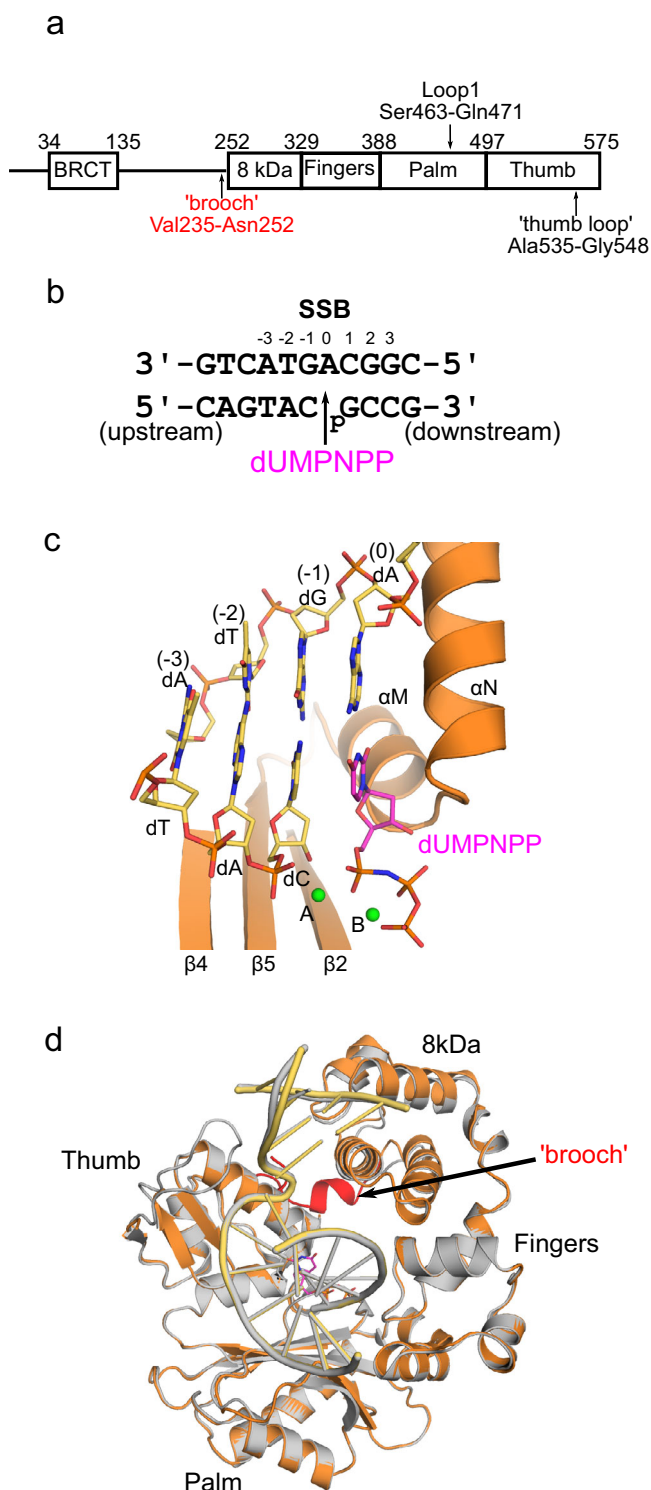
**Fig. 3** Determinants of DSB substrate specificity by Polλ. **a** Structural superposition of 1nt SSB crystal structures of Polλ (PDB ID code 1XSN<sup>11</sup>, protein in gray and DNA in transparent khaki surface) and Polμ (PDB ID code 4M04<sup>12</sup>, protein in light blue). Loop1 and the ‘thumb loop’ regions of Polλ are shown in cyan and magenta, respectively. **b** Structure-based sequence alignment of Loop1 (cyan) in Pols β, λ, and μ (structurally-equivalent regions highlighted in yellow). The Polλ ΔL1 variant was generated by deleting Ser463-Gln471 (underlined) and replacing them with Lys462-Thr465 from Polβ<sup>13</sup>. **c** Sequence alignment of ‘thumb loop’ (TL, magenta) of Pols β, λ and μ. The Polλ ΔTL variant was generated by deleting Val537-Val545 of Polλ (underlined) and fusing the ends of β-strands 7 and 8 with a single glycine residue. Thumb loop mutations used in this study are marked in green. **d** Interactions of the TL with an SSB substrate (PDB ID code 1XSN<sup>11</sup>, gray). Hydrogen bonds with the DNA template strand are shown as black dashed lines. **e** DNA fragments with partly complementary ends (GCG3′ overhangs) were introduced into *Polm*<sup>-/-</sup> *Pol*<sup>-/-</sup> cells (*Polm*<sup>-/-</sup>). **f** Representative gel electrophoresis (top) identification of the fraction of NHEJ that is Afel-sensitive and polymerase-dependent. Quantification of the average fraction of polymerase-dependent repair using GCG3′ DSB substrates from triplicate electroporations (bottom; error bars represent the standard deviation). **g** As in (e), except using noncomplementary blunt and GC3′ overhangs. **h** As in (f), except using PvuII sensitivity to identify polymerase-dependent repair.

specifically, in blunt-end substrate utilization. To further dissect the role of the ‘thumb loop’, alanine substitution mutations of Arg538, Lys544 and His541 were also tested. The H541A mutant showed a subtle decrease in blunt-end DSB repair, while the R538A and K544A mutants were nearly as detrimental as the full loop deletion (Fig. 3h). Observation of such a substantial effect for R538A was surprising, given that Arg538 interacts with the template strand in the SSB, but the same interaction would not exist since this phosphate is not present in the blunt-end DSB. End-joining by the H541A/K544A double mutant was nearly indistinguishable from that of K544A, suggesting that the double mutations’ effects are not additive and that Lys544 plays the more significant role in extension from a blunt end. Interestingly, comparison of the in vitro and cell-based gap-filling assays show that, while Polλ struggles to independently synapse and repair the blunt-end DSB substrate in vitro (Fig. 2b), this difficulty appears to be entirely overcome by the presence of the NHEJ complex, as suggested by the similar amounts of complementary and blunt-end repair by the wildtype enzyme in vivo (Fig. 3f, h).

**Synapsis by Polλ on a fully complementary DSB substrate.** Since Polλ was shown to be capable of synapsis and gap-filling in

the absence of other NHEJ binding partners, we attempted to characterize the different synaptic complexes structurally, using X-ray crystallography. The Family X members involved in end-joining, Polλ, Polμ, and TdT, contain a small, conserved structural motif—known as the ‘brooch’—immediately preceding the 8 kDa subdomain (Fig. 4a). This motif facilitates end-bridging and gap-filling of DSB-containing DNA substrates, and has been hypothesized to aid in maintaining a closed conformation of the polymerase domain throughout the catalytic cycle<sup>19</sup>. Prior to the elucidation of the role of this motif, the ‘brooch’ historically had not been included in catalytic domain constructs used for crystallization, which were designed based on the structure of Polβ<sup>11</sup>. In order to increase the likelihood of crystallizing more tenuous, unstable DSB complexes, we utilized a ‘brooch’-containing catalytic domain construct of Polλ (Val235-Trp575) for the current study. This construct was first co-crystallized in a pre-catalytic ternary complex with a single-nucleotide (1nt) gapped single-strand break (SSB, Fig. 4b, similar to Fig. 1a) substrate (1.57 Å, PDB ID code 7M07, Supplementary Table 1, Fig. 4, Supplementary Fig. 1a). Superposition of this pre-catalytic complex with that of a previously reported Polλ<sup>18</sup> pre-catalytic ternary complex lacking the brooch (PDB ID code 2PFO) shows that they are





**Fig. 4 Characterization of Polλ 'brooch'-containing constructs with a 1nt-gapped SSB.** **a** Domain organization of human Polλ, with key structural features indicated. **b** 1nt-gapped SSB DNA substrate co-crystallized with the Polλ catalytic domain containing the 'brooch' motif. For the pre-catalytic complex, the template (chain T, top), upstream primer (chain P, bottom left), and 5'-phosphorylated downstream primer (chain D, bottom right) are pre-annealed and complexed with Polλ and an incoming nonhydrolyzable dUMPNPP nucleotide (magenta). **c** Zoomed-in view of the pre-catalytic SSB active site (protein in orange, DNA and incoming nucleotide colored as in **(b)**, Mg<sup>2+</sup> as green spheres). **d** Global superposition of the pre-catalytic ternary complexes of the Polλ catalytic domains in the absence (PDB ID code 2PFO<sup>18</sup>, gray) and presence (PDB ID code 7M07, protein in orange, DNA in pale yellow, dUMPNPP in magenta) of the 'brooch' motif (red).

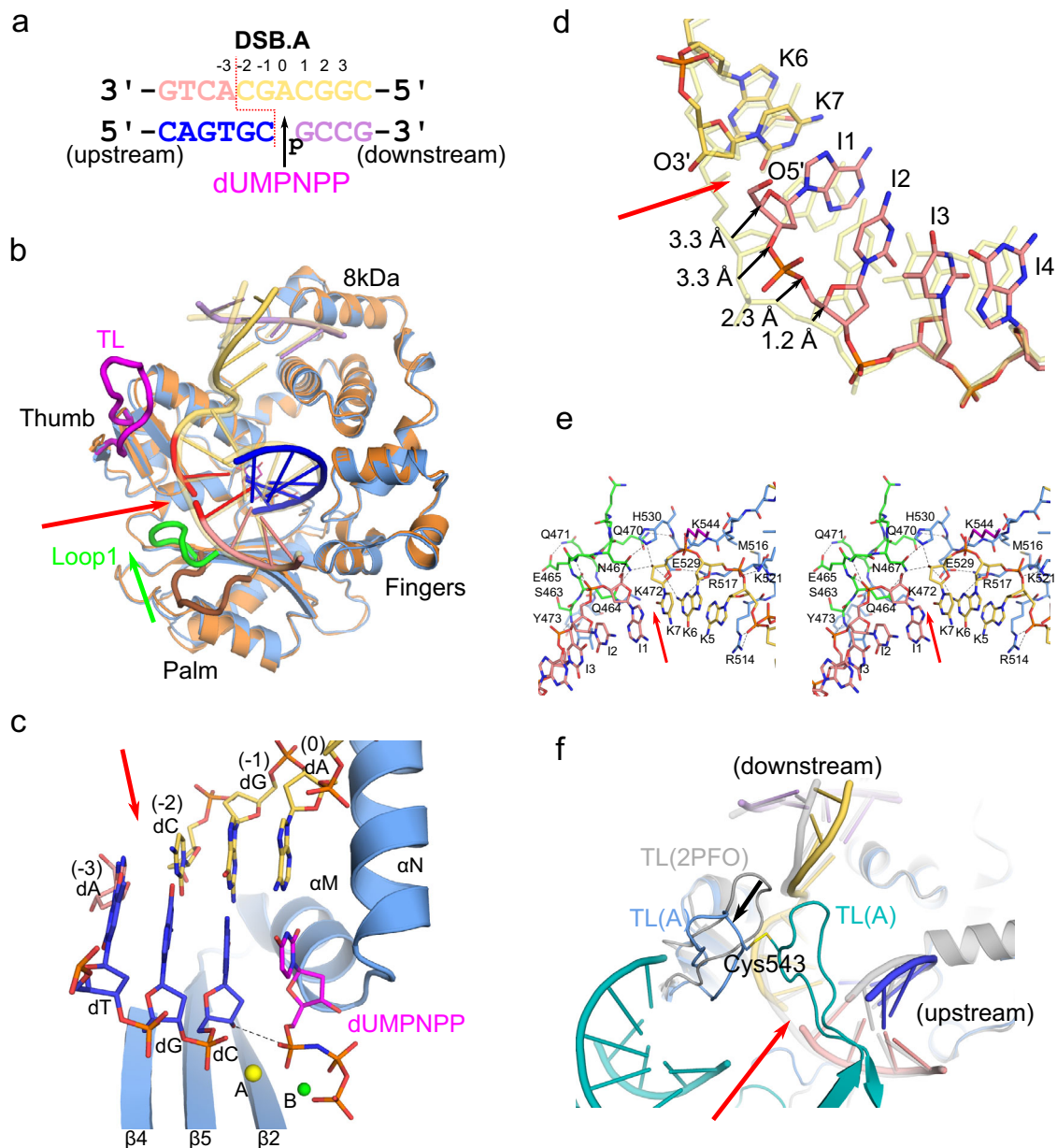
complexes per asymmetric unit, which display a high degree of structural similarity (RMSD of 0.21 Å over 272 Ca atoms, Supplementary Fig. 3b). The electron density for the active site in molecule B was slightly clearer (Supplementary Fig. 1b) than for molecule A and was therefore used for structural analysis. The active site of this pre-catalytic complex contains the nonhydrolyzable dUMPNPP bound in the nascent base pair binding site, along with two metals (Na<sup>+</sup> in metal A site, Mg<sup>2+</sup> in metal B site, Fig. 5c).

Superposition of the DSB.A and SSB complexes reveals global comparability (RMSD of 0.63 Å over 286 Ca atoms), though distinct differences are observed. Surprisingly, the template strand is shifted slightly from its position in the SSB complex (Figs. 5b and 5d). The templating base in the nascent base pair binding site (K7) and its immediate downstream neighbor (K6) occupy a very similar position, while the template strand upstream of the break site is rotated approximately 3 Å outward from its canonical position on the palm subdomain (Fig. 5d). The complementary primer strand nucleotides in this region maintain a similar position to those observed in the SSB complex, giving the appearance that the left-hand organization of the polymerase domain 'holds' tightly to the primer while allowing some movement of the template strand. The template strand shift coincides with a Loop1 (Ser463-Gln471, between β-strands 3 and 4, Fig. 5b, green) conformation different from those observed in either historical binary<sup>20</sup>/ternary<sup>18,21</sup> Polλ complexes, or the SSB in the current study (Fig. 5b, brown). This new position allows Loop1 to interact with both the thumb subdomain (Gln470 NE2 – Glu529 O) and the DSB.A substrate, proximal to the break site (Fig. 5e and Supplementary Table 2). Rather than the backbone amide nitrogens from Gln471 and Lys472 forming a pocket anchoring the phosphate between template nucleotides at the –4 and –3 positions, this role is now played by the backbone amides of Glu465 and Glu466, with the amide of Ser463 contacting the next upstream phosphate (I3). Asn467 mediates a hydrogen-bonding network between Gln464 and His530. His530, Asn467, and Glu529 then work in concert to position the 3'-OH on residue K7 at the break site. Glu529 then extends the hydrogen bonding network to the downstream side of the break, with Arg517. Lys544 assumes an extended conformation and lies within hydrogen-bonding distance of the nonbridging phosphate oxygens between template nucleotides K6 and K7. Whereas Lys521 normally interacts with the phosphate backbone immediately downstream of the 90° bend in the template strand in the SSB (equivalent to the phosphate between K3 and K4), it now interacts with the phosphate immediately upstream of the bend (between K5 and K6).

In addition to alterations in Loop1, there is also a distinct difference in the order and conformation of the 'thumb loop' (TL)

nearly indistinguishable (RMSD of 0.28 Å over 240 Ca, Fig. 4d and Supplementary Fig. 2), validating the use of the current SSB complex as a relevant reference model.

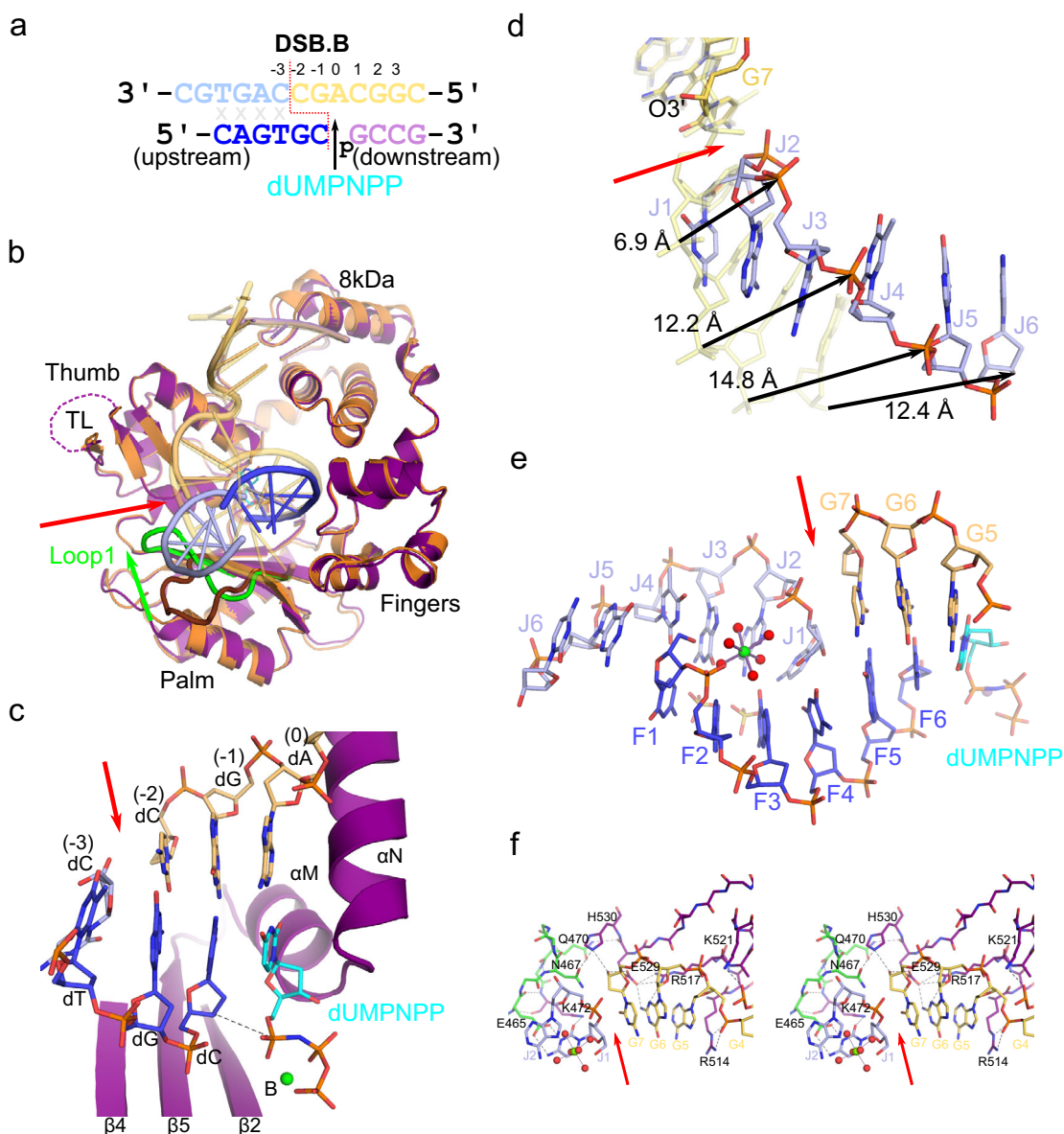
In order to understand how Polλ mediates synopsis of DSB ends during NHEJ, the 'brooch'-containing catalytic domain construct was crystallized with a fully complementary DSB substrate containing a 3nt downstream template 3'-overhang, placing the break site between the –2 and –3 positions (DSB.A, Fig. 5a, similar to Fig. 1b). The DSB.A pre-catalytic complex (1.8 Å, PDB ID code 7M0D, Supplementary Table 1, Fig. 5b and Supplementary Fig. 3a) crystallized with two quaternary



**Fig. 5 Polλ in complex with entirely complementary DSB.A.** **a** Complementary DSB.A substrate (template strand broken between the  $-2$  and  $-3$  positions) co-crystallized with the Polλ catalytic domain (molecule B). The break site is indicated by a red dashed line. Upstream (left) template (pink, chain I) and primer (blue, chain J) are annealed separately from the downstream (right) template (pale orange, chain K) and primer (lavender, chain L) strands, and synapsis is mediated entirely by the polymerase. The DSB.A pre-catalytic quaternary complex is formed by addition of the nonhydrolyzable incoming dUMPNNP (magenta) nucleotide. **b** Global superposition of the pre-catalytic SSB ternary (orange) and complementary DSB.A quaternary (protein in blue, DNA and incoming dUMPNNP colored as in **a**) complexes. Positional differences of Loop1 (green), the thumb loop (TL, magenta), and the DNA template strand between the DSB.A and SSB structures are highlighted. Break site is indicated by a red arrow. **c** Structure of the DSB.A substrate bound in the Polλ active site (protein in light blue; DNA and incoming nucleotide drawn in stick and colored as in **a**). The  $\text{Na}^+$  and  $\text{Mg}^{2+}$  ions are shown as yellow and green spheres, respectively. The path between the primer terminal 3'-OH (axial) and the  $\alpha$ -phosphate of the incoming dUMPNNP is shown (black dashed line). **d** Zoomed-in view of template strand shift in DSB.A (colored as in **a**), superimposed with the SSB (transparent khaki). **e** Stereo diagram of the interactions surrounding the break site (red arrow) in the DNA template strand (protein in light blue, DNA colored as in **a**), with hydrogen bonding interactions shown (black dashed lines). Residues from Loop1 and the 'thumb loop' are colored green and magenta, respectively. **f** Thumb loop in 7M0D (Ala535-Gly548) is ordered and differs in conformation from that observed in 2PFO<sup>18</sup> (gray). Thumb loop conformation may be influenced by crystal packing interactions, as the thumb loop in molecule A (teal) interacts with the end of the upstream duplex from molecule B (colored as in **a**), and vice versa. The disulfide bond between Cys543 in each loop is indicated.

between  $\beta$ -strands 7 and 8 in the thumb subdomain (Fig. 5b, magenta). The 'thumb loop' is often disordered in Polλ structures, but is ordered in both molecules within the asymmetric unit of the DSB.A pre-catalytic quaternary complex and adopts the same conformation in each (Supplementary Fig 3b), which differs from

that observed in other previously reported structures (PDB ID code 2PFO<sup>18</sup>, Fig. 5f, gray). Within the current lattice, the two 'thumb loops' are in close contact—mediated by a disulfide bond between the Cys543 sidechains—and interact with the DNA substrate.



**Fig. 6** Polλ mediates synapsis via primer terminal pairing and accommodates noncanonical structures. **a** DSB.B substrate co-crystallized with Polλ catalytic domain (molecule A), with break site indicated by a red dashed line. Upstream (left) primer (blue, chain F) was annealed separately from the downstream (right) template (pale orange, chain G) and primer (lavender, chain H) strands, and synapsis is mediated by the polymerase. Surrogate of upstream template (light blue, chain J) is the 5'-end of the upstream primer from a neighboring molecule, which mispairs (gray x) with the current upstream primer (blue). DSB.B pre-catalytic quaternary complex is formed by addition of nonhydrolyzable incoming dUMPNPP (cyan) nucleotide. **b** Global superposition of pre-catalytic SSB ternary (protein in orange, DNA in transparent yellow) and DSB.B synaptic (protein in purple, DNA and incoming dUMPNPP colored as in **a**) complexes. Positional differences of Loop1 (green) and the DNA template strand between the DSB.B and SSB structures are highlighted. Break site is indicated by a red arrow. **c** Structure of the DSB.B synaptic complex bound in the Polλ active site (protein in purple; DNA and incoming nucleotide colored as in **a**). While the metal A site is unoccupied, while the metal B site is fully occupied by a  $Mg^{2+}$  ion (green sphere). The path between the primer terminal 3'-OH (axial) and the  $\alpha$ -phosphate of the incoming dUMPNPP is shown (black dashed line). **d** Zoomed-in view of the template strand shift in DSB.B (colored as in **a**), superimposed with the template strand from the SSB (transparent khaki). **e** The upstream duplex (primer in blue, upstream template from neighboring molecule in light blue) forms an entirely mispaired duplex, mediated by hydrogen bonding interactions with a hydrated  $Mg^{2+}$  ion (green, with coordinating waters, red spheres) and alternate conformations of sulfate ions (yellow) from the mother liquor. Expanded views of individual mispairs are displayed in Supplementary Fig. 5. **f** Stereo diagram of interactions surrounding the break site (red arrow) in the DNA template strand (protein in dark purple, DNA colored as in **a**), with hydrogen bonding interactions drawn as black dashed lines. Residues from Loop1 are shown in green.

**Polλ tolerates upstream DSB synaptic noncomplementarity.** We next attempted to co-crystallize the Polλ catalytic domain in complex with a substrate similar to DSB.A, but with a single-nucleotide template gap at the -3 position (similar to Fig. 1c). However, the resulting crystals exhibited an entirely unexpected

—yet fortuitous—synaptic arrangement (DSB.B, Fig. 6a) *in trans*, where the upstream template oligo never annealed. These crystals were subsequently reproduced in the absence of the upstream template oligo (2.25 Å, PDB ID code 7M0E, Supplementary Table 1 and Fig. 6), wherein the 5'-end of the upstream primer



forms an entirely noncomplementary, mispaired duplex with the 5'-end of the same primer from a neighboring molecule in the asymmetric unit (Supplementary Fig. 4). The pre-catalytic complex was trapped by binding of the nonhydrolyzable dUMPNPP (Fig. 6b, c). Despite vastly different crystal packing, the DSB.B synaptic complex displays a high degree of global structural comparability to that of the SSB complex (RMSD of 0.53 Å over 281 Ca atoms, Fig. 6b). Components within the active site (primer terminus and correct base pairs at the 0, -1, and -2) are observed in their expected positions (Fig. 6c), however, the farther upstream along the DSB.B substrate from the catalytic center, the more this synaptic complex deviates from canonical geometry (Fig. 6d). For the upstream primer, a high degree of structural similarity is maintained while the substrate lies within the footprint of the palm subdomain (nucleotides F4-6), likely anchored by a myriad of hydrogen bonding interactions with the palm subdomain and coordination of the Na<sup>+</sup> ion in the Helix-hairpin-Helix (HhH) 2 site. The upstream primer progressively diverges from the canonical primer strand trajectory (distances of 1.1 Å, 2.4 Å, and 4.4 Å at nucleotides F3, F2, and F1, respectively). This deviation is likely due to mispairing of the upstream primer strand from the neighboring molecule, which serves as a surrogate for the missing upstream template strand. This surrogate template strand no longer occupies the conventional position of the template strand, as observed for the SSB or DSB.A complexes and has drastically shifted—by 6.9 Å at the break site to 14.8 Å closer to the upstream end (Fig. 6d). The noncomplementary interactions of the 5'-ends of these oligos creates an entirely mispaired region upstream of the break site (Fig. 6e and Supplementary Fig. 5), juxtaposed between the coordination of a solvent-exposed, hydrated Mg<sup>2+</sup> ion and partially ordered sulfate ions from the mother liquor.

Interactions downstream of the break site are largely conserved in the DSB.B synapse, similar to those observed in the DSB.A complex. Arg514 stacks with the templating base and anchors the phosphate backbone at the 90° bend between downstream template nucleotides G4-5 and Lys521 lies within hydrogen-bonding distance of the next phosphate downstream (Fig. 6f and Supplementary Table 2). Arg517 extends into the minor groove and interacts with the G6 base and the sidechain of Glu529, which participates in a small hydrogen bonding network with the G6 base and the 5'-OH of G7. His530 also helps to position the 5'-OH and mediates interactions with Asn467. Because the surrogate template residue J1 is rotated inward into the duplex, rather than maintaining the continuous flow of the phosphate backbone within the binding cleft, its base now participates in multiple hydrogen-bonding interactions, with the sidechains Lys472 and Glu465. Loop1 adopts a conformation different from either of those observed in the SSB and DSB.A complex (Fig. 6b, green), which cradles the unconventional DNA substrate configuration upstream of the break site.

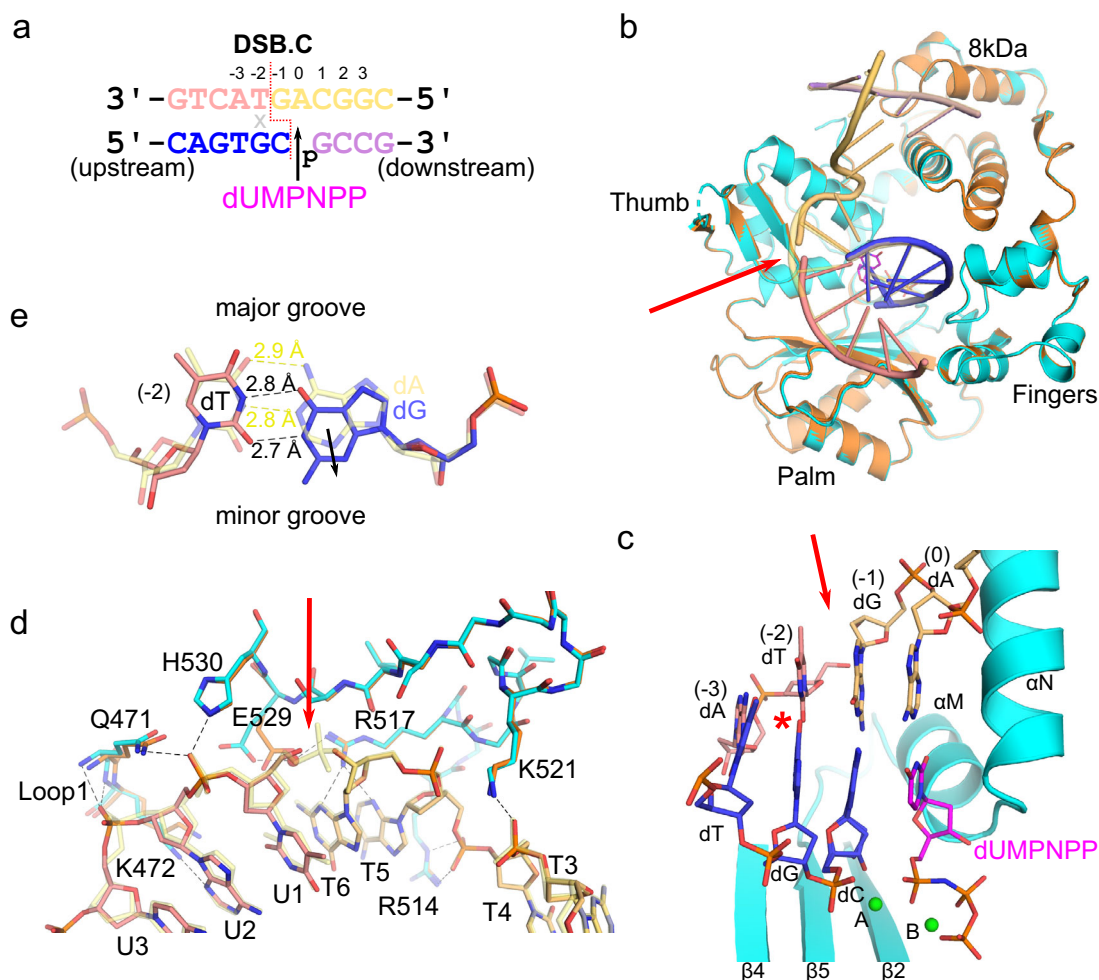
**Polλ can accommodate mispairing near the DSB break site.** To further explore Polλ's requirement for a paired primer terminus in gap-filling during NHEJ<sup>9,22</sup>, we next co-crystallized its catalytic domain in complex with a DSB substrate comprised of an upstream duplex and a 2nt downstream template 3'-overhang, yielding a break between the -1 and -2 positions (DSB.C, Fig. 7a, similar to Fig. 1e). Therefore, the polymerase may mediate DSB synopsis of the upstream and downstream duplexes, aided by a single correct base pair on the primer terminus. This substrate also contains a G:T (primer:template) mismatch immediately upstream of the break site, which could give insight into Polλ's tolerance for such nonhomology in DSB substrates. The pre-catalytic quaternary DSB.C complex was assembled in a

similar fashion to that of DSB.A-B but crystallized with the same space group and unit cell dimensions (2 Å, PDB ID code 7MOB, Supplementary Table 1 and Fig. 7) as the 1nt SSB. Structural superposition shows a high degree of similarity between the DSB.C pre-catalytic complex and that of the SSB (RMSD of 0.224 Å over 302 Ca atoms) (Fig. 7b). There is clear density for the backbone discontinuity between the -1 and -2 positions (Supplementary Fig. 1d). The DNA substrate in the DSB.C complex superimposes well with that of the SSB—therefore, interactions with the substrates are largely identical between these complexes (Fig. 7d and Supplementary Table 2). This observation suggests that Polλ likely utilizes comparable DNA substrate binding modes, whether the template strand is unbroken or contains a break within one or two positions upstream from the nascent base pair binding site (0 position). There are a few subtle differences observed between these complexes. Lys472 exhibits conformational flexibility, which often places the sidechain within hydrogen-bonding distance of template bases upstream of the break site. Glu529 is observed in distinct rotamers in each structure, which allow it to form different interactions in each case. In the current conformation, a water molecule mediates a hydrogen bond bridge between Glu529 and Arg517. The flexibility of this residue is intriguing, since previous studies have shown that Glu529 functions as a sensor preventing mutagenic bypass of a template strand 8-oxoG<sup>23</sup>.

The presence of the G:T mispair near the primer terminus does not overtly distort the upstream duplex, though the mispaired primer dG, is shifted very slightly (~1 Å) toward the minor groove in a wobble conformation (Fig. 7e). The incorrect G:T mispair allows for two hydrogen bonds between the bases, which are of similar distance (2.8 and 2.7 Å) and geometry to those in the correct A:T base pair in the SSB complex (2.9 and 2.8 Å). Both the primer dG and the template dT bases are observed in the *anti* conformation, which maintains the cognate base shape<sup>24</sup> of a purine:pyrimidine base pair. This suggests that Polλ would likely be tolerant of similar minimally-distorting mismatches proximal to DSB break sites.

**DSB synopsis by Polλ on a blunt-ended DSB substrate.** We then assessed how Polλ mediates end synopsis of a DSB substrate with a blunt-ended upstream duplex and a single nucleotide 3'-overhang from the downstream duplex (DSB.D, Fig. 8a, similar to Fig. 1g), as previous studies have shown that repair of this substrate configuration is primarily Polλ-dependent<sup>10</sup>. Since these quaternary complex crystals are grown in a mother liquor containing a high concentration of Ca<sup>2+</sup>, which is incompatible with nucleotide incorporation, we attempted to exchange the active site metals with Mg<sup>2+</sup> (pre-catalytic complex, 1.65 Å, PDB ID code 7M09, Supplementary Table 1), and then to exchange the nonhydrolyzable dUMPNPP with hydrolyzable dTTP to allow for *in crystallo* nucleotide incorporation (partially incorporated complex, 1.83 Å, PDB ID code 7M0A, Supplementary Table 1 and Fig. 8c). The structural superposition of the DSB.D pre-catalytic complex shows that it is nearly indistinguishable from the pre-catalytic SSB-bound complex (RMSD of 0.11 Å over 292 Ca atoms, Fig. 8b). There is a gap in the electron density between the templating base (0 position) and the -1 position (Supplementary Figs. 1e, inset and 1f). In the pre-catalytic complex, a mixture of metals is observed in the A site (refined with occupancies of 70% and 30% for Ca<sup>2+</sup> and Mg<sup>2+</sup> ions, respectively), indicating incomplete metal exchange at that location. Full occupancy for Mg<sup>2+</sup> is observed in the metal B site. The primer terminus is observed in a conformation intermediate between C2'-endo and C3'-endo, placing its 3'-OH neither fully axial nor fully equatorial (distance of 3.7 Å to α-phosphate of dUMPNPP,



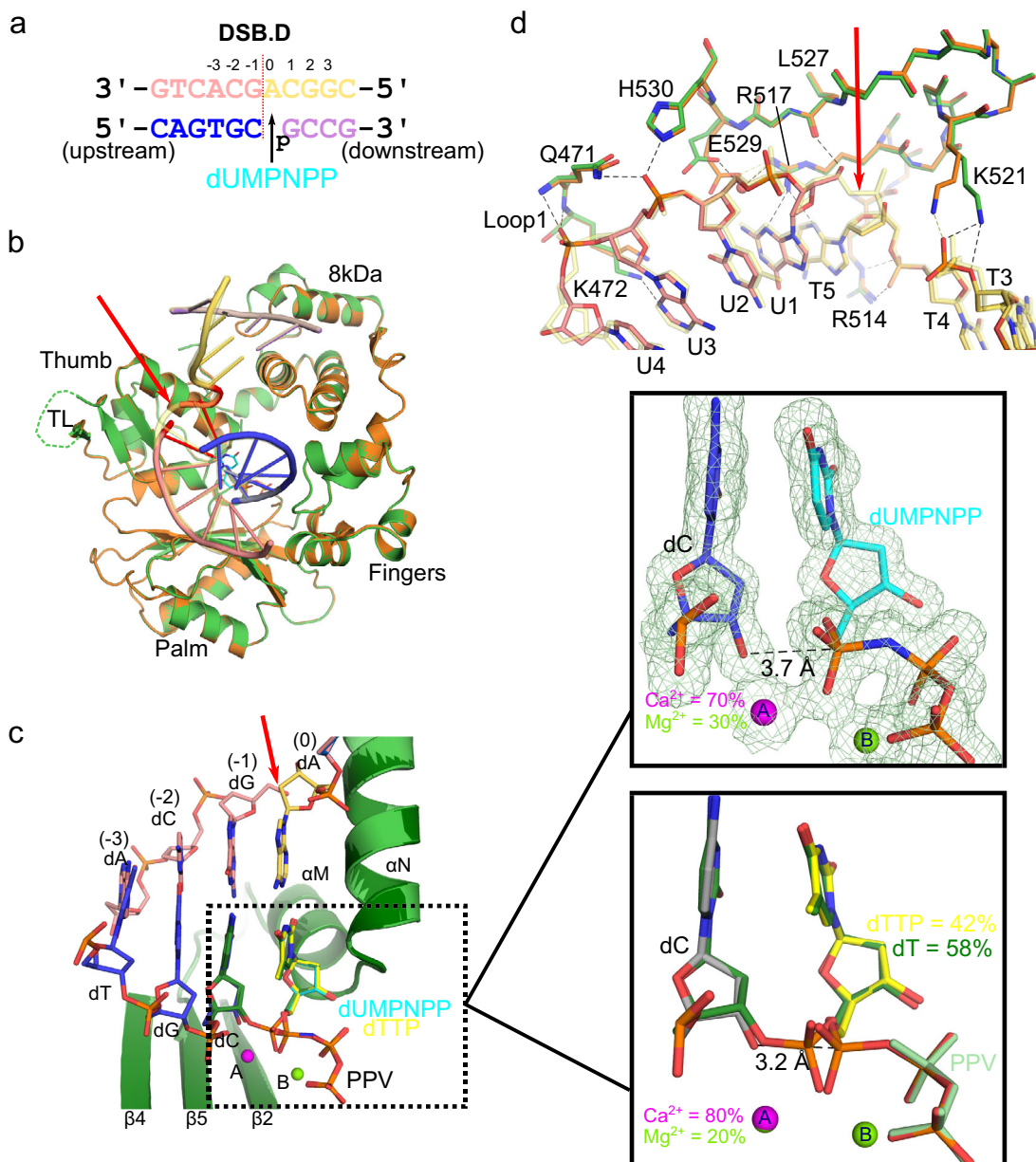


**Fig. 7 Pol $\lambda$  can accommodate mispairs proximal to the break site. a** DSB.C substrate, with template strand broken between the  $-1$  and  $-2$  positions, co-crystallized with the Pol $\lambda$  catalytic domain. This substrate also contains a G:T mispair (indicated by gray x) immediately upstream of the break site (red dashed line), at the  $-2$  position. Upstream (left) template (pink, chain U) and primer (blue, chain P) are annealed separately from the downstream (right) template (pale orange, chain T) and primer (lavender, chain D) strands, and synapsis is mediated by the polymerase. The DSB.C pre-catalytic quaternary complex is formed by the addition of the nonhydrolyzable incoming dUMPNPP (magenta) nucleotide. **b** Global superposition of the pre-catalytic SSB ternary (orange) and DSB.C quaternary (protein in cyan, DNA and incoming dUMPNPP colored as in **a**) complexes. **c** Structure of the mispaired (red asterisk) pre-catalytic DSB.C substrate bound in the Pol $\lambda$  active site. Mg $^{2+}$  ions are shown as green spheres. **d** Superposition of the pre-catalytic SSB (protein in orange, DNA template in transparent pale yellow) and mispaired DSB.C (protein in cyan, DNA colored as in **a**, water molecule as red sphere) complex for comparison of interactions surrounding the DSB break site (red arrow). Hydrogen bonding interactions with the mispaired DSB.C are drawn (black dashed lines). **e** Superposition of the base pair at the  $-2$  position in either the SSB (correct base pair, transparent yellow) or mispaired DSB (template dT:primer dG in pink and blue, respectively), with measured hydrogen bonding distances drawn as yellow (correct) or black (incorrect) dashed lines. Solid black arrow indicates subtle shift of the mispaired primer dG toward the minor groove.

Fig. 8c, upper inset). Interestingly, the electron density for this intermediate conformation is very clear and distinct, rather than exhibiting an elongated or diffuse appearance usually indicative of a mixture of axial and equatorial conformations. Exchange of the dUMPNPP for dTTP yielded a mixture of incorporated dTMP (58%) and unincorporated dTTP (42%) (Supplementary Fig. 1f). Similar to the pre-catalytic complex, a mixture of ions is again observed in the metal A site (80% and 20% occupancy for Ca $^{2+}$  to Mg $^{2+}$ , Fig. 8c, lower inset). Interestingly, the 3'-OH on the unincorporated primer terminus adopts the fully equatorial position, 3.2 Å from the  $\alpha$ -phosphate of the dTTP. The pyrophosphate leaving group (PPV) is still present and ordered in the active site.

The template strands in the DSB.D and SSB complexes superimpose well and therefore participate in largely similar interactions (Fig. 8d, Supplementary Table 2). Interactions common to both structures include a stacking interaction of the

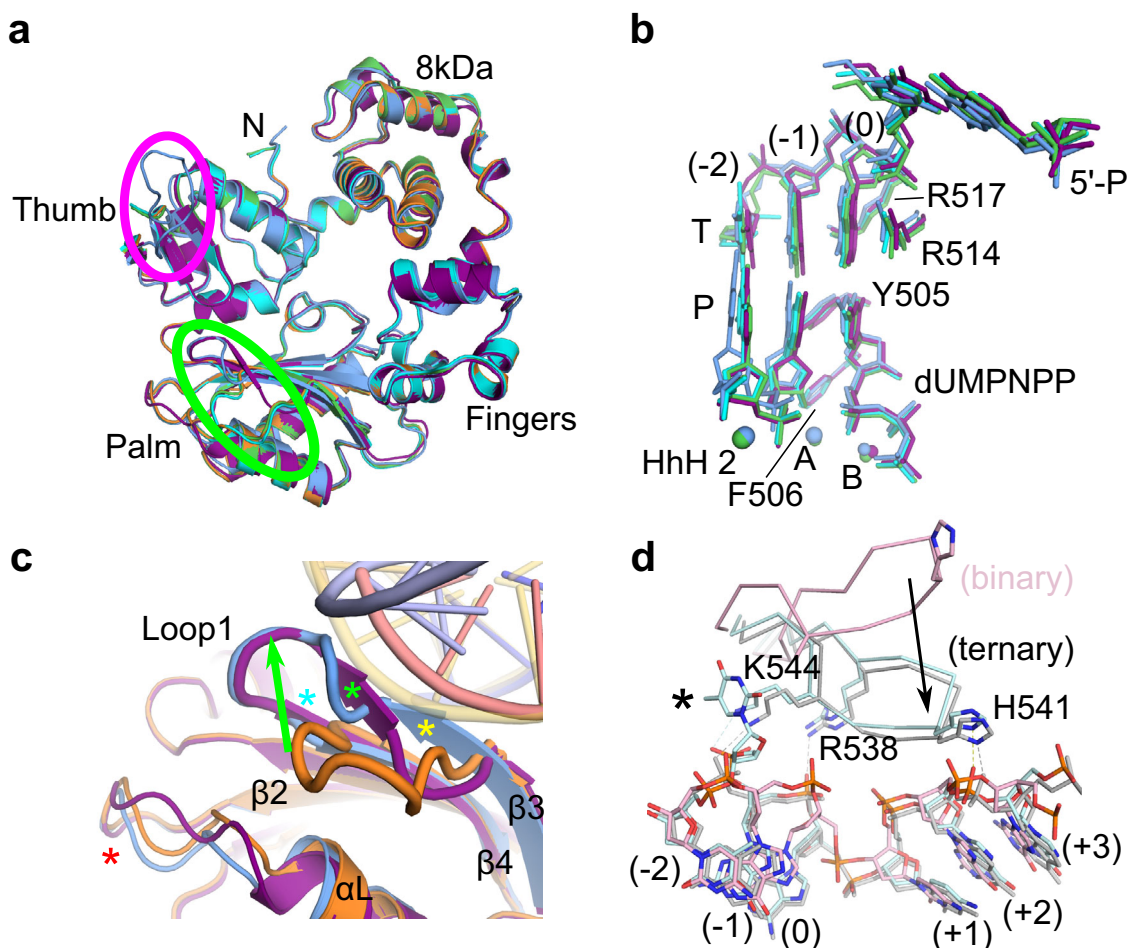
Arg514 sidechain with the templating base, as well as two putative hydrogen bonds between the sidechain and the nonbridging phosphate oxygens between nucleotides T4 and T5 (Supplementary Table 2). Arg517 extends along the minor groove and makes two sequence-nonspecific hydrogen bonds with the T5 and U1 bases, on either side of the break. The His530 and Gln471 sidechains each hydrogen bond with a nonbridging phosphate oxygen between upstream template U2 and U3. Backbone amides from Gln471 and Lys472 anchor the phosphate backbone between U3 and U4. There are a few subtle differences observed between the DSB.D and SSB complexes (Supplementary Table 2). In the DSB.D complex, the 5'-OH on the upstream template rotates outward toward solvent, rather than inward toward the protein, and makes a putative hydrogen bond with the backbone carbonyl of Leu527. The Glu529 sidechain makes a hydrogen bond with the Arg517 sidechain in the SSB complex but adopts a different conformation in the DSB.D complex, instead



**Fig. 8** Pol $\lambda$  can address blunt-ended DSBs and SSBs in a similar manner. **a** Blunt-ended DSB substrate (DSB.D) co-crystallized with the Pol $\lambda$  catalytic domain. The break site is indicated (red dashed line). Upstream (left) template and primer (pink and blue, respectively) are annealed separately from the downstream (right) template (pale orange) and primer (lavender) strands, and synapsis is mediated entirely by the polymerase. The DSB.D pre-catalytic quaternary complex is formed by addition of the nonhydrolyzable incoming dUMPNPP (cyan) nucleotide. **b** Global superposition of the pre-catalytic SSB ternary (orange) and blunt-ended DSB quaternary (protein in green, DNA and incoming dUMPNPP colored as in **a**) complexes. **c** Structure of the DSB.D substrate bound in the Pol $\lambda$  active site, with the pre- (protein in light green, DNA and incoming nucleotide colored as in **a**) and partially incorporated (protein and DNA in dark green, unincorporated hydrolyzable incoming dTTP nucleotide in yellow, PDB ID code 7MOA) substrates, with zoomed-in views of each complex included to the right (top, pre-catalytic, with  $2F_o-F_c$  electron density contoured to  $1\sigma$ ; bottom, partially incorporated, with unincorporated primer terminus and dTTP in gray and yellow, respectively). The  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  ions are shown as magenta and green spheres, respectively. Though the metal in the B site is modeled at full occupancy, based on geometry and coordination distances, a mixture of metals is observed in the A site. Occupancies for the metal A site mixture and newly incorporated nucleotide are indicated in each case. **d** Superposition of the pre-catalytic SSB (protein residues in orange, DNA template in transparent pale yellow) and blunt-ended DSB.D substrates (protein in green, DNA template strands colored as in **a**, water molecule as red sphere) for comparison of interactions surrounding the DSB break site (red arrow). Hydrogen bonding interactions with the DSB or the SSB are drawn as dashed lines (black or yellow, respectively).

utilizing a water molecule to bridge the interaction with Arg517 (as observed in the complex with DSB.C). Lysine residues 472 and 521 assume different conformations in each complex. In the SSB, Lys521 interacts with the nonbridging phosphate oxygen between T3 and T4. In the DSB.D complex, this lysine sidechain is shifted slightly farther downstream and could hydrogen bond with either

the T3-T4 nonbridging or bridging oxygens. Lys472 in the DSB.D lies within hydrogen-bonding distance of the U3 base but assumes a different conformation in the SSB-bound complex, where this interaction is not observed. Interestingly, the ‘thumb loop’ region is disordered in this structure, as it is in that of the SSB (Fig. 8b, green dashed line). It is possible that the ‘thumb



**Fig. 9 Loop1 and ‘thumb loop’ adopt varying conformations to cradle different substrate configurations.** Superpositions of the global protein (**a**) and active site components (**b**) of the pre-catalytic SSB (orange), complementary DSB.A (blue) and DSB.B (purple), mispaired DSB.C (cyan), and blunt-ended DSB.D (green) synaptic complexes. **c** The upstream template region of the DNA substrate adopts variable trajectories in different complexes, which are, in turn, cradled by varying Loop1 conformations (SSB, protein in orange, DNA in transparent yellow; DSB.A, protein in blue, DNA in pink; DSB.B, protein in purple, DNA in light blue). These conformations alter the length and configuration of  $\beta$ -strands 3 (yellow asterisk) and 4 (cyan asterisk) and intervening secondary structure (green asterisk), which likely influence the position of the loop between  $\beta$ -strand 2 and  $\alpha$ -helix L (red asterisk). **d** Structural superpositions of the Pol $\lambda$  ‘thumb loop’ observed in an ordered conformation in the canonical 1nt SSB binary (PDB ID code 1XSL<sup>11</sup>, pink) and 1nt SSB pre-catalytic ternary (PDB ID code 1XSN<sup>11</sup>, gray) complexes, relative to the 1nt strand-slippage intermediate (PDB ID code 2BCV<sup>25</sup>, light cyan, with extrahelical base marked with a black asterisk). Putative hydrogen bonding interactions are indicated by dashed lines in colors corresponding to the structure. Movement of the ‘thumb loop’ upon nucleotide binding is shown by a black arrow.

loop’ is required for interactions stabilizing the initial, tenuous steps of end-bridging, and is no longer required once a stable complex has been formed (consistent with the steady-state conditions within a crystal lattice).

### Discussion

Analysis of the SSB and DSB crystal structures presented in this study allows for a breakdown of the Pol $\lambda$  active center into its most critical components. Regardless of the substrate configuration, the protein subdomains, DNA substrates, and active site residues assume the canonical ‘active/closed’ conformation (Fig. 9a, b). In each case, the 5'-phosphate on the downstream duplex lies within the same pocket on the surface of the 8 kDa domain, likely aiding in binding and orientation of the downstream end of the gap. 5'-phosphate binding, juxtaposed with the 90° bend in the template phosphodiester backbone immediately upstream of the +1 position in the downstream duplex, properly positions the template base (0 position) in the nascent base pair binding site, stacked by Arg514. Arg517 extends into the minor groove, and Tyr505/Phe506 have shifted from their ‘inactive/

open’ to their ‘active/closed’ conformations observed upon nucleotide binding<sup>11</sup>. Coordination of the Na<sup>+</sup> ion in the HhH2 site and an intricate network of hydrogen bonding interactions anchor the upstream duplex to the palm subdomain through the primer strand. These interactions, aided by base stacking and pairing at the -1 position, correctly align the primer terminus for catalysis. The structural similarity of these components, despite varying end configurations and potential crystal packing influences, suggest that they are essential factors for synapsis by Pol $\lambda$ .

Detailed comparison of interactions between the Pol $\lambda$  catalytic domain and the various configurations of DSB substrates used in this study reveal that the hydrogen bonding networks downstream of the break are largely conserved (Figs. 5e, 6f, 7d, 8d and Supplementary Table 2), regardless of end configuration. However, interactions proximal to and upstream of the break site vary as the location of the template strand break moves farther from the nascent base pair binding site. Though previous studies have established that Pol $\lambda$  participates in repair of DSB substrates with a minimum of single nucleotide base pairing at the primer



terminus<sup>9,22</sup>, substrates with two or more base pair microhomology are also commonly used for *in vitro* and *in vivo* NHEJ assays<sup>10,19</sup>. Synaptic stability likely increases with increasing microhomology length, a supposition supported by the observation that Pol $\lambda$  generates more product using DSB substrates with 2nt rather than 1nt overlap (Fig. 2b, c). Comparison of the DSB.C crystal structure presented in this study with the recently reported complex of Pol $\mu$  bound to a similar end configuration (template strand discontinuity between the  $-1$  and  $-2$  positions, PDB ID code 6WIC<sup>17</sup>) shows a high degree of similarity between the two proteins in the vicinity of the template nucleotide opposite the primer terminus. A hydrogen bond exists between the protein and the base at that position in Pol $\lambda$  (Arg517 NH1—T6 N3) and a similar interaction is conserved in Pol $\mu$  (Arg445 NH1—T6 O2). Another hydrogen bond between Arg517 NH1 and the O4' of the ribose sugar is possible, which would be considered a long-distance interaction in Pol $\mu$ . Therefore, the means by which Pol $\lambda$  directly 'senses' the primer terminal base pair, and the structural underpinnings of its absolute requirement for this base pair, remains unclear.

Global superpositions of the pre-catalytic SSB and DSB synaptic complexes show that the Pol $\lambda$  catalytic domain adopts similar protein conformations, regardless of the DSB end configuration presented (Fig. 9a). The most notable structural differences are observed in Loop1 and the 'thumb loop'—alternations that could serve to stabilize a multitude of structurally diverse DSB end configurations (Fig. 9a, green and magenta circles, respectively). In the DSB.A complex, the upstream template strand is rotated slightly outward from the palm subdomain and there is a concomitant shift of Loop1. This shift lengthens both  $\beta$ -strands 3 and 4, which would create a steric clash with the template strand in the conventional location (Fig. 9c, yellow and cyan asterisks, respectively). In the DSB.B complex, Loop1 adopts a largely similar conformation to that observed in DSB.A, with the extension of  $\beta$ -strand 4. In DSB.B, however, there is a deviation from canonical geometry for  $\beta$ -strand 3 in the vicinity of Gln464–Glu465, which allows this region to interact with bases in the mispaired upstream duplex (Fig. 9c, green asterisk). Though the conformations of Loop1 in the DSB.A and DSB.B complexes are largely similar in trajectory, each would create a steric clash with the template strand of the other complex (Fig. 9c). The structures presented in the current study therefore provide structural evidence that Pol $\lambda$ 's Loop1 may play a similar stabilizing role for the upstream template region of DSB substrates, as has been hypothesized for the analogous loop in Pol $\mu$ <sup>16,17</sup>.

The 'thumb loop' is a structural feature unique to Pol $\lambda$ , compared to the other Family X polymerases, and is often disordered in crystal structures or displays substantial thermal motion when present. In the current study, the 'thumb loop' is ordered only in the DSB.A complex and makes contacts only with DNA template strands from other molecules in the asymmetric unit (Fig. 5f). In a similar fashion to that of Loop1, the 'thumb loop' has previously been shown to adopt different conformations in binary and pre-catalytic ternary SSB complexes (PDB ID codes 1XSL and 1XSN, respectively<sup>11</sup>), with the loop moving to cradle the template strand upstream and downstream of the bend in the DNA simultaneously (Fig. 9d). In the ordered conformation proximal to the DNA, residues in the loop (Arg538, His541, and Lys544) lie within hydrogen-bonding distance of the phosphate backbone. Additionally, crystal structures of Pol $\lambda$  in complex with a strand-slippage intermediate containing an extrahelical base between the  $-1$  and  $-2$  positions (PDB ID code 2BCV<sup>25</sup>) display a similar conformation of the 'thumb loop'. However, in this structure, Lys544 from the 'thumb loop' employs Van der Waals forces to interact with the planar face of the extruded base. Alanine

substitution of Lys544 markedly decreases activity on blunt-end DSB.D, with no deleterious effects on utilization of the complementary DSB with 2nt microhomology at the break site (Fig. 3e–h). In contrast, mutation of His541 has only minimal effect on usage of the blunt-end substrate. These results suggest that the limiting step of blunt-end DSB synapsis may lie with the difficulty of correctly positioning the upstream duplex and is strongly dependent on the presence of 'thumb loop' interactions for assembly.

In-depth structural comparisons of the variety of Pol $\lambda$  synaptic complexes, such as the ones presented in this study, provide invaluable insight into how this polymerase engages its DSB substrates and mediates end bridging in preparation for gap-filling during NHEJ. Such structural information presents a firm foundation for future understanding of DSB synapsis within the broader context of the full NHEJ complex.

## Methods

**Expression and purification of Pol $\lambda$  proteins.** Full-length (Met1–Trp575) or catalytic domain constructs (Val235–Trp575) of human DNA Pol $\lambda$  were cloned into the NotI/EcoRI restriction sites of the pGEXM vector<sup>12</sup>. Constructs were transformed into Rosetta2 (DE3) cells and expressed in large-scale in LB medium supplemented with 100  $\mu$ g/mL ampicillin and 35  $\mu$ g/mL chloramphenicol. Once the cultures reached an OD<sub>600nm</sub> of 0.8, the temperature was reduced to 18 °C for 30 min, when protein expression was induced by addition of isopropyl- $\beta$ -D-thiogalactoside (IPTG) to a final concentration of 0.4 mM, and continued overnight at 18 °C. The cells were pelleted by centrifugation and lysed by sonication in 25 mM Tris pH 8, 750 mM NaCl, 5% glycerol, 1 mM DTT. Polyethyleneimine was added to a final concentration of 0.1% (v/v), dropwise, while stirring at 4 °C for 5 min. The lysate was subsequently clarified by centrifugation, bound in-batch to glutathione sepharose 4B resin (Cytiva), and subjected to on-resin TEV protease cleavage in 25 mM Tris pH 8, 500 mM NaCl, 5% glycerol, 1 mM DTT overnight at 4 °C. Pol $\lambda$  proteins were further purified by size exclusion chromatography, dialyzed to storage buffer containing 25 mM Tris pH 8, 100 mM NaCl, 5% glycerol, 1 mM DTT and concentrated to 10–15 mg/mL. Concentrated proteins were flash frozen in liquid nitrogen, and stored at  $-80$  °C.

**Generation of full-length Pol $\lambda$  mutants.** Substitution mutants and deletion variants ( $\Delta$ L1 with Loop1 residues Ser463–Gln471 deleted and replaced with KGET loop sequence from Pol  $\beta$ <sup>13</sup>;  $\Delta$ TL, deletion of thumb loop residues Val537–Val545 and insertion of a single glycine residue to bridge the gap) were generated in the full-length human Pol $\lambda$  using QuikChange mutagenesis (Agilent). All mutant clones were fully sequenced by Genewiz. Wildtype, mutant or variant proteins were expressed from Rosetta2 (DE3) cells in Terrific broth (supplemented with 100  $\mu$ g/mL ampicillin and 35  $\mu$ g/mL chloramphenicol) to increase yield in small scale. All mutants behaved similarly to wild-type protein during size exclusion chromatography and are therefore likely to be properly folded. The final storage buffer contained 25 mM Tris pH 8, 200 mM NaCl, 5% (v/v) glycerol, and 1 mM DTT.

**Analysis of *in vitro* DSB gap-filling by Pol $\lambda$ .** The final reaction mixture (30  $\mu$ l) contained 50 mM Tris pH, 7.5–8, 1 mM DTT, 4% (v/v) glycerol, 0.1 mg/ml BSA, 10% (w/v) polyethylene glycol 8000, 2.5 mM MgCl<sub>2</sub>, 200 nM DSB DNA substrate, consisting either of the upstream and downstream oligonucleotide duplexes (DSB.A, DSB.C, or DSB.D substrates, Fig. 2a) or an upstream single-strand primer oligonucleotide and the downstream duplex (DSB.B substrate, Fig. 2a) and full-length Pol $\lambda$  at 600 nM. Individual components of the reaction mix were added incrementally. First, Pol $\lambda$  was added to the annealed downstream duplex. Following a 5 min incubation at 37 °C, to allow the enzyme to bind the 5'-phosphorylated terminus of the downstream duplex, the annealed upstream portion of the DNA substrate was added, and the mixture was incubated for another 5 min. After this incubation, the reaction was initiated by the addition of dTTP to a final concentration of 5  $\mu$ M. Ten microliter aliquots of the mixture were removed at different time points (as specified in Fig. 2b) and quenched with an equal volume of 99% formamide (v/v), 10 mM EDTA, and 0.1% (w/v) bromophenol blue. The products were resolved on a 16% (v/v) denaturing polyacrylamide gel, imaged using a Typhoon Imager (GE Healthcare), quantitated with ImageQuant TL, and graphed using GraphPad Prism (v. 9).

**Cell Lines.** WT (C57BL/6) or *Polm*<sup>-/-</sup> *Poll*<sup>-/-</sup> double knock-out murine fibroblast (MEF) cells (generously provided by Dr. L. Blanco) were derived from E14.5d embryos and immortalized by the introduction of SV40 large T-antigen as described<sup>10</sup>. MEF cells were maintained in DMEM supplemented with 10% (v/v) fetal bovine serum (FBS, Sigma), 100U/ml penicillin, 5 mM N-acetyl-L-cysteine (Sigma) at 37 °C and 5% CO<sub>2</sub>. These lines were confirmed by qPCR to be free of mycoplasma contamination.



**Cell-based extrachromosomal NHEJ repair assay.** Extrachromosomal DNA substrates were generated by annealing DNA strands (Ultrasomers, Integrated DNA Technologies) described in Supplementary Table 3 and electroporated with 600 ng of pMAX-GFP plasmid (Lonza) into 200,000 cells in a 10  $\mu$ L volume with one 1350 V 30 ms pulse (Neon, Invitrogen) as described previously<sup>10</sup>, except that in the current study, two substrates were introduced in each electroporation. One substrate had partially complementary (Fig. 3e) or non-complementary end structures (Fig. 3g), thus engaging polymerase activity during NHEJ. A second ‘spike-in’ control substrate containing fully complementary 5'GC overhang ends (repaired by NHEJ independently of polymerase activity) was introduced in all experiments at 1/20<sup>th</sup> the concentration of the polymerase-engaging substrates described above and was used to confirm consistent electroporation. Polymerase-deficient cells were complemented by the addition of 1  $\mu$ g of the noted purified polymerases to the substrate transfection solution just before electroporation. Following transfection, cells were incubated with Benzonase mix (HBSS, 0.5 M MgCl<sub>2</sub>, Benzonase) at 37 °C for 30 min. Cellular repair products were harvested using a QIAamp DNA mini kit (Qiagen). Each electroporation was reproduced in triplicate from three independent cell electroporations. Head-to-tail NHEJ repair products were amplified using 26 cycles using the primers 5'-TAAGCGATGCTCTCACCG-3' and 5'-GATGGGTGTGAGAGTGAAGATC-3', digested with restriction endonucleases specific for products of polymerase-dependent repair, as noted in Fig. 3f, h, and the fraction of polymerase-dependent repair determined by electrophoresis on a 10% (w/v) non-denaturing polyacrylamide gel, visualized using a Typhoon Imager (GE Healthcare). The intensities of digested and undigested bands were quantified using ImageQuant (v. 8.1), and resulting data were graphed and statistically analyzed using GraphPad Prism (v. 9).

**Crystallization of the 1nt-gapped SSB ternary complex.** Synthesized DNA oligonucleotides (Integrated DNA Technologies) were annealed to generate a 1nt-gapped SSB substrate (Fig. 4b): unbroken template (5'-CGGCAGTACTG-3'), upstream primer (5'-CAGTAC)-3' and 5'-phosphorylated downstream primer (5'-pGCCG-3'). Oligonucleotides were mixed in an equimolar ratio in 100 mM Tris pH 7.5, 40 mM MgCl<sub>2</sub> and annealed in a thermal cycler, with denaturation at 94 °C, followed by a slow temperature gradient from 90 °C to 4 °C. The annealed DNA was then mixed in a 3:1 molar ratio with concentrated PolI catalytic domain (Val235-Trp575, 14.47 mg/mL). The complex was incubated on ice at 4 °C for 1 hr, followed by the addition of the nonhydrolyzable incoming 2'-deoxyuridine-5'-( $\alpha,\beta$ -imido)triphosphate (dUMPNPP) nucleotide (0.91 mM final concentration). After another 1 h incubation at 4 °C, ternary complex crystals were grown at room temperature using the sitting drop vapor diffusion technique<sup>26</sup>. 300 nL of the protein/DNA complex were mixed with 300 nL of mother liquor (37.5 mM Na cacodylate pH 6.5, 150 mM KCl, 75 mM magnesium acetate, 7.5% (w/v) PEG 8000). The crystals were transferred from the drop to a solution containing 50 mM Na cacodylate pH 6.5, 200 mM KCl, 100 mM magnesium acetate, 50 mM NaCl, 10% (w/v) PEG 8000, and 1 mM dUMPNPP, then to a cryoprotectant (50 mM Na cacodylate pH 6.5, 200 mM KCl, 100 mM magnesium acetate, 50 mM NaCl, 20% (w/v) PEG 8000, 15% (v/v) ethylene glycol) in two steps. Crystals were flash frozen in liquid nitrogen.

**Crystallization of complementary DSB.A quaternary complex.** DNA oligonucleotides were used to generate the complementary DSB substrate (Fig. 5a): upstream template (5'-ACTG-3'), upstream primer (5'-CAGTGC)-3', downstream template (5'-CGGCAGC-3') and 5'-phosphorylated downstream primer (5'-pGCCG-3'), where the template strand break lies between the -2 and the -3 positions. Upstream and downstream DNA mixtures were separately annealed in a thermal cycler, as for the 1nt-gapped SSB substrate. The annealed DNA was then serially mixed in a 2.3:1 molar ratio with concentrated PolI catalytic domain (Val235-Trp575, 13.7 mg/mL)—first the downstream DNA, followed by the upstream DNA, and finally the incoming nonhydrolyzable dUMPNPP nucleotides (0.91 mM final concentration). The complex was incubated on ice at room temperature for 1 h, after each addition. Crystals of the quaternary complex were grown at 4 °C using the sitting drop vapor diffusion technique<sup>26</sup>. 400 nL of the protein/DNA complex were mixed with 200 nL of mother liquor (0.192 M tri-potassium citrate, 19.2% (w/v) PEG 3350). The growth drops were opened, and 500 nL of 0.188 M tri-potassium citrate, 18.8% (w/v) PEG 3350, 10 mM MgCl<sub>2</sub>, 0.5 mM dUMPNPP were added, followed by four subsequent 500 nL additions of cryoprotectant solution containing 20 mM Tris pH 8, 50 mM NaCl, 0.2 M tri-potassium citrate, 20% (w/v) PEG 3350, 25% (v/v) ethylene glycol, 10 mM MgCl<sub>2</sub>, 0.5 mM dUMPNPP, with 30–60 sec incubation between additions. Crystals were looped directly from this drop and flash frozen in liquid nitrogen.

**Crystallization of pre-catalytic DSB.B synaptic complex.** DNA oligonucleotides were used to generate the DSB.B substrate (Fig. 6a): upstream primer (5'-CAGTGC)-3', downstream template (5'-CGGCAGC-3') and 5'-phosphorylated downstream primer (5'-pGCCG-3'). Upstream and downstream DNA mixtures were separately annealed in a thermal cycler, as for the 1nt-gapped SSB substrate. The annealed DNA was then serially mixed in a 2.3:1 molar ratio with concentrated PolI catalytic domain (Pro233-Trp575, 13.7 mg/mL)—first the downstream DNA duplex, followed by the upstream primer, and finally the incoming nonhydrolyzable dUMPNPP nucleotide (0.91 mM final concentration). The complex was incubated on ice at 4 °C for 1 h, after each addition. Crystals of the quaternary

complex were grown at room temperature using the sitting drop vapor diffusion technique<sup>26</sup>. 300 nL of the protein/DNA complex were mixed with 300 nL of mother liquor (0.2 M sodium sulfate, 20% (w/v) PEG 3350). The growth drops were opened and 500 nL of mother liquor was added, followed by three subsequent 500 nL additions of cryoprotectant solution containing 0.138 M sodium sulfate, 13.8% (w/v) PEG 3350, 20 mM MgCl<sub>2</sub>, 25% (v/v) ethylene glycol, 0.5 mM dUMPNPP. The crystals were then transferred from the drop directly to the cryoprotectant solution, then flash frozen in liquid nitrogen.

**Crystallization of mismatched DSB.C quaternary complex.** DNA oligonucleotides were used to generate the DSB substrate containing a mispair upstream of the primer terminus (Fig. 7a): upstream template (5'-TACTG-3'), upstream primer (5'-CAGTGC)-3', downstream template (5'-CGGCAG-3') and 5'-phosphorylated downstream primer (5'-pGCCG-3'), where the template strand break lies between the -1 and -2 positions, and mispaired bases at the -2 position are indicated in bold. Upstream and downstream DNA mixtures were separately annealed in a thermal cycler, as for the 1nt-gapped SSB substrate. The annealed DNA was then serially mixed in a 2.3:1 molar ratio with concentrated PolI catalytic domain (Val235-Trp575, 14.47 mg/mL)—first the downstream DNA, followed by the upstream DNA, and finally the incoming nonhydrolyzable dUMPNPP nucleotides (0.91 mM final concentration). The complex was incubated on ice at 4 °C for 1 hr, after each addition. Crystals of the quaternary complex were grown at 4 °C using the sitting drop vapor diffusion technique<sup>26</sup>. 400 nL of the protein/DNA complex were mixed with 200 nL of mother liquor (76.5 mM Na cacodylate pH 6.5, 0.153 M ammonium sulfate, 22.95% (w/v), PEG 8 K, 13.5% (v/v) glycerol, 0.2 M NDSB-201). The crystals were transferred from the drop to a cryoprotectant solution containing 76.5 mM Na cacodylate pH 6.5, 0.153 M ammonium sulfate, 22.95% (w/v), PEG 8 K, 13.5% (v/v) glycerol, 20 mM MgCl<sub>2</sub>, 80 mM NDSB-201, and 0.5 mM dUMPNPP. Crystals were flash frozen in liquid nitrogen.

**Crystallization of blunt-ended DSB.D quaternary complex.** DNA oligonucleotides were used to generate the blunt-end DSB substrate (Fig. 8a): upstream template (5'-GCACTG-3'), upstream primer (5'-CAGTGC)-3', downstream template (5'-CGGCA-3') and 5'-phosphorylated downstream primer (5'-pGCCG-3'). Upstream and downstream DNA mixtures were separately annealed in a thermal cycler, as for the 1nt-gapped SSB substrate. The annealed DNA was then serially mixed in a 2.3:1 molar ratio with concentrated PolI catalytic domain (Val235-Trp575, 14.47 mg/mL)—first the downstream DNA, followed by the upstream DNA, and finally the incoming nonhydrolyzable dUMPNPP nucleotide (0.91 mM final concentration). The complex was incubated on ice at 4 °C for 1 hr, after each addition. Crystals of the quaternary complex were grown at 4 °C using the sitting drop vapor diffusion technique<sup>26</sup>. 500 nL of the protein/DNA complex were mixed with 100 nL of mother liquor (80 mM Na cacodylate pH 6.5, 0.16 M calcium acetate, 14.4% (w/v) PEG 8 K, 20% (v/v) glycerol). The crystals were transferred from the drop to a cryoprotectant solution containing 76 mM Na cacodylate pH 6.5, 0.152 M magnesium acetate, 15.2% (w/v), PEG 8 K, 19% (v/v) glycerol, and 0.5 mM dUMPNPP and soaked for 27 h at 4 °C. After soaking for 3 h, crystals were transferred to cryoprotectant solution containing 76 mM Na cacodylate pH 6.5, 0.152 M magnesium acetate, 15.2% (w/v), PEG 8 K, 19% (v/v) glycerol, and 5 mM dTTP and soaked for 24 h at 4 °C, yielding a mixture of incorporated and unincorporated dTTP within the crystal. Crystals were flash frozen in liquid nitrogen.

**Data collection and structural refinement.** Crystals were placed into a stream of nitrogen gas cooled to -180 °C for data collection. Data were collected on the Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source (APS) at Argonne National Laboratory. The data were integrated, and scaled using HKL2000<sup>27</sup> (v. 719.2). The crystal structure of the previously reported pre-catalytic 1nt-gapped SSB ternary complex (PDB ID code 2PFO<sup>18</sup>) was used as the search model for the molecular replacement of the current SSB complexes in Phaser<sup>28</sup> (v. 2.8.3). The structure of the current pre-catalytic ternary SSB complex (PDB ID code 7M07) was used as the molecular replacement model for all other complexes. In order to decrease potential model bias, the same  $R_{free}$  test set was used for all structures with this space group and unit cell dimensions. All structures were refined by iterative cycles of manual model building and refinement in COOT<sup>29,30</sup> (v. 0.8.9.2) and Phenix<sup>31</sup> (v. 1.19\_4092). TLS (Translation/Libration/Screw) vibrational motion refinement<sup>32</sup> was used for all structures. Data collection and refinement statistics are listed in Supplementary Table 1. Ramachandran statistics were generated using MolProbity<sup>33</sup>. All superpositions and structural figures were generated using PyMOL (v. 2.2.0).

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The data that support this study are available from the corresponding author upon reasonable request. Atomic coordinates and structure factors have been deposited in the Protein Data Bank (<https://www.rcsb.org/>) with ID codes 7M07, 7M09, 7M0A, 7M0B, 7M0D, 7M0E. Source data are provided with this paper.

Received: 8 November 2021; Accepted: 10 June 2022;

Published online: 01 July 2022

## References

- von Sonntag, C. *Free-Radical-Induced DNA Damage and its Repair* (Springer-Verlag, 2006).
- Jackson, S. P. & Bartek, J. The DNA-damage response in human biology and disease. *Nature* **461**, 1071–1078 (2009).
- Takata, M. et al. Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *EMBO J.* **17**, 5497–5508 (1998).
- Schatz, D. G. & Swanson, P. C. V(D)J recombination: mechanisms of initiation. *Annu. Rev. Genet.* **45**, 167–202 (2011).
- Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* **79**, 181–211 (2010).
- Bertocci, B., De Smet, A., Berek, C., Weill, J. C. & Reynaud, C. A. Immunoglobulin kappa light chain gene rearrangement is impaired in mice deficient for DNA polymerase mu. *Immunity* **19**, 203–211 (2003).
- Bertocci, B., De Smet, A., Weill, J. C. & Reynaud, C. A. Nonoverlapping functions of DNA polymerases mu, lambda, and terminal deoxynucleotidyltransferase during immunoglobulin V(D)J recombination in vivo. *Immunity* **25**, 31–41 (2006).
- Gilfillan, S., Benoist, C. & Mathis, D. Mice lacking terminal deoxynucleotidyl transferase: adult mice with a fetal antigen receptor repertoire. *Immunol. Rev.* **148**, 201–219 (1995).
- Nick McElhinny, S. A. et al. A gradient of template dependence defines distinct biological roles for family X polymerases in nonhomologous end joining. *Mol. Cell* **19**, 357–366 (2005).
- Pryor, J. M. et al. Essential role for polymerase specialization in cellular nonhomologous end joining. *Proc. Natl Acad. Sci. USA* **112**, E4537–E4545 (2015).
- García-Díaz, M., Bebenek, K., Krahn, J. M., Kunkel, T. A. & Pedersen, L. C. A closed conformation for the Pol lambda catalytic cycle. *Nat. Struct. Mol. Biol.* **12**, 97–98 (2005).
- Moon, A. F. et al. Sustained active site rigidity during synthesis by human DNA polymerase mu. *Nat. Struct. Mol. Biol.* **21**, 253–260 (2014).
- Bebenek, K., García-Díaz, M., Zhou, R. Z., Povirk, L. F. & Kunkel, T. A. Loop 1 modulates the fidelity of DNA polymerase lambda. *Nucleic Acids Res.* **38**, 5419–5431 (2010).
- Gosavi, R. A., Moon, A. F., Kunkel, T. A., Pedersen, L. C. & Bebenek, K. The catalytic cycle for ribonucleotide incorporation by human DNA Pol lambda. *Nucleic Acids Res.* **40**, 7518–7527 (2012).
- Delarue, M. et al. Crystal structures of a template-independent DNA polymerase: murine terminal deoxynucleotidyltransferase. *EMBO J.* **21**, 427–439 (2002).
- Juarez, R., Ruiz, J. F., Nick McElhinny, S. A., Ramsden, D. & Blanco, L. A specific loop in human DNA polymerase mu allows switching between creative and DNA-instructed synthesis. *Nucleic Acids Res.* **34**, 4572–4582 (2006).
- Kaminski, A. M. et al. Structural snapshots of human DNA polymerase mu engaged on a DNA double-strand break. *Nat. Commun.* **11**, 4784 (2020).
- García-Díaz, M., Bebenek, K., Krahn, J. M., Pedersen, L. C. & Kunkel, T. A. Role of the catalytic metal during polymerization by DNA polymerase lambda. *DNA Repair (Amst.)* **6**, 1333–1340 (2007).
- Martin, M. J. et al. A specific N-terminal extension of the 8 kDa domain is required for DNA end-bridging by human Polmu and Pollambda. *Nucleic Acids Res.* **41**, 9105–9116 (2013).
- García-Díaz, M. et al. A structural solution for the DNA polymerase lambda-dependent repair of DNA gaps with minimal homology. *Mol. Cell* **13**, 561–572 (2004).
- García-Díaz, M. et al. Template strand scrunching during DNA gap repair synthesis by human polymerase lambda. *Nat. Struct. Mol. Biol.* **16**, 967–972 (2009).
- Martin, M. J., Juárez, R. & Blanco, L. DNA-binding determinants promoting NHEJ by human Polmu. *Nucleic Acids Res.* **40**, 11389–11403 (2012).
- Burak, M. J., Guja, K. E., Hambardeeva, E., Derkunt, B. & García-Díaz, M. A fidelity mechanism in DNA polymerase lambda promotes error-free bypass of 8-oxo-dG. *EMBO J.* **35**, 2045–2059 (2016).
- Wang, W., Hellinga, H. W. & Beese, L. S. Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis. *Proc. Natl Acad. Sci. USA* **108**, 17644–17648 (2011).
- García-Díaz, M., Bebenek, K., Krahn, J. M., Pedersen, L. C. & Kunkel, T. A. Structural analysis of strand misalignment during DNA synthesis by a human DNA polymerase. *Cell* **124**, 331–342 (2006).
- Chayen, N. E. Comparative studies of protein crystallization by vapour-diffusion and microbatch techniques. *Acta Crystallogr. D: Biol. Crystallogr.* **54**, 8–15 (1998).
- Otwinowski, Z. M. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Zwart, P. H. et al. Automated structure solution with the PHENIX suite. *Methods Mol. Biol.* **426**, 419–435 (2008).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D: Biol. Crystallogr.* **60**, 2126–2132 (2004).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D: Biol. Crystallogr.* **66**, 486–501 (2010).
- Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D: Biol. Crystallogr.* **66**, 213–221 (2010).
- Urzhumtsev, A., Afonine, P. V. & Adams, P. D. TLS from fundamentals to practice. *Crystallogr. Rev.* **19**, 230–270 (2013).
- Lovell, S. C. et al. Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Proteins* **50**, 437–450 (2003).

## Acknowledgements

The authors thank Dr. J. Krahn for nucleotide parameter optimization, and D. Appel, Drs. A. Riccio and R.S. Williams for critical reading of the manuscript. This work was supported in part by Division of Intramural Research of the National Institute of Environmental Health Sciences, National Institutes of Health Grants IZIA ES102645 (to L.C.P.), U01 CA097096 (to D.A.R.) and Z01 ES065070 (to T.A.K.). Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences [W-31-109-Eng-38].

## Author contributions

A.M.K, K.K.C., and K.B. designed and performed research; A.M.K, K.K.C., D.A.R., T.A.K., L.C.P., and K.B. contributed to data analysis and manuscript preparation.

## Funding

Open Access funding provided by the National Institutes of Health (NIH).

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-31278-4>.

**Correspondence** and requests for materials should be addressed to Katarzyna Bebenek.

**Peer review information** *Nature Communications* thanks Luis Blanco and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2022