

# Human DNA polymerase θ grasps the primer terminus to mediate DNA repair

Karl E Zahn<sup>1</sup>, April M Averill<sup>1</sup>, Pierre Aller<sup>2</sup>, Richard D Wood<sup>3</sup> & Sylvie Doublie<sup>1</sup>

**DNA polymerase θ protects against genomic instability via an alternative end-joining repair pathway for DNA double-strand breaks. Polymerase θ is overexpressed in breast, lung and oral cancers, and reduction of its activity in mammalian cells increases sensitivity to double-strand break-inducing agents, including ionizing radiation. Reported here are crystal structures of the C-terminal polymerase domain from human polymerase θ, illustrating two potential modes of dimerization. One structure depicts insertion of ddATP opposite an abasic-site analog during translesion DNA synthesis. The second structure describes a cognate ddGTP complex. Polymerase θ uses a specialized thumb subdomain to establish unique upstream contacts to the primer DNA strand, including an interaction with the 3'-terminal phosphate from one of five distinctive insertion loops. These observations demonstrate how polymerase θ grasps the primer to bypass DNA lesions or extend poorly annealed DNA termini to mediate end-joining.**

Human polymerase (pol) θ is a multidomain protein of 2,590 amino acids (aa), with homologs throughout multicellular organisms<sup>1–3</sup>. The human *POLQ* gene encodes the pol θ protein, which contains conserved superfamily 2 helicase (SF2) and family-A DNA polymerase domains at the N and C termini, respectively, which are linked by a large central region<sup>4</sup>. In vertebrates, pol θ and pol ν (ref. 5) are the only family-A polymerases present in cell nuclei. The mitochondrial pol γ is distantly related to pol θ, and 5'-deoxyribose phosphate lyase activity has been attributed to both pol θ and pol γ (refs. 6,7). Pol θ, however, does not synthesize DNA with the accuracy or processivity of the mitochondrial replicase<sup>8,9</sup>. Several conserved insertion loops, absent from bacterial homologs, intervene within the family A-polymerase fold of pol θ. Deletion of residues 2264–2315, which include much of the penultimate insertion loop, has been shown to abrogate pol θ's ability to extend unannealed single-stranded oligonucleotides<sup>10</sup> and bypass abasic (AP) sites or sites of thymine glycol (Tg) damage<sup>11</sup>.

Early biochemical characterization of pol θ hinted at a role in translesion DNA synthesis (TLS), given pol θ's lesion-bypass activity<sup>12,13</sup>, and the enzyme indeed appears to be involved in TLS of oxidative DNA damage *in vivo*<sup>14</sup>. Assays in chicken DT40 cells deduced overlapping roles for pol θ and pol β in base excision repair<sup>15</sup>. Pol θ can also substitute for pol β during base excision repair in *Caenorhabditis elegans*<sup>16</sup>. It has been suggested that pol θ is involved in somatic hypermutation of immunoglobulin genes, but any role appears to be minimal<sup>17</sup>. Pol θ does, however, participate in a subset of immunoglobulin-gene class switch-recombination events in mouse B cells, in a manner dependent on an alternative (Ku-independent) DNA end-joining pathway<sup>18</sup>. Pol θ contributes to alternative end-joining

of double-strand breaks (DSBs) in *Drosophila melanogaster*<sup>19</sup> and in mammalian cells<sup>18</sup>. In *C. elegans*, pol θ limits extensive deletions at DNA replication-fork barriers but generates small indels templated by DNA adjacent to the excision site<sup>20,21</sup>. Pol θ also produces templated indels in *Drosophila*<sup>22,23</sup> and mice<sup>18</sup>. Furthermore, pol θ interacts with the origin recognition complex in human cells during G1 and may affect DNA replication timing<sup>24</sup>.

*POLQ* encodes the only nuclear DNA polymerase overexpressed in breast cancer, and higher expression correlates with unsuccessful clinical treatment<sup>25,26</sup>. Elevated *POLQ* expression also occurs in oral squamous-cell carcinomas<sup>27</sup>; however, it is also associated with disease-free survival in patients with lung cancers<sup>28</sup>. Knockout or knockdown of pol θ in mouse<sup>18,29</sup> and human<sup>30</sup> cells increases sensitivity to DSB-inducing agents, including ionizing radiation. Complementation of this sensitivity phenotype requires the polymerase activity of pol θ but not the ATPase activity of the helicase domain<sup>18</sup>. The presence of distinctive insertion loops in pol θ has encouraged speculation that these loops bestow unique properties upon the enzyme. However, the molecular mechanisms that confer pol θ's ability to prime DNA synthesis from nonoptimal base-pairing in the context of alternative end-joining have remained elusive until now. We set out to determine the crystal structures of two active constructs comprising the entire C-terminal polymerase domain and its associated inactivated exonuclease-like subdomain, which are reported here. The structural work, in conjunction with a biochemical analysis of point variants, illuminates specific interactions between the specialized thumb subdomain of pol θ and the primer terminus that are necessary for TLS and contribute to elongation of single-stranded DNA oligonucleotides.

<sup>1</sup>Department of Microbiology and Molecular Genetics, University of Vermont, Burlington, Vermont, USA. <sup>2</sup>Diamond Light Source, Didcot, Oxfordshire, UK.

<sup>3</sup>Department of Epigenetics & Molecular Carcinogenesis, University of Texas MD Anderson Cancer Center, Smithville, Texas, USA. Correspondence should be addressed to S.D. (sdoublie@uvm.edu).

Received 12 December 2014; accepted 13 February 2015; published online 16 March 2015; doi:10.1038/nsmb.2993

## RESULTS

## Human pol θ inserts ddATP opposite an AP site

To identify structural components providing for the unique enzymatic activities of pol θ, we determined the crystal structures of two DNA polymerase-domain constructs (**Table 1**). Molecular replacement<sup>31</sup>, with a ternary complex of Taq DNA polymerase (PDB 1QSY)<sup>32</sup> as the search model, allowed us to place the four similar molecules in the crystal asymmetric unit (ASU) of the tetrahydrofuran (THF)-ddATP complex, which captured human pol θ inserting ddATP opposite THF, a stable AP-site analog. The structure revealed an overall fold reminiscent of bacterial homologs—with exonuclease, thumb and fingers subdomains oriented about a right-hand palm subdomain (**Supplementary Table 1**)—but disrupted by five unique insertion loops (**Fig. 1a–c** and **Supplementary Table 2**). The THF-ddATP complex clearly showed the nascent base pair in the polymerase active site: the strictly conserved catalytic aspartate and glutamate residues (D2330, D2540 and E2541) of the palm subdomain coordinate a divalent  $\text{Ca}^{2+}$  ion associated with the triphosphate tail of the ddATP nucleotide (**Fig. 1d** and **Supplementary Fig. 1**). As a known inhibitor of DNA polymerases,  $\text{Ca}^{2+}$  was essential for trapping the closed complex, because the primer strand retains a 3'-hydroxyl moiety for nucleophilic attack (**Supplementary Fig. 2**). The highly conserved lysine (K2383) and arginine (R2379) residues of the fingers-subdomain O helix contact nonbridging oxygens of the  $\alpha$ - and  $\gamma$ -phosphates, respectively (**Supplementary Fig. 2**), as seen in other closed ternary complexes of family-A polymerases<sup>33</sup>. The conserved O-helix residue

Y2391 was fully displaced from its template-occluding position in open or ‘ajar’ structures<sup>34</sup> (**Fig. 1d** and **Supplementary Fig. 1**), and thus, to our knowledge, the pol θ structure represents the first fully closed model of any family-A DNA polymerase inserting adenine opposite a nontemplating DNA lesion. Previously, trapping ternary complexes destabilized by templating THF has required the purine analog 5-nitro-1-indolyl-2'-deoxyribose-5'-triphosphate (5-NITP) because of its enhanced capacity for base-stacking<sup>35,36</sup>. The 5-NITP and ddATP double rings show substantial overlap when the current pol θ model is superimposed onto these previous structures<sup>37</sup>.

Pol θ and pol v both contain unique signature residues at mutable O-helix sites<sup>38</sup>, adjacent to the essential arginine and lysine side chains that contact the triphosphate tail and make pyrophosphate a better leaving group<sup>39</sup>. A glutamine residue occupies position 2384 of pol θ (Q2384), where pol v has a conserved lysine (K679), and bacterial polymerases generally contain an alanine or threonine. Q2384 appeared in the vicinity of the Hoogsteen face of the incoming nucleotide, within hydrogen-bonding distance to the exocyclic N6 of the ddATP molecule (**Fig. 1d**). Mutation of the analogous residue to alanine in pol v (K679A) has been shown to reduce both bypass of Tg and pol v’s propensity for dTMP-dGTP mismatch formation<sup>38</sup>. An additional contact to the incoming nucleotide was established in pol θ between the  $\beta$ -phosphate and O-helix residue Y2387 (**Supplementary Fig. 1**), which is also conserved in DNA polymerases from T-odd bacteriophages and pol v (refs. 33,38).

## The fingers subdomain closes on a cognate base pair

We pursued a related pol θ structure in the presence of  $\text{Mg}^{2+}$ , the likely physiological divalent metal ion, with a cognate dCMP-ddGTP base pair in the active site. This crystallization construct encompassed an additional 27-aa truncation from the N terminus (**Fig. 1c**), and the DNA sequence was blunted to remove a 3' template overhang. We also modified the 5' sequence of the template to complement consecutive incoming ddGTP molecules, in order to promote enzymatic chain termination of the primer strand before trapping the closed ternary complex. These crystals diffracted to similar resolution as the  $\text{Ca}^{2+}$  crystals, although we deemed the diffraction data inferior, owing to an off-origin native Patterson map peak and corresponding noncrystallographic symmetrical (NCS) translation relating the two protein molecules of the ASU. Coupled with anisotropic diffraction, skewed intensity-distribution statistics led to systematic data loss and reduced data completeness, even for highly redundant data sets compiled with extensive cross-crystal data merging (**Table 1**). However, we readily obtained a molecular-replacement solution, based on the  $\text{Ca}^{2+}$  model, and refined it (**Table 1**). Subsequently, this approach yielded crystals of a selenomethionyl variant in the presence of  $\text{Mg}^{2+}$ , from which we calculated a 4.6-Å anomalous difference Fourier map pinpointing the incorporated selenium atoms. Superimposing these anomalous peaks onto the  $\text{Ca}^{2+}$  model added confidence to the overall trace of the model and aided in assignment of side chains in poorly conserved regions by anchoring the pol θ sequence (**Fig. 1c**). The  $\text{Mg}^{2+}$  complex appeared to be globally similar to the complex obtained with  $\text{Ca}^{2+}$ . The former depicted subtle rearrangements in the fingers subdomain, with an adjustment in the relative position of helix O relative to the primer-template and palm subdomain, so that the O helix appeared to have moved closer to the cognate dCMP-ddGTP pair than the THF-ddATP (**Fig. 1e**).

## Unique insertion loops decorate pol θ

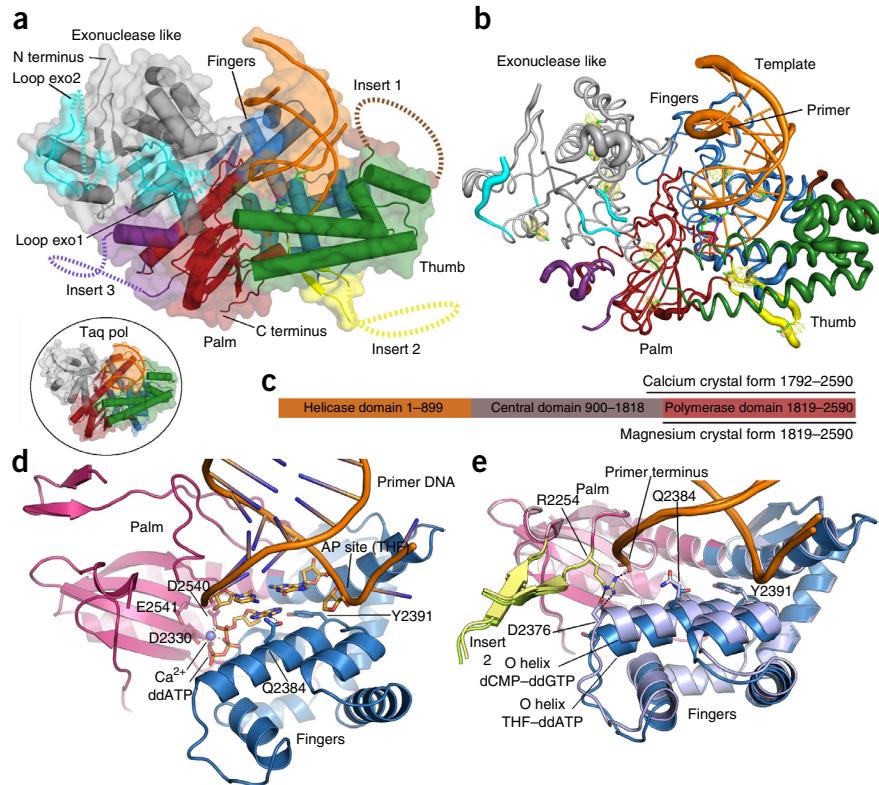
The locations of the three loop elements previously described in the palm and thumb subdomains<sup>12</sup> are now firmly established by the current pol θ structure in addition to two previously unreported

**Table 1** Data collection and refinement statistics

	Pol θ- $\text{Ca}^{2+}$ -THF-ddATP	Pol θ- $\text{Mg}^{2+}$ -dCMP-ddGTP	Pol θ-SeMet- $\text{Mg}^{2+}$ -dCMP-ddGTP
<b>Data collection</b>			
Space group	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>2</sub> 1	<i>P</i> 2 <sub>1</sub> 2 <sub>2</sub> 1
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	126.9, 137.0, 248.0	100.7, 135.8, 161.9	100.4, 134.8, 166.5
Resolution (Å)	50–3.9 (4.0–3.9) <sup>a</sup>	40–3.9 (4.0–3.9)	40–4.6 (4.8–4.6)
<i>R</i> <sub>merge</sub> (%)	20.5 (100)	19.6 (72.2)	15.5 (31.5)
<i>R</i> <sub>Friedel</sub> (%) <sup>b</sup>	8.4 (79.2)	9.0 (39.4)	16.9 (27.4)
<i>I</i> / $\sigma$ <i>I</i>	16.7 (2.1)	12.5 (1.9)	8.7 (4.3)
Completeness (%)	99.7 (99.1)	85.6 (64.0)	78.1 (53.1)
Redundancy	16.6 (14.0)	22.6 (13.4)	4.8 (3.7)
<b>Refinement</b>			
Resolution (Å)	30–3.91 (3.96–3.91)	30–3.90 (3.96–3.90)	
No. reflections	74,685 (2,403)	31,372 (1,987)	
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> (%)	24.1 (39.4) / 30.2 (43.9)	25.8 (35.0) / 31.6 (45.4)	
No. atoms	22,448	10,902	
Protein	19,900	9,950	
DNA	2,428	890	
Ligand/ion	120	62	
<i>B</i> factors (Å <sup>2</sup> )			
Protein	190	144	
DNA	215	166	
Ligand/ion	217	100	
r.m.s. deviations			
Bond lengths (Å)	0.015	0.014	
Bond angles (°)	1.362	1.169	

<sup>a</sup>Values in parentheses are for highest-resolution shells. Pol θ  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  data sets are the product of merging three and four crystals, respectively. The pol θ selenomethionine (SeMet)  $\text{Mg}^{2+}$  data set was collected on a single crystal. <sup>b</sup> $R_{\text{Friedel}} = \Sigma(|I_f - L_f|) / \Sigma |I_f|$

**Figure 1** Structure of human pol θ polymerase domain. (a) Overall THF-ddATP pol θ structure, shown in comparison to Taq polymerase (circled inset; PDB 1QSY<sup>32</sup>), with the five insertion loops identified: loops exo1 and exo2 (cyan), insert 1 (brown), insert 2 (yellow) and insert 3 (purple). Dotted lines represent regions of the loops not seen in the electron density map and thus not built in the crystallographic model. (b) Putty representation of pol θ, in the same orientation as in a, displaying the tube radius of the backbone trace proportionally to the refined atomic displacement parameters. Peaks from an NCS-averaged anomalous difference electron density map (yellow mesh, contoured at  $4\sigma$ ) pinpoint the locations of methionines. (c) Schematic of the domain architecture of full-length human pol θ and crystallization constructs, which encompass the entire C-terminal polymerase domain (residues 1819–2590), including the vestigial exonuclease-like subdomain (residues 1819–2090). (d) Close-up view of the pol θ active site, showing ddATP opposite THF in the closed conformation. Contacts (black lines) are mediated from the O-helix residue Q2384 to the incoming nucleobase. (e) Superimposition of the THF-ddATP (dark pink, dark blue and yellow-green) and dCMP-ddGTP (lighter hues) models, on the basis of palm-subdomain residues. Subtle rearrangements with cognate dCMP-ddGTP in the active site reposition the C-terminal end of the O helix, forming a putative salt bridge from R2254 of insert 2 to D2376 of the fingers. All molecular illustrations were made with PyMOL (<http://www.pymol.org/>).



insertion elements present in the N-terminal vestigial exonuclease-like subdomain (Fig. 1a and Supplementary Table 2). Intrinsic flexibility prevented modeling any of the pol θ insertion loops in their entirety (Fig. 1b). Nevertheless, the molecular-replacement solutions yielded electron density for the base of each loop; this served to orient the loops about the global polymerase fold and provided a partial description of their structures.

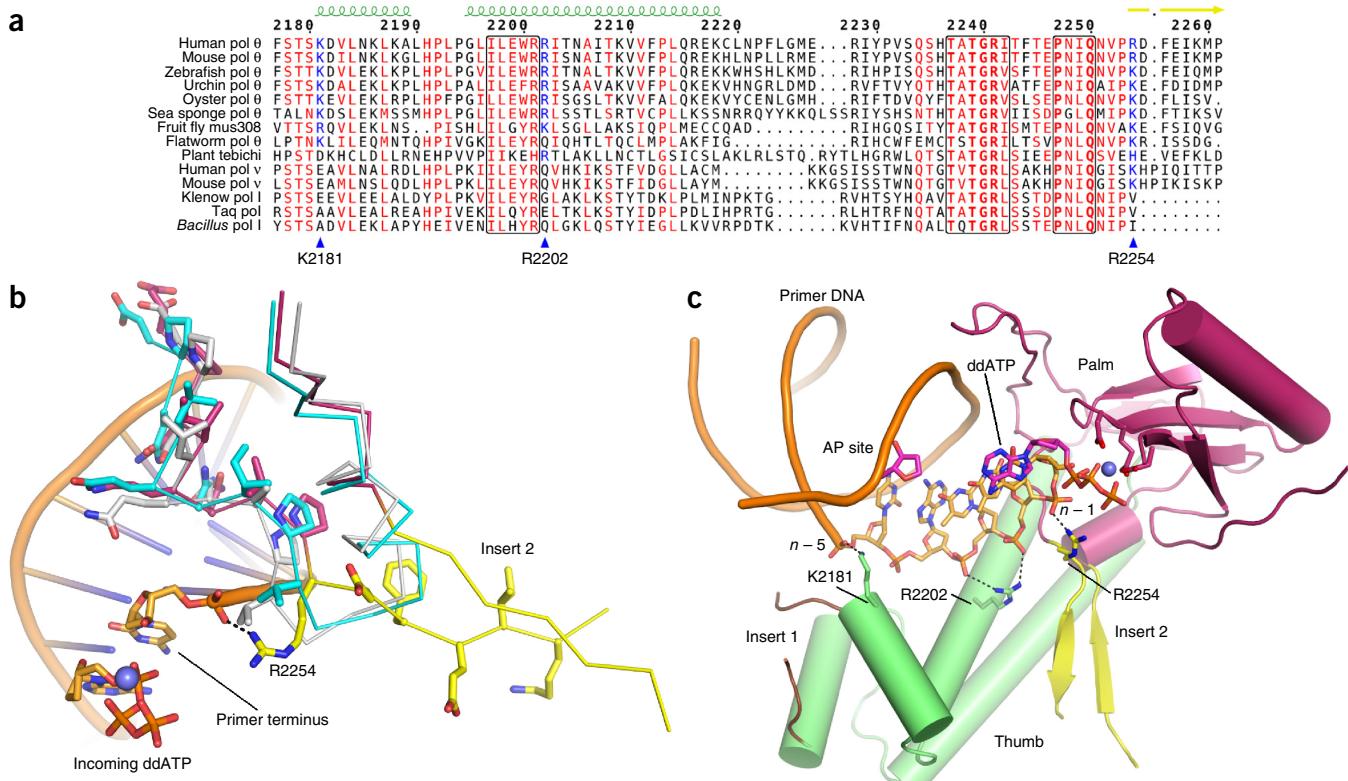
The tip of the thumb subdomain in pol θ resembles that of other homologous bacterial and phage polymerase structures, with two long helical segments separated by several flexible loops and smaller helices. Often in crystal structures of family-A DNA polymerases, the tip of the thumb develops relatively high *B* factors, owing to the flexibility of these loops<sup>32,33</sup>; here, pol θ proved to be no exception (Fig. 1b). Refinement produced electron density for only the first residues on either side of insert 1 (residues P2144–F2177), although the proximity of insert 1 to the DNA minor groove has demonstrated potential for contacts to the DNA<sup>9</sup>. The location of insert 1 resembles that of both the thioredoxin-binding motif<sup>12</sup> in T7 DNA polymerase<sup>33</sup> and the binding site of the pol γ processivity factor<sup>8</sup>, consistently with the ascribed role of insert 1 in the processivity of pol θ (ref. 11).

Insert 2 (residues R2254–S2313) departs from the palm subdomain near the primer terminus–binding site to join the thumb subdomain, where we observed electron density for two β-strands exiting the active site. These strands constitute the most highly conserved fragments of insertion loop 2 in pol θ and mus308-like proteins, from which a highly retained basic residue (R2254 in human pol θ; Fig. 2a) contacts the 3'-terminal phosphate of the primer DNA (Fig. 2b and Supplementary Fig. 2). A second contact of interest involving R2254 appeared only in the Mg<sup>2+</sup> crystal form, in which the O helix is closed further toward the active site. These subtle shifts reposition the acidic

C-terminal end of the O helix, from which a well-conserved aspartate residue (D2376) engages in a salt bridge with R2254 of insertion loop 2 (Fig. 1e), pinning the guanidinium moiety between the terminal phosphate and O helix. Unlike insert 3, which resides on the opposite site of the protein from the primer-template DNA, insert 2 is poised to align a poorly matched primer terminus for nucleophilic attack. Both inserts 2 and 3 have been shown to be necessary for TLS by pol θ, with insert 2 required for single-stranded-oligonucleotide extension<sup>10,11</sup>.

The N-terminal exonuclease domains of family-A DNA polymerases are generally divergent in sequence, despite conservation of the overall fold. A subset of these enzymes possess 3'-5'-exonuclease activity in this domain, such as *Escherichia coli* pol I, whereas this activity has been inactivated in other polymerases, such as Taq DNA polymerase. The crystallization construct used in this study includes approximately 300 residues upstream of the polymerase core, a region homologous to the *E. coli* pol I 3'-5'-exonuclease domain. The catalytic residues necessary for metal-ion binding and primer degradation (D355, E357 and D424 in *E. coli* pol I) are absent from the N-terminal subdomain of pol θ, and thus exonuclease activity has not been observed.

Density modification with cross-crystal NCS averaging provided detail to allow us to confidently place many side chains in regions of well-ordered secondary structure<sup>40</sup>, and this approach proved indispensable in modeling the N-terminal subdomain at 3.9-Å resolution, given the characteristic sequence divergence from available models. The current study identified two unreported insertions in the N-terminal subdomain of pol θ in addition to the three inserts in the core polymerase subdomains. The second insert of the N-terminal subdomain (loop exo2; Q1918–D1936) extends an established surface loop, which is only 10 aa in bacterial pol I (386–396 in *E. coli* pol I). Residues K1858–L1899 form the first insertion (loop exo1), of which



**Figure 2** DNA pol θ features insertions bearing conserved residues. (a) Structure-based sequence alignment illustrating the conservation of K2181, R2202 and R2254 (blue triangles) in multicellular organisms. Pol θ, mus308 and pol v have arginine or lysine at position 2254, whereas bacterial polymerases retain a hydrophobic side chain. Only pol θ-like enzymes (top nine proteins) have conserved basic residues at positions 2181 and 2202. (b) Superposition of pol θ (pink), Taq (PDB 1QS<sup>y32</sup>, cyan) and Klenow (PDB 1KLN<sup>41</sup>, light gray), illustrating how insertion loop 2 (yellow) departs from the canonical family A-polymerase fold. R2254 in human pol θ contacts the primer 3'-terminal ( $n - 1$ ) phosphate. (c) The palm (pink) and specialized thumb subdomain (green) of pol θ, illustrated to display unique contacts (black lines) to the primer DNA (orange sticks). Contacts of R2254 to the  $n - 1$  phosphate and R2202 to the  $n - 2$  and  $n - 3$  phosphates of the primer DNA are shown (dotted lines). K2181 establishes a putative salt bridge to the  $n - 5$  phosphate, just after insertion loop 1 (brown).

we could model only 8 aa. Loops exo1 and exo2 appear to be proximal to each other and to extend together from the conserved  $\beta$ -sheet of the N-terminal subdomain (Figs. 1a and 3a,b). In full-length pol θ, these insertion elements might provide contacts to the helicase domain or central region, given their proximity to the extreme N terminus of the polymerase-domain crystallization construct.

Insert 3 (S2503–F2534) terminates with two consecutive glycine residues that lead into a six-residue  $\beta$ -strand of the palm subdomain, directly before polymerase motif 5 (HDELLY) (Fig. 3a). Motif 5 contains two of the strictly conserved carboxylate residues, D2540 and E2541. A helical segment of insert 3 appeared at the interface between the N-terminal and the palm subdomains of pol θ, lying at the base of the cleft present in the N-terminal subdomain. In the *E. coli* pol I Klenow fragment, this cleft provides a path for DNA to migrate into the exonuclease proofreading active site<sup>41</sup>. In pol θ, a small loop (G2022–E2039) in the exonuclease-like domain, which would block DNA from taking a trajectory identical to that previously observed in Klenow fragment, was shifted (Fig. 3a).

#### The homologous *chaos1* variant (S1977P) catalyzes TLS

A previous genetic screen for genomic stability factors in mice implicated pol θ in DSB repair<sup>42,43</sup>. Mouse cells carrying the *Polq<sup>chaos1</sup>* allele have elevated spontaneous levels of micronuclei and also exhibit increased micronuclei after treatment with ionizing radiation or mitomycin C<sup>18,29,43</sup>. The *Polq<sup>chaos1</sup>* allele was subsequently verified

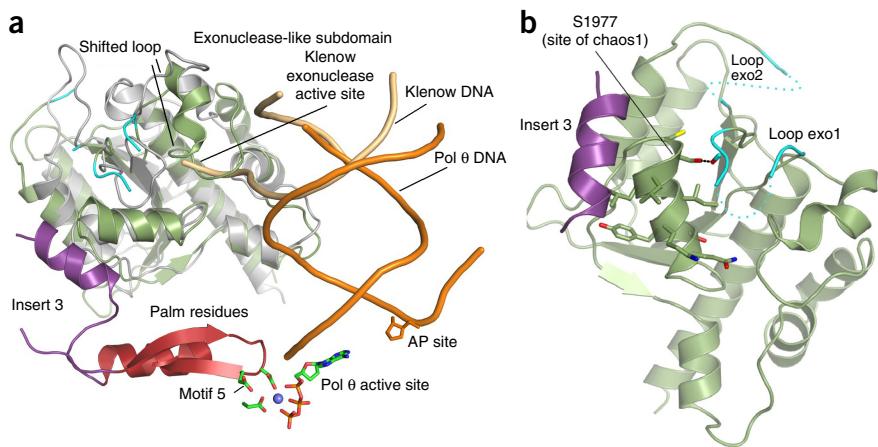
to encode a serine-to-proline missense mutation (S1977P in human pol θ) in the N-terminal exonuclease-like subdomain. Partial synthetic lethality results when the *Polq<sup>chaos1</sup>* allele is introduced into an *Atm*-knockout background, allowing only 10% of mice to survive past the neonatal period<sup>43</sup>.

The human pol θ model illustrates that S1977 caps the C-terminal end of a hydrophobic helix in the vestigial exonuclease-like subdomain. In our current model, S1997 could provide a hydrogen bond to the backbone carbonyl of D1897 of loop exo1 (Fig. 3b). We generated the S1977P mutant to biochemically characterize the homologous human *chaos1* variant. Assays aimed to evaluate TLS (Fig. 4a) and single-stranded primer extension (data not shown) failed to reveal a dramatic biochemical phenotype, however. These findings support the suggestion that cellular levels of the pol θ protein are depleted in the *Polq<sup>chaos1</sup>* mice, owing to poor *in vivo* expression or stability<sup>18</sup>.

#### Pol θ R2254V fails to bypass AP sites

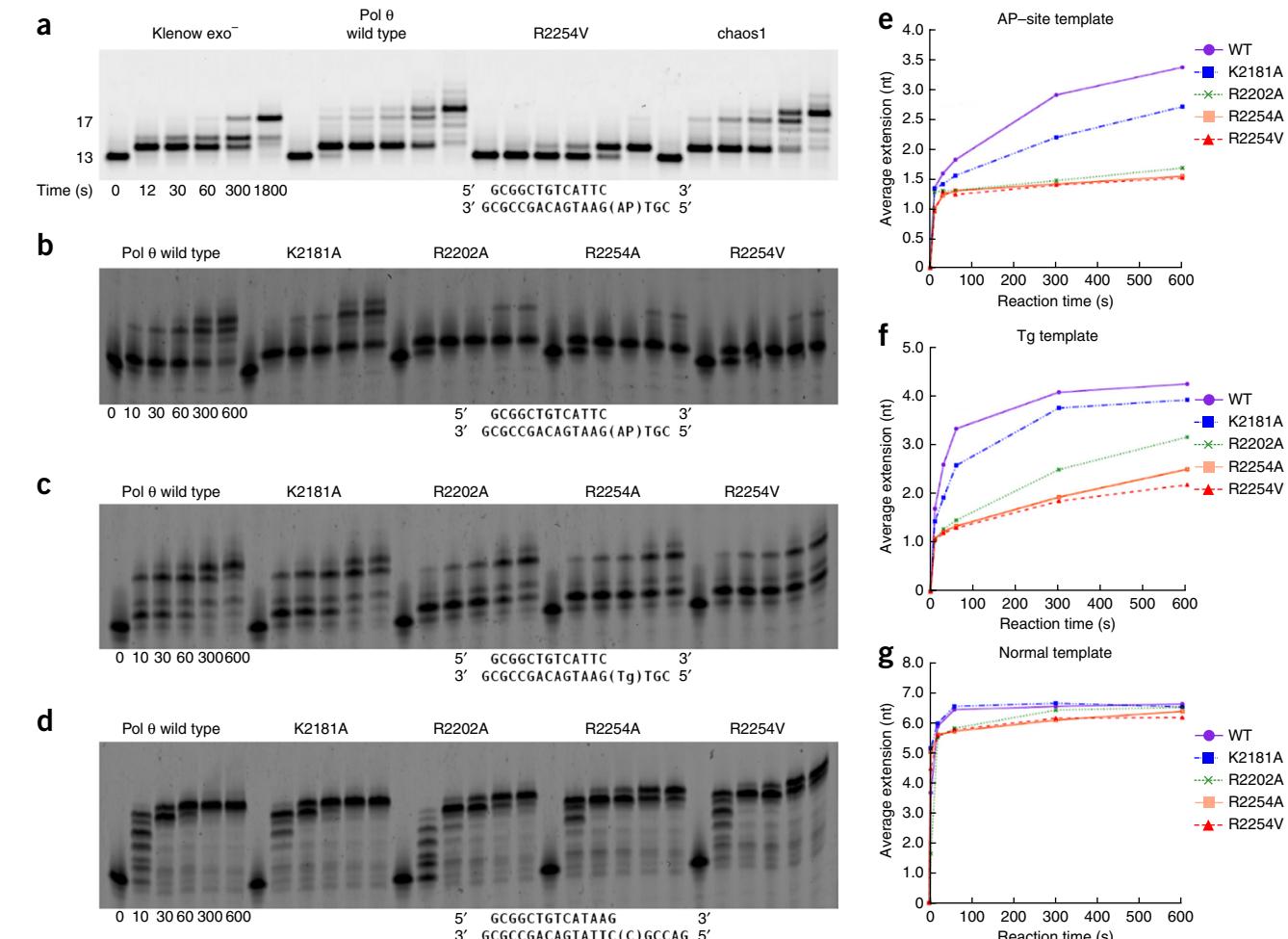
We designed the R2254V variant to evaluate how the conserved basic residue of insertion loop 2 contributes to pol θ's activity on single-stranded DNA oligonucleotides and its bypass of AP sites or Tg lesions. Family-A DNA polymerases from bacteria have a conserved hydrophobic amino acid (valine or isoleucine) at the equivalent position of R2254 (Fig. 2a), and the R2254V variant therefore mimics these bacterial enzymes. Pol θ R2254V, although active on double-stranded DNA, failed to bypass AP sites or Tg (Fig. 4) and was

**Figure 3** (a) The N-terminal inactivated exonuclease-like subdomain of pol θ (green), superimposed onto the 3'-5'-exonuclease domain of the Klenow fragment (gray, PDB 1KLN<sup>41</sup>). The Klenow structure shows the path of DNA into the exonuclease active site (tan), which is blocked by a shifted loop in the pol θ model. Insert 3 (purple) is resolved in proximity to the exonuclease-like subdomain and loops exo1 and exo2 (cyan). Insert 3 is sequentially close to polymerase motif 5 of the palm (red), thus providing a structural linkage between the exonuclease-like subdomain and polymerase active site. (b) Close-up view of loops exo1 and exo2 (cyan), with the location of S1977 indicated. S1977P is homologous to the mouse *Polgchaos1* allele. The current model indicates the potential for hydrogen-bonding (black dashes) from S1977 to a backbone carbonyl of loop exo1.



marginally hindered during extension of unannealed single-stranded DNA oligonucleotides, especially when provided with only pyrimidine nucleotides (Fig. 5). The salt bridge from R2254 to the primer

3'-terminal phosphate appears to be essential in compensating for interactions missing from the templating strand, owing to DNA lesions or distorted base-pairing.

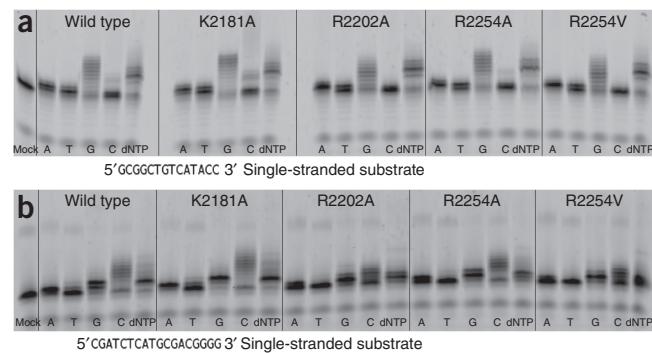


**Figure 4** Primer extension assays and quantifications. (a) Ability of different polymerase mutants to bypass the AP site during TLS (substrate shown at the bottom). Reactions used 125 nM enzyme, 250 nM primer-template and 500 μM of each nucleotide. Pol θ polymerase domain wild type, mutants R2254V and chaos1-allele homolog (pol θ S1977P), and Klenow fragment exo<sup>-</sup> were examined. (b–g) Primer extension assays comparing pol θ thumb-subdomain variants K2181A, R2202A, R2254A and R2254V, with substrates with an AP lesion (b,e), with a Tg lesion (c,f) or with an undamaged template (d,g); substrates are shown below each gel. All bands from primer extension assays in b–d are quantified to plot the average extension (nucleotides inserted past template) in e–g. WT, wild type.

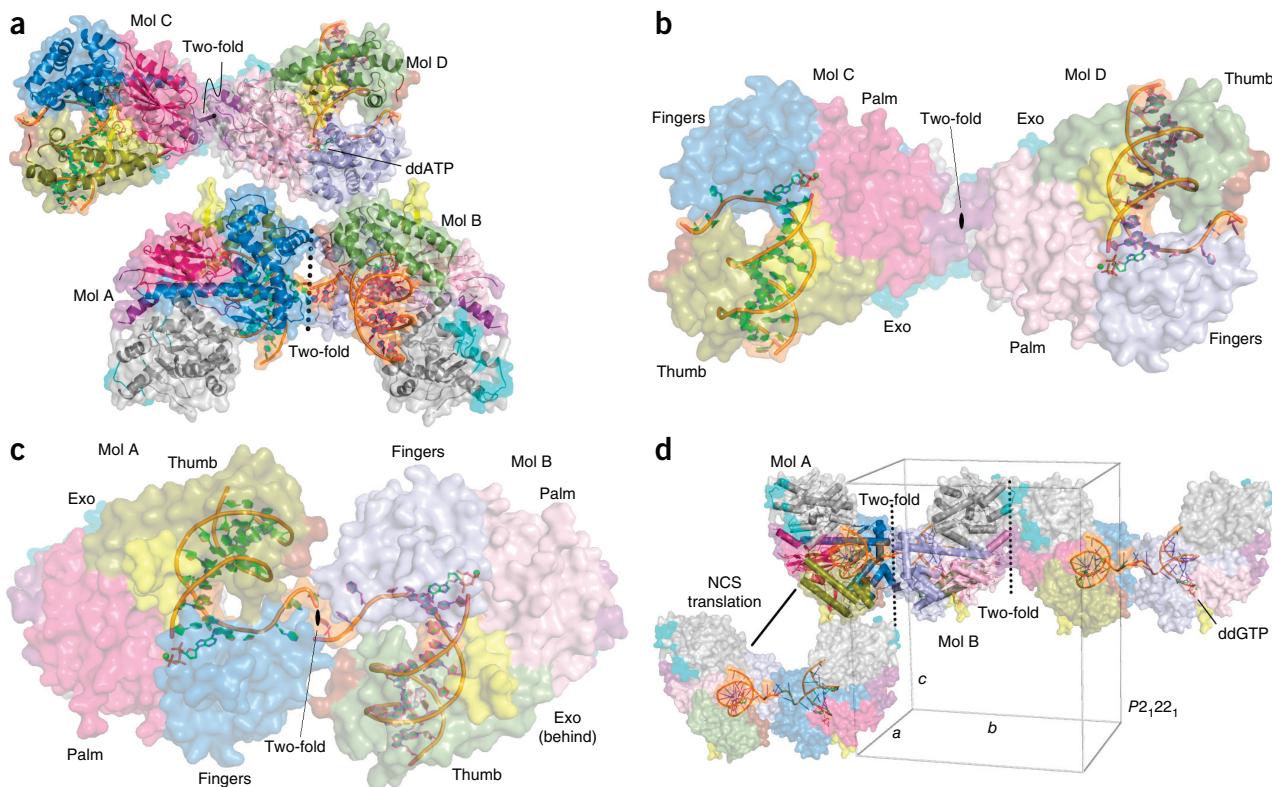
**Figure 5** Primer extension assays using pol θ polymerase domain variants and single-stranded substrates. (a) Single-stranded oligonucleotide-extension assay, showing that pol θ variants K2181A, R2202A, R2254A and R2254V are all able to catalyze in the indicated sequence context after 10 min. (b) Primer extension assay in an alternate sequence context, showing that variants of R2202 and R2254 produce shorter products when forced to use pyrimidines. In this sequence context, R2202A, R2254A and R2254V appear to be unable to use dTTP and show reduced incorporation of dCTP as compared to the parental enzyme.

The pol θ structure revealed other unique upstream contacts to the primer DNA strand, mediated by the specialized pol θ thumb subdomain (**Fig. 2a**). R2202 inserts its guanidinium moiety between the *n* – 2 and *n* – 3 primer-strand phosphates and is therefore poised to make two contacts to the DNA backbone. An additional contact is possible from K2181 of the thumb subdomain to the primer *n* – 5 phosphate (**Fig. 2c**). These contacts in addition to the contacts from R2254 to the *n* – 1 phosphate suggest that the pol θ thumb subdomain is specialized to provide salt bridges to the primer-strand phosphate backbone in excess of those conserved by all other family-A polymerases, including pol v (**Fig. 2a**). Like pol v and bacterial pol I, pol θ has retained arginine residues at positions 2201 and 2315 (631 and 690 in *E. coli* pol I), which provide absolutely conserved interactions to the *n* – 4 and *n* – 1 phosphate, respectively.

Generating alanine substitutions at K2181, R2202 or R2254 in pol θ, for the purpose of evaluating the contribution of the specialized thumb



subdomain during bypass of an AP site or Tg, revealed that the greatest TLS defect associates with a loss of contacts closest to the extreme primer terminus (**Fig. 4b–g**). Although the extension step of TLS opposite an AP site challenged all variants (**Fig. 4b,e**), the mutations made at R2254 (to alanine or valine) afforded the most dramatically deficient TLS phenotypes and were followed closely by R2202A. K2181A, as compared to the wild-type enzyme, reduced bypass of the AP site, although not to the same extent as other variants. When provided with a substrate placing Tg in the templating position (**Fig. 4c,f**), the same pattern emerged, although R2202A possessed a lesser defect, closer to that of K2181A. Even though pol v contains a lysine (K584) residue in insertion 2, homologous to R2254 in pol θ, R2202 and K2182 appear in



**Figure 6** The NCS two-fold axes in pol θ crystals, visualized and compared. (a) The ASU of the pol θ THF-ddATP model contains four protein–DNA complexes. Assembling the ASU requires two skew NCS two-fold axes (dotted lines). (b) The NCS two-fold axis relating chains C and D passes near insert 3 of the palm (pink) subdomain, adjacent to the N-terminal (gray) subdomain. (c) The 5'-template DNA–propagated NCS two-fold axis relating chains A and B, passing near the fingers (blue) and thumb (green) subdomains. (d) Crystal packing in the dCMP-ddGTP crystal form, showing that the two molecules per ASU pack differently from the four molecules of the THF-ddATP crystal form (described in a). The identical NCS two-fold axes are observed but are parallel to the  $2_1$  screw along cell edge *c*. The dimer at the right is generated by a crystallographic translation of the unit cell (green rectangular prism) along the *b* cell axis. The dimer at the bottom left is generated by the  $2_1$  screw along cell axis *c* (vertical axis), which in the context of the 5'-template DNA–propagated NCS two-fold axis gives rise to the NCS translation.

neither pol v nor bacterial homologs (**Fig. 2a**). These differences could explain why only pol θ grasps the primer tightly enough to bypass AP sites and extend certain minimally annealed primer-templates<sup>44</sup>. Pol v is less adept at bypassing blocking lesions, and it has been shown to bypass only 5S-Tg effectively<sup>13</sup>. Likewise, pol v does not extend single-stranded oligonucleotides *in vitro*, a reaction readily catalyzed by pol θ (ref. 10). Considering that pol θ has retained substantial ability to extend single-stranded substrates in the context of individual mutations at R2254, R2202 or K2181 *in vitro* (**Fig. 5a**), the transiently templated nature of this reaction is emphasized<sup>18</sup>, and pol θ must therefore use a different set of amino acid side chains to manipulate the single-stranded substrate as its own template for primer extension.

### Pol θ crystals reveal two modes of dimerization

The pol θ ASU in the Ca<sup>2+</sup> crystal form contains four protein-DNA complexes assembled as a dimer of two-fold dimers (**Fig. 6a**). Two dissimilar types of contacts mediated packing of the DNA ends distal from the active site: The 3' ends of each template DNA strand provide a dGMP overhang, two of which stack against conserved tryptophan residues (W1907) of adjacent protein molecules. The remaining template 3' ends appear in proximity, forming inter-DNA contacts outside the molecular footprint of pol θ.

Three perpendicular two-fold screw axes generate the space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>. However, two alternate pure two-fold axes assemble the ASU of the Ca<sup>2+</sup> crystal form, which exist as NCS operators. The first NCS two-fold axis passes near insert 3 and the N-terminal subdomain (**Fig. 6b**). The second NCS two-fold axis passes adjacent to the 5'-template DNA, upstream of the polymerase active site, and near the putative location of insertion loop 1 (**Fig. 6c**). The ASU in the current Ca<sup>2+</sup> model therefore presents two potentially biologically relevant assemblies, in which the 5'-template DNA's two-fold-symmetry axis relates protein chains A and B, and the two-fold-symmetry axis of insert 3 relates molecules C and D. Interestingly, the pseudo-symmetry observed in the Mg<sup>2+</sup> crystal form was a consequence of similar configurations within the ASU, in which the identical NCS two-fold dimeric relationships about the 5'-template DNA and insert 3 appeared once again. In this case, both NCS two-fold axes appeared nearly parallel to the crystallographic 2<sub>1</sub> screw axis along *c*, thus generating the NCS translation (**Fig. 6d**).

### DISCUSSION

An alternative DNA end-joining pathway in eukaryotes absolutely requires the C-terminal polymerase activity of the *POLQ* gene product, applied specifically in a role for which other DNA polymerases or helicases are unable to compensate<sup>18–20</sup>. The current study informs a mechanism by which the human pol θ polymerase domain maintains a tight hold on the primer DNA strand by establishing interactions from positively charged residues K2181, R2202 and R2254 to the phosphate backbone. The crystal structures also establish that the pol θ protein can dimerize. Although we did not observe a multimer in solution by gel-filtration chromatography, resolving identical NCS two-fold axes in two different crystal forms indicates that multimerization of the pol θ polymerase domain is possible at sufficient concentration. Analysis of these potential dimer interfaces in PISA<sup>45</sup> attributes 12,400 and 13,100 Å<sup>2</sup> of buried surface area for the 5'-DNA-propagated and insert-3 NCS two-fold interfaces (**Fig. 6b,c**), respectively, neglecting the flexible segments omitted from the models. In the context of alternative end-joining, pol θ multimerization might be essential for bringing two DNA ends together or sequestering free DNA ends from classical nonhomologous end-joining factors, such as Ku<sup>46</sup>.

Higher expression of *POLQ* correlates with decreased survival in patients with breast cancer<sup>47,48</sup>, and knockdown of the gene product in several malignant cell lines induces radiosensitivity<sup>30</sup>. Given that *POLQ* is nonessential in healthy cells<sup>43</sup>, the potential to pharmacologically target unique features of the pol θ protein, such as the specialized thumb subdomain revealed by this study, is an appealing approach for adjuvant radiation cancer therapy<sup>26</sup>. Future experiments will expand understanding of the mechanisms by which pol θ protects cancer cells from radiation or chemotherapeutics. The extent to which transient pol θ dimers might bridge DSBs, for example, must also be considered as *POLQ* is evaluated as a potential target for next-generation cancer drugs.

*Note added in proof:* During final revision of this manuscript, several additional papers were published on the function of pol θ in alternative end-joining of DSBs<sup>49–51</sup>. Kent *et al.*<sup>51</sup> suggest that a dimeric form of pol θ could function in repair, with dimerization mediated by insertion loop 2. Our crystal structures indicate multimerization of the pol θ polymerase domain (**Fig. 6**) but do not readily implicate insertion loop 2 in dimerization of this domain.

### METHODS

Methods and any associated references are available in the [online version of the paper](#).

**Accession codes.** Coordinates and structure factors for the ternary complexes of human pol θ obtained with Ca<sup>2+</sup> (THF-ddATP) and Mg<sup>2+</sup> (dCMP-ddGTP) have been deposited in the Protein Data Bank under accession codes 4X0P and 4X0Q, respectively.

*Note: Any Supplementary Information and Source Data files are available in the [online version of the paper](#).*

### ACKNOWLEDGMENTS

These studies were funded by US National Institutes of Health grants R01 CA052040 (S.D.) and CA097175 (R.D.W.), grant RP130297 from the Cancer Prevention and Research Institute of Texas (R.D.W.) and the Grady F. Saunders PhD. Distinguished Research Professorship (R.D.W.). The GM/CA beamline at the Advanced Photon Source has been funded in whole or in part with federal funds from the US National Cancer Institute (ACB-12002) and the US National Institute of General Medical Sciences (AGM-12006). This research used resources of the Advanced Photon Source, a US Department of Energy (DOE) Office of Science User Facility operated for the US DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357.

### AUTHOR CONTRIBUTIONS

K.E.Z. performed crystallization, data collection and structure refinement. A.M.A. expressed and purified native and selenomethionine pol θ protein. P.A. obtained preliminary diffraction images at Diamond Light Source beamline I04. K.E.Z. expressed and purified mutant proteins and performed biochemical experiments. S.D. and R.D.W. oversaw the project. K.E.Z. wrote the manuscript with S.D. and R.D.W. All authors discussed the results and commented on the manuscript.

### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

1. Harris, P.V. *et al.* Molecular cloning of *Drosophila* mus308, a gene involved in DNA cross-link repair with homology to prokaryotic DNA polymerase I genes. *Mol. Cell. Biol.* **16**, 5764–5771 (1996).
2. Inagaki, S. *et al.* *Arabidopsis* TEBICHI, with helicase and DNA polymerase domains, is required for regulated cell division and differentiation in meristems. *Plant Cell* **18**, 879–892 (2006).
3. Yousefzadeh, M.J. & Wood, R.D. DNA polymerase POLQ and cellular defense against DNA damage. *DNA Repair (Amst.)* **12**, 1–9 (2013).
4. Seki, M., Marini, F. & Wood, R.D. POLQ (Pol theta), a DNA polymerase and DNA-dependent ATPase in human cells. *Nucleic Acids Res.* **31**, 6117–6126 (2003).

5. Marini, F., Kim, N., Schuffert, A. & Wood, R.D. POLN, a nuclear PolA family DNA polymerase homologous to the DNA cross-link sensitivity protein Mus308. *J. Biol. Chem.* **278**, 32014–32019 (2003).
6. Longley, M.J., Prasad, R., Srivastava, D.K., Wilson, S.H. & Copeland, W.C. Identification of 5'-deoxyribose phosphate lyase activity in human DNA polymerase gamma and its role in mitochondrial base excision repair *in vitro*. *Proc. Natl. Acad. Sci. USA* **95**, 12244–12248 (1998).
7. Prasad, R. *et al.* Human DNA polymerase theta possesses 5'-dRP lyase activity and functions in single-nucleotide base excision repair *in vitro*. *Nucleic Acids Res.* **37**, 1868–1877 (2009).
8. Lee, Y.S., Kennedy, W.D. & Yin, Y.W. Structural insight into processive human mitochondrial DNA synthesis and disease-related polymerase mutations. *Cell* **139**, 312–324 (2009).
9. Arana, M.E., Seki, M., Wood, R.D., Rogozin, I.B. & Kunkel, T.A. Low-fidelity DNA synthesis by human DNA polymerase theta. *Nucleic Acids Res.* **36**, 3847–3856 (2008).
10. Hogg, M., Sauer-Eriksson, A.E. & Johansson, E. Promiscuous DNA synthesis by human DNA polymerase theta. *Nucleic Acids Res.* **40**, 2611–2622 (2012).
11. Hogg, M., Seki, M., Wood, R.D., Doublé, S. & Wallace, S.S. Lesion bypass activity of DNA polymerase theta (POLQ) is an intrinsic property of the pol domain and depends on unique sequence inserts. *J. Mol. Biol.* **405**, 642–652 (2011).
12. Seki, M. *et al.* High-efficiency bypass of DNA damage by human DNA polymerase Q. *EMBO J.* **23**, 4484–4494 (2004).
13. Takata, K., Shimizu, T., Iwai, S. & Wood, R.D. Human DNA polymerase N (POLN) is a low fidelity enzyme capable of error-free bypass of 5S-thymine glycol. *J. Biol. Chem.* **281**, 23445–23455 (2006).
14. Yoon, J.H., Roy Choudhury, J., Park, J., Prakash, S. & Prakash, L. A role for DNA polymerase theta in promoting replication through oxidative DNA lesion, thymine glycol, in human cells. *J. Biol. Chem.* **289**, 13177–13185 (2014).
15. Yoshimura, M. *et al.* Vertebrate POLQ and POLbeta cooperate in base excision repair of oxidative DNA damage. *Mol. Cell* **24**, 115–125 (2006).
16. Asagoshi, K. *et al.* Single-nucleotide base excision repair DNA polymerase activity in *C. elegans* in the absence of DNA polymerase beta. *Nucleic Acids Res.* **40**, 670–681 (2012).
17. Martomo, S.A., Saribasak, H., Yokoi, M., Hanaoka, F. & Gearhart, P.J. Reevaluation of the role of DNA polymerase theta in somatic hypermutation of immunoglobulin genes. *DNA Repair (Amst.)* **7**, 1603–1608 (2008).
18. Yousefzadeh, M.J. *et al.* Mechanism of suppression of chromosomal instability by DNA polymerase POLQ. *PLoS Genet.* **10**, e1004654 (2014).
19. Chan, S.H., Yu, A.M. & McVey, M. Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in *Drosophila*. *PLoS Genet.* **6**, e1001005 (2010).
20. Koole, W. *et al.* A Polymerase Theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites. *Nat. Commun.* **5**, 3216 (2014).
21. Roerink, S.F., van Schendel, R. & Tijsterman, M. Polymerase theta-mediated end joining of replication-associated DNA breaks in *C. elegans*. *Genome Res.* **24**, 954–962 (2014).
22. Yu, A.M. & McVey, M. Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions. *Nucleic Acids Res.* **38**, 5706–5717 (2010).
23. White, T.B. & Lambowitz, A.M. The retrohoming of linear group II intron RNAs in *Drosophila melanogaster* occurs by both DNA ligase 4-dependent and -independent mechanisms. *PLoS Genet.* **8**, e1002534 (2012).
24. Fernandez-Vidal, A. *et al.* A role for DNA polymerase θ in the timing of DNA replication. *Nat. Commun.* **5**, 4285 (2014).
25. Lemée, F. *et al.* DNA polymerase theta up-regulation is associated with poor survival in breast cancer, perturbs DNA replication, and promotes genetic instability. *Proc. Natl. Acad. Sci. USA* **107**, 13390–13395 (2010).
26. Higgins, G.S. *et al.* Overexpression of POLQ confers a poor prognosis in early breast cancer patients. *Oncotarget* **1**, 175–184 (2010).
27. Lessa, R.C. *et al.* Identification of upregulated genes in oral squamous cell carcinomas. *Head Neck* **35**, 1475–1481 (2013).
28. Allera-Moreau, C. *et al.* DNA replication stress response involving PLK1, CDC6, POLQ, RAD51 and CLASPIN upregulation prognoses the outcome of early/mid-stage non-small cell lung cancer patients. *Oncogenesis* **1**, e30 (2012).
29. Goff, J.P. *et al.* Lack of DNA polymerase theta (POLQ) radiosensitizes bone marrow stromal cells *in vitro* and increases reticulocyte micronuclei after total-body irradiation. *Radiat. Res.* **172**, 165–174 (2009).
30. Higgins, G.S. *et al.* A small interfering RNA screen of genes involved in DNA repair identifies tumor-specific radiosensitization by POLQ knockdown. *Cancer Res.* **70**, 2984–2993 (2010).
31. McCoy, A.J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D Biol. Crystallogr.* **63**, 32–41 (2007).
32. Li, Y., Mitaxov, V. & Waksman, G. Structure-based design of Taq DNA polymerases with improved properties of dideoxynucleotide incorporation. *Proc. Natl. Acad. Sci. USA* **96**, 9491–9496 (1999).
33. Doublé, S., Tabor, S., Long, A.M., Richardson, C.C. & Ellenberger, T. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* **391**, 251–258 (1998).
34. Wu, E.Y. & Beese, L.S. The structure of a high fidelity DNA polymerase bound to a mismatched nucleotide reveals an “ajar” intermediate conformation in the nucleotide selection mechanism. *J. Biol. Chem.* **286**, 19758–19767 (2011).
35. Zahn, K.E., Belrhali, H., Wallace, S.S. & Doublé, S. Caught bending the A-rule: crystal structures of translesion DNA synthesis with a non-natural nucleotide. *Biochemistry* **46**, 10551–10561 (2007).
36. Motea, E.A., Lee, I. & Berdis, A.J. Quantifying the energetic contributions of desolvation and pi-electron density during translesion DNA synthesis. *Nucleic Acids Res.* **39**, 1623–1637 (2011).
37. Obeid, S., Welte, W., Diederichs, K. & Marx, A. Amino acid templating mechanisms in selection of nucleotides opposite abasic sites by a family a DNA polymerase. *J. Biol. Chem.* **287**, 14099–14108 (2012).
38. Takata, K., Arana, M.E., Seki, M., Kunkel, T.A. & Wood, R.D. Evolutionary conservation of residues in vertebrate DNA polymerase N conferring low fidelity and bypass activity. *Nucleic Acids Res.* **38**, 3233–3244 (2010).
39. Castro, C. *et al.* Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nat. Struct. Mol. Biol.* **16**, 212–218 (2009).
40. Terwilliger, T.C. Statistical density modification with non-crystallographic symmetry. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 2082–2086 (2002).
41. Beese, L.S., Derbyshire, V. & Steitz, T.A. Structure of DNA polymerase I Klenow fragment bound to duplex DNA. *Science* **260**, 352–355 (1993).
42. Shima, N. *et al.* Phenotype-based identification of mouse chromosome instability mutants. *Genetics* **163**, 1031–1040 (2003).
43. Shima, N., Munroe, R.J. & Schimenti, J.C. The mouse genomic instability mutation *chaos1* is an allele of *Po1* that exhibits genetic interaction with *Atm*. *Mol. Cell. Biol.* **24**, 10381–10389 (2004).
44. Seki, M. & Wood, R.D. DNA polymerase theta (POLQ) can extend from mismatches and from bases opposite a (6–4) photoproduct. *DNA Repair (Amst.)* **7**, 119–127 (2008).
45. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
46. Ribes-Zamora, A., Indiviglio, S.M., Mihalek, I., Williams, C.L. & Bertuch, A.A. TRF2 interaction with Ku heterotetramerization interface gives insight into c-NHEJ prevention at human telomeres. *Cell Reports* **5**, 194–206 (2013).
47. Yang, S., Zhang, H., Guo, L., Zhao, Y. & Chen, F. Reconstructing the coding and non-coding RNA regulatory networks of miRNAs and mRNAs in breast cancer. *Gene* **548**, 6–13 (2014).
48. Santarpia, L. *et al.* DNA repair gene patterns as prognostic and predictive factors in molecular breast cancer subtypes. *Oncologist* **18**, 1063–1073 (2013).
49. Ceccaldi, R. *et al.* Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. *Nature* **518**, 258–262 (2015).
50. Mateos-Gomez, P.A. *et al.* Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. *Nature* **518**, 254–257 (2015).
51. Kent, T., Chandramouly, G., McDevitt, S.M., Ozdemir, A.Y. & Pomerantz, R.T. Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase θ. *Nat. Struct. Mol. Biol.* **22**, 230–237 (2015).

## ONLINE METHODS

**Crystallization of recombinant human pol θ.** The pol θ polymerase crystallization construct (residues 1792–2590) was expressed from the pSUMO3 vector in Rosetta2(DE3)pLysS cells (Novagen) by autoinduction and was purified to homogeneity with a nickel-NTA resin (Thermo Scientific), a HiTrap heparin column (GE Healthcare Life Sciences) and a Superdex 200 Increase GL gel-filtration column (GE Healthcare Life Sciences)<sup>11</sup>. The selenomethionyl pol θ protein was also autoinduced, but with the Overnight Express System 2 kit (EMD Millipore) with 125 mg/L of L-selenomethionine (Acros Organics). The protein was concentrated to 5 mg/mL in a buffer of 150 mM ammonium acetate, 150 mM KCl, 40 mM Tris-HCl buffer, pH 8.0, 2.5 mM TCEP and 1% (v/v) glycerol. Reacting pol θ (2.5 mg/mL) with a 13-mer to 18-mer oligonucleotide (50 μM) containing a THF abasic-site analog (Midland), with annealed primer (ggggctgttcattc) and template (cggt(THF)gaatgcacgcccc), in the presence of 1 mM ddATP, 300 μM sucrose monolaurate and 20 mM spermine tetrahydrochloride prepared the sample for hanging-drop vapor diffusion over a 1-mL reservoir containing 9.5% (v/v) PEG 2000 MME, 2% (v/v) MPD, 50 mM CaCl<sub>2</sub>, 200 mM KCl, and 100 mM Tris-HCl buffer, pH 8.5, by mixture of 0.6 μL of the reservoir solution with an equal part reaction solution. Narrow, long crystals of approximate final dimensions 400 × 60 × 60 μm<sup>3</sup> grew over the next 3 d. Cryoprotection was achieved by addition of 1 μL of a solution containing 25% (v/v) PEG 2000 MME, 25% (v/v) MPD, 20 mM CaCl<sub>2</sub>, 10 mM spermine tetrahydrochloride, 260 mM KCl and 1 mM nucleotide directly to the drop, before flash cooling into liquid nitrogen.

The alternative Mg<sup>2+</sup> crystal form was discovered in an attempt to obtain better-diffracting pol θ crystals. An additional 27 residues were cleaved from the N terminus of the crystallization construct, leaving residues 1819–2590. The DNA sequence was also altered to allow for formation of a Watson-Crick base pair in the active site and elimination of the 3'-template dGMP overhang. Two consecutive dCMPs were introduced in the template to allow for enzymatic chain termination of the primer strand by incubation with ddGMP, to yield the primer (ggggctgttcattc) and template (cggtccaaatgcacgcccc). Crystals were obtained from similar conditions as before, except that the reaction mixture was supplemented with 1 mM MgCl<sub>2</sub>, and ddGTP replaced the ddATP. The PEG 2000 MME in the reservoir was increased to 12% and the MPD to 10%. Although these crystals grew to similar length, they were wider (80 μm) but thinner (20 μm). Increasing the precipitant slightly to 12.5% allowed for crystallization similarly to the selenomethionyl variant of the Δ1818 construct in complex with the same primer-template, although these crystals were approximately half the size of the native Mg<sup>2+</sup> crystals. Both native and selenomethionyl forms were cryoprotected by the prior procedure, except the cryoprotection solution was supplemented with MgCl<sub>2</sub> instead of CaCl<sub>2</sub>.

**Data collection, structure solution and refinement.** Pol θ crystals of the Ca<sup>2+</sup> form diffracted at the APS synchrotron (beamline 23-ID-D) past 3.9 Å at 12 keV, and complete data sets were collected from these radiation-sensitive primitive orthorhombic (*P*2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>) crystals at 80 K by merging of approximately 20 image segments, as the 20-μm beam was translated along the length of the crystal. Owing to anisotropy of diffraction, multiple crystals were merged in order to obtain a complete data set with substantial redundancy to be used in structure solution and refinement. An identical strategy was used to collect native and derivative data sets at the selenium peak (12.67 keV) for the Mg<sup>2+</sup> crystal forms.

Structure solution of the Ca<sup>2+</sup> crystal form was accomplished by molecular replacement, with the Taq DNA polymerase ternary complex 1QSY<sup>32</sup> used to search for the four protein–DNA complexes per ASU. Long segments of secondary structure were rigid-body-fitted into a reduced-bias ‘prime and switch’ map<sup>52</sup> with NCS averaging. Rebuilding in Coot<sup>53</sup> and refinement in Phenix<sup>54</sup> allowed most of the DNA to be built around the active site, although uncertainty remained concerning insertion elements and in assigning sequence to the trace of the N-terminal vestigial exonuclease-like subdomain.

The Mg<sup>2+</sup> crystal form diffracted to similar resolution as the Ca<sup>2+</sup> form but was plagued by a strong off-origin native Patterson peak consistent with an NCS translation of (0.1, -0.5, 0.5) fractional units, which was verified by molecular replacement in the space group *P*2<sub>1</sub>2<sub>1</sub>, with the Ca<sup>2+</sup> structure as a search model. The identical NCS two-fold dimeric configurations were resolved, as observed of crystals grown in Ca<sup>2+</sup>, although in this case a single dimer constituted the ASU. The NCS two-folds occurred parallel to the crystallographic 2<sub>1</sub> screw operator along *c*, which gave rise to pseudosymmetry and potential twinning (Fig. 6d). Refinement against potential twin laws in all related monoclinic cells failed to improve refinement statistics, however, and twinning was ruled out, despite the presence of suspect intensity distribution. Moreover, the electron density maps were of lesser quality. Nevertheless, phases obtained by refining the rigid-body solution of the Mg<sup>2+</sup> crystal form in Phenix<sup>55</sup> proved useful in generating cross-crystal density-modified NCS-averaged maps, which provided high-quality electron density for side chain placement in the final model and helped resolve ambiguities in the tracing of the inserts and N-terminal exonuclease-like subdomain. Phases calculated from the transformed model also allowed visualization of anomalous peaks at 4.6 Å, owing to selenomethionine substitution in the SeMet Mg<sup>2+</sup> crystal form, and thereby provided invaluable verification of the backbone trace in the N-terminal domain and insertion element 2.

Concluding refinement steps were conducted in Phenix<sup>54</sup>, fitting protein domains and nucleic acid chains as rigid bodies. Torsion NCS, secondary structure and Ramachandran restraints were enforced during individual *xyz* refinement, before TLS and group\_AdP refinement, at which point the *R*<sub>work</sub> reached 28.3% (*R*<sub>free</sub> = 31.2%) for the Ca<sup>2+</sup> model. Finally, the individual\_AdP strategy was applied for two cycles of refinement to yield the completed model, for which statistics are reported. Ramachandran analysis places 98.6% of residues in the Ca<sup>2+</sup> model in favored or allowed regions of the plot, leaving 1.4% outliers. The final Mg<sup>2+</sup> model was obtained by trimming several residues from the base of insertion loop 1, and omitting 3 bp of the duplex DNA extruding from the molecule in the 5' primer direction, before refinement was run similarly in Phenix. The completed Mg<sup>2+</sup> model identifies 99.3% of residues in the favored or allowed region of the Ramachandran plot, with 0.7% outliers.

**Variant production and biochemical analysis.** Variant pol θ constructs were generated by site-directed mutagenesis with the QuikChange XL kit (Stratagene). Biochemical reactions were conducted by preincubation of the primer-template (250 nM) with pol θ variants (125 nM) in 20 mM Tris-HCl buffer, pH 8.0, 25 mM KCl and 1 mM β-mercaptoethanol, before addition of deoxynucleotides (500 μM each or individually) and 10 mM MgCl<sub>2</sub>. Aliquots were quenched at the indicated time points by mixture with equal parts of a quench solution made of 95% formamide, 20 mM EDTA and trace bromophenol blue. Products were separated on a SequaGel UreaGel (National Diagnostics) denaturing sequencing gel (12%) and visualized by excitation of a 5'-tetrachlorofluorescein tag on the primer strand at the 532-nm setting on a PharosFX (Bio-Rad) imaging device. Bands were quantified with QuantityOne (Bio-Rad) and plotted in GNUpot via the function  $\sum_i^n i \times \rho_i$ , where  $\rho_i$  is the normalized density of the band corresponding to the *i*th extension product, to yield the average extension of the primer strand in fractional nucleotides.

52. Terwilliger, T. SOLVE and RESOLVE: automated structure solution, density modification and model building. *J. Synchrotron Radiat.* **11**, 49–52 (2004).
53. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
54. Headd, J.J. et al. Use of knowledge-based restraints in phenix.refine to improve macromolecular refinement at low resolution. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 381–390 (2012).
55. Adams, P.D. et al. The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94–106 (2011).