Structure and mechanism of B-family DNA polymerase ζ specialized for translesion DNA synthesis

Radhika Malik1, Mykhailo Kopylov2,6, Yacob Gomez-Llorente1,6, Rinku Jain1,6, Robert E. Johnson3, Louise Prakash3, Satya Prakash3, Iban Ubarretxena-Belandia1,4,5 ✉ and Aneel K. Aggarwal1,6 ✉

DNA polymerase ζ (Polζ) belongs to the same B-family as high-fidelity replicative polymerases, yet is specialized for the extension reaction in translesion DNA synthesis (TLS). Despite its importance in TLS, the structure of Polζ is unknown. We present cryo-EM structures of the Saccharomyces cerevisiae Polζ holoenzyme in the act of DNA synthesis (3.1 Å) and without DNA (4.1 Å). Polζ displays a pentameric ring-like architecture, with catalytic Rev3, accessory Pol31, Pol32 and two Rev7 subunits forming an uninterrupted daisy chain of protein–protein interactions. We also uncover the features that impose high fidelity during the nucleotide-incorporation step and those that accommodate mismatches and lesions during the extension reaction. Collectively, we decrypt the molecular underpinnings of Polζ’s role in TLS and provide a framework for new cancer therapeutics.

The survival of all organisms depends critically on the ability to faithfully replicate DNA. However, cellular DNA is susceptible to damage by normal metabolic activities and environmental factors, such as ultraviolet light, ionizing radiation and industrial carcinogens, that can cause lesions that evade DNA repair and stall the replication machinery. To avoid compromising genomic integrity, both prokaryotes and eukaryotes possess specialized translesion synthesis (TLS) DNA polymerases (Pols) that can replicate through these lesions. Most of the TLS polymerases belong to the Y-family, which includes the single-subunit Polη, Polλ, Polκ and Rev1 in humans1–3. In contrast, Polζ is a multisubunit TLS polymerase containing catalytic Rev3 and accessory Rev7, Pol31 and Pol32 subunits. Rev3 belongs to the same B-family as Pol1, Pol2 and Pol3, the catalytic subunits of the high-fidelity eukaryotic replicative polymerases α, ε and δ, respectively4,5. However, unlike the replicative polymerases, Polζ is specialized for the extension step of lesion bypass, whereby it is recruited to add nucleotides once another TLS polymerase has added a nucleotide opposite a lesion4,5. The ability of Polζ to carry out synthesis downstream of mismatched termini and diverse DNA lesions is important in maintaining genome integrity and preventing cancer6. At the same time, human Polζ has emerged as an important determinant for tumor resistance to chemotherapeutic agents in various cancers4,5.

The Rev3 sequence differs from that of Pol1, Pol2 and Pol3 in containing a large insert which comprises the Rev7-binding sites (Fig. 1a). Rev7 (also known as MAD2B) increases the activity of Rev3 (ref. 6), and a mutation in mouse Rev7 that disrupts its association with Rev3 leads to defects in development and to the accumulation of DNA damage7. Rev7 is a member of the HORMA (Hop1, Rev7 and Mad2) family of proteins8, and has roles outside of Polζ, including association with the spindle assembly checkpoint protein Mad2 during chromosome segregation9. The Pol31–Pol32 subcomplex associates with Polζ via interactions between the Rev3 carboxy-terminal domain (CTD) and Pol31 (refs. 10,11). The Rev3 CTD contains two cysteine-rich metal-binding modules, CysAD and CysBD (Fig. 1a), analogous to the modules at the C termini of Pol1, Pol2 and Pol3.

Despite its importance in protecting eukaryotic cells from DNA damage, the structural basis of the ability of Polζ to function as the ‘master’ extender in TLS remains unknown. Available structural information is limited to structures of human Rev7 in complex with short human Rev3 peptides7,8, and to structures of Pol31 and Pol32, and their human counterparts (p50 and p66)9,12. There is no structural information on the catalytic subunit Rev3, which, due to its large size (1,504 amino acids in yeast Rev3 and 3,130 amino acids in human Rev3L)13 and tendency to aggregate, has resisted crystallization. A low-resolution (~22-Å) model of apo Polζ based on negatively stained electron microscopy provided some information on the overall shape of the holoenzyme, but none on subunit architecture or protein–protein and protein-DNA interactions14.

We present here cryo-electron microscopy (cryo-EM) structures of the complete yeast Polζ holoenzyme on (3.1-Å resolution) and off (4.1-Å resolution) DNA (Table 1 and Extended Data Figs. 1–3). The structures unveil the mechanism by which Polζ synthesizes DNA and resolve the longstanding conundrum of how it differs from replicative polymerases to perform TLS.

Results

Ring-like architecture of Polζ. The holoenzyme consists of one catalytic Rev3 subunit, two Rev7 (Rev7α and Rev7β) accessory subunits, one Pol31 accessory subunit and one Pol32 accessory subunit assembled in a pentameric ring-like architecture with approximate dimensions of 136 Å × 111 Å × 60 Å (Fig. 1b–d). This ring-like architecture is unique among multisubunit DNA polymerases and resembles an uninterrupted daisy chain of protein–protein interac-
Fig. 1 | The architecture of DNA-bound Polζ holoenzyme. a, Schematic of the primary structure of S. cerevisiae Polζ subunits. Different colors denote each subunit. b, Near-atomic-resolution cryo-EM density map of DNA-bound Polζ holoenzyme, colored by local resolution. Scale bar is in Å. c, The three-dimensional (3D) reconstruction of Polζ holoenzyme viewed perpendicular (left) and parallel (right) to the DNA axis. A red dashed connector represents disordered Pol32c, and the arrowhead marks the putative interaction location with PCNA. d, Cryo-EM structure of DNA-bound Polζ colored by domain and viewed from the same orientations as in c.
The catalytic subunit Rev3 alone makes all of the contacts to the DNA (Figs. 1c,d and 2 and Supplementary Fig. 1). The duplex portion of the template-primer has a B-DNA-like conformation, with average helical twist and rise values of 29.7° and 3.11 Å, respectively. For convenience, we refer to positions of nucleotides by T_{N−P}, where T and P refer to the template and primer strands, respectively, and the subscript N refers to the number of base pairs from the templating position (Fig. 2a). At the replicative end, the templating base G (position T_{N}) establishes Watson–Crick (W–C) base pairing with incoming deoxycytidine triphosphate (dTCTP) (position P_{N}).

Unexpectedly, two Rev7 subunits bind in a unique head-to-tail arrangement, of which there is no example in the HORMA family. Pol31 and Pol32_{rev7} (residues 1−115) are held rigidly in a radial arrangement with respect to the DNA duplex. Pol32_{rev7} (residues 116−350) is disordered, but a proliferating cell nuclear antigen (PCNA) interaction (PIP) motif at its C terminus can potentially extend to the downstream portion of the DNA duplex for putative interactions with PCNA (Fig. 1ac).

### Catalytic Rev3 structure

Rev3 embraces the template-primer with its palm (residues 329−373; 941−1043; 1098−1215), fingers (residues 302−328; 1044−1097), thumb (residues 1216−1372), exonuclease (exo; residues 662−894), and N-terminal (NTD; residues 1−301; 374−400; 895−940) domains (Figs. 1c and 2b and Supplementary Fig. 1). The palm interacts with the replicative end of the template-primer and carries the active site residues (D975 and D1144) towards the major groove (Fig. 2b). The inactive exonuclease domain lies on the side of the DNA opposite the thumb, extending towards the major groove (Fig. 2b). The NTD bridges the exonuclease and fingers domains and makes numerous contacts with the unpaired portion of the template strand (Figs. 1d and 2b).

Rev3 makes a broad range of interactions with incoming dCTP weaves a path between the fingers and palm active sites27,28. Strikingly, the β-hairpin is almost non-existent in most B-family polymerases that drape over the nascent base pair24, whereas in Rev3 (spot), β-hairpin in the palm domain differs by an extra helix (αG) and a long loop (palm loop) that drops across the connecting linker between the NTD and the palm (NTD–palm linker) (Fig. 3b). The fingers domain contains three α-helices (αF, αA and αB) (Fig. 2b). αA and αB are analogous to the two long α-helices in B-family polymerases that drape over the nascent base pair24, whereas αF is unique to Polζ and involved in contacts with the NTD. The NTD is much more extended than it is in Pol3, and contains additional secondary structural elements and loops that make contacts with the other domains of Rev3 (Extended Data Fig. 4).

The large insert in Rev3 between the NTD and the exonuclease domain comprises the Rev7 interaction region (RIR; residues 401−661). The RIR is disordered except for the region composed of residues 513−624, which contains the two Rev7-binding motifs (RBM1, residues 517−540; RBM2, residues 599−623) that interact with the Rev7α and Rev7β subunits, respectively (Figs. 1d and 2b and Supplementary Fig. 1). Of the two cysteine-rich metal-binding modules at the C-terminal end of Rev3, CysBD (residues 1381−1418) is disordered, whereas CysAD (residues 1419−1504) is ordered and serves to recruit Pol31 (Fig. 1c). CysBD contains a well-resolved δ-factor

---

### Table 1 | Cryo-EM data collection, refinement and validation statistics

|                     | Pol1−DNA-dCTP (EMD-21115, PDB 6V93) | Pol1−apo state (EMD-21108, PDB 6V8P) |
|---------------------|-------------------------------------|--------------------------------------|
| **Data collection and processing** |                                     |                                      |
| Magnification       | x22,500                             | x105,000                             |
| Voltage (kV)        | 300                                 | 300                                  |
| Electron exposure (e⁻/Å²) | 71.63                              | 87.62                                |
| Defocus range (µm) | −0.5 to −2.5                        | −1.5 to −2.5                         |
| Pixel size (Å)      | 0.537                               | 1.1                                  |
| Symmetry imposed    | C_{1}                               | C_{1}                                |
| Initial particle images (no.) | 205,914                           | 1,784,587                            |
| Final particle images (no.) | 156,067                           | 311,800                              |
| Map resolution (Å)  | 3.1                                 | 4.1                                  |
| FSC threshold       | 0.143                               | 0.143                                |
| Map resolution range (Å) | 2.5−5.0                           | 4.0−6.0                              |
| **Refinement**      |                                     |                                      |
| Initial model used  | PDB: 31AY, 3EOJ, 3ABD               | PDB: 6V93                           |
| Model resolution (Å) | 3.2                                | 4.3                                  |
| FSC threshold       | 0.5                                 | 0.5                                  |
| Map sharpening β factor (Å²) | −100                               | −224                                 |
| **Model composition** |                                     |                                      |
| Non-hydrogen atoms  | 18,105                              | 14,552                               |
| Protein residues    | 2,198                               | 2,117                                |
| DNA/other           | 27/3                                | 0/1                                  |
| Water               | 220                                 | 0                                    |
| **B factors (Å²)**  |                                     |                                      |
| Protein             | 57.06                               | 173.97                               |
| DNA/other           | 76.01/147.63                        | −132.86                              |
| Water               | 58.22                               | −                                    |
| r.m.s. deviations   |                                     |                                      |
| Bond lengths (Å)    | 0.002                               | 0.002                                |
| Bond angles (°)     | 0.519                               | 0.483                                |
| **Validation**      |                                     |                                      |
| MolProbity score    | 1.76                                | 1.60                                 |
| Clashscore          | 5.76                                | 3.96                                 |
| Poor rotamers (%)   | 0                                   | 0                                    |
| Ramachandran plot   |                                     |                                      |
| Favored (%)         | 93.00                               | 93.64                                |
| Allowed (%)         | 6.96                                | 6.31                                 |
| Disallowed (%)      | 0.05                                | 0.05                                 |
Fig. 2 | Structure and cryo-EM density details of Rev3. a, Close-up view of the active site of Rev3 depicting key residues forming the T0–P0 binding site, including ligands and metal ions. Highlighted on the right is the well-resolved density for the T1 and T2 positions (red) and the sequence of the palindromic DNA employed to form the ternary complex. The region of the template-primer duplex enclosed in the box was built into the final model. b, Structure of Rev3 colored by domain. Dark blue, brown, magenta, cyan, yellow, orange and gray denote, respectively, the N-terminal, RIR, exo, palm, fingers, thumb and C-terminal domains of Rev3. Cryo-EM density for selected regions of Rev3 that highlight the differences in sequence with Pol3 (PDB ID: 3IAY), including residues in the inactive exo domain, residues in close proximity to the CysBD and the near absence of the β-hairpin region in Rev3, are shown. Also shown are close-up views of the coordination around the 4Fe–4S cluster and the interfaces between the RIR and the palm and thumb domains.
Fig. 3 | Structural basis for fidelity and mismatch extension. a, Surface representation of a close-up view of the active sites of Rev3 and Pol3, depicting conserved residues interacting with the incoming nucleotide (dCTP) as well as the templating base (G). b, Superimposition of the palm domains of Rev3 and Pol3. The Rev3 template DNA strand is shown in red and the T₁ base is highlighted. In comparison with Pol3, the palm loop (β25 and β26) is a unique Rev3 structural element that interacts with the NTD–palm linker. Key residues in the NTD–palm linker are shown as sticks interacting with the palm-loop residues as well as the αxC palm helix. Another unique structural element, αG, which interacts with β27 and β28, is also highlighted. c, Superimposition of the exo domains of Rev3 and Pol3. Overlay of the exo domains of Rev3 and Pol3 shows a much shorter and disordered helical loop in Rev3 in comparison with a well-defined β-hairpin in Pol3. The Pol3 DNA is highlighted in red. d, Comparison of the T₁ binding site in Rev3 and Pol3, showing details of the interaction between the NTD–palm linker region and the T₁ base. The residue E954 in Rev3 points away from the T₁ base, whereas residue Y587 stacks against the sugar in Pol3. L953 in Rev3 is the only residue close to the T₁ base, resulting in a less constrained pocket in comparison with that in Pol3.
domains and draws the two domains together (Fig. 2a). Two calcium ions (A and B) are located between the triphosphate tail and the primer terminus, analogous to metals ‘A’ and ‘B’ in other DNA polymerases. Although calcium inhibits DNA polymerase activity, the positions of the catalytic residues (D975 and D1144) and the metal ions are appropriate for a two-metal mechanism of catalysis (Fig. 2a).
Fidelity and mismatch extension. The structure provides insights into why Pol3ζ is unable to incorporate nucleotides opposite DNA lesions, but is able to extend synthesis from mismatches and DNA lesions14. From the structure, Rev3’s inability to insert nucleotides opposite DNA lesions is due to residues L1087, N1090, V1091, Y1093 and G1094 from the fingers domain, and Y980 from the palm domain (Figs. 2a and 3a). L1087 and N1090 fit snugly atop the templating base, while Y980, Y1093 and G1094 impinge on the nascent base pair from the minor groove side (Figs. 2a and 3a). The contacts are primarily van der Waals in nature, and together they impose a strong preference for a W–C base pair at T1P0. These contacts are remarkably similar to those observed with Pol3 (Fig. 3a), suggesting that Pol3 operates in much the same way as a high-fidelity replicative polymerase during the nucleotide-incorporation step, sharing (in a steric sense) the same intolerance for mismatches and DNA lesions at the incipient T1P0 position.

The ability of Pol3ζ to tolerate mismatches and lesions at the T1-P1 position sets it apart from all other eukaryotic B-family polymerases23–25. We trace the ability of Pol3ζ to tolerate DNA distortions at the T1-P1 position and to extend synthesis from the aberrant junction to divergence in the path of the NTD–palm linker and its sequestration by the palm loop (Fig. 3b).

In Pol3, the NTD–palm linker spans the width of the template-primer, with Y587 lying flush against the sugar of the T1 nucleotide23 (Fig. 3d and Extended Data Fig. 5). Y587 is the only amino acid that makes direct contacts with the T1 nucleotide, whereas all of the other contacts are water-mediated. In Rev3, Y587 is replaced by a glutamate (E954), but most importantly, this segment of the NTD–palm linker traces a different path than in Pol3, in particular in the region that abuts against the T1 nucleotide in Pol3 (Fig. 3b). The linker is fixed in this alternative position in Rev3 by interaction with the palm loop, as well as helix αC of the palm domain. Notably, the palm loop is a unique structural feature of the Rev3 palm domain. The net result of this movement in the NTD–palm linker is the creation of additional space around the T1 nucleotide that, in principle, can more easily contain DNA distortion and deviations from W–C geometry at the T1-P0 position (Fig. 3b,d and Extended Data Fig. 5).

Pol3ζ incorporates two Rev7 subunits in a new head-to-tail arrangement. The presence of Rev7 in Pol3ζ is the main difference between Pol6 in terms of subunit composition13,26. Rev7 is a member of the HORMA family of proteins that can exist in topologically distinct open (O) and closed (C) states13,30,11. Until recently, Pol3ζ was thought to incorporate only a single copy of Rev7 (ref. 1). We show here that Pol3ζ contains in fact two copies of Rev7 (Fig. 1c,d), with the holoenzymes assembled as a five-subunit (Rev3–Rev7A–Pol31–Pol32) complex. The Rev7 monomers adopt a new head-to-tail arrangement unlike anything seen previously with other HORMA proteins11.

The N-terminal approximately two-thirds of Rev7 (residues 1–148) comprises the ‘core’, dominated by a three-stranded β-sheet (β4, β5 and β6) flanked by three α-helices (αA, αB and αC) from one side (Fig. 4a). The C-terminal region (residues 149–245; encompassing β6) is the ‘seatbelt’ that can adopt different topological conformations: packing (as strands β7 and β8) against the β6 side of the core in the open state, but rearranging (as strands β8’ and β8”) to the β5 side of the core in the closed state and encircling the bound peptide13,30,31 (Fig. 4a).

In Pol3ζ, both Rev7 subunits are in the closed state, with Rev7ζ binding RBM1 and Rev7ζ binding RBM2. However, the arrangement is fundamentally different from that observed with other HORMA proteins (Fig. 4a). Mad2, for example, has been resolved as both a symmetric (C-Mad2–C-Mad2)32 and an asymmetric (C-Mad2–O-Mad2) dimer (Fig. 4a)31, but the arrangement in both cases is antiparallel or head-to-head with the dimer interface dominated by helix αC from each monomer (the head is defined as the side of HORMA protein containing helix αC). By contrast, Rev7ζ and Rev7ζ adopt a parallel or head-to-tail arrangement, resulting in the C helices (one from each monomer) being far apart (Fig. 4a).

As such, the Rev7ζ–Rev7ζ interface is entirely symmetric with helix αC and the β2–β3 pseudo hairpin of Rev7ζ making hydrogen bonds and van der Waals contacts with the seafloor domain of Rev7ζ. The interface is rather sparse, burying ~524 Å2 of solvent-accessible surface area, as compared with 1,960 Å2 in C-Mad2–O-Mad2, with the Rev7 dimeric arrangement stabilized additionally by contacts with Rev3 as well as Pol31 and Pol32ζ (Fig. 4). As such, the Rev7ζ and Rev7ζ subunits act as a bridge between the catalytic Rev3 and the accessory Pol31 and Pol32 subunits (Figs. 1c,d and 4b).

Rev3-Rev7 interactions. Rev7ζ and Rev7ζ bind simultaneously to RBM1 and RBM2, capturing a HORMA dimeric arrangement with a peptide bound to each monomer (Fig. 4a). RBM2 folds into a β-strand, linker and an α-helix, whereas RBM1 consists of only the linker and α-helix (the β-strand in RBM1 and a portion of the seafloor of Rev7ζ, which are weakly distorted and are mostly disordered). The two RBMs contain consecutive prolines (P526 and P527 in RBM1 and P610 and P611 in RBM2) that are central in interactions with Rev7ζ and Rev7ζ (Fig. 4b). The aliphatic rings of P526 and P610 make hydrophobic contacts with the aromatic rings of Y57 and F141 of Rev7, whereas the aliphatic ring of P527 and P611 make hydrophobic contacts with the side chain of L54 (Fig. 4b and Extended Data Fig. 6). These hydrophobic contacts are augmented by a hydrogen bond between the main-chain carbonyl of P611 and the hydroxyl of Y27 of Rev7ζ. Overall, these interactions are similar to those observed in the structure of human Rev7 with a human Rev3 RBM1 or RBM2 peptide13,33, and are likely to extend to Pol3ζ from other eukaryotic species.

Strikingly, the entire segment of RIR connecting RBM1 and RBM2 (residues 513–624) is well-defined in our structure, weaving a path between Rev7ζ and Rev7ζ, as well as making contacts with the palm and thumb domains (Fig. 2b). Contacts with Rev7ζ are extensive, with amino acids such as V583, V590, V592 and F585 making hydrophobic contacts with residues from helices αA and αB of Rev7ζ (Fig. 4b and Extended Data Fig. 6). Overall, these contacts supplement (and even exceed) those made by RBM1 and RBM2 and appear to be critical in stabilizing the Rev7ζ–Rev7ζ homodimer in the non-canonical head-to-tail arrangement within Pol3ζ.

Pol31 and Pol32. Considering their different roles in DNA replication and repair, it is quite remarkable that Pol3ζ and Pol6 share the same accessory Pol31 and Pol32 subunits. Pol31 and Pol32ζ do not engage the DNA in either polymerase (Fig. 5a), despite containing domains (an oligonucleotide/oligosaccharide-binding (OB) fold and an inactive phosphodiesterase (PDE) domain) in Pol31 and a winged helix-turn-helix domain in Pol32ζ that are potentially capable of binding DNA (Figs. 1c,d and 5a). In both polymerases, CysBD at the C-terminus of the catalytic subunit interacts with both the OB and PDE domains of Pol31 (Fig. 5a), though the contacts vary between Pol3ζ and Pol6 (ref. 14) (Fig. 5b). The size and structure of CysBD is, however, similar in Rev3 and Pol3, and is composed of two long antiparallel α-helices and a 4Fe–4S cluster (Extended Data Fig. 7). By contrast, the equivalent domains in the catalytic subunits of Pol6 and Pole are larger and observed to bind to a single divalent Zn2+ ion14–16.

Although Pol31 and Pol32ζ are situated radially with respect to the DNA in Pol3ζ and Pol6, their exact position and orientation differs in the two enzymes (Fig. 5a). In particular, to make room for Rev7ζ in Pol3ζ, Pol31 and Pol32ζ rotate by ~18° and translate by ~10 Å relative to their positions in Pol6 (Fig. 5a). This motion is along a direction roughly perpendicular to the DNA axis, and it positions Pol31ζ closer to the exonuclease domain in Pol3ζ than to that in
Polδ (Fig. 5a). Thus, whereas there is a sizeable gap at the interface between Pol31OB and the exonuclease domain in Polδ, the interface is more tightly packed in Polζ and includes an α-helix (αL) from Rev3 not present in Polδ (Fig. 5a).

Overall, Pol31 and Pol32N are held much more rigidly than in Polδ (ref. 19). Besides contacts with CysBD and the exonuclease domain of Rev3 (Fig. 5b), Pol31 and Pol32N also make numerous contacts with Rev7N (Fig. 4b), with ~695 Å² of surface area buried at this interface (Fig. 5b). This rigidity is highlighted by the resolution of Pol31 and Pol32N in the Polζ cryo-EM map, which is comparable with that of Rev3 and Rev7 (Fig. 1b). By contrast, in the Polδ cryo-EM structure19, the resolution of Pol31 and Pol32N was relatively low and improved only after multibody refinement (reflecting flexibility).

Conformational changes upon DNA binding. A hallmark of replicative DNA polymerases is a conformational change in the fingers domain, from an open to a closed state, on correct dNTP
binding. The apo Polζ structure reveals the fingers domain in the open conformation, with helices αF, αA and αB rotated outwards by 15° compared with their positions in the ternary complex (Fig. 6). Thus, analogous to replicative polymerases, Polζ’s fidelity during the nucleotide-incorporation step appears to be augmented by the opening and closing of the fingers domain. Also, in the apo structure, parts of the thumb domain that contact the DNA minor groove in the ternary complex are disordered or located away from the DNA (Fig. 6). Thus, many of the positively charged residues that interact with the DNA sugar-phosphate backbone in the ternary complex are either completely disordered (such as K1280 and K1283) or are far away from the DNA and partially disordered (such as K1272, K1273, R1309 and R1357) (Fig. 6).

Discussion

Polζ stands out as the central DNA polymerase for the extension step in the bypass of the vast majority of DNA lesions formed in eukaryotic cells. For decades, structural studies of the Polζ holoenzyme, or even just its large catalytic subunit Rev3, have been hampered by low yields and the unattainability of well-diffracting crystals. Here, we employ cryo-EM to present the near-atomic-resolution structures of the complete yeast Polζ holoenzyme, with and without bound DNA. The structures reveal a pentameric ring-like architecture for Polζ, whereby the subunits form an uninterrupted daisy chain of protein–protein interactions.

The inability of Polζ to insert nucleotides opposite DNA lesions is readily understood from the DNA-bound structure. The juxtaposition between the fingers helices and the nascent base pair (position T1–P1) is remarkably similar to that observed with Pol3 (refs. 42,43), and the sheer density of these contacts is incompatible with DNA distortion or deviations from W–C geometry at the T1–P1 position. In contrast to Rev3, DNA Pol II has proofreading activity and a β-hairpin44, but an alteration in the position of the β-hairpin has been suggested to increase the dwell time of the DNA substrate in the Pol II polymerase active site for TLS to occur45. The near absence of the β-hairpin in Rev3 should, in principle, increase the dwell or residence time of a DNA substrate in the Rev3 polymerase active site, which coupled to fewer overall contacts to the template strand may increase the opportunity for TLS to occur. In future studies, it will be interesting to probe the consequences on TLS of altering the lengths and sequences of the NTD–palm linker and the β-hairpin in Rev3.

Rev7 is the only subunit of Polζ that does not have a counterpart in other B-family polymerases, such as Polx, Pole and Polβ. Until recently, the identification of a single RBM in human Rev3 (residues 1877–1887) pointed to human Polζ incorporating only 1 copy of Rev7 (ref. 1). In 2015, a second RBM was characterized in human Rev3 (residues 1993–2003)46, leading to the suggestion that Polζ might actually incorporate 2 copies of Rev7 (refs. 1,46). The equivalent RBM motifs had been difficult to discern in yeast Rev3, but from the structure we can now identify them as residues 517–540 (RBM1) and 599–623 (RBM2). The two RBMs bind simultaneously to the two Rev7 subunits, which captures a HORMA dimeric arrangement with a peptide bond to each monomer. Rev7a and Rev7b arrange in a non-canonical head-to-tail configuration, lending to an interface that is fundamentally different from that observed in C-Mad2–O-apoMad2 or C-apoMad2–C-apoMad2 (refs. 42,43) (Fig. 4a and Extended Data Fig. 6). Notably, the solvent-accessible surface area buried at the Rev7a–Rev7b interface is markedly less than that typically observed in oligomeric proteins47, implying that the Rev7a–Rev7b homodimer is stabilized in the head-to-tail arrangement (within Polζ) by additional contacts that Rev7b establishes with the RIR region connecting RBM1 and RBM2, and that Rev7a establishes with Pol31 and Pol32N (Figs. 1d and 4b). Interestingly, Rev7 has been proposed to homodimerize independently of Polζ (ref. 48), although the dimer interaction appears to be weak compared with that of Mad2 (ref. 49). Rev7a has also been suggested, from biophysical studies, to dimerize in a canonical head-to-head arrangement when in complex with RBM1 and RBM2 of human Rev3 (ref. 1). An intriguing question for future work is whether Rev7 dimerizes in a head-to-tail or a head-to-head configuration independently of Polζ. An interesting possibility is that Rev7 forms weak head-to-head dimers independently of Polζ, but switches to a non-canonical head-to-tail arrangement in the context of Polζ, following interactions with RBM1 and RBM2, the RIR between RBM1 and RBM2, and Pol31 and Pol32N.

What is the role of Rev7a and Rev7b in Polζ activity? The two subunits emerge from our structure as the organizing center of Polζ. Most importantly, the incorporation of two copies of Rev7 in Polζ increases the potential surface area available for interactions with other components of the TLS machinery, including DNA polymerase Rev1. The C-terminal domain of Rev1 (Rev1-CTD) can interact simultaneously with Rev7 and Y-family polymerases and serves to link the master extender Polζ to inserters Polη, Polκ or Polλ for lesion bypass. Intriguingly, when we model Rev1-CTD on Rev7a and Rev7b, only Rev7a is capable of accommodating Rev1
(Extended Data Fig. 9). This stoichiometry is consistent with a biophysical study showing that only a single Rev1-CTD is capable of binding human Rev7 when tethered to RBM1 and RBM2 (ref. 18). We find potentially stabilizing interactions between Rev1-CTD and Pol32N (Extended Data Fig. 9), complementing evidence that Rev1-CTD can interact with a portion of Pol32C (ref. 48). In all, the ability of Rev7 to assemble as a head-to-tail homodimer has major implications for how Polζ interacts with other components of the TLS machinery.

Polζ has emerged as an important determinant for tumor resistance to chemotherapeutic agents and radiotherapy in several types of cancers6,8,9. Indeed, in mouse cancer models, depletion of
Rev3 sensitizes non-small-cell lung and prostate cancers to conventional chemotherapy\textsuperscript{40,49}. Similarly, depletion of Rev7 sensitizes ovarian cancer cells to chemotherapy\textsuperscript{41}, and enhances the sensitivity of glioma cells to ionizing radiation\textsuperscript{42}. However, the development of inhibitors of Pol\textsubscript{ζ} has been hampered by the lack of structural information on the polymerase. As such, most of the effort to date has been directed at the Rev7–RB1M and Rev7–Rev1 interfaces for which structural information is available\textsuperscript{50,51}. The structure of Pol\textsubscript{ζ} presented here will spur new efforts to develop inhibitors of this unique polymerase. We identify several new protein–protein interfaces that can potentially serve as models for targeting by small molecules or stapled peptides, including the Rev\textsubscript{7}–Pol31, Rev\textsubscript{7}–Pol32, Rev\textsubscript{7}–RIR and Pol31\textsubscript{on}–exonuclease interfaces. The structure of Rev3 itself offers new opportunities to directly target the catalytic activity of Pol\textsubscript{ζ}. The Rev3 fingers and palm domains and the NTD are markedly different from those in other polymerases and may provide a basis for the selective binding of small molecules to Rev3. Although human Rev3 (containing 3,130 amino acids) is about twice the length of yeast Rev3, this difference is due almost entirely to increase in the length of the RIR, which, except for RBM1 and RBM2, is predicted to be disordered. Indeed, deletion of ~1,000 residues (from amino acids 526 to 1588) of human Rev3’s RIR has no marked effect on the biochemical activities of human Pol\textsubscript{ζ} (ref. \textsuperscript{33}). Importantly, all of the domains and protein–protein interaction modalities identified here are also present in human Pol\textsubscript{ζ}.

Altogether, we unveil here for the first time the pentameric ring-like architecture of Pol\textsubscript{ζ} and identify structural elements that allow this unique polymerase to synthesize DNA and perform TLS. The structures provide an unprecedented new framework for genetic and biochemical studies aimed at understanding the role of Pol\textsubscript{ζ} in protecting organisms from environmental and cellular genotoxic stresses, and a framework for the discovery of therapeutics in the treatment of chemotherapy-resistant tumors.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41594-020-0476-7.

Received: 23 April 2020; Accepted: 29 June 2020; Published online: 17 August 2020

References
1. Prakash, S., Johnson, R. E. & Prakash, L. Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. Annu. Rev. Biochem. 74, 317–353 (2005).
2. Jain, R., Aggarwal, A. K. & Rechkhobit, O. Eukaryotic DNA polymerases. Curr. Opin. Struct. Biol. 53, 77–87 (2018).
3. Makarova, A. V. & Burgers, P. M. Eukaryotic DNA polymerase ζ. DNA Repair (Amst.) 29, 47–55 (2015).
4. Martin, S. K. & Wood, R. D. DNA polymerase ζ: the role of the catalytic subunit in cellular signaling. J. Cell Biol. 211, 745–755 (2015).
5. Baranovskiy, A. G. et al. DNA polymerase δ and ζ switch by sharing accessory subunits of DNA polymerase δ. J. Biol. Chem. 287, 17281–17287 (2012).
6. Johnson, R. E., Prakash, L. & Prakash, S. Pol31 and Pol32 subunits of yeast DNA polymerase δ are also essential subunits of DNA polymerase ζ. Proc. Natl Acad. Sci. USA 109, 12455–12460 (2012).
7. Makarova, A. V., Stodola, J. L. & Burgers, P. M. A four-subunit DNA polymerase ζ complex containing Pol δ accessory subunits is essential for PCNA-mediated mutagenesis. Nucleic Acids Res. 40, 11618–11626 (2012).
8. Hara, K. et al. Crystal structure of human REV7 in complex with a human REV3 fragment and structural implication of the interaction between DNA polymerase ζ and REV1. J. Biol. Chem. 285, 12299–12307 (2010).
9. Rizzo, A. A. et al. Rev7 dimerization is important for assembly and function of the Rev1/Polδ translesion synthesis complex. Proc. Natl Acad. Sci. USA 115, E191–E200 (2018).
10. Jain, R. et al. Cryo-EM structure and dynamics of eukaryotic DNA polymerase δ holozyme. Nat. Struct. Mol. Biol. 26, 955–962 (2019).
11. Baranovskiy, A. G. et al. X-ray structure of the complex of regulatory subunits of human DNA polymerase δ. Cell Cycle 7, 3026–3036 (2008).
12. Lee, Y. S., Gregory, M. T. & Yang, W. Human Pol ζ purified with accessory subunits is active in translesion DNA synthesis and complements Pol η in cisplatin bypass. Proc. Natl Acad. Sci. USA 111, 2954–2959 (2014).
13. Gómez-Llortere, Y. et al. The architecture of yeast DNA polymerase ζ. Cell Rep. 7, 79–86 (2013).
14. Swan, M. K., Johnson, R. E., Prakash, L., Prakash, S. & Aggarwal, A. K. Structural basis of high-fidelity DNA synthesis by yeast DNA polymerase δ. Nat. Struct. Mol. Biol. 16, 979–986 (2009).
15. Doublel, S. & Zahn, K. E. Structural insights into eukaryotic DNA replication. Front. Microbiol. 5, 444 (2014).
16. Burgers, P. M. I. & Kunkel, T. A. Eukaryotic DNA replication fork. Annu. Rev. Biochem. 86, 417–438 (2017).
17. Hogg, M., Aller, P., Konigsberg, W., Wallace, S. S. & Doublel, S. Structural and biochemical investigation of the role in proofreading of a β hairpin loop found in the exonuclease domain of a replicative DNA polymerase of the B family. J. Biol. Chem. 282, 1432–1444 (2007).
18. Stocki, S. A., Nonay, R. L. & Reha-Krantz, L. J. Dynamics of bacteriophage T4 DNA polymerase function: identification of amino acid residues that affect switching between polymerase and 5′ → 5′ exonuclease activities. J. Mol. Biol. 254, 15–28 (1995).
19. Siltez, T. A. DNA polymerases: structural diversity and common mechanisms. J. Biol. Chem. 274, 13795–13798 (1999).
20. Mapelli, M. & Musacchio, A. MAD contortions: conformational dimerization boosts spindle checkpoint signaling. Curr. Opin. Struct. Biol. 17, 716–725 (2007).
21. Luo, X. & Yu, H. Protein metabolism: the two-state behavior of Mad2. Structure 16, 1616–1625 (2008).
22. Yang, M. et al. Insights into Mad2 regulation in the spindle checkpoint revealed by the crystal structure of the symmetric Mad2 dimer. PLoS Biol. 6, e50 (2008).
23. Mapelli, M., Massimiliano, L., Santaguida, S. & Musacchio, A. The Mad2 conformational dimer: structure and implications for the spindle assembly checkpoint. Cell 131, 730–743 (2007).
24. Baranovskiy, A. G. et al. Crystal structure of the human Pole B subunit in complex with the C-terminal domain of the catalytic subunit. J. Biol. Chem. 292, 15717–15730 (2017).
25. Klingle, S., Núñez-Ramírez, R., Jorca, O. & Pellegrini, L. 3D architecture of DNA Pol α reveals the functional core of multi-subunit replicative polymerases. EMBO J. 28, 1978–1987 (2009).
26. Suwa, Y. et al. Crystal structure of the human Pol β subunit in complex with the C-terminal domain of the catalytic subunit. J. Biol. Chem. 290, 14328–14337 (2015).
27. Joyce, C. M. & Benkowski, S. J. DNA polymerase fidelity: kinetics, structure, and checkpoints. Biochemistry 43, 14317–14324 (2004).
28. Doublel, S., Savaya, M. R. & Ellenberger, T. An open and closed case for all polymerases. Structure 7, R31–R35 (1999).

NATURE STRUCTURAL & MOLECULAR BIOLOGY | VOL 27 | OCTOBER 2020 | 913–924 | www.nature.com/nsmb
39. Wang, F. & Yang, W. Structural insight into translesion synthesis by DNA Pol II. *Cell* **139**, 1279–1289 (2009).
40. Tomida, J. et al. REV7 is essential for DNA damage tolerance via two REV3L binding sites in mammalian DNA polymerase ζ. *Nucleic Acids Res.* **43**, 1000–1011 (2015).
41. Janin, J., Miller, S. & Chothia, C. Surface, subunit interfaces and interior of oligomeric proteins. *J. Mol. Biol.* **204**, 155–164 (1988).
42. Hara, K. et al. Purification, crystallization and initial X-ray diffraction study of human REV7 in complex with a REV3 fragment. *Acta Crystallogr. F Struct. Biol. Cryst. Commun.* **65**, 1302–1305 (2009).
43. Hanafusa, T. et al. Overlapping in short motif sequences for binding to human REV7 and MAD2 proteins. *Genes Cells* **15**, 281–296 (2010).
44. Wojtaszek, J. et al. Structural basis of Rev1-mediated assembly of a quaternary vertebrate translesion polymerase complex consisting of Rev1, heterodimeric polymerase (Pol) ζ, and Pol κ. *J. Biol. Chem.* **287**, 33836–33846 (2012).
45. Pustovalova, Y., Bezsonova, I. & Korzhnev, D. M. The C-terminal domain of human Rev1 contains independent binding sites for DNA polymerase η and Rev7 subunit of polymerase ζ. *FEBS Lett.* **586**, 3051–3056 (2012).
46. Pozhidaeva, A. et al. NMR structure and dynamics of the C-terminal domain from human Rev1 and its complex with Rev1 interacting region of DNA polymerase η. *Biochemistry* **51**, 5566–5570 (2012).
47. Ohashi, E. et al. Identification of a novel Rev1-interacting motif necessary for DNA polymerase κ function. *Genes Cells* **14**, 101–111 (2009).
48. Pustovalova, Y. et al. Interaction between the Rev1 C-terminal domain and the PolD3 subunit of Polζ suggests a mechanism of polymerase exchange upon Rev1/Polζ-dependent translesion synthesis. *Biochemistry* **55**, 2043–2053 (2016).
49. Doles, J. et al. Suppression of Rev3, the catalytic subunit of Polζ, sensitizes drug-resistant lung tumors to chemotherapy. *Proc. Natl Acad. Sci. USA* **107**, 20786–20791 (2010).
50. Xu, X. et al. Enhancing tumor cell response to chemotherapy through nanoparticle-mediated codelivery of siRNA and cisplatin prodrug. *Proc. Natl Acad. Sci. USA* **110**, 18638–18643 (2013).
51. Niimi, K. et al. Suppression of REV7 enhances cisplatin sensitivity in ovarian clear cell carcinoma cells. *Cancer Sci.* **105**, 545–552 (2014).
52. Zhao, J. et al. Mitotic arrest deficient protein MAD2B is overexpressed in human glioma, with depletion enhancing sensitivity to ionizing radiation. *J. Clin. Neurosci.* **18**, 827–833 (2011).
53. Wojtaszek, J. L. et al. A small molecule targeting mutagenic translesion synthesis improves chemotherapy. *Cell* **178**, 152–159.e11 (2019).
54. Actis, M. L. et al. Identification of the first small-molecule inhibitor of the Rev7 DNA repair protein interaction. *Bioorg. Med. Chem.* **24**, 4339–4346 (2016).

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020
Methods

Protein expression and purification. *S. cerevisiae* Polζ holoenzyme, composed of the full-length Rev3 (residues 1–1,504), Rev7 (residues 1–245), Pol31 (residues 1–487) and Pol32 (residues 1–350) subunits, was expressed in yeast from plasmid pBJ1462 and purified as described.† Rev3 and Pol31 subunits harboring PreScission protease cleavable FLAG-MAT and glutathione S-transferase (GST) tags, respectively. Protein was purified as described with several modifications. In brief, yeast cells were lysed with a mechanical bead beater using 0.5-mm precoated zirconia beads. Lysis was done in the presence of β-mercaptoethanol and the protease inhibitors benzamidine, leupeptin, aprotinin and pepstatin. A. After lysis, polyethyleneimine was added to a final concentration of 0.04%. Proteins in the clarified cell extract were precipitated with 48% ammonium sulfate. Pellets containing the Polζ holoenzyme were then solubilized and purified over manually packed Glutathione Sepharose (GE Healthcare) column and the bound protein was eluted with 40 mM glutathione. The eluate was then passed through an anti-FLAG agarose (Sigma) column. The GST and FLAG tags were cleaved with GST-PreScission Protease and free GST derived from the Pol31 subunit and PreScission Protease were removed by further incubation with GST beads. The protein was concentrated and purified (Supplementary Fig. 2) over a Superose-6 Gel Filtration (GE Healthcare) column. The complex with DNA was prepared by incubating the Polζ holoenzyme in CaCl2 supplemented buffer with a HPLC-purified (Integrated DNA technologies) palindromic DNA (5′-TAATGGTAGGGGGAATCCCTCCTCCCTAC-3′) added in 1.5 molar excess, yielding G as the templating base. The addition of incoming dCTP (1 mM) allowed the capture of Polζ in the act of DNA synthesis. Final buffer for the ternary complex in vitreous ice was 200 mM NaCl, 5 mM CaCl2, 10 mM Tris pH 8.0, 2 mM MgCl2, 0.2 M sucrose, 2.2 mM glutathione, 2 mM TCEP. The sample of Polζ without DNA was obtained fortuitously. The Polζ holoenzyme was prepared with a short primer-template (11 nt–16 nt; 5′-TAACCGGTTTC-3′–5′-CTCTTGAACCGGGTTA-3′) in the absence of incoming nucleotide but a subsequent cryo-EM map, at a nominal resolution of 4.1 Å, revealed it as structure of the apo holoenzyme (described below). The switch from a short DNA fragment to a long DNA fragment required pico-liter sample volumes for vitrification (https://www.sptlabtech.com/solarus). Polζ was tilted to an angle of 40° and data were collected following a recently reported strategy. As a part of this strategy, we compensated for the loss of signal-to-noise ratio, resulting from the increase at high tilt angles in the path of electrons through vitreous ice, by recording movies at a frame rate of 200 ms for 12 s (yielding a total accumulated dose of 87.62 e−/Å2). Initial frames were collected with the microscope operating an energy filter. Another consequence of the high tilt was, in general, a notable increase in drift between frames compared to the unitless data. Frame alignment at a single pixel level with MotionCor2 improved data quality. CTF estimation was done on a per-particle basis using GCTF7 to account for particles at different Z′-heights due to the specimen tilt.

Image processing and 3D reconstruction. Particle picking for the ternary complex was initially done using template-based picking (FindEM) in Appion with reprojections from the negative-stain reconstruction of Polζ (ref. 2). Initial processing of micrographs from the first session in cryoSPARC3 resulted in a 3D model with a sphericity of 0.93 (Extended Data Fig. 3). From the CTF estimates shown (Fig. 1b) were calculated using ResMap73. Frame alignment at a single pixel level with MotionCor2 improved data quality. Particle picking from the apo-state of Polζ was done with FindEM using templates from the negative-stain Polζ reconstruction. Particles were picked on the basis of a relaxed threshold in an attempt to include all the possible particles on the micrographs. Particles were then extracted using RELION2 for further processing.

Re-extracted particles for the ternary complex were subjected to multiple rounds of 2D classification (2D) in cryoSPARC. A final set of 181,726 particles was subjected to an ab initio clean-up, which allowed for removal of particles corresponding to a low-resolution model with preferred orientation. The resulting particles were refined in cryoSPARC to give a 3D reconstruction at nominal resolution of 3.2 Å, based on the Fourier shell correlation (FSC) value of 0.143 between independently refined half-sets.8 The cryo-EM map was checked for directional anisotropy (https://3dsc.salk.edu) and gave a value of 0.96 out of 1. Processing in RELION3 with the Topaz picked particles also gave similar results and map quality. In order to improve the resolution further, separate masks for Rev3 and Rev7A–Rev7B–Pol31–Pol32N region, were generated using Slicer, an extension of UCSF Chimera9. Masks were applied to the volume and 3D refinement without particle extraction was done for each region. The resulting consensus maps were at a FSC1/2 of 3.02 Å and 3.08 Å for Rev3 and Rev7A–Rev7B–Pol31–Pol32N, respectively (Extended Data Fig. 2). The maps were then combined in PHENIX10 to generate a composite map which was used for model building. Local resolution estimates shown (Fig. 1b) were calculated using ResMap. As a part of apo Polζ particles extracted with RELION2 were subjected to iterative rounds of 2D classification in cryoSPARC (Extended Data Fig. 3). A set of 457,987 particles was selected from the 2D class averages, which were then used to do an ab initio clean-up to generate an initial model. This model from 357,328 particles was subjected to 3D refinement resulting in a 3D reconstruction with a sphericity of 0.92 out of 1.80,000 particles was exported from cryoSPARC using pyem script (https://github.com/asarnow/pyem). Per-particle CTF refinement was done in cisEM11 resulting in the cryo-EM map with a sphericity of 0.95 that used for model building (Extended Data Fig. 3).

Model building, refinement and analysis. The structure of the Pol domain of Rev3 (excluding residues 303–400) was built manually using the cryo-EM map of the ternary complex. Model building was done in COOT12 using a structure-based sequence alignment of Rev3 with Pol3 (PDB ID: 3AY1) as a guide. In comparison with Pol3, the well-resolved side-chain densities guide the building of various important regions, Rev3 NTD, exomesh, palm and the thymine dimerous density for 22 bases of the DNA, incoming nucleotide and metal ions helped in building them accurately. The density for the 4Fe–4S cluster and the CTD of Rev3 was also well-defined, especially for the region closest to the interface of Rev3 with Pol31, Pol32A, Pol31 and Pol32B, were built based on homology models from their human homologs, p50 and p66, (PDB ID: 35I9) as well as the complex of Pol1 CTD with the B-subunit in yeast Polk (PDB ID: 3FL0). We built Rev7 by using a homology model of human Rev7 (PDB ID: 3ABD) and placing it in the density for the domain next to Pol32B. Rev7 had weaker density in comparison with Rev7 but was identified as a Mad2 fold using BALBES.13 A careful analysis of the side-chain density confirmed it to be a second Rev7 monomer in the Polζ holoenzyme.

The refined cryo-EM structure of the ternary complex displays clear density (Fig. 1b–d) for almost all of the secondary structures and ~85% of the side chains, 22 of the 26 bases of the DNA duplex, templating base G, DCTP and 220 solvent molecules, as well as the active site metal ions. The model for the apo structure was built by fitting the coordinates of the ternary complex into the cryo-EM map using
COOT and Chimera. Manual adjustments and deletions of disordered regions were done in COOT.

Both models were refined using real-space refinement in PHENIX and validated using MolProbity and EMRinger. Figures were prepared using Chimera and PyMOL. Superimposition of structures was performed in COOT. Sequence alignments were done in Clustal Omega, PHYRE2 (ref. 23) and PROMALS3D. Buried surface area was calculated in Chimera.

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

**Data availability**

Accession codes for the Polζ–DNA–dNTP model and map are Protein Data Bank (PDB) 6V93 and EMD-21115, respectively. Accession numbers of the Polζ (apo form) model and map are PDB 6V97 and EMD-21108, respectively.

**References**

55. Green, A. A. & Hughes, W. L. Protein fractionation on the basis of solubility in aqueous solutions of salts and organic solvents. *Methods Enzymol.* 1, 67–90 (1955).
56. Jain, T., Sheehan, P., Crum, J., Carragher, B. & Potter, C. S. Spotiton: a prototype for an integrated inkjet dispenser and vitrification system for cryo-TEM. *J. Struct. Biol.* 179, 68–75 (2012).
57. Wei, H. et al. Optimizing “self-wicking” nanowire grids. *J. Struct. Biol.* 202, 170–174 (2018).
58. Lander, G. C. et al. Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.* 166, 95–102 (2009).
59. Suloway, C. et al. Automated molecular microscopy: the new Legion system. *J. Struct. Biol.* 151, 41–60 (2005).
60. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods* 14, 331–332 (2017).
61. Rohou, A. & Grigorieff, N. CTFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* 192, 216–221 (2015).
62. Tan, Y. Z. et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* 14, 793–796 (2017).
63. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* 193, 1–12 (2016).
64. Roseman, A. M. FindEM — a fast, efficient program for automatic selection of particles from electron micrographs. *J. Struct. Biol.* 145, 91–99 (2004).
65. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* 14, 290–296 (2017).
66. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* 180, 519–530 (2012).
67. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* 7, e42166 (2018).
68. Bepler, T. et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* 16, 1153–1160 (2019).
69. Scheres, S. H. W. & Chen, S. X. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* 9, 853–854 (2012).
70. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* 333, 721–745 (2003).
71. Pettersen, E. F. et al. UCSF chimera — a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612 (2004).
72. Klajnjak, B. P. Deriving and refining atomic models in crystallography and cryo-EM: the latest Phenix tools to facilitate structure analysis. *Acta Crystallogr. D Struct. Biol.* 75, 878–881 (2019).
73. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* 11, 63–65 (2014).
74. Grant, T., Rohou, A. & Grigorieff, N. cuTEM, user friendly software for single-particle image processing. *Elife* 7, e35383 (2018).
75. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132 (2004).
76. Brown, A. et al. Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D Biol. Crystallogr.* 71, 136–153 (2015).
77. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 66, 12–21 (2010).
78. Barad, B. A. et al. EMRinger: side chain directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* 12, 943–946 (2015).
79. DeLano, W. L. & Lam, J. W. PyMOL: a communications tool for computational models. *Abstr. Pap. Am. Chem. Soc.* 230, U1371–U1372 (2005).
80. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539 (2011).
81. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858 (2015).
82. Pei, J. & Grishin, N. V. PROMALS: towards accurate multiple sequence alignments of distantly related proteins. *Bioinformatics* 23, 802–808 (2007).

**Acknowledgements**

We thank B. Carragher, C. Potter and E. Eng for helpful advice and discussion throughout the project. We also thank Z. Zhang and D. Bobe for help in grid preparation, Y. Z. Tan for help in collecting tilted cryo-EM data, A. Brown and T. Terwilliger for help in implementing software and D. Nair for help in model building. This work was primarily funded by grant R01-GM124047 from the National Institutes of Health (NIH). I.U.-B. was supported by a grant P41OD019994 from the Spanish State Research Agency and by the Basque Excellence Research Centre program. Initial EM screening was performed at the Icahn School of Medicine microscope facility supported by a shared instrumentation grant from the NIH (1S1OR026473). Most of the cryo-EM work was performed at the Simons Electron Microscopy Center and National Resource for Automated Molecular Microscopy, located at the New York Structural Biology Center, supported by grants from the Simons Foundation (SF349247), NYSTAR and the NIH National Institute of General Medical Sciences (GM103310), with additional support from Agouron Institute (F00316), NIH (OD019994) and NIH (RR029300). Computing resources needed for this work were provided in part by the High Performance Computing facility of the Icahn School of Medicine at Mount Sinai. Molecular graphics and analyses were performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311.

**Author contributions**

A.K.A. conceived the project; A.K.A., I.U.-B. and R.M. designed the experiments; R.E.J. expressed Polζ; F.Y. Tan for help in grid preparation, T. Terwilliger for help in implementing software and D. Nair for help in model building. This work was primarily funded by grant R01-GM124047 from the National Institutes of Health (NIH). I.U.-B. was supported by a grant P41OD019994 from the Spanish State Research Agency and by the Basque Excellence Research Centre program. Initial EM screening was performed at the Icahn School of Medicine microscope facility supported by a shared instrumentation grant from the NIH (1S1OR026473). Most of the cryo-EM work was performed at the Simons Electron Microscopy Center and National Resource for Automated Molecular Microscopy, located at the New York Structural Biology Center, supported by grants from the Simons Foundation (SF349247), NYSTAR and the NIH National Institute of General Medical Sciences (GM103310), with additional support from Agouron Institute (F00316), NIH (OD019994) and NIH (RR029300). Computing resources needed for this work were provided in part by the High Performance Computing facility of the Icahn School of Medicine at Mount Sinai. Molecular graphics and analyses were performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311.

**Competing interests**

The authors declare no competing interests.

**Additional information**

Extended data is available for this paper at https://doi.org/10.1038/s41594-020-0476-7. Supplementary information is available for this paper at https://doi.org/10.1038/s41594-020-0476-7.

**Correspondence and requests for materials** should be addressed to I.U.-B. or A.K.A.

**Peer review information** Beth Moorefield was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

**Reprints and permissions information** is available at www.nature.com/reprints.
a. Preferred orientation of the Polζ-DNA-dCTP complex

b. Preferred orientation of the Polζ-Apo state

Extended Data Fig. 1 | Preferred specimen orientation. a. Data collected at 0° stage angle resulted in disproportionally low number of classes for side-views of the ternary complex of Polζ depicted in the 2D class averages. This resulted in a ‘smeared 3D model’ as shown by the anisotropic 3DFSC plot. Scale bar = 123 Å. b. Data collected at a stage angle of 0° for the apo-state of Polζ also had preferred set of views as shown in the 2D class averages. The final construction was anisotropic as depicted by the directional FSC plot. Scale bar = 123 Å.
a. Data collection on Chameleon grids and particle selection with FindEM template picker

b. Combining with a second dataset and particle selection with Topaz neural network picker

c. Focused refinement in cryoSPARC and final map

Extended Data Fig. 2 | See next page for caption.
Extended Data Fig. 2 | Cryo-EM data collection and processing of Polζ-DNA-dCTP complex. a. Data were collected on Chameleon grids and particles from one session were picked with template based picker (FindEM) and processed in cryoSPARC to give a consensus map with a FSC$_{0.143}$ of 3.57 Å. Major stages of processing are shown schematically and particles involved at each stage are highlighted in green. Scale bar = 137 Å. b. Final particles from two sessions were merged and used to train Topaz. Data processing from Topaz picked particles in cryoSPARC2 improved the sphericity. A schematic representation of the improved consensus map displaying a FSC$_{0.143}$ of 3.2 Å is shown. Scale bar = 137 Å. c. Focused refinement of the final volume was done in cryoSPARC2. Masks were created (along the blue dashed line) for 3D refinement of Rev3 and accessory subunits separately to give consensus maps at 3.02 Å and 3.08 Å, respectively.
Extended Data Fig. 3 | Cryo-EM data collection and processing for Polζ apo state. a. Data were collected at a 40° tilt angle and processed in cryoSPARC to give a good distribution of particles (green) with different views depicted in the 2D class averages. The final 3D reconstruction displaying a FSC$_{0.143}$ of 4.1 Å showed an isotropic map amenable for model building. Scale bar = 123 Å. b. Per-particle CTF refinement of the map improved the sphericity further as shown by the 3DFSC plot.
Extended Data Fig. 4 | Comparison of the NTD of Rev3 and Pol3. The NTD in Rev3 and Pol3, is composed of three motifs (I, II, III) but is much more elaborate and extended in Rev3. Loop 1 and Loop 2 contact all three motifs and connect the NTD to the fingers and palm domains, respectively.
Extended Data Fig. 5 | Surface representation of the T1 binding site. Residues around the T1 site are shown (sticks) for Rev3 (left) and Pol3 (right). Surface for the palm domains and DNA are shown in cyan and grey, respectively. The T1 base (red) and the key residues are highlighted in dots.
Extended Data Fig. 6 | Comparison of yeast and human Rev7-RIR complexes. 

a. Sequence alignment of yeast and human RBM1 and RBM2 regions of Rev3. Conserved prolines within RBM1 and RBM2 are highlighted in green. Also, highlighted are the conserved residues among the yeast and human homologs within the RIR region.

b. Structural comparison of the yeast and human RBM1 and RBM2. Individual structures of human Rev7 with RBM1 peptide (hRev7:RBM1; PDB ID: 3ABD) and RBM2 peptide (hRev7:RBM2; PDB ID: 6BC8) are compared to the corresponding sub-regions (yRev7A:RBM1; yRev7B:RBM2) in the yeast Polζ holoenzyme. The protein residues involved in the interactions are highlighted in green and the RIR is shown in brown. The interactions of Rev7A and Rev7B with the RIR segment connecting RBM1 and RBM2 (yRev7A:yRev7B:RIR_int) is also depicted.
Extended Data Fig. 7 | Comparison between the CysBD of Polζ and Polδ. A superimposition of the CysBD of the Polζ (left; grey in color) and Polδ (right; yellow in color; PDB ID: 6P1H) shows conservation in its overall topology. Notably, helix αXM in Polζ CysBD has been substituted by a loop in Polδ (PDB ID: 6P1H). All the four cysteines interacting with the 4Fe–4S cluster in Rev3 are also highlighted.
Extended Data Fig. 8 | Comparison of Rev3 and Pol II. Overlay of the palm domains of Rev3 and Pol II show a similar trajectory for the NTD-palm linker. In Rev3, this trajectory is coupled to interactions with the Palm-loop. The Pol II template DNA strand is shown in yellow (PDB ID: 3K5M). A close-up view of the looped-out abasic lesion and the adjoining 5′ guanine nucleotide. Notably, the guanine base clashes with the backbone carbonyl of E954 in Rev3.
Extended Data Fig. 9 | Docking of Rev1 CTD on the Pol\textsubscript{ζ} holoenzyme. a, Superimposition of human Rev7-RBM1-Rev1\textsubscript{CTD} (PDB ID: 4EXT) on Rev7\textsubscript{α} shows close proximity to Pol32\textsubscript{N} (shown in yellow), highlighting the importance of Pol32\textsubscript{N} in stabilizing this interaction. b, Superimposition of the human Rev7-RBM1-Rev1\textsubscript{CTD} on Rev7\textsubscript{α} shows clashes of Rev1\textsubscript{CTD} with various secondary structure elements of Rev7\textsubscript{α} (shown in yellow).
Corresponding author(s): Aneel K. Aggarwal, Iban Ubarretxena-Belandia
Last updated by author(s): Jun 12, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted. Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection
- Chameleon robot was used for sample preparation
- Leginon and Appion pipeline for collecting and processing EM images

Data analysis
- CTFFind4 was used for CTF determination
- GCTF was used for per-particle CTF estimation
- MotionCor2 was used for aligning cryoEM images
- Topaz was used for picking particles from EM images
- FindEM was used for picking particles from EM images
- RELION 2.0, RELION 3.0 beta and cryoSPARC were used for 2D classification, 3D classification, EM map calculation
- Directional anisotropy of the EM maps were calculated using the server- https://3dfsc.salk.edu/
- ResMAP and Relion 3.0 was used for estimating the local resolution of EM maps
- BALBES was used to determine the protein fold
- COOT was used for building structure models into the cryoEM map
- PHENIX was used for real space refinement of structure against the cryoEM map
- Chimera was used for map visualization and analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.
Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The cryo-EM density map has been deposited in the Electron Microscopy Data Bank (https://www.ebi.ac.uk/pdbe/emdb/) under accession numbers EMD-21115 and EMD-21108, respectively. Atomic coordinates have been deposited in the Protein Data Bank (https://www.rcsb.org) with accession numbers 6V93 and 6V8P, respectively. All reagents and relevant data are available from the authors upon reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- [x] Life sciences
- [ ] Behavioural & social sciences
- [ ] Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size          | No sample size calculation was required |
|----------------------|-----------------------------------------|
| Data exclusions      | EM images were excluded based on the confidence level of the CTF fit. |
| Replication          | Multiple data sets were collected on EM grids prepared using different instrumentation for sample preparation. The 2D class averages from these data sets agree with each other, and all show the pentameric ring-like architecture. Structure of the accessory subunits (Rev7, Pol31 and Pol32) obtained from the cryoEM map compares very well with the previously published x-ray and cryoEM structures of Pol31 and Pol32 and the x-ray structure of human Rev7. Finally, Fourier Shell correlation value of 0.143 between independently refined half sets was used to estimate the resolution of the cryoEM map. |
| Randomization        | This is not relevant to our study. A maximum likelihood based method was used for 2D classification of the particles. |
| Blinding             | This is not relevant to our study |

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

| Materials & experimental systems | Methods |
|----------------------------------|---------|
| n/a | Involved in the study | n/a | Involved in the study |
| [x] | Antibodies | [x] | ChiP-seq |
| [x] | Eukaryotic cell lines | [x] | Flow cytometry |
| [x] | Palaeontology | [x] | MRI-based neuroimaging |
| [x] | Animals and other organisms | | |
| [x] | Human research participants | | |
| [x] | Clinical data | | |