







# Cryo-EM structure of translesion DNA synthesis polymerase $\zeta$ with a base pair mismatch

Radhika Malik<sup>1</sup>, Robert E. Johnson<sup>2</sup>, Louise Prakash<sup>2</sup>, Satya Prakash <sup>2</sup>, Iban Ubarretxena-Belandia <sup>1,3,4</sup>  & Aneel K. Aggarwal <sup>1</sup> 

The B-family multi-subunit DNA polymerase  $\zeta$  (Pol $\zeta$ ) is important for translesion DNA synthesis (TLS) during replication, due to its ability to extend synthesis past nucleotides opposite DNA lesions and mismatched base pairs. We present a cryo-EM structure of *Saccharomyces cerevisiae* Pol $\zeta$  with an A:C mismatch at the primer terminus. The structure shows how the Pol $\zeta$  active site responds to the mismatched duplex DNA distortion, including the loosening of key protein-DNA interactions and a fingers domain in an “open” conformation, while the incoming dCTP is still able to bind for the extension reaction. The structure of the mismatched DNA-Pol $\zeta$  ternary complex reveals insights into mechanisms that either stall or favor continued DNA synthesis in eukaryotes.

<sup>1</sup>Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, NY, USA. <sup>2</sup>Department of Biochemistry and Molecular Biology, 301 University Blvd. University of Texas Medical Branch, Galveston, TX, USA. <sup>3</sup>Instituto Biofisika (UPV/EHU, CSIC), University of the Basque Country, E-48940 Leioa, Spain. <sup>4</sup>Ikerbasque, Basque Foundation for Science, 48013 Bilbao, Spain. email: [iban.ubarretxena@gmail.com](mailto:iban.ubarretxena@gmail.com); [aneel.aggarwal@mssm.edu](mailto:aneel.aggarwal@mssm.edu)

Assembled from catalytic Rev3 and accessory Rev7, Pol31, and Pol32 subunits, the translesion DNA synthesis (TLS) polymerase  $\zeta$  (Pol $\zeta$ ) plays an important role in the replication of damaged or mismatched DNA in eukaryotic cells<sup>1–4</sup>, and in the prevention of cancer<sup>5,6</sup>. We recently reported the cryo-EM structures of *S.cerevisiae* Pol $\zeta$  holoenzyme without DNA (4.1 Å) and in the act of DNA synthesis (3.1 Å)<sup>7</sup>. The structures decrypt a pentameric ring-like architecture for Pol $\zeta$ , with the catalytic Rev3 and the accessory Pol31, Pol32, and two Rev7 subunits (Rev7<sub>A</sub> and Rev7<sub>B</sub>) forming a daisy chain of protein-protein interactions<sup>7</sup>. Rev3 makes contacts to the DNA via its palm, fingers, thumb, exonuclease, and N-terminal (NTD) domains. While in apo Pol $\zeta$  the Rev3 fingers domain adopts an “open” conformation, in the matched DNA bound form it lies flush against the nascent base pair in a “closed” conformation and provides a basis for Pol $\zeta$ 's high fidelity during the nucleotide insertion step<sup>7</sup>. The ability of Pol $\zeta$  to tolerate mismatches and lesions at the primer terminus appears to derive in part from the path divergence of the linker between the NTD and the palm, which creates space to better accommodate deviations from Watson-Crick (W-C) base pair geometry<sup>7</sup>.

We present here, at a nominal resolution of 3.05 Å, a cryo-EM structure of Pol $\zeta$  with an A:C mismatch at the primer terminus (Fig. 1). The structure of the mismatched DNA-Pol $\zeta$  ternary complex provides a basis for understanding what makes Pol $\zeta$  more adept than other eukaryotic B-family DNA polymerases at extending DNA synthesis past mismatched base pairs.

## Results and discussion

**Structure of the mismatched DNA-Pol $\zeta$  ternary complex.** The Rev3, Rev7<sub>A</sub>, Rev7<sub>B</sub>, Pol31, and Pol32 subunits are organized around the mismatched duplex DNA in the same pentameric ring-like architecture as the matched complex (Fig. 1), with catalytic Rev3 alone making all of the contacts to the DNA. Rev3 embraces the mismatched template-primer with its palm (residues 329–373; 941–1043; 1098–1215), fingers (residues 302–328; 1044–1097), thumb (residues 1216–1372), inactive exonuclease (residues 662–894) domains, and the NTD (residues 1–301, 374–400, 895–940). The palm, thumb, and the exonuclease domains, as well as the NTD, occupy positions that are essentially identical to those observed with the matched DNA<sup>7</sup> (Fig. 1). That is, the palm interacts with the replicative end of the template-primer and carries the active site residues (Asp<sup>975</sup> and Asp<sup>1144</sup>), the thumb grips the duplex portion of the primer-template and makes contacts through the minor groove, the inactive exonuclease domain extends towards the major groove, and the NTD makes numerous contacts with the unpaired portion of the template strand. However, unlike the matched DNA-Pol $\zeta$  ternary complex, the fingers domain of Pol $\zeta$  on the mismatched template adopts an open conformation<sup>8</sup>, wherein the fingers helices  $\alpha$ F,  $\alpha$ XA and  $\alpha$ XB rotate outwards by  $\sim 15^\circ$  from the palm domain (Fig. 1), reminiscent of their position in the Pol $\zeta$  apo structure<sup>7</sup> (Supplementary Fig. 1). Interestingly, the fingers helices are less defined in the cryo-EM density compared to the matched complex (Supplementary Fig. 2a), suggesting motion around the open conformation. The replicative end of the mismatched DNA is also less defined in cryo-EM density than the matched complex (Supplementary Fig. 2b), indicating an increase in motion throughout the Pol $\zeta$  active site when containing a mismatch at the primer terminus.

**Pol $\zeta$  active site response to the mismatched DNA.** Intriguingly, even though the fingers domain is open to the same extent as in the apo Pol $\zeta$  structure<sup>7</sup> (Supplementary Fig. 1), Rev3 is observed with the incoming nucleotide dCTP (position P<sub>0</sub>) opposite the

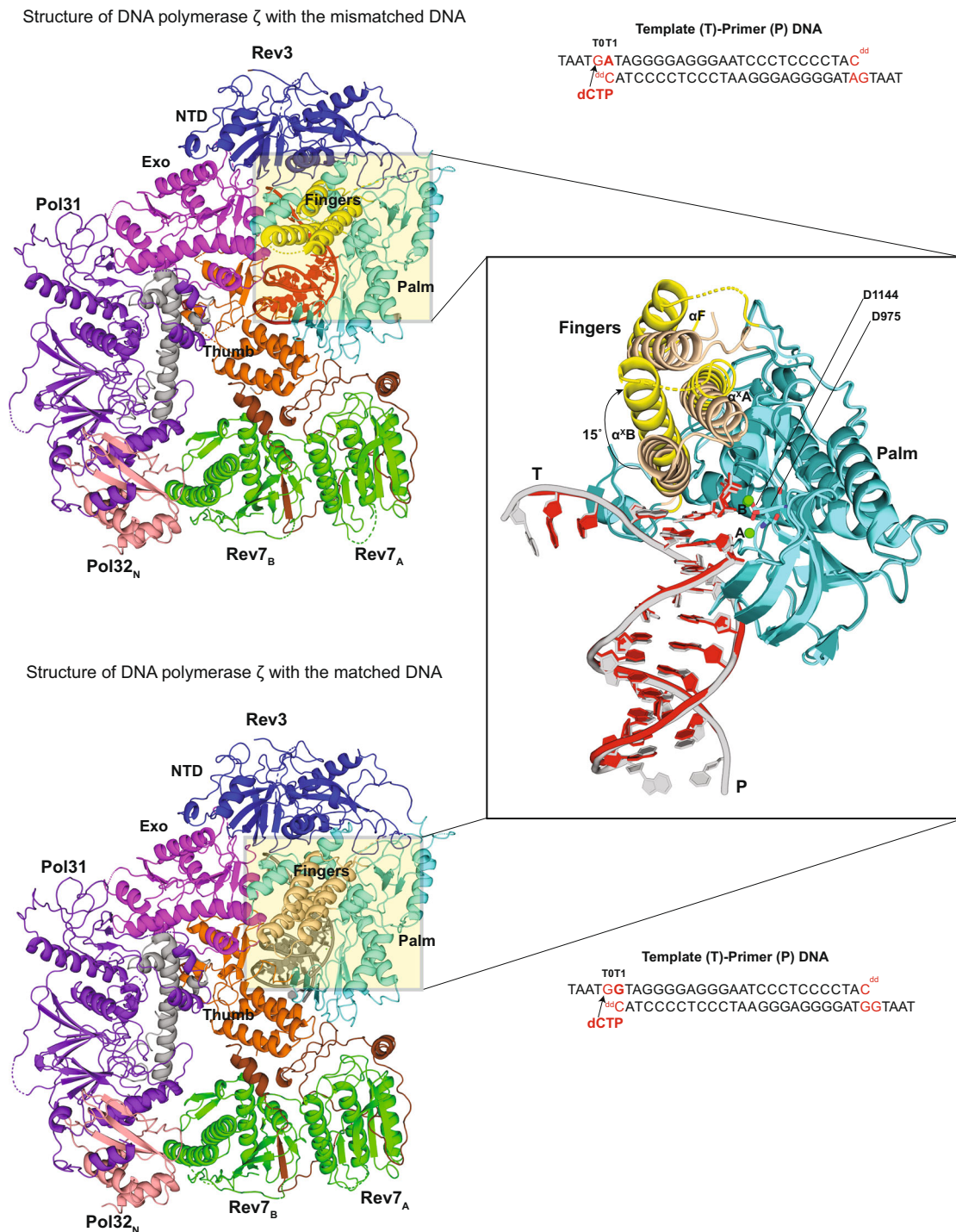
templating base G (position T<sub>0</sub>), establishing standard W-C base pairing for the nascent base pair (Fig. 2a and Supplementary Fig. 3). The dCTP triphosphate moiety treks between the fingers and palm domains, but many of the contacts observed in the matched complex with the fingers helices are lost; including, for example, hydrogen bonds between Lys<sup>1086</sup> and Arg<sup>1057</sup> and the  $\alpha$ - and  $\gamma$ -phosphates of dCTP, respectively (Fig. 2a, b). The dCTP is anchored loosely in the active site by inter-base hydrogen bonds with templating G, and by the contact it maintains with the palm domain, including stacking interactions between its sugar and Tyr<sup>980</sup>, and hydrogen bonds between its  $\beta$ - and  $\gamma$ -phosphates and the main chain amides of Leu<sup>979</sup> and Ser<sup>978</sup>, respectively. Of the two metals “A” and “B” observed in the active site of the matched complex<sup>7</sup> and associated with a two-metal ion mechanism of catalysis<sup>9,10</sup>, only metal B, a Ca<sup>2+</sup> ion, appears to be coordinated to the dCTP triphosphate moiety (Fig. 2a, b). The absence of metal A likely reflects the overall openness and mobility of the Rev3 active site when accommodating a mismatch at the primer terminus.

An open fingers domain creates space above the nascent G:C base pair and many of the van der Waals contacts observed in the matched complex are lost, including those from Leu<sup>1087</sup> and Val<sup>1091</sup> of the fingers helix  $\alpha$ XB (Fig. 2a, b). But, despite the open conformation of the fingers domain, Tyr<sup>1093</sup> and Gly<sup>1094</sup> from helix  $\alpha$ XB continue to impinge on the minor groove side of the nascent base pair and provide a basis for some fidelity for W-C base pairing at the insertion position (Fig. 2a).

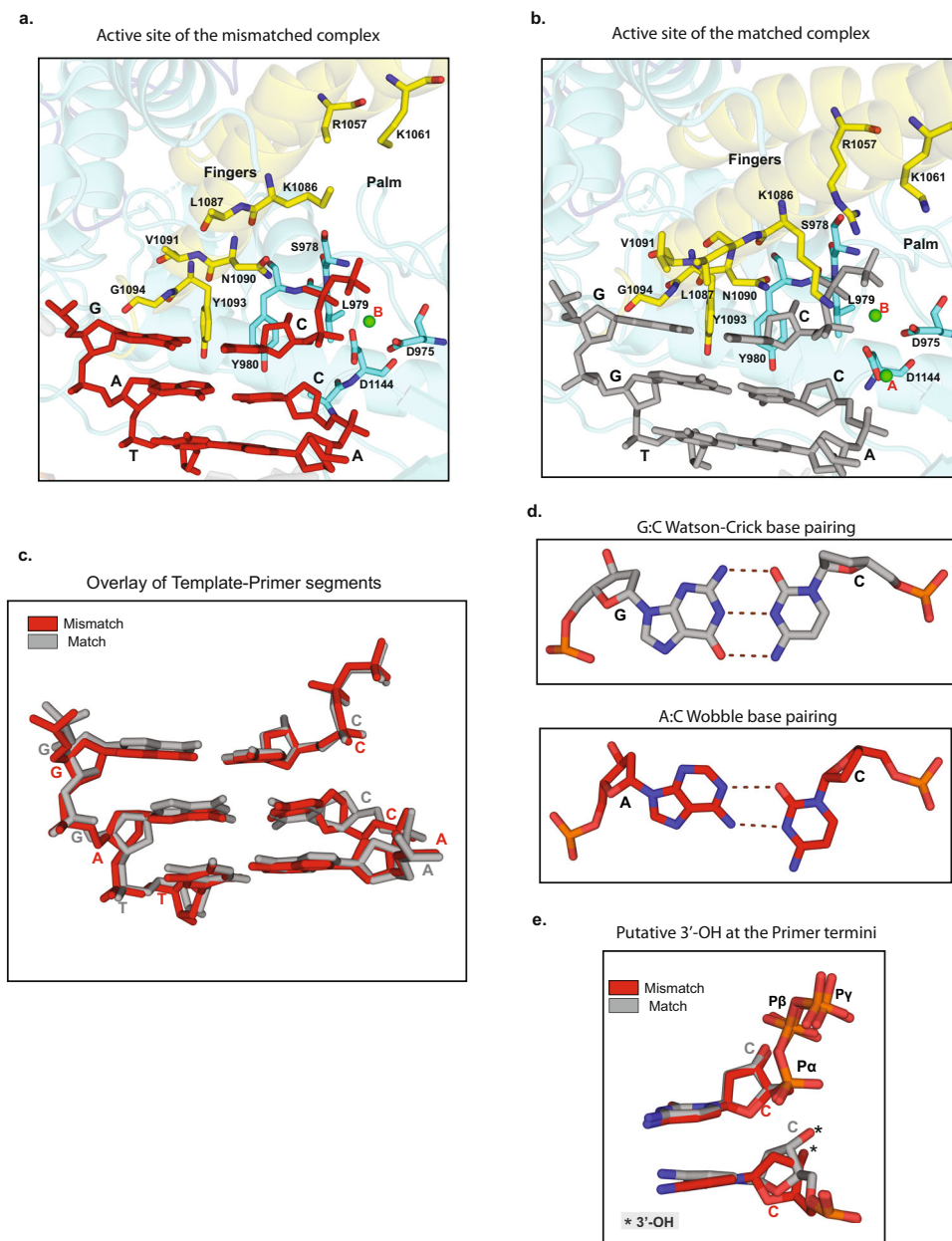
Amongst the various hydrogen bonding schemes that have been considered for an A:C mismatch within a DNA duplex, the DNA cryo-EM density we observe is most consistent with a “wobble” base pairing, in which the N<sup>6</sup> and N<sup>1</sup> atoms of adenine make putative hydrogen bonds with N<sup>3</sup> and O<sup>2</sup> atoms of cytosine, respectively (Fig. 2c, d). This would be the same configuration described by Kennard and co-workers<sup>11</sup> for an A:C mismatch in a DNA duplex, implying protonation of N<sup>1</sup> of adenine. For the wobble pairing, the cytosine base shifts towards the major groove by  $\sim 1.2$  Å, and the A:C base pair as a whole is much more propeller twisted ( $-17.8^\circ$ ) than the G:C base pair ( $-3.8^\circ$ ) in the matched complex (Fig. 2d). The increase in propeller twist (and buckling) extends to the neighboring base pairs, including the nascent base pair. These small perturbations in the conformation of the mismatched DNA duplex appear to hamper the ability of the Rev3 fingers domain to adopt the closed conformation observed in the matched DNA-Pol $\zeta$  ternary complex<sup>7</sup>.

The active site geometry is less primed for the nucleotidyl transfer reaction than in the matched complex. In particular, the A:C wobble pairing results in a shift in the cytidine sugar, which displaces the putative primer 3'OH by  $\sim 1.2$  Å from its position in the matched DNA-Pol $\zeta$  ternary complex<sup>7</sup> and increases the distance from  $\sim 3.8$  Å to 4.4 Å to the dCTP  $\alpha$ -phosphorous atom - making it less amenable for a nucleophilic attack (Fig. 2e). This increase in distance and the absence of catalytic metal A in the active site may explain the  $\sim 20$ -fold reduction in the ability of Pol $\zeta$  to extend DNA synthesis from mismatched A:C versus matched A:T at the primer terminus<sup>12</sup>.

**What makes Pol $\zeta$  a better extender of DNA synthesis past mismatched base pairs than other eukaryotic B-family polymerases?** It is likely that the overall mobility we observe at the primer terminus (Supplementary Fig. 2), while limiting the rate of catalysis on the one hand, may also facilitate the sampling of catalytically competent conformers for the nucleotidyl transfer reaction. Pol $\zeta$  is also notably different from most B-family polymerases in lacking proofreading exonuclease activity<sup>2–4</sup>, attributed to the absence of catalytic carboxylates in the Rev3



**Fig. 1 Structure of mismatched (A:C) DNA-Pol $\zeta$  ternary complex.** For the mismatched (A:C) DNA-Pol $\zeta$  ternary complex, the NTD (N-terminal domain), Exo (Exonuclease), RIR (Rev7 interacting region), fingers, palm, thumb, CTD (C-terminal domain) of Rev3 are shown in blue, magenta, brown, yellow, cyan, orange, and light gray, respectively; Rev7<sub>A</sub> and Rev7<sub>B</sub> in green; Pol31 and Pol32<sub>N</sub> in purple and salmon, respectively; and the DNA and incoming dCTP in red. For comparison, the matched (G:C) DNA-Pol $\zeta$  ternary complex<sup>7</sup> (PDB ID: 6V93) is shown below following the same color scheme, except for the fingers domain depicted in wheat color and the DNA and incoming dCTP in gray. The figure also displays the palindromic template-primer DNA sequences and the dCTP used to reconstitute the ternary complexes with Pol $\zeta$ . The right panel displays a close-up view of the region around the DNA, upon an  $\sim 90^\circ$  clockwise rotation of the structure along the y-axis, depicting the fingers helices  $\alpha F$ ,  $\alpha^A$ , and  $\alpha^B$  in a closed (matched DNA complex) and open (mismatched DNA complex) conformation. An arc arrow highlights the  $\sim 15^\circ$  outward movement of the fingers domain in the mismatched DNA complex relative to the matched DNA complex. Also highlighted in sticks are the active site residues D975 and D1144. Green spheres depict the active site metal B in the mismatched DNA complex and the active site metals A and B in the matched DNA complex.



**Fig. 2 Active sites of the Pol $\zeta$  complexes with mismatched and matched DNA.** **a** Close-up view of the Pol $\zeta$  active site in the mismatched DNA complex depicting the key residues from the fingers (yellow) and palm (blue) domains in close proximity to the replicative end of the template-primer DNA (red) and incoming dCTP (red). A green sphere depicts metal ion B. **b** Close-up view of the Pol $\zeta$  active site in the matched DNA complex (PDB ID: 6V93). The coloring scheme is the same as in **a**, except that the template-primer and dCTP are shown in gray. The complete side chains for few residues including Lys<sup>1061</sup>, Lys<sup>1086</sup>, and Arg<sup>1057</sup> have not been built in mismatched DNA complex due to less well-defined cryo-EM density. **c** Overlay of equivalent template-primer segments and incoming dCTP for the mismatched (red) and matched (gray) DNA complexes. **d** Base pairing at the primer terminus in the matched (G:C, gray color) and mismatched (A:C, red color) complexes. **e** Position of putative 3'-OH (marked by a star) at the primer terminus relative to the incoming dCTP in the mismatched (red) and matched (gray) complexes.

exonuclease domain and by the near absence of a  $\beta$ -hairpin substructure<sup>7</sup> (Supplementary Fig. 4). The  $\beta$ -hairpin in these B-family polymerase exonuclease domains is postulated to facilitate the transfer of a mismatched primer from the polymerase to exonuclease active site<sup>13,14</sup>, and its near absence in Rev3 may preferably partition the mismatched primer in the polymerase active site for the extension reaction. The mobility of the primer terminus and the near absence of the  $\beta$ -hairpin may combine to lend Pol $\zeta$  the ability to better extend synthesis from mismatched base pairs compared to Pol $\delta$  or Pole. This is likely coupled to the path divergence of the NTD-palm linker in Pol $\zeta$ , creating extra

space for DNA lesions and mismatches<sup>7</sup>. It is striking that despite the DNA distortion from the A:C mismatch and the open fingers domain, the incoming dCTP is still able to bind the Rev3 active site, with its triphosphate moiety oriented in the same manner as in the matched complex (Fig. 2a). Whether the fingers domain remains open or transiently closes during the covalent addition of a nucleotide from a mismatch is an intriguing question that remains to be answered. In either case, the binding/entry of metal A would be a prerequisite for activation of the primer 3'-OH for the nucleotidyl transfer reaction. Curiously, an open fingers domain has also been observed in the structure of the human



**Table 1 Summary of structural and refinement statistics.**

	Pol $\zeta$ -mismatch (A:C)-dCTP
Data collection and processing	
Magnification	$\times 105,000$
Voltage (kV)	300
Pixel size ( $\text{\AA}/\text{pixel}$ )	1.096
Electron dose ( $e^-/\text{\AA}^2$ )	64.82
Defocus range ( $\mu\text{m}$ )	0.5–2.5
Number of micrographs	2279
Number of particles	120,985
Symmetry imposed	$C_1$
Nominal map resolution ( $\text{\AA}$ )	3.05
FSC threshold	0.143
Map sharpening B-factor ( $\text{\AA}^2$ )	–102
Refinement (phenix)	
Initial model used (PDB code)	6V93
Model resolution ( $\text{\AA}$ )	3.4
FSC threshold	0.5
Model composition	
Non-hydrogen atoms	17,099
Protein residues	2180
DNA/other	24/3
B-factors ( $\text{\AA}^2$ )	
Protein	79.09
DNA/other	118.32/92.67
R.M.S. deviations	
Bond length ( $\text{\AA}$ )	0.002
Bond angles ( $^\circ$ )	0.495
Validation	
Molprobrity score	1.74
Clashscore	5.98
Rotamer outliers (%)	0.00
$C\beta$ outliers (%)	0.00
Ramachandran statistics (%)	
Favored	93.87
Allowed	6.08
Outliers	0.05

Pola catalytic domain in complex with a DNA duplex and dCTP, and suggested as one reason for the lower fidelity of Pola in incorporating nucleotides after the first dNTP<sup>15</sup>.

It has proven difficult indeed to capture structures of ternary complexes between wild-type B-family DNA polymerases and mismatched duplex DNA. By contrast, the A-family DNA polymerase from *B.stearothermophilus* has been characterized with numerous mismatches, revealing DNA distortions that extend up to six base pairs from the primer terminus<sup>16</sup>. The available structures of replicative B-family polymerase from bacteriophage RB69 with DNA mismatches, for example, have been largely derived with a quadruple mutant of the enzyme designed specifically for low base selectivity<sup>17</sup>. The structure presented here constitutes an early example of how the advent of cryo-EM methods offers unprecedented opportunities to capture and compare B-family polymerases with mismatches for a fuller understanding of the mechanisms that either stall or favor continued DNA synthesis.

## Methods

**Protein expression and purification.** *Saccharomyces cerevisiae* Pol $\zeta$  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 plasmids pBJ1462 and pBJ1524 and purified as described<sup>7,18</sup>. The complex with the mismatched duplex DNA was prepared by incubating the Pol $\zeta$  holoenzyme in a  $\text{CaCl}_2$  supplemented buffer with an HPLC purified (Integrated DNA technologies) palindromic DNA (in 1.5 molar excess), yielding an A:C mismatch at the extension ( $T_1$ ) position (5'-TAATGATAGGGGAGGGAATCCCTCCCTAC<sup>dd</sup>-3) and G as the templating base. Incoming dCTP was added (1 mM) to capture of Pol $\zeta$  in the act of DNA synthesis past an A:C mismatch.

**Cryo-EM specimen preparation.** Specimen preparation for Pol $\zeta$  holoenzyme with the mismatched DNA was done on 300-mesh gold-coated quantifoil grids of 1.2  $\mu\text{m}$  hole size and 1.3  $\mu\text{m}$  spacing. The grids were plasma cleaned using Ar and  $\text{O}_2$  for 8 s using a solarus plasma cleaner (Gatan) prior to loading 2.5  $\mu\text{l}$  of the sample. Back blotting followed by vitrification in liquid ethane was performed with a Leica EM GP2 plunge freezer (Leica microsystems).

**Cryo-EM data collection.** The mismatched DNA-Pol $\zeta$  ternary complex was imaged on a Titan Krios microscope (Thermo Fisher Scientific) operated at 300 kV and equipped with a K2 direct electron detector (Gatan) operating in super-resolution mode at a calibrated pixel size of 0.548  $\text{\AA}$ . The data were subsequently binned by 2 during frame alignment to yield a pixel size of 1.096  $\text{\AA}$ . Movies were recorded at a frame rate of 200 ms for 10 s for a total accumulated dose of 64.82  $e^-/\text{\AA}^2$ . A nominal defocus range of 0.5–2.5  $\mu\text{m}$  was employed, and the movies were automatically acquired using Legion to control both the microscope and the K2 camera<sup>19,20</sup>. Frames were aligned using MotionCor2 with dose weighting<sup>21</sup> and the contrast transfer function (CTF) estimation was performed with CTFIND4<sup>22</sup>. A total of 2279 images were collected over two sessions and processed to obtain the final three-dimensional (3D) reconstruction (Table 1).

**Cryo-EM data processing.** Particle picking for the mismatch complex was done with FindEM in Appion<sup>23</sup> using reprojections from the negative-stain reconstruction of Pol $\zeta$ <sup>24</sup>. Particles picked were subjected to multiple rounds of two-dimensional (2D) classification in cryoSPARC2<sup>25</sup>. Selected Pol $\zeta$  particles from about half of the micrographs (Session 1: 1008) were used to obtain a 3D reconstruction with a sphericity of 0.90 out of 1. The other half of the data (Session 2: 1171) was trained with topaz, a neural network-based particle picker<sup>26</sup> implemented within cryoSPARC2. The micrographs were binned by four and used with resnet8 neural-network architecture. Pi, the expected fraction of positive pixels, was set to 0.027, and the radius parameter, which sets the number of pixels around a labeled particle coordinate, was set to 3. A total of 200 iterations were used to fit the topaz model.

Selected particles from both sessions were combined and iterative rounds of 2D classification were performed. A final set of 210,505 particles was subjected to ab-initio clean-up in cryoSPARC2, which allowed the removal of low-resolution models with preferred orientation. The ab-initio model was refined using non-uniform refinement to generate a 3D reconstruction at a nominal resolution of 3.1  $\text{\AA}$  based on the Fourier shell correlation (FSC) value of 0.143 between independently refined half sets<sup>27,28</sup>. The cryo-EM map was checked for directional anisotropy (<https://3dfsc.salk.edu>) and gave a value of 0.967 out of 1. The particles in this reconstruction were subjected to local CTF refinement followed by local refinement within cryoSPARC2, which lead to a 3D reconstruction with an  $\text{FSC}_{0.143}$  value of 3.05  $\text{\AA}$  (Table 1 and Supplementary Fig. 5).

**Model building and refinement.** Model building was performed in COOT<sup>29</sup> using the matched DNA-Pol $\zeta$  ternary complex as the initial model (PDB ID: 6V93). The 3D reconstruction was subjected to local map sharpening using the fitted model coordinates in locscale (ccpem)<sup>30,31</sup> to improve the contrast of the cryo-EM density, especially in the flexible regions of the map. The model was refined by multiple rounds of real-space refinement in Phenix<sup>32</sup> using the locscale sharpened map. The model was validated using Molprobrity<sup>33</sup> and the data was analyzed in Chimera<sup>34</sup>. The figures were prepared in Pymol<sup>35</sup>.

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

## Data availability

The data that support this study are available from the corresponding author upon reasonable request. The cryo-EM density map generated in this study has been deposited in the Electron Microscopy Data Bank (EMDB) under accession number EMD-24793. The resulting atomic coordinates have been deposited in the Protein Data Bank (PDB) with accession number 7S0T. The atomic coordinates used in the study are available in the PDB with accession numbers 6V93 and 6V8P.

Received: 31 August 2021; Accepted: 31 January 2022;

Published online: 25 February 2022

## References

- Prakash, S., Johnson, R. E. & Prakash, L. Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. *Annu Rev. Biochem.* **74**, 317–353 (2005).
- Jain, R., Aggarwal, A. K. & Rechkoblit, O. Eukaryotic DNA polymerases. *Curr. Opin. Struct. Biol.* **53**, 77–87 (2018).
- Makarova, A. V. & Burgers, P. M. Eukaryotic DNA polymerase zeta. *DNA Repair* **29**, 47–55 (2015).

4. Martin, S. K. & Wood, R. D. DNA polymerase zeta in DNA replication and repair. *Nucleic Acids Res.* **47**, 8348–8361 (2019).
5. Lange, S. S., Takata, K. & Wood, R. D. DNA polymerases and cancer. *Nat. Rev. Cancer* **11**, 96–110 (2011).
6. Sharma, S., Helchowski, C. M. & Canman, C. E. The roles of DNA polymerase zeta and the Y family DNA polymerases in promoting or preventing genome instability. *Mutat. Res.* **743–744**, 97–110 (2013).
7. Malik, R. et al. Structure and mechanism of B-family DNA polymerase zeta specialized for translesion DNA synthesis. *Nat. Struct. Mol. Biol.* **27**, 913–924 (2020).
8. Doublet, S., Sawaya, M. R. & Ellenberger, T. An open and closed case for all polymerases. *Structure* **7**, R31–R35 (1999).
9. Steitz, T. A. & Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl Acad. Sci. USA* **90**, 6498–6502 (1993).
10. Steitz, T. A. DNA polymerases: structural diversity and common mechanisms. *J. Biol. Chem.* **274**, 17395–17398 (1999).
11. Hunter, W. N., Brown, T., Anand, N. N. & Kennard, O. Structure of an adenine-cytosine base pair in DNA and its implications for mismatch repair. *Nature* **320**, 552–555 (1986).
12. Johnson, R. E., Washington, M. T., Haracska, L., Prakash, S. & Prakash, L. Eukaryotic polymerases  $\iota$  and  $\zeta$  act sequentially to bypass DNA lesions. *Nature* **406**, 1015–1019 (2000).
13. Hogg, M., Aller, P., Konigsberg, W., Wallace, S. S. & Doublet, S. Structural and biochemical investigation of the role in proofreading of a beta hairpin loop found in the exonuclease domain of a replicative DNA polymerase of the B family. *J. Biol. Chem.* **282**, 1432–1444 (2007).
14. 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 3'→5' exonuclease activities. *J. Mol. Biol.* **254**, 15–28 (1995).
15. Baranovskiy, A. G. et al. Activity and fidelity of human DNA polymerase  $\alpha$  depend on primer structure. *J. Biol. Chem.* **293**, 6824–6843 (2018).
16. Johnson, S. J. & Beese, L. S. Structures of mismatch replication errors observed in a DNA polymerase. *Cell* **116**, 803–816 (2004).
17. Xia, S. & Konigsberg, W. H. RB69 DNA polymerase structure, kinetics, and fidelity. *Biochemistry* **53**, 2752–2767 (2014).
18. Johnson, R. E., Prakash, L. & Prakash, S. Pol31 and Pol32 subunits of yeast DNA polymerase  $\delta$  are also essential subunits of DNA polymerase  $\zeta$ . *Proc. Natl Acad. Sci. USA* **109**, 12455–12460 (2012).
19. Lander, G. C. et al. Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.* **166**, 95–102 (2009).
20. Suloway, C. et al. Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).
21. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
22. Rohou, A. & Grigorieff, N. CTFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
23. Roseman, A. M. FindEM - a fast, efficient program for automatic selection of particles from electron micrographs. *J. Struct. Biol.* **145**, 91–99 (2004).
24. Gomez-Llorente, Y. et al. The Architecture of Yeast DNA Polymerase  $\zeta$ . *Cell Rep.* **5**, 79–86 (2013).
25. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290 (2017).
26. Bepler, T. et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* <https://doi.org/10.1038/s41592-019-0575-8> (2019).
27. Scheres, S. H. W. & Chen, S. X. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9**, 853–854 (2012).
28. 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).
29. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
30. Burnley, T., Palmer, C. M. & Winn, M. Recent developments in the CCP-EM software suite. *Acta Crystallogr. D Struct. Biol.* **73**, 469–477 (2017).
31. Jakobi, A. J., Wilmanns, M. & Sachse, C. Model-based local density sharpening of cryo-EM maps. *Elife* <https://doi.org/10.7554/eLife.27131> (2017).
32. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
33. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
34. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
35. DeLano, W. L. & Lam, J. W. PyMOL: a communications tool for computational models. *Abstr. Pap. Am. Chem. Soc.* **230**, U1371–U1372 (2005).

## Acknowledgements

This work was funded by grants R01-GM124047 (A.K.A. and L.P.) and R35-GM13170 (A.K.A.) from the National Institutes of Health (NIH). I.U.-B was supported by a grant PID2019-104423GB-I00/AEI/10.13039/501100011033 from the Spanish State Research Agency and by the Basque Excellence Research Centre program. 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 Bio-computing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311.

## Author contributions

A.K.A. and R.M. designed the experiments. R. E. J. expressed Pol $\zeta$  in yeast. R.M. purified Pol $\zeta$  and optimized grid preparation for cryo-EM studies. R.M. collected and processed the cryo-EM data. R.M. built and refined the atomic model. A.K.A. and I.U.-B guided the overall project. S.P. and L.P. guided the protein expression studies. A.K.A. and R.M. and I.U.-B prepared the manuscript with input from all authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-28644-7>.

**Correspondence** and requests for materials should be addressed to Iban Ubarretxena-Belandia or Aneel K. Aggarwal.

**Peer review information** *Nature Communications* thanks Yoshizumi Ishino and the other anonymous reviewer(s) for their contribution to the peer review this work.

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

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



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

© The Author(s) 2022