nature structural & molecular biology

Article

https://doi.org/10.1038/s41594-025-01514-8

Human polymerase θ helicase positions DNA microhomologies for double-strand break repair

Received: 11 September 2024

Accepted: 10 February 2025

Published online: 28 February 2025

Check for updates

Christopher J. Zerio[1], Yonghong Bai[2], Brian A. Sosa-Alvarado[2], Timothy Guzi[2] & Gabriel C. Lander[1]

DNA double-strand breaks occur daily in all human cells and must be repaired with high fidelity to minimize genomic instability. Deficiencies in high-fidelity DNA repair by homologous recombination lead to dependence on DNA polymerase θ, which identifies DNA microhomologies in 3′ single-stranded DNA overhangs and anneals them to initiate error-prone double-strand break repair. The resulting genomic instability is associated with numerous cancers, thereby making this polymerase an attractive therapeutic target. However, despite the biomedical importance of polymerase θ, the molecular details of how it initiates DNA break repair remain unclear. Here, we present cryo-electron microscopy structures of the polymerase θ helicase domain bound to microhomology-containing DNA, revealing DNA-induced rearrangements of the helicase that enable DNA repair. Our structures show that DNA-bound helicase dimers facilitate a microhomology search that positions 3′ single-stranded DNA ends in proximity to align complementary bases and anneal DNA microhomology. We characterize the molecular determinants that enable the helicase domain of polymerase θ to identify and pair DNA microhomologies to initiate mutagenic DNA repair, thereby providing insight into potentially targetable interactions for therapeutic interventions.

Rapid and accurate repair of DNA double-strand breaks is crucial for maintaining genome integrity. Because chromosome breaks, deletions and rearrangements are enabling characteristics of cancer, robust and precise DNA repair mechanisms are necessary to minimize carcinogenesis[1]. The majority of DNA double-strand breaks are repaired with high fidelity through nonhomologous end joining or homologous recombination (HR)[2,3]. Mutations in DNA damage repair genes cause increased dependency on an alternative, error-prone DNA repair pathway called theta-mediated end joining (TMEJ)[4,5]. These mutations are accompanied by an upregulation of DNA polymerase θ (Polθ), a multidomain protein that is critical for performing TMEJ[6,7] and, therefore, associated with genome instability and carcinogenesis in HR-deficient cells[8–10]. Polθ is overexpressed in many cancers, including breast, ovarian, lung, and colorectal cancers, and blocking Polθ expression or activity in the context of breast cancer susceptibility protein (BRCA1/2)-deficient tumors is synthetic lethal[5,8,9]. While TMEJ allows cancer cells to repair DNA damage in an HR-deficient setting, this dependence on TMEJ also renders tumors particularly vulnerable to Polθ inhibition[4]. Thus, Polθ inhibitors hold therapeutic promise in combatting HR-deficient cancers[11–15].

Polθ is a 290-kDa protein consisting of an N-terminal superfamily 2 helicase (PolθH) that is tethered to a C-terminal A-family polymerase (PolθP) through an unstructured 920-amino acid central domain (PolθC)[16]. Polθ engages and anneals DNA microhomologies 2–6 base

[1]Department of Integrative Structural and Computational Biology, Scripps Research, La Jolla, CA, USA. [2]MOMA Therapeutics, Cambridge, MA, USA. e-mail: glander@scripps.edu

https://doi.org/10.1038/s41594-025-01514-8

pairs in length within resected 3′ single-stranded (ss)DNA overhangs at the site of a DNA double-strand break to initiate TMEJ[17–19]. Despite extensive study of this DNA repair system, the molecular mechanism by which Polθ identifies and anneals DNA microhomologies is not well established[4,20,21].

Both PolθP and PolθH are required for efficient TMEJ[22] but the role of PolθH is poorly understood. Prior studies hypothesized that translocation of PolθH along ssDNA may remove DNA-binding proteins[23,24], position PolθH to facilitate 3′ DNA microhomology annealing[21,23,25], or suppress intrastrand base pairing and snapback replication by PolθP[22]. These mixed observations, along with the fact that PolθH displays weak DNA-unwinding activity[26], raise questions about the role of PolθH in 3′ DNA microhomology annealing and the interactions between PolθH and DNA that facilitate TMEJ. As a result, we sought to investigate the molecular determinants of the PolθH–DNA interaction.

Results

To investigate the mechanistic relationship between PolθH and DNA microhomologies, we recombinantly expressed and purified human PolθH (residues 2–894) (Extended Data Fig. 1a) for single-particle cryo-electron microscopy (cryo-EM) analyses (Table 1). In the absence of DNA, PolθH protomers assemble not only as a tetramer (Fig. 1a), which is consistent with prior crystallographic studies[25], but also as a dimer (Fig. 1b,c), as recently observed for the cryo-EM structure of PolθH bound to inhibitor novobiocin[27]. Our ~3.3-Å-resolution structure of the tetramer assembly adhered to D_2 symmetry as a 'dimer of dimers', with an interdimer rotation of about 4° compared to the tetramer crystal structure (Extended Data Fig. 1b)[25]. Our isolated dimer structure, resolved to ~3.5-Å resolution, is structurally consistent with both dimers that comprise the tetramer. Although neither of the previous PolθH structural studies detected a mixed population of assemblies[25,27], we surmise that the dimeric PolθH may be the functional population, as the two helicase protomers are suited to handle the two resected ssDNA overhangs that are substrates for DNA repair by TMEJ.

To gain further insight into PolθH function, we used a native PAGE assay to identify DNA species that bind PolθH. Our assay showed that PolθH preferentially binds DNA oligonucleotides with microhomology in a 3′ overhang (Extended Data Fig. 2a). This prompted us to introduce a stem-loop DNA with microhomology embedded in a 3′ overhang to mimic the end-resected TMEJ DNA substrate (Extended Data Fig. 2b). This DNA species bound PolθH in the native PAGE assay and we expected it to interact with PolθH in a fashion that would recapitulate the initial stage of Polθ-mediated TMEJ with sufficient stability for cryo-EM structure determination.

We were immediately surprised to see in our cryo-EM images that the PolθH–DNA complex consisted entirely of dimers, indicating that binding of this DNA to PolθH drives dissociation of the tetrameric assembly. We determined a ~3.5-Å-resolution cryo-EM reconstruction of dimeric PolθH bound to DNA, enabling us to build an atomic model detailing the interactions between the helicase dimer and two strands of DNA that contain regions of 3′ microhomology (Fig. 1b,d and Table 2). We identified density corresponding to ssDNA spanning the DNA tunnel of each protomer that was sufficiently resolved to discriminate between purines and pyrimidines. Following this ssDNA toward one of the 5′ ends, where one DNA strand presumably enters the tunnel, we observed density consistent with a double-stranded (ds)DNA stem-loop. In the opposite direction, toward the center of the PolθH dimer, the 3′ end of the ssDNA exits the tunnel and connects to density corresponding to two 3′ ssDNA overhangs (one exiting from each PolθH protomer) annealed to form dsDNA microhomology. This annealed microhomology is flexibly positioned above an open valley between the two PolθH protomers that lacks any stabilizing protein–DNA interactions and, thus, could only be resolved at moderate resolution. However, given that the DNA density is connected to

Table 1 | Apo PolθH cryo-EM data collection, analysis, and modeling statistics

	Apo tetramer	Apo dimer
PDB identifier	**9BH6**	**9BH7**
EMDB identifier	**EMD-44534**	**EMD-44535**
Data collection		
Microscope	Titan Krios	
Camera	K3 BioQuantum	
Camera mode	Counting	
Magnification (nominal/at detector)	105,000/60,024	
Voltage (kV)	300	
Data acquisition software	Leginon	
Exposure navigation	Image shift to 54	
Total electron exposure (e^- per Å^2)	50	
Exposure rate (e^- per pixel per s)	38.8	
Frame length (ms)	30	
Number of frames	30	
Pixel size (Å per pixel)	0.833	
Defocus range (μm)	−0.3 to −2.0	
Slit width (eV)	N/A	
Micrographs collected (no.)	3,906	
Data analysis		
Total extracted picks (no.)	1,474,207	
Particles used for 3D (no.)	1,436,257	
Final particles (no.)	208,545[a]	138,712[a]
Symmetry	D_2	C_2
Resolution (global) (Å)		
FSC=0.5 (unmasked/masked)	4.4[a]/3.7[a]	4.7[a]/4.0[a]
FSC=0.143 (unmasked/masked)	3.8[a]/3.3[a]	4.1[a]/3.5[a]
Local resolution range (Å)	2.9[a]–4.4[a]	3.1[a]–6.3[a]
3DFSC sphericity (%)	97.7[a]	86.0[a]
Map sharpening *B*-factor (Å^2)	−90.2[a]	−112.2[a]
Model composition		
Chains	4	2
Nonhydrogen atoms	24,388	12,166
Protein residues	3,096	1,544
Model refinement		
Refinement package	ISOLDE/PHENIX	ISOLDE/PHENIX
MapCC (volume/mask)	0.82/0.81	0.77/0.78
B-factors (Å^2)		
Protein	54.43	55.33
Nucleotide	N/A	N/A
Root-mean-square deviations		
Bond lengths (Å)	0.003	0.003
Bond angles (°)	0.614	0.634
Validation		
Map-to-model FSC 0.5	3.5[a]	3.7[a]
Ramachandran plot (%)		
Favored (%)	99.38	99.21
Allowed (%)	0.49	0.66
Outliers (%)	0.13	0.13
MolProbity score	0.88	0.90
Clashscore	1.44	1.58
Poor rotamers (%)	0.00	0.00
CaBLAM outliers (%)	0.30	0.40
EMRinger score	3.48	2.24
Q-score	0.53	0.48

[a]Refers to asymmetric units after symmetry expansion (single protomer). N/A, not applicable.

https://doi.org/10.1038/s41594-025-01514-8

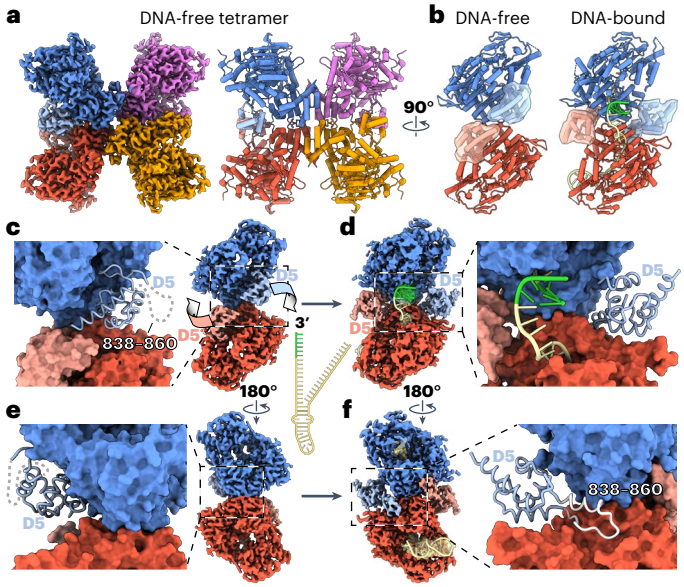


Fig. 1 | PolθH D5 rotates outward to accommodate DNA microhomology annealing. a, Apo PolθH tetramer cryo-EM map and atomic model where helices are represented as tubes, colored by protomer. D5 in each blue and red protomer is colored using a lighter shade. **b**, Atomic models of the apo PolθH (left) and DNA-bound PolθH dimer (right) shown using a cartoon representation. The D5 in each protomer is highlighted using a semitransparent surface. The annealed 3′ DNA microhomology (green) is shown between two helicase protomers in the DNA-bound dimer. **c**, Close-up view of the apo PolθH dimer atomic model shown as a molecular surface, with D5 shown using a licorice representation and the disordered residues 838–860 as a dashed gray line. Right, cryo-EM density of the apo PolθH colored as in **b**, with arrows showing the trajectory of the D5 rearrangement that occurs upon interaction with DNA, along with a schematic of the DNA oligomer used for this study, with the color green denoting the region of microhomology. **d**, Cryo-EM density of PolθH bound to DNA, with density corresponding to DNA extending from each DNA tunnel shown as a semitransparent isosurface and the modeled DNA shown as a cartoon within (yellow with green microhomology). Right, close-up view emphasizing the new position of D5 after having rotated away from the center of the dimer to accommodate DNA pairing between protomers. **e**,**f**, Views of the PolθH dimer maps, rotated 180° relative to the corresponding panels in **c** (**e**) and **d** (**f**). The position of D5 before and after DNA binding is observable, with **f** emphasizing how repositioning of D5 causes residues 838–860 (white), which are disordered in the apo form, to adopt a loop–helix structure that docks into a pocket in the opposite protomer, stabilizing the DNA-bound dimer.

both incoming ssDNA oligonucleotides and the size of the structural feature can accommodate a span of 4–6 base-paired nucleotides, we are confident that the density between the protomers corresponds to the joined 3′ dsDNA microhomology. No adenosine triphosphate (ATP) was added to solve the PolθH–DNA structures, strongly indicating that PolθH ATPase activity is dispensable for PolθH identifying, aligning, and annealing 3′ DNA microhomology during TMEJ in the absence of ssDNA-binding proteins[21–23].

PolθH D5 rearrangements facilitate microhomology annealing

We next compared our DNA-free and DNA-bound PolθH structures to investigate conformational changes associated with PolθH–DNA binding. This comparison revealed a striking rearrangement of the PolθH C-terminal domain 5 (D5; residues 790–894). In our DNA-free PolθH structure, both D5s are positioned adjacent to each other in the intradimer valley, occupying the space where we observe DNA microhomology annealed in our DNA-bound structure (Fig. 1c). This adjacent positioning of the D5s at the protomer–protomer interface is consistent with the orientation of D5 in previous PolθH crystal structures[25]. However, in our DNA-bound structure, joining of DNA microhomology is associated with a repositioning of D5, whereby each D5 undergoes a ~150° rotation away from the dimeric axis, clearing the region between the two PolθH protomers to accommodate DNA microhomology annealing (Fig. 1d and Supplementary Video 1). AlphaFold 3 fails to predict this D5 rotation when provided two PolθH protomers and two DNA substrates[28]. This D5–DNA relationship contrasts with the D5 equivalent in archaeal PolθH homolog *Archaeoglobus fulgidus* Hel308 (~30% sequence similarity to human PolθH), which directly interacts with ssDNA[29]. A more subtle and distinct D5 rearrangement was previously observed in the cryo-EM structure of PolθH bound to novobiocin[27], suggesting that D5 is a mobile element that responds to or regulates functional states.

The large D5 rotation we observe is likely triggered by interactions formed by DNA as it traverses the PolθH DNA tunnel. We posit that, as an incoming DNA strand exits the PolθH DNA tunnel, it encounters a negatively charged patch on D5 of the protomer and resulting repulsive electrostatic interactions cause the outward swinging of D5 (Extended Data Fig. 3). This DNA-promoted D5 rearrangement establishes new intraprotomer and interprotomer interactions within the dimer, including the ordering of D5 residues 838–860, residues that were found to bind Rad51 recombinase in a previous study[9], as a loop–helix that interacts with a pocket on the back side of the opposing protomer (Fig. 1f and Supplementary Video 1). Conversely, in the DNA-free PolθH structure, these residues are flexible and not observed and the interaction pocket on the back side of the dimer is not accessed (Fig. 1e). Formation of this loop–helix within this pocket on the opposite helicase protomer simultaneously locks D5 in the outward position to accommodate the annealed dsDNA while also stabilizing the interprotomer interactions of the DNA-bound dimer.

The DNA-promoted rearrangement of D5 is concomitant with a larger-scale rearrangement of the PolθH dimer that disrupts the tetrameric assembly (Fig. 1f and Extended Data Fig. 4a,b). Upon DNA binding, the angle between the subunits of the PolθH dimer increases by ~14°, with the conserved protomer–protomer interface serving as a hinge point. Notably, the DNA-promoted rigid-body motions of the subunits, along with the new D5-mediated interprotomer interactions, together destabilize the PolθH tetramer by sterically obstructing the interprotomer interactions between the diagonal and adjacent protomers responsible for tetramer formation (Extended Data Fig. 4b). These conformational changes of PolθH are consistent with a mechanism whereby PolθH binding DNA and DNA traversing the PolθH tunnel to the dimeric interface trigger a rearrangement of D5 that promotes tetramer dissociation, explaining why there are no tetrameric species in the DNA-bound PolθH population.

DNA interactions with PolθH

D1, D2, and D4 contribute to the DNA tunnel of PolθH (Fig. 2a,b) and the quality of our density in this region enabled us to identify key interactions between DNA and PolθH residues (Fig. 2 and Extended Data Fig. 5).

At the entrance of the DNA tunnel, three lysine residues (K348, K352, and K497) are positioned for electrostatic interactions with the dsDNA backbone (Fig. 2c). Proximal to the tunnel entrance, a small loop in D2 comprising residues G465, G466, and R467 is positioned adjacent to the duplex DNA of the stem-loop, at the branch point between the last paired DNA bases and the first unpaired bases of the 3′ DNA overhang (Fig. 2d). This nucleoprotein arrangement disrupts base pairing and stacking while directing the 3′ ssDNA overhang into the DNA tunnel, similarly to the corresponding β-hairpin DNA-unwinding loop in *A. fulgidus* Hel308 (Extended Data Fig. 6a,b)[29]. However, this loop is smaller in PolθH than in the Hel308 family helicases and lacks the aromatic residues that stack with the last of the paired DNA bases (Extended Data Fig. 6b)[29]. These differences may explain why PolθH does not possess the robust DNA-unwinding activity displayed by *A. fulgidus* Hel308 (refs. 4,26,30).

https://doi.org/10.1038/s41594-025-01514-8

Table 2 | DNA-bound PolθH cryo-EM data collection, analysis, and modeling statistics

	DNA searching	DNA aligning	DNA annealed
PDB identifier	9BH8	9BH9	9BHA
EMDB identifier	EMD-44536	EMD-44537	EMD-44538
Data collection			
Microscope	Titan Krios		
Camera	K3 BioQuantum		
Camera mode	Counting		
Magnification (nominal/at detector)	105,000/60,024		
Voltage (kV)	300		
Data acquisition software	Leginon		
Exposure navigation	Image shift to 63		
Total electron exposure (e^- per Å^2)	50		
Exposure rate (e^- per pixel per s)	24.8		
Frame length (ms)	30		
Number of frames	47		
Pixel size (Å per pixel)	0.833		
Defocus range (μm)	−0.3 to −2.5		
Slit width (eV)	20		
Micrographs collected (no.)	12,440		
Data analysis			
Total extracted picks (no.)	5,424,691		
Particles used for 3D (no.)	5,050,169		
Final particles (no.)	102,983	107,353	88,854
Symmetry	C_1	C_1	C_1
Resolution (global) (Å)			
FSC=0.5 (unmasked/masked)	5.0/4.0	4.6/4.0	4.5/3.7
FSC=0.143 (unmasked/masked)	4.0/3.5	4.0/3.4	3.8/3.3
Local resolution range (Å)	3.1–7.4	3.0–6.8	2.9–7.7
3DFSC sphericity (%)	79.7	81.3	87.4
Map sharpening *B*-factor (Å^2)	−102.1	−102.6	−108.7
Model composition			
Chains	4	4	4
Nonhydrogen atoms	13,051	13,155	13,270
Protein residues	1,574	1,574	1,586
Nucleotides	44	50	49
Model refinement			
Refinement package	ISOLDE/PHENIX	ISOLDE/PHENIX	ISOLDE/PHENIX
MapCC (volume/mask)	0.74/0.76	0.74/0.77	0.74/0.77
B-factors (Å^2)			
Protein	69.39	67.19	53.26
Nucleotide	33.75	42.19	35.80
Root-mean-square deviations			
Bond lengths (Å)	0.003	0.003	0.003
Bond angles (°)	0.640	0.680	0.654
Validation			
Map-to-model FSC 0.5	3.8	3.7	3.7
Ramachandran plot (%)			
Favored (%)	99.29	99.35	99.42
Allowed (%)	0.71	0.65	0.51
Outliers (%)	0.00	0.00	0.06
MolProbity score	0.90	0.93	0.92
Clashscore	1.53	1.75	1.70
Poor rotamers (%)	0.15	0.07	0.00
CaBLAM outliers (%)	0.52	0.39	0.45
EMRinger score	2.33	3.06	2.78
Q score	0.46	0.48	0.47

https://doi.org/10.1038/s41594-025-01514-8

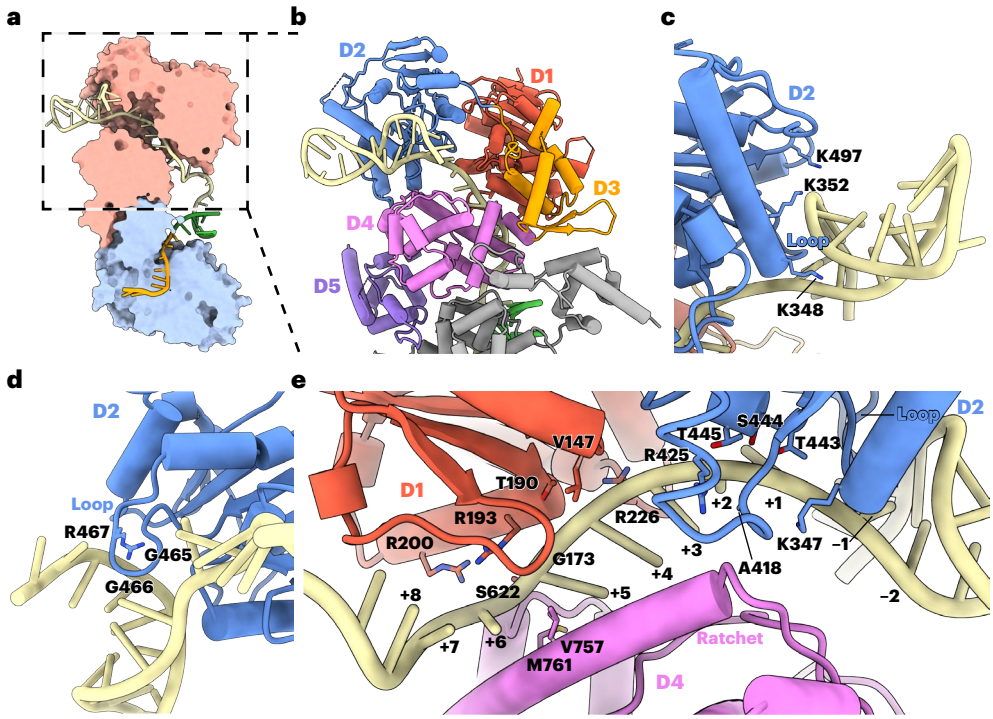


Fig. 2 | Characterization of PolθH–DNA binding. a, Cross-section of the PolθH–DNA model showing that PolθH binds two copies of the DNA oligonucleotide (one orange and one yellow) and facilitates annealing of the self-complementary 3′ microhomology (green). **b**, Each domain of a single PolθH protomer is individually colored and labeled. **c**, dsDNA interacts with residues in PolθH D2 adjacent to the ssDNA tunnel entrance. **d**, A small PolθH loop facilitates ssDNA entry into the DNA tunnel. **e**, DNA interacts with residues in PolθH D1, D2, and D4 as it traverses the DNA tunnel. DNA bases are labeled starting two bases upstream of the small PolθH unwinding loop from 5′ (−2) to 3′ (+8).

As ssDNA traverses the DNA tunnel, it contacts the ratchet helix, an ATP-coupled structural feature that has been shown to promote unidirectional (3′ to 5′) ssDNA translocation in Hel308 family helicases (Fig. 2e)[29,31]. In *A. fulgidus* Hel308, the ratchet helix ensures unidirectional ssDNA translocation through two residues (R592 and W599) that stack with the bases of an incoming ssDNA substrate (Extended Data Fig. 6c)[29]. The PolθH ratchet helix, however, does not appear to introduce such base-stacking interactions with the DNA substrate, although residues V757 and M761 on this helix appear to wedge between two DNA bases, which may prevent backtracking to facilitate processive 3′-to-5′ ssDNA translocation (Extended Data Fig. 6d)[24,25].

The ssDNA is guided through the PolθH DNA tunnel by a series of base-agnostic interactions with the phosphate backbone (Fig. 2e). For clarity, we number the DNA bases from 5′ to 3′ with base −2 two bases upstream of the unwinding loop and base +8 exiting the DNA tunnel. Adjacent to the unwinding loop, K347 coordinates the phosphates of DNA bases +1 and +2, T443 and the backbone amine of A418 are positioned for hydrogen bonding interactions with the backbone of base +2, and T445 and R425 stabilize the backbone of DNA base +3. As the DNA continues through the tunnel, the DNA backbone approaches D1, where R226 and the backbone amine of V147 stabilize DNA base +4, T190 and the backbone amine of G173 stabilize base +5, and R193 and R200 are positioned for ionic interactions with the backbone of base +6. As the DNA extends toward the tunnel exit, S622 stabilizes the base +6 backbone and the bases point toward the D4 ratchet helix where V757 and M761 wedge between DNA bases +5 and +6 (Extended Data Fig. 6d). DNA bases +7 and +8 make no notable interactions with PolθH as they exit the DNA tunnel between D1 and D4.

Multistage positioning of DNA for microhomology annealing

We can unambiguously trace ssDNA exiting the DNA tunnel from both PolθH protomers and joining as a large globular density in the intradimer valley. The location of this annealed DNA, combined with the external rotation of each D5 clearing a path for 3′ ssDNA overhangs to exit the DNA tunnels and extend toward one another, indicates that the intradimer valley serves as a nexus for microhomology pairing and associated allosteric rearrangements. Thus, we performed additional focused image analyses on this region, giving rise to PolθH reconstructions with 3′ DNA density in three distinct conformations between the PolθH protomers. The different positions of the DNA in these states present a putative three-step process of 3′ DNA microhomology pairing. In state 1, which we refer to as a microhomology 'searching' state, we observe density corresponding to each 3′ overhang exiting the DNA tunnel and extending toward the first helix (residues 792–799) of the protomer's D5 (Fig. 3a). In this configuration, the two ssDNA strands are positioned in an antiparallel arrangement ~30 Å apart. This arrangement enables ssDNA to simultaneously traverse both PolθH protomer DNA tunnels and pass each other with the capacity to pair if there is complementary microhomology. Microhomology-driven interactions between the strands lead to state 2, which represents what we refer to as a 'microhomology aligning' state (Fig. 3b). Here, both 3′ ssDNA overhangs exit the DNA tunnel and approach D5 but then turn away from D5 toward each other to establish a nascent pairing in preparation for state 3, which we call the 'microhomology annealed' state (Fig. 3c). In this state, the ssDNA moves away from D5 and the annealed microhomology forms a linear duplex between the DNA tunnel exits of both PolθH protomers, priming Polθ to continue TMEJ.

Collectively, our structural data (Extended Data Figs. 7–10) present a mechanism by which PolθH fosters ATP-independent 3′ ssDNA microhomology pairing and we sought to support this mechanism biochemically. We observed that PolθH alone, as well as PolθH incubated with DNA containing 3′ overhangs without microhomology, struggles to migrate into a native gel (Fig. 3d). Conversely, PolθH readily runs into a native gel upon addition of DNA containing 3′ overhangs with microhomology. We also noted that introducing 5′ secondary

https://doi.org/10.1038/s41594-025-01514-8

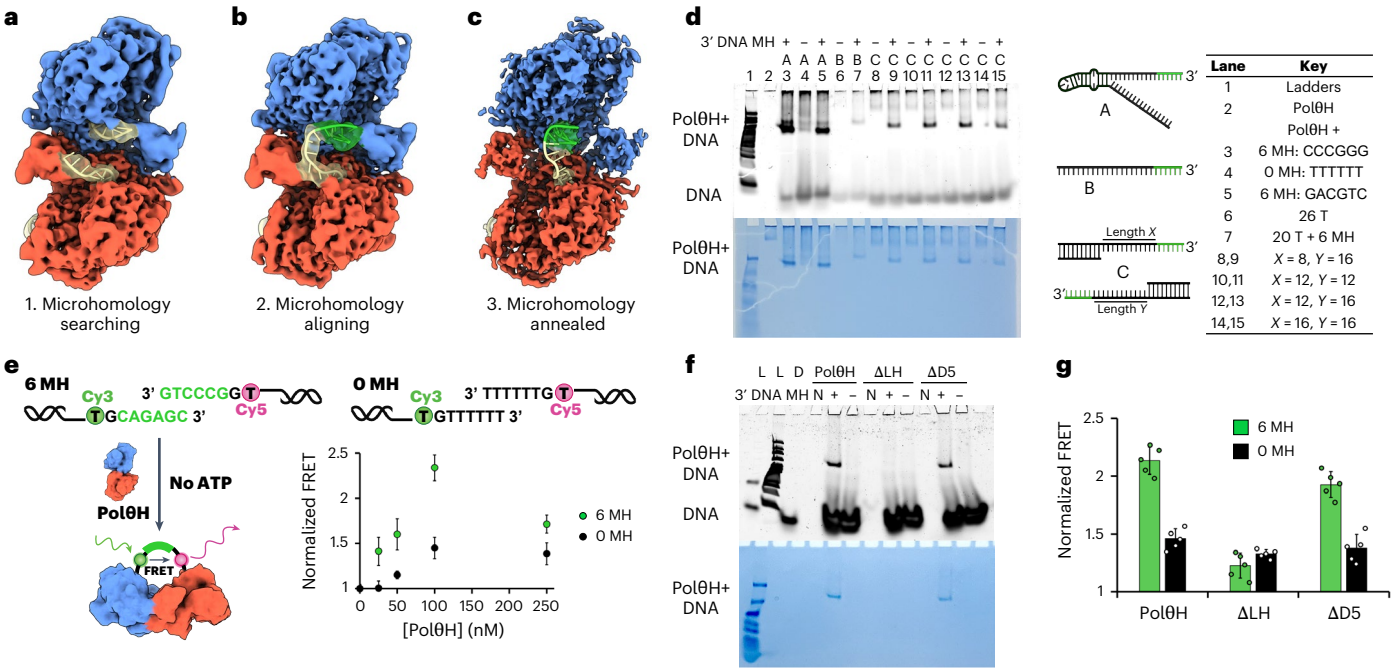


Fig. 3 | PolθH anneals 3′ DNA microhomology in three steps, as revealed by cryo-EM. a–**c**, Cryo-EM maps of DNA-bound PolθH shown in the 3′ microhomology (green) pairing states: searching (**a**), aligning (**b**), and annealed (**c**). The cryo-EM density corresponding to the DNA extending from each DNA tunnel is shown as semitransparent, with the modeled DNA shown as a cartoon within. The searching and aligning maps are shown locally filtered by resolution to emphasize DNA position. **d**, Native PAGE assay illustrating that PolθH migrates more robustly into a native gel if there is 3′ DNA microhomology present. MH, microhomology. A key showing the types of oligonucleotides used is on the right. **e**, FRET assay showing that PolθH brings 3′ DNA overhangs close together more effectively if there is 3′ DNA microhomology present. Data are presented as the mean values ± s.d. ($n = 5$ biological replicates). **f**, Native PAGE assay showing that PolθH and PolθH ΔD5 migrate into a native gel if there is 3′ DNA MH present, while PolθH ΔLH does not. L, ladder; D, DNA only; N, no DNA; +, DNA with 3′ microhomology; −, DNA without 3′ microhomology. **g**, FRET assay showing that PolθH and PolθH ΔD5 bring 3′ DNA overhangs close together more effectively if there is 3′ DNA microhomology present, while PolθH ΔLH does not. Data are presented as bars for mean values ± s.d., along with a point for each biological replicate ($n = 5$ biological replicates).

structure into the DNA results in a more stable interaction with PolθH, possibly because formation of a DNA stem-loop prevents PolθH from fully translocating DNA, giving rise to a stalled complex. We concluded that PolθH can only form a stable complex and migrate into a native gel if there is 3′ microhomology present in the DNA substrate.

We next tested the capacity of PolθH to bring together 3′ ssDNA overhangs with and without microhomology by measuring fluorescence resonance energy transfer (FRET) levels arising from PolθH-mediated joining of Cy3-labeled and Cy5-labeled DNA substrates. PolθH facilitated increased FRET with both oligonucleotide pairs in a concentration-dependent manner in the absence of ATP, but we observed substantially higher FRET for the oligonucleotide pair containing 3′ microhomology (Fig. 3e). Notably, FRET increased with increasing concentrations of PolθH but decreased at a protein concentration higher than 100 nM, indicating that, at higher concentrations, PolθH dimers sequester individual free DNA ends, reducing the level of microhomology annealing[21]. Our observation that PolθH exhibits optimal ssDNA binding and annealing at concentrations ranging from 25 to 100 nM is consistent with previous fluorescence-based studies[21,22,25]. Our FRET data indicate that PolθH brings 3′ DNA ends together independent of DNA microhomology; however, if the DNA overhangs contain microhomology, PolθH facilitates annealing, affording a more stable interaction consistent with a more robust FRET signal. Conversely, the minor FRET signal arising from the DNA without microhomology likely corresponds to a population of PolθH in the DNA searching state, where the overhangs are brought in proximity but are unable to anneal (Fig. 3a).

We additionally leveraged our FRET assay to test the role of D5 in positioning 3′ ssDNA overhangs for annealing using two PolθH mutants: one with D5 completely removed (ΔD5) and another where we only removed the D5 loop residues that organize into a loop–helix as part of the PolθH D5 rearrangement (residues 838–860; ΔLH). Surprisingly, our PolθH ΔD5 construct maintained the ability to facilitate 3′ DNA microhomology annealing in our native PAGE (Fig. 3f) and FRET assays (Fig. 3g). However, our PolθH ΔLH construct did not promote microhomology annealing. On the basis of these data, we surmise that the absence of the D5 barrier enables ssDNA to transverse and exit the PolθH tunnel unencumbered, allowing incoming DNA microhomologies to pair. However, if D5 is present but unable to become locked in an outward position away from the DNA tunnel (Fig. 1c), this domain acts as a molecular barrier that prevents microhomology annealing. These findings are consistent with D5 having a regulatory role in 3′ DNA microhomology searching and annealing, whereby ordering of the 838–860 loop–helix to lock D5 in the outward conformation and stabilize the DNA-bound PolθH dimer (Fig. 1d,f) serves as a key step for initiating the DNA repair process. Together with our structural data, we conclude that the PolθH dimer functions as a platform capable of scanning ssDNA to identify, align, and anneal 3′ DNA microhomologies to initiate repair by TMEJ.

Discussion

Cells carrying mutations in genes involved in DNA damage repair by HR, such as *BRCA1/2*, upregulate Polθ to repair DNA double-strand breaks by TMEJ, a process that is strongly correlated with HR-deficient cancers[9,32]. A crucial step in TMEJ is the Polθ-mediated pairing of 3′ ssDNA microhomologies. Our studies enable us to propose a mechanism by which PolθH searches 3′ DNA to identify, align, and anneal 3′ DNA microhomology without requiring ATP. Furthermore, our native PAGE and FRET assays show that the robustness of this activity is dependent on the presence of 3′ DNA microhomology.

https://doi.org/10.1038/s41594-025-01514-8

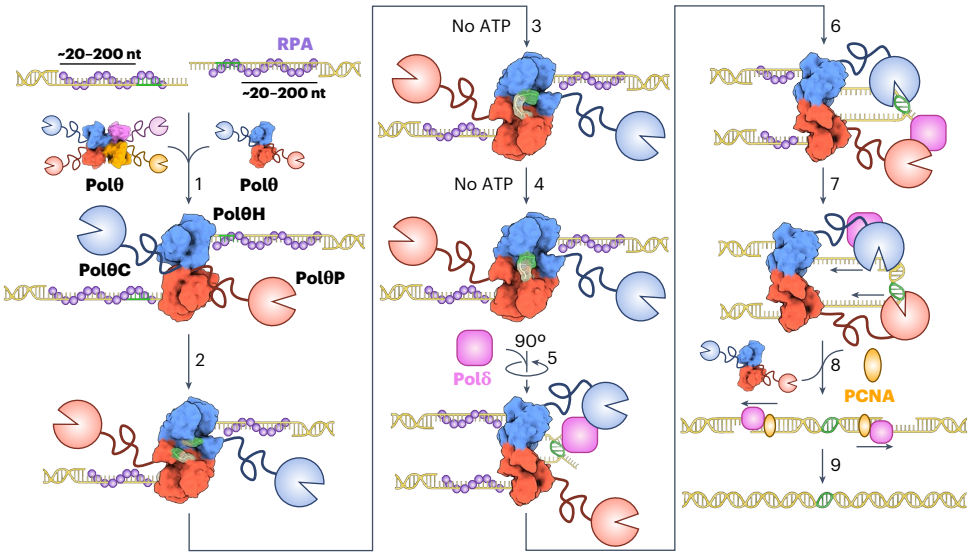


Fig. 4 | Proposed model for TMEJ, initiated by annealing of 3′ DNA microhomology by PolθH. Short-range DNA end resection after a double-strand break leaves the TMEJ substrate with two 3′ ssDNA overhangs coated in ssDNA-binding proteins, such as RPA. PolθH exists in equilibrium between tetrameric and dimeric forms before encountering 3′ ssDNA. (1) PolθH binds each 3′ ssDNA terminus. (2) PolθH translocates 3′ to 5′ on the DNA and removes RPA in an ATP-dependent manner. This DNA translocation facilitates PolθH D5 swinging open to accommodate the DNA. PolθH aligns (3) and anneals (4) 3′ DNA microhomologies in an ATP-independent manner. (5) Polδ trims one 3′ DNA flap and may stabilize the joined microhomology. (6) PolθP attaches to the available 3′ primer terminus formed by the joined microhomology and stabilizes it, and Polδ may trim the opposite 3′ DNA flap. (7) PolθP begins nonprocessive DNA synthesis. This may occur in only one direction or the other copy of PolθP may begin nonprocessive DNA synthesis of the other side of the microhomology. (8) After minimal synthesis by PolθP, Polδ and PCNA perform processive, strand-displacement synthesis before final resolution of the double-strand break (9).

In the context of prior studies, our findings present an updated mechanism for the initial steps of TMEJ (Fig. 4 and Supplementary Video 1). In the absence of DNA, PolθH exists in an equilibrium of tetrameric and dimeric states, both of which have accessible DNA tunnels for binding and initial translocation of the 3′ ends of resected ssDNA. Upon DNA binding and full traversal of the tunnel, interaction between the incoming ssDNA and the PolθH D5 C-terminal helix triggers an outward rotation of each D5. This rearrangement is concomitant with ordering of residues 838–860 into a loop–helix that docks into into the back of the PolθH dimer and locks both D5 protomers in the outward conformation. This D5 arrangement stabilizes the PolθH dimer, disrupts dimer–dimer interactions in the tetrameric form, and allows 3′ DNA exit from the tunnel to establish the microhomology searching conformation observed in our cryo-EM data.

We posit that this dimeric conformation enables 3′-to-5′ DNA translocation and displacement of replication protein A (RPA) from the ssDNA by PolθH in an ATP-dependent manner[23,24]. PolθH serves as a platform to bring DNA ends together; given the 93% likelihood that a 3-base pair microhomology is present within 15 bases of any given pair of DNA ends, DNA microhomology is likely identified and annealed almost immediately after DNA translocation begins[17]. However, given the limited number of base pairing interactions, microhomology annealing is transient and prone to strand dissociation without stabilization by other TMEJ factors such as PolθP or DNA polymerase δ (Polδ) (refs. 21,33). Thus, PolθH has a key role in identifying and positioning regions of 3′ microhomology for subsequent annealing and stabilization for TMEJ, a conclusion that is supported by recent single-molecule FRET studies that characterized the contributions of PolθH and PolθP to TMEJ[21]. Mounting evidence suggests that after microhomology annealing, Polδ trims one unpaired 3′ ssDNA end (flap) and may bind Polθ in a manner that stabilizes the annealed microhomology[33]. After flap trimming, one of the two tethered PolθPs may recognize a 3′ dsDNA microhomology primer, further stabilizing the annealed microhomology, and subsequently begin nonprocessive synthesis[21,34]. Once the first polymerase has gap-filled a sufficient distance from the microhomology, Polδ may trim the other unpaired 3′ ssDNA flap so that the second tethered copy of PolθP can begin synthesis on the opposite 3′ end of the microhomology. After less than 15 nucleotides of synthesis by PolθP, Polδ can perform processive strand-displacement synthesis in complex with the proliferating cell nuclear antigen (PCNA) processivity clamp, triggering final resolution of the DNA break[33].

In the context of this proposed mechanism, Polδ likely works in complex with PolθH and PolθP to stabilize the annealed microhomology, trim unpaired 3′ DNA flaps, and remain in proximity to perform DNA synthesis after PolθP[33]. Interaction with Polδ may also facilitate Polθ exit from the DNA substrate, likely before PolθH encounters 5′ dsDNA. This may explain why Polθ, with its small unwinding loop, has no need to act as a processive helicase. Our proposed mechanism also implicates a role for the disordered PolθC, which likely functions as a tether of defined length that maintains PolθP in proximity to PolθH to recognize paired microhomology while granting the polymerase the degrees of freedom necessary to robustly grasp, stabilize, and start nonprocessive synthesis. We anticipate that these insights into the molecular determinants of the PolθH–DNA interaction, which outline the critical steps required to initiate TMEJ, provide the mechanistic details to enable rational design of future PolθH-targeted cancer therapeutics.

Online content

Any methods, additional references, Nature Portfolio 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/s41594-025-01514-8.

References

1. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
2. Scully, R., Panday, A., Elango, R. & Willis, N. A. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat. Rev. Mol. Cell Biol.* **20**, 698–714 (2019).

3. Ceccaldi, R., Rondinelli, B. & D'Andrea, A. D. Repair pathway choices and consequences at the double-strand break. *Trends Cell Biol.* **26**, 52–64 (2016).
4. Ramsden, D. A., Carvajal-Garcia, J. & Gupta, G. P. Mechanism, cellular functions and cancer roles of polymerase-θ-mediated DNA end joining. *Nat. Rev. Mol. Cell Biol.* **23**, 125–140 (2022).
5. Wood, R. D. & Doublié, S. DNA polymerase θ (POLQ), double-strand break repair, and cancer. *DNA Repair* **44**, 22–32 (2016).
6. Chan, S. H., Yu, A. M. & McVey, M. Dual roles for DNA polymerase θ in alternative end-joining repair of double-strand breaks in *Drosophila*. *PLoS Genet.* **6**, e1001005 (2010).
7. Yousefzadeh, M. J. et al. Mechanism of suppression of chromosomal instability by DNA polymerase POLQ. *PLoS Genet.* **10**, e1004654 (2014).
8. Mateos-Gomez, P. A. et al. Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. *Nature* **518**, 254–257 (2015).
9. Ceccaldi, R. et al. Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. *Nature* **518**, 258–262 (2015).
10. Wyatt, D. W. et al. Essential roles for polymerase θ-mediated end joining in the repair of chromosome breaks. *Mol. Cell* **63**, 662–673 (2016).
11. Zatreanu, D. et al. Polθ inhibitors elicit *BRCA*-gene synthetic lethality and target PARP inhibitor resistance. *Nat. Commun.* **12**, 3636 (2021).
12. Zhou, J. et al. A first-in-class polymerase θ inhibitor selectively targets homologous-recombination-deficient tumors. *Nat. Cancer* **2**, 598–610 (2021).
13. Study of orally administered MOMA-313 in participants with advanced or metastatic solid tumors (NCT06545942). https://clinicaltrials.gov/study/NCT06545942 (2024).
14. A study of ART4215 for the treatment of advanced or metastatic solid tumors (NCT04991480). https://clinicaltrials.gov/study/NCT04991480 (2021).
15. A study to investigate the safety, tolerability, pharmacokinetics (PK), and preliminary anticancer activity of GSK4524101 alone or with niraparib in participants with solid tumors (NCT06077877). https://clinicaltrials.gov/study/NCT06077877 (2023).
16. Seki, M., Marini, F. & Wood, R. D. POLQ (Pol θ), a DNA polymerase and DNA-dependent ATPase in human cells. *Nucleic Acids Res.* **31**, 6117–6126 (2003).
17. Carvajal-Garcia, J. et al. Mechanistic basis for microhomology identification and genome scarring by polymerase θ. *Proc. Natl Acad. Sci. USA* **117**, 8476–8485 (2020).
18. He, P. & Yang, W. Template and primer requirements for DNA Pol θ-mediated end joining. *Proc. Natl Acad. Sci. USA* **115**, 7747–7752 (2018).
19. Luedeman, M. E. et al. Poly(ADP) ribose polymerase promotes DNA polymerase θ-mediated end joining by activation of end resection. *Nat. Commun.* **13**, 4547 (2022).
20. Kent, T., Chandramouly, G., McDevitt, S. M., Ozdemir, A. Y. & Pomerantz, R. T. Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase θ. *Nat. Struct. Mol. Biol.* **22**, 230–237 (2015).
21. Fijen, C. et al. Sequential requirements for distinct Polθ domains during theta-mediated end joining. *Mol. Cell* **84**, 1460–1474 (2024).
22. Black, S. J. et al. Molecular basis of microhomology-mediated end-joining by purified full-length Polθ. *Nat. Commun.* **10**, 4423 (2019).
23. Mateos-Gomez, P. A. et al. The helicase domain of Polθ counteracts RPA to promote alt-NHEJ. *Nat. Struct. Mol. Biol.* **24**, 1116–1123 (2017).
24. Schaub, J. M., Soniat, M. M. & Finkelstein, I. J. Polymerase θ-helicase promotes end joining by stripping single-stranded DNA-binding proteins and bridging DNA ends. *Nucleic Acids Res.* **50**, 3911–3921 (2022).
25. Newman, J. A., Cooper, C. D. O., Aitkenhead, H. & Gileadi, O. Structure of the helicase domain of DNA polymerase θ reveals a possible role in the microhomology-mediated end-joining pathway. *Structure* **23**, 2319–2330 (2015).
26. Ozdemir, A. Y., Rusanov, T., Kent, T., Siddique, L. A. & Pomerantz, R. T. Polymerase θ-helicase efficiently unwinds DNA and RNA–DNA hybrids. *J. Biol. Chem.* **293**, 5259–5269 (2018).
27. Guo, H. et al. Cryo-EM structure of DNA polymerase θ helicase domain in complex with inhibitor novobiocin. Preprint at *bioRxiv* https://doi.org/10.1101/2023.01.20.524915 (2023).
28. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
29. Büttner, K., Nehring, S. & Hopfner, K.-P. Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nat. Struct. Mol. Biol.* **14**, 647–652 (2007).
30. Richards, J. D. et al. Structure of the DNA repair helicase Hel308 reveals DNA binding and autoinhibitory domains. *J. Biol. Chem.* **283**, 5118–5126 (2008).
31. Gyimesi, M., Sarlós, K. & Kovács, M. Processive translocation mechanism of the human Bloom's syndrome helicase along single-stranded DNA. *Nucleic Acids Res.* **38**, 4404–4414 (2010).
32. Schrempf, A., Slyskova, J. & Loizou, J. I. Targeting the DNA repair enzyme polymerase θ in cancer therapy. *Trends Cancer* **7**, 98–111 (2021).
33. Stroik, S. et al. Stepwise requirements for polymerases δ and θ in theta-mediated end joining. *Nature* **623**, 836–841 (2023).
34. Zahn, K. E., Averill, A. M., Aller, P., Wood, R. D. & Doublié, S. Human DNA polymerase θ grasps the primer terminus to mediate DNA repair. *Nat. Struct. Mol. Biol.* **22**, 304–311 (2015).

Methods

Mutagenesis, protein expression, and protein purification

The 10xHis-SUMO-Avi-tagged human PolθH (residues 2–894) was used as parental DNA to generate PolθH mutants. PolθH ΔLH was generated with site-directed deletion of amino acids 838–860 and PolθH ΔD5 was generated by introducing a Q790STOP mutation.

The 10xHis-SUMO-Avi-tagged human PolθH and PolθH ΔLH were expressed in *Escherichia coli* Rosetta (DE3) cells. The cultures were grown in 2× yeast extract–tryptone broth to an optical density at 600 nm of 0.8 at 37 °C, transferred to 16 °C for 1 h and then induced with 0.2 mM IPTG for 18 h. Cells were pelleted and resuspended in lysis buffer (40 mM HEPES pH 7.5, 0.5 M NaCl, 5% glycerol, 0.5 mM TCEP and 1 mM PMSF) and lysed by sonication. The lysate was clarified by centrifugation and the supernatant was incubated with Ni-NTA resin (Qiagen) at 4 °C for 1 h. The mixture was applied to a gravity column, the beads were washed with lysis buffer and lysis buffer plus 20 mM imidazole and the protein was eluted with lysis buffer plus 250 mM imidazole. The eluted proteins were digested with His-tagged tobacco etch virus protease during dialysis in dialysis buffer (40 mM HEPES pH 7.5, 0.5 M NaCl, 5% glycerol, 0.5 mM TCEP, 3 mM β-mercaptoethanol and 0.5 mM EDTA) at 4 °C for 12 h to remove the His–SUMO–Avi tag. The protein mixture was incubated on Ni-NTA resin and loaded onto a gravity column, the flowthrough was collected and the resin was washed with dialysis buffer. PolθH ΔD5 was expressed and recovered the same way, except the lysis and dialysis buffers contained 250 mM NaCl.

For PolθH and PolθH ΔLH, no-salt lysis buffer (40 mM HEPES pH 7.5, 5% glycerol and 0.5 mM TCEP) was added to dilute the protein to 100 mM NaCl and the protein was loaded onto a 5-ml HiTrap Heparin column (Cytiva). The column was washed with heparin buffer A (50 mM HEPES pH 7.5, 100 mM NaCl, 5% glycerol and 0.5 mM TCEP), PolθH was eluted from the column with a gradient of 0–50% heparin buffer B (50 mM HEPES pH 7.5, 1 M NaCl, 5% glycerol and 0.5 mM TCEP) and the column was washed with 50% and then 100% heparin buffer B. Fractions containing PolθH or PolθH ΔLH were pooled and further purified by size-exclusion chromatography (SEC) using a Superdex 200 10/300 GL (Cytiva) column and SEC buffer (10 mM HEPES pH 7.5, 250 mM NaCl and 0.5 mM TCEP). PolθH or PolθH ΔLH were collected, concentrated, aliquoted, flash-frozen, and stored in SEC buffer at −70 °C.

For PolθH ΔD5, no-salt lysis buffer was added to dilute the protein to 50 mM NaCl and the protein was loaded onto the 5-ml HiTrap Heparin column. The column was washed with Heparin buffer A with 50 mM NaCl, PolθH ΔD5 was eluted from the column with a gradient of 0–20% heparin buffer B and the column was washed with 20% and then 100% heparin buffer B. Fractions containing PolθH ΔD5 were pooled and further purified by SEC using the Superdex 200 10/300 GL column and SEC buffer with 150 mM NaCl. PolθH ΔD5 was collected, concentrated, aliquoted, flash-frozen, and stored in 150 mM NaCl SEC buffer at −70 °C.

Native PAGE

All DNA oligonucleotides used in this study are shown in Supplementary Table 1. Frozen PolθH protein in SEC buffer was thawed on ice and 2× preannealed oligonucleotides (Integrated DNA Technologies) were added. The mixture was incubated on ice for 20 min, NaCl concentration was lowered to 100 mM over the course of 1 h with addition of no-salt SEC buffer (10 mM HEPES–NaOH pH 7.5 and 0.5 mM TCEP) and the mixture was incubated on ice for 20 min. Glycerol was added to a final concentration of 5% and samples were loaded on a 4.5% native PAGE gel that had prerun at 75 V for 90 min in 0.2× TBE buffer pH 7.6 at 4 °C. The gel ran at 75 V for 90 min at 4 °C, was rinsed with water, and was then placed in a 50-ml solution of 1x Diamond nucleic acid dye (Promega) in 0.2× TBE buffer pH 7.6 to rock in the dark for 30 min. DNA bands were visualized on a ChemiDoc MP imaging system (Bio-Rad). The gel was then rinsed with water, incubated in Coomassie protein stain, and destained; protein bands were visualized on a lightbox.

Preparation of PolθH–DNA complex

PolθH was thawed on ice and 2× preannealed oligonucleotides (Integrated DNA Technologies) were added. The mixture incubated on ice for 30 min, NaCl concentration was lowered to 100 mM over the course of 1 h with addition of no-salt SEC buffer, and the mixture incubated on ice for 30 min. The resulting complex was loaded onto a Superdex 200 10/300 GL (Cytiva) column pre-equilibrated with low-salt SEC buffer (10 mM HEPES–NaOH pH 7.5, 100 mM NaCl and 0.5 mM TCEP) for SEC at a flow rate of 0.5 ml min^{-1}. Fractions containing protein were collected and analyzed using native PAGE. Fractions with both protein and DNA were collected and concentrated. Concentration was determined using a BCA assay and samples were immediately prepared for cryo-EM.

Preparation of samples for cryo-EM

Samples were kept at 4 °C throughout the sample preparation process. Apo PolθH was diluted to 0.9 mg ml^{-1} in SEC buffer, PolθH–DNA complex was diluted to 1.25 mg ml^{-1} in low-salt SEC buffer, and samples were subjected to centrifugation at 15,000*g* for 10 min. Quantifoil 300 mesh R 1.2/1.3 holey carbon films were glow-discharged under vacuum for 30 s at 15 mA in a Pelco easiGlow 91000 glow discharge cleaning system (Ted Pella). The sample (3 µl) was applied to the surface of the grid and blotted with Whatman 1 filter paper until 2 s after the liquid spot on the filter paper stopped spreading. The grid was then immediately plunged into a liquid ethane pool cooled by liquid nitrogen using a manual plunge-freezer in a 4 °C cold room with >95% humidity. Grids were then clipped under liquid nitrogen in preparation for imaging.

Cryo-EM data acquisition

Cryo-EM data were acquired using a Titan Krios transmission electron microscope with a field emission gun operating at 300 keV. Movies were collected using a Gatan K3 BioQuantum direct electron detector with the energy filter set to 20 eV, operated in electron counting mode. The Leginon data collection software (version 3.6)[35] was used to collect micrographs at 105,000× nominal magnification (0.833 Å per pixel at the specimen level) with a nominal defocus set to 1.0 µm under focus. Apo PolθH videos were collected with the slit out at an exposure rate of 38.8 e^- per pixel per s, with 30 frames (30 ms each) over 0.9 s, for a total electron exposure of 50 e^- per $Å^2$. PolθH–DNA videos were collected with the slit in at an exposure rate of 24.8 e^- per pixel per s, with 47 frames (30 ms each) over 1.4 s, for a total electron exposure of 50 e^- per $Å^2$. Stage movement was used to target the center of four 1.2-µm holes for focusing and an image shift was used to acquire high-magnification images in the center of a larger array of targeted holes. Detailed dataset specifics are available in Tables 1 and 2.

Data analysis of apo PolθH

The image-processing pipeline for the apo dataset is shown in Extended Data Figs. 7 and 8. Frames from cryo-EM videos were aligned and combined applying a dose-weighting scheme using MotionCor2 (ref. 36) with default settings and transferred to cryoSPARC live versions 4.2.1–4.3.0 (ref. 37) for contrast transfer function (CTF) estimation with patchCTF using default parameters. A total of 4,164 micrographs were collected, micrographs with reported CTF resolutions lower than 6.5 Å were discarded, and 3,906 micrographs were exported to cryoSPARC versions 4.2.1–4.3.0 (ref. 38). Particles were picked using a blob picker and an initial round of two-dimensional (2D) classification was used to generate templates for template picking with a 130-Å particle diameter and 0.7× diameter minimum separation distance. A total of 1,643,557 particles were picked using template-based picking and 1,474,207 particles were extracted with a 320-pixel box without any Fourier binning. All particles were subject to minimal cleanup based on 2D classification using 70 classes and four initial models were generated ab initio from 1,436,257 particles contributing to selected class averages showing detailed secondary structure.

Tetrameric, dimeric, and junk populations were identified from the same population of selected particles using two rounds of heterogeneous refinement (401,032 tetramer and 399,289 dimer particles). Each population was refined using nonuniform refinement with optimized per-particle defocus and optimized per-group CTF parameters, with and without symmetry applied (tetramer, D_2; dimer, C_2)[39]. Symmetry was validated by rigid-body fitting the corresponding C_1 maps into the symmetrized maps in ChimeraX version 1.6.1 (UCSF)[40,41]. The symmetry-imposed dimer and tetramer particles were symmetry-expanded (1,604,128 tetramer protomer particles and 798,578 dimer protomer particles) and a mask around a single protomer in each structure was used to perform a local refinement with rotation and shift search extents of 1° and a maximum alignment resolution of 0.1° to improve the resolution of the protomer. The refined tetramer protomer density was further refined using three-dimensional (3D) variability analysis in cluster mode using ten clusters to remove junk protomers. A total of 1,233,292 intact tetramer protomers were subjected to 3D classification without alignment using three classes, a target resolution of 8 Å, an initial O-EM learning rate of 0.2, principal component analysis (PCA) initialization, forced hard classification, and a class similarity of 0.1. Then, 3D variability analysis was performed in cluster mode using ten clusters. The refined dimer protomer density was refined using 3D variability analysis in cluster mode with five clusters.

A small subset of protomer particles from the tetrameric and dimeric populations (208,545 tetramer and 138,712 dimer particles) were chosen for masked local refinement to generate final protomer maps on the basis of homogeneous structural density across the entire particle, quality of map features, and estimated resolution by Fourier shell correlation (FSC) at a cutoff of 0.143. For each population, a locally refined protomer map and a protomer mask were opened in ChimeraX and the command 'volume multiply' was performed to multiply the map and mask values pointwise. The resulting protomer maps were copied (four copies for the tetramer and two copies for the dimer) and fit into the corresponding D_2 and C_2 symmetric maps in ChimeraX. The command 'volume resample' on the combined map grid was used to produce a combined tetramer map and a combined dimer map. These maps were deposited to the EM Data Bank (EMDB) along with the D_2 symmetric tetramer map, the C_2 symmetric dimer map, the locally refined tetramer protomer map, and the locally refined dimer protomer map, as recommended by cryo-EM experts[42]. All map-based parameters for the apo dataset refer to the locally refined individual protomer maps. Sphericity and resolution of protomer maps were evaluated by the remote 3DFSC processing server and PHENIX mtriage[43,44]. Local resolution plots were generated within cryoSPARC for each reconstruction and are found in Supplementary Fig. 1. These plots were used to determine local resolution ranges of both protomer maps using 'report value at mouse position' within the ChimeraX surface color tool.

Data analysis of PolθH–DNA

The PolθH–DNA dataset image-processing pipeline is shown in Extended Data Figs. 9 and 10. Frames from cryo-EM movies were aligned and combined applying a dose-weighting scheme using MotionCor2 (ref. 36) and transferred to cryoSPARC live versions 4.2.1–4.3.0 (ref. 37). A total of 12,440 micrographs were collected, CTF estimation was performed with live patchCTF using default parameters, and micrographs with reported CTF resolutions lower than 5.7 Å were discarded, leaving 12,201 micrographs. Particles were picked using a blob picker and an initial round of live 2D classification was used to generate templates for template picking with a 120-Å particle diameter and 0.7× diameter minimum separation distance. A total of 5,424,691 particles were picked using template picker, extracted with a 324-pixel box without any Fourier binning, and exported to cryoSPARC for minimal cleanup based on 2D classification with 120 classes. Three maps were generated ab initio from a subset of the 5,050,169 particles contributing to selected class averages and one was chosen as an initial model on the basis of the presence of DNA density between two PolθH protomers. This initial model, along with trash classes, was used to seed iterative heterogeneous refinement of all quality particles to remove classes with no density for DNA or D5. The remaining 1,111,072 particles were subject to 3D variability analysis in cluster mode using ten clusters, which revealed different DNA positions among PolθH–DNA classes. A total of 413,973 particles in the microhomology annealed state were subject to an additional round of 3D variability analysis in cluster mode using eight clusters and then masked 3D classification was performed without alignment using two classes, a target resolution of 10 Å, an initial O-EM learning rate of 0.2, PCA initialization, forced hard classification, and a class similarity of 0.1 to identify a population with the strongest density for annealed microhomology.

Populations in nonannealed DNA states were combined and subject to 3D variability analysis in cluster mode using five clusters and then heterogeneous refinement to separate populations in DNA searching and aligning states. Subsets of particles in each microhomology pairing state were chosen as final models selected for homogeneity across the entire particle, quality of map features, and estimated resolution by FSC at a cutoff of 0.143. Per-particle defocus and up to fourth-order optical aberrations were corrected in cryoSPARC's nonuniform refinement protocol to generate sharpened maps of each particle population. Local resolution plots were generated within cryoSPARC for each reconstruction and are found in Supplementary Fig. 1. These plots seeded cryoSPARC local filtering jobs to produce maps that were locally filtered by resolution and were used to determine local resolution ranges of each sharpened map using 'report value at mouse position' within the ChimeraX surface color tool. Sphericity and resolution of protomer maps were evaluated by the remote 3DFSC processing server and PHENIX mtriage[43,44].

Atomic model building and refinement

Model building and refinement were initiated with published models for PolθH (Protein Data Bank (PDB) 5A9J)[25] and *A. fulgidus* Hel308 (PDB 2P6R)[29]. For the apo reconstructions, one protomer was modeled into its corresponding protomer map and the resulting models were symmetry-expanded to fit their combined maps in ChimeraX. The combined models were then further refined to fit their respective combined maps. For all models, iterative rounds of model building and refinement were performed in Coot version 0.9.8.1 EL (ref. 45), ISOLDE version 1.6.0 (ref. 46), and PHENIX version 1.20.1 (ref. 47) until reasonable agreement between the model and data was achieved. Finally, after Ramachandran parameters, rotamers, and clashes were satisfied in ISOLDE, ISOLDE command 'isolde write phenixRsrInput #<model><map resolution> #<map>' was used to export the model and generate a rigid-body refinement settings file for PHENIX real-space refinement. PHENIX real-space refinement was performed with the output model, rigid-body settings file, and sharpened map and PHENIX comprehensive validation (cryo-EM) was performed using the map and output model to yield map and model parameters. Quality of model fitting to the map was evaluated in per-residue correlation coefficient (CC) outputs from Phenix as shown in Supplementary Fig. 2, as well as scores from MolProbity[48], EMRinger[49], and Q score[50] shown in Tables 1 and 2. Q scores were calculated with $\sigma = 0.4$. ChimeraX was used to interpret the EM reconstructions and atomic models and to generate figures.

FRET

Frozen PolθH protein in SEC buffer was thawed on ice. Individually annealed Cy3 and Cy5 oligonucleotides (Integrated DNA Technologies) were mixed in a 1:1 ratio, protein was added, and the samples were incubated on ice for 20 min. FRET buffer (25 mM HEPES–NaOH pH 7.5, 100 mM NaCl, 1 mM $MgCl_2$, 0.01% NP-40, 5% glycerol and 2 mM DTT) was added to dilute the sample to the indicated protein concentration along

with 25 nM FRET oligonucleotide substrate. The samples equilibrated at room temperature for 20 min. Then, 50 µl of each sample was added to a black 384-well plate (Greiner Bio-One 781900). FRET (excitation, 550 nm; emission, 668 nm) was read on a Biotek Synergy H1 microplate reader (Agilent) with Gen5 software version 3.02.

Reporting summary

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

Data availability

Cryo-EM maps and associated atomic models were deposited to the EMDB and the PDB, respectively, with the following accession codes: apo PolθH tetramer, EMD-44534, PDB 9BH6; apo PolθH dimer, EMD-44535, PDB 9BH7; PolθH–DNA microhomology searching, EMD-44536, PDB 9BH8; PolθH–DNA microhomology aligning, EMD-44537, PDB 9BH9; PolθH–DNA microhomology annealed, EMD-44538, PDB 9BHA. The D_2 symmetric tetramer map and the locally refined tetramer protomer map are available as additional maps under EMD-44534. The C_2 symmetric dimer map and the locally refined dimer protomer map are available as additional maps under EMD-44535. Source data are provided with this paper.

References

35. Cheng, A. et al. Leginon: new features and applications. *Protein Sci.* **30**, 136–150 (2021).
36. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
37. Punjani, A. Real-time cryo-EM structure determination. *Microsc. Microanal.* **27**, 1156–1157 (2021).
38. 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).
39. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221 (2020).
40. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
41. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
42. Kleywegt, G. J. et al. Community recommendations on cryoEM data archiving and validation. *IUCrJ* **11**, 140–151 (2024).
43. Tan, Y. Z. et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017).
44. Afonine, P. V. et al. New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr. D* **74**, 814–840 (2018).
45. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
46. Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. D* **74**, 519–530 (2018).
47. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D* **74**, 531–544 (2018).
48. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
49. Barad, B. A. et al. EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015).
50. Pintilie, G. et al. Measurement of atom resolvability in cryo-EM maps with Q-scores. *Nat. Methods* **17**, 328–334 (2020).

Acknowledgements

We thank J. C. Ducom at Scripps Research high-performance computing and C. Bowman at Scripps Research for computational support, as well as W. Lessin at the Scripps Research EM facility for microscopy support. Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health (NIH) under award number F32CA288144 (C.J.Z.). G.C.L. is supported by NIH grant GM14305 and the work used equipment supported by NIH grant S10OD032467.

Author contributions

C.J.Z. prepared all the cryo-EM samples, collected the data, produced the high-resolution structures, built all the atomic models, performed all the biochemical experiments, and wrote the paper. Y.B., B.A.S.-A., and T.G. provided initial support on the PolθH–DNA binding studies and the native PAGE assay. C.J.Z. and G.C.L. designed all the experiments and performed all the mechanistic interpretation. G.C.L. provided guidance in the cryo-EM data collection and analyses and edited the paper.

Competing interests

Y.B., B.A.S.-A. and T.G. are employees of MOMA Therapeutics. The remaining authors declare no competing interests.

Additional information

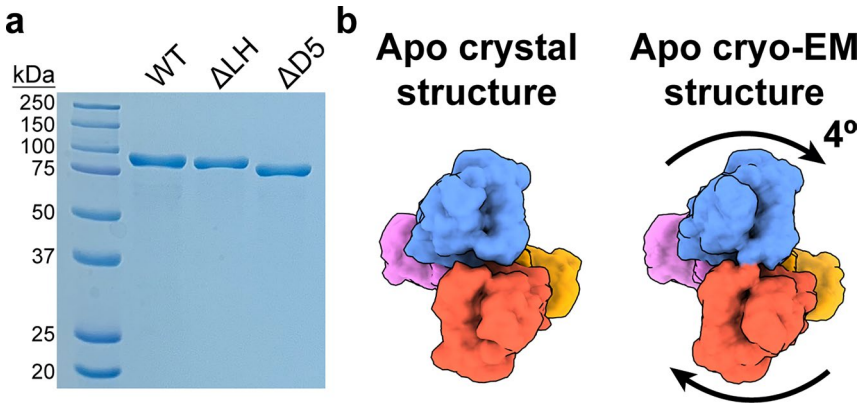
Extended data is available for this paper at https://doi.org/10.1038/s41594-025-01514-8.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41594-025-01514-8.

Correspondence and requests for materials should be addressed to Gabriel C. Lander.

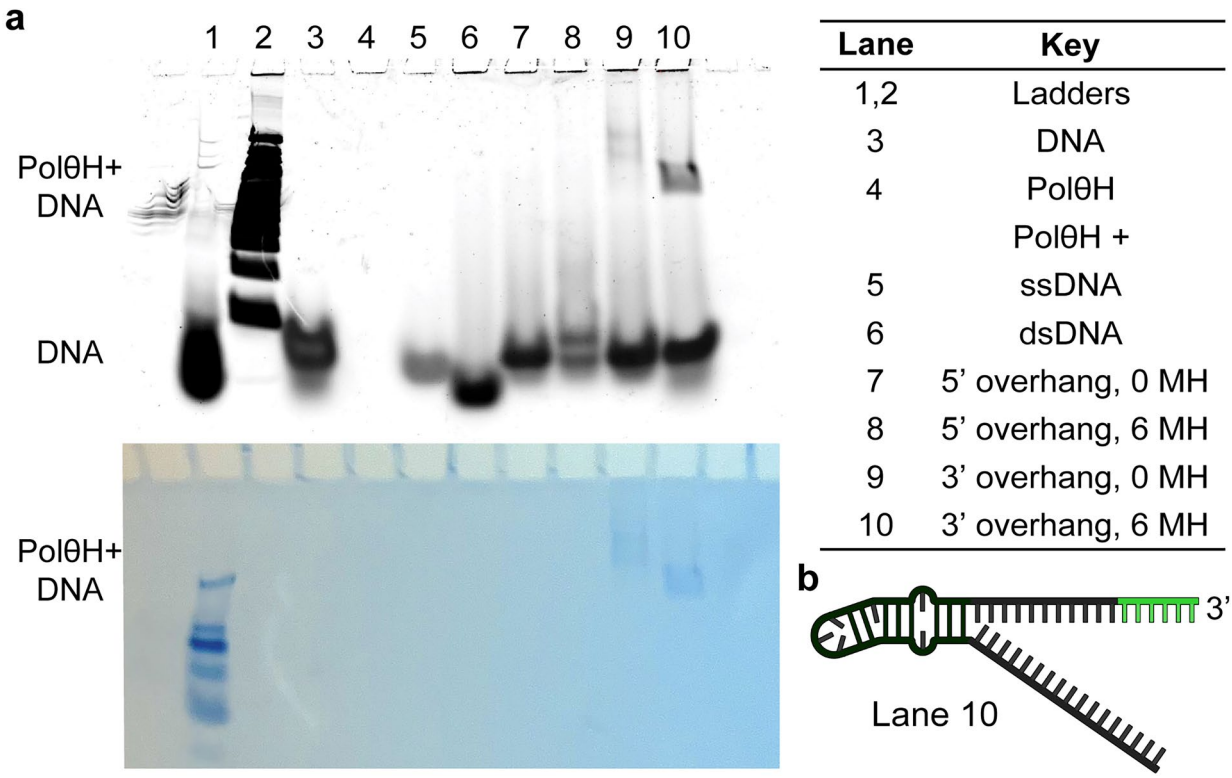
Peer review information *Nature Structural & Molecular Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Dimitris Typas, in collaboration with the *Nature Structural & Molecular Biology* team.

https://doi.org/10.1038/s41594-025-01514-8



Extended Data Fig. 1 | Comparison of apo PolθH tetramer cryo-EM and crystal structures. a, Coomassie-stained SDS-PAGE of proteins used in this study. This experiment was independently repeated three times with similar results. WT: PolθH amino acids 2–894. ΔLH: PolθH with amino acids 838–860 deleted. ΔD5: PolθH amino acids 2–789. 2 µg of protein was loaded in each well. **b**, In the PolθH tetramer cryo-EM structure, two dimeric PolθH subunits are rotated about 4° with respect to the crystal structure (PDB: 5A9J). Residues 35–66 are omitted from the crystal structure representation.

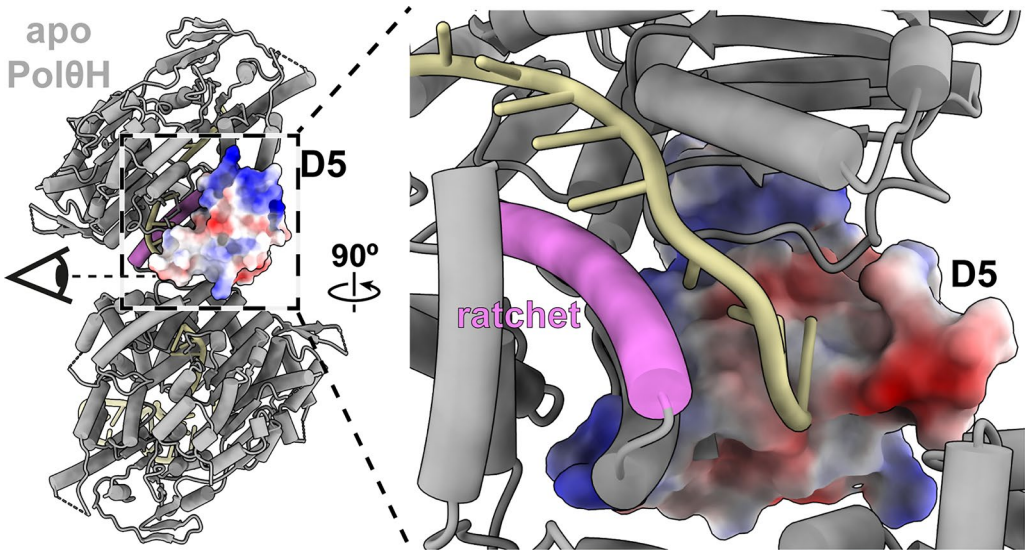
https://doi.org/10.1038/s41594-025-01514-8



Lane	Key
1,2	Ladders
3	DNA
4	PolθH
	PolθH +
5	ssDNA
6	dsDNA
7	5' overhang, 0 MH
8	5' overhang, 6 MH
9	3' overhang, 0 MH
10	3' overhang, 6 MH

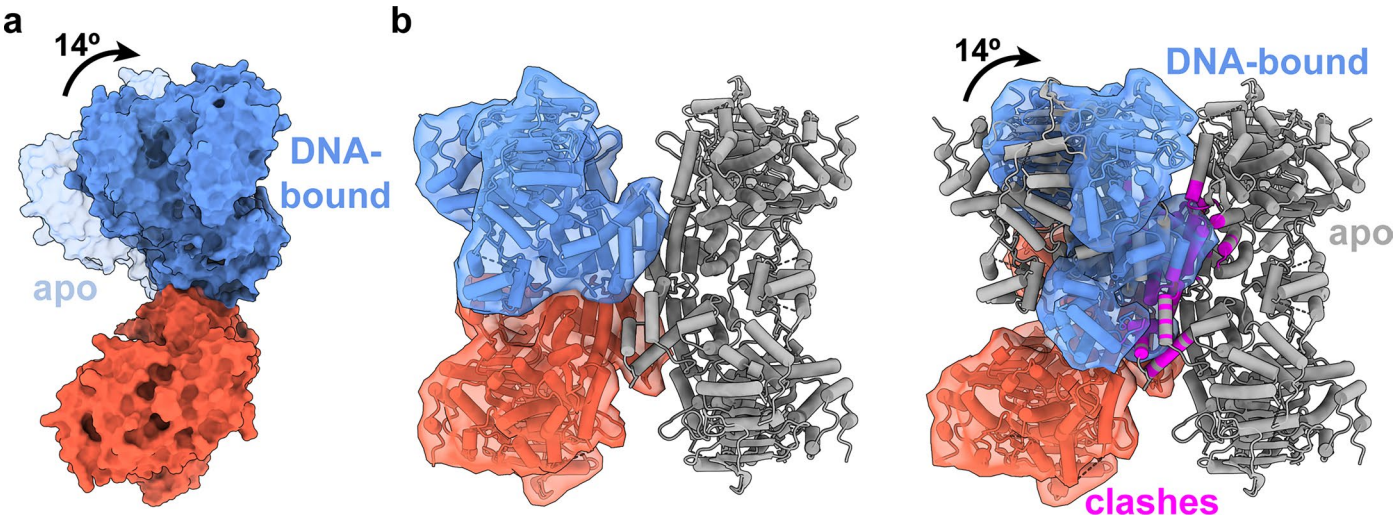
Extended Data Fig. 2 | Native PAGE screening for suitable substrates for PolθH-DNA cryo-EM studies. a, An example gel from Native PAGE screening for DNA substrates that bind PolθH, consecutively stained with Diamond DNA stain (top) and Coomassie protein stain (bottom). This experiment was independently repeated three times with similar results. Gel migration of each species is labeled on the left, and the key is on the right. MH = microhomology. **b**, The stem-loop DNA species pursued for structural studies, with 6 bases of self-complementary microhomology (green) at the end of the 3′ overhang.

https://doi.org/10.1038/s41594-025-01514-8



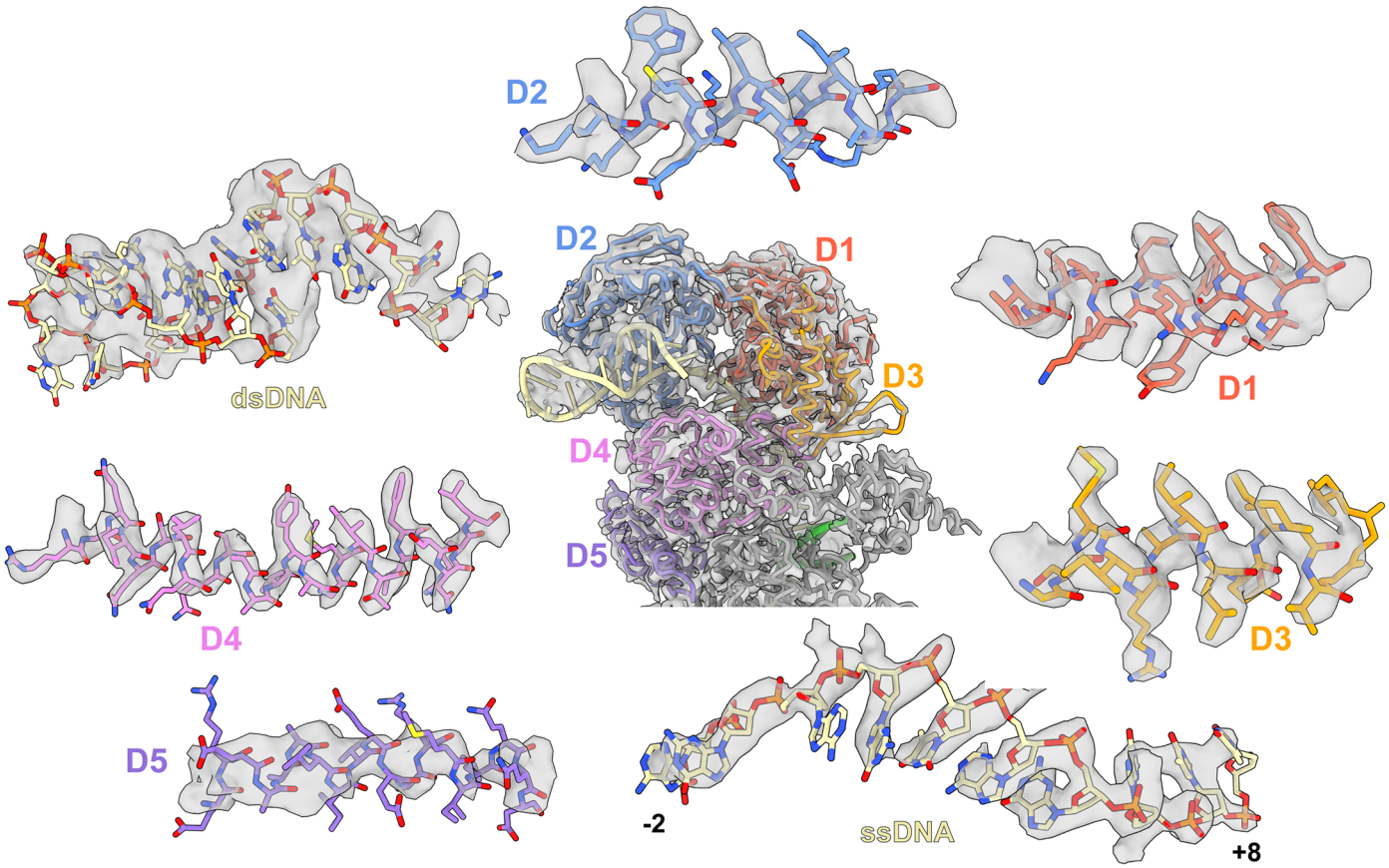
Extended Data Fig. 3 | 3′ ssDNA encounters a negatively charged patch on the C-terminal helix of D5 at the PolθH DNA tunnel exit. DNA from the PolθH-DNA searching model is merged with the apo PolθH dimer model. The D5 surface of one PolθH protomer is colored by electrostatic potential, and the ratchet helix is colored pink.

https://doi.org/10.1038/s41594-025-01514-8



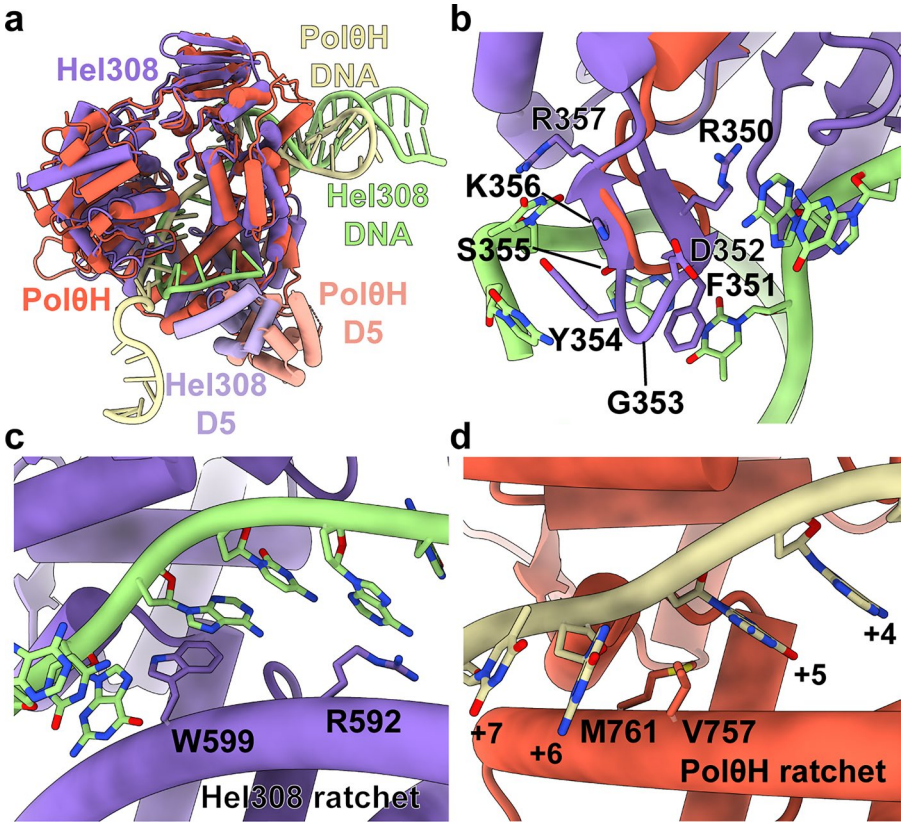
Extended Data Fig. 4 | DNA binding to PolθH causes two protomers to hinge apart, destabilizing the PolθH tetramer. a, Surface representations of the atomic models of the apo and DNA-bound PolθH dimers are shown after aligning the lower subunit (colored red) of each. The upper subunits are colored blue, with the apo subunit semi-transparent. Upon DNA-binding, the two PolθH protomers hinge 14° apart from each other. For clarity, DNA and D5 (residues 790–894) are omitted from each surface representation. **b**, On the left, the apo tetramer is shown as a gray tube representation, with one dimer colored as in **a** and highlighted with a semi-transparent molecular surface. On the right, a DNA-bound dimer (colored and highlighted with DNA omitted) is overlaid on an apo tetramer (gray), aligned to the lower red protomer. The overlay demonstrates that upon DNA binding, the hinging and the D5 conformational change introduce steric clashes with the adjacent and diagonal tetramer protomers (clashing residues colored magenta), destabilizing the PolθH tetramer. Atoms less than 2 Å apart are considered to be clashing.

https://doi.org/10.1038/s41594-025-01514-8



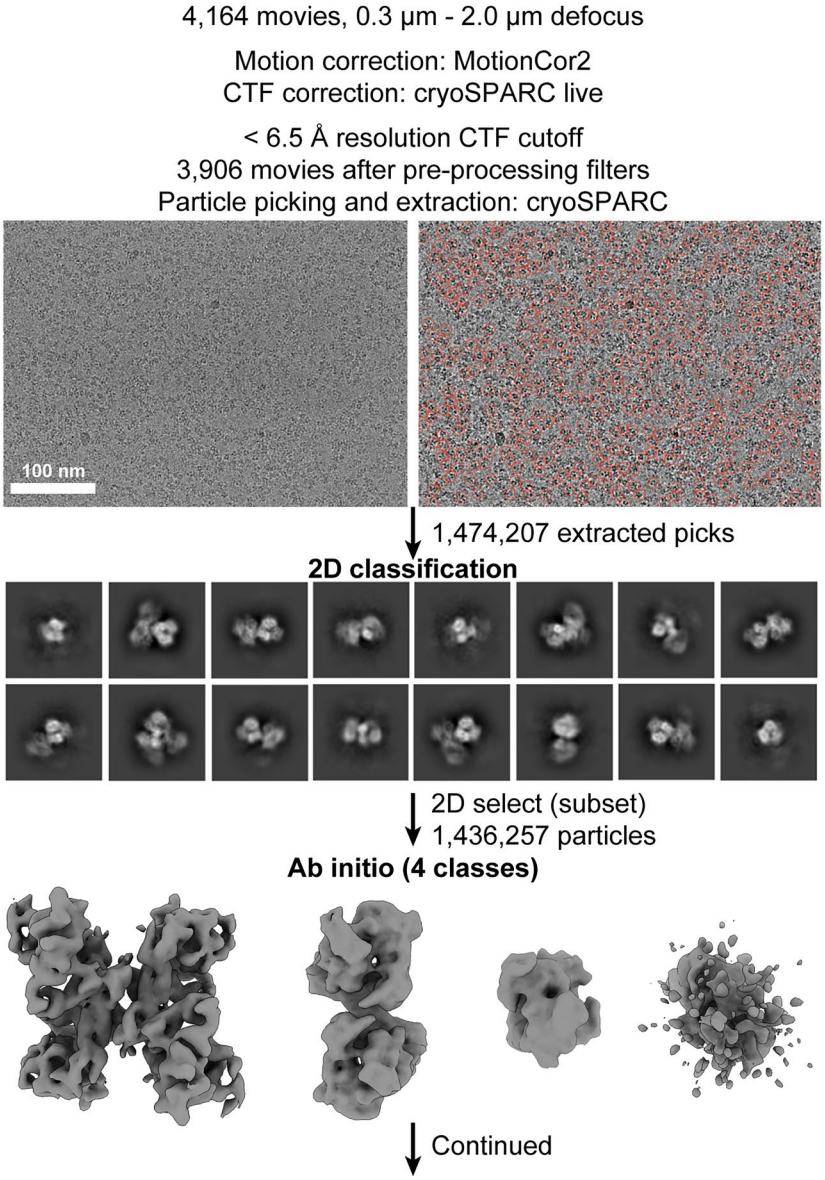
Extended Data Fig. 5 | Representative cryo-EM density of PolθH. Cryo-EM map (DNA microhomology annealed state) and model of one DNA-bound PolθH protomer with domains colored, surrounded by representative map density from alpha helices in each PolθH domain. Domain 1 (D1): residues 147–161. Domain 2 (D2): residues 347–360. Domain 3 (D3): residues 529–540. Domain 4 (D4): residues 745–769. Domain 5 (D5): residues 872–890. ssDNA: bases -2 to +8. dsDNA is in the same pose as in the central protomer.

https://doi.org/10.1038/s41594-025-01514-8



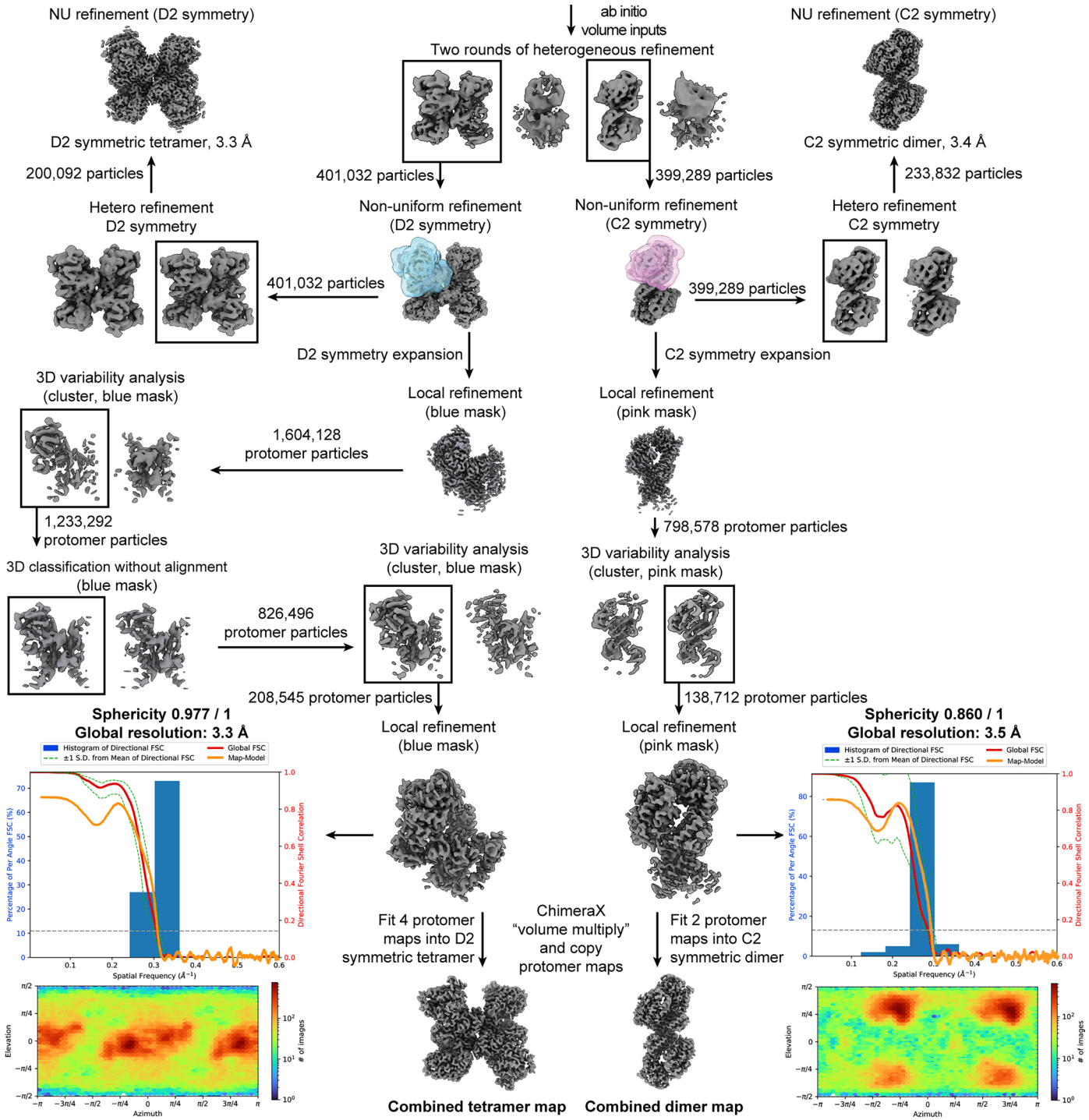
Extended Data Fig. 6 | Comparison of DNA-bound structures of PolθH and *A. fulgidus* Hel308. a, PolθH-DNA structure (red and yellow) overlaid with *A. fulgidus* Hel308-DNA structure (PDB: 2P6R, purple and green). **b**, The unwinding loop of *A. fulgidus* Hel308 (purple, with residues shown) is larger than the equivalent PolθH loop (red). In the *A. fulgidus* Hel308 unwinding loop, F351 and Y354 are positioned to form pi-stacking interactions with DNA bases. Some bases in the DNA duplex have been hidden for clarity. **c**, R592 and W599 in the *A. fulgidus* Hel308 ratchet helix are positioned to stack with DNA. **d**, V757 and M761 in the PolθH ratchet helix wedge between DNA bases +5 and +6.

https://doi.org/10.1038/s41594-025-01514-8



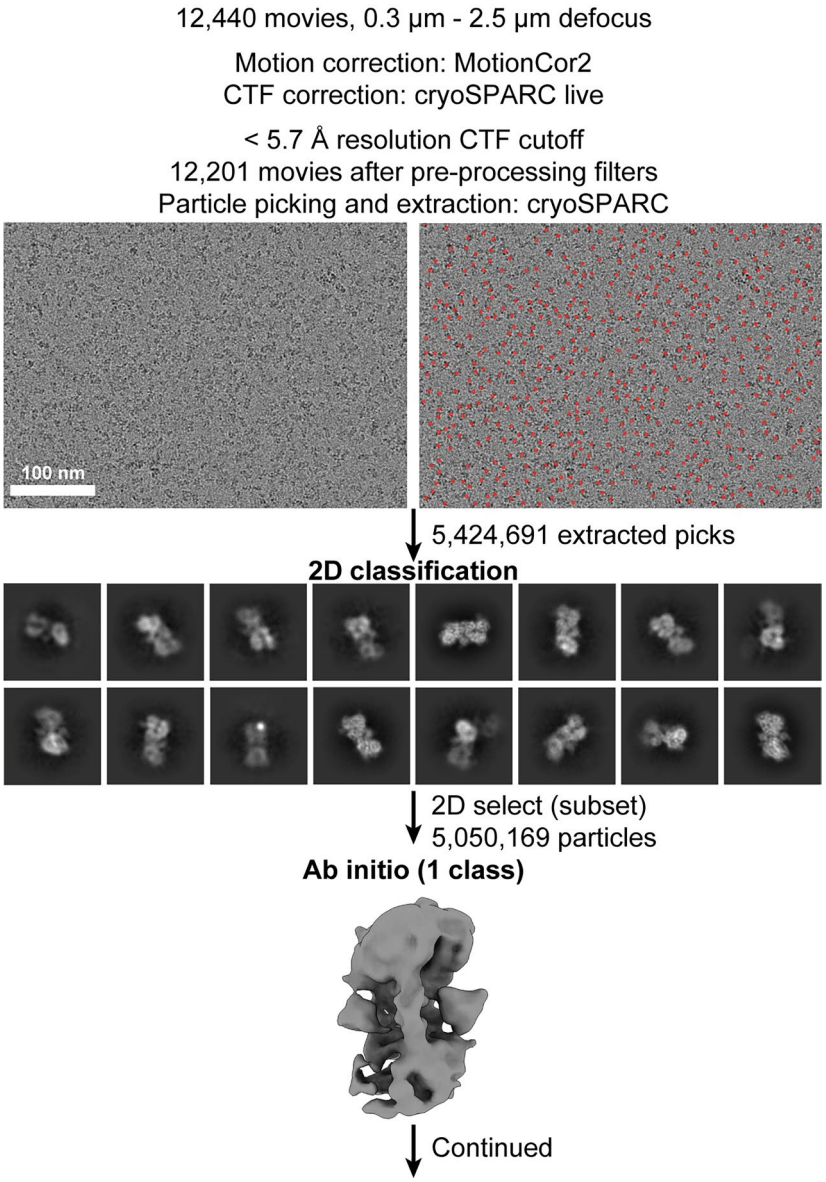
Extended Data Fig. 7 | Apo cryo-EM data collection and initial processing. Data collection and initial processing described for apo PolθH. CTF correction was performed in cryoSPARC live, and micrographs were exported to cryoSPARC for particle picking, extraction, and further processing based on particle presence in 2D and map homogeneity in 3D as described in the methods. One representative micrograph is shown with picked particles in red. For 2D sorting jobs, representative classes are shown. Initial models were created Ab initio from particles.

https://doi.org/10.1038/s41594-025-01514-8



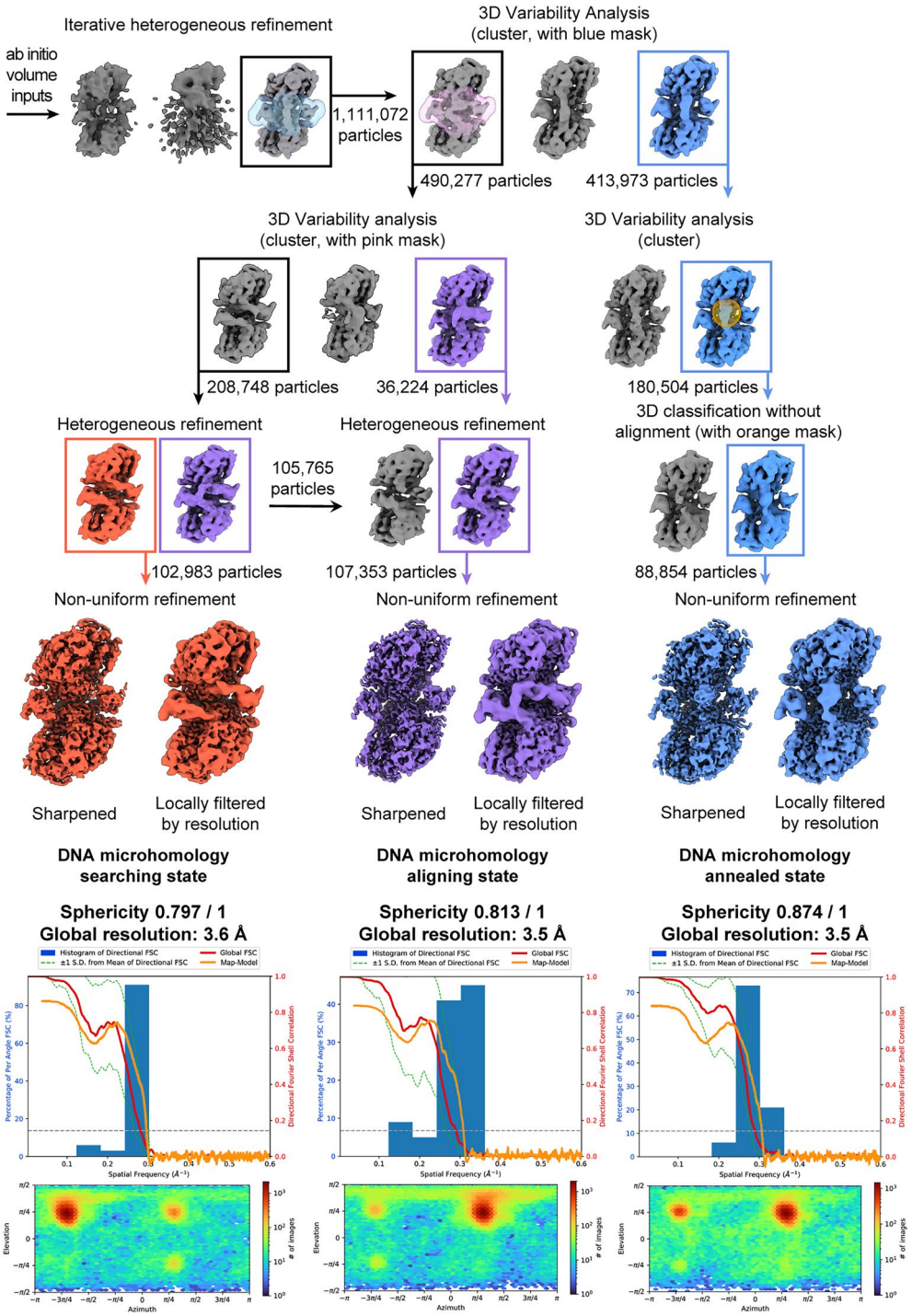
Extended Data Fig. 8 | Apo cryo-EM image analysis workflow. Representative processing workflow described for apo PolθH. For 3D sorting jobs, representative classes are shown. In this scheme, Ab initio models were utilized as volume inputs for iterative heterogeneous refinement, which separated tetrameric and dimeric PolθH particles. Non-uniform refinement was performed with symmetry imposed for both species to produce symmetric reconstructions. In addition, one protomer from each species was masked and symmetry expanded, subject to 3D variability analysis and/or 3D classification without alignment, and a consensus protomer of both species was locally refined. This yielded final protomer reconstructions both shown adjacent to an angular distribution plot and a 3DFSC histogram with the map-to-model FSC overlaid. To generate each final 'combined' map, the locally refined protomer was multiplied by its protomer mask, copied, and fit to the appropriate symmetric map.

https://doi.org/10.1038/s41594-025-01514-8



Extended Data Fig. 9 | PolθH-DNA cryo-EM data collection and initial processing. Data collection and initial processing described for DNA-bound PolθH. CTF correction was performed in cryoSPARC live, and micrographs were exported to cryoSPARC for particle picking, extraction, and further processing based on particle presence in 2D and map homogeneity in 3D as described in the methods. One representative micrograph is shown with picked particles in red. For 2D sorting jobs, representative classes are shown. The initial model was created Ab initio from particles.

https://doi.org/10.1038/s41594-025-01514-8



Extended Data Fig. 10 | PolθH-DNA cryo-EM image analysis workflow. Representative processing workflow described for DNA-bound PolθH. For 3D sorting jobs, representative classes are shown. In this scheme, the Ab initio model was utilized as a volume input for iterative heterogeneous refinement. Particles with apparent DNA density were subject to iterative 3D variability analysis with a mask encompassing the most heterogeneous regions to eliminate particles with no D5 density and poor DNA density. Particles were sorted into three groups based on DNA conformation (colored) and subject to non-uniform refinement to produce final reconstructions of PolθH in three different DNA pairing states, each above an angular distribution plot and a 3DFSC histogram with the map-to-model FSC overlaid. As some DNA density disappears upon map sharpening, a locally filtered map is also shown for each reconstruction.

nature portfolio

Corresponding author(s): Gabriel Lander

Last updated by author(s): 01/09/2024

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our Editorial Policies 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.

n/a	Confirmed
☐	☒ 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	Leginon 3.6, Biotek Gen5 3.02
Data analysis	MotionCor2, CryoSPARC 4.2.1 - 4.3.0, ChimeraX 1.6.1, Coot 0.9.8.1 EL, ISOLDE 1.6.0, Phenix 1.20.1

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Cryo-EM maps and associated atomic models were deposited to the Electron Microscopy Databank (EMDB) and the Protein Databank (PDB), respectively, with the following EMDB and PDB IDs: apo PolθH tetramer - EMD-44534, 9BH6; apo PolθH dimer - EMD-44535, 9BH7; PolθH-DNA microhomology searching - EMD-44536, 9BH8; PolθH-DNA microhomology aligning - EMD-44537, 9BH9; PolθH-DNA microhomology annealed - EMD-44538, 9BHA.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

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.

☒ 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	For cryo-EM, enough particles were collected to achieve necessary resolution to answer the biological question, as is standard in the field. For FRET, five biological replicates were collected for each experiment.
Data exclusions	No data were excluded from analyses
Replication	Each cryo-EM grid square was collected on once because the holes can only be collected on once. This is standard in the field. Each FRET experiment was performed as five biological replicates, and each native PAGE gel was repeated at least once.
Randomization	Randomization is not applicable to this study.
Blinding	Blinding is not relevant for this study because the experimental conditions must be known to interpret the results.

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

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A