

# Molecular basis for the initiation of DNA primer synthesis

<https://doi.org/10.1038/s41586-022-04695-0>

Received: 10 August 2021

Accepted: 28 March 2022

Published online: 4 May 2022

 Check for updates

Arthur W. H. Li<sup>1,5</sup>, Katerina Zabradý<sup>1,5</sup>, Lewis J. Bainbridge<sup>1</sup>, Matej Zabradý<sup>1</sup>, Sehr Naseem-Khan<sup>2</sup>, Madison B. Berger<sup>2,3</sup>, Peter Kolesar<sup>1,4</sup>, G. Andrés Cisneros<sup>2,3</sup> & Aidan J. Doherty<sup>1</sup>✉

During the initiation of DNA replication, oligonucleotide primers are synthesized de novo by primases and are subsequently extended by replicative polymerases to complete genome duplication. The primase-polymerase (Prim-Pol) superfamily is a diverse grouping of primases, which includes replicative primases and CRISPR-associated primase-polymerases (CAPPs) involved in adaptive immunity<sup>1–3</sup>. Although much is known about the activities of these enzymes, the precise mechanism used by primases to initiate primer synthesis has not been elucidated. Here we identify the molecular bases for the initiation of primer synthesis by CAPP and show that this mechanism is also conserved in replicative primases. The crystal structure of a primer initiation complex reveals how the incoming nucleotides are positioned within the active site, adjacent to metal cofactors and paired to the templating single-stranded DNA strand, before synthesis of the first phosphodiester bond. Furthermore, the structure of a Prim-Pol complex with double-stranded DNA shows how the enzyme subsequently extends primers in a processive polymerase mode. The structural and mechanistic studies presented here establish how Prim-Pol proteins instigate primer synthesis, revealing the requisite molecular determinants for primer synthesis within the catalytic domain. This work also establishes that the catalytic domain of Prim-Pol enzymes, including replicative primases, is sufficient to catalyse primer formation.

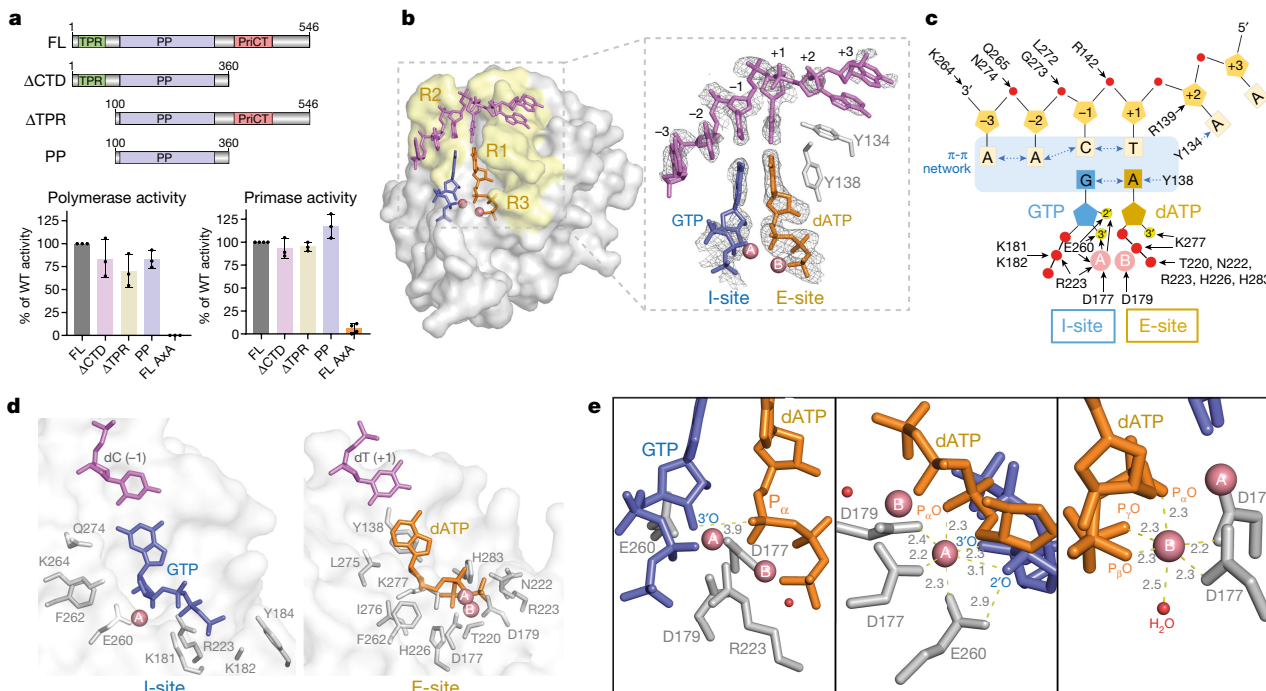
DNA replication is initiated by primases that synthesize short RNA/DNA primers, which are subsequently extended by processive polymerases. In bacteria, primer synthesis is undertaken by DnaG primases belonging to the TOPRIM family<sup>4,5</sup>. In eukaryotes, archaea and some viruses, replicative priming is performed by specialized primases from the Prim-Pol superfamily<sup>1</sup>, formerly known as archaeo-eukaryotic primases<sup>1,2</sup>. Prim-Pol proteins also undertake more diverse roles, including DNA repair, damage tolerance and adaptive immunity<sup>1,3,6–8</sup>. Although primases and polymerases have a common metal-dependent mechanism<sup>9</sup>, little is known about the de novo initiation step of primer synthesis<sup>10,11</sup>. The catalytic domain of Prim-Pol proteins (PP) is proficient in polymerase and translesion synthesis activities<sup>6,7,9,12–15</sup>. However, Prim-Pol enzymes reportedly require additional modules, in conjunction with their PP domain, to facilitate dinucleotide formation, the first step of primer synthesis<sup>16–22</sup>. Given the fundamental importance of priming for genome duplication, it is critical to understand the mechanism of primer synthesis.

## The CAPP PP domain produces primers

To determine the molecular requisites for primer synthesis, we studied a CAPP from *Marinitoga piezophila* (Mp), a Prim-Pol protein

implicated in CRISPR–Cas spacer acquisition<sup>3</sup>. MpCAPP possesses both DNA primase and polymerase activities and consists of a tetratricopeptide repeat (TPR), a PP domain and a predicted PriCT motif within the C-terminal domain (CTD) (Fig. 1a, top) containing a putative iron–sulfur cluster-binding motif, potentially required for primer synthesis (Extended Data Fig. 1a–d). Although mutating this motif ablated iron binding (Extended Data Fig. 1b–d), synthesis activities were unaffected (Extended Data Fig. 1e, f, lanes 6–9). To evaluate which domains are required for DNA synthesis, truncations ( $\Delta$ TPR, PP and  $\Delta$ CTD) were assessed for primer extension activities (Fig. 1a, bottom left, and Extended Data Fig. 1g, lanes 6–17). The PP and  $\Delta$ CTD truncations retained efficient polymerase activity, indicating that only PP is essential for primer extension. Given that CAPP's CTD has similarities with Pri2/L and PriCTs, proposed to be involved in primer synthesis<sup>3</sup>, we tested the truncations for primase activity. The  $\Delta$ CTD and PP truncations both exhibited robust primer synthesis (Fig. 1a, bottom right, and Extended Data Fig. 1h, lane 6–9 and 14–17), establishing that CAPP's catalytic domain is sufficient for primer initiation and extension. This was unexpected given that Prim-Pol enzymes reportedly require additional modules to facilitate primer synthesis<sup>10,11,16–20</sup>.

<sup>1</sup>Genome Damage and Stability Centre, School of Life Sciences, University of Sussex, Brighton, UK. <sup>2</sup>Department of Chemistry, University of North Texas, Denton, TX, USA. <sup>3</sup>Present address: Department of Physics and Department of Chemistry and Biochemistry, University of Texas at Dallas, Richardson, TX, USA. <sup>4</sup>Present address: National Centre for Biomolecular Research, Masaryk University, Brno, Czech Republic. <sup>5</sup>These authors contributed equally: Arthur W. H. Li, Katerina Zabradý. ✉e-mail: [ajd21@sussex.ac.uk](mailto:ajd21@sussex.ac.uk)



**Fig. 1 | MpCAPP's PP domain is polymerase and primase proficient.** **a**, Top, schematic representations of MpCAPP fragments with the highlighted domains. Bottom left, quantification of the polymerase activity of full-length wild-type MpCAPP (FL), its fragments or the full-length D177A/D179A mutant (FL AXA) (125 nM protein). Bottom right, quantification of the primase activity of MpCAPP FL, its fragments or the FL AXA mutant (2 μM protein). Data represent the mean ± s.d. from three independent experiments, except for FL and FL AXA in the primase assay, where four independent experiments were performed. Representative gels are shown in Extended Data Fig. 1g, h. Black dots represent individual values. WT, wild type. **b**, Left, overall structure of a primer initiation complex of MsCAPP (grey) bound to ssDNA (pink), GTP (blue), dATP (orange) and Co(II) (spheres) highlighting DNA-interacting regions R1–R3 (yellow). Right, MsCAPP active site showing a simulated annealing  $F_o - F_c$  omit map of the ssDNA template, I-site GTP and E-site dATP (contoured at  $2\sigma$  at a

resolution of 1.90 Å), along with the side chains of Y134, Y138 and Co(II). **c**, Schematic showing the protein–DNA and protein–nucleotide interactions in the primer initiation complex. Pentagon, sugar; square, base; red sphere, phosphate; yellow sphere, OH group; peach sphere, metal ion. Black arrows indicates an interaction between an amino acid and DNA, nucleotides or metal ions; blue dashed arrows indicate  $\pi$ – $\pi$  stacking. **d**, Residues interacting with nucleotides bound in the I-site (left) and E-site (right). **e**, Left, active site with the oxygen of the 3'-OH of GTP about to initiate a nucleophilic attack on the  $\alpha$ -phosphate of dATP. Middle, A-site Co(II), showing octahedral coordination to nucleotides and surrounding acidic residues with distances labelled in angstroms. E260 interacts with oxygen from the 2'-OH of GTP. Right, B-site Co(II), showing octahedral coordination to dATP, the DXD motif and a water molecule, with distances labelled in angstroms.

### Structure of a primer initiation complex

To understand the active site architecture of CAPPs, we elucidated the crystal structures for the PP domain of *Marinitoga* sp. 1137 CAPP, amino acids 111–328 (MsPP<sub>111–328</sub>), in its apo form (resolution of 2.97 Å) and in complex with dGTP and manganese (Mn(II)) ions (resolution of 1.28 Å) (Extended Data Fig. 2a and Extended Data Table 1). MsPP<sub>111–328</sub> is >98% identical to *M. piezophila* PP<sub>111–328</sub> (Extended Data Fig. 2b) and also possesses primase activity (Extended Data Fig. 2c, lanes 6–9). The MsPP domain is composed of an  $\alpha/\beta$  domain (residues 111–164 and 274–278) and an RNA recognition motif (RRM) domain (residues 169–262 and 291–328)<sup>2</sup> (Extended Data Fig. 3a, b). This structure has substantial homology with those for other Prim-Pol enzymes, including PriI (ref. 23) and PrimPol<sup>14</sup> (Extended Data Fig. 3c). In the complex with dGTP, nucleotide and Mn(II) ions were bound in an active site cleft, proposed to be the elongation (E) site (Extended Data Figs. 2a and 3a). D177 and D179 (motif I) and E260 (motif III) interacted with the two divalent metals. H226 (motif II) was hydrogen-bonded to the triphosphate tail of dGTP.

Prim-Pol proteins initiate primer synthesis by catalysing dinucleotide bond formation. To understand the molecular basis of this de novo synthesis step, we elucidated the structure of a ternary primer initiation complex performing dinucleotide synthesis at a resolution of 1.9 Å (Fig. 1b and Extended Data Table 1). The complex consists of the MsPP domain bound to GTP in the initiating site (I-site), a

non-hydrolysable dATP analogue (dAMPNPP) in the E-site, two cobalt ions<sup>3</sup> and a single-stranded DNA (ssDNA) template (9-mer) (Fig. 1b–d). The incoming bases (G and dA) form Watson–Crick pairings with their respective templating bases, with the 3'-OH nucleophilic group of the initiating base (GTP) positioned within attacking distance (3.9 Å) of the  $\alpha$ -phosphate of the elongating nucleotide (dATP) (Fig. 1e, left), in readiness for dinucleotide bond formation using a two-metal-dependent mechanism<sup>24</sup>. The Co(II) ions are bound with octahedral symmetry in the A- and B-sites (Fig. 1e, middle and right; Extended Data Fig. 3d shows an omit map at  $5\sigma$ ). Metal A is coordinated by D177 and D179 (DXD, motif I), E260 (motif III), the  $\alpha$ -phosphate group of dATP (E-site) and the 2'- and 3'-OH groups of GTP (I-site). The 2'-OH of the initiating nucleotide is weakly coordinated to metal A (3.1 Å) and stabilized by E260, suggesting why a ribonucleotide is requisite for primer initiation<sup>3</sup>. Metal B coordinates to the DXD sequence (motif I), phosphate groups (E-site dATP) and a water molecule, in the active site. This is similar to the coordination of Mn(II) in the dGTP complex, with hydroxyl groups from the cryoprotectant ethylene glycol replacing the hydroxyl groups of the ribose group from GTP (Extended Data Fig. 3e). Binding of PP to the template strand (–3 to +1) predominantly involves residues from region 1 (R1, hairpin 130–142) and region 2 (R2, loop 263–274) (Fig. 1b, c), which have moved to accommodate the ssDNA template (Extended Data Fig. 3f). Y134 pivots on its C $\beta$  atom to form a  $\pi$ – $\pi$  stacking interaction with the adenine base at the +2 position, wedging the template open, while R139 stabilizes the oxygen of the deoxyribose ring of the +2

templating nucleotide. There are no interactions with the nucleotide at position +3 and those further downstream (template strand).

PP interacts with the E-site nucleotide (dATP) via basic residues (R223, H226, K277 and H283), along with polar residues (T220 and N222), which directly contact the phosphate groups (Fig. 1c, d). E-site nucleotide binding also has a stabilizing effect on region 3 (R3, loop 283–286), which is unresolved in the apo structure (Extended Data Fig. 3f). The backbone amine group of K277 interacts with the 3'-OH group of dATP's ribose, while F262, L275 and I276 form a hydrophobic pocket for the base and sugar (Extended Data Fig. 3g). The 2' position of the dATP ribose ring fits snugly against hydrophobic residues L275 and Y138, which cannot accommodate a 2'-OH group (NTPs), explaining the preferred specificity for dNTPs in the E-site<sup>3</sup>. The adenine base of dATP bound in the E-site forms  $\pi$ - $\pi$  stacking interactions with Y138 and the guanine base of GTP at the I-site.

A notable feature of the initiation complex is how remarkably few contacts are made between the enzyme and GTP (I-site) (Fig. 1c, d). Instead, this interaction relies mainly on  $\pi$ - $\pi$  stacking interactions between the guanine base and neighbouring purine bases, provided by dATP (E-site) and an adenine base (-2) on the template strand. Together with Y138, these bases form an extended  $\pi$ - $\pi$  stacking network that stabilizes GTP binding within the I-site and also PP's binding to the template strand (Fig. 1c). These interactions are reminiscent of stabilizing contacts made within the primer initiation complex of RNA polymerases<sup>25–27</sup>. K181, K182 and R223 form a positively charged pocket that, along with metals A and B, binds to the triphosphate tail of the initiating GTP. Notably, the triphosphate tail in an alternative initiation complex adopts a different orientation and binds an additional metal ion (Extended Data Fig. 3h).

Molecular modelling studies investigating intermolecular interactions within the active site showed strong electrostatic stabilization between dATP and R223, with smaller stabilizing contributions from dispersion and induction (Extended Data Fig. 4). Dispersion and induction interactions between Y138 and GTP and between GTP and dATP were observed to be stabilizing for these pairs, consistent with an extended  $\pi$ - $\pi$  stacking network (Extended Data Fig. 4a, c). However, when only pairwise interactions were considered, an overall destabilizing interaction between GTP and dATP was observed (Extended Data Fig. 4c). Although calculation of pairwise interactions indicates a repulsive interaction between the triphosphates of the I-site and E-site nucleotides, interaction of the metal dications, Y138, dT (+1) and dATP, together with GTP, suggests that the intermolecular interaction of these fragments with GTP is stabilizing, with a major overall contribution from induction and dispersion, as well as electrostatic interactions (Extended Data Fig. 4c).

## Structure of a PP domain bound to double-stranded DNA

To compare the primer initiation intermediate with a primer–template complex, we elucidated the structure of CAPP's PP domain bound to double-stranded DNA (dsDNA, 6-mer) and Co(II) ions at a resolution of 2.02 Å (Fig. 2a and Extended Data Table 1). The base pairings at positions -5, -4, -1 and 1 of the blunt-ended dsDNA are formed from standard Watson–Crick hydrogen bonds, while the bases at -3 and -2 form purine–pyrimidine (G–T) mismatched pairings, which induce a slight kinking of the B-form helix. The PP domain showed relatively minor overall differences from the other complexes (Extended Data Fig. 5). Binding of PP to the template strand (-3 to +1) is remarkably similar to the initiation complex, which is held in position by residues from R1 and R2 (Fig. 2b, c). A notable feature was the limited contacts made between PP and the primer strand (Fig. 2b, right). H226 (motif II) and H283 (R3) are in close proximity to the primer strand and, along with K277, position the hydroxyl group of the 3' nucleotide (+1) in place. Both metal ions interact with the phosphodiester bond at +1 and -1,

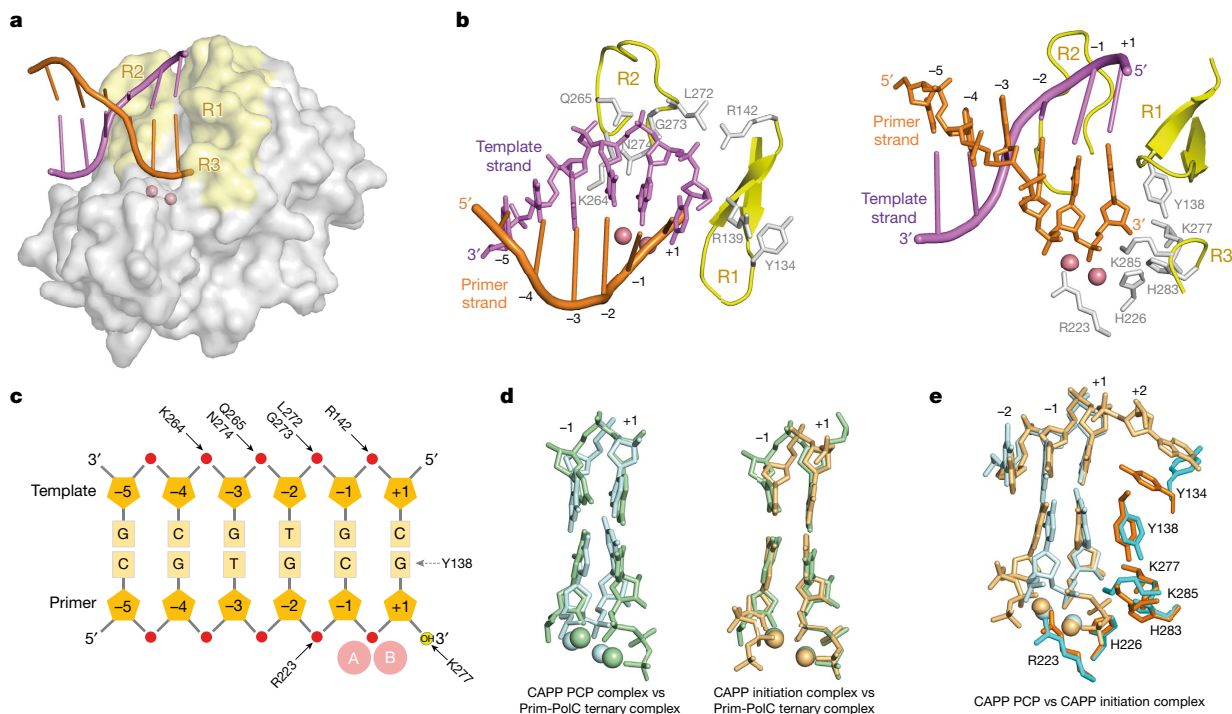
and R223 contacts the phosphodiester bond at -1 and -2 of the primer strand. Y138 (R1) stacks against the base at +1 (primer strand). R223, which sits between the phosphate tails of the initiating and elongating nucleotides in the primer initiation complex, has moved further towards the initiation site. It now holds the phosphate of the phosphodiester linkage between positions -2 and -1 in place, suggesting a role in translocating the priming strand during extension. The 3' nucleotides (primer strand), template nucleotides and metal ions superpose onto a Prim-PolC ternary complex<sup>9</sup> with root-mean-square deviation (r.m.s.d.) of about 2.0 Å (Fig. 2d, left), indicating that this represents a postcatalytic product complex. Comparison of the primer initiation complex with either the Prim-PolC ternary complex (Fig. 2d, right) or the CAPP postcatalytic product complex (Fig. 2e) shows similar active site configurations, indicating that a shared two-metal-dependent mechanism catalyses both dinucleotide synthesis and primer extension.

## Structure–function analyses of the PP domain

To examine the roles of specific residues in primer synthesis and/or extension, we mutated residues interacting with incoming nucleotides, metal ions or DNA (Fig. 3a) and evaluated the effect on MpCAPP's PP synthesis activities (Fig. 3b and Extended Data Fig. 6a, b). Mutation of the metal-binding residues of motif I (AXA, D177A/D179A) ablated synthesis activities. Mutation of R223, which interacts with the dinucleotide/primer, resulted in almost complete loss of primase and polymerase synthesis activities; although the mutant protein was still capable of some dinucleotide synthesis, it only extended primers by 1 or 2 nucleotides, intimating a role in primer translocation. Mutation of H226 (motif II), which interacts with the incoming dNTP (+1), resulted in a reduction in polymerase activity, similar to that seen with R223A, but the mutant protein was significantly more deficient in dinucleotide synthesis. Mutation of other contacts with the incoming nucleotides (Y138A, E260A, F262A, K277A and K181A/K182A (KK)) also reduced primase and polymerase activities. Although the polymerase activity of the Y138A mutant was only modestly compromised, this mutant's ability to synthesize dinucleotides was significantly diminished. This indicates that  $\pi$ - $\pi$  stacking between Y138 and the incoming dNTP (E-site) is critical for the dinucleotide formation process but less essential for primer elongation. E260 makes contacts with metal A as well as with the 2'-OH and 3'-OH of the initiating GTP base (-1). The E260A mutant was barely active in primer extension assays, compared with the R223A or H226A mutant, and its inability to synthesize dinucleotides was comparable to that of the AXA mutant. This intimates that E260 has crucial roles in both priming and extension, owing to its binding to both metal A and the 2'-OH of the incoming NTP (-1) during the initiation of dinucleotide synthesis and to the 3'-OH of the deoxynucleotide (-1) of the primer during extension. Interaction of E260 with the 2'-OH of the incoming NTP explains why this residue is crucial for the first step of priming, as it appears to stabilize the NTP in a catalytically competent orientation for dinucleotide bond formation. Mutation of R2 residues (K264A, Q265A or N274A), which interact with the template strand, slightly reduced primer extension activity, but the resulting mutants exhibited strongly reduced primase activity (Fig. 4b). Although a triple R2 mutant (KQN) could perform limited primer extension, its priming activity was severely compromised, establishing the importance of R2 for both priming and polymerase activities.

## Nucleotides affect PP binding to ssDNA

To investigate the mechanism of nucleotide and template binding, we determined the affinities of McCAPP's PP domain for GTP, dATP and ssDNA template (8-mer) using fluorescence polarization (FP) to measure dissociation constants ( $K_d$ ). PP's binding affinity was relatively weak for GTP ( $K_d = 23.70 \mu\text{M}$ ) but was much stronger for dATP ( $K_d = 1.40 \mu\text{M}$ ) (Fig. 3c). Although PP alone bound weakly to ssDNA,



**Fig. 2 | Structure of a post-ternary complex of PP bound to dsDNA. a**, A surface representation of the MsCAPP PP domain (grey) bound to dsDNA (pink and orange) forming a post-ternary complex, highlighting important DNA-interacting regions R1–R3 (yellow). **b**, Molecular models showing protein–DNA interactions with the template strand (left) and the primer strand (right). **c**, Schematic showing the important protein–DNA and protein–nucleotide interactions formed in the PP–dsDNA complex. Pentagon, sugar

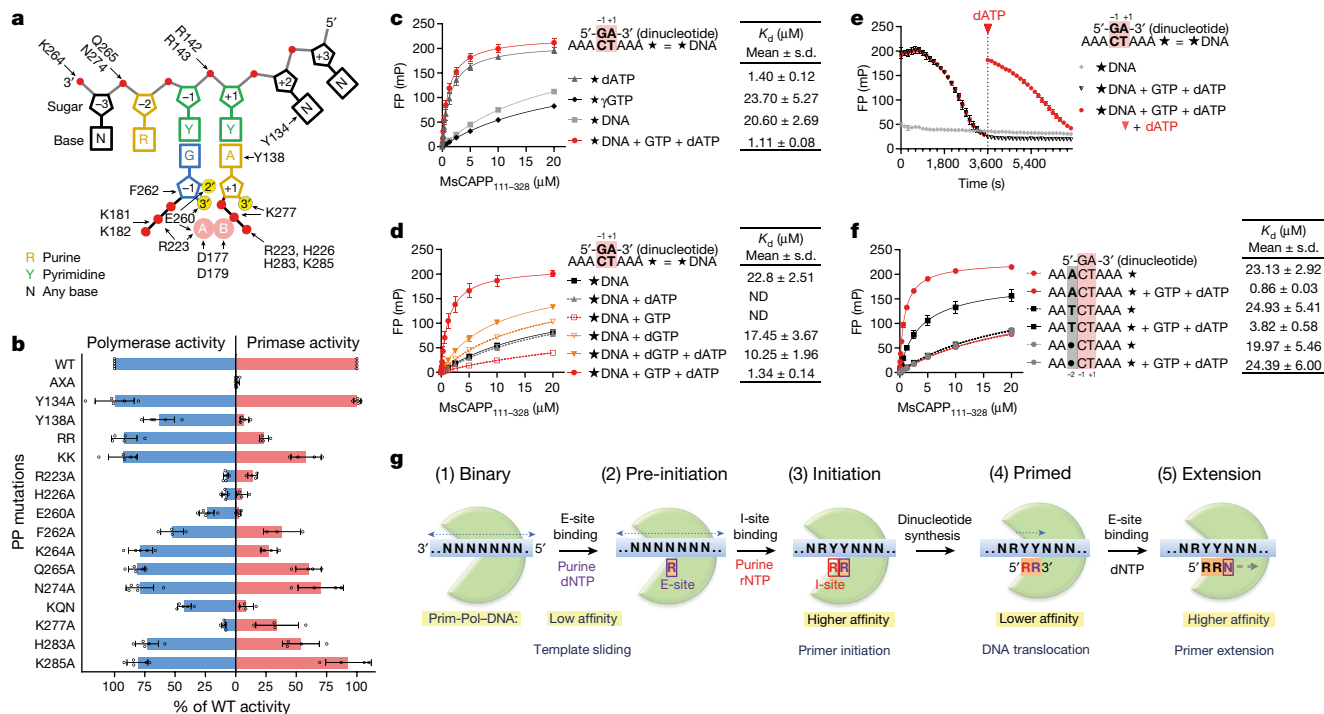
(deoxyribose); square, base; red sphere, phosphate; yellow sphere, OH group; peach sphere, metal ion. Black arrows indicate interactions between indicated amino acids and DNA. **d**, Superposition of active site bases and metal ions of the CAPP–dsDNA (cyan) and Prim-PoIC–DNA (PDB 6SA0) (green) complexes (left) and the CAPP primer initiation (orange) and Prim-PoIC–DNA (green) complexes (right). **e**, Superposition of the active sites of the CAPP–dsDNA (cyan) and CAPP primer initiation (orange) complexes.

addition of both GTP (1 mM) and dATP (0.1 mM), in the presence of metal ions, strongly enhanced its affinity for DNA ( $K_d$  of about 1  $\mu$ M). Modelling studies were also consistent with these findings (Extended Data Fig. 4). Together, these results underscore the importance of the divalent cations in stabilization of the negative charges of the two triphosphate moieties in close proximity. This increased affinity is not due to PP binding to newly synthesized dinucleotide, as addition of pppGpdA (riboG–deoxyA; rG–dA) dinucleotide did not increase the affinity for template (Extended Data Fig. 6c). Addition of GTP or dATP alone also did not stimulate template binding (Fig. 3d). Although addition of dGTP (1 mM) with dATP (0.1 mM) stimulated DNA binding by PP ( $K_d = 10.25 \mu$ M), this was about 10-fold lower than was observed in the presence of both GTP and dATP, which supports dinucleotide synthesis (Fig. 3d). This suggests that the 2'-OH group (GTP) stabilizes the first base (I-site) next to the dNTP (E-site) on the template strand. Together, these data indicate that the strongest affinity of PP for template occurs when both nucleotides and template are bound within PP's active site before turnover. This conclusion was further supported by assays that showed that PP binding to template, under conditions that allow turnover, decreased over time to levels observed in the absence of GTP and dATP (Fig. 3e). However, addition of dATP (0.1 mM at 60 min) increased the affinity of PP to a level similar to that observed at time 0, indicating that the loss of binding affinity was due to lack of dATP, which is required for dinucleotide synthesis and primer extension.

Next, we tested how template sequence influences PP's affinity for DNA. Binding of PP to a poly(dA) template containing a single dC or dT was not stimulated by the presence of both GTP and dATP, in contrast to a template containing 3'-dCdT-5' ( $K_d$  of about 1  $\mu$ M), indicating that nucleotide-stimulated affinity of PP for ssDNA is sequence dependent (Extended Data Fig. 6d). As the primer initiation structure showed  $\pi$ - $\pi$  stacking between the template dA (-2) and GTP in the I-site, we

examined whether the pre-initiation template base (-2) influences PP binding in the presence of incoming nucleotides. A template purine base at -2 should stabilize GTP (purine) more effectively than a pyrimidine base (Fig. 3f), which has reduced stacking potential<sup>25</sup>. When the template dA (-2) was substituted with dT, PP's DNA binding affinity was about 4-fold lower ( $K_d = 3.82 \mu$ M) than for a template containing dA at this position ( $K_d$  of about 1  $\mu$ M). PP binding to a template containing an abasic site at -2 was not stimulated by addition of GTP and dATP ( $K_d$  of about 20  $\mu$ M). Similar results were obtained when templates containing four different bases or an abasic site at the -2 position were tested in primer assays in the presence of GTP and dATP. Dinucleotide synthesis was highest when a purine was at -2 (template) (Extended Data Fig. 6e). Modelling calculations (SAPT0) showed that the interaction between a dA base at position -2 and GTP exhibits an important induction and dispersion contribution (Extended Data Fig. 4c). Together, these results establish that the pre-initiation template base (-2) also has an important role in influencing PP's affinity for the primer initiation nucleotide and probably explains the influence this 'cryptic' site exerts on primer synthesis.

MsPP exhibited greater affinity for DNA substrates with increasing length (Extended Data Fig. 6f), suggesting that low-affinity template sliding probably occurs before binding of incoming nucleotides<sup>28</sup> (Fig. 3g). The relatively high affinity of MpCAPP for dATP, similar to PP's affinity for template in the presence of both nucleotides, suggests that a dNTP (preferably a purine) is bound first in the E-site, possibly as the enzyme slides on ssDNA. PP's affinity for template increases about 20-fold in the presence of both nucleotides, suggesting that a purine ribonucleotide binds into the I-site as the last component, 'locking' the enzyme onto the primer initiation site. Following turnover, the enzyme's affinity for DNA decreases, enabling primer translocation to occur and the next round of nucleotide binding and addition to proceed (Fig. 3g).



**Fig. 3 | Structure–function analyses of MsCAPP PP primer synthesis and extension activities.** **a**, Schematic representation of the interactions of MsCAPP with template and nucleotides in the primer initiation complex. Red sphere, phosphate; yellow sphere, OH group; peach sphere, metal ion. Black arrows indicate interactions. **b**, Effects of MsCAPP PP domain mutations on polymerase (left) and primase (right) activities. AXA, D177A/D179A; RR, R142A/R143A; KK, K181A/K182A; KQN, K264A/Q265A/N274A. Data represent the mean  $\pm$  s.d. from five (polymerase) and four (primase) independent experiments. Representative gels are shown in Extended Data Fig. 6a, b. Black circles indicate individual values. **c**, Affinity of MsCAPP PP for template, dATP and GTP. FP assays contained 0–20  $\mu$ M protein, 25 nM FAM–dATP (★dATP) or FAM– $\gamma$ GTP (★ $\gamma$ GTP) or 50 nM FAM–DNA (★DNA, oKZ409)  $\pm$  1 mM GTP and 0.1 mM dATP. **d**, Presence of GTP and dATP stimulates affinity of MsCAPP PP for template. FP assays contained 0–20  $\mu$ M protein, 50 nM FAM–DNA

(oKZ409)  $\pm$  0.1 mM dATP and/or 1 mM GTP/dGTP. **e**, Affinity of MsCAPP PP for template decreased over time in the presence of GTP and dATP. FP assays contained 10  $\mu$ M protein, 50 nM FAM–DNA (oKZ409)  $\pm$  1 mM GTP and 0.1 mM dATP. FP was measured every 5 min for 2 h; 0.1 mM dATP was added at 60 min (dashed line indicated by a red arrowhead). **f**, The MsCAPP PP domain prefers to bind template containing a purine at the –2 position. FP assays contained 0–20  $\mu$ M protein, 50 nM FAM–DNA (oKZ409, oKZ417 or oKZ418)  $\pm$  1 mM GTP and 0.1 mM dATP. A star indicates FAM labelling. Data in **c–f** represent the mean  $\pm$  s.d. from four independent experiments. ND, not determined. **g**, A model of Prim-Pol proteins' primer initiation and extension mechanism. Green crescent, Prim-Pol; blue rectangle, ssDNA template; peach rectangle, newly primed strand; red R, purine ribonucleotide. Primed strand deoxyribonucleotides are shown in purple.

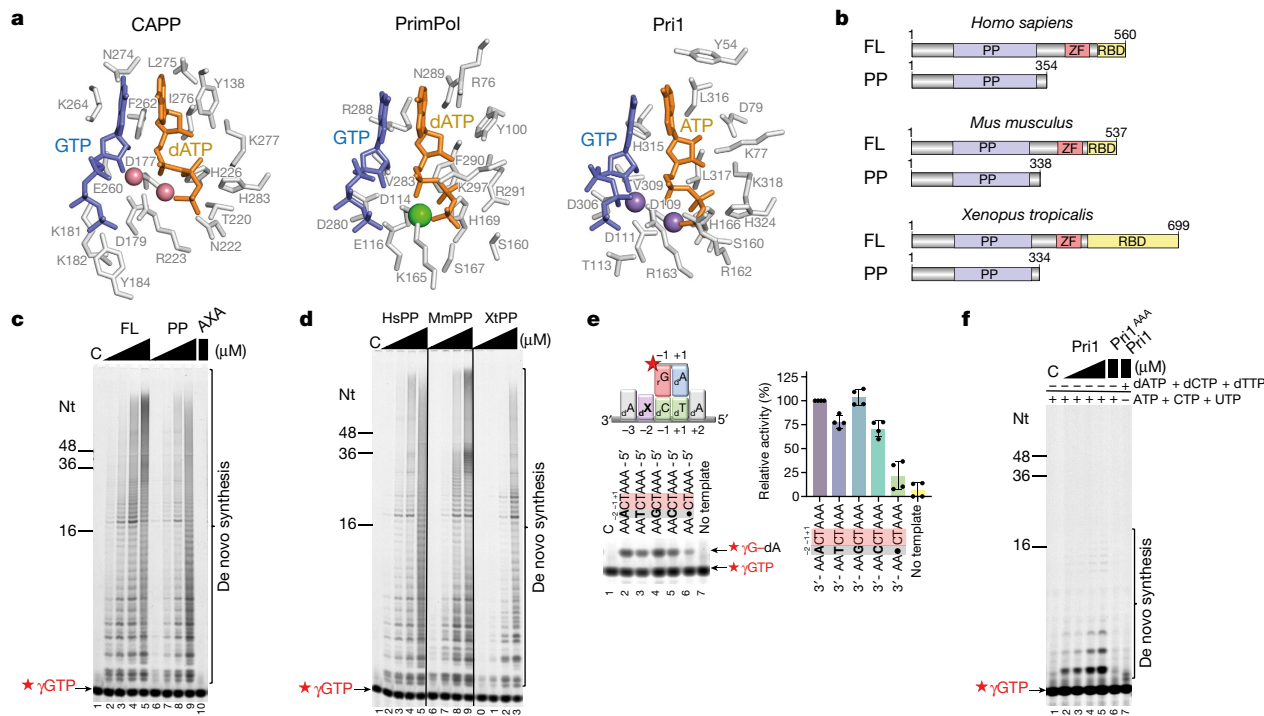
## PP domains share homology

Docking the primer initiation complex, using the bound nucleotides (E-site) and metal ions from structures of human PriI (ref. 23) and Prim-Pol<sup>14</sup>, enabled identification of putative binding sites for the I-site and ssDNA template within these replicative primases. The overall structures and active site residues surrounding the I- and E-sites are remarkably similar to those of MsCAPP (Fig. 4a and Extended Data Fig. 7a). In MsCAPP and PriI, motifs I (DXD) and II (G/SXH) are conserved in both structures, with H226 (CAPP) and H166 (PriI) both interacting with the phosphate tail of the E-site nucleotide. Motif III (hd/Eh, in which 'h' is a hydrophobic residue) is also present in PriI, although E260 (CAPP) is replaced by D306 (PriI). H283 (CAPP) and H324 (PriI) also interact with the phosphate tail of the E-site nucleotide. R223 (CAPP) also has a counterpart in PriI (R163). L275 and K277 (CAPP) have counterparts L316 and K318 (PriI). Both proteins have a tyrosine in R1, of which Y138 (CAPP) stacks with the base in the E-site and Y54 (PriI) is close to the active site, although it does not interact with the base. PriI uses ribonucleotides for extension, although it can bind to dNTPs<sup>23</sup>. By contrast, CAPP's E-site binds to dNTPs<sup>3</sup>. When examining residues surrounding the 2' position of the E-site nucleotide, Y138 and L275 form a more hydrophobic environment in CAPP, whereas D79 forms a hydrogen bond with the 2'-OH in PriI (Extended Data Fig. 7b).

The CAPP and PrimPol active sites are also highly similar (Fig. 4a). Motif I (DXD in CAPP; DXE in PrimPol), motif II (G/SXH) and motif III (hEh in CAPP and hDh in PrimPol) are all present in both structures (Fig. 4a). Many key basic residues are also conserved, including R223, K277 and H283 (CAPP) and K165, R291 and K297 (PrimPol). R76 in R1 of PrimPol fulfils a role similar to that of Y138 in R1 of CAPP. Y100 (PrimPol) is also involved in selecting for dNTP binding in the E-site<sup>29</sup>. Together, these structural similarities intimate that the mechanism of primer initiation is probably conserved between these related DNA Prim-Pol proteins.

## Eukaryotic PP domains can prime

Given CAPPs' overt structural similarities with replicative Prim-Pol proteins, we investigated whether eukaryotic PP domains could also catalyse primer synthesis independently. PrimPol consists of a PP domain and a C-terminal zinc finger, previously considered critical for priming (Fig. 4b)<sup>13,16</sup>. As with CAPP, we observed that the PP domain (HsPrimPol<sub>1–354</sub>) of human PrimPol alone was sufficient for primer synthesis (Fig. 4c, lanes 6–9, and Extended Data Fig. 8a), although the activity was decreased in comparison with full-length protein, suggesting some stimulatory/stabilization role for the zinc-finger domain in priming<sup>13,16</sup>. All catalytic activities were ablated in a catalytic mutant (D114A/E116A) (Fig. 4c, lane 10). The equivalent PP domains of mouse (MmPrimPol<sub>1–338</sub>) and *Xenopus tropicalis* (XtPrimPol<sub>1–334</sub>) PrimPol



**Fig. 4 | The PP domains of eukaryotic replicative primases are primase proficient.** **a**, I- and E-sites of CAPP (left), human PrimPol (middle) and human Pri1 (right). GTP (CAPP primer initiation complex) was superposed onto human Prim-Pol structures to identify their I-sites. **b**, Schematic representations of human, mouse and *X. tropicalis* full-length Prim-Pol proteins and their PP domains. ZF, zinc finger; RBD, RPA-binding domain. **c**, The human PrimPol PP is priming proficient. Primase assays contained 0.125, 0.25, 0.5 or 1  $\mu$ M full-length HsPrimPol (lanes 2–5), 0.125, 0.25, 0.5 or 1  $\mu$ M HsPrimPol<sub>1–354</sub> (PP; lanes 6–9) or 1  $\mu$ M HsPrimPol<sub>1–354</sub> D114A/E116A (AXA; lane 10), 1  $\mu$ M ssDNA (oKZ388), 2.5  $\mu$ M FAM- $\gamma$ GTP ( $\star\gamma$ GTP) and 100  $\mu$ M dNTP. **d**, Mouse and *X. tropicalis* PPs are priming proficient. Assays contained 0.125, 0.25, 0.5 or 1  $\mu$ M HsPrimPol<sub>1–354</sub> (HsPP; lanes 2–5), MmPrimPol<sub>1–338</sub> (MmPP; lanes 6–9) or XtPrimPol<sub>1–334</sub> (XtPP; lanes 10–13) under the conditions described in **c**. **e**, The

HsPrimPol PP prefers a purine base at the –2 position (template) for dinucleotide synthesis. Reactions contained 0.25  $\mu$ M HsPP, 1  $\mu$ M ssDNA template (lane 2, oKZ435; lane 3, oKZ447; lane 4, oKZ449; lane 5, oKZ450; lane 6, oKZ448), 2.5  $\mu$ M FAM- $\gamma$ GTP and 100  $\mu$ M dNTP. Left, representative gel. Right, quantification. Dinucleotide signal was normalized to that in the presence of 3'-AACTAAA-5' (100%). Data represent the mean  $\pm$  s.d. Black dots indicate individual values. **f**, Human Pri1 exhibits priming activity. Reactions contained 1, 2, 4 or 8  $\mu$ M wild-type Pri1 (Pri1; lanes 2–5, 7) or 8  $\mu$ M Pri1 D109A/D111A/D306A (Pri1<sup>AAA</sup>; lane 6), 1  $\mu$ M ssDNA (oKZ388), 10  $\mu$ M FAM- $\gamma$ GTP, and 100  $\mu$ M ATP, CTP and UTP (lanes 1–6) or 100  $\mu$ M dATP, dCTP and dTTP (lane 7). Results are representative of three (**d**, **f**) or four (**c**, **e**) independent experiments. Nt, oligonucleotide length marker; 'C' indicates control without protein.

proteins are also primase proficient, despite lacking their auxiliary domains (Fig. 4b, d, lanes 6–9 and 10–13, respectively). Decreasing concentrations of labelled GTP in the reaction caused loss of detectable PrimPol priming activity (Extended Data Fig. 8b), in agreement with the FP studies on MpCAPP, indicating that the I-site nucleotide can be readily outcompeted by unlabelled dNTPs, if present at substoichiometric concentrations (Extended Data Fig. 8c, d). Therefore, discrepancies with previous studies are probably due to more physiologically relevant concentrations of labelled initiating nucleotide being used in the current study, allowing priming to be more readily observed.

Mutation of corresponding residues in HsPrimPol<sub>1–354</sub>, shown to be important for CAPP's primase and/or polymerase activity, also had a significant negative effect on its synthesis activities (Extended Data Fig. 8e–g). While HsPrimPol<sub>1–354</sub> could produce primers using only dNTPs, primer synthesis was stimulated by the addition of purine NTPs, particularly GTP (Extended Data Fig. 8h), which was incorporated as the initiating primer nucleotide (Extended Data Fig. 8i). Similarly to CAPP, HsPrimPol<sub>1–354</sub> also primes most efficiently when a purine base is located at the –2 position, suggesting that a similar  $\pi$ - $\pi$  stacking network is also involved in primer initiation by other Prim-Pol proteins (Fig. 4e).

Eukaryotic and archaeal replicative primases (PriI/PriS) reportedly require a second subunit (Pri2/PriL) to initiate primer synthesis<sup>30</sup>. However, when we assayed human Pri1 for primer synthesis activity, de novo primer synthesis was evident (Fig. 4f, lanes 2–5)<sup>31</sup>, although this was less efficient than with Pri2 (ref. <sup>32</sup>). As with CAPP and PrimPol,

Pri1 also prefers to initiate primer synthesis with a 5' GTP over ATP (Extended Data Fig. 8j) but extends with NTPs, rather than dNTPs (Fig. 4f, compare lanes 5 and 7)<sup>30</sup>. Together, these findings establish that the PP domains of replicative primases are sufficient to catalyse de novo primer synthesis, supporting a conserved mechanism across the Prim-Pol superfamily.

## Discussion

Here we present the molecular basis for primer synthesis by a DNA primase, uncovering the mechanism of de novo dinucleotide bond formation that initiates priming. This study provides compelling evidence that all the molecular determinants required to undertake the critical steps of primer synthesis reside within the catalytic domain of Prim-Pol proteins. The first key step of primer synthesis involves sliding of Prim-Pol along the ssDNA, which binds the elongating dNTP into an active site pocket (E-site) (Fig. 3g). The next step involves preferential binding of a purine ribonucleotide in the primer initiation site (I-site). Prim-Pol makes only limited contacts with the I-site nucleotide, and adjacent nucleotides and metal ions are therefore crucial for stabilizing its binding to the I-site. Most of the critical interactions with the initiating nucleotide occur between adjacent bases that form a  $\pi$ - $\pi$  stacking network, which stabilizes initial binding of the incoming nucleotide (I-site), enabling it to form a stable Watson-Crick pairing with the template strand. A common feature of most primases is their preference to

incorporate a purine as the initiating base. In the primer initiation complex, the incoming purine (I-site) stacks against a purine base (–2) from the template strand, which stabilizes its binding and probably explains the preferential binding of purines, over pyrimidines, at the I-site. These  $\pi$ – $\pi$  contacts may also influence nucleotide docking (E-site) as a result of template stabilization. The specificity for ribonucleotide binding (I-site) appears to be conferred by specific interactions between the 2'-OH of the ribose moiety with metal A and its primary liganding residue (E260 in MscAPP). RNA-dependent RNA polymerases undertake dinucleotide synthesis using an analogous, convergent priming mechanism during transcription initiation and replication<sup>25–27</sup>.

This study also demonstrates that the PP domains of replicative primases initiate de novo primer synthesis in the absence of ancillary modules. Prim-Pol enzymes almost certainly evolved from primordial RNA replicases, which were subsequently repurposed to undertake more specialized cellular roles, including primer synthesis. As de novo synthesis is a relatively inefficient step<sup>28</sup>, other modules were probably acquired during primase evolution to stabilize the precarious initiation intermediate, ensuring more efficient primer synthesis and extension<sup>13,22,33,34</sup>. These modules may act to enhance DNA and dinucleotide binding but may also regulate the primer initiation step to prevent unlicensed priming or ensure efficient termination. Notably, Prim-Pol enzymes involved in DNA repair synthesis, which are primase deficient, lack equivalent modules<sup>6,7,9</sup>. Having established the mechanism for initiating primer synthesis within the catalytic core of a Prim-Pol protein, further studies are now required to determine how these catalytic steps are influenced by ancillary modules and how primer termination is achieved.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-022-04695-0>.

- Guilliam, T. A., Keen, B. A., Brissett, N. C. & Doherty, A. J. Primase-polymerases are a functionally diverse superfamily of replication and repair enzymes. *Nucleic Acids Res.* **43**, 6651–6664 (2015).
- Iyer, L. M., Koonin, E. V., Leipe, D. D. & Aravind, L. Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. *Nucleic Acids Res.* **33**, 3875–3896 (2005).
- Zabradý, K., Zabradý, M., Kolesar, P., Li, A. W. H. & Doherty, A. J. CRISPR-associated primase-polymerases are implicated in prokaryotic CRISPR–Cas adaptation. *Nat. Commun.* **12**, 3690 (2021).
- Bouché, J. P., Zechel, K. & Kornberg, A. *dnaG* gene product, a rifampicin-resistant RNA polymerase, initiates the conversion of a single-stranded coliphage DNA to its duplex replicative form. *J. Biol. Chem.* **250**, 5995–6001 (1975).
- Aravind, L., Leipe, D. D. & Koonin, E. V. Toprim—a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. *Nucleic Acids Res.* **26**, 4205–4213 (1998).
- Płociński, P. et al. DNA ligase C and Prim–PolC participate in base excision repair in mycobacteria. *Nat. Commun.* **8**, 1251 (2017).
- Pitcher, R. S., Brissett, N. C. & Doherty, A. J. Nonhomologous end-joining in bacteria: a microbial perspective. *Annu. Rev. Microbiol.* **61**, 259–282 (2007).

- Bainbridge, L. J., Teague, R. & Doherty, A. J. Repriming DNA synthesis: an intrinsic restart pathway that maintains efficient genome replication. *Nucleic Acids Res.* **49**, 4831–4847 (2021).
- Brissett, N. C. et al. Molecular basis for DNA repair synthesis on short gaps by mycobacterial primase–polymerase C. *Nat. Commun.* **11**, 4196 (2020).
- Bell, S. D. Initiating DNA replication: a matter of prime importance. *Biochem. Soc. Trans.* **47**, 351–356 (2019).
- Boudet, J., Devillier, J.-C., Allain, F. H.-T. & Lipps, G. Structures to complement the archaeo-eukaryotic primases catalytic cycle description: what's next? *Comput. Struct. Biotechnol. J.* **13**, 339–351 (2015).
- Bianchi, J. et al. PrimPol bypasses UV photoproducts during eukaryotic chromosomal DNA replication. *Mol. Cell* **52**, 566–573 (2013).
- Keen, B. A., Jozwiakowski, S. K., Bailey, L. J., Bianchi, J. & Doherty, A. J. Molecular dissection of the domain architecture and catalytic activities of human PrimPol. *Nucleic Acids Res.* **42**, 5830–5845 (2014).
- Rechkoblit, O. et al. Structure and mechanism of human PrimPol, a DNA polymerase with primase activity. *Sci. Adv.* **2**, e1601317 (2016).
- Rechkoblit, O. et al. Structural basis of DNA synthesis opposite 8-oxoguanine by human PrimPol primase-polymerase. *Nat. Commun.* **12**, 4020 (2021).
- Martínez-Jiménez, M. I., Calvo, P. A., García-Gómez, S., Guerra-González, S. & Blanco, L. The Zn-finger domain of human PrimPol is required to stabilize the initiating nucleotide during DNA priming. *Nucleic Acids Res.* **46**, 4138–4151 (2018).
- Holzer, S., Yan, J., Kilkenny, M. L., Bell, S. D. & Pellegrini, L. Primer synthesis by a eukaryotic-like archaeal primase is independent of its Fe–S cluster. *Nat. Commun.* **8**, 1718 (2017).
- Liu, B. et al. A primase subunit essential for efficient primer synthesis by an archaeal eukaryotic-type primase. *Nat. Commun.* **6**, 7300 (2015).
- Boudet, J. et al. A small helical bundle prepares primer synthesis by binding two nucleotides that enhance sequence-specific recognition of the DNA template. *Cell* **176**, 154–166 (2019).
- Beck, K., Vannini, A., Cramer, P. & Lipps, G. The archaeo-eukaryotic primase of plasmid pRN1 requires a helix bundle domain for faithful primer synthesis. *Nucleic Acids Res.* **38**, 6707–6718 (2010).
- Baranovskiy, A. G. et al. Crystal structure of the human primase. *J. Biol. Chem.* **290**, 5635–5646 (2015).
- Baranovskiy, A. G. et al. Insight into the human DNA primase interaction with template-primer. *J. Biol. Chem.* **291**, 4793–4802 (2016).
- Holzer, S. et al. Structural basis for inhibition of human primase by arabinofuranosyl nucleoside analogues fludarabine and vidarabine. *ACS Chem. Biol.* **14**, 1904–1912 (2019).
- Steitz, T. A. & Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl Acad. Sci. USA* **90**, 6498–6502 (1993).
- Basu, R. S. et al. Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. *J. Biol. Chem.* **289**, 24549–24559 (2014).
- Butcher, S. J., Grimes, J. M., Makeyev, E. V., Bamford, D. H. & Stuart, D. I. A mechanism for initiating RNA-dependent RNA polymerization. *Nature* **410**, 235–240 (2001).
- Appleby, T. C. et al. Structural basis for RNA replication by the hepatitis C virus polymerase. *Science* **347**, 771–775 (2015).
- Sheaff, R. J. & Kuchta, R. D. Mechanism of calf thymus DNA primase: slow initiation, rapid polymerization, and intelligent termination. *Biochemistry* **32**, 3027–3037 (1993).
- Díaz-Talavera, A. et al. A cancer-associated point mutation disables the steric gate of human PrimPol. *Sci. Rep.* **9**, 1121 (2019).
- Copeland, W. C. & Wang, T. S. Enzymatic characterization of the individual mammalian primase subunits reveals a biphasic mechanism for initiation of DNA replication. *J. Biol. Chem.* **268**, 26179–26189 (1993).
- Lee, J.-G. et al. Structural and biochemical insights into inhibition of human primase by citrate. *Biochem. Biophys. Res. Commun.* **507**, 383–388 (2018).
- Copeland, W. C. Expression, purification, and characterization of the two human primase subunits and truncated complexes from *Escherichia coli*. *Protein Expr. Purif.* **9**, 1–9 (1997).
- Lee, S.-J., Zhu, B., Hamdan, S. M. & Richardson, C. C. Mechanism of sequence-specific template binding by the DNA primase of bacteriophage T7. *Nucleic Acids Res.* **38**, 4372–4383 (2010).
- Baranovskiy, A. G. et al. Mechanism of concerted RNA–DNA primer synthesis by the human primosome. *J. Biol. Chem.* **291**, 10006–10020 (2016).

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

© The Author(s), under exclusive licence to Springer Nature Limited 2022

## Methods

**Cloning, expression and purification of recombinant proteins**

A description of all constructs, their cloning details and a list of all primers used can be found in Supplementary Table 1 and Supplementary Note 1.

MpCAPP FL WT, MpCAPP FL AXA, MpCAPP FL CC, MpCAPP fragments  $\Delta$ CTD,  $\Delta$ TPR and PP WT (and mutants), MsCAPP PP<sub>100–359</sub> and MsCAPP PP<sub>111–328</sub> were expressed from plasmids pKZ43, pKZ60, pKZ125, pKZ38, pKZ121, pKZ39, pPK247 and pAL101, respectively, in the BL21(DE3) *Escherichia coli* strain. The transformed cell cultures were grown in standard TB medium to OD<sub>600</sub> of 0.8–1. Expression of all proteins was induced by adding IPTG to a final concentration of 0.5 mM, followed by incubation for 3 h at 37 °C.

MpCAPP FL WT and mutants (FL AXA and FL CC) and the  $\Delta$ TPR fragment were fused to MBP via their C terminus and purified as described for MpCAPP FL WT previously<sup>3</sup> with one modification: all purification procedures were done under deoxygenated conditions with N<sub>2</sub>-purged solutions and in a glove bag under an N<sub>2</sub> atmosphere. In brief, collected cells were resuspended in buffer A (50 mM HEPES pH 7.5, 500 mM NaCl, 10% (vol/vol) glycerol, 1 mM TCEP, 10 mM imidazole) containing protease inhibitors, sonicated and cleared by ultracentrifugation. The supernatant was incubated with cobalt resin and eluted in buffer A containing 300 mM imidazole. Eluted protein was loaded onto amylose resin and washed in amylose wash buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 10% (vol/vol) glycerol, 1 mM TCEP), and the bound protein was eluted with amylose wash buffer supplemented with 10 mM maltose. Eluted protein was concentrated in a Vivaspin 20 column (Sartorius), frozen in liquid nitrogen and stored at –80 °C.

For purification of MpCAPP fragments  $\Delta$ CTD, PP WT (and mutants) and MsCAPP PP, the cell pellet was resuspended in buffer B (50 mM HEPES pH 7.5, 250 mM NaCl, 10 mM imidazole, 10% (vol/vol) glycerol, 0.5 mM TCEP) containing protease inhibitors, sonicated and ultracentrifuged. The supernatant was incubated with cobalt resin and extensively washed, and the protein was eluted in buffer B containing 300 mM imidazole. Eluted protein was loaded onto a 5-ml HiTrap Q HP column (Cytiva), and the resulting flow was immediately loaded onto a 5-ml HiTrap SP FF column (Cytiva) (before loading of the MpCAPP  $\Delta$ CTD fragment onto Q and SP columns, the salt concentration of the sample was decreased to 150 mM to allow binding to the SP column). The SP column was developed with a 50-ml gradient of 250–600 mM NaCl in buffer B for MpCAPP PP WT and MsCAPP PP and 150–600 mM NaCl for MpCAPP  $\Delta$ CTD. The peak fractions were pooled, concentrated in a Vivaspin 20 column (Sartorius), aliquoted, frozen in liquid nitrogen and stored at –80 °C. All MpCAPP PP mutants were expressed and purified as MpCAPP PP WT.

Full-length HsPrimPol was expressed from pET28a-HsPrimPol and purified as described previously<sup>13</sup>. In brief, the protein was expressed in SHuffle T7 *E. coli* cells overnight at 16 °C. The protein was purified using Ni-NTA affinity resin (Generon) followed by a HiTrap Heparin HP column (Cytiva) and finally size exclusion chromatography on a Superdex 75 column (Cytiva).

HsPrimPol<sub>1–354</sub> and HsPrimPol<sub>1–354</sub> mutants were expressed from plasmids pET28a-HsPrimPol<sub>1–354</sub> and pLB38-46 in BL21(DE3) *E. coli* cells. The transformed cell cultures were grown in standard TB medium to an OD<sub>600</sub> of 3. Expression of the proteins was induced by adding IPTG to a final concentration of 1 mM, followed by incubation for 3 h at 37 °C. Collected cells were resuspended in buffer C (180 mM phosphate citrate pH 7.0, 30 mM imidazole, 10% (vol/vol) glycerol, 0.5 mM TCEP) containing 0.1% Tween-20 and 0.5 mg ml<sup>–1</sup> lysozyme, sonicated and cleared by ultracentrifugation. The supernatant was loaded onto a Ni-NTA column and washed with 180 mM phosphate citrate, pH 7.0, and then 90 mM phosphate citrate, pH 7.0. The protein was eluted directly into a HiTrap SP HP column with buffer D (90 mM phosphate citrate pH 7.0, 500 mM imidazole) and washed with 180 mM phosphate citrate,

pH 7.0. The protein was eluted in buffer E (180 mM phosphate citrate pH 7.0, 100 mM potassium glutamate, 250 mM NaCl, 0.5 mM TCEP). The protein was concentrated in a Vivaspin 500 column (Sartorius), diluted with glycerol (50% final), aliquoted, frozen in liquid nitrogen and stored at –80 °C.

MmPrimPol<sub>1–338</sub> and XtPrimPol<sub>1–334</sub> were expressed from pET28a-MmPrimPol<sub>1–338</sub> and pET28a-XtPrimPol<sub>1–338</sub>, respectively, in BL21 cells (the amino acid sequence of *X. tropicalis* PrimPol can be found in Supplementary Note 2). The transformed cell cultures were grown in standard TB medium to OD<sub>600</sub> of 0.6. Expression of the proteins was induced by adding IPTG to a final concentration of 0.4 mM, followed by overnight incubation at 20 °C (MmPrimPol<sub>1–338</sub>) or 25 °C (XtPrimPol<sub>1–334</sub>). In brief, collected cells were resuspended in buffer F (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 30 mM imidazole, 10% (vol/vol) glycerol, 17  $\mu$ g ml<sup>–1</sup> PMSF, 34  $\mu$ g ml<sup>–1</sup> benzamidine) with 0.1% IGEPAL and 1 mg ml<sup>–1</sup> lysozyme, sonicated and cleared by ultracentrifugation. The supernatant was loaded onto a Ni-NTA column, washed with 5% buffer G (same as buffer F with 300 mM imidazole and 2 mM  $\beta$ -mercaptoethanol) and eluted in buffer G. To load the proteins onto a HiTrap Heparin HP column, the samples were diluted 1:10 in buffer H (50 mM Tris-HCl pH 7.5, 50 mM NaCl, 10% (vol/vol) glycerol, 2 mM DTT). The proteins were eluted using a gradient of up to 1 M NaCl. MmPrimPol<sub>1–338</sub> was further purified with an SP column, using the same protocol as for the HiTrap Heparin HP column above. XtPrimPol<sub>1–334</sub> was further purified by size exclusion chromatography on a Superdex 75 10/300 GL gel filtration column (Cytiva) using buffer J (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 10% (vol/vol) glycerol, 0.5 mM TCEP). The proteins were frozen in liquid nitrogen and stored at –80 °C.

HsPriI WT and the D109A/D111A/D306E mutant (plasmids pKZ241 and pKZ248, respectively) were expressed in BL21(DE3) *E. coli* cells. The culture was grown in TB medium at 37 °C to OD<sub>600</sub> of 0.8–1. Protein expression was induced with 1 mM IPTG, and the culture was further incubated at 16 °C for 16 h. The first purification step by IMAC was performed as for the MpCAPP PP fragment as described above. After elution from cobalt resin, the protein was diluted 1:5 with ion exchange wash buffer (50 mM HEPES pH 7.5, 10% (vol/vol) glycerol, 0.5 mM TCEP) and loaded onto a 5-ml HiTrap Q HP column (Cytiva). The protein was eluted with a 100-ml gradient of 0–750 mM NaCl. The peak fractions containing PriI were pooled, diluted 1:5 with ion exchange wash buffer and loaded on a 5-ml HiTrap SP HP column (Cytiva). The bound protein was eluted with a 100-ml gradient of 0–750 mM NaCl. The eluted protein was further purified using a HiLoad 16/600 Superdex 200 pg column (Cytiva) in buffer containing 50 mM HEPES pH 7.5, 250 mM NaCl, 10% (vol/vol) glycerol and 0.5 mM TCEP. The fractions containing PriI were pooled, concentrated in a Vivaspin 20 column (Sartorius), diluted with glycerol (50% final), aliquoted, frozen in liquid nitrogen and stored at –80 °C.

Gels showing all purified proteins used in this study are available in Extended Data Fig. 9.

**Polymerase assays**

Polymerase assays were performed as described previously<sup>3</sup> with minor changes. Twenty-microlitre reactions contained DNA substrate (FAM-labelled DNA primer and DNA template (oligonucleotides oPK405 + oPK404 or oNB2 + oNB1)) and 100  $\mu$ M dNTPs in MpPolBuffer (10 mM Bis-Tris propane pH 7.0, 10 mM MgCl<sub>2</sub>, 10 mM NaCl and 0.5 mM TCEP) for MpCAPP or EuPolBuffer (10 mM Bis-Tris propane pH 7.0, 10 mM MgCl<sub>2</sub>, 0.5 mM TCEP, 0.1 mg ml<sup>–1</sup> BSA) for HsPrimPol. Reactions were supplemented with MpCAPP or HsPrimPol variants (protein concentrations are given in the figure legends) and incubated at 37 °C for the time indicated. Reactions were quenched with 20  $\mu$ l stop buffer (92.5% formamide, 5 mM EDTA, 0.025% SDS) and boiled for 3 min before electrophoresis on a denaturing gel containing 15% polyacrylamide (19:1 acrylamide:bis-acrylamide, 7 M urea, 1 $\times$  TBE buffer). The gel was run at 25 W for 1.5 h in 1 $\times$  TBE. Extended fluorescent primers



were imaged using a Typhoon FLA 9500 scanner (Cytiva). For figures, contrast was adjusted in the linear range using ImageJ.

ImageJ was used for quantification of primer extension products using unmodified original scans. The signal of each band in the sample lane was assigned a number corresponding to the number of nucleotides added (ligated) to the fluorescently labelled primer (no extension = 0 and full template-dependent primer extension =  $x$  nucleotides). The weighted average was used to calculate the average extension length for each sample, where each extension (band;  $0-x$ ) had the importance of the value for the intensity of its fluorescence signal. The average extension of wild-type protein was used to standardize each gel (WT = 100%). Data represent the mean  $\pm$  s.d. Source data are presented in Supplementary Table 2. The sequences for DNA oligonucleotides used for this assay are available in Supplementary Table 1.

### Primase assays

For MpCAPP and MsCAPP, 20- $\mu$ l reactions contained protein at the concentration indicated in the figure legend, 1  $\mu$ M ssDNA oligonucleotide template, and 2.5  $\mu$ M dATP, dTTP, dGTP and FAM-labelled dCTP (NU-809-5FM, Jena Bioscience) and 100  $\mu$ M non-labelled GTP or 100  $\mu$ M dNTP mix and 10  $\mu$ M FAM- $\gamma$ GTP (NU-834-6FM, Jena Bioscience) in MpPrimBuffer (10 mM Bis-Tris propane pH 7.0, 0.5 mM TCEP, 10 mM MgCl<sub>2</sub>, 100  $\mu$ M ZnCl<sub>2</sub>). Reactions were incubated at 50 °C for 30 min.

For eukaryotic PrimPol proteins, unless otherwise stated in the figure legend, 20- $\mu$ l reactions contained protein at the concentration indicated, 1  $\mu$ M ssDNA oligonucleotide template, and 100  $\mu$ M dNTPs and 2.5  $\mu$ M fluorescently labelled FAM- $\gamma$ GTP or FAM- $\gamma$ ATP (NU-833-6FM, Jena Bioscience) or 2.5  $\mu$ M unlabelled dCTP, dTTP and dGTP, 2.5  $\mu$ M labelled FAM-dATP (NU-835-6FM, Jena Bioscience) and 100  $\mu$ M individual NTPs in EuPrimBuffer (10 mM Bis-Tris propane pH 7.0, 10 mM MnCl<sub>2</sub>, 0.5 mM TCEP, 0.1 mg ml<sup>-1</sup> BSA). Reactions were incubated at 37 °C for 30 min.

For human PriI, 20- $\mu$ l reactions contained PriI at the concentration indicated in the figure legend, 1  $\mu$ M ssDNA oligonucleotide template (oKZ388), and 100  $\mu$ M non-labelled ATP, UTP and CTP or dATP, dTTP and dCTP and 10  $\mu$ M FAM- $\gamma$ GTP or 100  $\mu$ M non-labelled GTP, UTP and CTP and 10  $\mu$ M FAM- $\gamma$ ATP in PriIPrimBuffer (10 mM Tris-HCl pH 8.0, 0.5 mM TCEP, 5 mM MnCl<sub>2</sub>, 0.2 mg ml<sup>-1</sup> BSA). Reactions were incubated at 25 °C for 30 min.

All primase reactions (20- $\mu$ l volume) were precipitated by adding 20  $\mu$ l CTAB solution (200  $\mu$ M CTAB, 30 mM ammonium sulfate, 25 mM EDTA) and centrifuged at room temperature at 16,000g for 10 min. The pellet was resuspended in 25  $\mu$ l loading buffer (92.5% formamide, 25 mM EDTA, 0.5% Ficoll 400). Samples were boiled for 3 min before 20  $\mu$ l was loaded on a 24% (if not indicated otherwise in the figure legend) urea-PAGE gel (19:1 acrylamide:bis-acrylamide, 8 M urea (20% gel) or 6 M urea (24% gel), 1 $\times$  TBE buffer) and run at 25 W for 2 h and 15 min in 1 $\times$  TBE. Products were imaged using a Typhoon FLA 9500 scanner (Cytiva). For figures, contrast was adjusted in the linear range using ImageJ.

ImageJ was used to quantify products of the primase assay using unmodified original scans. The signal of the sample in the whole lane excluding fluorescently labelled mononucleotide was quantified and the background signal was subtracted (signal of control lane without protein excluding the signal of labelled mononucleotide). The primer synthesis signal of wild-type protein was used to standardize each gel (WT = 100%). Data are the mean  $\pm$  s.d. Source data are presented in Supplementary Table 2. Sequences for DNA templates used in primer assays can be found in Supplementary Table 1.

### Crystallization, data collection and structure determination

The construct for *Marinitoga* sp. 1137 CAPP<sub>100-359</sub> was expressed and purified in the same way as MpCAPP PP WT, with an additional step of size exclusion chromatography using a HiLoad 26/60 Superdex 75 prep-grade column (Cytiva) with buffer containing 25 mM HEPES pH 7.4, 250 mM NaCl and 0.5 mM TCEP. The construct for *Marinitoga*

sp. 1137 CAPP<sub>111-328</sub> was cloned into pOPINF<sup>35</sup> with a cleavable His tag, which was removed via overnight incubation with 3C protease (home-made) at 4 °C, and was otherwise purified using the same method as for MsCAPP<sub>100-359</sub>. Crystal screening experiments were set up with 6 $\times$ His-tagged *Marinitoga* sp. 1137 CAPP (residues 100-359 for the apo form and dGTP complex, residues 111-328 for DNA complexes), and matrix screens (Molecular Dimensions, Hampton Research) were performed using the sitting drop vapor diffusion method with equal volumes of protein solution (381  $\mu$ M for the apo form and dGTP complex, 90  $\mu$ M for DNA complexes) and reservoir buffer. Apo crystals were grown in 0.1 M propionate, cacodylate, Bis-Tris propane (PCTP) pH 4.0 and 25% PEG 1500. For the dGTP complex, CAPP was co-crystallized with 2 mM dGTP and 10 mM MnCl<sub>2</sub> in 20% PEG 3350 and 0.2 M potassium thiocyanate. For the primer initiation complex, CAPP was co-crystallized with 200-400  $\mu$ M 5'-AAAATCAA-3' DNA oligonucleotide (ATDBio), 0.5 mM dAMPNPP (NU-443-1, Jena Bioscience), 2-4 mM GTP and 2 mM CoCl<sub>2</sub> in 0.1 M HEPES/MOPS pH 7.5, 20% ethylene glycol, 10% PEG 8000, 0.1 M diethylene glycol, 0.1 M triethylene glycol, 0.1 M tetraethylene glycol, 0.1 M pentaethylene glycol and 140 mM potassium glutamate. For the dsDNA complex, CAPP was co-crystallized with 180  $\mu$ M 5'-CGTGGC-3' DNA oligonucleotide (ATDBio), 2 mM GTP and 5 mM CoCl<sub>2</sub> in 0.05 M sodium cacodylate trihydrate, 10% PEG 4000, 0.1 M sodium chloride and 0.5 mM spermine.

Apo crystals were cryoprotected in the mother liquor with 30% glycerol; dGTP and dsDNA co-crystals were cryoprotected in the mother liquor with 25% ethylene glycol; and the crystal for the primer initiation complex was cryoprotected in the mother liquor alone. Diffraction data were collected at beamlines I03 and I04 of Diamond Light Source (Didcot, UK).

The diffraction data were processed with xia2. The initial phase solution was obtained with SHELXC/D/E<sup>36</sup> using weak anomalous signal from manganese atoms for the dGTP complex dataset. The statistics for data processing are summarized in Extended Data Table 1. Automated model building was performed with ARP/wARP<sup>37</sup>, followed by alternate rounds of manual and automated refinement of the model using Coot<sup>38</sup> and phenix.refine<sup>39</sup>. Molecular replacement with Phaser<sup>40</sup> was performed on the apo and DNA complex datasets, using the dGTP complex model as a template. The refinement procedure for the apo and DNA complex datasets was the same as for the dGTP complex. Molecular graphics were generated with PyMOL (Schrödinger), with hydrophobic surfaces generated using a normalized consensus hydrophobicity scale<sup>41</sup>.

### Fluorescence polarization assays

Fifty-microlitre reactions contained MsCAPP<sub>111-328</sub> protein at the concentration indicated in the figure legend, 10 mM Bis-Tris propane pH 7.0, 10 mM MgCl<sub>2</sub>, 0.1 mM ZnCl<sub>2</sub>, 50 mM NaCl, 0.05% Tween-20 and 50 nM FAM-ssDNA oligonucleotide template with or without 1 mM GTP/dGTP and 0.1 mM dATP or 25 nM FAM-dATP or 25 nM FAM- $\gamma$ GTP and measured immediately at room temperature or as otherwise indicated in the figure legend. FP (excitation filter, 482-16 nm; dichroic filter, LP 504 nm; emission filter, 530-540 nm) was measured using a CLARIOstar multimode plate reader (BMG Labtech). Background signal (no-protein sample) was subtracted from each value, and the data obtained were analysed using GraphPad Prism (v.9.0, GraphPad Software). Data are the mean  $\pm$  s.d. The equation for specific binding with a Hill slope was used for data curve fitting to obtain  $K_d$  values. Calculated  $K_d$  values that had a 95% interval of confidence containing negative values were excluded and labelled as not determined (ND). The equation [agonist] versus response - variable slope was used to calculate the half-maximal effective concentration (EC<sub>50</sub>) and the equation [inhibitor] versus response - variable slope was used to calculate the half-maximal inhibitory concentration (IC<sub>50</sub>). Source data used in GraphPad Prism are presented in Supplementary Table 2. All data were obtained from at least three independent assays. Sequences of the DNA templates used in FP assays are listed in Supplementary Table 1.

## Computational methods

Non-covalent interaction (NCI) analysis of the active site using a cut-off of 8 Å was performed with the NCIPLLOT programme<sup>42</sup>. The NCI analysis method is based on a reduced density gradient and the electron density, allowing attractive forces to be distinguished from repulsive forces. These forces were represented as isosurfaces using the visualization programme VMD (Visual Molecular Dynamics)<sup>43</sup>. The default red–green–blue (RGB) colour code for isosurfaces was used to represent attractive interactions (hydrogen bonds) in blue, repulsive interactions (steric effects in rings) in red and weak attractive interactions (that is, van der Waals forces) in green.

Symmetry-adapted perturbation theory (SAPT) calculations, which are based on Rayleigh–Schrödinger perturbation theory, were performed to investigate individual contributions to the total intermolecular energy<sup>44</sup>. In SAPT, the total intermolecular energy can be expressed as the sum of electrostatic, exchange repulsion, induction and dispersion contributions. SAPT can be truncated at different orders of inter- and intramolecular perturbation, offering several levels of SAPT<sup>45,46</sup>. We carried out SAPTO calculations for selected pairs of residues present in the active site, using the def2-SV(P) basis set with PSI4 code<sup>47</sup>. The DNA bases (dA, dC and dT) and amino acids (Y138 and R223) in the active site were fragmented and completed with a methyl group to only consider the side chains for SAPTO calculations, and Mg(II) was used as a surrogate for the divalent metal ions (Extended Data Fig. 4b). The charge transfer energy was computed following the Stone–Misquitta (SM09) definition<sup>48</sup>.

## Reporting summary

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

## Data availability

The coordinates and crystallographic structure factors for MscCAPP have been deposited at PDB under accession codes 7NQG (apo), 7NQE (dGTP complex), 7NQF (postcatalytic complex), 7P9J (primer initiation complex) and 7QAZ (primer initiation complex (alternate model)). All other data needed to evaluate the conclusions in this study are present in the manuscript and/or its supplementary information and tables. Uncropped versions of all gels are provided in Supplementary Fig. 1. Source data for graphs (gel-based assays and FP assays) can be found in Supplementary Table 2.

35. Berrow, N. S. et al. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.* **35**, e45 (2007).

36. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D* **66**, 479–485 (2010).
37. Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171–1179 (2008).
38. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
39. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352–367 (2012).
40. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
41. Eisenberg, D., Schwarz, E., Komaromy, M. & Wall, R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* **179**, 125–142 (1984).
42. Contreras-García, J. et al. NCIPLLOT: a program for plotting noncovalent interaction regions. *J. Chem. Theory Comput.* **7**, 625–632 (2011).
43. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
44. Jeziorski, B. et al. SAPT: a program for many-body symmetry-adapted perturbation theory calculations of intermolecular interaction energies. *Methods Tech. Comput. Chem.* **B**, 79–129 (1993).
45. Parker, T. M., Burns, L. A., Parrish, R. M., Ryno, A. G. & Sherrill, C. D. Levels of symmetry adapted perturbation theory (SAPT). I. Efficiency and performance for interaction energies. *J. Chem. Phys.* **140**, 094106 (2014).
46. Naseem-Khan, S., Gresh, N., Misquitta, A. J. & Piquemal, J.-P. Assessment of SAPT and supermolecular EDA approaches for the development of separable and polarizable force fields. *J. Chem. Theory Comput.* **17**, 2759–2774 (2021).
47. Turney, J. M. et al. Psi4: an open-source ab initio electronic structure program. *WIREs Comput. Mol. Sci.* **2**, 556–565 (2012).
48. Stone, A. J. & Misquitta, A. J. Charge-transfer in symmetry-adapted perturbation theory. *Chem. Phys. Lett.* **473**, 201–205 (2009).

**Acknowledgements** The laboratory of A.J.D. was supported by grants from the Biotechnology and Biological Sciences Research Council (BB/S008691/1 and BB/P007031/1). L.J.B. was supported by a PhD studentship funded by an Institutional Strategic Support Fund 2 grant from the Wellcome Trust (204833/Z/16/Z). The computational simulations were funded by NIH grant R01GM108583. Computational time was provided by the University of North Texas CASCAs CRUNTCh3 high-performance cluster partially supported by NSF grant CHE-1531468 and XSEDE supported by project TG-CHE160044. We thank M. Roe for assistance with phasing, Diamond Light Source for beamtime (proposal MX20145) and the staff of beamlines I03, I04 and I04-1 for assistance with data collection.

**Author contributions** A.J.D. designed the project and directed the experimental work, and co-wrote the manuscript with A.W.H.L., K.Z. and L.J.B. A.W.H.L., K.Z., L.J.B., M.Z. and P.K. contributed to project design and performed and analysed the experiments: A.W.H.L. performed all crystallographic experiments; K.Z. performed MpCAPP, MsCAPP and HsPri1 polymerase and primase assays, MscCAPP FP assays and MpCAPP iron-binding assays; L.J.B. performed primase assays with eukaryotic PrimPol proteins; and M.Z. performed initial primase assays with MscCAPP and HsPrimPol PPs. P.K. performed initial activity assays with MscCAPP. S.N.-K., M.B.B. and G.A.C. designed, performed and analysed the *in silico* molecular modelling experiments.

**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/s41586-022-04695-0>.

**Correspondence and requests for materials** should be addressed to Aidan J. Doherty.

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

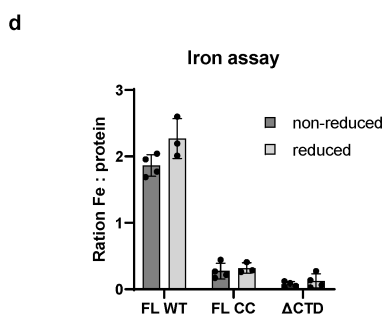
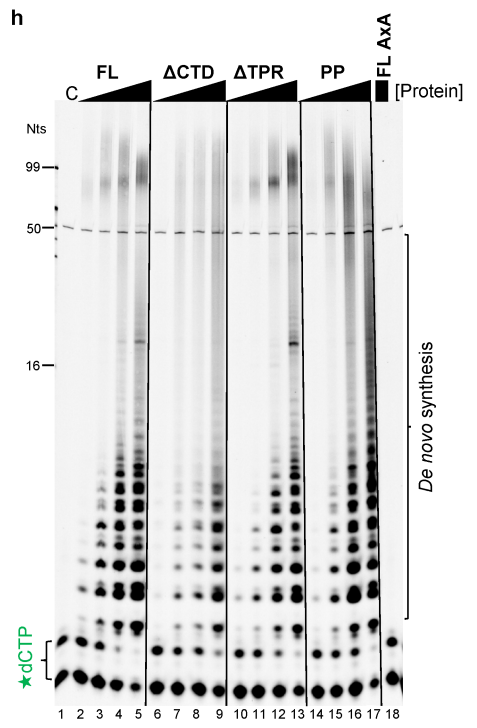
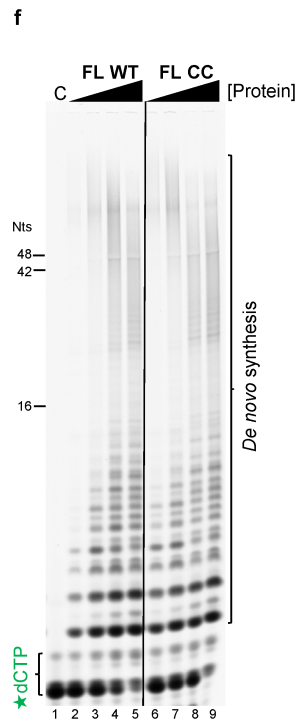
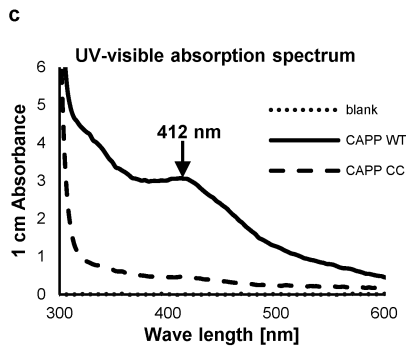
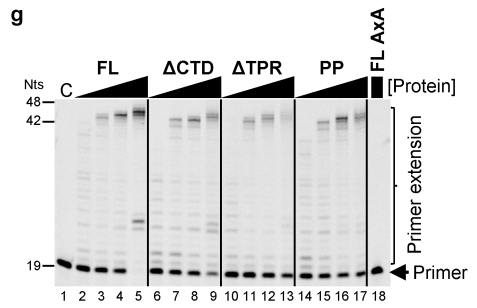
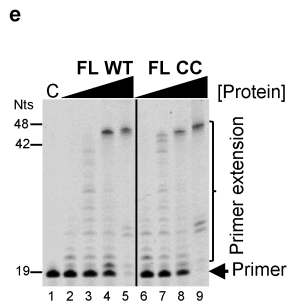
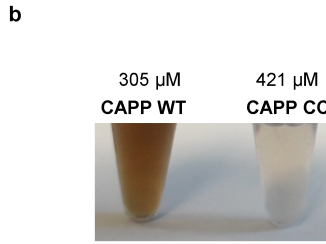
**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

**a** **C-terminal domain (CTD)**

Sc aa330 SSNYPLIKNLMGLKKNHHLRYYGRQQLSLFLKGLGSLADEALKFWSEAxxxR aa396  
 Lp aa258 IKNFPAIQEILSKLKEGQNLVHTERLYIYVWFLNALKYPEEEIINIFSALPDFD aa311  
 Fg aa401 TLLESSFVVDQIMRKIET-LAHISAEEEKILVSLKHLPLNGYGVAKKVKLEKTI aa453  
 Tp aa360 NTMIQAAYPLKQIYEKIQK-LGYINEDEEYIITIAACVAMENSKEFLSDFLKKTI aa412  
 Mp aa370 TQMIEGSHILKQINEKIEK-EAYITEEEMIFIKSLANLENSKFFIEMKLSQTI aa422

Sc aa397 YSFRHNYGLEGNRINYKPWDHTI--LSKPRPGRGDYHGC PFRDxxxIAC TKVF aa478  
 Lp aa312 RDKTGYQVRFKAKKGYTPYSKTLK-----TYSLLGKFKFKDKLLEGY aa355  
 Fg aa454 DYSEEKVRNLFDEVGNIPTAEEIKALATSLNIPLDLSKCVRFKQVFNTPLVN aa506  
 Tp aa413 NYSQSRHLHALMNRVGTITTEETKRIILSKSLALNIDQNNKFNELVNTPLCL aa466  
 Mp aa423 NFSKKRIDMIKQSLGVPITTEETRKIILNKDISLLENTEKFKQGNYSPIYAI aa476

C462 C464



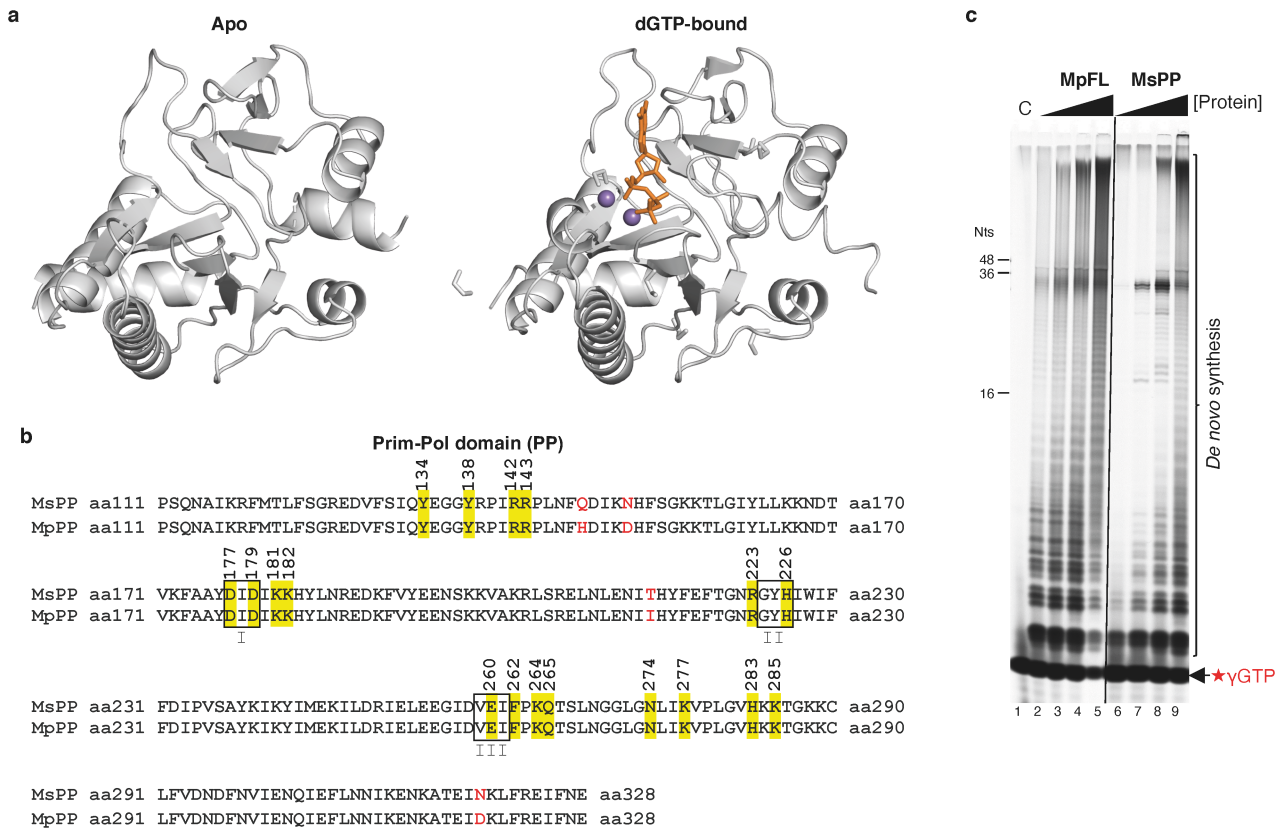
**Extended Data Fig. 1** | See next page for caption.

# Article

## Extended Data Fig. 1 | CAPP's Prim-Pol domain alone is sufficient for polymerase and primase synthesis.

**a.** Alignment of the C-terminal domains (CTD) of different primases or CAPPs. Sc: *Saccharomyces cerevisiae* PriL (Pri2p), Ls: *Lokiarchaeum sp. GC14\_75* PriL, Mp: *Marinitoga piezophila* CAPP, Tp: *Thermotoga profunda* CAPP, Fg: *Fervidobacterium gondwanense* CAPP. Conserved motifs: cysteines – red; basic residues – blue; hydrophobic residues – green; prolines – yellow. **b.** Visualisation of MpCAPP full length wild-type (WT) and C462S, C464S (CC) proteins after two-step purification. **c.** UV-visible absorption spectrum of MpCAPP WT and CC. CAPP WT, but not CC, exhibited a major absorbance peak at 412 nm. **d.** Comparison of three different MpCAPP constructs using an Iron assay kit (MAK025, Sigma). MpCAPP FL WT sample had a significantly higher concentration of iron than the MpCAPP CC mutant or CAPP  $\Delta$ CTD fragment (aa 1-360). Data are the mean of four measurements, except reduced FL WT and FL CC – three measurements. Error bars indicate the mean  $\pm$  standard deviation. Individual values are presented as black dots. Reduced – sample after treatment with an iron reducing agent; Non-reduced – sample without treatment. **e.** MpCAPP CC mutant has similar polymerase activity as WT. 1, 5, 25 and 125 nM MpCAPP FL WT (lanes 2-5) or FL CC (lanes 6-9)

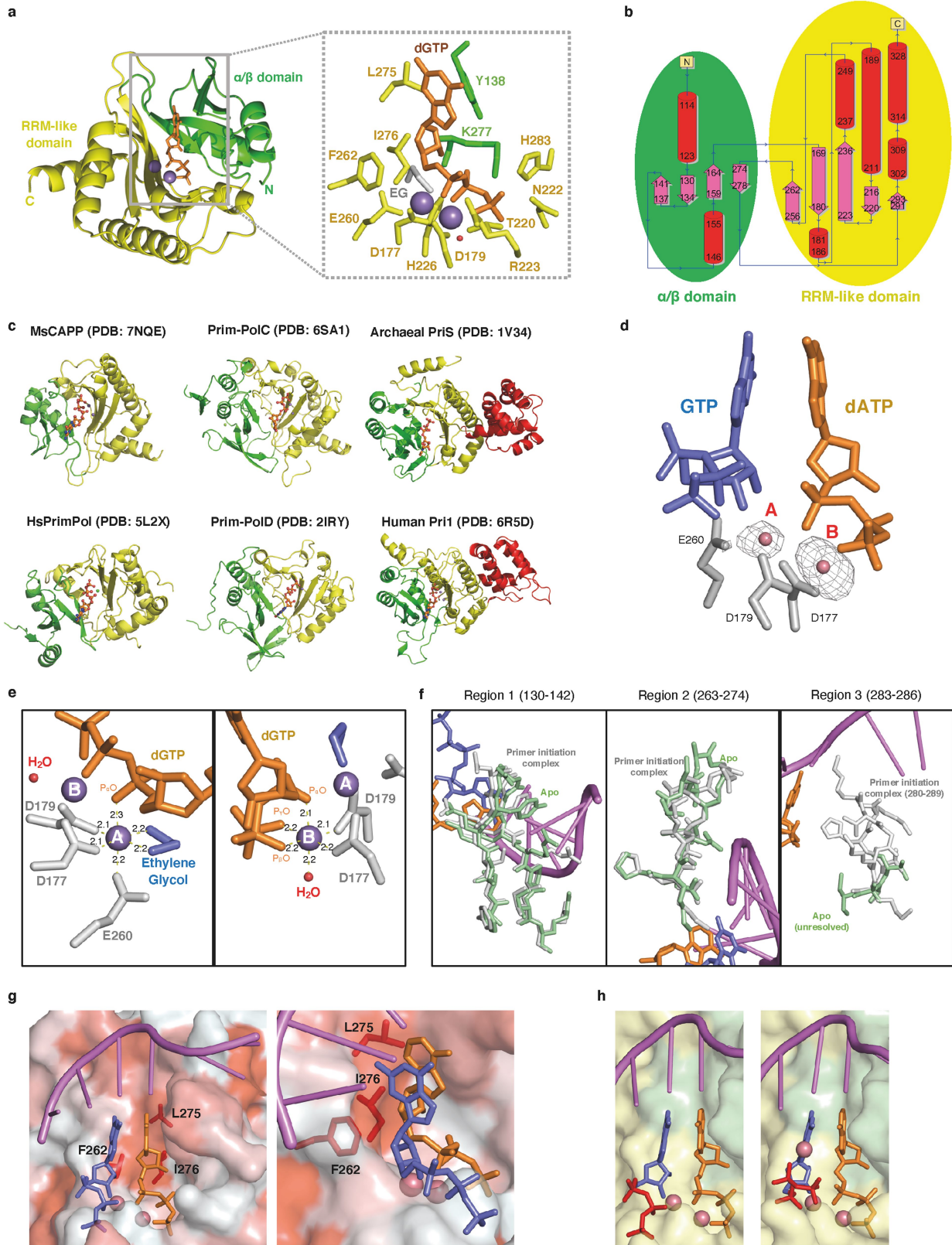
was added into 30 nM DNA substrate (DNA template – oPK404 + FAM-labelled DNA primer – oPK405) and 100  $\mu$ M dNTPs. Reactions were incubated at 37 °C for 30 min. **f.** MpCAPP CC mutant has comparable priming activity as WT. 0.5, 1, 2 and 4  $\mu$ M MpCAPP FL WT (lanes 2-5) or FL CC (lanes 6-9) was added into reactions containing 1  $\mu$ M ssDNA template (oKZ388), 2.5  $\mu$ M non-labelled dATP, dTTP, dGTP, 2.5  $\mu$ M FAM-dCTP ( $\star$ dCTP) and 100  $\mu$ M GTP. The reactions were incubated at 50 °C for 30 min. The products were resolved on 20% urea-PAGE gel. **g.** MpCAPP PP domain exhibits efficient polymerase activity. 1, 5, 25 and 125 nM MpCAPP full length wild-type (FL) (lanes 2-5) and its fragments (lanes 6-17) or 125 nM D177A, D179A full-length mutant (FL AxA) (lane 18) were tested as described in panel e. **h.** MpCAPP PP domain exhibits strong primase activity. 0.25, 0.5, 1 and 2  $\mu$ M of MpCAPP FL (lanes 2-5) and its fragments (lanes 6-17) or 2  $\mu$ M FL AxA (lane 18) were tested as described in panel f. 'C' indicates a control reaction without protein. Oligonucleotide (Nts) length markers are shown on the left of the gel. Results shown in panels e-h are representative of three independent repeats, except PP and FL AxA in polymerase assay – four independent repeats.



**Extended Data Fig. 2 | MsCAPP PP structure and its primase activity.**

**a.** Structure of MsCAPP PP domain in cartoon representation showing the apo protein (left) and the dGTP complex with Mn(II) ions (right). Protein – grey, dGTP – orange, Mn(II) – purple spheres. **b.** Protein sequence alignment of MpCAPP<sub>111-328</sub> (MpPP) and *Marinitoga sp. 1137* CAPP<sub>111-328</sub> (MsPP). Conserved amino acids mutated in this study are shown in yellow. Non-conserved amino acids are shown in red. Conserved catalytic motifs I, II and III are in black

rectangles as indicated. **c.** MsPP possesses priming activity. 0.5, 1, 2 and 4  $\mu$ M MpCAPP full-length (MpFL) (lanes 2-5) or MsCAPP<sub>111-328</sub> (MsPP) (lanes 6-9) was added into the reaction containing 1  $\mu$ M ssDNA mixed sequence template (oKZ388), 100  $\mu$ M non-labelled dNTP mix and 10  $\mu$ M FAM- $\gamma$ GTP ( $\star\gamma$ GTP). The reactions were incubated at 50  $^{\circ}$ C for 30 min. 'C' indicates a control reaction without protein. Oligonucleotide (Nts) length markers are shown on the left of the gel. Results shown are representative of three independent repeats.

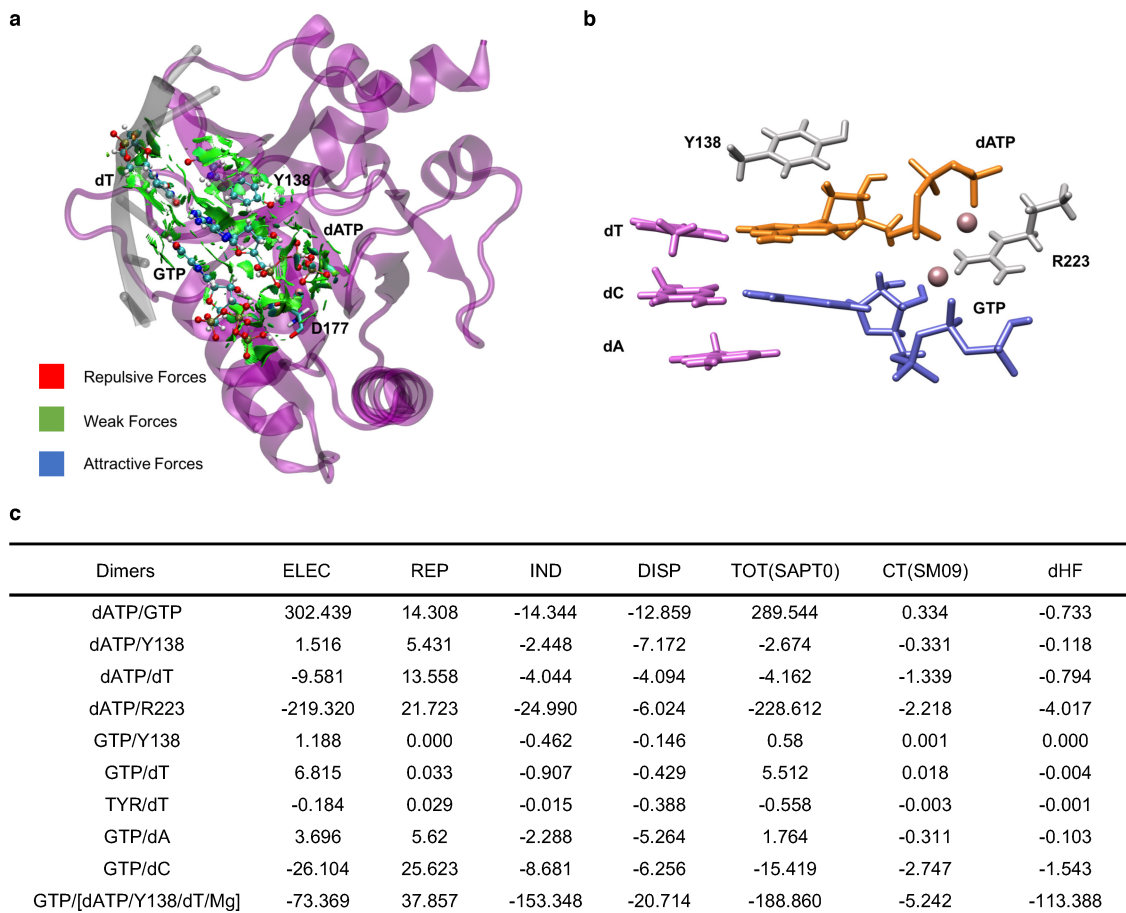


Extended Data Fig. 3 | See next page for caption.

**Extended Data Fig. 3 | MsCAPP PP domain structure analyses and comparison.**

**a.** Overall structure of dGTP-complexed MsCAPP PP domain, with N-terminal  $\alpha/\beta$  domain coloured in green and RRM-like domain coloured in yellow (left), and a close-up view showing dGTP (orange), Mn(II) ions (spheres) and residues lining the active site pocket. **b.** Architecture of MsCAPP PP domain showing the secondary structural elements within the  $\alpha/\beta$  domain (aa 111-164; aa 274-278) in green and the RRM-like domain (aa 169-262; aa 291-328) in yellow. **c.** Side by side comparison of PP domains from various Prim-Pols with a single nucleotide (ball-and-stick model) bound to the elongation site. N-terminal RRM-like domain (yellow),  $\alpha/\beta$  domain (green) and helical domain (red). **d.** Simulated annealing Fo-Fc omit map of Co(II) ions in the active site of MsCAPP (contoured at  $5\sigma$ -level at 1.90 Å resolution), along with GTP (blue), dATP (orange), and acidic residues D177, D179 and E260 in stick representation. **e.** Close-up view of Mn(II) bound in the A site of the MsCAPP dGTP complex, showing octahedral coordination to dGTP, ethylene glycol and surrounding acidic residues with distances labelled in Å (left). Close-up view of Mn(II) bound in the B site of MsCAPP dGTP complex, showing octahedral coordination to dGTP, DxD motif and a water molecule, with distances labelled in Å (right).

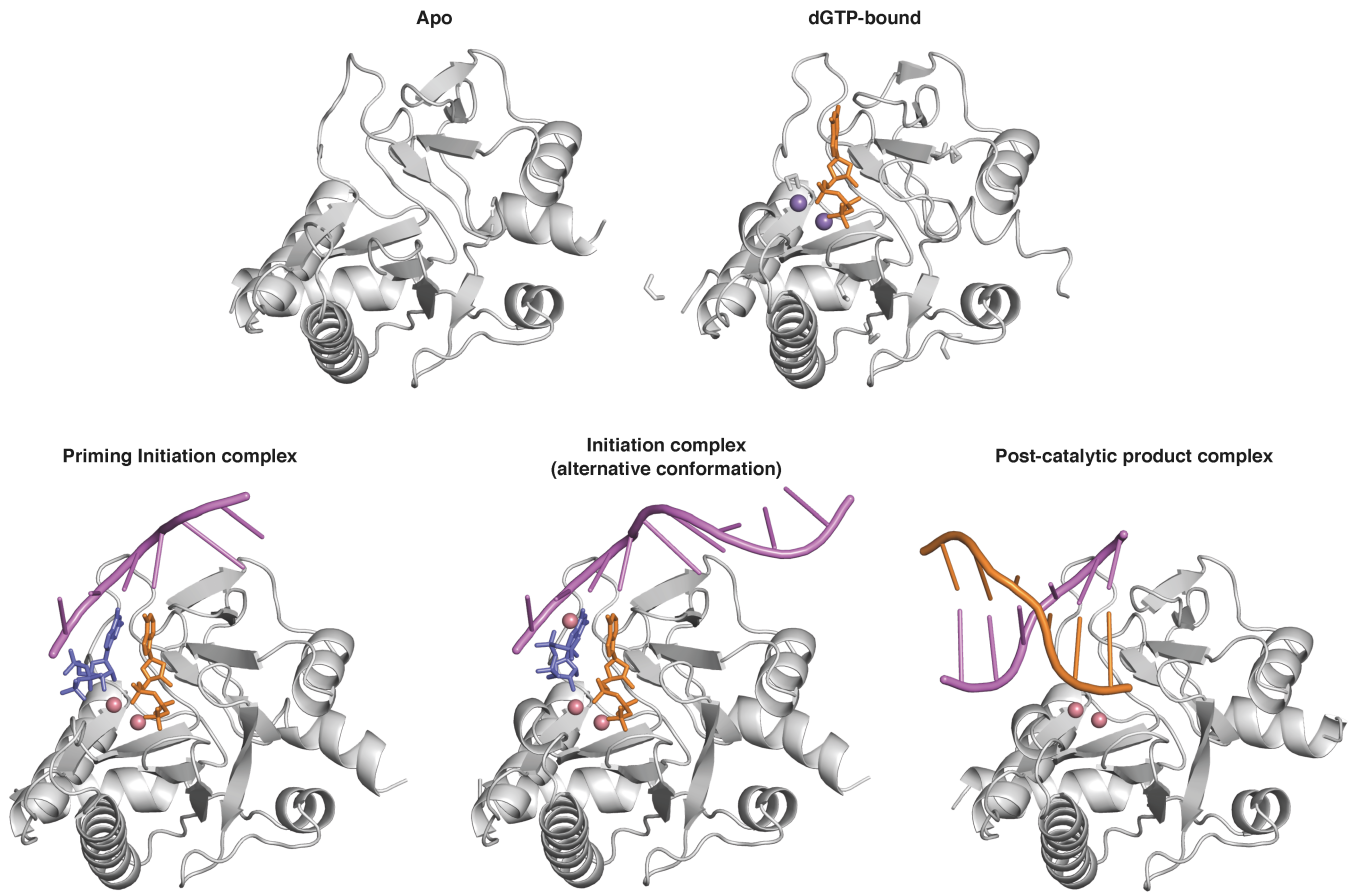
Mn(II) – purple spheres. **f.** Overlay of Region 1 (aa 130-142) (left), Region 2 (aa 263-274) (middle) and residues around Region 3 (aa 280-289) (right) from the structures of MsCAPP apo (green) and primer initiation complex (grey). Residues 283-286 in Region 3 are flexible and not resolved in the apo structure. **g.** MsCAPP active site with surface coloured according to hydrophobicity, with regions of high hydrophobicity coloured in red. Three core residues that form part of the active site hydrophobic pocket (L275, I276 and F262) are shown in stick representation (red). I-site GTP (blue) and E-site dATP (orange) are shown in stick representation. Templating DNA (pink) is shown in cartoon representation. **h.** Comparison of CAPP primer initiation complex active site. Figure on the left – CAPP primer initiation complex shown in the main figures (PDB: 7P6J). Figure on the right shows the structure of an alternative CAPP primer initiation complex (PDB: 7QAZ), where the phosphate tail (red) of I-site GTP adopts a different conformation and coordinate to an extra Co(II) ion. The rest of the GTP molecule is shown in blue and dATP in orange. Templating DNA in pink, surface of  $\alpha/\beta$  domain is coloured in green and RRM-like domain is shown in yellow.



**Extended Data Fig. 4 | Intermolecular interaction analyses of the active site of the primer initiation complex.** **a.** Non-Covalent Interaction (NCI) analysis of the active site (isosurface = 0.35, cut-off = 8 Å). **b.** Representation of the active site used for Symmetry-Adapted Perturbation Theory (SAPT0) calculations. For dA, dC, dT, Y138 and R223, only the side chains have been

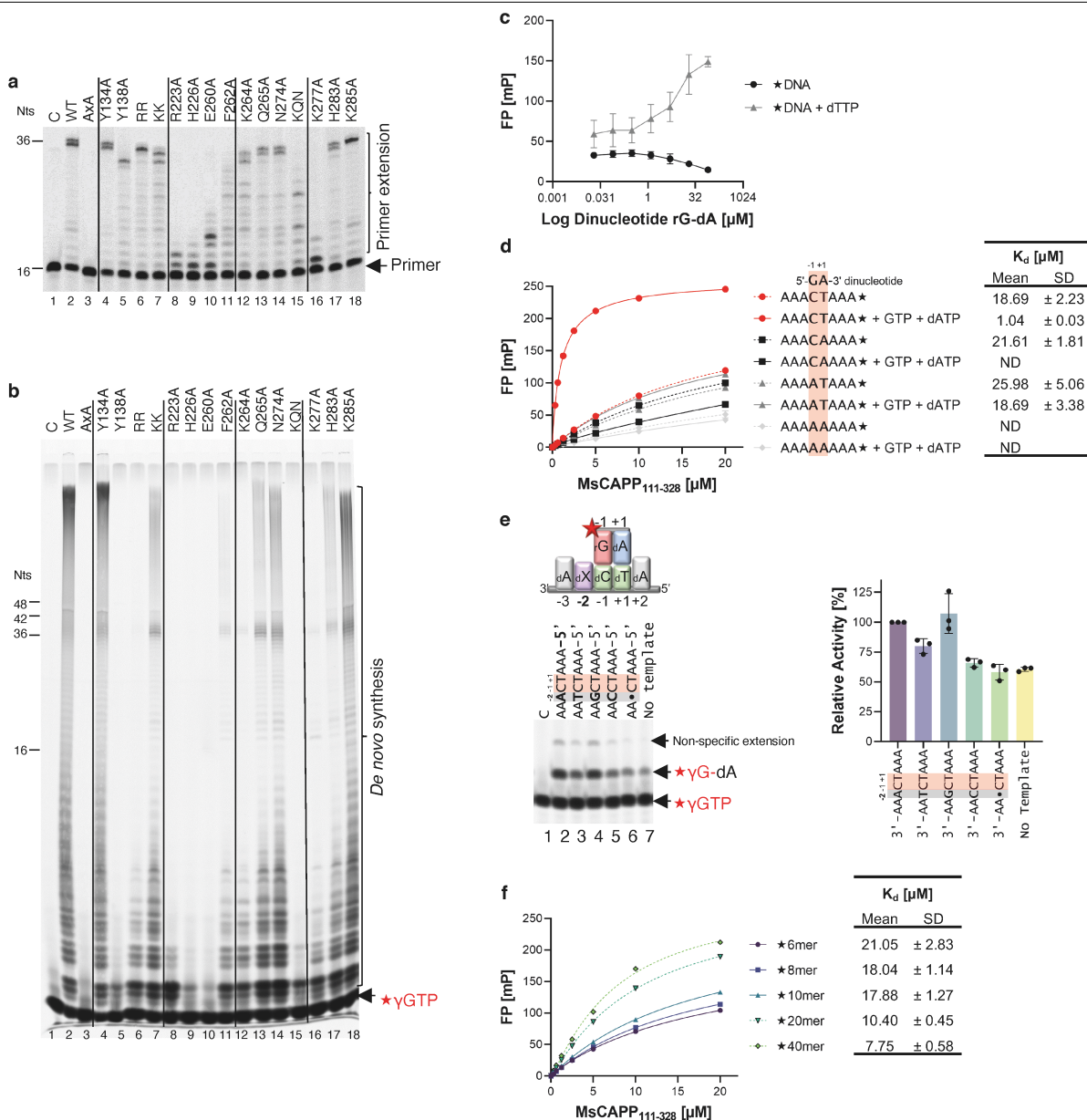
considered. **c.** Results from SAPTO calculations for the indicated pairs using the def2-SV(P) basis set in kcal/mol. The final row indicates the SAPTO/def2-SV(P) calculation for GTP interacting with all other fragments in the system (dATP, Y138, dT and both Mg(II) ions).





**Extended Data Fig. 5 | Comparison of the different crystal structures of the MsCAPP domain.** From left to right, top to bottom; apo, dGTP-bound, primer initiation complex, primer initiation complex (alternative conformation) and post-ternary complex. Protein – grey cartoon,

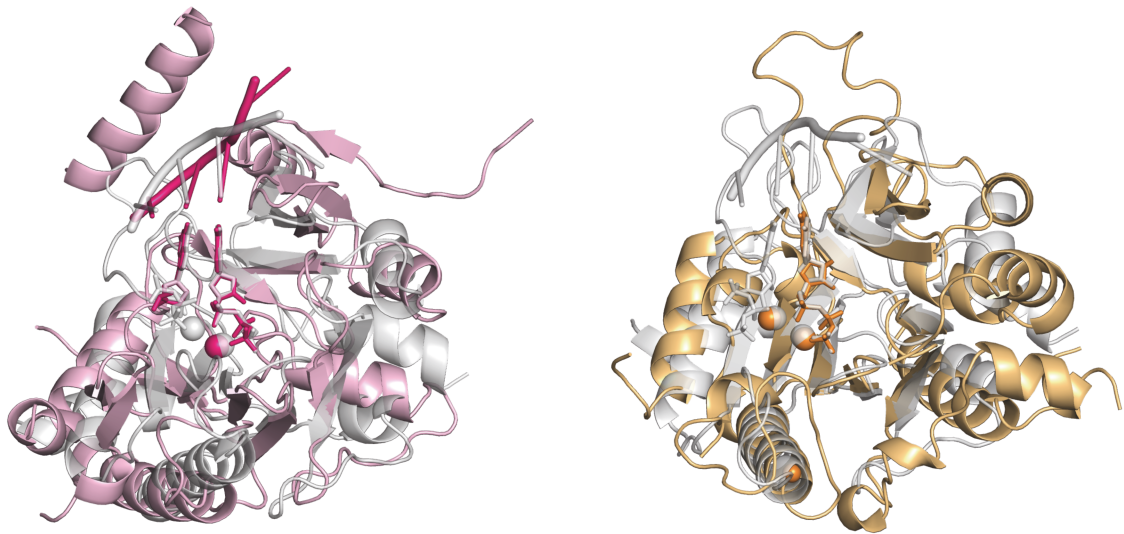
deoxyribonucleotide (dGTP / dATP) – orange sticks, ribonucleotide (GTP) blue sticks, templating DNA strand – pink cartoon, primer DNA strand – orange cartoon, Mn(II) – purple spheres, Co(II) – pink spheres.



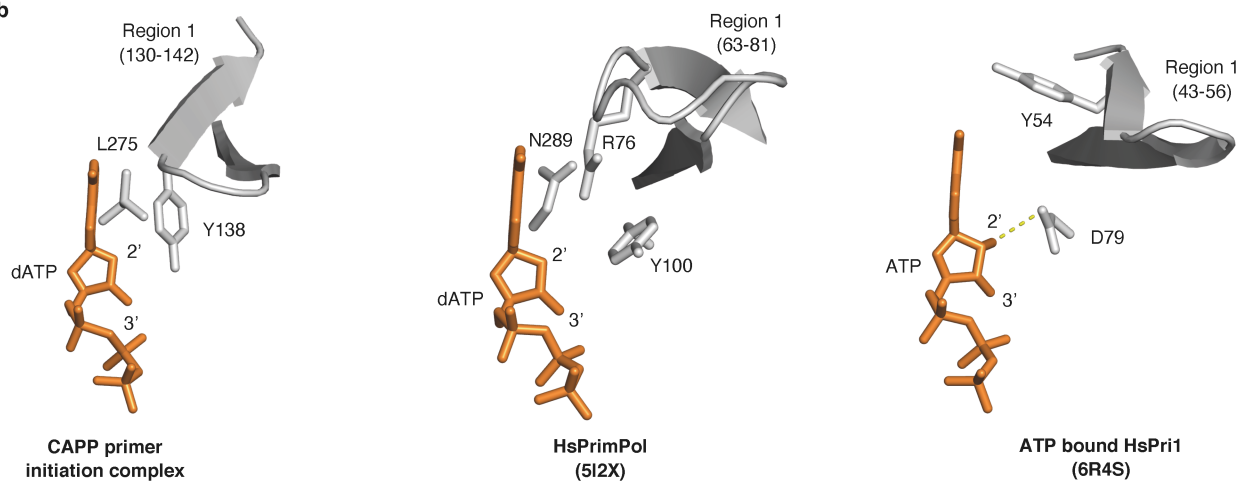
**Extended Data Fig. 6 | Structure-function and binding studies on the PP domain of MsCAPP.** **a.** Effect of mutations of MpCAPP<sub>100-360</sub> (MpPP) on its polymerase activity. 50 nM MpPP wild-type (WT) (lane 2) or its mutated variants (lanes 3-18) were added to 50 nM DNA substrate (DNA template – oNB1 + FAM-labelled DNA primer – oNB2) and 100  $\mu$ M dNTPs. The reactions were incubated at 37 °C for 15 min. **b.** Primase activity of MpPP WT and its mutants. 2  $\mu$ M MpPP WT protein (lane 2) or its mutants (lanes 3-18) were added to reactions containing 1  $\mu$ M DNA substrate (oKZ388), 100  $\mu$ M dNTP mix and 10  $\mu$ M FAM-labelled GTP (fused via  $\gamma$ -phosphate) (\* $\gamma$ GTP). The reactions were incubated at 50 °C for 30 min. The products were resolved on 20% urea-PAGE gel. 'C' indicates a control reaction without protein. AxA – D177A, D179A, RR – R142A, R143A, KK – K181A, K182A, KQN – K264A, Q265A, N274A. Results are representative of five (panel a) and four (panel b) independent repeats. Oligonucleotide length marker (Nts) is shown on the left of the gel. 'C' indicates control reaction without protein. **c.** Fluorescence polarization assays (FP) reveal that the presence of dinucleotide (rG-dA) does not stimulate binding of MsCAPP<sub>111-328</sub> (MsPP) to template. FP: 0–80  $\mu$ M 5'-<sub>3p</sub>rG-dA-3' (ATDBio) dinucleotide was added to 5  $\mu$ M MsPP and 50 nM FAM-DNA (\* DNA, oKZ409) in presence or absence of 0.1 mM dTTP. **d.** Stimulation of PP DNA binding affinity

in presence of nucleotides is dependent on the template sequence. FP: 0–20  $\mu$ M MsPP was added to 50 nM \* DNA templates (oKZ409, oKZ413, oKZ414 and oKZ416)  $\pm$  1 mM GTP and 0.1 mM dATP. **e.** The efficiency of PP dinucleotide formation is dependent on the –2 base on the template. 1  $\mu$ M MsPP was added into the reaction containing 1  $\mu$ M template (lane 2 - oKZ435, lane 3 - oKZ447, lane 4 - oKZ449, lane 5 - oKZ450, lane 6 - oKZ448), 100  $\mu$ M dATP and 10  $\mu$ M \* $\gamma$ GTP. The reactions were incubated at 50 °C for 30 min. The gel is representative of three independent repeats (left). Signal of synthesized dinucleotides were normalized to signal of dinucleotide in presence of 3'-AAACTAAA-5' ssDNA template (100%). Data represent the mean  $\pm$  standard deviation from three independent experiments. Black dots – individual values. 'C' indicates control reaction without protein. **f.** Affinity of PP to template increases with the template length. FP – 0–20  $\mu$ M MsPP was added to 50 nM \* DNA templates (oKZ408–oKZ412). Data representing the mean  $\pm$  standard deviation from four independent experiments (Panels c, d and f). The mean values were used to calculate the dissociation constants ( $K_d$ ) shown on the right (Panel d, f); SD – standard deviation of calculated  $K_d$ ; ND – not determined (Panels d, f).

a

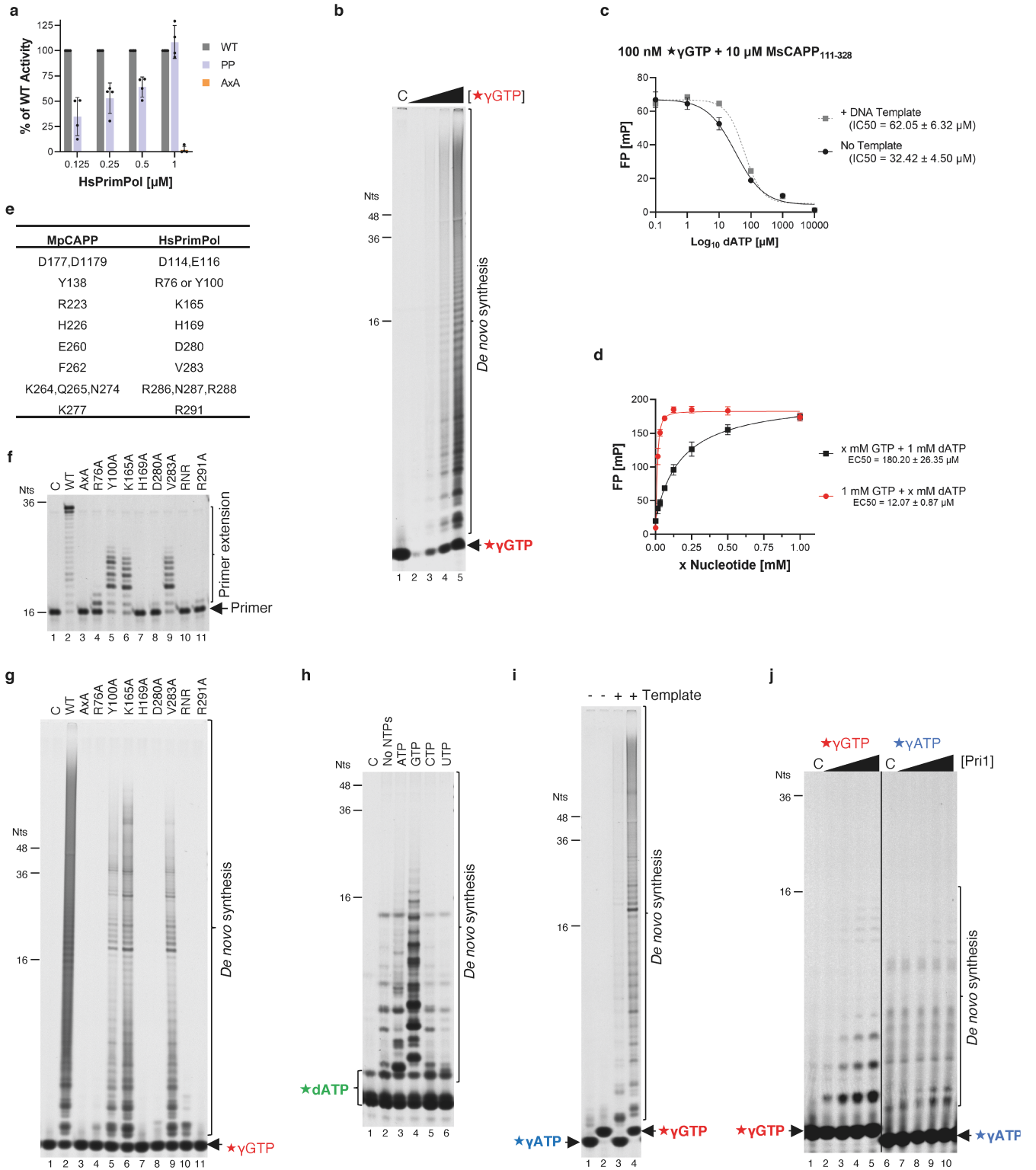


b



**Extended Data Fig. 7 | Comparison of MsCAPP and PP domains of human Prim-Pols. a.** Overlay of HsPrimPol (pink) and MsCAPP (grey) PP domains (left), and of HsPri1 (orange) and MsCAPP (grey) PP domains (right). Protein and DNA strands are displayed in cartoon representation, nucleic acids are displayed in stick representation, and metal ions are displayed as spheres.

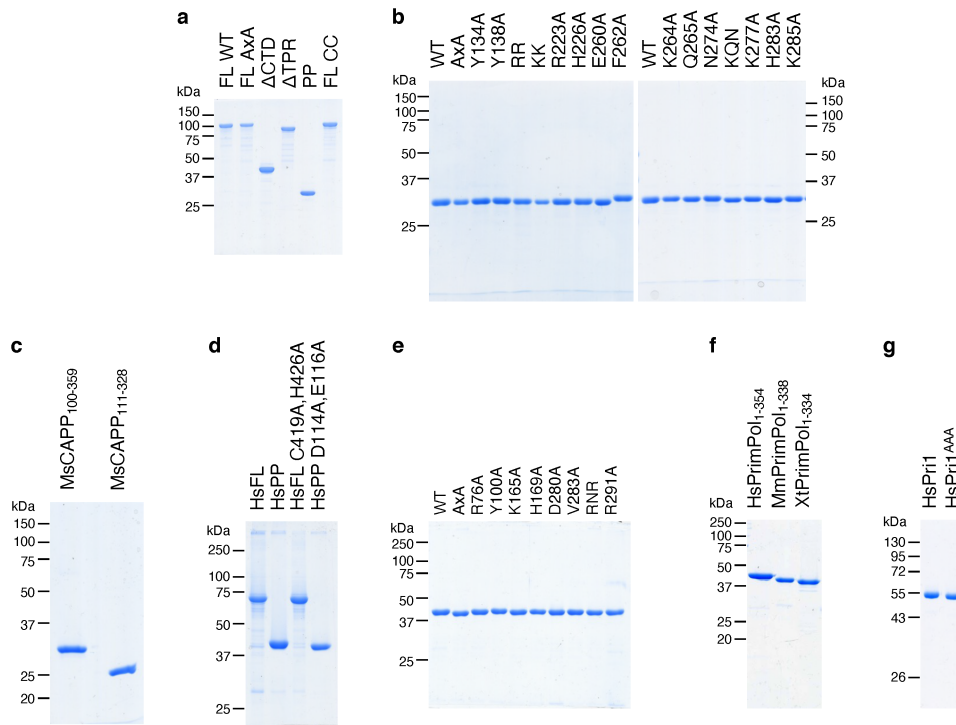
**b.** Close-up view of the E-site nucleotide, amino acid residues close to the 2' position of the ribose ring, and Region 1 of MsCAPP (left), HsPrimPol (middle) and HsPri1 (right). Interaction between D79 of HsPri1 and 2'-OH is displayed with a dashed line.



Extended Data Fig. 8 | See next page for caption.

**Extended Data Fig. 8 | Analyses of PrimPol mutations and how incoming nucleotides influence the primase activities of different Prim-Pols.** **a.** The primase activity of HsPrimPol<sub>1-354</sub> (PP) is reduced compared to the full-length enzyme (WT) at lower concentrations. HsPrimPol<sub>1-354</sub> D114A, E116A (AxA) (1 μM) exhibits no primase activity. Quantification of primase assays represented in Fig. 4c. Data represent the mean ± standard deviation from four independent experiments. **b.** Efficiency of priming by HsPrimPol<sub>1-354</sub> is dependent on FAM-γGTP (★γGTP) concentration. Primase reactions contained varying concentrations of ★γGTP (0.02, 0.1, 0.5 and 2.5 μM), 100 μM dNTP mix and 1 μM ssDNA template (oKZ388). **c.** GTP is outcompeted by high concentrations of dATP from MsCAPP's active site. FP – 10 μM MsCAPP<sub>111-328</sub> (MsPP) was added to 100 nM ★γGTP ± 1 μM DNA template (oKZ435) with increasing concentrations of dATP (0.1 μM – 10 mM). **d.** The effect of different concentration of GTP and dATP on MsPP affinity to DNA. FP – 5 μM MsCAPP<sub>111-328</sub> (MsPP) was added to 50 μM DNA template (oKZ409) in presence of 1 mM GTP/dATP and increasing concentrations (x mM) of dATP/GTP (15.625 μM – 1 mM). Data were obtained from four independent repeats (Panels c and d). Error bars in the graphs show the mean ± standard deviation. IC50 and EC50 values were calculated as described in materials and methods. **e.** Structurally equivalent residues of

MpCAPP and HsPrimPol. **f.** Polymerase activity of HsPrimPol<sub>1-354</sub> is severely disrupted by point mutations in key catalytic residues. AxA – D114A, E116A, RNR – R286A, N287R, R288A. 200 nM protein was incubated with 50 nM DNA substrate (oNB1 + oNB2) for 30 min at 37 °C. **g.** Primase activity of HsPrimPol<sub>1-354</sub> is significantly disrupted by point mutations in key catalytic residues. The reactions contained 4 μM protein, 10 μM ★γGTP, 100 μM dNTP mix and 1 μM DNA template (oKZ388). **h.** HsPrimPol preferentially utilizes GTP to initiate primer synthesis. Primase assay reaction contained 1 μM HsPrimPol<sub>1-354</sub> (HsPP), 1 μM DNA template (oKZ388), 2.5 μM FAM-dATP (★dATP), 2.5 μM dCTP, dGTP, and dTTP and 100 μM individual NTPs. **i.** HsPrimPol preferentially initiates primer synthesis with GTP over ATP. Primase assay reaction contained 1 μM HsPP, 1 μM DNA template (oKZ388), 100 μM dNTPs and 2.5 μM FAM-γATP (★γATP) or ★γGTP. **j.** Human Pri1 prefers GTP over ATP as the primer initiation base. Reactions containing 1, 2, 4 and 8 μM protein were incubated with 1 μM DNA template (oKZ388), 10 μM ★γGTP or ★γATP and 100 μM CTP, ATP and UTP or 100 μM CTP, GTP and UTP, respectively, at 25 °C for 30 min. Results shown in panel b and f; j are representative of three independent repeats. 'C' indicates control reaction without protein. Oligonucleotide (Nts) length marker is shown on the left of the gels.



**Extended Data Fig. 9 | Qualitative gel-based analysis of purified proteins.**

**a.** MpCAPP fragments and FL mutants. 1 μg of each purified MpCAPP variant was resolved on 12% SDS-PAGE and Coomassie stained. Note: FL WT, FL CC and ΔTTPR fragment are fused to MBP. FL WT – full-length wild type WT, FL AxA – full-length D177A, D179A, ΔCTD – aa1-360, ΔTTPR – aa100-546, PP – aa100-360, FL CC – full-length C462S, C464S. **b.** Mutants of MpCAPP PP domain. 1 μg of each purified MpCAPP PP mutant was resolved using 12% SDS-PAGE and Coomassie stained. AxA – D177A, D179A, RR – R142A, R143A, KK – K181A, K182A, KQN – K264A, Q265A, N274A. **c.** MsCAPP fragments. 2 μg of purified fragments

of MsCAPP<sub>100-359</sub> and MsCAPP<sub>aa111-328</sub> were resolved using 12% SDS-PAGE and Coomassie stained. **d.** HsPrimPol full-length (HsFL) and HsPrimPol<sub>1-354</sub> fragment (HsPP) and mutants. 2 μg of each purified variant was resolved using 12% SDS-PAGE and Coomassie stained. Note: AxA – D114A, E116A, RNR – R286A, N287A, R288A. **f.** Eukaryotic PrimPols. 2 μg of each purified PP was resolved using 15% SDS-PAGE and Coomassie stained. **g.** HsPriI. 1 μg of HsPriI wild-type (HsPriI) or HsPriI D109A, D111A, D306A (HsPriI<sup>AAA</sup>) was resolved using 12% SDS-PAGE and Coomassie stained.

**Extended Data Table 1 | X-Ray Data Collection and Structure Refinement Statistics**

	Apo CAPP	dGTP complex	Post-catalytic complex	Primer initiation complex	Primer initiation complex (alternate)
Deposition ID (PDB)	7NQD	7NQE	7NQF	7P9J	7QAZ
<b>Data Collection</b>					
Wavelength (Å)	0.97624	0.97625	0.91589	0.97950	0.97950
Space group	P 6 <sub>5</sub>	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Cell parameters					
a, b, c (Å)	97.4, 97.4, 109.0	40.0, 62.7, 82.5	77.0, 160.4, 39.8	227.8, 39.5, 75.0	227.9, 39.6, 79.8
α, β, γ (°)	90, 90, 120	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution limits (Å)	45.79 – 2.97	49.90 – 1.28	80.18 – 2.02	45.35 – 1.90	56.98 – 2.11
(high-resolution bin)*	(3.02 – 2.97)	(1.33 – 1.28)	(2.06 – 2.02)	(1.93 – 1.90)	(2.15 – 2.11)
Completeness (%)	100.00 (99.67)	99.20 (90.64)	99.8 (99.9)	76.4 (78.2)	94.4 (69.6)
Multiplicity	6.75 (6.88)	5.90 (3.52)	4.3 (4.4)	2.5 (2.1)	3.8 (1.9)
CC-half	0.9924 (0.3348)	0.9987 (0.2840)	0.995 (0.321)	0.997 (0.282)	0.992 (0.298)
I/sigma	8.55 (1.01)	9.58 (1.00)	5.6 (0.7)	7.7 (0.5)	6.5 (0.6)
R <sub>merge</sub>	0.2191 (1.7515)	0.0805 (1.6421)	0.159 (1.986)	0.070 (1.373)	0.151 (1.042)
<b>Refinement</b>					
Protein monomers per asymmetric unit	2	1	2	3	3
Total nonhydrogen atoms	3375	2198	4268	5813	6224
Water molecules	0	191	141	154	395
R <sub>cryst</sub> (%)	21.60	16.29	21.21	22.75	20.97
R <sub>free</sub> (%)	23.99	19.01	25.93	25.90	26.08
Ramachandran analysis (%)					
Most favoured	96.34	97.70	97.89	97.21	97.38
Outliers	0.24	0	0	0.16	0
Rmsds					
Bonds (Å)	0.003	0.006	0.008	0.009	0.007
Angles (°)	0.620	1.007	0.942	1.347	0.979
Planes (Å)	0.004	0.007	0.008	0.008	0.007
Mean atomic b value (Å <sup>2</sup> )	79.50	14.19	34.34	37.49	31.36

\* Values in parentheses are for the highest-resolution shell.

## 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 Crystal data were initially collected with GDA and auto-processed with xia2.

Data analysis Assay data processed with Prism 9 (GraphPad) and ImageJ (open source). Crystal data were processed using SHELXC/D/E, ARP/wARP, Coot, Phenix, Phaser, and visualized with Pymol (Schrödinger, LLC). Molecular modeling calculations processed using PSI4 1.2.1 and NCIPLOT 4 and outputs visualized with VMD (Visual Molecular Dynamics).

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](#)

Data Availability. The coordinates and crystallographic structure factors for MsCAPP have been deposited at the PDB under accession codes 7NQD (apo), 7NQE (dGTP-complex), 7NQF (post-catalytic complex), 7P9J (primer initiation complex) and 7QAZ (primer initiation complex (alternate model)).



## 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](https://www.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	<input type="text" value="Our study does not include population / cohort analysis."/>
Data exclusions	<input type="text" value="In one FP assay one data point was excluded after manual inspection."/>
Replication	<input type="text" value="Assays were repeated in at least three independent replicates."/>
Randomization	<input type="text" value="The experiments were not randomized."/>
Blinding	<input type="text" value="This study did not require blinding as does not include human subjects or live animals."/>

## 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	Involvement
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involvement
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging