Structural basis for the increased processivity of D-family DNA polymerases in complex with PCNA

Clément Madru, Ghislaine Henneke, Pierre Raia, Inès Hugonneau-Beaufet, Gérard Pehau-Arnaudet, Patrick England, Erik Lindahl, Marc Darue, Marta Carroni,及 Ludovic Sauguet

Replicative DNA polymerases (DNAPs) have evolved the ability to copy the genome with high processivity and fidelity. In Eukarya and Archaea, the processivity of replicative DNAPs is greatly enhanced by its binding to the proliferative cell nuclear antigen (PCNA) that encircles the DNA. We determined the cryo-EM structure of the DNA-bound PolD–PCNA complex from Pyrococcus abyssi at 3.77 Å. Using an integrative structural biology approach — combining cryo-EM, X-ray crystallography, protein–protein interaction measurements, and activity assays — we describe the molecular basis for the interaction and cooperativity between a replicative DNAP and PCNA. PolD recruits PCNA via a complex mechanism, which requires two different PIP-boxes. We infer that the second PIP-box, which is shared with the eukaryotic Polα replicative DNAP, plays a dual role in binding either PCNA or primase, and could be a master switch between an initiation and a processive phase during replication.
DNA replication is one of the most important functions in living organisms and viruses. It ensures the integrity of the genome and the accurate transfer of genetic information. DNA polymerases (DNAPs) are the key enzymes of DNA replication and diverse DNA repair processes. Cellular organisms typically use multiple DNAPs, which have been grouped into different families based on their sequence alignments: PolA, PolB, PolC, PolD, PolX, PolY, and reverse transcriptase. Genomic DNA replication is carried out by the so-called replicative DNAPs, which have evolved to copy the genome with high processivity and fidelity. The main replicative DNAPs from Eukarya are found in family B, from Bacteria in family C, and from Archaea in families B and D. Across every domain of life, polymerase holoenzyme accessory proteins play an integral role in achieving the extraordinary efficacy and accuracy of the replicative polymerase complex. These include a sliding clamp that encircles the DNA and greatly enhances the processivity. The bacterial sliding clamp is referred to as the clamp, whereas the eukaryotic and archaean sliding clamp protein is called the proliferative cell nuclear antigen (PCNA). Clamps are constructed from either two (β) or three monomers (PCNA) to yield a ring composed of six domains, which share similar protein folds.

In eukaryotes, PCNA stimulates processive DNA synthesis of both lagging and leading strands upon association with DNAPs (Pol δ and ε), respectively. PCNA inhibition is therefore considered as a valuable anticancer strategy. In Archaea, PCNA has been shown to recruit replicative DNAPs of both B- and D-families, respectively, named PolB and PolD. Organisms within the archaean domain of life possess a simplified version of the eukaryotic DNA replication machinery. The archaean PCNA shares 25% identity with the human PCNA and PolD, despite having the two-barrel fold of multi-subunit RNA polymerases for its catalytic domain, shares intriguing similarities with the three main multi-subunit eukaryotic replicative DNAPs: Polα, Polβ, and Polε. In particular, the PolD DP1 subunit and the C-terminal domain of the DP2 subunit are homologous to the regulatory B-subunit and the C-terminal domain of the catalytic A-subunit, which are found in all eukaryotic replicative DNAPs. PoID is an archaean replicative DNAP, which is widely distributed among Archaea (except in crenarchaea) and has been shown to be essential for cell viability. PoID is strongly stimulated through its interaction with PCNA. PCNA-binding partners carry short motifs known as the PCNA-interacting protein box (PIP-box), but sequence divergent motifs have been reported to bind to the same binding pocket. Although the PIP-boxes are the best known PCNA-interacting peptides, other motifs including RIR and MIP motifs have been reported. Since the first structures of sliding clamps were determined, about 100 structures have been reported, in their apo form, bound with DNA, or in complex with various PIP-boxes and other PCNA-interacting motifs. Only two structures of full-length replicative DNAPs bound with PCNA and DNA have been reported to date: the Pyrococcus furiosus PoID–PCNA-DNA ternary complex that was determined by negative-staining electron microscopy and cryo electron microscopy of the Human Polδ–PCNA holoenzyme, which was published while this manuscript was under revision.

Here we present the cryo-EM structure of the DNA-bound PoID–PCNA complex from Pyrococcus abyssi at 3.77 Å using an integrative structural biology approach, combining cryo-EM, X-ray crystallography, protein–protein interaction measurements, and activity assays. This structure unveils the molecular basis for the interaction and cooperativity between the whole replicative polymerase and PCNA with an unprecedented level of detail. PoID recruits PCNA via a complex mechanism, which requires two different PIP-box motifs: a C-terminal and an internal one that has never been characterized so far. We infer that the C-terminal PIP-box, which is shared with the eukaryotic Polα replicative DNAP, plays a dual role in binding either PCNA or primase, and could be a master switch between an initiation phase and a processive phase during replication.

Results
Architecture of the DNA-bound PoID–PCNA processive complex. The P. abyssi PoID processive complex was reconstructed by incubating PCNA with the PoID exonuclease-deficient variant (DP1 H451A) in a 3:1 ratio, in the presence of an 18-mer primed DNA duplex with a 7-nucleotide overhang and a non-hydrolyzable nucleotide analog. The reconstituted complex (317 kDa) was vitrified and its structure was determined using single-particle cryo-EM. The map was solved at an average resolution of 3.77 Å (Table 1 and Supplementary Figs. 1 and 2). The essential PoID and PCNA DNA-binding regions, as well as the DP1–DP2 and DP2–PCNA interface regions showed a higher resolution map at 3.0–3.5 Å (Fig. 1a and Supplementary Fig. 2). In these regions, the density map of the DNA-bound PoID–PCNA complex was of sufficient quality to allow de novo building of the majority of the protein. The map includes several regions for which no atomic model was known before, such as regions neighboring the active site and the DP1–DP2 interface. In the peripheral region of the complex, the DP2 KH domain, the DP1 OB domain, and some regions of the PCNA were found to be flexible and the local resolution map ranged between 4.0 and 6.0 Å (Supplementary Fig. 2). In these regions, crystal structures of PoID DP1 (144–619) and DP2 (1–1050) individual subunits, and the structure of the P. abyssi PCNA (from this study using X-ray crystallography at 2.3 Å resolution) were used in model building. DNA was docked into the cryo-EM map, guided by the density for the duplex region showing minor and major grooves, as well as the unambiguous position of purines and pyrimidines (Supplementary Fig. 3). However, no obvious density for single-stranded DNA and the incoming nucleotide was observed in the DP2 active site.

A defining feature of the PoID–PCNA–DNA ternary complex is its compactness (Supplementary Movie 1): the radius of the PCNA ring and the clamp-like PoID DNA-binding domain match perfectly (Fig. 1a). The structure of PCNA in the complex is not distorted compared with the structure of free PCNA and we conclude that the cryo-EM structure represents a stable interaction of DNA-bound PoID with a closed PCNA clamp. Away from the active site, PCNA surrounds one helix turn of the nascent DNA duplex, which is located at the center of the PCNA ring. The nascent DNA duplex runs straight through PCNA and adopts an almost perpendicular orientation (~80°) with respect to the DNA (Fig. 1b). Interaction with PCNA nearly doubles the positively charged surface formed by the PoID active site, making a 60 Å-long DNA-binding site (Fig. 1c). This structure thus rationalizes how PCNA enhances the processivity of PoID interaction with PCNA perpetuates the interactions with the nascent DNA duplex when it exits the PoID clamp, thereby preventing the polymerase from falling off prematurely.

PoID–DNA complex. The primer-template is held in position in the PoID active site by a bipartite clamp domain. Clamp-1 and clamp-2 domains contribute a central cleft located upstream from the DP2 catalytic center, bordered with positively charged side chains, which encircles one helix turn of the nascent DNA (Fig. 2a). Although the DNA bound by the PoID–PCNA complex is predominantly in the B-form, interaction with the PoID clamp
A clamp-1 domain contains a Zn-binding module, showing a decreased helical twist and a widened minor groove on the opposite side, the clamp-2 domain binds to the minor groove whereas K1125 and K1145 contact the template strand. On the other side, the clamp-2 domain binds to the minor groove and to participate in the catalytic efficiency, which is essential for the DNA interactions and base selectivity. Similarly, the structure of PolD shows two canonical residues K1157 and Y1158 pointing towards universal hydrogen bond acceptors at purines N3 and pyrimidines O2 positions, which may be important for the catalytic efficiency and fidelity of PolD.

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

| PolD–PCNA-DNA complex (EMDB-10401) (PDB 6TH8) |
|-----------------------------------------------|
| Data collection and processing                 |
| Magnification                                  | 165,000 |
| Voltage (kV)                                   | 300     |
| Electron exposure (e−/Å²)                      | 40      |
| Defocus range (µm)                             | −0.5 to −3.5 |
| Pixel size (Å)                                 | 0.83    |
| Symmetry imposed                              | C1      |
| Initial particle images (no.)                  | 269,251 |
| Final particle images (no.)                    | 147,511 |
| Map resolution (Å)                             | 3.77    |
| FSC threshold                                  | 0.143   |
| Map resolution range (Å)                       | 3 to 6  |
| Refinement                                     |         |
| Initial model used (PDB code)                  | 5IJL, SIHE, 6T7X |
| Model resolution (Å)                           | 3.77    |
| FSC threshold                                  | 0.143   |
| Model resolution range (Å)                     | 3 to 6  |
| Map sharpening B factor (Å²)                    | 145.192 |
| Model composition                              |         |
| Non-hydrogen atoms                             | 19,398  |
| Protein residues                               | 2347    |
| Ligands                                        | 5       |
| B factors (Å²)                                 |         |
| Protein                                       | 81.31   |
| Ligand                                        | 131.54  |
| R.m.s. deviations                              |         |
| Bond lengths (Å)                               | 0.012   |
| Bond angles (°)                                | 1.258   |
| Validation                                     |         |
| MolProbity score                               | 2.58    |
| Clashscore                                     | 38.66   |
| Poor rotamers (%)                              | 0.94    |
| Ramachandran plot                              |         |
| Favored (%)                                    | 91.21   |
| Allowed (%)                                    | 8.66    |
| Disallowed (%)                                 | 0.13    |

PCNA–DNA binding interactions. Interestingly, PCNA and PolD adopt strikingly different DNA-binding modes. Although PolD extensively interacts with the nascent DNA duplex, the PCNA channel exposes several positively charged residues, which are directed towards the DNA backbone and make labile transient polar contacts with DNA. Consistently, unlike the DNA region present in the PolD active site, the structure of DNA does not appear to be influenced by its interaction with PCNA and adopts a perfect B-form DNA architecture (Fig. 2b). The local resolution of the outer perimeter of PCNA is about 4.0–6.0 Å, substantially lower than the average resolution of the consensus map (3.77 Å), indicating a greater flexibility in the PCNA–DNA interactions compared with the PolD–DNA interactions. To characterize further the molecular determinants of the PCNA sliding movement, we performed a focused three-dimensional (3D) classification on PCNA and identified three classes showing extra densities between DNA and the PCNA residues K84 and K86 (Fig. 2c). Interestingly, DNA was found to be in contact with different monomers of PCNA in the three 3D classes, suggesting that all three PCNA monomers contribute to DNA binding through short-lived polar contacts. Such labile transient interactions between DNA and PCNA have already been observed in another study using an integrative structural biology approach combining nuclear magnetic resonance and molecular dynamics simulations. Thus, PCNA provides an electrostatic cushion for the DNA to pass through as it leaves the PolD active site, thereby allowing it to rapidly slide onto DNA, pulled by PolD.

PolD–PCNA interface. The cryo-EM map shows how PCNA is tethered to PolD through multiple contacts that involve both clamp-1 and clamp-2 domains of DP2 (Fig. 3a). First, the α-helices, α-40, and α-41 of the clamp-1 domain are connected by a loop of 17 amino acids, which is hooked into PCNA. This loop binds to the canonical PCNA PIP-binding pocket through an internal PIP-box (iPIP), which has never been identified so far (Fig. 3b). Six residues within this iPIP fill the PCNA PIP-binding pocket. Among them, the side chain of Q1198 penetrates deep into PCNA, making contacts with the canonical PCNA residue P245 (Fig. 3c). In addition, the bulky side chains of L1199, L1201, and L1202 make extensive contacts with hydrophobic residues, which line the PCNA PIP-binding pocket. Interestingly, the PolD iPIP shows a non-canonical structure compared with other PIP-boxes, lacking a 4-residue 310-helix turn, which has been observed in the structures of all PIP–PCNA complexes determined so far. Second, the N-terminal region of the interdomain-connecting loop of the adjacent PCNA monomer binds to the DP2 clamp-2 domain (Fig. 3a). This interaction is mediated through polar contacts between residues E692, K727, Y781, and K896 from clamp-2 and PCNA residues H75, D117, and E119 (Supplementary Fig. 4). This PCNA–clamp-2 interaction, which includes a large proportion of polar contacts, contrasts with the PCNA–iPIP site-specific interaction described above. The PCNA–clamp-2 interaction may thus be easily broken or profoundly remodeled, enabling the PCNA ring to form a new interface with PolD, when the polymerase encounters a damage and adopts an editing mode. Such a conformational transition has been proposed for the PolB–PCNA complex, when PolB switches from the polymerase state to the proofreading state.
**Fig. 1** Cryo-EM structure of the DNA-bound PolD–PCNA processive complex.  
**a** Two orthogonal views of the cryo-EM density map (left) and cartoon representations (right) of the DNA-bound PolD–PCNA complex.  
**b** Orientation of the DNA duplex with respect to the PCNA threefold symmetry axis.  
**c** Cutaway front view of the PolD–PCNA–DNA complex showing the electrostatic surface potential with negative, neutral, and positive charges represented in red, white, and blue, respectively.

**Fig. 2** Structural basis for DNA binding by the PolD–PCNA complex.  
**a** View of the DP2 clamp-1 (light blue) and clamp-2 (dark blue) domains surrounding one turn of the DNA duplex (orange). The two catalytic double-psi β-barrels DPBB-1 and DPBB-2 are represented in red and yellow, respectively. Template and primer strands are indicated by T and P, respectively.  
**b** Interaction with the PolD clamp causes a local distortion of the DNA (orange) compared with an ideal B-form DNA (white). The phosphates of the DNA duplexes are shown as spheres.  
**c** PCNA makes labile contacts with DNA. Focused classification on PCNA resulted in three 3D classes showing extra-densities between DNA and PCNA residues K84 and K86, which are underlined by arrows.
PolD uses two distinct PIP-boxes for recognition of PCNA. In addition to the iPIP that was identified from our cryo-EM structure, the DP2 subunits of PolD from *P. furiosus* and *P. abyssi* have been shown to host a C-terminal PIP-box. This second PIP-box (hereafter referred to as canonical PIP-box (cPIP)), is connected to DP2 by a 40-residue linker, which is variable in both length and amino-acid composition across archaea. Strikingly, both cPIP and the linker are not visible in the cryo-EM density of the DNA-bound PolD–PCNA complex. To better characterize the role of cPIP, we co-crystallized PCNA from *P. abyssi* with a 12 amino-acid peptide mimicking the DP2 cPIP and solved its structure at 2.7 Å resolution. The final model includes 9 of the 12 amino acids of the co-crystallized peptide. In contrast to the structure of iPIP, which differs from other structures of PIP-boxes, cPIP shares the same overall fold. Sequence similarities are highlighted with light blue boxes and conserved residues are highlighted with dark blue boxes. 

**Fig. 3 PolD uses two distinct PIP-boxes for molecular recognition of PCNA.** a Two views of the cryo-EM map showing the interfacial region between the DP2 clamp-1 (left) and clamp-2 (right) binding to PCNA. DP2 clamp-1 and clamp-2 domains are represented in light and dark blue, respectively; DP1 is shown in yellow; DNA in orange; PCNA in green. b Multiple-sequence alignment of the C-terminal region in Thermococcus species: *P. abyssi* (P.ab), *Thermococcus barophilus* (T.ba), *Thermococcus litoralis* (T.li), *Thermococcus thioreducens* (T.re), *Thermococcus kodakarencis* (T.ko), and *Thermococcus nautili* (T.na). Internal (iPIP) and C-terminal (cPIP) PIP-boxes are framed in blue and purple, respectively, with secondary structure elements shown above. Sequence similarities are highlighted with light blue boxes and conserved residues are highlighted with dark blue boxes. c Detailed view of the iPIP–PCNA interaction in the cryo-EM map contoured at a level of 7σ. d Detailed view of the cPIP–PCNA interaction in the 2Fo-Fc X-ray electron density map contoured at a level of 1.5σ.
cPIP to PCNA are mutually exclusive. One may ask whether cPIP could not bind to one of the neighboring PCNA subunits, but no extra density that could be accounted for by cPIP was found in the cryo-EM map. Furthermore, the linker region connecting cPIP to DP2 can hardly cover the 70 Å distance, which separates the last defined residue of DP2 from the nearest unoccupied PIP-binding pocket. Altogether, these observations strongly suggest that iPIP and cPIP interact with PCNA through different mechanisms.

To assess the role of these two PIP-boxes, primer extension reactions were carried out in the presence of a primer-template of large size and increasing amounts of PCNA, to stimulate full-length DNA synthesis (Fig. 4a, b). Using primed-M13mp18 template (7 nM), hybridized to a fluorescent-labeled primer. Reactions contained 25 nM of DNA, in the presence or absence of PCNA (75, 150, and 300 nM). a Quantification of the stimulation by PCNA of PolD wild type, PolDΔcPIP, and PolDΔiPIP-ΔcPIP. Histograms show full-length 7249 bp (%) for extension reaction in the presence of PCNA relative to full-length 7249 bp (%) without PCNA for each PolD. PolD wild type (n = 3), PolDΔcPIP (n = 3), and PolDΔiPIP-ΔcPIP (n = 3). Error bars represent 1SD. c Specific binding of immobilized-DP2(1196-1270) to PCNA measured by biosensor interferometry (BLI). Steady-state analysis was performed using the average signal measured at the end of the association step (between 290 and 300 s). d Comparative binding of immobilized-DP2(1196-1270) and -DP2(1196-1253) to PCNA measured by BLI. The range of concentrations used in the binding experiments are listed in the Methods section. e Hypothetical two-steps mechanism for PCNA recruitment by PolD. Source data are provided as a Source Data file.
cPIP is a dual PCNA/primase-binding peptide shared with Polα. PolD shares, with its eukaryotic counterparts, unifying features of their subunit organization that reveal a clear evolutionary relationship. The eukaryotic replicative DNAPs Polα, Polδ, and Polε possess a catalytic subunit, often referred to as the A-subunit, constituvely associated with different cohorts of regulatory proteins among which B-subunits are present in all three DNAP assemblies. Both OB and PDE domains of DP1 share a remarkable degree of 3D structural similarity with the regulatory B-subunits of all eukaryotic replicative DNAPs. In addition, the C-terminal region of their catalytic subunits, which is dedicated to interaction with their B-subunits, resembles the C-terminal region of DP2 that is required for interaction with DP1 (ref.16). In addition to their structural similarities (Fig. 5d), PolD shares common functional features with Polα, which is tightly associated with the DNA primase in a complex called primosome that is required for initiating DNA replication in eukaryotic cells. Similarly, PolD has been shown to interact with the DNA primase and is able to extend RNA primers, suggesting that PolD is required for initiating DNA replication in archaea. Previously, a short conserved motif located at the extreme C terminus of Polα was shown to be critical for the interaction with the primase. We tested whether the C-terminal region of PolD, which is homologous to that of Polα (Fig. 5e), could host a similar primase-interacting peptide.

To assess the role of the C terminus of PolD in the interactions with PCNA and primase, we performed biolayer interferometry (BLI) experiments using His$_{6}$-tagged maltose-binding protein (MBP) fusions of the C-terminal region of DP2, which were captured via surface-linked Ni-NTA. As expected, the MBP–iPIP–cPIP fusion (DP2:1196–1270) was found to readily bind to PCNA, with a $K_D$ of 472 ± 120 nM (Fig. 4c). Interestingly, the same construct was also able to interact with primase, with a measured $K_D$ of 237 ± 22 nM (Fig. 5a), which is very similar to the $K_D$ of 245 nM that was reported for the interaction within the C terminus of yeast Polα and the primase. Deleting cPIP in the MBP–iPIP–ΔcPIP fusion (DP2:1196–1253) strongly impaired binding to PCNA and abrogated binding to the DNA primase (Figs. 4d and 5b), showing that cPIP is a dual PCNA/primase-binding peptide. Using His$_{6}$-tagged PCNA and His$_{6}$-tagged primase captured via surface-linked Ni-NTA, we then performed surface plasmon experiments, allowing us to measure the ability of PCNA and primase to bind a synthetic peptide encompassing the cPIP. The cPIP binds to PCNA with a lower affinity ($K_D$ of 49 ± 4 μM), which differs by two orders of magnitude from the one observed for the MBP–iPIP–cPIP fusion (Supplementary Fig. 7), suggesting that the primase-binding peptide binds onto a protein–protein interface. The ability of cPIP to recruit both PCNA and primase is consistent with the dual role of PolD in DNA replication initiation and elongation, which requires interaction with both partners.

Interestingly, comparing the structures of the cPIP of PolD and the primase-binding motif of Polα reveals that both peptides fold in a one-turn $3_{10}$ helix (Fig. 5f). This structural similarity is underpinned by the conservation of four hydrophobic and aromatic residues. The side chains of these four conserved hydrophobic residues become buried at the hydrophobic protein–peptide interface in both PolD–PCNA (Fig. 3d) and Polα–primase. Moreover, hotspot residues that were shown to be important for primase binding such as L1451 and F1455, are particularly well-conserved in archaea (Fig. 5f). In archaea and eukaryotes, the primase forms a heterodimer composed of a small PriS subunit with the polymerase activity and a larger regulatory PriL subunit. The Pol primase-binding peptide binds onto a hydrophobic edge of the PriL subunit. Interestingly, in archaeal PriL, the exposed hydrophobic surface is buried by a short α-helix, which fills the space occupied by the DNAP in the Pol–primase complex. This suggests that either cPIP binds to another site in the archaeal primase or that the corresponding region is remodeled upon cPIP binding. Such differences in the mode of interaction between the primase and these two polymerases may be accounted for by the fact that although eukaryotic Polα forms a stable and constitutive complex with primase, PolD and primase form only a transient complex.

**Discussion**

In contrast with cellular transcriptases and ribosomes, which evolve by accretion of complexity from a conserved catalytic core, it is striking that DNA replication was reinvented several times during evolution and that no replicative DNAP family is universally conserved. Bacteria, Archaea, and Eukarya have evolved three distinct protein folds to replicate their genomes as follows: (i) the Polβ-like fold—found in bacterial Pol-III, (ii) the Klenow-like fold—found both in archaeal/eukaryotic B-family DNAPs and in bacterial Pol-1, and (iii) the two-barrel fold—the third structural class of DNAPs that was recently unveiled with the structure of archaeal PolD. Although structurally divergent, all replicative DNAPs share unifying features. Hence, across every domain of life, the extraordinary efficacy of the replicative polymerase complex is dependent on their interaction with sliding clamps, which encircle DNA and greatly enhance their processivity.

The structure of the DNA-bound PolD–PCNA complex from *P. abyssi* unveils the molecular basis for the interaction and cooperativity between the PolD replicative DNAP and PCNA. Away from the PolD active site, PCNA surrounds one helix turn of the nascent DNA duplex, perpetuating the interactions with the nascent DNA duplex, thereby preventing the polymerase from falling-off prematurely. Although PolD makes extensive and strong contacts with the DNA minor groove, PCNA contributes to DNA binding through short-lived polar contacts, which provides an electrostatic cushion for the DNA to pass through as it leaves the PolD active site, thereby allowing PCNA to rapidly slide onto DNA, pulled by PolD. Despite belonging to structurally distinct classes of DNAPs, the archaeal PolD–PCNA complex (a two-barrel fold DNAP), the bacterial PolIII–clamp–exonuclease–τ complex (a Polβ-like fold DNAP), and the eukaryotic Polδ–PCNA holoenzyme (a Klenow-like fold DNAP) share intriguing structural features. In all three structures, the nascent DNA duplex runs straight through PCNA and adopts an almost perpendicular orientation with respect to the DNA. This view contrasts with the highly tilted double-stranded DNA in the crystal structure of the β-clamp DNA complex (~22°) and in the PCNA–DNA complex (~40°), suggesting that DNA binding by the clamp is versatile and strongly influenced by the polymerase.

PIP-boxes often exist in multiple copies in DNAPs. Polh and Pol6 have three PIP-boxes, which contribute differentially to distinct biological functions. We have shown that PolD uses two distinct PIP-boxes for molecular recognition of PCNA, which are located in the C-terminal region of their DP2 subunit. Strikingly, these two PIP-boxes contribute differentially to PCNA recruitment. Based on an integrative structural biology approach, combining cryo-EM, X-ray crystallography, protein–protein interaction measurements, and activity assays, we hypothesize that PolD may be recruited by PCNA through a two-step mechanism (Fig. 3e). First, PCNA is recruited by PolD through its interaction with the DP2 cPIP. Once the PolD–PCNA complex is
loaded on DNA, the complex is stabilized by an interaction between PCNA and iPIP, as observed in the cryo-EM structure, whereas cPIP becomes dispensable. This mechanism is supported by a former study on PolD from *P. abyssi*, showing that removing the cPIP did not disrupt the physical interaction with PCNA, when both partners are bound to DNA.

In addition, we have shown that cPIP is not only important for DNA repair protein interactions showing that PIP-boxes are a much more conserved domain in archaeal PolD and eukaryotic Polα than previously thought. This finding expands current views on DNA primase polymerase. Recent gene deletion studies on hyperthermophilic Euryarchaeota have demonstrated that only PolD is required for viability, suggesting that PolD is solely responsible for DNA replication, whereas PolB may be required for DNA repair. It is noteworthy that the situation is different in Crenarchaeota, which do not possess PolD. Due to the multiple biological reactions required during DNA replication, PolD must be able to switch from one replication factor to another in a spatially and temporally regulated process. Indeed, our work shows that cPIP has overlapping specificities and is capable of binding both PCNA and primase. Hence, PolD must be able to interact with the primase during the initiation of DNA replication and with PCNA to ensure processive extension of both leading and lagging strands. The versatility of cPIP may be instrumental in such process. This finding expands current views on PCNA interactions showing that PIP-boxes are a much more conserved domain in archaeal PolD and eukaryotic Polα than previously thought.
broader class of motifs than initially thought, which form the network of interacting proteins responsible for DNA replication and repair. In eukaryotes, chromosomal replication is accomplished primarily by three distinct DNAPs, which play different roles in DNA replication: Polα, Polδ, and Polε. Polα is tightly associated with the primase in a constitutive complex, named the primosome, which is responsible for initiating DNA replication. Polδ and Polε have been shown, in a series of experiments, to be responsible for lagging and leading strand replication, respectively. Although they have diverged to acquire specific biological activities, all these three polymerases share unifying structural features that they most probably inherited from a common ancestor with PolD (Fig. 6). Both OB and PDE domains of DP1 share a remarkable degree of 3D structural similarity with the regulatory B-subunits of all eukaryotic replicative DNAPs. In addition, the C-terminal region of their catalytic subunits, which is dedicated to the two-barrel D-family catalytic core found in PolD, was conserved in Polδ, PCNA DNA-bound complex (this study). No structures are currently available for the PCNA–Polδ complex, but its schematic representation is supported by the literature.

**Methods**

**Cloning, protein expression, and purification.** PolD from P. abyssi was co-expressed by 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) induction in *Escherichia coli* strains BL21 (DE3) grown overnight in Lysogeny broth (LB) at 20 °C and copuriﬁed with Ni-NTA and heparin chromatography (GE Healthcare), followed by tobacco etch virus (TEV) cleavage and size-exclusion chromatography. The purified PolD was concentrated to 2 mg/ml in 20 mM Tris HCl (pH 8), 200 mM NaCl, 3 mM MgCl₂ storage buffer. The DNA coding sequence of DP2 (1196–1253) and DP2(1196–1270) were inserted in a pVEX His-MBP-TEV plasmid allowing the expression of a TEV-degradable N-terminal His₆-MBP tag. The open reading frame (ORF) of the PCNA gene from *P. abyssi* was optimized and synthesized by GeneArt (Thermo Fisher) and inserted into pRSETa(+) as a polyclinogenic N-terminal 6 × His tag. The ORFs of the PriS and PriL (1–210) genes from *P. abyssi* were also optimized and synthesized commercially by GeneArt (Thermo Fisher) and inserted into pRSETa(+) as a polyclinogenic construct with a TEV-degradable N-terminal His₆-tagged PriS fusion protein. Production and purification of PolD constructions, PCNA, and PriS-PriL(1–210) were performed as follows: proteins were expressed in BL21–CodonPlus (DE3)–RIPL strain from *E. coli* (Agilent) at 37 °C in LB medium supplemented with 100 μg ml⁻¹ of antibiotic (kanamycin or ampicillin, depending on the plasmid) and 25 μg ml⁻¹ chloramphenicol. Recombinant protein expression was induced by adding 1 mM IPTG. Cells were then incubated overnight at 20 °C, collected by centrifugation, resuspended in buffer A (50 mM Na·HEPES at pH 8, 500 mM NaCl, 0.5 M imidazole). The protein fractions were then combined, dialyzed in buffer C (20 mM Na·HEPES pH 8, 0.1 M NaCl) and mixed with 5 mM Heparin Hitrap columns (GE Healthcare) and eluted with a linear gradient, by mixing buffer C with buffer D (20 mM Na·HEPES pH 8, 1 M NaCl). Purifications were finally polished using exclusion-size chromatography in buffer E (20 mM Na·HEPES pH 8, 0.15 M NaCl) on a Superdex 75 10/300 or Superdex 200 10/300 (GE Healthcare) depending on the purified protein molecular weight.

**Sample preparation for Cryo-EM.** The DNA duplex was purified by mixing equivalent molar amounts of primer (5′-GGGCGGCGGAGCCTGTCGCTG-3′) and template (5′-ACTGCTGTAGGCTGCGGCGCAAGGCGCGGG-3′). The mix was then heated 3 min at 90 °C and slowly cooled to room temperature. PolD–PCNA–DNA complexes were associated by mixing 0.4 mM PolD and 1.2 mM PCNA with 0.7 μM DNA in the presence of 0.1 mM DAMPCPP in buffer F (20 mM Na·HEPES pH 8, 0.01 mM NaCl, 2 mM Mg·Acetate). The mixture was incubated for 20 min at 4 °C and pipetted onto glow-discharged holey carbon cryo-EM grids (C-flat 2/2, 4Cu, 2/2, 4/1) for quick drying.
50. Grids were frozen in liquid ethane by using a Vitrobot Mark IV (ThermoFisher) at 100% humidity, 22 °C temperature, blotting force 20, and blotting time of 4 s.

Cryo-EM data acquisition and image processing. Movies were collected using EPU software on a Titan Krios electron microscope (Thermo Fisher Scientific) operated at 300 kV, on a GatanK2 Summit direct electron detector coupled with a Bioquantum energy filter with 20 eV slit. The defocus range was between −0.5 and −3.5 μm, the pixel size was 0.83 Å/pixel and the total dose was ~40 electrons/Å² distributed into 40 frames. Image pre-processing until two-dimensional (2D) classification was performed during data acquisition using the Scipion package. Images were imported and movie frames were aligned using MotionCor2 with dose compensation applied. The contrast transfer function (CTF) was estimated with Gctf and particles were automatically picked using the Xmipp supervised picker after training was provided to tell particle from no-particles over ~1000 manually picked molecules. Particles were extracted 2x binned into a 150 × 150 pixel box to perform 2D classification for cleaning on-the-fly on consecutive batches of 20,000 particles each. Around 700,000 particles were automatically picked from 4602 micrographs and after 2D classification around 270,000 particles were kept, after excluding those belonging to poorly-resolved 2D classes.

Three initial models were generated by ab initio reconstruction using stochastic gradient descent in Relion 3.0. At this point, good particles were again extracted without building 300 × 300 pixel box size and 3D refinement was run using one of the initial models. After 3D refinement, a 3D classification with 3 classes and local search was performed, to separate possible conformations. The classification resulted in two classes displaying sharp details and a third class with less-defined features. No obvious conformational differences were noticed and the 3D classification was combined and CTF refinement was confined to a final envelope of 3.77 Å (gold standard 0.143 Fourier shell correlation (FSC) criterion) resolution from around 150,000 particles. CTF-refinement and Bayesian polishing were also performed, but they did not improve the overall resolution nor the quality of the map.

To separate different conformational states, focused 3D classification upon signal subtraction was performed using masks to focus on the PCNA part of the complex, followed by another 2D classification of 20,000 particles each. Around 700,000 particles were automatically picked from 4602 micrographs and after 2D classification around 270,000 particles were kept, after excluding those belonging to poorly-resolved 2D classes.

Building and refinement of cryo-EM model. The density map of the DNA-bound Polδ–PCNA complex was of sufficient quality to allow de novo building of the majority of the protein in COOT. In the peripheral region of the complex, the DP2 KH domain, the DP1 OB domain and some regions of the PCNA were found to be more flexible and the local resolution map ranged between 4.0–4.5 Å. In these regions, model building was guided by the crystal structures of Polδ DP1 (144–619) and DP2 (1–1050) individual subunits, and the structure of the P. abyssi PCNA, which was solved in this study using X-ray crystallography at 2.3 Å resolution (PDB IDs 5IJL, 5HE and 6T7X). Concerning the DNA building, an ideal B-form DNA duplex was docked in the density as a starting point. The initial model was then subjected to global real-space refinement using the PHENIX suite using secondary structure restraints. The refined model was further manually inspected and adjusted in COOT. The final model was validated with statistics from Ramachandran plots and MolProbity scores. All figures were prepared using UCSF Chimera and UCSF Chimera X.

Crystallization, X-ray data collection, and processing. PCNA crystallization trials were performed at 18 °C using the hanging drop vapor diffusion technique in 2 μl drops (1:1 reservoir to protein ratio) equilibrated against 500 μl of reservoir solution. For the PCNA–cPIP complex, PCNA was pre-incubated 30 min with a twofold molar excess of cPIP (KRRVILEEFS) (Smart Bioscience) in buffer E prior to crystallization trials. PCNA crystals were obtained in 20% PEG 400, 0.2 M CaCl₂, and 0.1 M MES pH 5.5 with a PCNA solution at 5 mg/ml, whereas PCNA–cPIP complex crystals were obtained in 30% PEG 400, 0.2 M MgCl₂, and 0.1 M Bis-Tris pH 7.1 with a PCNA–cPIP complex solution at 10 mg/ml. The crystals were cryoprotected by soaking in a 1:1 paraffin : paratone oil mix.

X-ray data were collected at the European Synchrotron Radiation Facility on beamlines ID23 and ID29 and at theSOLEIL synchrotron on beamlines PX1 and PX2. Data sets were indexed using XDS, scaled and merged with Aimless (from the CCP4 suite) (Crystallization Computational Protocol 1994) and for anisotropy with the STARANISO server (staraniso.globalphasing.org). PCNA X-ray structure was solved by molecular replacement using the structure of PCNA from P. furiosus (PDB ID: 5AUJ). Molecular replacements were carried out with the Phaser program from Phenix and subsequent rebuilding and refinement were achieved with COOT and BUSTER. Coordination and structure factors for the PCNA and cPIP-bound PCNA structures from P. abyssi were deposited in the Protein Data Bank under accession codes 6T7X and 6T7Y, respectively.

BLI assays. BLI experiments were performed on an Octet RED968 instrument (ForteBio). His-MBP-fused DP2 constructs were captured at a 1.5 nm density on Ni-NTA biosensors. Binding interactions were monitored for 300 s at 25 °C in buffer E supplemented with 0.2 mg/mL bovine serum albumin. Seven proteins concentrations were assayed (31.25, 62.5, 125, 250, 1000, and 2000 nM) and a buffer-only reference was subtracted from all curves. Affinities were determined by fitting the concentration dependence of the experimental steady-state signals, using the Octet RED data analysis v11 software (ForteBio).

Surface plasmon resonance assays. Surface plasmon resonance experiments were performed using a Biacore T200 instrument (GE Healthcare). All measurements were performed at 25 °C in buffer E supplemented with 100 μM EDTA. A series S sensor chip NTa (GE Healthcare) was used to immobilize ~2000 RU of His-tagged PCNA and PriS-PriL(1–210) on two of the four flowcells, and His-tagged MBP on a third as a reference. Ten concentrations (0, 0.75, 1.5, 3, 6, 12, 25, 50, 100, and 200 μM) of cPIP-peptide (KRRVILEEFS) (Smart Bioscience) were injected for 30 s at 30 μl/min over the three flowcells. The raw sensograms were processed by subtracting both the signals measured on the reference flowcell and the signals measured for blank injections. Corrected data were analyzed with the BIACalculator evaluation software (GE Healthcare), by fitting the concentration dependence of the experimental steady-state signals.

Primer extension assays. Extension reactions of the fluorescent-labeled 32-mer primer 5′-Cy5-TGCGAACATGTCATGCCTGACGTGACACTCTA-3′ annealed to a single-stranded circular M13mp18 template were performed in 12.5 μl of 50 mM Tris (pH 8.0), 1 mM dithiothreitol, 50 mM NaCl, 5 mM MgCl₂, and 200 μM each of dNTPs in the presence or absence of PCNA at the indicated concentrations. DNA polymerization was initiated by addition of DNAPs (25 nM) and was conducted at 60 °C for 30 min. Reactions were quenched on ice by addition of one volume of 90% deionized formamide and 20 μl EDTA, before heating at 95 °C for 5 min. Products were resolved on 1% (w/v) alkaline agarose gel. DNA markers (2-Log DNA ladders, New England Biolabs) were loading in 45% acrylamide 10% glycerol). Gels were stained with SYBR Gold (Invitrogen). Gels were first scanned with a Mode Imager Typhoon 9500 (GE Healthcare) and Cy5 was visualized using the Cy5-labeled products and then scanned for SYBR Gold for detecting the ladders. Full-length 7249 bp (%) corresponds to the intensity of 7249 bp bands as a percentage of total lane intensity. In all cases, the background value was subtracted. The mean of percentage ± SD of full-length 7249 bp (%) from three independent experiments were obtained (the raw data are provided in Supplementary Table 2).

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

Data availability Coordinates and structure factors for the PCNA and the PCNA–cPIP complex were deposited in the Protein Data Bank under the accession codes PDB 6T7X and PDB 6T7Y, respectively. The cryo-EM map of the Polδ–PCNA–DNA ternary complex is deposited in the EM DataBank under accession code EMD-10401. The atomic coordinate of the complex is deposited in the Protein Data Bank under accession code PDB 6T8H. The source data underlying Figs. 4c, d and 5a–c, and Supplementary Figs. 6 and 7, and Supplementary Table 2 are provided as a Source Data file. Other data are available from the corresponding authors.

Received: 21 November 2019; Accepted: 5 March 2020; Published online: 27 March 2020

References
1. Kornberg, A. & Baker, T. A. DNA Replication (University Science Books, 2005).
2. Braithwaite, D. K. & Ito, J. Compilation, alignment, and phylogenetic relationships of DNA polymerases. Nucleic Acids Res. 21, 787–802
3. Delarue, M., Poch, O., Tordo, N., Moras, D. & Argos, P. An attempt to unify the structure of polymerases. Protein Eng. 3, 461–467 (1990).
4. Raia, P., Delarue, M. & Sauguet, L. An updated structural classification of replicative DNA polymerases. Biochem. Soc. Trans. 47, 239–249 (2019).
5. Sukenberg, P. T., Studwell-Vaughan, P. S. & O’Donnell, M. Mechanism of the sliding beta-clamp of DNA polymerase III holoenzyme. J. Biol. Chem. 266, 11328–11334 (1991).
6. Indiani, C. & O’Donnell, M. The replication clamp-loading machine at work in the three domains of life. Nat. Rev. Mol. Cell Biol. 7, 751–761 (2006).
27. Gueneau, E. et al. Structure of the MutL

24. Castrec, B. et al. Binding to PCNA in Euryarchaeal DNA replication requires

26. Dherin, C. et al. Characterization of a highly conserved binding site of Mlh1

28. Prestel, A. et al. The PCNA interaction motifs revisited: thinking outside the

29. Mayanagi, K. et al. Architecture of the DNA polymerase B-proliferating cell

34. Wing, R. A., Bailey, S. & Steitz, T. A. Insights into the replisome from the

33. Doublié, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. Crystal

10. Burgers, P. M. J. & Kunkel, T. A. Eukaryotic DNA replication fork.

15. Tori, K., Kimizu, M., Ishino, S. & Ishino, Y. DNA polymerases BI and D from

19. Ishino, Y., Komori, K., Cann, I. K. & Koga, Y. A novel DNA polymerase

46. McNally, R., Bowman, G. D., Goedken, E. R., O'Donnell, M. & Kunkel, T. A. Evolution of DNA polymerases: an inactivated polymerase-exonuclease

52. Yan, J. et al. Identification and characterization of a heterotrimeric archaeal

53. Pursell, Z. F., Isoz, I., Lundström, E.-B., Johansson, E. & Kunkel, T. A. Yeast

55. Tahirov, T. H., Makarova, K. S., Rogozin, I. B., Pavlov, Y. I. & Koonin, E. V. Translesion and repair DNA polymerases: diverse structure and mechanism.

59. Abrishami, V. et al. A pattern matching approach to the automatic selection of

31. Lacey, C. et al. Structure of the processive human Pol δ holoenzyme. Nat. Commun. 11, 1109 (2020).

32. Prestel, A. et al. The PCNA interaction motifs revisited: thinking outside the

35. De March, M. et al. Structural basis of human PCNA sliding on DNA. Nat. Commun. 8, 13935 (2017).

38. Mizu-Falconi, M., Giannattasio, M., Foiari, M. & Plevani, P. The DNA polymerase alpha-prime complex: multiple functions and interactions. ScientificWorldJournal 3, 21–33 (2003).

40. Baranovsky, A. G. et al. Mechanism of concerted RNA-DNA primer synthesis by the human primosome. J. Biol. Chem. 291, 10006–10200 (2016).

41. Kilkenny, M. L., De Piccoli, G., Perera, R. L. J., Labib, K. & Pellegrini, L. A conserved motif in the C-terminal tail of DNA polymerase α tethers primase to the eukaryotic replisome. J. Biol. Chem. 287, 23740–23747 (2012).

43. Sauguet, L. The extended ‘two-barrel’ superfamily: structure, function and evolution. J. Mol. Biol. 431, 4167–4183 (2019).

44. Fernandez-Leiro, R., Conrad, J., Scheres, S. H. & Lamers, M. H. Cyo-EM structures of the E. coli replicative DNA polymerase reveal its dynamic interactions with the DNA sliding clamps, exonuclease and e. elife 4, e11134 (2015).

45. Georgescu, R. E. et al. Structure of a sliding clamp on DNA. Cell 132, 43–54 (2008).

46. McNally, R., Bowman, G. D., Goedken, E. R., O'Donnell, M. & Kunkel, T. A. Analysis of the role of PCNA-DNA contacts during replamaging. BMC Struct. Biol. 10, 3 (2010).

47. Masuda, Y. et al. Different types of interaction between PCNA and PIP boxes contribute to distinct cellular functions of Y-family DNA polymerases. Nucleic Acids Res. 43, 7898–7910 (2015).

48. Acharya, N., Kläßen, R., Johnson, R. E., Prakash, L. & Prakash, S. PCNA binding domains in all three subunits of yeast DNA polymerase Δ6 modulate its function in DNA replication. Proc. Natl Acad. Sci. USA 108, 17927–17932 (2011).

49. Lindahl, T. Instability and decay of the primary structure of DNA. Nature 362, 709–715 (1993).

50. Yang, W. & Gao, Y. Translesion and repair DNA polymerases: diverse structure and mechanism. Annu. Rev. Biochem. 87, 239–261 (2018).

51. Kushida, T. et al. Pol B, a family B DNA polymerase, in Thermococcus kodakarenensis is important for DNA repair, but not DNA replication. Microbes Environ. 34, 316–326 (2019).

52. Yan, J. et al. Identification and characterization of a heterotrimeric archaeal DNA polymerase holoenzyme. Nat. Commun. 8, 15075 (2017).

53. Pursell, Z. F., Isoz, I., Lundström, E.-B., Johansson, E. & Kunkel, T. A. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. Science 317, 127–130 (2007).

54. Makarova, K. S., Krupovic, M. & Koonin, E. V. Evolution of replicative DNA polymerases in archaea and their contributions to the eukaryotic replisome machinery. Front. Microbiol. 5, 354 (2014).

55. Tahirov, T. H., Makarova, K. S., Rogozin, I. B., Pavlov, Y. I. & Koonin, E. V. Evolution of DNA polymerases: an inactivated polymerase-exonuclease module in Pol e and a chimeric origin of eukaryotic polymerases from two classes of archaeal ancestors. Biol. Direct 4, 11 (2009).

56. de la Rosa-Trevín, J. M. et al. Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. J. Struct. Biol. 195, 93–99 (2016).

57. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat. Methods 14, 331–332 (2017).

58. Zhang, K. Gete Real-time CTF determination and correction. J. Struct. Biol. 193, 1–12 (2016).

59. Abrishami, V. et al. A pattern matching approach to the automatic selection of particles from low-contrast electron micrographs. Bioinformatics 29, 2460–2468 (2013).

60. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. elife 7, e42166 (2018).

61. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. 66, 486–501 (2010).

62. Adams, P. D. et al. The Phenix software for automated determination of macromolecular structures. Acta Crystallogr. D Biol. Crystallogr. 70, 3065–3072 (2014).

63. Davis, I. W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. 35, W375–W383 (2007).
Acknowledgements

We wish to thank Dr Ravi Nookala for helpful advices and Jéril Degrouard (Laboratoire de Physique du Solide, Orsay, France) for help with preparing the samples. The cryo-EM data were collected at the Swedish National Cryo-EM Facility funded by the Knut and Alice Wallenberg Foundation, Family Erling Persson and Kempe Foundation, SciLifeLab, Stockholm University, and Umeå University. We thank Ahmed Haouz and the Pasteur Institute crystallization platform for help with preparing the crystals, and the staff at Soleil (PX2 beamline) and European Synchrotron Radiation Facility (ESRF) synchrotrons (ID23-1 beamline) for data collection. We thank Dr Ezazhisai Kandiah and Dr Gregory Effantin at the CM01 cryo-EM facility at ESRF, where initial screening and data collection were performed. The work and post-doctoral fellowship of C.M. are funded by an ANR JCJC grant ANR-17-CE11-0005-01. The work is also funded by Institut Pasteur, Ifremer, and the Swedish National Research Council (grant number 2017-04644). The fellowship of P.R. was funded by Sorbonnes University ED515 and Fondation pour la recherche médicale.

Author contributions

C.M., G.H., P.E., G.A., E.L., M.D., M.C., and L.S. conceived the experiments and analyzed the data. C.M., P.R., and I.H. performed molecular cloning and protein purification. C.M., G.A. and P.R. prepared the cryo-EM grids. M.C. collected the EM data, analyzed the data, and reconstructed the cryo-EM map. C.M. and L.S. prepared the figures. C.M. determined the atomic-scale models. All authors made suggestions and approved the final manuscript. G.H., M.C., M.D., and L.S. supervised the project.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-15392-9.

Correspondence and requests for materials should be addressed to M.C. or L.S.

Peer review information Nature Communications thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

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

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

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

© The Author(s) 2020