HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2023 June 15.

Published in final edited form as:

Nature. 2022 August ; 608(7924): 826–832. doi:10.1038/s41586-022-05040-1.

Structures of the human CST-Polα–primase complex bound to telomere templates

Qixiang He[1], **Xiuhua Lin**[1], **Bianca L. Chavez**[1], **Sourav Agrawal**[1], **Benjamin L. Lusk**[1], **Ci Ji Lim**[1,✉]

[1]Department of Biochemistry, University of Wisconsin–Madison, Madison, WI, USA.

Abstract

The mammalian DNA polymerase-α–primase (Polα–primase) complex is essential for DNA metabolism, providing the de novo RNA–DNA primer for several DNA replication pathways[1–4] such as lagging-strand synthesis and telomere C-strand fill-in. The physical mechanism underlying how Polα–primase, alone or in partnership with accessory proteins, performs its complicated multistep primer synthesis function is unknown. Here we show that CST, a single-stranded DNA-binding accessory protein complex for Polα–primase, physically organizes the enzyme for efficient primer synthesis. Cryogenic electron microscopy structures of the CST-Polα–primase preinitiation complex (PIC) bound to various types of telomere overhang reveal that template-bound CST partitions the DNA and RNA catalytic centres of Polα–primase into two separate domains and effectively arranges them in RNA–DNA synthesis order. The architecture of the PIC provides a single solution for the multiple structural requirements for the synthesis of RNA–DNA primers by Polα–primase. Several insights into the template-binding specificity of CST, template requirement for assembly of the CST-Polα–primase PIC and activation are also revealed in this study.

Polα–primase initiates DNA synthesis during genome replication[2,5]. The four-subunit Polα–primase (PRIM1–PRIM2–POLA1–POLA2, also known as p49–p58–p180–p70; Fig. 1a) complex does this by using its primase domain (PRIM1–PRIM2) to de novo synthesize a

✉**Correspondence and requests for materials** should be addressed to Ci Ji Lim. ciji.lim@wisc.edu.

Author contributions Q.H. and C.L. designed the expression constructs for recombinant protein production, with support from B.L.L., and Q.H. and X.L. performed insect cell culture and baculovirus production for insect cell infection. Q.H. expressed and purified the recombinant human protein complexes. Q.H., B.L.C. and C.L. conducted the negative-stain EM data collection and sample screening. Q.H. and C.L. prepared the cryo-EM samples for data collection. C.L. processed and analysed the cryo-EM datasets with support from Q.H. C.L. and Q.H. performed model building and refinement. C.L. and X.L. performed the enzyme assays, and C.L., X.L. and Q.H. analysed the data. Q.H., X.L., S.A. and B.L.C. performed the electrophoresis mobility shift assays, and Q.H., S.A., X.L. and B.L.C. analysed the data. C.L. directed the project and designed the experiments with input from Q.H., X.L., B.L.C. and S.A. C.L. wrote the manuscript with input from all authors.

Competing interests The authors declare no competing interests.

Reporting summary

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

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-05040-1.

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

short 7–11-nucleotide (nt) RNA primer, which is then extended by its DNA polymerase domain (POLA1)[2,5]. The enzyme requires modulation by highly conserved single-stranded DNA (ssDNA)-binding proteins (accessory proteins) such as the replication protein A (RPA)[6] and CTC1–STN1–TEN1 (CST; Fig. 1a; ref. [7]) complexes. These two protein complexes share functions[8,9] and structural homology[10].

At mammalian telomeres, CST inhibits telomerase extension of telomere overhangs[11,12] and coordinates the subsequent complementary strand synthesis with Polα–primase. This latter step is known as the telomere C-strand fill-in process and is essential for telomere stability[3,13]. Beyond telomere maintenance, the CST-Polα–primase machinery is also critical for restarting stalled DNA replication forks genome-wide and regulating DNA end resectioning at double-stranded DNA (dsDNA) breaks[11,14–19]. Despite a central and essential role in eukaryotic DNA metabolism, how Polα–primase performs its complicated enzymatic actions and how its accessory proteins modulate each step are unknown[2,5].
In this work we report the cryogenic electron microscopy (cryo-EM) structures of the human CST-Polα–primase preinitiation complex (PIC) bound to three distinct types of telomere overhang, which reveal several key structure–function insights into how the enzyme synthesizes de novo RNA–DNA primer with the aid of CST.

Template length requirements for PIC

To reconstitute the human CST-Pol–α–primase co-complex for structural determination, we purified recombinant CST and Polα–primase complexes from insect cells and combined them stoichiometrically (Fig. 1a,b and Extended Data Fig. 1a). We found that CST and Polα–primase directly interact, regardless of nucleic acids or the intrinsically disordered $POLA1_{1–337}$ domain of Polα–primase that interacts with other accessory proteins[6,20] (Extended Data Fig. 1b). In agreement with our finding, a recent structural study showed that CST-Polα–primase can form a recruitment co-complex in the absence of a template DNA or the disordered amino-terminal region of POLA1 (ref. [21]). Our recombinant Polα–primase is enzymatically active, as tested using telomeric ssDNA templates, and the enzyme can be stimulated by CST (Fig. 1c). However, the stimulation effect by CST was dependent on the length of the telomeric template (Fig. 1d and Extended Data Fig. 1c,d); a six-repeat telomeric template (6×TEL, $(TTAGGG)_6$), but not a three-repeat telomeric template (3×TEL, $(TTAGGG)_3$), resulted in stimulation of Polα–primase by CST. This result indicates that CST stimulation of Polα–primase activity requires a telomeric ssDNA template longer than 18 nt, a dimension that CST, alone, can bind tightly[11,22,23].

Next we investigated whether CST and Polα–primase form a new co-complex structure, with or without the presence of telomeric templates. Negative-stain EM analysis showed that CST-Polα–primase formed a new co-complex structure upon adding telomeric templates (Extended Data Fig. 2a), one that is different from that of CST or Polα–primase alone (Extended Data Fig. 2b). This new structure (hereafter termed PIC, as no nucleotides were added) was formed using telomeric ssDNA templates as short as 15 nt ($(TTAGGG)_2TTA$; Extended Data Fig. 2a). Without template added, the CST-Polα–primase co-complex failed to yield this new PIC structure (Extended Data Fig. 2b).

All of the above findings show that a short telomeric template (≥15 nt) is sufficient to trigger CST and Polα–primase to assemble into a preinitiation complex, but a longer template (>18 nt) is required to stimulate enzyme activity.

Structure of the human CST-Polα–primase PIC

To investigate the mechanism underlying the assembly of the CST-Polα–primase PIC and the stimulation mechanism, we used cryo-EM single-particle analysis to solve the cryo-EM structures of CST-Polα–primase bound to three different telomere templates: 3×TEL, 4×TEL-FB (a 15-base-pair foldback stem-loop preceding a $(TTAGGG)_4$ 3′ overhang) and 6×TEL (Supplementary Table 1). The three structures were solved at a global resolution (reported at a Fourier shell correlation (FSC) level of 0.143) of 3.6 Å, 3.4 Å and 3.7 Å for the 3×TEL, 4×TEL-FB and 6×TEL co-complexes, respectively (Supplementary Figs. 1–6 and Supplementary Table 2).

The three cryo-EM maps shared the same PIC structure, with map-to-map FSC values of about 3.7–3.8 Å (Extended Data Fig. 3), consistent with our negative-stain EM analysis (Extended Data Fig. 2a). The overall structural similarity of the cryo-EM maps shows that CST-Polα–primase forms the same PIC structure regardless of the telomeric template length or structure (single-stranded or overhang). As the three co-complexes share the same consensus structure, hereafter, unless otherwise stated, we used the 4×TEL-FB co-complex as the representative model to describe our findings (Fig. 1e).

The cryo-EM map resolution allowed us to dock most, if not all, of the seven subunits of CST and Polα–primase (from AlphaFold models[24] and past structural studies[2,10,25,26]), and de novo build the protein-bound DNA template, thus revealing the overall architecture of the human CST-Polα–primase PIC structure (Fig. 1e,f).

Polymerase and primase separated by CST

From the cryo-EM structure, we gained several key insights into the architecture of the PIC. The PIC has a 1:1:1 molar stoichiometry of CST/Polα–primase/template (Fig. 1e,f). Interestingly, the template-bound CST caused Polα–primase to undergo a large conformational change in the PIC structure, changing the enzyme from a compact shape to a segregated form (Fig. 1f and Extended Data Fig. 4). This conformational change led to a separation of the RNA[26] and DNA[27] catalytic centres of Polα–primase, exposing the DNA catalytic centre of POLA1 that was buried in the apo state of the enzyme[2,25] (Extended Data Fig. 4).

The template-bound CST serves as a scaffold to mediate the conformational change of Polα–primase, and this is achieved through two spatially separated interaction sites between CST and Polα–primase; $STN1_N$–TEN1 binds the primase domain (PRIM1–PRIM2–$POLA1_{CTD}$–POLA2), and CTC1 and $STN1_C$ support the polymerase domain ($POLA1_{EXO}$–$POLA1_{CAT}$; Fig. 1f). TEN1 seemingly makes contacts with three of the four (PRIM1, $PRIM2_N$ and POLA2) subunits of the primase domain, with the $PRIM1_{84–101}$ loop sitting in the interface between $STN1_N$ and TEN1 (Fig. 1f and Extended Data Fig. 5a–d).

The template-bound CST interacts with the polymerase domain through several contact points (Fig. 1f): the antiparallel β-strands of $CTC1_{OB\text{-}G}$ interact with the flexible loop between residues 547 and 556 of $POLA1_{EXO}$ (Extended Data Fig. 5e); the template interacts with $POLA1_{EXO}$, whereby $POLA1_{EXO}$ uses an α-helical region and other nearby charged residues to contact the ssDNA template, as seen in yeast Pola1 (ref. [28]; Extended Data Fig. 5f); and $STN1_C$ interacts with $POLA1_{EXO}$ (Extended Data Fig. 5g)—coulombic surface analysis reveals a positively charged surface that $POLA1_{EXO}$ potentially uses to interact with a negatively charged surface on $STN1_C$ (Extended Data Fig. 5h,i).

We found that the $PRIM2_N$ conformation is slightly changed in the PIC structure (as compared to its apo-state structure[25]) to accommodate the primase domain conformation, with the large subdomain of $PRIM2_N$ (ref. [29]) rotated about 40° from the small subdomain that interacts with PRIM1 (Extended Data Fig. 6a).

The cryo-EM map densities of the thumb domain of $POLA1_{CAT}$ and $PRIM2_C$ are poorly resolved (Supplementary Fig. 7), indicating substantial flexibility in these domains. This flexibility of the thumb domain of $POLA1_{CAT}$ in the PIC allows the thumb domain to curl up when $POLA1_{CAT}$ binds to a primer–template molecule[27,28,30,31] (Extended Data Fig. 6b). The flexibility of the $PRIM2_C$ domain is consistent with it being flexibly tethered to the rest of the primase domain by an 18-amino-acid linker (Fig. 1a). Below we will further analyse and discuss the PIC structure flexibility and functions (see the section entitled $PRIM2_C$ is poised for primer synthesis).

Molecular basis for telomere binding by CST

Fifteen nucleotides of the 4×TEL-FB template were resolved and modelled in the cryo-EM structure of our PIC (Fig. 2a). The telomeric sequence, 5′-TAGGGTTAGGGTTAG-3′, assigned to the 15-nt template model, was based on the best model-to-map cross-correlation value (Extended Data Fig. 7a,b). This sequence assignment is further supported by DNA-binding experiments (Fig. 2b,c and Extended Data Fig. 7c). The template strand polarity is further confirmed by the foldback dsDNA density at the template 5′ end in the cryo-EM map (Fig. 1e,f).

Given that the telomeric ssDNA portion of the 4×TEL-FB template is 24 nt, there remains 8 nt of unresolved template at the 3′ end (based on the assigned sequence). This unresolved portion of the template, which extends beyond $STN1_N$ into the primase region, is presumably flexible, either bound or unbound by a protein. In support of this view, $PRIM2_C$, the domain of primase that engages this part of the template during primer synthesis, binds the template poorly[32] and is flexibly linked to the rest of the primase[29]. The $PRIM2_C$ domain is also poorly resolved in the cryo-EM map of our PIC structure (Supplementary Fig. 7).

Our template model extends from the previously determined ssDNA-binding anchor site of CST[10] and reveals how CST uses other binding sites to engage a longer telomeric ssDNA. The bound template adopts an extended conformation that spans 5′ to 3′ from $CTC1_{OB\text{-}F}$ to $STN1_N$. These protein–DNA interactions are further supported by interactions of the

POLA1$_{EXO}$ domain (Fig. 2d). The modelled template conformational homogeneity and its contact with the CTC1, STN1$_N$ and POLA1$_{EXO}$ domains (Fig. 2a,d) suggest that this portion of the template is stably bound by CST-Polα–primase.

These protein–ssDNA interactions are clustered across three separate ssDNA-binding sites in CTC1$_{OB-F/G}$, CTC1$_{OB-G}$ and STN1$_N$ (Extended Data Fig. 7d). The first binding site in CTC1$_{OB-F/G}$ shares similar interactions with the previously determined ssDNA-binding anchor site of CST[10]. This binding site utilizes a network of hydrogen bonds to contact the phosphate groups and bases of A2 and G3 of the template (Fig. 2d and Extended Data Fig. 7e,f): Y983, S977 and N981 of CTC1 hydrogen bond with A2-p and G3-p; E951, S979, E1162 and K1167 of CTC1 hydrogen bond with the A2 and G3 bases; and K1164 of CTC1 seemingly forms a cation–π interaction with the G3 base (Extended Data Fig. 7f). Extensive hydrogen bonds are also observed at the second binding site at CTC1$_{OB-G}$ (Fig. 2d and Extended Data Fig. 7g): E1120, R1175, Q1177, R1193 and R1195 of CTC1 hydrogen bond with the bases of A8, G9, G10 and G11. This binding site is further supported by the POLA1(EXO) domain binding to several template residues in its proximity: N652 of POLA1 hydrogen bonds with A8-p, and Q554 and K672 of POLA1 hydrogen bond with T7 and G11 bases, respectively (Fig. 2d and Extended Data Fig. 7h).

The third binding site in STN1(N) uses a wider variety of interactions to engage the template (Fig. 2d and Extended Data Fig. 7i,j): K55 and R139 of STN1 form salt bridges with T12-p and T13-p, respectively; R69 of STN1 forms a hydrogen bond with the G15 base; and W89 and Y141 form a π–π interaction with the A14 and G15 bases, respectively. Coulombic surface analysis confirmed that the interactions utilized by the first two binding sites of CTC1 are electrostatic (highly positively charged surface), whereas the third binding site in STN1$_N$ is neutral (Extended Data Fig. 7k). Amino acid sequence conservation analysis[33] suggests that most of the PIC residues identified for ssDNA binding are well conserved (Supplementary Fig. 8). As similarly shown for ssDNA molecules bound by POT1 (ref. [34]) or RPA[35], the extended conformation of the template is supported by several DNA base-stacking interactions: G3 stacks with G4, G5 stacks with T6, T7 stacks with A8 and G9, G10 stacks with G11, and T13 stacks with A14 (Fig. 2d).

The function of several of these identified DNA-binding residues has been demonstrated: most of the site 1 residues have been rigorously validated as the ssDNA-binding anchor site of CST[10], the alterations to R1193 and R1195 of CTC1 in site 2 were shown to abolish ssDNA binding by CST[10], and the alterations to W89, R139 and Y141 of STN1 in site 3 reduced the affinity of CST for short telomeric DNA[36].

Collectively, the above protein–DNA analysis provides a structural basis for CST binding to telomeric ssDNA: the modelled 15-nt telomeric template (5′-TAGGGTTAGGGTTAG-3′) explains why CST prefers G-rich sequences when binding short ssDNA molecules[11,22]. The preference of CST for G-rich sequences is defined by the specific and nonspecific interactions between CST and the template, which explain why CST can or cannot tolerate base changes at specific template positions[22]. For example, the T12 and T13 residues of the template do not participate in base stacking or protein–DNA interactions; thus, their base changes can be tolerated[22].

PIC directs the template 3′ end to PRIM1

The polymerase domain sits on top of the template-bound CST (Fig. 1f), and this interaction is mediated by the $POLA1_{EXO}$ binding to the template (see the second binding site in Fig. 2d), CTC1 and $STN1_C$ (Fig. 1f and Extended Data Fig. 5g). Surface representation of the PIC structure revealed that this assembly formed a tight-fit tunnel for the CST-bound template (Fig. 3a), directing the 3′ end of the template towards the RNA catalytic centre of PRIM1 (Fig. 3b, c). This tunnel was also recently reported for the *Tetrahymena* telomere replication holoenzyme[37], suggesting that this molecular tunnel is likely to be a conserved structural feature for CST-Polα–primase in the eukaryotic kingdom.

The function of this molecular tunnel regarding RNA–DNA primer synthesis is unclear, but we can draw some insights from our PIC structure. On the basis of our protein–ssDNA interaction analysis and PIC architecture (Figs. 1f and 2d), the positioning of the polymerase domain on top of the template-bound CST may help stabilize the template binding of CST or direct the template 3′ end towards the RNA catalytic centre of PRIM1 (Fig. 3b, c) or both. Further analysis of the PIC structure flexibility yielded more insights into other possible functions of the tunnel (see the section below entitled POLA1 swivels on top of CST–template).

The PIC structure is not compatible with the ability of CST to form a decameric nucleoprotein complex[10]: the polymerase domain on top of CST would be a steric hindrance for CST to form an inverted dimer, which is the first step of CST oligomerization (Extended Data Fig. 8).

Catalytic centres are in synthesis order

RNA–DNA primers made by Polα–primase are in the order of RNA to DNA synthesis[2,5], and CST stimulates this complicated process at several steps[7,38,39]. Our PIC structure reveals how CST, as an ssDNA-binding accessory protein of Polα–primase, arranges the template and Polα–primase into a conformation that stimulates enzyme activity (Fig. 3c,d). Specifically, CST positions the DNA and RNA catalytic centres of the enzyme in the order of 5′ to 3′ along an extended template.

We propose that this unique arrangement can stimulate Polα–primase activity at several enzymatic steps by increasing the template binding affinity, removing potential template secondary structures and facilitating efficient switching of RNA to DNA synthesis. First, the PIC structure shows that the template is bound mainly by CST (Fig. 2d), which provides the molecular reasoning for: how CST modulates Polα–primase template affinity[39], how CST recruits Polα–primase to telomeres[13], and why DNA binding is necessary for stimulation of Polα–primase activity[40]. Second, the structure also reveals that the ssDNA template ahead of Polα–primase is unfolded (inferred by the extended conformation of the template) by the CST-polymerase domain (Fig. 3a), which would explain how CST-Polα–primase resolves ssDNA secondary structures at stalled DNA replication forks[15,41]. Third, positioning the RNA catalytic centre of PRIM1 downstream of the template after the DNA catalytic centre of POLA1 (Fig. 3d) provides the correct sequence of action for RNA and DNA

synthesis. Presumably, this order facilitates an efficient handover of RNA primer–template from PRIM1 to POLA1 for DNA elongation (for further analysis and discussion, see the section below entitled $PRIM2_C$ is poised for primer synthesis).

POLA1 swivels on top of CST–template

Three-dimensional (3D) variability analysis (3DVA)[42] of the cryo-EM structure of our PIC revealed that the polymerase domain swivels on top of the template-bound CST (Extended Data Fig. 9 and Supplementary Video 1), with the pivot point being the interaction between $POLA1_{EXO}$ and CTC1 (Extended Data Fig. 5e). The swivel range is seemingly restricted and controlled by the $STN1_C$ domain; a flexible 7-amino-acid linker connects this domain to the $STN1_N$ domain (Extended Data Fig. 9). This large swivelling motion is consistent with the modest interactions between $POLA1_{EXO}$ of the polymerase domain and its three contact points: CTC1, $STN1_C$ and template (Extended Data Fig. 5e–g). The function of this elaborate mechanism is unclear but may be related to template movement during RNA or DNA synthesis. A speculative but plausible explanation is that this swivelling motion comes from a Brownian ratchet process used by CST-Polα–primase to facilitate template translocation during RNA primer synthesis.

$PRIM2_C$ is poised for primer synthesis

The carboxy-terminal domain of PRIM2, $PRIM2_C$, facilitates primer initiation by binding the 5′-triphosphate of the primer and coordinating the handover of the RNA primer–template to POLA1 (refs. [25,26,32,43]). These functions depend on the domain flexibility of $PRIM2_C$; $PRIM2_C$ is linked to $PRIM2_N$ through an 18-amino-acid linker[2,32]. How the flexibility of the $PRIM2_C$ domain mediates the multiple functions of $PRIM2_C$ is unclear.

3DVA analysis revealed that the $PRIM2_C$ domain is flexible compared to the rest of the rigid PIC structure (Fig. 4a,b and Supplementary Video 2). From the range of flexibility, we found that $PRIM2_C$ can position itself in front of the template, the RNA catalytic centre of PRIM1 or DNA catalytic centre of POLA1. These are key positions where $PRIM2_C$ has to be to perform its multiple functions (Fig. 4b). We thus propose that the PIC structure represents a single architectural platform that CST-Polα–primase uses to meet the multiple conformational requirements for RNA–DNA primer synthesis (Fig. 4c). The $PRIM2_C$ position and flexibility would enable it to find and bind the template that protrudes from the CST-polymerase domain tunnel and initiate RNA primer synthesis with the nearby PRIM1 (as shown in the next section). After maturation of the RNA primer, the $PRIM2_C$-bound primer–template requires only a simple tilt for the conveniently positioned $POLA1_{CAT}$ to engage the 3′ end of the primer.

Template-bound primase structure in PIC

One of the $PRIM2_C$ positions in our 3DVA analysis showed $PRIM2_C$ in contact with the RNA catalytic centre of PRIM1 (Fig. 4a, pink-coloured density, and Supplementary Video 1), suggesting a state in which $PRIM2_C$ could bind both the unresolved 3′ end of the

template and PRIM1—an advanced PIC state in which the primase is poised to incorporate the first set of NTPs.

Thus, using further 3D heterogeneity analysis[44,45], we attempted to filter out a conformational state in which PRIM1 and $PRIM2_C$ are in contact. The analysis yielded a subset of particles, leading to a 4.3-Å (FSC = 01.43) cryo-EM map (Fig. 5a and Extended Data Fig. 10a). This map has the $PRIM2_C$ density sufficiently well resolved for us to dock in a $PRIM2_C$ atomic model. Structural studies have shown that the $PRIM2_C$ domain exists as two structural variants: an all-helical variant form[46] and a β-sheet-containing version[47]. The all-helical variant is proposed to be the variant that is more capable of binding ssDNA[46]. We found that the all-helical variant docks well into our $PRIM2_C$ density (Extended Data Fig. 10b), revealing the human primase structure that is poised to initiate RNA primer synthesis (Fig. 5b).

In this advanced preinitiation state, $PRIM2_C$ binds both PRIM1 and template (Fig. 5b,c and Extended Data Fig. 10c,d). The binding of $PRIM2_C$ stabilizes the previously unresolved 3′ portion of the template (Fig. 2) and allows us to extend the template by a further 5 nt (5′-GGTTA-3′; Extended Data Fig. 10e). We now see that the template 3′ end reaches into the RNA catalytic crevice that is formed by PRIM1 and $PRIM2_C$ (Extended Data Fig. 10f). The template entry into the crevice is seemingly guided by several protein–DNA interactions that are contributed by both proteins (Fig. 5c and Extended Data Fig. 10d), with many of these participating residues well conserved (Supplementary Fig. 9). In particular, the identified R56 residue of PRIM1 has been shown to be critical for primase RNA primer synthesis[26], but its underlying mechanism is unknown. Our structure reveals that the $PRIM1_{R56}$ plays an essential role in template binding and positioning the template for RNA primer initiation.

The template extension into the PRIM1–$PRIM2_C$ crevice revealed that the A20 residue of the template (the last modelled nucleotide) is positioned in front of the RNA catalytic centre of PRIM1 (Fig. 5d) and is thus poised for PRIM1 to catalyse the formation of the first NTP of de novo RNA primer synthesis. This extended template path connects seamlessly with the template path seen in a previously solved structure of a template-bound $PRIM2_C$ (ref. [25]; Extended Data Fig. 10g). By combining our PIC structure and that structure, we can visualize how a more prolonged template traces along the molecular surface of a CST-Polα–primase preinitiation complex (Fig. 5e); the template makes a sharp turn on entering the PRIM1–$PRIM2_C$ crevice and is guided to exit the PRIM1 catalytic centre through its interaction with $PRIM2_C$.

The flexibility of the $PRIM2_C$ domain (Fig. 4a and Supplementary Video 2) suggests that this primase preinitiation structure is a transient one, which is consistent with the rate-limiting steps of primase RNA primer synthesis[32,48]. As the template threads through the crevice formed between PRIM1 and $PRIM2_C$ (Fig. 5e and Extended Data Fig. 10f), it is likely that PRIM1 or $PRIM2_C$ first binds the template before interacting with the other protein to form the observed structure.

The 4×TEL-FB template is nearly sufficient in length (about 24 nt of telomere overhang) to reach into the RNA catalytic centre of PRIM1–$PRIM2_C$ (Fig. 5d). In other words, for

CST-Polα–primase PIC to be productive, it requires a template of at least this length. A shorter template, such as 3×TEL, would not be productive as it will be too short to reach the primase after CST binding. The 3×TEL template can still trigger CST-Polα–primase to form the PIC (because 3×TEL is still long enough for CST to bind; Fig. 2), just not a productive one (Fig. 1c,d). Longer templates, such as 6×TEL, will result in the formation of a PIC that is productive (Fig. 1d).

Discussion

Our work provides the structural basis to explain how the ssDNA-binding accessory protein complex CST dictates the specificity of Polα–primase for the template and organizes the enzyme to perform its complicated multistep RNA–DNA primer synthesis during telomere replication. CST achieves this by driving Polα–primase to form a preinitiation complex structurally poised to stimulate the enzyme activity at several levels. This co-complex structure also provides an elegantly simple solution to the complex changing needs of the biphasic Polα–primase functions. Given the structural similarity of CST to RPA[6,10], it will be interesting to determine whether RPA shares the same stimulatory mechanism.

The CST-Polα–primase PIC structure ties in several previous studies and places them in a coherent framework that can be used to generate new structure-guided hypotheses and questions about the downstream catalytic mechanisms of the enzyme. How the swivelling action of the polymerase domain participates in template translocation during RNA primer synthesis and how the primase determines the length of the RNA primer and performs primer handover are directions for future research. Notably, our PIC structure is compatible with the archaeal primase models[49,50]. Future structure–function studies involving further enzymatically advanced co-complexes will yield answers to this next set of questions.

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/s41586-022-05040-1.

Methods

Cloning

Human *POLA1* (NP_058633.2) and $POLA1_{\Delta 1–337}$ were cloned into the pOET1 transfer vector (Mirus Bio). Human *PRIM1* (NP_000937.1), *PRIM2* (NP_000938.2) and *POLA2* (NP_002680.2) cDNAs were individually cloned into the pFastBac1 expression vector (Invitrogen). The POLA1 and $POLA1_{\Delta 1–337}$ expression vectors have N-terminal twin-Strep-tags whereas those of the POLA2, PRIM1 and PRIM2 constructs have N-terminal hexa-histidine (6×His) tags. The cDNAs of the human CST subunits, *CTC1* (AAI11784), *STN1* (NP_079204) and *TEN1* (NP_001106795) were cloned into a single pBAC4x-1 transfer vector (Novagen). The CTC1 open reading frame has an N-terminal 3×Flag whereas both STN1 and TEN1 have 6×His tags.

Insect cell culture and baculovirus generation

Recombinant baculovirus encoding CST or POLA1 (or its truncation mutant) was produced using the flashBAC ultra system (Mirus Bio). The baculoviruses encoding PRIM1, PRIM2 and POLA2 were made using the Bac-to-Bac system (Invitrogen). All viruses were made using *Spodoptera frugiperda* (SF9) cells (Invitrogen). The viruses were amplified to a titre of >1.0×10^8 pfu ml^{-1} before using them for insect cell infection. The virus titres were measured by a flow cytometry assay (Expression Systems).

Expression and purification of CST and Polα–primase

One or two litres of *Trichoplusia* (Tni) cells (Expression System) were infected with recombinant baculoviruses at a cell density of 1.5–2.0 × 10^6 cells per millilitre. A single baculovirus stock was used for CST expression, and four baculoviruses (POLA1, POLA2, PRIM1 and PRIM2) were used to co-express Polα–primase. Each infection was carried out at a multiplicity of infection of 1. The infected Tni cells were further incubated in an orbital shaker (27 °C, 130 r.p.m.) for 66–68 h before collection.

The CST protein complex was purified according to an established protocol[10]. The infected Tni cells were collected by centrifugation at 1,500*g* for 30 min at 4 °C. The cell pellets were resuspended in lysis buffer (25 mM HEPES–NaOH pH 7.5, 300 mM NaCl, 15 mM imidazole, 1 mM dithiothreitol (DTT), 1× Xpert Protease Inhibitor Cocktail Solution (GenDEPOT)) at 50 ml lysis buffer per litre of cells and subjected to ultrasonication for cellular lysis. The cell lysate was clarified by high-speed centrifugation, and the clarified lysate was incubated with Ni-NTA agarose resin (Qiagen) for 2 h at 4 °C using a rotator. The resin was washed three times with wash buffer (same as the lysis buffer but without protease inhibitors) and eluted with Ni-NTA elution buffer (25 mM HEPES–NaOH pH 7.5, 300 mM NaCl, 300 mM imidazole, 1 mM DTT and protease inhibitors). The eluate was then incubated with anti-Flag resin (GenScript) overnight at 4 °C. The resin was washed three times with wash buffer, and the CST complex was eluted with Flag elution buffer (25 mM HEPES–NaOH pH 7.5, 300 mM NaCl, 1 mM DTT, 0.4 mg ml^{-1} 3×Flag peptide (APExBIO)).

The protocol for purification of Polα–primase is the same as that for purification of CST except that the buffers contain 150 mM NaCl instead of 300 mM NaCl and a twin-Strep-tag pulldown was performed instead of a Flag-tag pulldown. The Ni-NTA eluate was incubated with Strep-Tactin XT 4Flow-resin (IBA LifeScience) overnight at 4 °C. The resin was washed three times with wash buffer (25 mM HEPES–NaOH pH 7.5, 150 mM NaCl, 1 mM DTT) before eluting the Polα–primase complex with 1× BXT elution buffer (IBA LifeScience). The purity and integrity of both protein complexes were checked with SDS–polyacrylamide gel electrophoresis (PAGE) analysis before they are concentrated to ≥5.0 mg ml^{-1}. The complexes were used in experiments or snap-frozen in small aliquots for long-term storage at −80 °C.

Co-complex formation

Purified CST and Polα–primase were mixed together at a 2:1 molar ratio and incubated at 4 °C overnight. Strep-Tactin XT 4Flow-resin was added to the mixture and incubated for 1

h at 4 °C. The resin was washed four times with wash buffer (25 mM HEPES–NaOH pH 7.5, 150 mM NaCl, 1 mM DTT) before the co-complex was eluted with 1× BXT elution buffer supplemented with 1 mM $MgCl_2$ (0.5 h incubation). The integrity and purity of the co-complex were confirmed with SDS–PAGE before experimental use or snap-freezing for long-term storage at −80 °C.

Pulldown assay

The protein bait and prey samples were incubated for 1 h at 4 °C before Strep-Tactin XT 4Flow-resin was added to capture the twin-Strep-tagged bait proteins. The resin was washed three times with wash buffer before the captured samples were eluted with 1× BXT elution buffer (incubated for 0.5 h). The pulled-down protein samples were analysed with Coomassie-stained SDS–PAGE.

Agarose gel binding assay

Radioactive oligonucleotides, 5′-labelled with ^{32}P, were used in the binding assays. The estimated specific activity was at least 200,000 c.p.m. $pmol^{-1}$, and 500 c.p.m. was used for each reaction. For binding, purified protein complexes and DNA were mixed and incubated as a 10 µl reaction volume in binding buffer (25 mM HEPES–Na pH 7.5, 150 mM NaCl, 2 mM $MgCl_2$, 0.2 mM EGTA, 0.1% NP-40, 10% glycerol, 1 mM TCEP) for 2 h on ice. The samples were then loaded into a pre-chilled 1× TBE 0.7% agarose gel and the gel was run at 7 V cm^{-1} for 1.5 h in a cold room (4 °C). The gels were vacuum-dried at 80 °C for 1.5 h before exposure to a storage phosphor screen overnight. The screen was then imaged with a Typhoon FLA 9000 scanner (GE Lifesciences). Binding analysis was carried out using GelAnalyzer software (GelAnalyzer 19.1; www.gelanalyzer.com), and curve fitting was performed with OriginPro software (OriginLab, USA). The Hill binding equation was used for the curve fitting.

Enzyme direct assay

A standard reaction (20 µl) consisted of NTPs, dNTPs, 5 µM DNA template and 50 nM Polα–primase in direct assay buffer (of 50 mM of HEPES–NaOH pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$ and 2 mM DTT). Addition of CST concentrations was as indicated in the figure legends or annotations. The NTP mixture used was ATP rich; 2 mM ATP, and 50 µM UTP and 50 µM CTP. The dNTP mixture used was: 0.13 µM [^{32}P]dCTP, 2.37 µM dCTP, 50 µM dATP and 50 µM dTTP. DNA templates of various telomeric repeats $(TTAGGG)_n$ were ordered from a vendor (IDT). The enzymatic reactions were initiated by adding the DNA template. The reactions were incubated at 37 °C for 1 h before they were quenched by DNA precipitation. A random-sequence 80-nt oligonucleotide was added as a loading control in this step. After DNA precipitation, each sample was dissolved in 20 µl 1× formamide loading dye and 10 µl of it was loaded into each well of a 10% 1× TBE 7 M urea PAGE gel. Electrophoresis was performed at a constant 45 W until the bromophenol blue dye reached a third of the way from the gel bottom. The gels were vacuum-dried at 80 °C for 1.5 h before exposure to a storage phosphor screen. After adequate exposure, the gels were imaged on a Typhoon FLA9000 gel imager. Enzyme activity analysis was carried out using the GelAnalyzer software.

Negative-stain EM sample preparation, data collection and image analysis

Samples were immediately diluted to 75–100 nM before depositing onto a glow-discharged EM grid (CF400-Cu-UL, EMS). The grids were then subjected to 2% (w/v) uranyl formate (Structure Probe) staining using an established protocol[51]. The negative-stain EM sample grids were imaged on a 120 kV transmission electron microscope (Talos L120C, Thermo Fisher). Image datasets were collected with SerialEM[52] using low-dose mode. 2D classification analysis was performed using cryoSPARC2 (ref. [53]).

Cryo-EM sample preparation and data collection

Cryo-EM sample grids for the co-complex were prepared as described for CST[10]. Briefly, 1.2/1.3 μm 300-mesh holey grids (C-Flat or Quantifoil) were glow-discharged for 20 s at 15 mA (PELCO easiGlow) before use. Protein samples were supplemented with 4 mM CHAPSO detergent[54] before being added to freshly glow-discharged EM grids and then vitrified using Vitrobot (Thermo Fisher) with a 4.5 s blot time at 4 °C and 95% relative humidity.

Suitable cryo-EM sample grids for data collection were first screened on an Arctica 200 kV transmission electron microscope (Thermo Fisher) equipped with a Bioquantum K3 energy filter-detector (Gatan). The CST-Polα–primase–3×TEL sample dataset (6,252 videos) was collected at the National Cryo-EM Facility at the Frederick National Laboratory for Cancer Research. The videos were collected using Latitude software at a calibrated pixel size of 1.08 Å per pixel, a defocus range of −1.0 to −2.5 μm, a total dose of 50.0 e^{-} $Å^{-2}$ (40 frames) and the K3 detector (20-eV energy slit) in counting mode. The CST-Polα–primase–4×TEL-FB sample dataset (7,794 videos) was collected at the University of Wisconsin–Madison Cryo-EM Research Center. The videos were collected using SerialEM software[52] (version 4.0) at a calibrated pixel size of 1.1 Å per pixel, a defocus range of −0.7 to −2.8 μm, a total dose of 50.0 e^{-} $Å^{-2}$ (40 frames) and the K3 detector (20 eV energy slit) in counting mode. The CST-Polα– primase–4×TEL-FB sample dataset (8,658 videos) was collected at the University of Wisconsin–Madison Cryo-EM Research Center. The videos were collected using SerialEM software at a calibrated pixel size of 1.1 Å per pixel, a defocus range of −0.7 to −2.8 μm, a total dose of 51.1 e^{-} $Å^{-2}$ (40 frames) and the K3 detector (20 eV energy slit) in counting mode.

Cryo-EM single-particle analysis

In general, all three datasets were processed using similar image processing pipelines (Supplementary Figs. 5–7). All video frames were motion-corrected (5 × 5 patch)[55] and their contrast transfer function (CTF)[56] was estimated in Relion. The motion-corrected micrographs were curated, on the basis of their CTF parameters, and subjected to particle picking using the Topaz software[57]. The picked particles were extracted with a binning factor of 5, and particles corresponding to intact samples were selected through two rounds of hetero refinement (six classes) in the cryoSPARC software[53]. The selected particles were re-extracted and centred at the original pixel size before 3D reconstruction using non-uniform refinement[45] in cryoSPARC. The refined particles were exported back to Relion for several rounds of CTF refinement (global and per-particle) and particle polishing[55,58]. These particles were then imported back to cryoSPARC for final reconstruction to obtain the

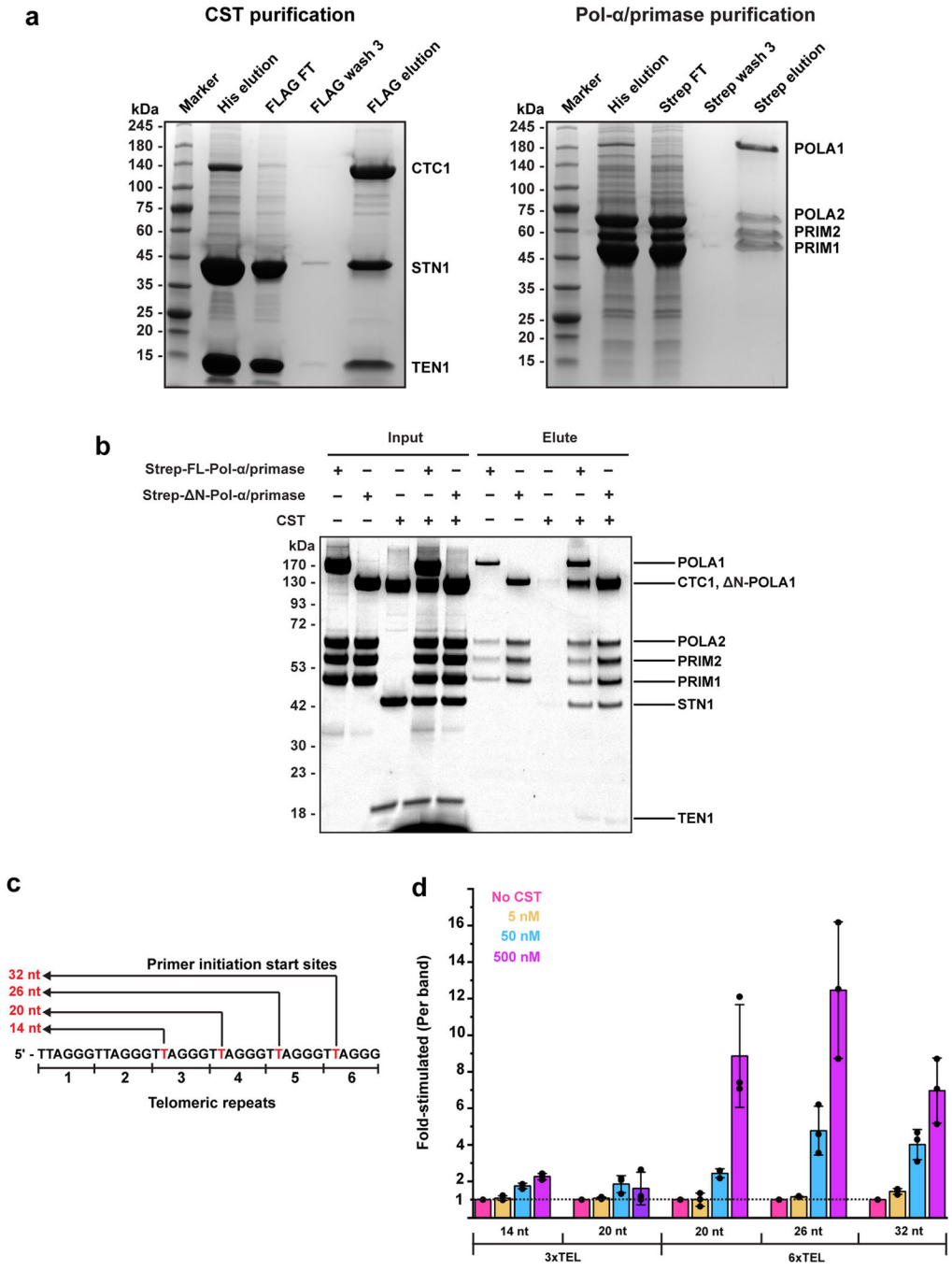
cryo-EM maps of the consensus PIC structure. The final reconstruction was performed using the non-uniform refinement algorithm in CryoSPARC.

To improve the resolution of the consensus maps, masked 3D classification and local refinement were performed individually for the CTC1–$STN1_N$–ssDNA, polymerase–$STN1_C$ and primase domains (Supplementary Figs. 4–6). The masks were generated in ChimeraX software[59,60] using the docked atomic models as segmentation guides. Each domain was subjected to particle subtraction and 3D classification (eight classes) in Relion to identify a subset of particles that yielded an intact domain conformation. During particle subtraction, the particles were recentred and re-extracted at a smaller box size (256 pixels) for 3D classification. Correspondingly, a matching 3D model was reconstructed at a similar dimension to that of the subtracted particle images and filtered to 12 Å. This 3D model was then used as an initial model for the eight-class 3D classification step. The 3D classification was performed with either no alignment or with a small-angle search with local searches enabled. Appropriate 3D class(es) were selected and reverted to their original images for a final 3D reconstruction. The three local refined domain EM densities were combined using the consensus map as a guide. The conversion of files from cryoSPARC to Relion was performed using UCSF PyEM scripts (https://github.com/asarnow/pyem).

Model building, refinement and validation

All seven protein subunits of the PIC were built using computed models from the AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk/). The models were first manually moved into their respective positions in the 4×TEL-FB cryo-EM map and then flexibly fitted using the Jiggle fit function in Coot[61]. The ssDNA molecule was built starting from the 4-nt model shown in the previous CST structure[10]. The assembled PIC model was visually checked in Coot before subjecting it to real-space refinement in PHENIX[62] and visually checked again using Coot. Model-to-map validations were performed using the PHENIX cryo-EM comprehensive validation program.

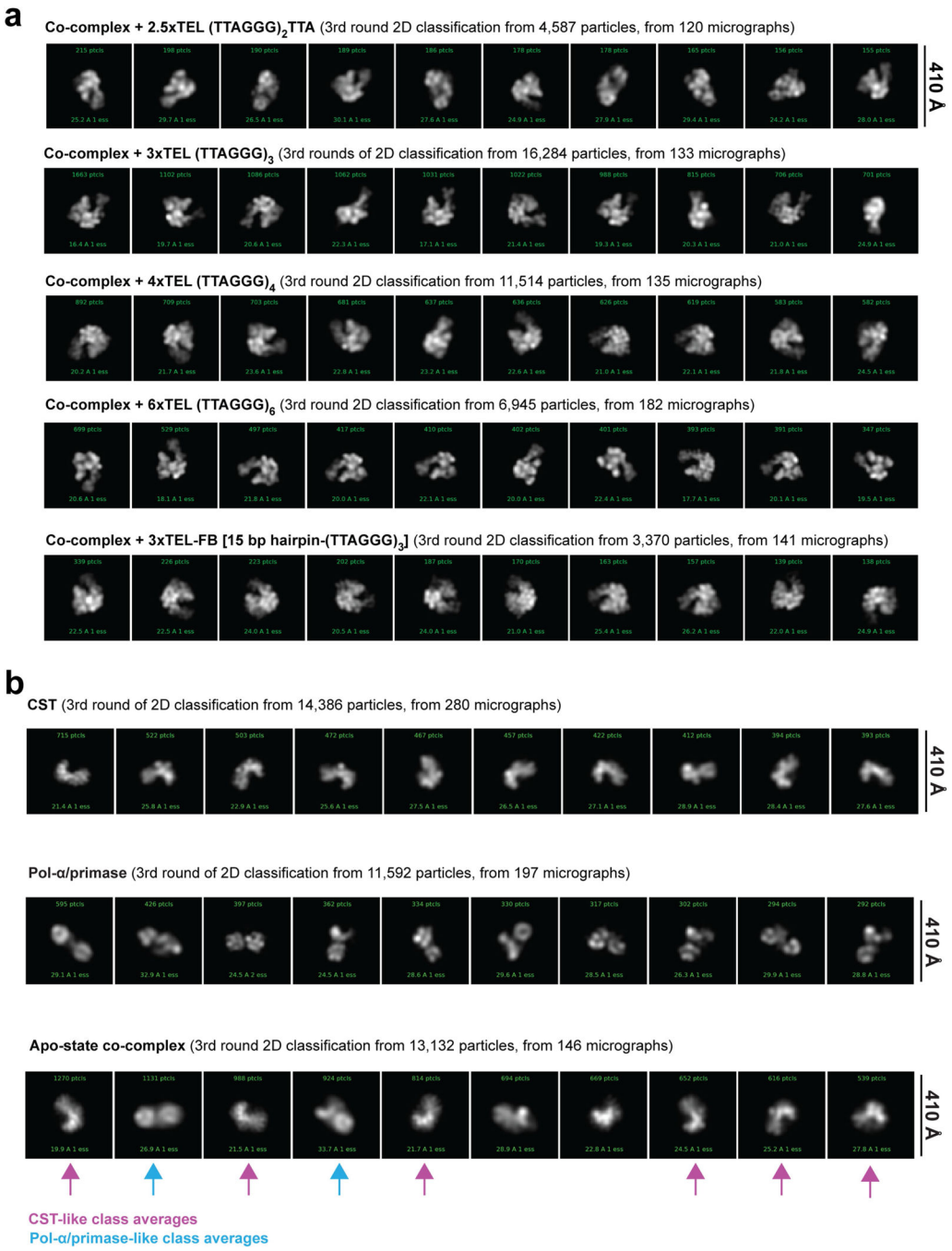
Extended Data



Extended Data Fig. 1 |. Recombinant human CST and pol-α/primase purification and their biochemical characterization assays.

(**a**) Tandem affinity tags purification of CST and pol-α/primase complexes from baculovirus-infected insect cells. CST heterotrimeric complex was purified by first using Ni-NTA beads to pull down 6xHis-STN1 and 6xHis-TEN1 and then using anti-FLAG beads to pull-down 3xFLAG-CTC1. Pol-α/primase 4-subunit complex was purified by first using Ni-NTA beads to pull down 6xHis-PRIM1, 6xHis-PRIM2, and 6xHis-POLA2 and

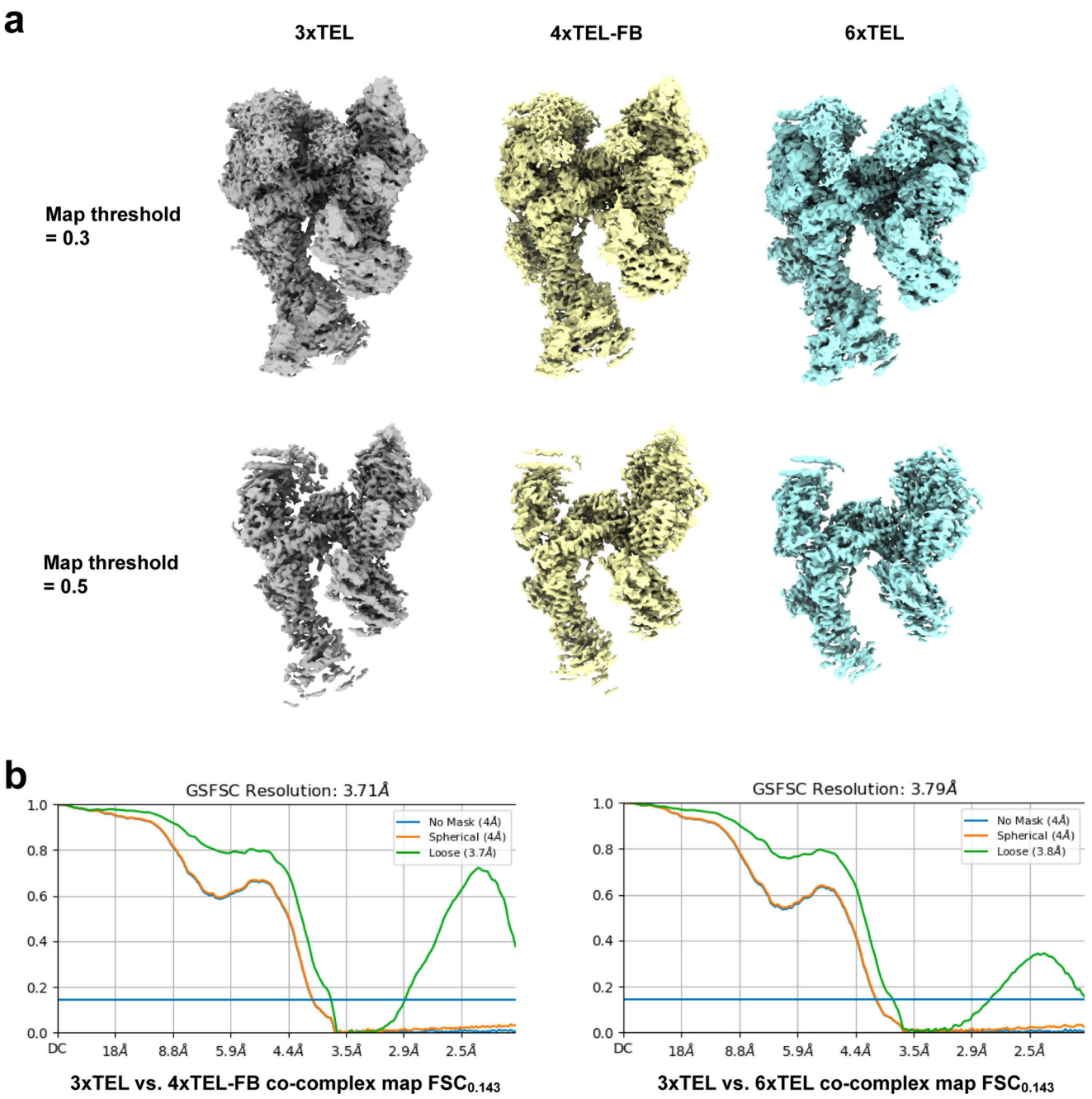
then using Strep beads to pull-down Strep-POLA1. FT: flowthrough. Trident (GeneTex) high-range prestained protein marker was used in both coomassie-stained denaturing protein gels. The results are reproducible across multiple independent experiments ($n > 5$). (**b**) Pull-down assay to characterize direct protein-protein interactions between CST and full-length or truncated ($\Delta POLA1_{1–337}$) pol-α/primase (ΔN-Pol-α/primase). Strep beads were used with pol-α/primase as bait and CST as prey. All input lanes are loaded as 10% of the total sample. Molecular weight markers are annotated for size reference. The pull-down results are reproducible across multiple independent experiments ($n = 3$). (**c**) Predicted RNA-DNA primers that are made by pol-α/primase using the repetitive telomeric DNA template in ATP-rich NTP reaction conditions. Primer product sizes are calculated assuming the polymerase reaches the end of the template. (**d**) Fold-stimulation analysis of direct enzyme assays done at predicted RNA-DNA primer product level (see Fig. 1c). CST does not stimulate pol-α/primase with the 3xTEL template, whether measured at the individual primer product level or total activity (see Fig. 1d). CST can stimulate pol-α/primase using the 6xTEL template. Data are plotted using bars for mean values and error bars for standard deviation as calculated from three independent experiments ($n = 3$). Data points from each experiment are represented by circle-filled markers in each corresponding bar. The dashed line marked fold-stimulation of one (no stimulation) and values were calculated by dividing the band intensity counts over the corresponding counts from the band and lane without CST added.



Extended Data Fig. 2 |. Negative-stain electron microscopy 2D classification analysis of CST-pol-α/primase co-complex bound to different telomeric ssDNA template lengths and structures.

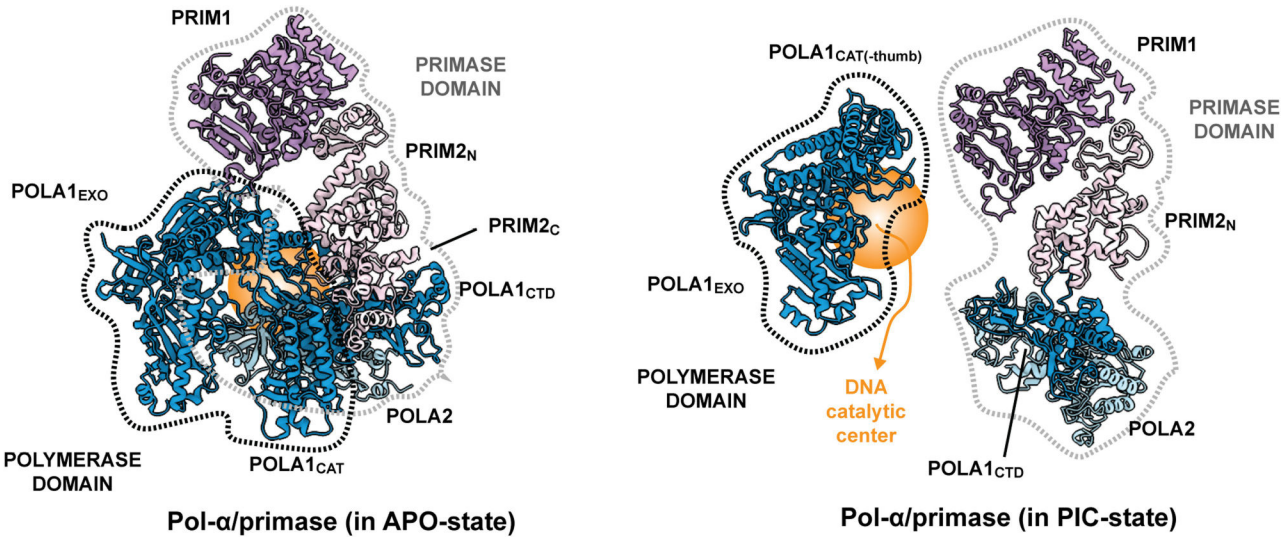
(**a**) Top ten (by distribution) class averages of CST-pol-α/primase preincubated with various lengths of telomeric single-stranded DNA, from 2.5 to 6 repeats of TTAGGG, and structures (ssDNA versus ss/dsDNA). 3xTEL-FB is a DNA template that has a 15 bp hairpin structure at the 5' end of the $GG(TTAGGG)_2TTAG$ overhang. (**b**) Top ten (by distribution) 2D class averages of each apo-state CST, pol-α/primase, and CST-pol-α/primase co-complex. No new PIC structure was observed in the absence of a template. Instead, individual CST or

pol-α/primase 2D class averages were seen in the co-complex analysis. CST-like and pol-α/primase class averages are pointed out by color-coded arrows for visual comparison.



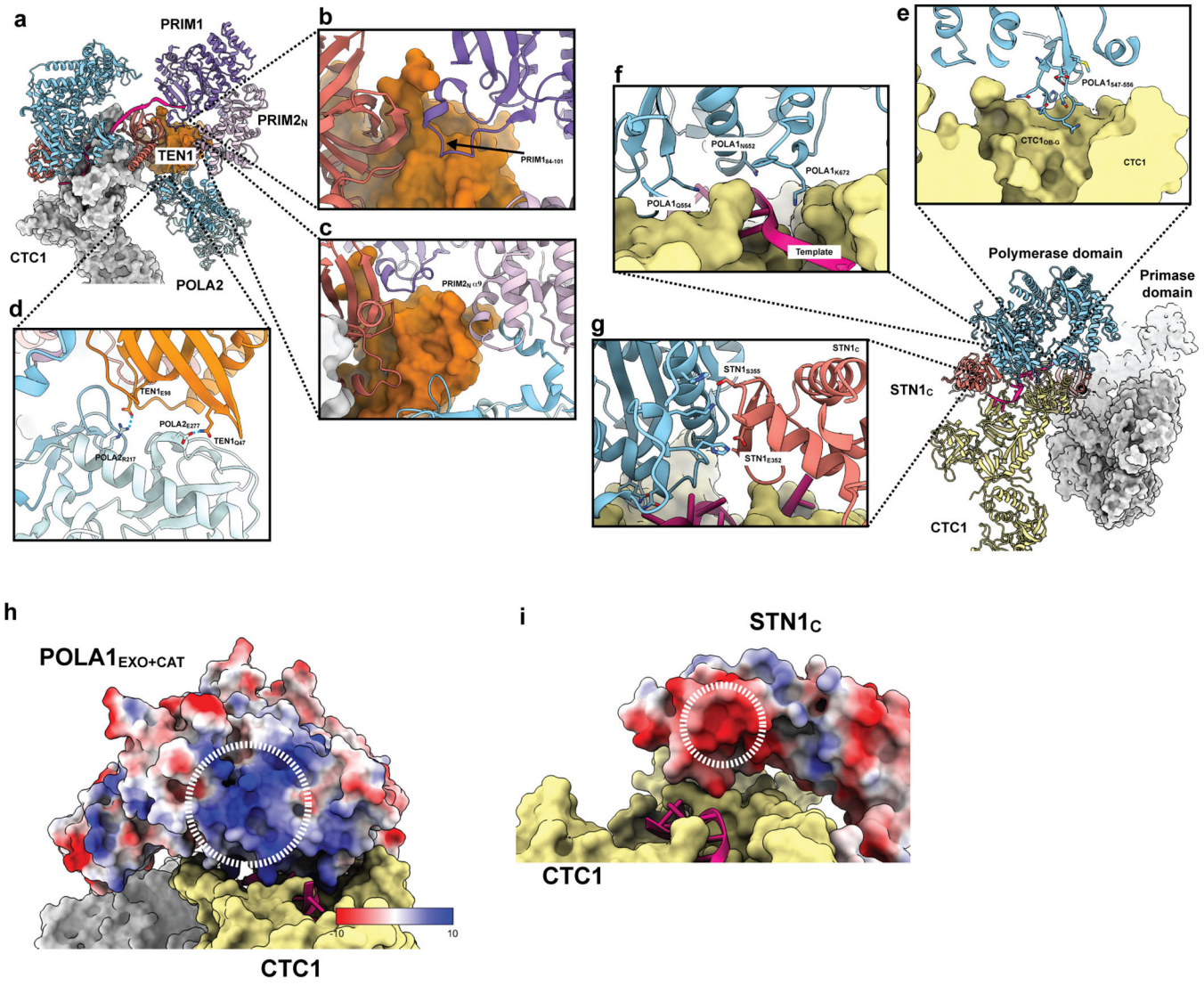
Extended Data Fig. 3 |. 3xTEL, 4xTEL-FB, and 6xTEL co-complexes share the same overall PIC structure.

(**a**) Comparison of the three co-complexes consensus cryo-EM maps at two different map thresholds. (**b**) Fourier-shell correlation (FSC) values of the 3xTEL PIC cryo-EM map aligned against the 4xTEL-FB map or 6xTEL map. The FSC values are reported at the level of 0.143.



Extended Data Fig. 4 |. Pol-α/primase conformation comparison in its apo-state vs. its preinitiation-state.

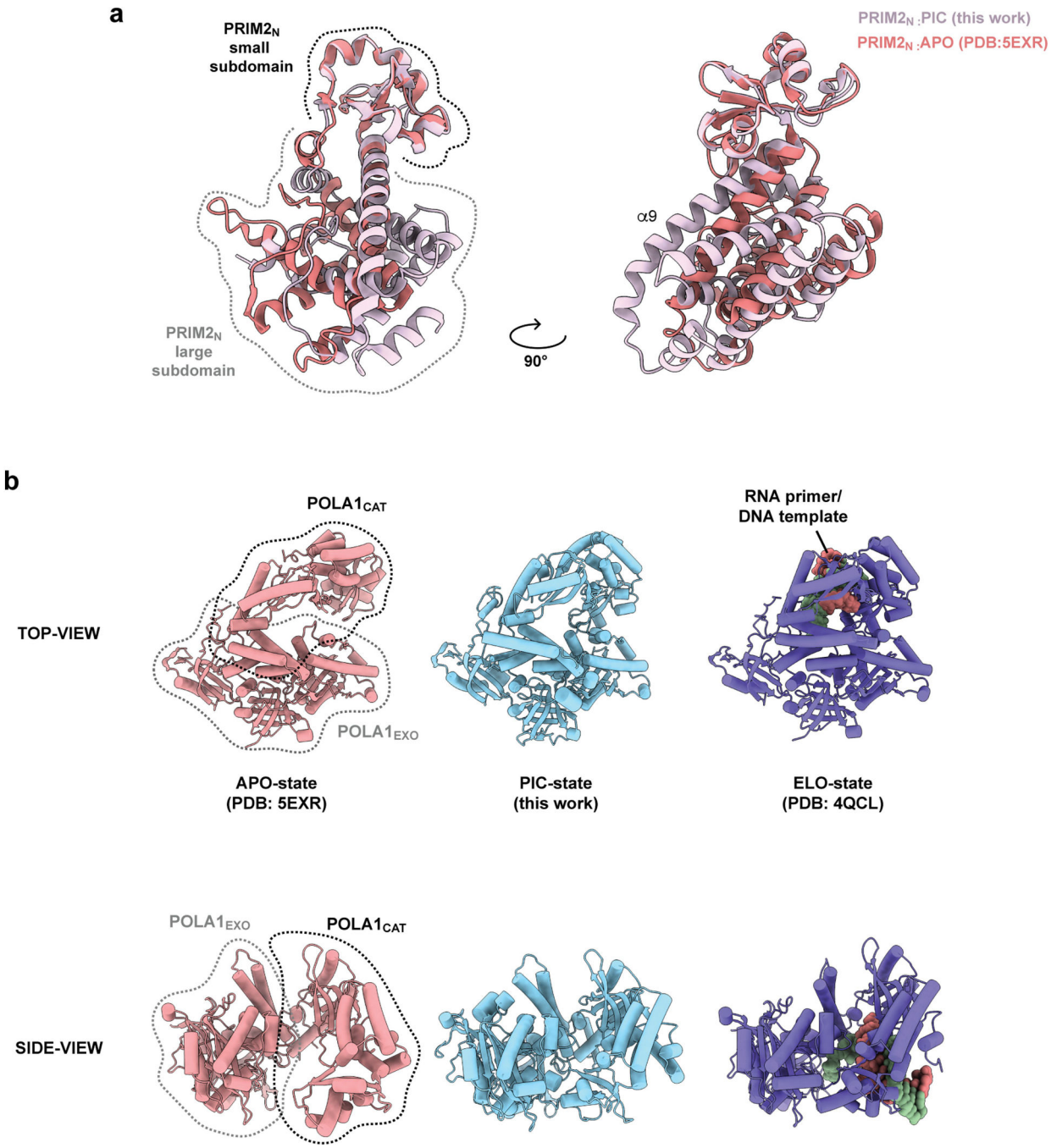
Pol-α/primase architecture changes from a compact shape to a segregated form upon forming the preinitiation complex (PIC) with a template-bound CST. The APO and PIC structures are depicted as ribbon cartoons. The CST complex and template are hidden in the PIC state to better demonstrate the segregated architecture of the PIC structure – separated into polymerase and primase domains. The lobe boundaries are loosely defined by the dashed lines – black for the polymerase domain and grey for the primase domain. The POLA1 DNA catalytic center is highlighted in orange to illustrate its accessibility upon forming the PIC. $POLA1_{CAT}$ thumb subdomain and $PRIM2_C$ domain are not illustrated in the PIC structure because of their flexibility (see main text).



Extended Data Fig. 5 |. CST and Pol-α/primase interactions.

(**a**) Overview of TEN1 interactions with primase domain. TEN1 is depicted as surface representation while the primase subunits are drawn as ribbons. (**b**) TEN1 and $STN1_N$ interface sits the $PRIM1_{84–101}$ loop. (**c**) TEN1 also made contacts with the alpha-helix nine (α9) of the $PRIM2_N$ domain. (**d**) Two hydrogen bonds were identified between TEN1 and POLA2 subunit; $TEN1_{E98}$ to $POLA2_{R217}$ and $TEN1_{Q47}$ to $POLA2_{E277}$. (**e**) The loop between residues 547–556 of POLA1 sits on the $CTC1_{OB-G}$. The POLA1 loop is depicted with heteroatom-coloured sidechains. (**f**) The template bound by CTC1 forms hydrogen bonds with POLA1 charged residues Q554, N652, and K672. (**g**) $STN1_C$ interacts with POLA1 using negatively charged residues E352 and S355. On the opposite surface, POLA1 uses the positively charged residues H627, H658, and K661. (**h**, **i**) The surfaces where POLA1 and $STN1_C$ contact has complementary charged residues. In panel **h**, the POLA1 $STN1_C$-interacting surface, encircled by a white dashed circle, is positively charged. The colour key for the coulombic electrostatic potential is shown at the bottom right of the panel. In panel **i**, for $STN1_C$, its POLA1-interacting surface, circle by a white dashed circle, is

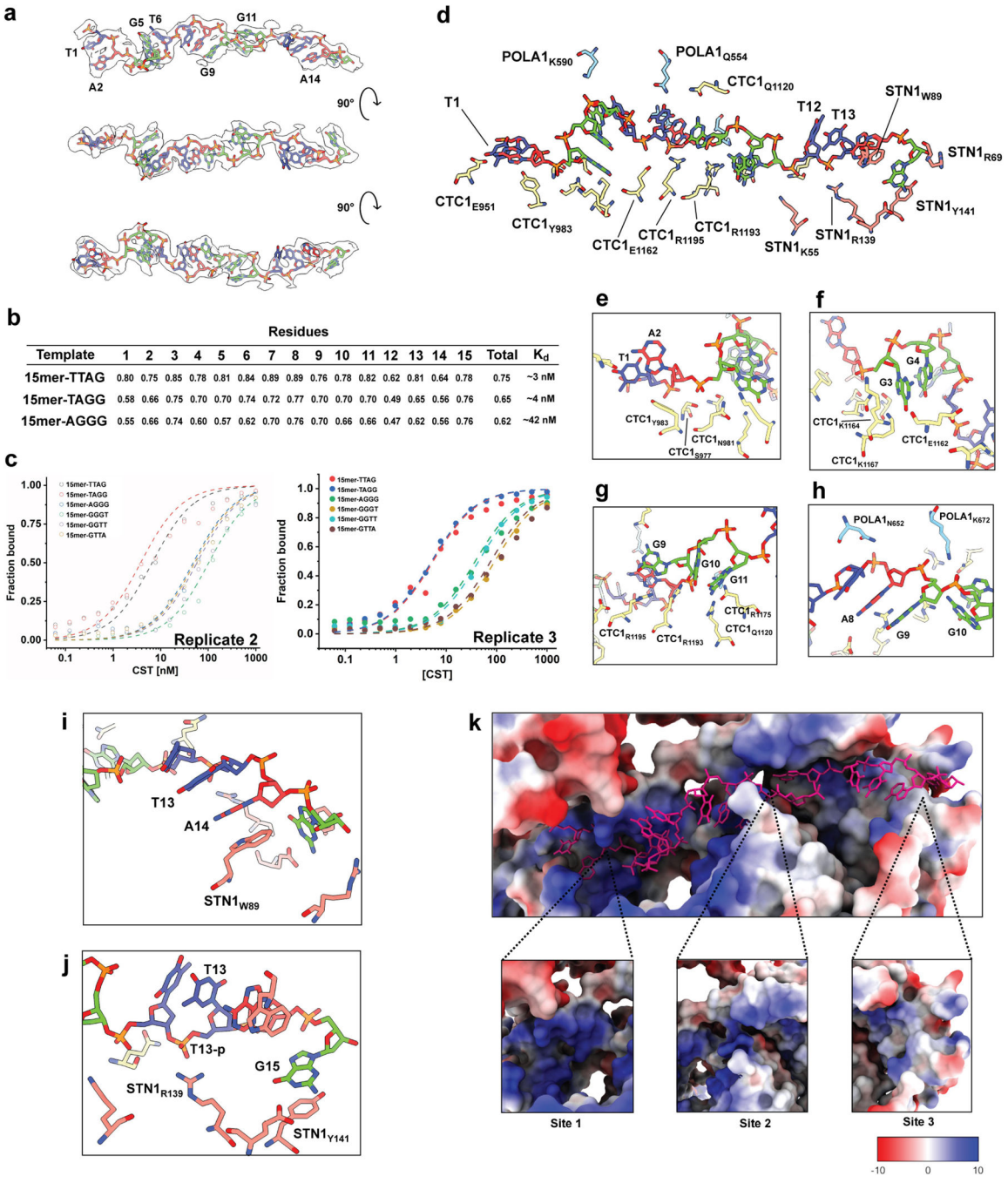
negatively charged. The white dashed circles represent the area of contact for the two protein subunits.



Extended Data Fig. 6 |. Comparisons of $PRIM2_N$ conformation in enzyme apo and preinitiation states and $POLA1_{EXO-CAT}$ domain conformation in the enzyme apo, preinitiation and elongation state.

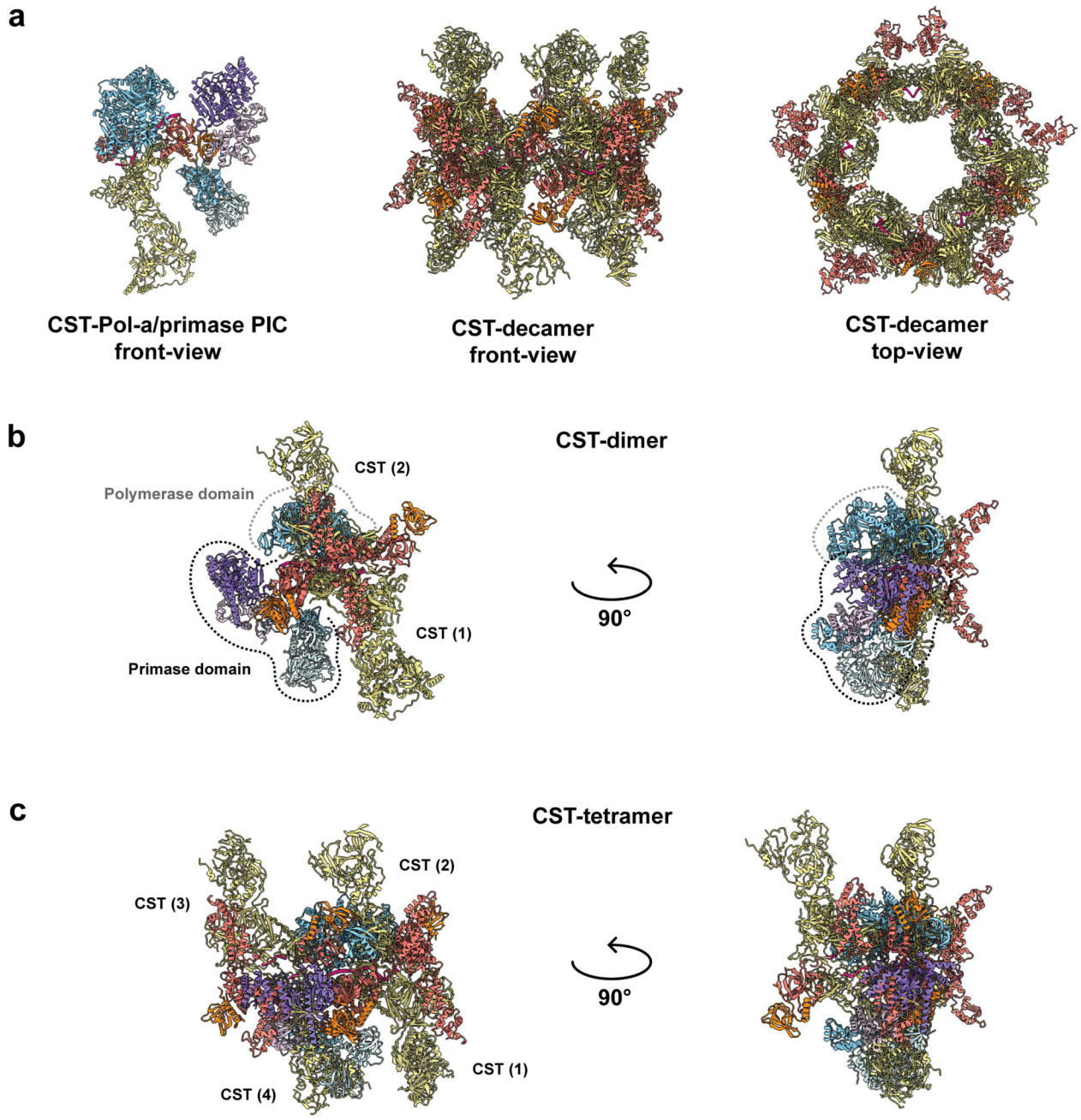
(**a**) The $PRIM2_N$ structure from our PIC model was compared against the pol-α/primase model in the APO-state (PDB: 5EXR). The $PRIM2_N$ small and large subdomains were the same for both enzymatic states but the domain relative position had rotated by about 40°.

The structures were aligned using the small subdomain to illustrate the large domain shift. (**b**) Top-down view of the POLA1$_{EXO-CAT}$ domain at various enzymatic states; apo (APO), preinitiation (PIC), and DNA elongation (ELO). To note, the APO- and ELO-state structures were obtained from either pol-α/primase (PDB: 5EXR) or POLA1 catalytic domain (PDB: 4QCL) alone. The PIC-state was derived from the CST-pol-α/primase co-complex structure. The RNA-DNA primer-template structure of the ELO-state is depicted as a space-filled model. The side-view perspective of each enzyme state is also shown for comparison.



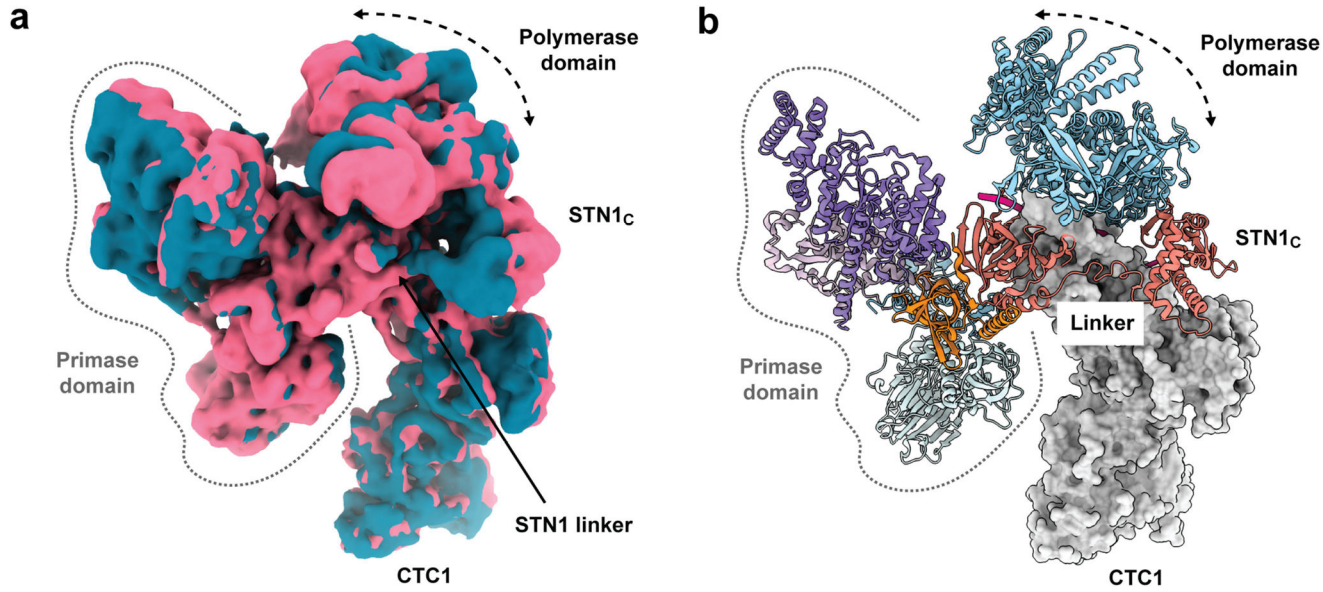
Extended Data Fig. 7 |. Analysis of PIC-template interactions.

(**a**) Model of the 15 nt telomeric sequence built into template cryo-EM density (shown as semi-transparent grey coloured volume). The model is depicted as an atomic model that is color-coded T=blue, A=red, and G=green. The modelled sequence is 5'-TAGGGTTAGGGTTAG – 3' (15mer-TTAG). (**b**) Model-to-map cross-correlation (CC) analysis reported at per residue level and chain average. The experimental binding affinities of CST to these templates are provided. The template 15mer-TAGG and 15mer-AGGG sequences are 5'- AGGGTTAGGGTTAGG – 3' and 5'- GGGTTAGGGTTAGGG – 3' respectively. (**c**) Independent experiments and quantification for Fig. 2b. See Supplementary Table 1 for template sequences. (**d**) Overview of protein-DNA interactions involved in template binding. (**e**, **f**) $CTC1_{OB\text{-}F}$ interactions with the template at the binding site 1. (**g**) CTC1OB-G residues binding to the template at binding site two are shown. (**h**) POLA1 residues binding to the template at binding site 2 are shown. (**i**) STN1 W89 base stacks with A14 of the template. This is supported by T13-T14 base stacking. (**j**) Other protein-DNA interactions between $STN1_N$ and template at binding site 3. STN1 R139 hydrogen bonds with T13 phosphate backbone and STN1 Y141 base stacks with G15 aromatic base. **(k)** The 15 nt DNA template model is depicted as a deep pink atomic model. Sites 1 and 2 are positively charged while site 3 is more neutral. Site 1 DNA-binding residues are mostly made from $CTC1_{OB\text{-}F}$, site 2 residues are from $CTC1_{OB\text{-}G}$ and POLA1, and site 3 residues are from the $STN1_N$ domain. The electrostatic potential colour key is in units of kcal/(mol·*e*) at 298 K.

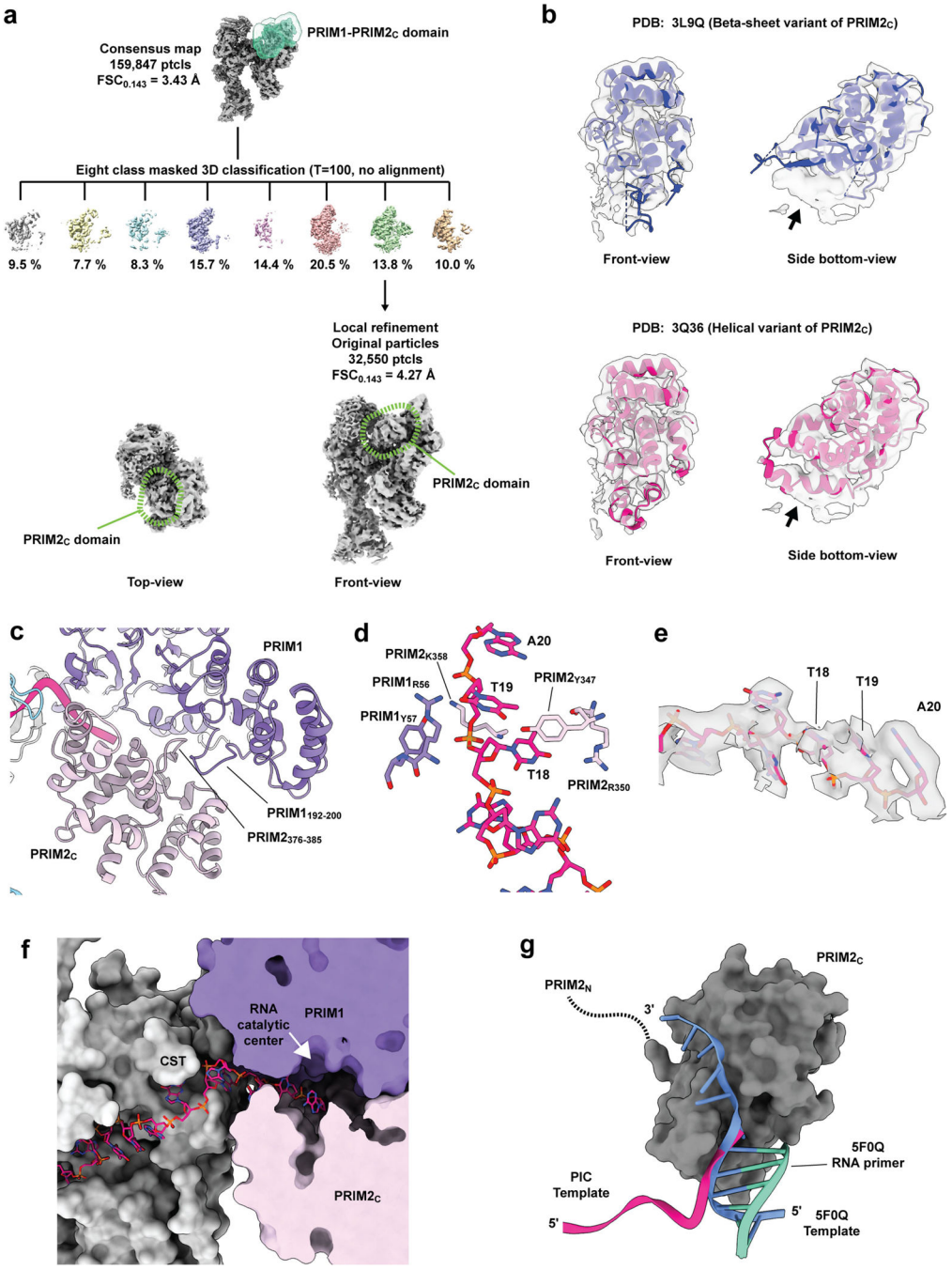


Extended Data Fig. 8 |. CST-pol-α/primase PIC structure is incompatible with CST decamer formation at early oligomerization states.

(**a**) Dimension comparison of the PIC model to the CST decamer model (ten copies of CST; PDB: 6W6W). (**b**) The PIC primase domain is not sterically hindered by the CST inverted dimer and thus could exist. However, its polymerase domain directly clashes with CST dimer formation. (**c**) The PIC primase and polymerase domains clash with CST tetramer formation. As in panel **b**, the polymerase domain clashes with the dimer formation, while the primase domain is now in the way of the fourth CST in the CST tetramer model. Subunits are coloured as in Fig. 1 for all panels.



Extended Data Fig. 9 |. Polymerase domain swivels on top of the template-bound CST.
(**a**) Three-dimensional (3D) variable analysis (3DVA) principal component analysis revealed a mode that showed the polymerase domain rotating about on the top of the template-bound CST. The first and last frames of the PCA mode are shown here as pink and navy blue cryo-EM maps. The cryo-EM maps are filtered to a resolution of 8 Å. The polymerase domain swivelling motion is seemly controlled by the STN1C domain and rotation angle restrained by the STN1 linker. (**b**) The corresponding components in A depicted on the PIC atomic model. CTC1 is represented as surface and coloured light grey. The rest of the PIC are drawn as ribbons. The color scheme follows Fig. 1.



Extended Data Fig. 10 |. Cryo-EM heterogeneity analysis reveals template-bound human primase structure at an advanced preinitiation state.

(**a**) The 4xTEL-FB PIC structure was masked with a PRIM1-PRIM2$_C$ for particle subtraction and 3D classification to isolate a subset of particles that allowed higher-resolution 3D reconstruction of the PRIM2$_C$ domain. An eight classes 3D classification step without image alignment was used to sieve out a conformation (~13.8%) that has an extra EM density in front of the PRIM1 subunit. This subtracted particles subset (32,550 particles) was reverted to their original images before subjecting to 3D refinement to obtain a 4.3 Å

cryo-EM map. The resolved $PRIM2_C$ domain density is circled in green dashed lines. (**b**) Two crystal structure variants of the $PRIM2_C$ domain are docked into the cryo-EM density that was resolved in Fig. 5 and panel **a**. The all-helical variant (PDB: 3Q36) is a better fit than the β-sheet-containing version (PDB: 3L9Q), as can be seen in the region pointed out by the black arrow. (**c**) PRIM1 and $PRIM2_C$ have minimal contacts except for two loop regions; $PRIM1_{192–200}$ and $PRIM2_{376–385}$. (**d**) PRIM1 and PRIM2 interactions with the template region into the primase crevice. (**e**) The five additional template residues were built based on the resolved primase density. The last three nucleotides are annotated. (**f**) The PRIM1 and $PRIM2_C$ domains form a crevice that the template sits in for RNA primer synthesis. The crevice has the PRIM1 RNA catalytic center facing the template. (**g**) The 3' end of the extended 20 nt template model in the PIC structure connects seamlessly with the template model in the previously solved RNA-DNA primer-template-bound $PRIM2_C$ crystal structure (PDB: 5F0Q).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank T. Tahirov and A. Baranovskiy from the University of Nebraska Medical Center for initial discussion and suggestions in setting up our experiments. We thank T. Cech and A. Zaug from the University of Colorado Boulder for help, discussion and suggestions. We also thank our departmental colleagues S. Butcher and J. Kimble for helpful input, discussion and critical reading of the manuscript. We also thank the members of the laboratory of C.L. for helpful discussions and suggestions. This research was, in part, supported by the Cryo-EM Research Center in the Department of Biochemistry at the University of Wisconsin–Madison and the National Cancer Institute's National Cryo-EM Facility at the Frederick National Laboratory for Cancer Research under contract 75N91019D00024. Support for this research was provided to C.L. by the National Institutes of Health, the National Institute of General Medical Sciences (R00GM131023) and the University of Wisconsin–Madison, Office of the Vice-Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation and the Department of Biochemistry.

Data availability

The described cryo-EM maps and coordinate files have been deposited in the Electron Microscopy Data Bank and the Protein Data Bank (PDB) under the accession codes EMD-27097 (co-complex–3×TEL-consensus); EMD-27099 (co-complex–3×TEL-local-refined-merged); EMD-27104, PDB ID 8D0B (co-complex–4×TEL-local-refined-merged and consensus maps and model); EMD-27107, PDB ID 8D0K (co-complex–4×TEL–$PRIM2_C$-advanced PIC-local-refined-merged and consensus maps and model); and EMD-27109 (co-complex–6×TEL-local-refined-merged and consensus). The coordinates were assembled and built from individual predicted models from the AlphaFold Protein Structure Database https://alphafold.ebi.ac.uk/.

References

1. Attali I, Botchan MR & Berger JM Structural mechanisms for replicating DNA in eukaryotes. Annu. Rev. Biochem. 90, 77–106 (2021). [PubMed: 33784179]
2. Baranovskiy AG & Tahirov TH Elaborated action of the human primosome. Genes 8, 62 (2017). [PubMed: 28208743]

3. Lim C & Cech TR Shaping human telomeres: from shelterin and CST complexes to telomeric chromatin organization. Nat. Rev. Mol. Cell Biol. 22, 283–298 (2021). [PubMed: 33564154]
4. Bonnell E, Pasquier E & Wellinger RJ Telomere replication: solving multiple end replication problems. Front. Cell Dev. Biol. 9, 668171 (2021). [PubMed: 33869233]
5. Pellegrini L The Pol alpha-primase complex. Subcell. Biochem. 62, 157–169 (2012). [PubMed: 22918585]
6. Braun KA, Lao Y, He Z, Ingles CJ & Wold MS Role of protein-protein interactions in the function of replication protein A (RPA): RPA modulates the activity of DNA polymerase alpha by multiple mechanisms. Biochemistry 36, 8443–8454 (1997). [PubMed: 9214288]
7. Casteel DE et al. A DNA polymerase-α·primase cofactor with homology to replication protein A-32 regulates DNA replication in mammalian cells. J. Biol. Chem. 284, 5807–5818 (2009). [PubMed: 19119139]
8. Miyake Y et al. RPA-like mammalian Ctc1-Stn1-Ten1 complex binds to single-stranded DNA and protects telomeres independently of the Pot1 pathway. Mol. Cell 36, 193–206 (2009). [PubMed: 19854130]
9. Surovtseva YV et al. Conserved telomere maintenance component 1 interacts with STN1 and maintains chromosome ends in higher eukaryotes. Mol. Cell 36, 207–218 (2009). [PubMed: 19854131]
10. Lim C et al. The structure of human CST reveals a decameric assembly bound to telomeric DNA. Science 368, 1081–1085 (2020). [PubMed: 32499435]
11. Chen LY, Redon S & Lingner J The human CST complex is a terminator of telomerase activity. Nature 488, 540–544 (2012). [PubMed: 22763445]
12. Zaug AJ et al. CST does not evict elongating telomerase but prevents initiation by ssDNA binding. Nucleic Acids Res. 49, 11653–11665 (2021). [PubMed: 34718732]
13. Kelich JM, Papaioannou H & Skordalakes E Pol alpha-primase dependent nuclear localization of the mammalian CST complex. Commun. Biol. 4, 349 (2021). [PubMed: 33731801]
14. Wang F et al. Human CST has independent functions during telomere duplex replication and C-strand fill-in. Cell Rep. 2, 1096–1103 (2012). [PubMed: 23142664]
15. Stewart JA et al. Human CST promotes telomere duplex replication and general replication restart after fork stalling. EMBO J. 31, 3537–3549 (2012). [PubMed: 22863775]
16. Mirman Z et al. 53BP1–RIF1–shieldin counteracts DSB resection through CST- and Polα-dependent fill-in. Nature 560, 112–116 (2018). [PubMed: 30022158]
17. Gu P et al. CTC1-STN1 coordinates G- and C-strand synthesis to regulate telomere length. Aging Cell 17, e12783 (2018). [PubMed: 29774655]
18. Gu P & Chang S Functional characterization of human CTC1 mutations reveals novel mechanisms responsible for the pathogenesis of the telomere disease Coats plus. Aging Cell 12, 1100–1109 (2013). [PubMed: 23869908]
19. Chen LY, Majerska J & Lingner J Molecular basis of telomere syndrome caused by CTC1 mutations. Genes Dev. 27, 2099–2108 (2013). [PubMed: 24115768]
20. Zhu W et al. Mcm10 and And-1/CTF4 recruit DNA polymerase alpha to chromatin for initiation of DNA replication. Genes Dev. 21, 2288–2299 (2007). [PubMed: 17761813]
21. Cai SW et al. Cryo-EM structure of the human CST-Polalpha/primase complex in a recruitment state. Nat. Struct. Mol. Biol. 10.1038/s41594-022-00766-y (2022).
22. Hom RA & Wuttke DS Human CST prefers G-rich but not necessarily telomeric sequences. Biochemistry 56, 4210–4218 (2017). [PubMed: 28726394]
23. Bhattacharjee A, Wang Y, Diao J & Price CM Dynamic DNA binding, junction recognition and G4 melting activity underlie the telomeric and genome-wide roles of human CST. Nucleic Acids Res. 45, 12311–12324 (2017). [PubMed: 29040642]
24. Jumper J et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021). [PubMed: 34265844]
25. Baranovskiy AG et al. Mechanism of concerted RNA-DNA primer synthesis by the human primosome. J. Biol. Chem. 291, 10006–10020 (2016). [PubMed: 26975377]

26. Kilkenny ML, Longo MA, Perera RL & Pellegrini L Structures of human primase reveal design of nucleotide elongation site and mode of Pol alpha tethering. Proc. Natl Acad. Sci. USA 110, 15961–15966 (2013). [PubMed: 24043831]
27. Coloma J, Johnson RE, Prakash L, Prakash S & Aggarwal AK Human DNA polymerase alpha in binary complex with a DNA:DNA template-primer. Sci. Rep. 6, 23784 (2016). [PubMed: 27032819]
28. Perera RL et al. Mechanism for priming DNA synthesis by yeast DNA polymerase alpha. Elife 2, e00482 (2013). [PubMed: 23599895]
29. Baranovskiy AG et al. Crystal structure of the human primase. J. Biol. Chem. 290, 5635–5646 (2015). [PubMed: 25550159]
30. Baranovskiy AG et al. Activity and fidelity of human DNA polymerase alpha depend on primer structure. J. Biol. Chem. 293, 6824–6843 (2018). [PubMed: 29555682]
31. Baranovskiy AG et al. Structural basis for inhibition of DNA replication by aphidicolin. Nucleic Acids Res. 42, 14013–14021 (2014). [PubMed: 25429975]
32. Baranovskiy AG et al. Insight into the human DNA primase interaction with template-primer. J. Biol. Chem. 291, 4793–4802 (2016). [PubMed: 26710848]
33. Ashkenazy H et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 44, W344–W350 (2016). [PubMed: 27166375]
34. Lei M, Podell ER & Cech TR Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. Nat. Struct. Mol. Biol. 11, 1223–1229 (2004). [PubMed: 15558049]
35. Fan J & Pavletich NP Structure and conformational change of a replication protein A heterotrimer bound to ssDNA. Genes Dev. 26, 2337–2347 (2012). [PubMed: 23070815]
36. Bhattacharjee A, Stewart J, Chaiken M & Price CM STN1 OB fold mutation alters DNA binding and affects selective aspects of CST function. PLoS Genet. 12, e1006342 (2016). [PubMed: 27690379]
37. He Y, Song H, Chan H et al. Structure of *Tetrahymena* telomerase-bound CST with polymerase α-primase. Nature 10.1038/s41586-022-04931-7 (2022).
38. Goulian M, Heard CJ & Grimm SL Purification and properties of an accessory protein for DNA polymerase alpha/primase. J. Biol. Chem. 265, 13221–13230 (1990). [PubMed: 2165497]
39. Goulian M & Heard CJ The mechanism of action of an accessory protein for DNA polymerase alpha/primase. J. Biol. Chem. 265, 13231–13239 (1990). [PubMed: 2376593]
40. Zaug AJ, Goodrich KJ, Song JJ et al. Reconstitution of a telomeric replicon organized by CST. Nature 10.1038/s41586-022-04930-8 (2022).
41. Zhang M et al. Mammalian CST averts replication failure by preventing G-quadruplex accumulation. Nucleic Acids Res. 47, 5243–5259 (2019). [PubMed: 30976812]
42. Punjani A & Fleet DJ 3D variability analysis: resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. J. Struct. Biol. 213, 107702 (2021). [PubMed: 33582281]
43. Zerbe LK & Kuchta RD The p58 subunit of human DNA primase is important for primer initiation, elongation, and counting. Biochemistry 41, 4891–4900 (2002). [PubMed: 11939784]
44. Scheres SH et al. Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. Nat. Methods 4, 27–29 (2007). [PubMed: 17179934]
45. Punjani A, Zhang H & Fleet DJ Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. Nat. Methods 17, 1214–1221 (2020). [PubMed: 33257830]
46. Agarkar VB, Babayeva ND, Pavlov YI & Tahirov TH Crystal structure of the C-terminal domain of human DNA primase large subunit: implications for the mechanism of the primase-polymerase alpha switch. Cell Cycle 10, 926–931 (2011). [PubMed: 21346410]
47. Vaithiyalingam S, Warren EM, Eichman BF & Chazin WJ Insights into eukaryotic DNA priming from the structure and functional interactions of the 4Fe-4S cluster domain of human DNA primase. Proc. Natl Acad. Sci. USA 107, 13684–13689 (2010). [PubMed: 20643958]

48. Sheaff RJ & Kuchta RD Mechanism of calf thymus DNA primase: slow initiation, rapid polymerization, and intelligent termination. Biochemistry 32, 3027–3037 (1993). [PubMed: 7681326]

49. Yan J, Holzer S, Pellegrini L & Bell SD An archaeal primase functions as a nanoscale caliper to define primer length. Proc. Natl Acad. Sci. USA 115, 6697–6702 (2018). [PubMed: 29891690]

50. Greci MD, Dooher JD & Bell SD The combined DNA and RNA synthetic capabilities of archaeal DNA primase facilitate primer hand-off to the replicative DNA polymerase. Nat. Commun. 13, 433 (2022). [PubMed: 35064114]

51. Rames M, Yu Y & Ren G Optimized negative staining: a high-throughput protocol for examining small and asymmetric protein structure by electron microscopy. J. Vis. Exp. 10.3791/51087 (2014).

52. Mastronarde DN Automated electron microscope tomography using robust prediction of specimen movements. J. Struct. Biol. 152, 36–51 (2005). [PubMed: 16182563]

53. Punjani A, Rubinstein JL, Fleet DJ & Brubaker MA cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods 14, 290–296 (2017). [PubMed: 28165473]

54. Chen J, Noble AJ, Kang JY & Darst SA Eliminating effects of particle adsorption to the air/water interface in single-particle cryo-electron microscopy: bacterial RNA polymerase and CHAPSO. J. Struct. Biol. X 1, 100005 (2019). [PubMed: 32285040]

55. Zivanov J, Nakane T & Scheres SHW A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. IUCrJ 6, 5–17 (2019).

56. Rohou A & Grigorieff N CTFFIND4: fast and accurate defocus estimation from electron micrographs. J. Struct. Biol. 192, 216–221 (2015). [PubMed: 26278980]

57. Bepler T et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. Nat. Methods 16, 1153–1160 (2019). [PubMed: 31591578]

58. Zivanov J et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. Elife 7, e42166 (2018). [PubMed: 30412051]

59. Pettersen EF et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci. 30, 70–82 (2021). [PubMed: 32881101]

60. Goddard TD et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. Protein Sci. 27, 14–25 (2018). [PubMed: 28710774]

61. Emsley P, Lohkamp B, Scott WG & Cowtan K Features and development of Coot. Acta Crystallogr. D 66, 486–501 (2010). [PubMed: 20383002]

62. Afonine PV et al. Real-space refinement in PHENIX for cryo-EM and crystallography. Acta Crystallogr. D 74, 531–544 (2018).

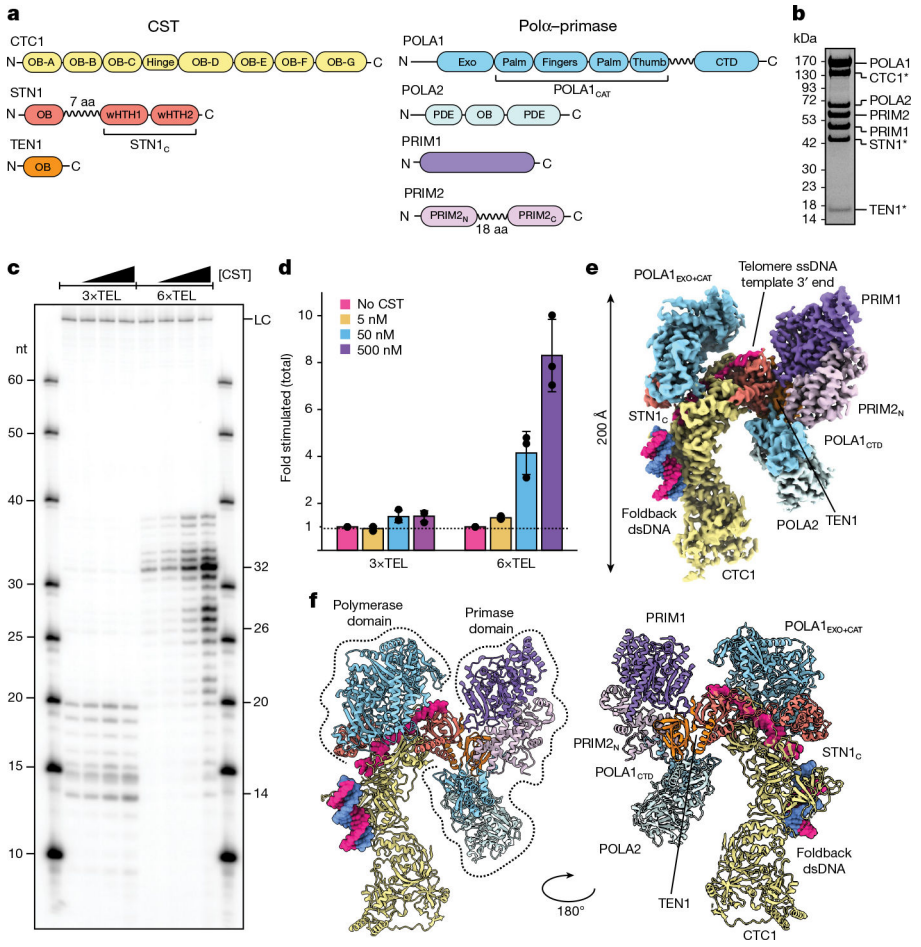


Fig. 1 |. The architecture of the human CST-Polα–primase PIC bound to a telomere overhang. **a**, The structures of the CST and Polα–primase domains. The zigzag lines indicate flexible linkers connecting domains. aa, amino acids. **b**, Coomassie-stained denaturing protein gel of 20 μg of the purified reconstituted CST-Polα–primase co-complex. The labels marked with asterisks are the CST subunits; all other labels are the Polα–primase subunits. Molecular mass markers are shown on the left. **c**, Direct assay of Polα–primase using telomeric templates with the titration of CST (0, 5, 50 and 500 nM) for enzyme stimulation. The first and last lanes are labelled oligonucleotide ladders. The numbers on the left are oligonucleotide ladder sizes; the numbers on the right are predicted sizes of RNA–DNA primers made by the enzyme using the repetitive telomeric sequence templates (see Extended Data Fig. 1). **d**, Total lane activity quantified from **c**. The bars represent the mean values and the error bars show the standard deviation as calculated from several independent experiments ($n = 3$). Data points from each experiment are represented by circles in each corresponding bar. The dotted line marks the fold stimulation of 1 (no stimulation), and the values were calculated by dividing the raw counts from each relevant lane by the counts without CST added. **e**, The front view of the cryo-EM map of the CST-Polα–primase PIC bound to the 4×TEL-FB template. **f**, A ribbon representation of the atomic model built from the cryo-EM map in **e**. For **e**,**f**, each domain is coloured as in **a**. The foldback dsDNA in **e**,**f** is modelled on the basis of the cryo-EM map in Fig. 4a.

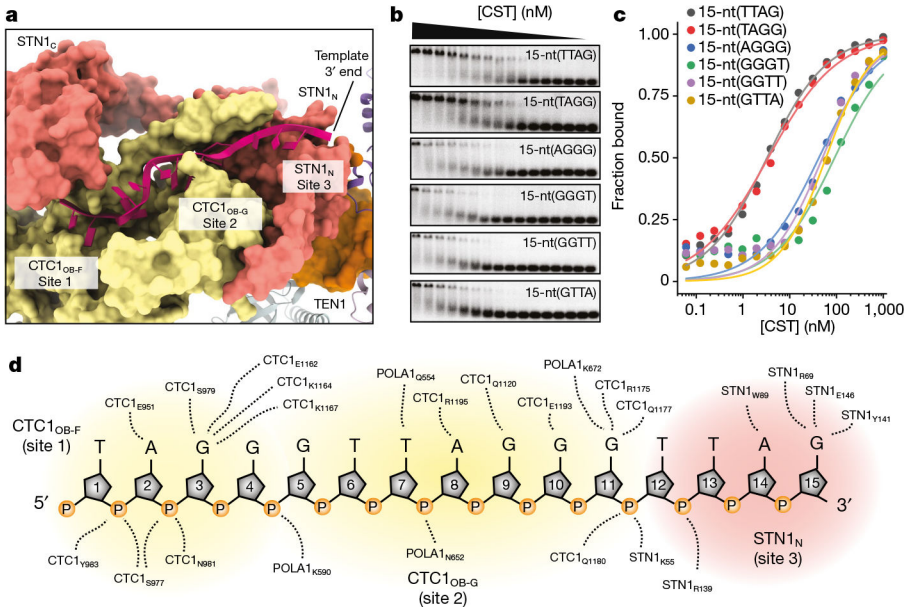


Fig. 2 |. The molecular and structural basis for the binding preference of CST to telomeric sequences.

a, Template bound by the CTC1 and STN1 subunits of CST in the PIC structure. CTC1, STN1 and TEN1 are depicted as surfaces; the rest of the PIC structure is represented as ribbon diagrams. The 15-nt modelled template, coloured deep pink and illustrated as filled nucleotides, is bound by three binding sites spanning the $CTC1_{OB\text{-}F}$, $CTC1_{OB\text{-}G}$ and $STN1_N$ domains. These binding sites are labelled sites 1, 2 and 3, respectively. **b**, Gel electrophoresis mobility shift assays to determine the binding affinity of CST to six different permutations of a 15-nt telomeric sequence. The highest CST concentration used was 1 μM and a twofold serial dilution was used for the protein titration. **c**, Binding analysis and curve fitting of the electrophoresis mobility shift assay data in **b**. The circles are the data points for the fraction bound; the lines are fitted curves. Two other independent experiments are shown in Extended Data Fig. 7c. **d**, Protein–DNA interactions as shown in **a**.

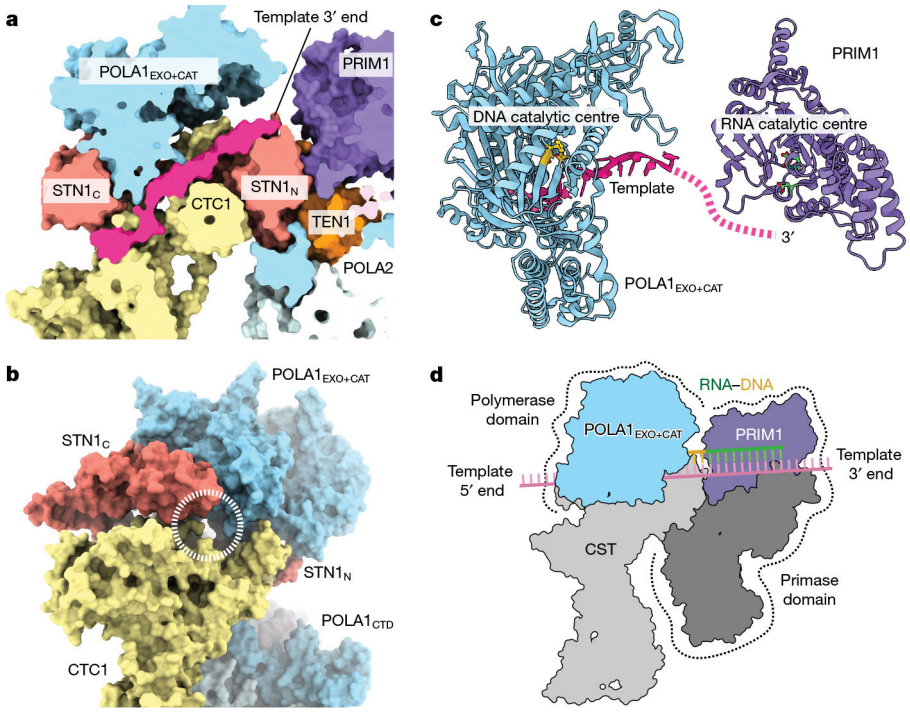


Fig. 3 |. CST arranges the catalytic centres of Polα–primase in RNA–DNA synthesis order along the template.

a, A cross-sectional view of the molecular tunnel formed by CST and the polymerase domain, with the ssDNA (deep pink) template threading through. **b**, The molecular tunnel viewed from the 5′ end of the template but with the template model hidden to reveal the tight-fitting tunnel. The entrance of the tunnel is marked by a white dashed circle. **c**, CST separation of the polymerase and primase domains of Polα–primase puts the RNA and DNA catalytic centres in the correct RNA–DNA primer synthesis order. Catalytic side chains are shown as green and yellow for the RNA and DNA catalytic centres, respectively. Only the $POLA1_{EXO+CAT}$ domain, PRIM1 subunit and template are shown. The rest of the PIC structure is hidden for clarity. A dashed deep pink line is drawn to illustrate a longer template reaching across PRIM1. **d**, A model illustrating how CST segregates Polα–primase into the polymerase and primase domains (dashed lines) and arranges them along an extended DNA template in the RNA-to-DNA synthesis order.

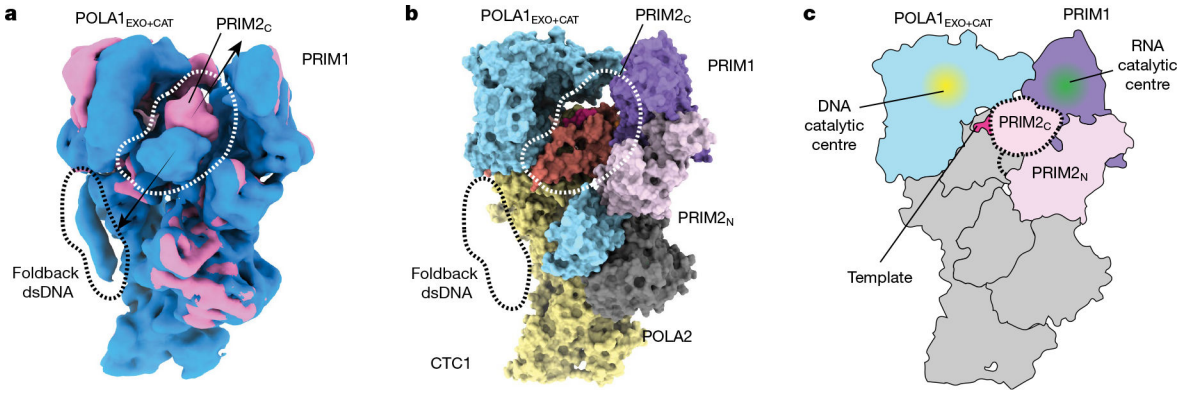


Fig. 4 |. PRIM2$_C$ is flexibly poised to perform de novo RNA primer synthesis and facilitate primer handover.

a, A 3DVA end-to-end model of the PIC structure revealing flexibility in the EM density of PRIM2$_C$ and the foldback dsDNA. The first and last frames (EM maps) of the principal component analysis mode are shown as blue and pink surfaces. The black dashed line encircles the dsDNA region of motion; the white dashed line represents the boundary of the PRIM2$_C$ domain motion. The black arrows represent the direction of motion of the PRIM2$_C$ domain. These maps are filtered at a nominal resolution of 8 Å. **b**, A direct comparison of the consensus PIC model, depicted as a surface representation, with the flexibility range of the foldback dsDNA and PRIM2$_C$ domain overlaid. **c**, A model of the PIC structure illustrating the proximity of the CST-bound template and the RNA and DNA catalytic centres of Polα–primase to the flexible PRIM2$_C$ domain (dashed outline). The template is coloured deep pink; the RNA and DNA catalytic centres are highlighted with green and yellow spots, respectively. The rest of the PIC is coloured grey.

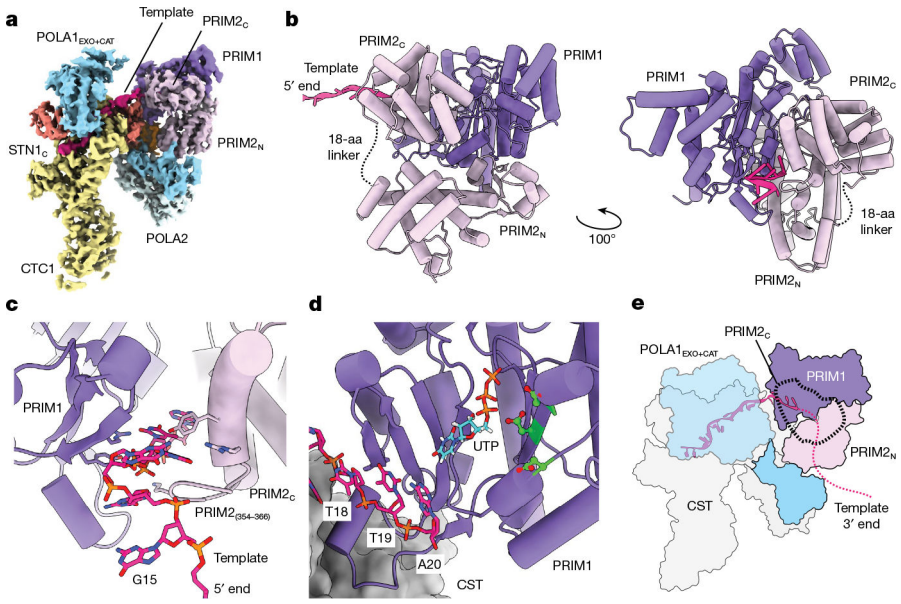


Fig. 5 |. The human primase structure that is poised for RNA primer synthesis.

a, The cryo-EM structure of a PIC structure trapped in an advanced preinitiation state, as compared to the consensus PIC structure in Fig. 1e. The PRIM2$_C$ domain density is sufficiently well resolved to be docked with its atomic model. **b**, A model of the primase structure poised to initiate RNA primer synthesis. The template (deep pink) is shown as a ladder of nucleotides; the PRIM1 and PRIM2 subunits are depicted as ribbon–tube helix representations. The rest of the PIC structure is hidden for visual clarity. **c**, Template entry into the PRIM1–PRIM2$_C$ crevice. The side chains of protein residues involved in template interactions are shown. The template and side chains are coloured by heteroatom. **d**, The template reaches in front of RNA catalytic residues of PRIM1 (side chains shown in green and coloured by heteroatom). The 3′-end residue, A20, of the ssDNA template is in position for base pairing with a UTP RNA molecule bound by the RNA catalytic centre of PRIM1. The UTP molecule is from the 4BPW model in the Protein Data Bank[26] and is superimposed on our structure to illustrate its proximity to our template model. **e**, A model illustrating how the template is bound by CST-Polα–primase in an advanced preinitiation state poised for NTP incorporation. The 5′-end region of the template is stably bound by the CST-polymerase domain (the polymerase domain is on top of the template and thus is shown semi-transparent for visual clarity) and its 3′ end is engaged by the primase (PRIM1–PRIM2$_C$) for primer initiation. The template makes a sharp turn on contacting the primase. The black dashed circle depicts the PRIM2$_C$ position. The deep pink dotted line represents the predicted extended pathway of a longer template. The POLA1$_{CTD}$ domain is drawn as a solid sky-blue shape.