

# The architecture of a eukaryotic replisome

Jingchuan Sun<sup>1,2</sup>, Yi Shi<sup>3</sup>, Roxana E Georgescu<sup>3,4</sup>, Zuanning Yuan<sup>1,2</sup>, Brian T Chait<sup>3</sup>, Huilin Li<sup>1,2</sup> & Michael E O'Donnell<sup>3,4</sup>

At the eukaryotic DNA replication fork, it is widely believed that the Cdc45–Mcm2–7–GINS (CMG) helicase is positioned in front to unwind DNA and that DNA polymerases trail behind the helicase. Here we used single-particle EM to directly image a *Saccharomyces cerevisiae* replisome. Contrary to expectations, the leading strand Pol  $\epsilon$  is positioned ahead of CMG helicase, whereas Ctf4 and the lagging-strand polymerase (Pol)  $\alpha$ –primase are behind the helicase. This unexpected architecture indicates that the leading-strand DNA travels a long distance before reaching Pol  $\epsilon$ , first threading through the Mcm2–7 ring and then making a U-turn at the bottom and reaching Pol  $\epsilon$  at the top of CMG. Our work reveals an unexpected configuration of the eukaryotic replisome, suggests possible reasons for this architecture and provides a basis for further structural and biochemical replisome studies.

DNA is replicated by a multiprotein machinery referred to as a replisome<sup>1,2</sup>. Replisomes contain a helicase that unwinds DNA, DNA polymerases that synthesize the leading and lagging strands, and a primase that makes short primed sites that facilitate initiation of DNA synthesis on both strands. The eukaryotic helicase is an 11-subunit CMG complex that encircles the leading strand during unwinding and consists of the Mcm2–7 motor subunits, the GINS heterotetramer and the Cdc45 subunit<sup>3–5</sup>. In eukaryotes, the leading and lagging strands are replicated by Pol  $\epsilon$  and Pol  $\delta$  (refs. 6,7), and the lagging strand is repeatedly primed by Pol  $\alpha$ , a polymerase–primase that synthesizes short RNA–DNA primers<sup>1,8</sup>. Several studies have indicated that Pol  $\epsilon$  is responsible for bulk synthesis of the leading strand and that Pol  $\delta$  performs bulk synthesis of the lagging strand<sup>6,9–12</sup>. Consistently with a dedicated role of Pol  $\epsilon$  on the leading strand, Pol  $\epsilon$  binds directly to CMG helicase, which encircles the leading strand<sup>13</sup>. In addition, a trimer of the Ctf4 protein binds both CMG and Pol  $\alpha$  (refs. 14,15). Structures of replisome components, such as the Pol  $\epsilon$  catalytic domain, Pol  $\delta$  and Pol  $\alpha$  have been reported<sup>16–18</sup>, and EM three-dimensional (3D) reconstruction has outlined the structure of *Drosophila melanogaster* (*Dm*) CMG<sup>19,20</sup>.

Owing to the low cellular abundance and dynamic nature of replisomes, it has not been feasible to purify an intact and homogeneous replisome complex for structural characterization in any system, prokaryotic or eukaryotic. DNA polymerases require single-stranded (ss) DNA to function, and thus it has long been assumed that the leading- and lagging-strand polymerases trail behind the helicase<sup>1,2</sup>. Interestingly, the loading process of the Mcm2–7 hexamer (i.e., the CMG helicase motor proteins) onto the replication origin is also a dynamic process that has been recalcitrant to structural characterization. Our recent work has captured several intermediate structures of single and double Mcm2–7 hexamers loading onto origin DNA,

by single-particle EM<sup>21,22</sup>. We therefore applied the same approach in this report to progressively build up and elucidate the eukaryotic replisome structure.

## RESULTS

### 3D reconstruction of ScCMG

We first imaged the purified *S. cerevisiae* (*Sc*) CMG helicase 11-subunit complex in the presence of a small 80- and 75-mer (80/75-mer) forked DNA to which CMG binds under the conditions used in this study (Supplementary Fig. 1a,b). The ScCMG EM structure (Fig. 1) is remarkably similar to that of the previously reported *Dm*CMG<sup>19,20</sup>. We confirmed the positional assignments of CMG subunits by chemical cross-linking in combination with mass spectrometric readout, as discussed below. The subunits follow the same arrangement as those determined for the *Dm*CMG<sup>20</sup>. The structure of the Mcm2–7 hexamer in the ScCMG complex agrees with the Mcm2–7 structures of single and double Mcm2–7 hexamers loaded onto duplex DNA<sup>21,22</sup>. Further, the assigned GINS and Cdc45 densities (Fig. 1) are consistent with the homolog crystal structures outlined for *Dm*CMG<sup>20</sup>. An additional density, unoccupied by the human GINS structure, may belong to Cdc45 or the C-terminal domain (CTD) of Psf1, which is too mobile to visualize in the GINS crystal structure<sup>23–25</sup>, as suggested in the *Dm*CMG analysis<sup>20</sup>. The interface of GINS–Cdc45 with Mcm2–Mcm5 appears to form a second hole, or channel, through ScCMG, as noted for *Dm*CMG<sup>19</sup> (Fig. 1a).

### 3D reconstruction of CMGE

We have recently demonstrated that Pol  $\epsilon$  binds ScCMG, thus forming a complex denoted CMGE<sup>13</sup>. Addition of ScPol  $\epsilon$  to ScCMG revealed particles containing an extra density unmistakably belonging to Pol  $\epsilon$ , in addition to the characteristic CMG particle, in averaged EM

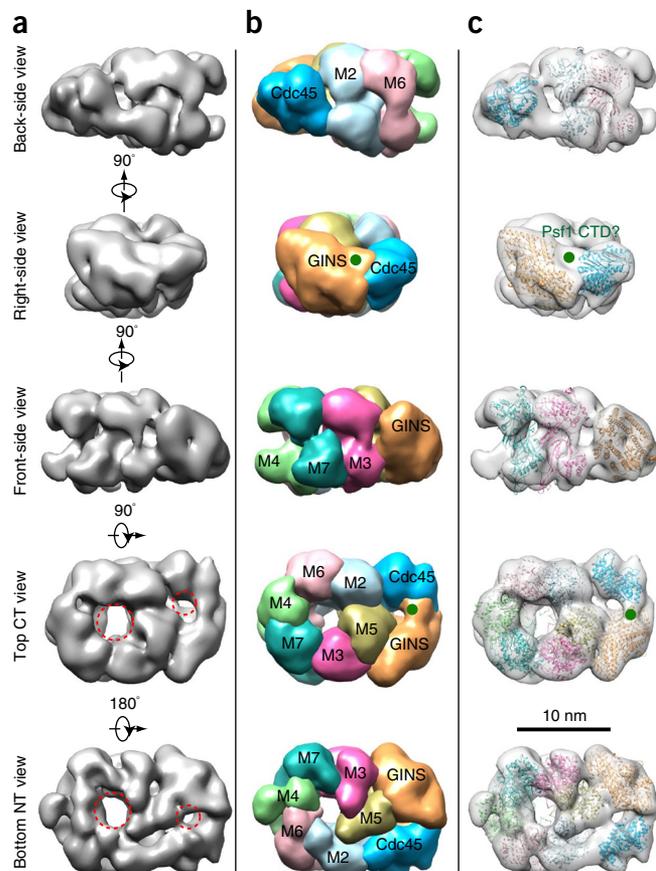
<sup>1</sup>Biosciences Department, Brookhaven National Laboratory, Upton, New York, USA. <sup>2</sup>Department of Biochemistry and Cell Biology, Stony Brook University, Stony Brook, New York, USA. <sup>3</sup>DNA Replication Laboratory, Rockefeller University, New York, New York, USA. <sup>4</sup>Howard Hughes Medical Institute, Rockefeller University, New York, New York, USA. Correspondence should be addressed to M.E.O'D. (odonnel@rockefeller.edu) or H.L. (hli@bnl.gov).

Received 20 August; accepted 21 September; published online 2 November 2015; doi:10.1038/nsmb.3113

**Figure 1** Structure of the *S. cerevisiae* CMG helicase. (a) Surface-rendered and segmented *Sc*CMG structure in different views of the 3D EM map. The larger dashed red circle marks the apparent hole in the middle of Mcm2–7, and the smaller circle indicates the apparent second hole between GINS–Cdc45 and Mcm2–7. (b) Corresponding views of the segmented map. (c) Docking of the crystal structures of the GINS (PDB 2E9X), Cdc45 homolog (PDB 1IR6) and six copies of the *Sulfolobus solfataricus* MCM monomer crystal structure (PDB 3F9V). The green dots in b and c mark the unoccupied density between assigned GINS and Cdc45 that may be the truncated CTD of GINS subunit Psf1 but may also belong to Cdc45. CT, C terminal; NT, N terminal.

images<sup>16,26</sup> (Fig. 2a and Supplementary Fig. 2). Although CMGE can form without the 80/75-mer forked DNA<sup>13</sup>, DNA was required to visualize CMGE through EM and thus appears to stabilize CMGE in the negative-stain procedure. We then determined the EM structure of the CMGE complex (Fig. 2b and Supplementary Fig. 2). Many features such as the N-terminal zinc-finger domains in Mcm subunits were resolved (Fig. 2b and Supplementary Video 1). Similarly to the CMG map (Fig. 1), the crystal structure of the human GINS complex fit remarkably well into the segmented GINS density in the CMGE 3D map (Fig. 3a), and the RecJ homolog structure fit into the density of Cdc45 (refs. 23–25) (Fig. 3b,c).

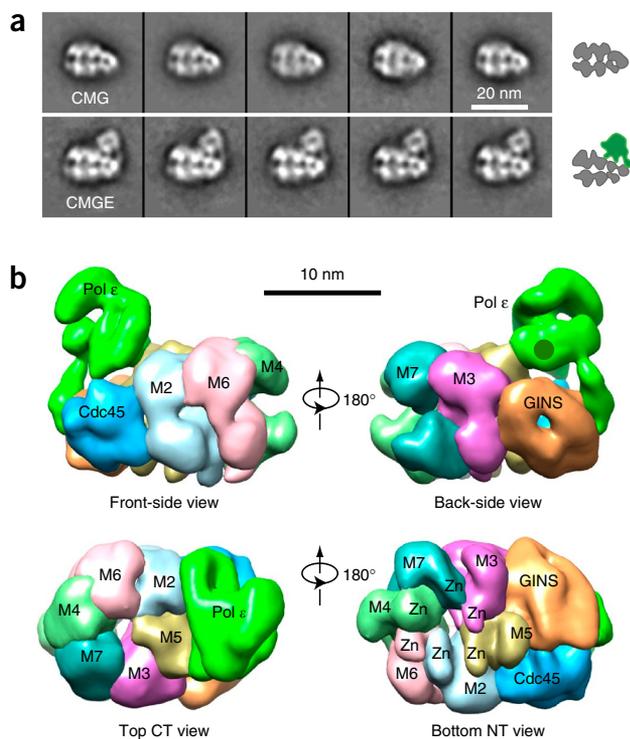
Comparison of the structures of *Sc*CMGE and *Sc*CMG revealed the full density of Pol  $\epsilon$  (Fig. 2c). The structure of the catalytic N-terminal half of the large Pol2 subunit of Pol  $\epsilon$  in complex with primed DNA and dNTP has recently been solved<sup>16,27</sup>. Despite the presence of apo-Pol  $\epsilon$  in CMGE, the Pol2–DNA–dNTP crystal structure docked reasonably well into a region of EM density with CMGE that contains a large groove suggestive of the polymerase active site (Fig. 3d). However, the mass of Pol  $\epsilon$  in CMGE was underestimated in the EM density by about 30%, thus lending uncertainty to this assignment. A small percentage of CMG particles contaminating the CMGE data set may have contributed to a somewhat lower density of Pol  $\epsilon$  in CMGE. For example, the profiles of the top and bottom views of CMG and CMGE were quite similar because Pol  $\epsilon$  was positioned directly over CMG. Thus, computational



sorting of some views was dependent upon slight intensity differences due to Pol  $\epsilon$  rather than to the particle profile. Unintended inclusion of CMG images during 3D CMGE reconstruction would have had the net effect of downweighting the Pol  $\epsilon$  density and size in the final CMGE 3D map. Alternatively, the peripheral features of a large complex, such as Pol  $\epsilon$  holoenzyme at the edge of CMG may have some flexibility, thus leading to a reduction in the overall density and size of Pol  $\epsilon$ . It is also possible that individual subunit flexibility may contribute to a lower density of Pol  $\epsilon$ , thus leading to loss of a particular region (discussed further below). The crystal structure of the Dls1–Dpb4 complex of the CHRAC nucleosome remodeler<sup>28</sup> is homologous to the Dpb3–Dpb4 histone-fold heterodimer of Pol  $\epsilon$ , and it may form the extended arm of density between GINS and Cdc45 (Fig. 3b). However, given their small size and EM resolution, assignment of Dpb3–Dpb4 within the Pol  $\epsilon$  density must await higher-resolution analysis<sup>28</sup>.

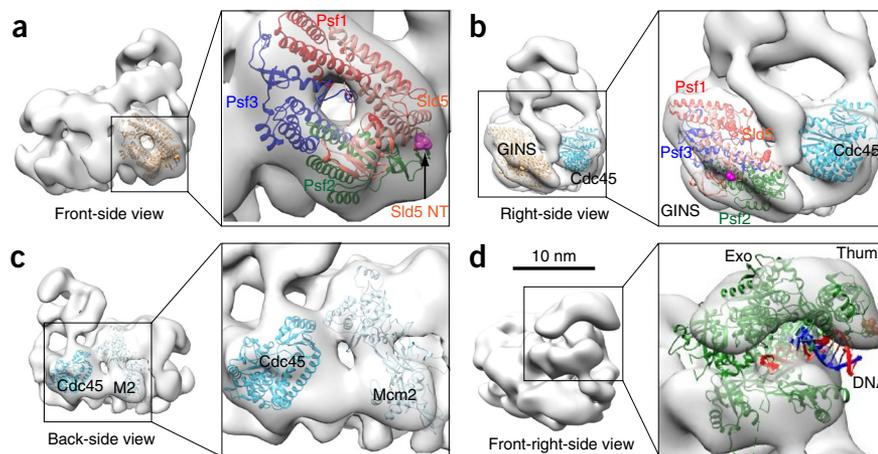
### Pol $\epsilon$ is on the C side of CMG

The Mcm subunits are composed of C- and N-terminal domains, thus giving the Mcm 2–7 complex the appearance of two rings stacked on



**Figure 2** Structure of the *S. cerevisiae* CMGE leading-strand helicase-polymerase. (a) Selected side views of reference-free class averages of CMG (top row) compared to CMG complexed with Pol  $\epsilon$  (bottom row). (b) Surface-rendered and segmented EM map in four views. The Pol  $\epsilon$  density is shown in green. CT, C terminal; NT, N terminal. The distinct groove may be the polymerase active site, and the arm of density extending down between GINS and Cdc45 is the shape expected for the Dpb3–Dpb4 histone-fold heterodimer subunits. Assignments of subunits in the Pol  $\epsilon$  density are not certain, as explained in the text. The six ‘Zn’ labels in the bottom NT view denote the N-terminal zinc-finger domains of Mcm proteins (Supplementary Video 1).

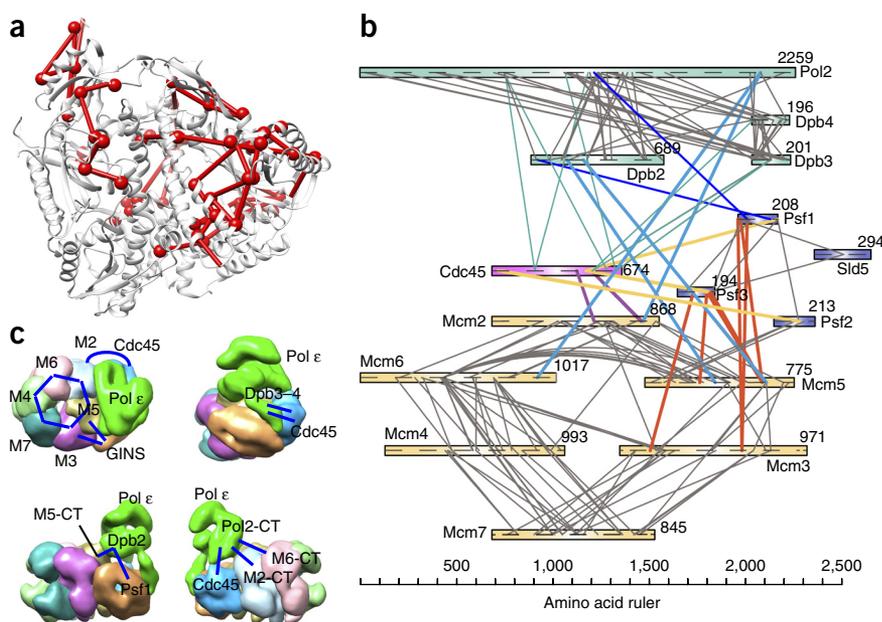
**Figure 3** Rigid-body docking of CMG subunits into the CMGE density map with available crystal structures. (a) The crystal structures of human GINS complex (PDB 2E9X) fitted in the EM density. The GINS subunits are colored red (Psf1), green (Psf2), blue (Psf3) and orange (Sld5). The red spheres show the last residue in the CTD-truncated Psf1 crystal structure. The magenta spheres show the first N-terminal resolved residue (Leu21) of the Sld5 subunit. (b) Side view showing the docking of human GINS and the catalytic core of the RecJ exonuclease homolog to Cdc45 (PDB 1IR6, cyan). (c) Back-side view of the catalytic core of the RecJ exonuclease homolog to Cdc45 (PDB 1IR6, cyan) adjacent to the *S. solfataricus* MCM monomer crystal structure (PDB 3F9V), labeled 'M2'. (d) Crystal structure of the Pol2 catalytic N-terminal domain complexed with primed DNA and dNTP (PDB 4M80), docked such that DNA aligns into the large groove; this is speculative and should be regarded as tentative, owing to the underweight density of Pol  $\epsilon$  in the CMGE structure. Template- and primer-strand DNA are in red and blue, respectively. NT, N terminus; exo, exonuclease.



top of each other<sup>22,29</sup>. Pol  $\epsilon$  contacts CMG at the C-terminal AAA+ domains of Mcm2 and Mcm5, Cdc45 and GINS. This orientation is supported by the location of the Mcm Zn<sup>2+</sup> fingers, which are in the N-terminal domains on the opposite side of the Mcms from Pol  $\epsilon$  (Supplementary Video 1). The position of Pol  $\epsilon$  density on the C side of CMG is also consistent with the subunit assignments determined in the *Dm*CMG study, which would require a different arrangement in *Sc*CMG to position the Pol  $\epsilon$  density on the N surface of CMG.

To further dissect the architecture of CMGE in a residue-specific manner and independently assess the EM subunit assignments, we chemically conjugated the reconstituted complexes with the amine-specific cross-linker disuccinimidyl suberate (DSS) and applied high-resolution mass spectrometry to identify the cross-linked peptides<sup>30</sup>. For a control for this strategy, we mapped the experimentally determined cross-links onto the crystal structure of the catalytic domain of Pol2 (Fig. 4a). The majority of the cross-linked lysines fell within 15–20 Å (Euclidean C $\alpha$ -C $\alpha$  distance), and fully 96% fell within 30 Å,

which is the maximum reach of two lysine side chains coupled by DSS (Supplementary Fig. 3). This result confirmed that only nearby lysines were cross-linked and thus validated the use of intersubunit cross-links (summarized in Fig. 4b) to report on subunit proximity and arrangement within CMGE. We identified 553 unique cross-linked peptides, 189 of which were intersubunit cross-links (comprehensive list in Supplementary Table 1). Our intersubunit cross-link connectivity map (Fig. 4b) recapitulated the overall topology of CMGE deduced from the prior evidence available to us and the newly acquired EM data provided here. Thus, for example, the major cross-links between Mcm subunits confirmed the established order of Mcm2–7 subunits<sup>31</sup>. In addition, we detected several cross-links across the Mcm ring (i.e., between Mcm6 and Mcm3), thus indicating that these specific regions are in proximity (<30 Å). Cross-links of the Mcms to the CMG accessory factors were also consistent with the subunit arrangement of *Dm*CMG<sup>20</sup>. Hence, Mcm2 is the only Mcm subunit that forms cross-links to Cdc45; the Mcm3 and Mcm5



**Figure 4** Subunit proximities within CMGE determined by chemical cross-linking with mass spectrometry readout (CX-MS). CMGE was cross-linked with a lysine-specific bifunctional cross-linker, then fragmented by proteolysis, and cross-linked peptides were identified by mass spectrometry. (a) Overview of cross-links observed within the region of Pol2 corresponding to the crystal structure (PDB 4M80). The cross-linked lysine residues are presented as red spheres. Straight lines represent DSS cross-links. (b) Intersubunit cross-links between subunits of the 15-protein CMGE complex. The lengths of the subunits correspond to the lengths of the colored rectangles that they represent: yellow, Mcms; blue, GINS; purple, Cdc45; green, Pol  $\epsilon$ . (c) Top left, major cross-links observed to connect GINS and Cdc45 to Mcm3–Mcm5 and Mcm2. Remaining views, cross links between Pol  $\epsilon$  and CMG: putative Dpb3–Dpb4 cross-linked to Cdc45 (top right); Pol2 C-terminal region cross-linked to the C-terminal regions of Mcm2 and Mcm6 and Cdc45 (bottom right); Dpb2 cross-linked to the Mcm5 C-terminal region and to Psf1 of GINS (bottom left). M, Mcm; CT, C terminus.

**Figure 5** Staged assembly of the eukaryotic replisome. (a–c) Replisome reconstitution. Selected side views of reference-free class averages of CMG mixed with Ctf4 and 80/75-mer (a), CMG mixed with Ctf4, Pol  $\alpha$  and 80/75-mer (b) and CMG mixed with Pol  $\epsilon$ , Ctf4 and 80/75-mer (c). (d) CMG mixed with Pol  $\epsilon$  and the 160/91-mer primed fork. (e) CMG mixed with Pol  $\epsilon$ , Ctf4, Pol  $\alpha$  and the 160/91-mer primed fork.

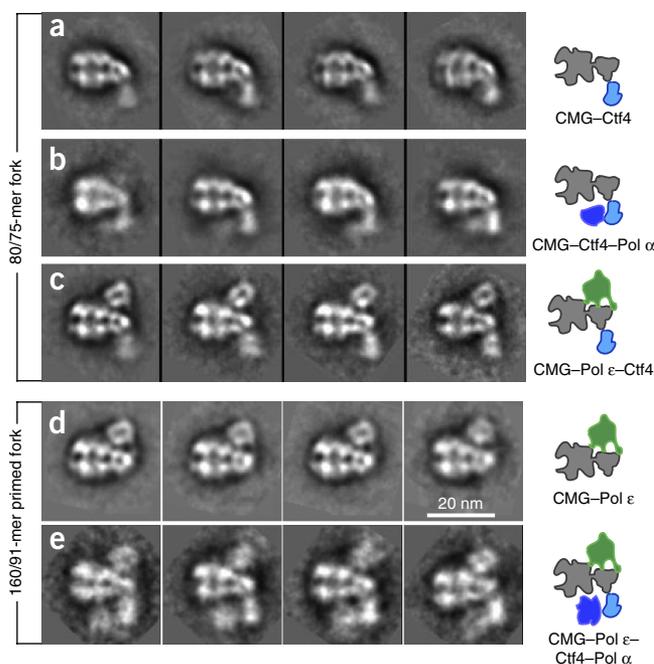
subunits cross-link to the GINS subunits; and the C terminus of Psf1 cross-links to Cdc45 (Fig. 4b,c).

Chemical cross-linking with mass spectrometric readout (CX-MS) confirmed that Pol  $\epsilon$  lies on the C-terminal side of the Mcm ring and that the four subunits of Pol  $\epsilon$  cross-linked to CMG subunits in the regions expected from the Pol  $\epsilon$  EM density (Fig. 4c). The C-terminal domains of Mcm2 and Mcm6 cross-linked to the C half of Pol2. The N half of Pol2 showed no cross-links to Mcm subunits, thus indicating that Pol2-Mcm interactions occur via the C half of Pol2, which has been proposed to encode an inactive polymerase. The C-terminal domain of Mcm5 cross-linked to Dpb2, and there were no cross-links of any of the Pol  $\epsilon$  subunits to the N-terminal domains of any of the Mcm subunits. The N terminus of Dpb2 also cross-linked to Pol2 and the C terminus of Psf1, an interaction previously characterized with isolated domains<sup>32</sup>. Psf1 is located at the C side of CMG (Figs. 1–3), and thus Dpb2-Psf1 cross-links support the position of Pol  $\epsilon$  density on the C-terminal side of CMG. Pol2 and the Pol  $\epsilon$  accessory factors Dpb3 and Dpb4 cross-linked to Cdc45, although the Cdc45 cross-links did not provide information on the location of Pol  $\epsilon$  on the N or C sides of CMG because the structure of Cdc45 is unknown.

At the current level of resolution, we could not precisely locate the active site of Pol  $\epsilon$ , although the distinct groove in the EM structure may correspond to it. It is also possible that the two polymerase regions in the N and C halves of Pol2 define individual domains that are close but separated by a short linker region. This could explain the dearth of cross-links between the N and C halves of Pol2. A previous cryo-EM study of Pol2 has shown a single globular structure, and therefore if the two polymerases within Pol2 are in separate domains, they must be close to each other<sup>24</sup>. In any case, the CX-MS data show that both the N and C halves of Pol2 are in proximity to the accessory subunits of Pol  $\epsilon$  holoenzyme, and this restricts the nonvisible region of Pol  $\epsilon$  density to a region adjacent to the observed density of Pol  $\epsilon$ . The reasoning is as follows. The cross-linking data showed that all the subunits of Pol  $\epsilon$  form an extensive network of many cross-links among one another across the entire length of each subunit, including the active site N half of Pol2, which has several cross-links to each of the Dpb2, Dpb3 and Dpb4 subunits (Fig. 4 and Supplementary Fig. 4). This is what would be expected for any multiprotein holoenzyme (i.e., that the subunits are in proximity and would cross-link to one another). Given the short cross-linker, the nonvisible portion of Pol  $\epsilon$  must be proximal to the observed density. Thus, the N region of Pol2 containing the active site must be (i) within the observed density of Pol  $\epsilon$ , (ii) partially within the observed density or (iii) adjacent to the observed density. Thus, even if the N half of Pol2 is flexibly attached to the C half of Pol2, given the short cross-linker and the many cross-links between the N half of Pol2 and other subunits of Pol  $\epsilon$ , the N region of Pol2 cannot be 120–150 Å separate from the other subunits of Pol  $\epsilon$ , which is the distance required to reach the bottom of CMG.

#### Pol $\alpha$ -Ctf4 attaches to the opposite side of CMG from Pol $\epsilon$

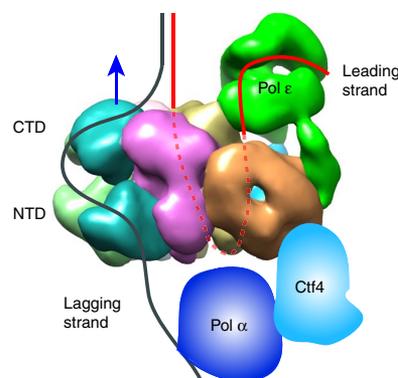
Ctf4 has recently been shown to form a trimer that can bind to both Pol  $\alpha$  and the Sld5 subunit of GINS<sup>14</sup>. Therefore, the position of Ctf4 on CMG serves as a proxy for the location of Pol  $\alpha$  and the lagging



strand. By adding purified Ctf4 trimer to CMG, we observed a new structure that we refer to as CMG-Ctf4 (Fig. 5a and Supplementary Fig. 5b). Addition of both Ctf4 and Pol  $\alpha$  to CMG resulted in a CMG-Ctf4-Pol  $\alpha$  complex, although the Pol  $\alpha$  density adjacent to Ctf4 was very weak (Fig. 5b and Supplementary Fig. 5c). Addition of both Pol  $\epsilon$  and Ctf4 to CMG resulted in a superternary complex that we refer to as CMGE-Ctf4, with densities on diametrically opposed sides of CMG (Fig. 5c and Supplementary Fig. 5e). The N-terminal half of Ctf4 is connected by a flexible linker to the C-terminal half of Ctf4, which forms the trimer<sup>14</sup>, thus possibly accounting for the fuzzy appearance of Ctf4. Most of these particles were side views on the carbon substrate, thus making 3D reconstruction of the complexes unfeasible. Well-defined two-dimensional (2D) class averages of the CMG-Ctf4 complex, the superternary complexes CMGE-Ctf4 and CMG-Ctf4-Pol  $\alpha$ , and the superquaternary complex CMGE-Ctf4-Pol  $\alpha$  established the relative position of Ctf4 and Pol  $\alpha$  with respect to the CMG architecture (Fig. 5a–c and Supplementary Fig. 5a–e).

The architecture of a replisome has been unknown for any cell type: bacteria, archaea or eukaryotes. Current models of replisome action in all cells have placed the polymerases and primase in back of the helicase because these enzymes require ssDNA as a template<sup>1,13,20,33–35</sup>. However, the eukaryotic replisome structure reported here indicates that Pol  $\epsilon$  and Pol  $\alpha$  are on opposite sides of the helicase, thus suggesting that one polymerase rides ahead of the helicase, not behind it. It is possible that we have trapped a particular conformer of the replisome with polymerases on opposite sides of the helicase and that a gross rearrangement occurs at a moving fork. For example, the observed complex could be an intermediate in replisome assembly at an origin or at a stalled fork where Pol  $\epsilon$  triggers a checkpoint response. However, we note that Pol  $\epsilon$  binds CMG through several observable points of contact, and thus a gross relocation of Pol  $\epsilon$  to the Pol  $\alpha$ -Ctf4 side of CMG would require breaking these multiple connections and establishing new ones. To determine whether a gross rearrangement occurs upon engaging an active DNA fork, we assembled CMG-Pol  $\epsilon$ -Ctf4-Pol  $\alpha$  on a 160- and 91-mer (160/91-mer) primed DNA fork that we have previously validated to be active in replication assays<sup>35</sup>. We also confirmed that CMGE binds the 160/91-mer

**Figure 6** Architecture of the eukaryotic replisome. Replisome structure and the proposed DNA path through the replisome. Pol  $\alpha$  is shown in blue. Ctf4 is in cyan. Red and black lines illustrate possible leading- and lagging-strand DNA. The blue arrow indicates the direction of replisome movement on DNA. The diagram indicates a long path of leading-strand DNA through the entire Mcm ring and then bending back up to Pol  $\epsilon$ , requiring about 40 nucleotides of ssDNA. Leading ssDNA is illustrated as going completely through the Mcm2–7 complex and then bending up through the second ‘accessory’ channel of CMG, but this path is speculative. Other DNA paths are possible. Additional details in main text and **Supplementary Figure 6**.



primed fork under the conditions used here (**Supplementary Fig. 1c**). Examination of 2D class-average images of CMGE and the CMGE–Ctf4–Pol  $\alpha$  complex showed the same general outline as complexes using the unprimed 80/75-mer fork (**Fig. 5d,e**).

## DISCUSSION

### Pol $\epsilon$ rides ahead of the unwinding point at the fork

The current study reveals that the leading-strand Pol  $\epsilon$  and lagging-strand Pol  $\alpha$  are located on opposite sides of CMG helicase. Hence, one of these polymerases must be on top of the helicase and ride ahead of the unwinding point of the parental duplex, an unexpected position relative to that in decades of textbook drawings that show both polymerases behind the helicase. Determination of which polymerase rides ahead of the fork requires knowledge of the DNA path through the Mcm complex. Viewed from their side, the Mcm complex appears as two stacked rings composed of the CTD tier and the NTD tier, owing to the bilobed structure of the Mcm subunits<sup>22,29</sup> (**Fig. 1a,c**). The widely accepted view of Pol  $\epsilon$  below CMG would require the leading strand to enter the Mcms from the NTD tier. However, studies of both archaeal Mcm and *Dm*CMG have demonstrated that the leading strand enters the CTD tier of the Mcm complex<sup>20,36</sup>. This DNA path places the leading-strand Pol  $\epsilon$  ahead of the forked junction (**Fig. 6**), thus producing a completely unanticipated replisome architecture.

Given the surprising arrangement of Pol  $\epsilon$  ahead of CMG helicase, and the reported DNA path, the leading-strand ssDNA would need to traverse the ~110-Å central chamber of the Mcms, then bend back an additional 110 Å to reach Pol  $\epsilon$ , for a total of 220 Å, or ~40 nucleotides. The ssDNA may traverse the outside of CMG to reach Pol  $\epsilon$ , or it could thread through the second channel in CMG formed by the Cdc45 and GINS accessory proteins (**Fig. 6**). Indeed, a recent study of *Dm*CMG has demonstrated that the leading strand can occupy the second channel under particular conditions<sup>37</sup>. Alternatively, the leading ssDNA may take an ~20-nucleotide path by exiting the Mcm channel at an internal position (for example, at the Mcm2–Mcm5 ‘gate’) and then bending up toward Pol  $\epsilon$  (**Supplementary Fig. 6a**). If the N half of Pol2 is a separate domain and flexibly connected to the C half, its location within cross-linking distance of Pol  $\epsilon$  holoenzyme subunits would place it near the visible density of Pol  $\epsilon$  at the top of CMG, and DNA would still be required to make a U-turn from the MCMs (**Supplementary Fig. 6b**). Interestingly, experimental support for a 20- to 40-nucleotide leading-strand ssDNA gap at the fork has been provided by studies in the *Xenopus* system in which replisome advance was blocked with an interstrand cross-link<sup>34</sup>. The 10- to 30-min time points show 20- to 40-nucleotide ssDNA gaps upon blocking the leading strand. Further studies will be required to define the DNA path and architecture of a moving replisome, but if the present findings prove correct, the current view that polymerases trail behind the helicase will need to be updated<sup>1,13,20,33–35</sup>.

### To what extent does Pol $\epsilon$ perform leading-strand synthesis?

Pol  $\epsilon$  was initially discovered as a third essential polymerase in budding yeast<sup>38</sup>, and its role in replication continues to be studied extensively. Recent studies have shown that mutations in Pol  $\epsilon$  are associated with some types of cancer, similarly to the association of cancers with mismatch repair and Pol  $\eta$  defects<sup>39</sup>. Mutations in active site residues of Pol  $\epsilon$  are lethal, thus indicating that Pol  $\epsilon$  is required to synthesize DNA during chromosome replication<sup>40</sup>. Interestingly, the N-terminal region of Pol2 containing the active DNA polymerase can be deleted, and cells still survive but are severely compromised in S-phase progression; this result suggests that another DNA polymerase can substitute for Pol  $\epsilon$  but probably does not reflect normal replication<sup>40–42</sup>. These observations are reminiscent of genetic studies in *Escherichia coli* in which *dnaE*, encoding the polymerase that normally duplicates both strands of the chromosome, is mutated; these cells survive via replication by Pol I, but they grow slowly<sup>43</sup>. Interestingly, the inactive C half of Pol2 is essential, presumably serving a structural role<sup>38,39</sup>.

Several genetic studies using a slightly altered Pol  $\epsilon$  that provides a mutation signature on the DNA product have indicated that Pol  $\epsilon$  predominates over Pol  $\delta$  on the leading strand and that Pol  $\delta$  predominates over Pol  $\epsilon$  on the lagging strand<sup>6,9–12</sup>. This is consistent with studies of proofreading mutants in Pals  $\epsilon$  and  $\delta$  that have concluded that the two Pals act on different strands<sup>44</sup>. Use of an altered Pol  $\delta$  indicates its predominant role in bulk lagging-strand synthesis in both *S. cerevisiae* and *Schizosaccharomyces pombe*<sup>6,10–12</sup>. Use of a Pol  $\epsilon$  mutant that frequently misincorporates rNMPs has enabled genome-wide analysis of strand bias by Pol  $\epsilon$  and supports the conclusion that Pol  $\epsilon$  performs bulk leading-strand synthesis in both budding and fission yeast<sup>11,12</sup>. Results from biochemical studies of Pals  $\epsilon$  and  $\delta$  with *Sc*CMG are consistent with these cellular findings, because Pol  $\epsilon$  binds CMG directly, is more active than Pol  $\delta$  in leading-strand synthesis with CMG and is less active than Pol  $\delta$  on the lagging strand of CMG-driven forks<sup>13,35,45</sup>. Earlier biochemical assays of Pals  $\epsilon$  and Pol  $\delta$  have also indicated that their properties are most consistent with leading- and lagging-strand synthesis, respectively<sup>46,47</sup>. Genome-wide chromatin immunoprecipitation assays of polymerase occupancy in budding yeast have shown that Pol  $\epsilon$  cross-links specifically to the leading strand, whereas Pol  $\delta$  cross-links to the lagging strand<sup>48</sup>. However, it has recently been argued that cross-linking studies may bias Pol  $\delta$  to the lagging strand because it must extend multiple Okazaki fragments<sup>49,50</sup>. An interesting recent report has arrived at a different conclusion from all the previous work, indicating that Pol  $\delta$  performs bulk leading- and lagging-strand synthesis similar to that in the SV40 viral system<sup>49</sup>. The different conclusions are explained by the cellular studies having been performed

in strain backgrounds with mutations in various repair pathways, which may have introduced strand bias into the results<sup>50</sup>. Hence, the extent to which Pol  $\epsilon$  and Pol  $\delta$  perform leading-strand synthesis remains an open question that requires further study.

### Function of the replisome architecture

The counterintuitive position of Pol  $\epsilon$  at the 'top' of CMG suggests that an unforeseen function may underlie the unexpected replisome architecture. Although the function that this architectural facet serves cannot be ascertained a priori, there are several possibilities. For example, the arrangement segregates the two daughter strands above and below CMG, and this may help organize daughter strands during replication. Second, the requirement for leading ssDNA to transit over (or through) GINS–Cdc45 might enable CMG subunits to serve a surveillance role to recognize template lesions or misincorporated nucleotides (for example, dUMP and rNMP) before they enter the Pol  $\epsilon$  active site. Yet another possibility is that Pol  $\epsilon$  is pushed by CMG ATP hydrolysis, and this may cause Pol  $\epsilon$  to become a strand-displacement enzyme at the prow of the fork (**Supplementary Fig. 6c**). A most intriguing possible function of the architecture is suggested by the genetics of Pol  $\epsilon$ . Particular mutations in Pol2, or deletions of Dpb3 or Dpb4, result in loss of epigenetic silencing in yeast<sup>51</sup>. Indeed, Pol  $\epsilon$  has been shown to bind histones, whose modifications underlie epigenetic silencing<sup>52</sup>. We note that Dpb3–Dpb4 is a histone-fold heterodimer, and both TFIID of RNA Pol II and the CHRAC chromatin remodeler contain histone-fold heterodimers that are thought to help mobilize nucleosomes. Hence, placement of Pol  $\epsilon$  ahead of the helicase may facilitate replisome function with nucleosomes during replication, possibly directing asymmetric epigenetic states in the two daughter cells (i.e., asymmetric cell division during development). Clearly, further studies will be required to understand the functional implications of the unexpected replisome architecture reported here.

### METHODS

Methods and any associated references are available in the [online version of the paper](#).

**Accession codes.** The 3D EM maps of CMG and CMGE have been deposited at the EMDB database under accession codes EMD-6463 and EMD-6465, respectively.

*Note: Any Supplementary Information and Source Data files are available in the [online version of the paper](#).*

### ACKNOWLEDGMENTS

We would like to thank O. Yurieva and D. Zhang for the purification of CMG, Pol  $\alpha$ , Pol  $\epsilon$  and Ctf4. This work was funded by the US National Institutes of Health (GM103314 and GM109824 to B.T.C.; GM74985 and AG29979 to H.L.; and GM38839 to M.E.O'D.) and the Howard Hughes Medical Institute (M.E.O'D.).

### AUTHOR CONTRIBUTIONS

J.S., R.E.G., Y.S., B.T.C., H.L. and M.E.O'D. designed experiments. J.S., Y.S., R.E.G. and Z.Y. performed experiments. J.S., R.E.G., Y.S., Z.Y., B.T.C., H.L. and M.E.O'D. analyzed the data. H.L. and M.E.O'D. prepared the manuscript with input from all authors.

### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

1. O'Donnell, M., Langston, L. & Stillman, B. Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harb. Perspect. Biol.* **5**, a010108 (2013).

2. Johansson, E. & Dixon, N. Replicative DNA polymerases. *Cold Spring Harb. Perspect. Biol.* **5**, a012799 (2013).
3. Bell, S.D. & Botchan, M.R. The minichromosome maintenance replicative helicase. *Cold Spring Harb. Perspect. Biol.* **5**, a012807 (2013).
4. Ilves, I., Petojevic, T., Pesavento, J.J. & Botchan, M.R. Activation of the MCM2–7 helicase by association with Cdc45 and GINS proteins. *Mol. Cell* **37**, 247–258 (2010).
5. Moyer, S.E., Lewis, P.W. & Botchan, M.R. Isolation of the Cdc45/Mcm2–7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc. Natl. Acad. Sci. USA* **103**, 10236–10241 (2006).
6. Kunkel, T.A. & Burgers, P.M. Dividing the workload at a eukaryotic replication fork. *Trends Cell Biol.* **18**, 521–527 (2008).
7. Waga, S. & Stillman, B. The DNA replication fork in eukaryotic cells. *Annu. Rev. Biochem.* **67**, 721–751 (1998).
8. MacNeill, S. *The Eukaryotic Replisome: a Guide to Protein Structure and Function* (Springer, New York, 2012).
9. Pursell, Z.F., Isoz, I., Lundstrom, E.B., Johansson, E. & Kunkel, T.A. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science* **317**, 127–130 (2007).
10. Nick McElhinny, S.A., Gordenin, D.A., Stith, C.M., Burgers, P.M. & Kunkel, T.A. Division of labor at the eukaryotic replication fork. *Mol. Cell* **30**, 137–144 (2008).
11. Clausen, A.R. *et al.* Tracking replication enzymology *in vivo* by genome-wide mapping of ribonucleotide incorporation. *Nat. Struct. Mol. Biol.* **22**, 185–191 (2015).
12. Miyabe, I., Kunkel, T.A. & Carr, A.M. The major roles of DNA polymerases epsilon and delta at the eukaryotic replication fork are evolutionarily conserved. *PLoS Genet.* **7**, e1002407 (2011).
13. Langston, L.D. *et al.* CMG helicase and DNA polymerase epsilon form a functional 15-subunit holoenzyme for eukaryotic leading-strand DNA replication. *Proc. Natl. Acad. Sci. USA* **111**, 15390–15395 (2014).
14. Simon, A.C. *et al.* A Ctf4 trimer couples the CMG helicase to DNA polymerase  $\alpha$  in the eukaryotic replisome. *Nature* **510**, 293–297 (2014).
15. Kang, Y.H. *et al.* Interaction between human Ctf4 and the Cdc45/Mcm2–7/GINS (CMG) replicative helicase. *Proc. Natl. Acad. Sci. USA* **110**, 19760–19765 (2013).
16. Hogg, M. *et al.* Structural basis for processive DNA synthesis by yeast DNA polymerase  $\epsilon$ . *Nat. Struct. Mol. Biol.* **21**, 49–55 (2014).
17. Swan, M.K., Johnson, R.E., Prakash, L., Prakash, S. & Aggarwal, A.K. Structural basis of high-fidelity DNA synthesis by yeast DNA polymerase  $\delta$ . *Nat. Struct. Mol. Biol.* **16**, 979–986 (2009).
18. Klinge, S., Nunez-Ramirez, R., Llorca, O. & Pellegrini, L. 3D architecture of DNA Pol alpha reveals the functional core of multi-subunit replicative polymerases. *EMBO J.* **28**, 1978–1987 (2009).
19. Costa, A. *et al.* The structural basis for MCM2–7 helicase activation by GINS and Cdc45. *Nat. Struct. Mol. Biol.* **18**, 471–477 (2011).
20. Costa, A. *et al.* DNA binding polarity, dimerization, and ATPase ring remodeling in the CMG helicase of the eukaryotic replisome. *eLife* **3**, e03273 (2014).
21. Sun, J. *et al.* Cryo-EM structure of a helicase loading intermediate containing ORC-Cdc6-Cdt1–MCM2–7 bound to DNA. *Nat. Struct. Mol. Biol.* **20**, 944–951 (2013).
22. Sun, J. *et al.* Structural and mechanistic insights into Mcm2–7 double-hexamer assembly and function. *Genes Dev.* **28**, 2291–2303 (2014).
23. Chang, Y.P., Wang, G., Bermudez, V., Hurwitz, J. & Chen, X.S. Crystal structure of the GINS complex and functional insights into its role in DNA replication. *Proc. Natl. Acad. Sci. USA* **104**, 12685–12690 (2007).
24. Choi, J.M., Lim, H.S., Kim, J.J., Song, O.K. & Cho, Y. Crystal structure of the human GINS complex. *Genes Dev.* **21**, 1316–1321 (2007).
25. Kamada, K., Kubota, Y., Arata, T., Shindo, Y. & Hanaoka, F. Structure of the human GINS complex and its assembly and functional interface in replication initiation. *Nat. Struct. Mol. Biol.* **14**, 388–396 (2007).
26. Asturias, F.J. *et al.* Structure of *Saccharomyces cerevisiae* DNA polymerase epsilon by cryo-electron microscopy. *Nat. Struct. Mol. Biol.* **13**, 35–43 (2006).
27. Jain, R. *et al.* Crystal structure of yeast DNA polymerase epsilon catalytic domain. *PLoS ONE* **9**, e94835 (2014).
28. Hartlepp, K.F. *et al.* The histone fold subunits of *Drosophila* CHRAC facilitate nucleosome sliding through dynamic DNA interactions. *Mol. Cell Biol.* **25**, 9886–9896 (2005).
29. Li, N. *et al.* Structure of the eukaryotic MCM complex at 3.8 Å. *Nature* **524**, 186–191 (2015).
30. Shi, Y. *et al.* Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. *Mol. Cell Proteomics* **13**, 2927–2943 (2014).
31. Davey, M.J., Indiani, C. & O'Donnell, M. Reconstitution of the Mcm2–7p heterohexamer, subunit arrangement, and ATP site architecture. *J. Biol. Chem.* **278**, 4491–4499 (2003).
32. Sengupta, S., van Deursen, F., de Piccoli, G. & Labib, K. Dpb2 integrates the leading-strand DNA polymerase into the eukaryotic replisome. *Curr. Biol.* **23**, 543–552 (2013).
33. Yardimci, H. *et al.* Bypass of a protein barrier by a replicative DNA helicase. *Nature* **492**, 205–209 (2012).
34. Fu, Y.V. *et al.* Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase. *Cell* **146**, 931–941 (2011).
35. Georgescu, R.E. *et al.* Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. *Nat. Struct. Mol. Biol.* **21**, 664–670 (2014).

36. Rothenberg, E., Trakselis, M.A., Bell, S.D. & Ha, T. MCM forked substrate specificity involves dynamic interaction with the 5'-tail. *J. Biol. Chem.* **282**, 34229–34234 (2007).
37. Petojevic, T. *et al.* Cdc45 (cell division cycle protein 45) guards the gate of the Eukaryote Replisome helicase stabilizing leading strand engagement. *Proc. Natl. Acad. Sci. USA* **112**, E249–E258 (2015).
38. Morrison, A., Araki, H., Clark, A.B., Hamatake, R.K. & Sugino, A. A third essential DNA polymerase in *S. cerevisiae*. *Cell* **62**, 1143–1151 (1990).
39. Henninger, E.E. & Pursell, Z.F. DNA polymerase epsilon and its roles in genome stability. *IUBMB Life* **66**, 339–351 (2014).
40. Dua, R., Levy, D.L. & Campbell, J.L. Analysis of the essential functions of the C-terminal protein/protein interaction domain of *Saccharomyces cerevisiae* pol epsilon and its unexpected ability to support growth in the absence of the DNA polymerase domain. *J. Biol. Chem.* **274**, 22283–22288 (1999).
41. Kesti, T., Flick, K., Keranen, S., Syvaaja, J.E. & Wittenberg, C. DNA polymerase epsilon catalytic domains are dispensable for DNA replication, DNA repair, and cell viability. *Mol. Cell* **3**, 679–685 (1999).
42. Ohya, T. *et al.* The DNA polymerase domain of pol(epsilon) is required for rapid, efficient, and highly accurate chromosomal DNA replication, telomere length maintenance, and normal cell senescence in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **277**, 28099–28108 (2002).
43. Niwa, O., Bryan, S.K. & Moses, R.E. Alternate pathways of DNA replication: DNA polymerase I-dependent replication. *Proc. Natl. Acad. Sci. USA* **78**, 7024–7027 (1981).
44. Shcherbakova, P.V. & Pavlov, Y.I. 3'→5' exonucleases of DNA polymerases epsilon and delta correct base analog induced DNA replication errors on opposite DNA strands in *Saccharomyces cerevisiae*. *Genetics* **142**, 717–726 (1996).
45. Georgescu, R.E. *et al.* Reconstitution of a eukaryotic replisome reveals suppression mechanisms that define leading/lagging strand operation. *eLife* **4**, e04988 (2015).
46. Garg, P., Stith, C.M., Sabouri, N., Johansson, E. & Burgers, P.M. Idling by DNA polymerase delta maintains a ligatable nick during lagging-strand DNA replication. *Genes Dev.* **18**, 2764–2773 (2004).
47. Burgers, P.M. Polymerase dynamics at the eukaryotic DNA replication fork. *J. Biol. Chem.* **284**, 4041–4045 (2009).
48. Yu, C. *et al.* Strand-specific analysis shows protein binding at replication forks and PCNA unloading from lagging strands when forks stall. *Mol. Cell* **56**, 551–563 (2014).
49. Johnson, R.E., Klassen, R., Prakash, L. & Prakash, S. A major role of DNA polymerase delta in replication of both the leading and lagging DNA strands. *Mol. Cell* **59**, 163–175 (2015).
50. Stillman, B. Reconsidering DNA polymerases at the replication fork in eukaryotes. *Mol. Cell* **59**, 139–141 (2015).
51. Iida, T. & Araki, H. Noncompetitive counteractions of DNA polymerase epsilon and ISW2/yCHRAC for epigenetic inheritance of telomere position effect in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **24**, 217–227 (2004).
52. Tackett, A.J. *et al.* Proteomic and genomic characterization of chromatin complexes at a boundary. *J. Cell Biol.* **169**, 35–47 (2005).



(CMG + DNA  $\leftrightarrow$  CMG-DNA), and the  $K_d$  is defined by the equation  $K_d = [\text{CMG}][\text{DNA}]/[\text{CMG-DNA}]$ , where [CMG], [DNA] and [CMG-DNA] are the molar concentrations of free CMG, free DNA and bound complex at equilibrium, respectively. Because the gel-shift experiment monitors the fraction ( $f$ ) of bound DNA rather than the free protein concentration, the fraction of bound DNA is related to  $K_d$  by the equation  $f = [\text{CMG-DNA}]/([\text{DNA}] + [\text{CMG-DNA}]) = [P_t]/([P_t] + K_d) = 1/(1 + (K_d/[P_t]))$ , where  $P_t$  is the total CMG concentration. This equation assumes that the DNA is in trace amounts, because the titration is performed at 1 nM primed fork DNA, which is approximately 30 fold below the  $K_d$ , such that  $P_t$  approximates the free protein concentration at equilibrium. The data were fitted with Matlab software. The observed  $K_d$  value of yeast CMG to primed fork DNA was  $29.7 \pm 2.2$ . This value is similar to the  $K_d$  values of CMG-DNA binding obtained in studies of *Dm*CMG that have estimated a binding affinity of *Dm*CMG to forked DNA of approximately 10–20 nM (ref. 4).

**Chemical cross-linking with mass spectrometry readout (CX-MS).** The CX-MS procedure was performed essentially as previously described<sup>30,57</sup>. CMGE was formed as described above for electron microscopy, except 20 mM HEPES-OH, pH 7.5 was used in place of 20 mM Tris Acetate. CMGE was cross-linked with 2 mM disuccinimidyl suberate (Creative Molecules) for 30 min at 25 °C with constant agitation (1,200 r.p.m.). The reaction was then quenched in 50 mM ammonium bicarbonate. Approximately 50–100  $\mu$ g cross-linked complex was resuspended and heated in 100–200  $\mu$ l  $\times$  LDS loading buffer (Life Technologies). The sample was cooled at room temperature for cysteine alkylation (50 mM iodoacetamide, Sigma) and separated by electrophoresis in a 4–12% SDS PAGE gel. The gel region above 350 kDa was sliced and digested in gel with trypsin to release the cross-linked peptides. The resulting proteolytic peptide mixture was dissolved in 20  $\mu$ l of a solution containing 30% 5 mM Tris(2-carboxyethyl)phosphine (Sigma) and 0.2% formic acid (FA) and fractionated by peptide size-exclusion chromatography (SEC) (Superdex Peptide PC 3.2/30 GE Healthcare) with offline HPLC separation with an autosampler (Agilent Technologies). Three SEC fractions in the molecular-mass range of ~2.5 kDa to 8 kDa were collected and analyzed by LC/MS.

For identification of cross-linked peptides, the purified peptides were dissolved in the sample loading buffer (5% MeOH, 0.2% FA) and analyzed by an LTQ Velos Orbitrap Pro mass spectrometer or an Orbitrap Q Exactive (QE) Plus mass spectrometer (Thermo Fisher). For the analysis with the Velos Orbitrap mass spectrometer, the peptides were loaded by a pressure Baume onto a self-packed PicoFrit column with an integrated electrospray ionization emitter tip (360 O.D, 75 I.D with 15- $\mu$ m tip, New Objective). The column was packed with 8 cm of reverse-phase C18 material (3  $\mu$ m porous silica, 200- $\text{\AA}$  pore size, Dr. Maisch). Mobile phase A consisted of 0.5% acetic acid, and mobile phase B consisted of 70% ACN with 0.5% acetic acid. The peptides were eluted in a 120-min LC gradient (8% B to 50% B, 0–93 min, followed by 50% B to 100% B, 93–110 min and equilibrated with 100% A until 120 min) with an HPLC system (Agilent), and analyzed with a LTQ Velos Orbitrap Pro mass spectrometer. The flow rate was ~200 nL/min. The spray voltage was set at 1.9–2.3 kV. The capillary temperature was 250 °C, and ion transmission on Velos S lenses was set at 45%. The instrument was operated in the data-dependent mode, in which the top eight most-abundant ions were fragmented by higher-energy collisional dissociation (HCD) (HCD normalized energy 29, 0.1-ms activation time) and analyzed in the Orbitrap mass analyzer. The target resolution for MS1 was 60,000 and was 7,500 for MS2. Ions (370–1,700  $m/z$ ) with a charge state of >3 were selected for fragmentation. A dynamic exclusion of (15 s / 2 / 60 s) was used. Other instrumental parameters included: 'lock mass' at 371.1012 Da, a mass-exclusion window of 1.5 Th, and a

minimal threshold of 5,000 to trigger an MS/MS event. Ion trap-accumulation limits (precursors) were  $1 \times 10^5$  and  $1 \times 10^6$  for the linear ion trap and Orbitrap, respectively. For MS2, the Orbitrap ion accumulation limit was  $5 \times 10^5$ . The maximum ion-injection time for the Orbitrap was 500–700 ms. The QE plus instrument was directly coupled to an EasyLC system (Thermo Fisher), and experimental parameters were similar to those of the Velos Orbitrap. The cross-linked peptides were loaded onto an Easy-Spray column heated at 35 °C (C18, 3  $\mu$ m particle size, 200- $\text{\AA}$  pore size, and 50  $\mu$ m  $\times$  15 cm, Thermo Fisher) and eluted with a 120-min LC gradient (2% B to 10% B, 0–6 min; 10% B to 35% B, 6–102 min; 35% B to 100% B, 102–113 min); this was followed by equilibration, in which mobile phase A consisted of 0.1% formic acid, and mobile phase B consisted of 0.1% formic acid in acetonitrile. The flow rate was ~300 nl/min. The spray voltage was 2.0 kV, and the top 10 most abundant ions (charge stage of 3–7) were selected and fragmented by HCD (normalized HCD energy 28).

The raw data were transformed to MGF (mascot generic format) and searched by pLink software<sup>58</sup> with a database containing sequences of the 15 protein subunits of yeast CMG-Pol  $\epsilon$  complexes. Other search parameters included mass accuracy of MS1  $\leq 10$  p.p.m. and MS2  $\leq 20$  p.p.m. for the initial database search, cysteine carboxymethylation as a fixed modification, methionine oxidation as a variable modification, and a maximum of one trypsin miscleavage. The results were filtered at 5% false discovery rate (FDR), and false positives were then identified by manual verification as previously described<sup>30,59</sup>. Briefly, the primary mass spectrometry data were initially analyzed by software that predicted a 5% FDR as an initial filter. Because many of these were still false positives, the data were manually inspected for verification of spectra. We reason that an important additional caveat for FDR estimation for cross-linked peptides that is not generally taken into account lies in the large background of peptides that are not cross-linked, including the multitude of low-abundance peptide species that probably arise from residual undercleaved tryptic peptides, nontryptic peptides, chemically modified species, in-source fragmentations and combinations of these as well as combinations with cross-linked peptides. For this reason, we stringently post-filter the 5% FDR data (requiring, for example, extensive fragmentation coverage of both peptide chains, and a minimum of four amino acids for both of the cross-link peptide chains), with the result that we generally discard an additional 20–30% of these data<sup>10,12</sup>. We adopt this stringent post-filtering strategy to reduce the likelihood of false-positive cross-link identifications. Thus, the final FDR in our cross-link data set is expected to be substantially smaller than 1%. A total of 553 unique cross-linked peptides were identified as a result (**Supplementary Table 1**).

53. Morrison, A., Bell, J.B., Kunkel, T.A. & Sugino, A. Eukaryotic DNA polymerase amino acid sequence required for 3'→5' exonuclease activity. *Proc. Natl. Acad. Sci. USA* **88**, 9473–9477 (1991).
54. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
55. Scheres, S.H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
56. Pettersen, E.F. *et al.* UCSF Chimera: a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
57. Leitner, A. *et al.* Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Mol. Cell Proteomics* **11**, M111.014126 (2012).
58. Yang, B. *et al.* Identification of cross-linked peptides from complex samples. *Nat. Methods* **9**, 904–906 (2012).
59. Cevher, M.A. *et al.* Reconstitution of active human core Mediator complex reveals a critical role of the MED14 subunit. *Nat. Struct. Mol. Biol.* **21**, 1028–1034 (2014).