BNL-112308-2016-JA

The Architecture of a Eukaryotic Replisome

Jingchuan Sun[1,2], Yi Shi[3], Roxana E. Georgescu[3,4], Zuanning Yuan[1,2], Brian T. Chait[3], Huilin Li*[1,2], Michael E. O'Donnell*[3,4]

Submitted to Nature Structural & Molecular Biology

November 2, 2015

Department/Division/Office

Brookhaven National Laboratory

U.S. Department of Energy
DOE Office of Basic Energy Sciences

DISCLAIMER

The Architecture of a Eukaryotic Replisome

Jingchuan Sun[1,2], Yi Shi[3], Roxana E. Georgescu[3,4], Zuanning Yuan[1,2], Brian T. Chait[3], Huilin Li*[1,2], Michael E. O'Donnell*[3,4]

[1] Biosciences Department, Brookhaven National Laboratory, Upton, New York, USA

[2] Department of Biochemistry & Cell Biology, Stony Brook University, Stony Brook, New York, USA.

[3] The Rockefeller University, 1230 York Avenue, New York, New York, USA.

[4] Howard Hughes Medical Institute

*Correspondence and requests for materials should be addressed to M.O.D. (odonnel@rockefeller.edu) or H.L. (hli@bnl.gov)

ABSTRACT

At the eukaryotic DNA replication fork, it is widely believed that the Cdc45-Mcm2-7-GINS (CMG) helicase leads the way in front to unwind DNA, and that DNA polymerases (Pol) trail behind the helicase. Here we use single particle electron microscopy to directly image a replisome. Contrary to expectations, the leading strand Pol ε is positioned ahead of CMG helicase, while Ctf4 and the lagging strand Pol α-primase (Pol α) are behind the helicase. This unexpected architecture indicates that the leading strand DNA travels a long distance before reaching Pol ε, it first threads through the Mcm2-7 ring, then makes a U-turn at the bottom to reach Pol ε 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.

INTRODUCTION

DNA is replicated by a multi-protein machinery referred to as a replisome [1,2]. Replisomes contain a helicase to unwind DNA, DNA polymerases that synthesize the leading and lagging strands, and a primase that makes short primed sites to initiate 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 [3-5]. In eukaryotes, the leading and lagging strands are replicated by Pol ε and Pol δ [6,7] and the lagging strand is repeatedly primed by Pol α, a polymerase-primase that makes short RNA-DNA primers [1,8]. Several studies indicate that Pol ε is responsible for bulk synthesis of the leading strand, and that Pol δ performs bulk synthesis of the lagging strand [6,9-12]. Consistent with a dedicated role of Pol ε on the leading strand, Pol ε binds directly to CMG helicase, which encircles the leading strand [13]. In addition, a trimer of the Ctf4 protein binds both CMG and Pol α [14,15]. Structures of replisome components, such as the Pol ε catalytic domain, Pol δ, and Pol α have been reported [16-18], and EM 3D reconstruction has outlined the structure of *Drosophila melanogaster* CMG (DmCMG) [19,20].

Due 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-strand (ss) DNA for function and thus it has long been assumed that the leading and lagging strand polymerases trail behind the helicase [1,2]. 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 was 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 electron microscopy [21,22]. 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 *Saccharomyces cerevisiae* CMG helicase 11-subunit complex (ScCMG) in the presence of a small 80/75mer forked DNA to which it binds under the

conditions used in this study (**Supplementary Fig. 1a,b**). The ScCMG EM structure (**Fig. 1)** is remarkably similar to the previously reported *Drosophila melanogaster* CMG (DmCMG) [19,20]. We have confirmed the positional assignments of CMG subunits by chemical crosslinking in combination with mass spectrometric readout, discussed later in this report. The subunits follow the same arrangement as those determined for the DmCMG [20]. The Mcm2-7 hexamer structure in the ScCMG complex agrees with the Mcm2-7 structures of single and double Mcm2-7 hexamers loaded onto duplex DNA [21,22]. Further, the assigned GINS and Cdc45 densities are consistent with the homolog crystal structures **(Fig. 1)** as outlined for DmCMG [20]. An additional density, unoccupied by the human GINS structure may belong to Cdc45, or the C-terminal domain (CTD) of Psf1 that is too mobile to visualize in the GINS crystal structure [23-25], as suggested in the DmCMG analysis [20]. The interface of GINS/Cdc45 with Mcm2/5 appears to form a second hole, or channel through ScCMG as noted for DmCMG [19] (**Fig. 1a**).

3D reconstruction of CMGE

We recently showed that Pol ε binds ScCMG to form a CMGE complex [13]. Addition of ScPol ε to ScCMG revealed particles containing an extra density that unmistakably belongs to Pol ε in averaged EM images [16,26], in addition to the characteristic CMG particle (**Fig. 2a, Supplementary Fig. 2**). Although CMGE can form without the 80/75mer forked DNA [13], DNA was required to visualize CMGE in the EM and thus appears to stabilize CMGE to the negative stain procedure. We then determined the EM structure of the CMGE complex (**Fig. 2b,c**, **Supplementary Fig. 2**). Many features such as the N-terminal Zn-finger domains in Mcm subunits are resolved (**Fig. 2b,c and Supplementary video**). Similar to the CMG map (**Fig. 1**), the crystal structure of the human GINS complex fits remarkably well into the segmented GINS density in the CMGE 3D map (**Fig. 3a**) and the RecJ homologue structure fits into the density of Cdc45 (**Fig. 3b,c**) [23-25].

Comparing the structures of ScCMGE and ScCMG revealed the full density of Pol ε (**Fig. 2c**). The structure of the catalytic N-terminal half of the large Pol2 subunit of Pol ε in complex with primed DNA and dNTP has recently been solved [16,27]. Despite use of apo Pol ε in CMGE, the Pol2-DNA-dNTP crystal structure docks 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 ε in CMGE is underestimated in the EM density by about 30%, lending uncertainty to this assignment. Contributing to a somewhat lower density of Pol ε in CMGE may be a small percentage of CMG particles contaminating the CMGE dataset. For example, the profile of the top and bottom views of CMG and CMGE are quite similar because Pol ε is positioned directly over CMG. Thus computational sorting of some views is dependent upon slight intensity differences due to Pol ε, rather than the particle profile. Unintended inclusion of CMG images during 3D CMGE reconstruction would have the net effect of down-weighting the Pol ε density/size in the final CMGE 3D map. Alternatively, the peripheral features of a large complex, such as Pol ε holoenzyme at the edge of CMG may have some flexibility, leading to a reduction in the overall density/size of Pol ε. It is also possible that individual subunit flexibility may contribute to a lower density of Pol ε, leading to loss of a particular region (discussed again later). The crystal structure of the Dls1-Dpb4 complex of the CHRAC nucleosome remodeler [28] is homologous to the Dpb3-Dpb4 histone fold heterodimer of Pol ε, and may form the extended arm of density between GINS and Cdc45 (**e.g. Fig. 3b**). However, given their small size and EM resolution, assignment of Dpb3-Dpb4 within the Pol ε density must await higher resolution analysis [28].

Pol ε is on the C-side of CMG

The Mcm subunits are composed of C- and N-terminal domains, giving the Mcm 2-7 complex the appearance of two rings stacked on top of one another [22,29]. Pol ε 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^{2+} fingers which are in the N-terminal domains on the opposite side of the Mcms from Pol ε (**Supplementary video**). The position of Pol ε density on the C-side of CMG is also consistent with the subunit assignments determined in the DmCMG study, which would require a different arrangement in ScCMG to position the Pol ε 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 using the amine-specific cross-linker disuccinimidyl suberate (DSS) and applied high-resolution mass spectrometry to identify the cross-linked peptides [30]. As a control for this strategy, **Fig. 4a** shows the experimentally determined cross-links mapped onto the crystal structure of the catalytic domain of Pol2. The majority of the cross-linked lysines fall within 15-20 Å (Euclidean Cα-Cα distance) and fully 96% fall within 30 Å, which is the maximum reach of two lysine side chains coupled by DSS (**Supplementary Fig. 3)**. This result confirms that only nearby lysines are cross-linked and thus supports the use of intersubunit cross-links to report on subunit proximity and arrangement within CMGE. The intersubunit cross-links within CMGE are summarized in **Fig. 4b**; 553 unique cross-linked peptides were identified, 189 of which were intersubunit cross-links (comprehensive list is in **Supplementary Table 1**). Our intersubunit cross-link connectivity map (**Fig. 4b**) recapitulates the overall topology of CMGE deduced from the prior evidence available to us including the newly acquired EM data provided here. Thus, for example, the major cross-links between Mcm subunits confirm the established order of Mcm2-7 subunits [31]. In addition, several cross-links were detected across the Mcm ring (i.e, between Mcm6 and Mcm3) indicating that these specific regions are in close proximities (< 30 Å). Cross-links of the Mcms to the CMG accessory factors are also consistent with the subunit arrangement of DmCMG [20]. Hence, Mcm2 is the only Mcm subunit that forms cross-links to Cdc45, the Mcm3 and Mcm5 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) confirms that Pol ε lies on the C-terminal side of the Mcm ring and that the four subunits of Pol ε cross-link to CMG subunits in the regions expected from the Pol ε EM density (**Fig. 4c**). Thus, the C-terminal domains of Mcm2, Mcm6 cross-link to the C-half of Pol2; the N-half of Pol2 shows no cross-links to Mcm subunits, indicating that Pol2-Mcm interactions occur via the C-half of Pol2, which is proposed to encode an inactive polymerase. The C-terminal domain of Mcm5 cross-links to Dpb2, and there are no cross-links of any of the Pol ε subunits to the N-terminal domains of any of the Mcm subunits. The N-terminus of Dpb2 also cross-links to Pol2 and the C-terminus of Psf1, an interaction previously characterized using isolated domains [32]. Psf1 is located at the C-side of CMG (**Figs. 1-3**), and thus Dpb2-Psf1 cross-links support the position of Pol ε density is on the C-terminal side of CMG. Pol2 and the Pol ε accessory factors, Dpb3 and Dpb4, cross-link to Cdc45, although the Cdc45 cross-links do not provide information on the location of Pol ε on the N- or C-side of CMG because the structure of Cdc45 is unknown.

At the current level of resolution we cannot precisely locate the active site of Pol ε, although the distinct groove in the EM structure may correspond to this. 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. The previous cryoEM study of Pol2 showed a single globular structure, and therefore if the two polymerases within Pol2 are in separate domains, they must be close to one another[24]. In any case, the CX-MS data show that both the N- and C-halves of Pol2 are in close proximity to the accessory subunits of Pol ε holoenzyme and this restricts the non-visible region of Pol ε density to a region adjacent to the observed density of Pol ε. The reasoning is as follows. The cross-linking data show that all the subunits of Pol ε 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 Dpb2,3,4 subunits (**Fig. 4 and Supplementary Fig. 4**). This is what one would expect 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 non-visible portion of Pol ε must be proximal to the observed density. Thus the N-region of Pol2 containing the active site must be either: 1) within the observed density of Pol ε, 2) partially within the observed density, or 3) 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-inker, and the many cross-links between the N-half of Pol2 and other subunits of Pol ε, the N-region of Pol2 can not be 120-150 Å separate from the other subunits of Pol ε, which is the distance required to reach the bottom of CMG.

Pol α-Ctf4 attach to the opposite side of CMG from Pol ε

Ctf4 has recently been shown to form a trimer that can bind to both Pol α and the Sld5 subunit of GINS [14]. Therefore the position of Ctf4 on CMG serves as a proxy for the location of Pol α 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**). Adding both Ctf4 and Pol α to CMG resulted in a CMG-Ctf4-Pol α complex, although the Pol α density adjacent to Ctf4 was very weak (**Fig. 5b**, **Supplementary Fig. 5c**). Adding both Pol ε and Ctf4 to CMG resulted in a super-ternary complex we refer to as CMGE-Ctf4, with densities on diametrically opposed sides of CMG (**Fig. 5c, Supplementary Fig. 5e**). The N-terminal half of Ctf4 is connected by a flexible linker to the C-terminal half of Ctf4 that forms the trimer [14], and this may account for the fuzzy appearance of Ctf4. Most of these particles were side views on the carbon substrate, making 3D reconstruction of the complexes unfeasible. Well-defined 2D class averages of the CMG-Ctf4 complex, the super ternary complexes CMGE-Ctf4 and CMG-Ctf4-Pol α, and the super quaternary CMGE-Ctf4-Pol α complex, establish the relative position of Ctf4 and Pol α with respect to the CMG architecture (**Fig. 5a-c and Supplementary Fig. 5a-e**).

The architecture of a replisome is unknown for any cell type; bacteria, archaea or eukaryotes. Current models of replisome action in all cells place the polymerases and primase in back of the helicase because these enzymes require ssDNA as a template [1,13,20,33-35]. The eukaryotic replisome structure reported here finds that Pol ε and Pol α are on opposite sides of the helicase, 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 Pols 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 ε triggers a checkpoint response. However, we note that Pol ε binds CMG through several observable touch points, and thus a gross relocation of Pol ε to the Pol α-Ctf4 side of CMG would require breaking these several connections and establishing new ones. To determine

whether a gross rearrangement occurs upon engaging an active DNA fork, we assembled CMG-Pol ε-Ctf4-Pol α on a 160/91mer primed DNA fork that we have previously validated to be active in replication assays [35]. We also confirmed that CMGE binds the 160/91mer primed fork under the conditions used here (**Supplementary Fig. 1c)**. Examination of 2D class average images of CMGE and the CMGE-Ctf4-Pol α complex show the same general outline as complexes using the unprimed 80/75mer fork (**Fig. 3d,e**).

DISCUSSION

Pol ε rides ahead of the unwinding point at the fork

The current study reveals that the leading strand Pol ε and lagging strand Pol α 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 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, due to the bilobed structure of the Mcm subunits (**Figs. 1a and c**) [22,29]. The widely accepted view of Pol ε below CMG would require the leading strand to enter the Mcms from the NTD tier. However, studies of both the archeal Mcm and *Drosophila* CMG demonstrate that the leading strand enters the CTD tier of the Mcm complex [20,36]. This DNA path places the leading strand Pol ε ahead of the forked junction (**Fig. 6**), for a completely unanticipated replisome architecture.

Given the surprising arrangement of Pol ε ahead of CMG helicase, and reported DNA path, the leading strand ssDNA will need to traverse the ~ 110 Å central chamber of the Mcms, then bend back an additional 110 Å to reach Pol ε for a total of 220 Å, or about 40 nucleotides. The ssDNA may traverse the outside of CMG to reach Pol ε, or could thread through the second channel in CMG formed by the Cdc45/GINS accessory proteins (**as illustrated in Fig. 6**). Indeed, a recent study of *Drosophila* CMG demonstrates the leading strand can occupy the second channel under particular conditions [37]. Alternatively, the leading ssDNA may take a ~20 nucleotide path by exiting the Mcm channel at an internal position (e.g. at the Mcm2/5 "gate"), and to bend up toward Pol ε (**Supplementary Fig. 6a**). In the event the N-half of Pol2 is a separate domain and flexibly connected to the C-half, its location within cross-linking distance of Pol ε holoenzyme subunits place it near the visible density of Pol ε at the top of CMG and still requires DNA to make a U-turn from the MCMs (**Supplementary Fig. 6b**). Interestingly, experimental support for a 20-40 nucleotide leading strand ssDNA gap at the fork stem from studies in the *Xenopus* system that block replisome advance with an interstrand cross-link [34]. The 10-30 min time points show 20-40 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 needs to be updated [1,13,20,33-35].

To what extent does Pol ε perform leading strand synthesis?

Pol ε was initially discovered as a third essential polymerase in budding yeast [38], and its role in replication continues to be studied extensively. Recent studies show that mutations in Pol ε are associated with some types of cancer, similar to cancers associated with mismatch repair

and Pol η defects [39]. Mutations in active site residues of Pol ε are lethal, indicating that Pol ε is required to synthesize DNA during chromosome replication [40]. 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, suggesting that another DNA polymerase can substitute for Pol ε, but probably does not reflect normal replication [40-42]. These observations are reminiscent of genetic studies in *E. coli* in which *dnaE*, encoding the polymerase that normally duplicates both strands of the chromosome, can be mutated and the cells survive via replication by Pol I, but the *dnaE* mutant cells grow slowly [43]. Interestingly, the inactive C-half of Pol2 is essential, presumably serving a structural role [38,39].

Several genetic studies using a slightly altered Pol ε that provides a mutation signature on the DNA product indicate that Pol ε predominates over Pol δ on the leading strand and that Pol δ predominates over Pol ε on the lagging strand [6,9-12]. This is consistent with studies of proofreading mutants in Pols ε and δ that conclude that the two Pols act on different strands [44]. Use of an altered Pol δ indicates its predominant role in bulk lagging stand synthesis in both *S. cerevisiae* and *Schizosaccharomyces pombe* [6,10-12]. Use of a Pol ε mutant that frequently misincorporates rNMPs enabled genome wide analysis of strand bias by Pol ε and supports the conclusion that Pol ε performs bulk leading strand synthesis in both budding and fission yeast [11,12]. Biochemical studies of Pols ε and δ with ScCMG are consistent with these cellular findings, as Pol ε binds CMG directly, is more active in leading strand synthesis with CMG compared to Pol δ, and Pol δ is more active on the lagging strand of CMG driven forks than Pol ε [13,35,45]. Earlier biochemical assays of Pols ε and Pol δ also indicated their properties were most consistent with leading and lagging strand synthesis, respectively [46,47]. Genome wide chip assays of polymerase occupancy in budding yeast show that Pol ε cross-links specifically to the leading strand, while Pol δ cross-links to the lagging strand [48]. However, it has recently been argued that cross-linking studies may bias Pol δ to the lagging strand because it must extend multiple Okazaki fragments [49,50]. An interesting recent report arrives at a different conclusion from all the previous work, indicating that Pol δ performs bulk leading and lagging strand synthesis, similar to the SV40 viral system [49]. The different conclusions are explained by the fact that the cellular studies are performed in strain backgrounds with mutations in various repair pathways, and that these mutations may introduce strand bias into the results [50]. Hence, the extent to which Pol ε and Pol δ perform leading strand synthesis remains an open question that requires further study.

Function of the replisome architecture

The anti-intuitive position of Pol ε at the "top" of CMG suggests that an unforeseen function may underlie the unexpected replisome architecture. While one cannot *a priori* ascertain the function this architectural facet serves, 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 (e.g. dUMP, rNMP) before they enter the Pol ε active site. Yet another possibility is that Pol ε is pushed by CMG ATP hydrolysis, and this may cause Pol ε 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 ε. Particular mutations in Pol2, or deletions of Dpb3 or Dpb4 result in loss of epigenetic silencing in yeast [51]. Indeed, Pol ε has been shown to bind histones, modifications of

which underlie epigenetic silencing [52]. We note that Dpb3/4 is a histone fold heterodimer, and both TFIID of RNAP II, and the CHRAC chromatin remodeler contain histone fold heterodimers that are thought to help mobilize nucleosomes. Hence, placement of Pol ε 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 are required to understand the functional implications of the unexpected replisome architecture reported here.

METHODS

Methods and any associated references are available in the ONLINE METHODS SECTION.

Accession Codes The 3D EM maps of CMG and CMGE have been deposited at the EMDB database with accession codes 6463 and 6465.

Acknowledgements We would like to thank O. Yurieva and D. Zhang for the purification of CMG, Pol α, Pol ε 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.O.D.) and Howard Hughes Medical Institute (M.O.D.)

Author Contributions J.S., R.G., Y.S, B.T.C. H.L. and M.O.D. designed experiments. J.S., Y.S., R.G. and Z.Y. performed experiments. J.S., R.G., Y.S., Z.Y., B.T.C., H.L. and M.O.D. analyzed the data. H.L and M.O.D. prepared the manuscript with input from all authors.

Competing financial interests. The authors declare no competing financial interests.

Author Information Correspondence and request for materials should be addressed to M.O.D. (odonnell@rockefeller.edu) or H.L. (hli@bnl.gov).

REFERENCES

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**(2013).
2. Johansson, E. & Dixon, N. Replicative DNA polymerases. *Cold Spring Harb Perspect Biol* **5**(2013).
3. Bell, S.D. & Botchan, M.R. The minichromosome maintenance replicative helicase. *Cold Spring Harb Perspect Biol* **5**(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**, 247-58 (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 U S A* **103**, 10236-41 (2006).
6. Kunkel, T.A. & Burgers, P.M. Dividing the workload at a eukaryotic replication fork. *Trends Cell Biol* **18**, 521-7 (2008).

7. Waga, S. & Stillman, B. The DNA replication fork in eukaryotic cells. *Annu Rev Biochem* **67**, 721-51 (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**, 127-30 (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**, 137-44 (2008).
11. Clausen, A.R. et al. Tracking replication enzymology in vivo by genome-wide mapping of ribonucleotide incorporation. *Nat Struct Mol Biol* **22**, 185-91 (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 U S A* **111**, 15390-5 (2014).
14. Simon, A.C. et al. A Ctf4 trimer couples the CMG helicase to DNA polymerase alpha in the eukaryotic replisome. *Nature* **510**, 293-7 (2014).
15. Kang, Y.H. et al. Interaction between human Ctf4 and the Cdc45/Mcm2-7/GINS (CMG) replicative helicase. *Proc Natl Acad Sci U S A* **110**, 19760-5 (2013).
16. Hogg, M. et al. Structural basis for processive DNA synthesis by yeast DNA polymerase varepsilon. *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**, 979-86 (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**, 1978-87 (2009).
19. Costa, A. et al. The structural basis for MCM2-7 helicase activation by GINS and Cdc45. *Nat Struct Mol Biol* **18**, 471-7 (2011).
20. Costa, A. et al. DNA binding polarity, dimerization, and ATPase ring remodeling in the CMG helicase of the eukaryotic replisome. *Elife*, 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**, 944-51 (2013).
22. Sun, J. et al. Structural and mechanistic insights into Mcm2-7 double-hexamer assembly and function. *Genes Dev* **28**, 2291-303 (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 U S A* **104**, 12685-90 (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**, 1316-21 (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**, 388-96 (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**, 9886-96 (2005).
29. Li, N. et al. Structure of the eukaryotic MCM complex at 3.8 A. *Nature* **524**, 186-91 (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**, 2927-43 (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**, 4491-9 (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**, 543-52 (2013).
33. Yardimci, H. et al. Bypass of a protein barrier by a replicative DNA helicase. *Nature* **492**, 205-9 (2012).
34. Fu, Y.V. et al. Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase. *Cell* **146**, 931-41 (2011).
35. Georgescu, R.E. et al. Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. *Nat Struct Mol Biol* **21**, 664-70 (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**, 34229-34 (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 U S A* **112**, E249-58 (2015).
38. Morrison, A., Araki, H., Clark, A.B., Hamatake, R.K. & Sugino, A. A third essential DNA polymerase in S. cerevisiae. *Cell* **62**, 1143-51 (1990).
39. Henninger, E.E. & Pursell, Z.F. DNA polymerase epsilon and its roles in genome stability. *IUBMB Life* **66**, 339-51 (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**, 22283-8 (1999).
41. Kesti, T., Flick, K., Keranen, S., Syvaoja, J.E. & Wittenberg, C. DNA polymerase epsilon catalytic domains are dispensable for DNA replication, DNA repair, and cell viability. *Mol Cell* **3**, 679-85 (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**, 28099-108 (2002).
43. Niwa, O., Bryan, S.K. & Moses, R.E. Alternate pathways of DNA replication: DNA polymerase I-dependent replication. *Proc Natl Acad Sci U S A* **78**, 7024-7 (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**, 717-26 (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**, 2764-73 (2004).
47. Burgers, P.M. Polymerase dynamics at the eukaryotic DNA replication fork. *J Biol Chem* **284**, 4041-5 (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**, 551-63 (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**, 163-75 (2015).
50. Stillman, B. Reconsidering DNA Polymerases at the Replication Fork in Eukaryotes. *Mol Cell* **59**, 139-41 (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**, 217-27 (2004).
52. Tackett, A.J. et al. Proteomic and genomic characterization of chromatin complexes at a boundary. *J Cell Biol* **169**, 35-47 (2005).

FIGURE LEGENDS

Figure 1. Structure of the *S. cerevisiae* CMG helicase. (a) Surface-rendered and segmented ScCMG 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 ID 2E9X), Cdc45 homolog (PDB ID 1IR6), and six copies of the *Sulfolobus solfataricus* MCM monomer crystal structure (PDB ID 3F9V). (**d**) Docking in the Mcm2-7 region of the NTD hexamer crystal structure of the *Sulfolobus solfataricus* MCM (PDB ID 2VL6). The green dot in (**b**) and (**c**) marks the unoccupied density between assigned GINS and Cdc45 that may be the truncated CTD of GINS subunit Psf1 (Psf1-CTD), but may also belong to Cdc45.

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 ε (bottom row). (**b**) Surface rendered EM map in five views. Display threshold is set to include the predicted molecular mass. (**c**) Segmented map. The Pol ε density is shown in green. 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 Dpb3/4 histone fold heterodimer subunits. Assignments of subunits in the Pol ε density are not certain as explained in the text. The six "Zn" labels in the bottom NT view denote the N-terminal Zn finger domains of Mcm proteins (see also the Supplemental video).

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 ID 2E9X) fit well in the EM density. The four GINS subunits are colored red (Psf1), green (Psf2), blue (Psf3), and orange (Sld5) respectively. The red spheres show the last residue in the CTD-truncated Psf1 crystal structure. The magenta spheres show the N-terminal first 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 ID 1IR6, cyan). **(c)** Back side view of the catalytic core of the RecJ exonuclease homolog to Cdc45 (PDB ID 1IR6, cyan) adjacent to the *Sulfolobus solfataricus* MCM monomer crystal structure (PDB ID 3F9V) labeled "M2". **(d)** Crystal structure of the Pol2 catalytic N-terminal domain complexed with primed DNA and dNTP (PDB ID 4M8O) is docked such that DNA aligns into the large groove; this is speculative and should be regarded as tentative due to the underweight density of Pol ε in the CMGE structure. Template and primer strand DNA is in red and blue, respectively.

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 followed by identification of cross-linked peptides by mass spectrometry. **(a)** Overview of cross-links observed within the region of Pol2 corresponding to the crystal structure (PDB 4M8O). 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: Mcms (yellow), GINS (blue), Cdc45 (purple), and Pol ε (green). **(c)** The upper left view illustrates major cross-links observed to connect GINS and Cdc45 to Mcm3/5 and Mcm2, respectively. Remaining views illustrate cross links between Pol ε and CMG, clockwise: Putative Dpb3/Dpb4 to Cdc45. Pol2 C-terminal region cross-links to the C-terminal regions of Mcm2 and Mcm6, and Cdc45. Dpb2 cross-links to the Mcm5 C-terminal region and to Psf1 of GINS.

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

Figure 6. Architecture of the eukaryotic replisome. Replisome structure and the proposed DNA path through the replisome. Pol α 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 leading strand DNA path through the entire Mcm ring and then bending back up to Pol ε, requiring about 40 ntd of ssDNA. Leading ssDNA is illustrated as going completely through the Mcm2-7 complex, then bending up through the second "accessory" channel of CMG, but this path is speculative. Other DNA paths are possible. See text and supplementary Fig. 6 for further details.

ONLINE METHODS

Sample preparation and electron microscopy

Saccharomyces cerevisiae CMG, Pol ε, Pol α, and Ctf4 were purified according to published procedures [13,35]. Pol ε contained two amino acid replacements that eliminate the proofreading exonuclease activity (D275S, E277S) [53]. A small unprimed 80/75mer fork DNA was prepared upon annealing equimolar amounts of two DNA oligomers: 80mer leading strand oligo (5' - cccccaccgatgtggtaggaagtgagaattggagagtgtgtttttttttttttttttttttttttttttttttttttttttttt - 3') with lagging strand 75mer oligo (5' - ttttttttttttttttttttttttttttttttttttttttttcacactctccaattctcacttcctaccacatcggt - 3') to form a fork with a 40-bp duplex and arms of 40 dT and 35 dT. A primed 160/91mer fork was prepared by annealing the NB1T 91mer (lagging strand), NB160mer (leading strand) and C2 37mer primer as described in our earlier report [2]. The primed 160/91mer fork was demonstrated to be active with CMG and Pol ε in replication assays as described[2]. Proteins were freshly dialyzed just before each sample preparation to remove the glycerol present in the samples to a value less than 0.5% (concentration verified using an Abbé refractometer). The dialysis buffer contained 20 mM Tris-Acetate, 40 mM K-Glutamate, 2 mM DTT and 0.1 mM EDTA. We first mixed 3 µl of 0.75 µM (0.6 mg/ml) CMG with 0.5 µl of 9 µM 80/75mer unprimed fork DNA (or equivalent 160/91mer primed fork, as indicated), then added AMP-PNP and MgOAc to 1mM and 8 mM, respectively, in a final volume of 3.8 µl and incubated the mixture on ice for 20 min. To form larger CMG-containing complexes, we subsequently added one or more components: (1) 0.4 µl of purified Pol ε at 12.5 µM; (2) 0.5 µl of purified Ctf4 trimer at 7.5 µM; (3) 0.9 µl of Pol α at 5 µM. In the case of the 160/91mer primed fork, 60 µM dATP and dCTP, 0.5 µl 1 µM RFC and 0.5 µl 8 µM PCNA were present [13,35]. To prepare EM grids, we diluted each sample with fresh dialysis buffer used in glycerol removal. We applied 3 µl of CMG sample at a final concentration of 45 µg/ml (57 nM) to glow-discharged carbon coated EM grids. For EM grid preparation of larger CMG-containing complexes, we maintained their concentrations similar to the CMG sample. We left the sample droplet on grids for 1 min, then blotted with a piece of filter paper and immediately added 3 µl of 1% uranyl acetate aqueous solution, then blotted the stain, and added a second drop of stain solution, and waited for 1 min, and finally blotted most of the stain solution with the edge of the filter paper, leaving only a very thin layer of stain solution on the grid. We fanned the EM grid with a sheet of paper to speed the drying process. JEM-2010F with a Gatan 626 holder was used to exam the EM grids. The microscope was operated at 200 kV in the low dose mode. Images were recorded in a defocus range of -0.5 to -4 µm on a 4k by 4k Gatan Ultrascan CCD camera at the instrument magnification of 50,000 with an electron dose of 20 $e^-/Å^2$. The image sampling size was 2.12 Å/pixel at this magnification.

Image processing and 3D reconstruction

We collected > 300 4k x 4k CCD images for each sample, ensuring that there are more than 10,000 raw particle images in each dataset. Image processing was done mainly in program EMAN2.1 [54], and RELION1.3 [55]. Several particle-picking methods were tried, including manually or automatically with different options. The “GAUSS” option in EMAN2 "interactive Boxing-e2boxer" worked best for the data and we set “Threshold Low” to 1.15, “Threshold High” to 1.8, and used default values for other parameters. If too many noise regions were picked, “Threshold Low” would be increased, and if the image contained dark stain or large aggregates the “Threshold High” would be reduced. If particles selected with this option do not result in meaningful 2D class averages, it was concluded that the quality of raw images was

insufficient for further analysis and therefore we would return to wet-lab experiments to purify and improve the protein sample and EM grid preparation. When the particle selection result was satisfactory, we saved the particle coordinates without saving the raw particle images. We then used command line program "e2boxer.py" with gauss option to apply the coordinates and window out particles from all micrographs. Selecting particles by this method takes less than 10 minutes of manual analysis, after which the program will run automatically. This approach not only saves time but also avoids potential bias when particles are picked manually or based on references. After contrast transfer function (CTF) correction, all the particles were combined into one data set, mean-shrunk by a factor of 2 to 4.24 Å/pixel to speed up image classification. Reference free 2D classification and averaging were performed multiple times to check the quality of the data and also clean up bad particles or contamination. If conformational flexibility or different structures was observed at this stage, "e2refinemult" in EMAN2 were used to separate the data into several subsets.

After reference-free 2D classification-based particle cleanup, the dataset of CMG, CMG-E had 13483, 38425 particles, respectively. Since the 2D averages of CMG in different views were essentially the same as the published Drosophila CMG data [20], we initially used the low-pass filtered (60 Å) Drosophila CMG model (EMD-2772) as a starting model for multiple-model 3D classification. We derived four 3D models from the CMG dataset: two models were highly similar and their associated particles were combined for further refinement; two remaining models were distorted and those particles were discarded. For CMGE, we found that three models were either broken complexes or entirely missing the Pol ε density; only one model had full Pol ε density (Extended Data Fig. 5). The later model and associated particles were selected for further refinement. After 3D classification, the final dataset used for 3D reconstruction of the CMG and CMGE structures had 8781 and 18721 particles, respectively. Single model refinement was then performed with the dataset separated from the preceding 3D classification. The 3D maps of CMG and CMGE were both refined in RELION 1.3 and had an estimated resolution of 18 Å and 16 Å, respectively, based on a 0.143 threshold in the gold-standard Fourier shell correlation (Extended Data Fig. 2, 6).

UCSF Chimera was used for surface rendering of the 3D maps and docking of the crystal structures [56]. Density segmentation used the watershed algorithm based built-in function Segger (v1.6) in Chimera. Grouping of segmented density and identity assignment in the CMG region largely followed the assigned architecture of DmCMG [20]. In the Pol ε region, the rod like density between GINS and Cdc45 was assigned Dpb3/4 based on the similarity to the crystal structure of the homologous Drosophila Dls1/Dpb4 crystal structure. Rigid-body docking of the crystal structure of the Pol2 NTD was tentative as the structure did not fit well in the EM density. This was likely due to the fact that Pol2 in our CMGE structure was in the apo form, a state that was different from the primed DNA-bound state in the crystal structure (PDB ID 4M8O).

Five structures were characterized only at the level of 2D classification and averaging without further 3D reconstruction, because their corresponding samples were either too heterogeneous or had preferred orientations on EM grids. They were CMG-Ctf4, CMGE-Ctf4, and/91mer primed fork DNA. After 2D classification-based particle cleanup, the dataset for samples containing 80/75mer fork DNA had 12295 (CMG-Ctf4), 25482 (CMGE-Ctf4), 33253 (CMG-Ctf4-Pol α) particles, respectively. The particle number for the dataset containing primed forked DNA was 20188 (CMGE) and 8469 (CMGE-Ctf4-Pol α), respectively. Computational image analyses of these structures were done in EMAN2.

CMG interaction with forked-DNA templates by electrophoresis mobility shift assay (EMSA)

The binding of CMG to the 80/75mer unprimed fork and the 160/91mer primed fork was tested in EMSA assays. For EMSA assays the 5' terminus of the leading strand of each forked DNA was labeled with ^{32}P-γ-ATP using T4-PNKinase (NEB). First we determined the K_d, using a low concentration of primed forked DNA, and then performed EMSAs at high concentrations, about 2-fold lower than used to prepare samples for the EM. For the K_d measurement, The CMG titration was performed by adding increasing amounts of CMG into reactions containing 1 nM of primed forked-DNA in a buffer containing 25 mM Hepes pH 7.5, 10% glycerol, 40 mM NaAcetate 8 mM MgAcetate 1 mM DTT, 250 mg/mL insulin and 0.2 mM PMSF in a total volume of 10 ml. In all reactions, we added 0.5 mM AMP-PNP and samples were incubated a further 30 minutes, then loaded onto a 4% native PAGE-TBE gel prepared with 8 mM MgAcetate. The gel was run at 120V for 2 hours using 0.5X TBE running buffer supplemented with 8 mM MgAcetate. We then performed EMSA assays using 80/75mer fork and 160/91mer primed fork DNAs under the conditions and higher DNA/CMG concentrations similar to those used for EM. Thus we mixed DNA and CMG (100:200 nM) using the same buffer conditions as those used in preparation of EM samples. The EMSA of these reactions are shown in **Supplemental Figure 1**, which shows nearly complete binding of the primed and unprimed forked DNAs to CMG. The EM conditions used 6.8-12-fold higher concentrations of DNA and 1.5-3-fold higher CMG, which should push the equilibrium even further to the full complex.

To derive the K_d value of CMG binding to DNA from the EMSA titration, the native gel autoradiograph was analyzed assuming a 1:1 binding event (CMG + DNA ↔ CMG-DNA) and the equilibrium dissociation constant, K_d, is defined by the equation: K_d = [CMG][DNA]/[CMG-DNA], where [CMG], [DNA], and [CMG-DNA] are the molar concentrations of free CMG, free DNA and bound complex at equilibrium, respectively. Since 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 = [CMG-DNA]/([DNA] + [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, as the titration is performed at 1 nM primed forked DNA which is approximately 30 fold below the K_d, such that P_t approximates the free protein concentration at equilibrium. The data was fitted using the Matlab software. The observed K_d value of yeast CMG to primed forked DNA was 29.7 +/- 2.2 This value is similar to the K_d values of CMG-DNA binding obtained in studies of *Drosophila* CMG that estimate a binding affinity of Dm CMG to forked DNA of approximately 10-20 nM [4].

Chemical cross-linking with mass spectrometry readout (CX-MS)

The CX-MS procedure was performed essentially as described [30,57]. CMG-E 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. CMG-E was cross-linked using 2 mM disuccinimidyl suberate (Creative Molecules) for 30 minutes at 25 °C with constant agitation (1,200 rpm). The reaction was then quenched in 50 mM ammonia bicarbonate. Approximately 50-100 μg cross-linked complex was resuspended and heated in 100-200 μl 2X 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 kD was sliced and digested in-gel using trypsin to release the cross-link peptides. The resulting proteolytic peptide mixture was dissolved in 20 μl 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) using off-line HPLC separation with an auto sampler (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 cross-linked peptide identification, the purified peptides were dissolved in the sample loading buffer (5% MeOH, 0.2% FA) and analyzed by a LTQ Velos Orbitrap Pro mass spectrometer or an Oribtrap Q Exactive (QE) Plus mass spectrometer (Thermo Fisher). For the analysis by 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 μm tip, New Objective). The column was packed with 8 cm of reverse-phase C18 material (3μm porous silica, 200 Å pore size, Dr. Maisch GmbH). Mobile phase A consisted of 0.5% acetic acid and mobile phase B of 70% ACN with 0.5% acetic acid. The peptides were eluted in a 120 minute LC gradient (8% B to 50% B, 0-93 minutes, followed by 50% B to 100% B, 93-110 minutes and equilibrated with 100% A until 120 minutes) using a 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, where 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 7,500 for MS2. Ions (370-1700 m/z) with charge state of > 3 were selected for fragmentation. A dynamic exclusion of (15 s / 2 / 60 s) was used. Other instrumental parameters include: “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 x 10^5 and 1 x 10^6 respectively for the linear ion trap and Orbitrap. For MS2, the Orbitrap ion accumulation limit was 5 x 10^5. The maximum ion injection time for Orbitrap was 500 - 700 milliseconds. 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 μm particle size, 200 Å pore size, and 50 μm × 15 cm, Thermo fisher) and eluted using a 120-min LC gradient (2% B to 10% B, 0-6 minutes, 10% B – 35% B, 6-102 minutes, 35% B- 100% B, 102- 113 minutes followed by equilibration, where mobile phase A consisted 0.1% formic acid and mobile phase B consisted 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 [58] with a database containing sequences of the 15 protein subunits of yeast CMG-E complexes. Other search parameters included: mass accuracy of MS1 ≤ 10 ppm (parts per million) and MS2 ≤ 20 ppm for the initial database search, cysteine carboxymethylation as a fixed modification, methionine oxidation as a variable modification, and a maximum of one trypsin miscleavages was allowed. The results were filtered at 5% false discovery rate (FDR) and false positives were then identified by manual verification as previously described [30,59]. Briefly, the primary mass spectrometry data was initially analyzed by software that predicted a 5% FDR as an initial filter. As many of these are still false positives, the data is 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 likely arise from residual under-cleaved tryptic peptides, non-tryptic 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, e.g., extensive fragmentation coverage of both peptide chains, and a minimum of 4 amino acids for both of the crosslink peptide chains), with the result that we generally discard an additional 20-30% of this data[10,12]. We adopt this stringent post-filtering strategy to reduce the likelihood of false positive cross-link identifications. Thus the final FDR in our crosslink dataset is expected to be significantly smaller than 1%. A total of 553 unique cross-linked peptides were identified as a result (**Supplemental Table I**).

References for Online Methods

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 U S A* **88**, 9473-7 (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**, 519-30 (2012).
56. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-12 (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**, 904-6 (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**, 1028-34 (2014).

FIGURES

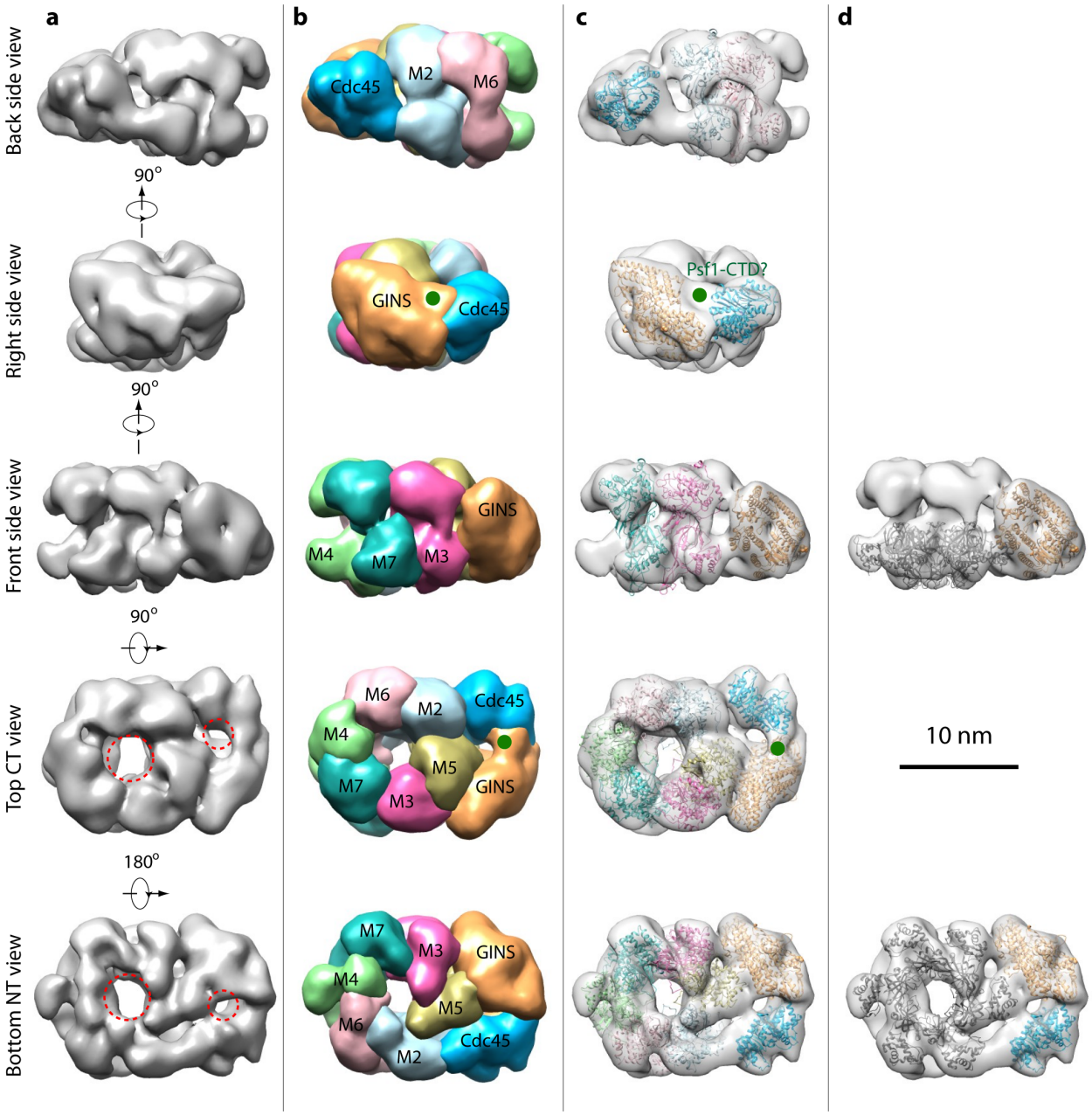

Figure 1. Structure of the *S. cerevisiae* CMG helicase. (a) Surface-rendered and segmented ScCMG structure in different views of the 3D EM map. The larger dashed red circle marks the apparent hole in the middle of Mcm2-7, and the smaller circle indicates the apparent second hole between GINS-Cdc45 and Mcm2-7. (**b**) Corresponding views of the segmented map. (**c**) Docking of the crystal structures of the GINS (PDB ID 2E9X), Cdc45 homolog (PDB ID 1IR6), and six copies of the *Sulfolobus solfataricus* MCM monomer crystal structure (PDB ID 3F9V). (**d**) Docking in the Mcm2-7 region of the NTD hexamer crystal structure of the *Sulfolobus solfataricus* MCM (PDB ID 2VL6). The green dot in (**b**) and (**c**) marks the unoccupied density between assigned GINS and Cdc45 that may be the truncated CTD of GINS subunit Psf1 (Psf1-CTD), but may also belong to Cdc45.

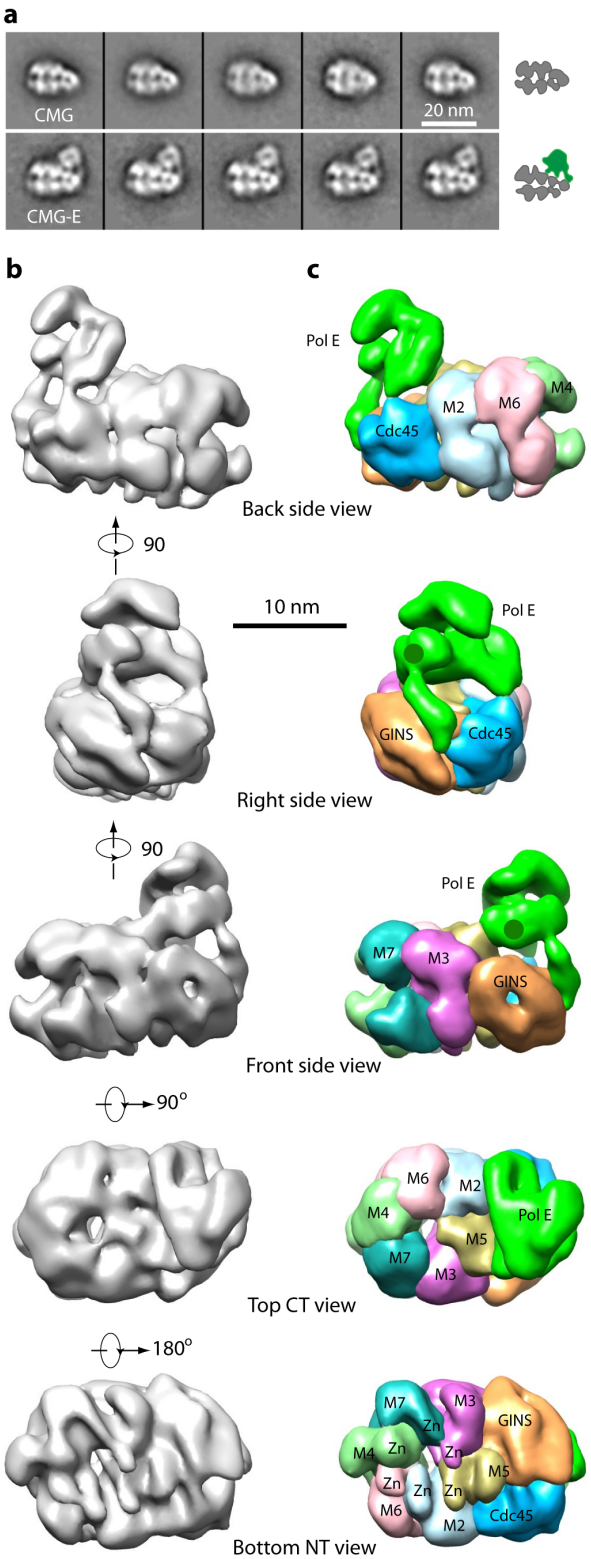

Figure 2. Structure of the *S. cerevisiae* CMGE leading strand helicase-polymerase (**a**) Selected side views of reference-free class averages of CMG (top row) compared to CMG complexed with Pol ε (bottom row). (**b**) Surface rendered EM map in five views. Display threshold is set to include the predicted molecular mass. (**c**) Segmented map. The Pol ε density is shown in green. 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 Dpb3/4 histone fold heterodimer subunits. Assignments of subunits in the Pol ε density are not certain as explained in the text. The six "Zn" labels in the bottom NT view denote the N-terminal Zn finger domains of Mcm proteins (see also the Supplemental video).

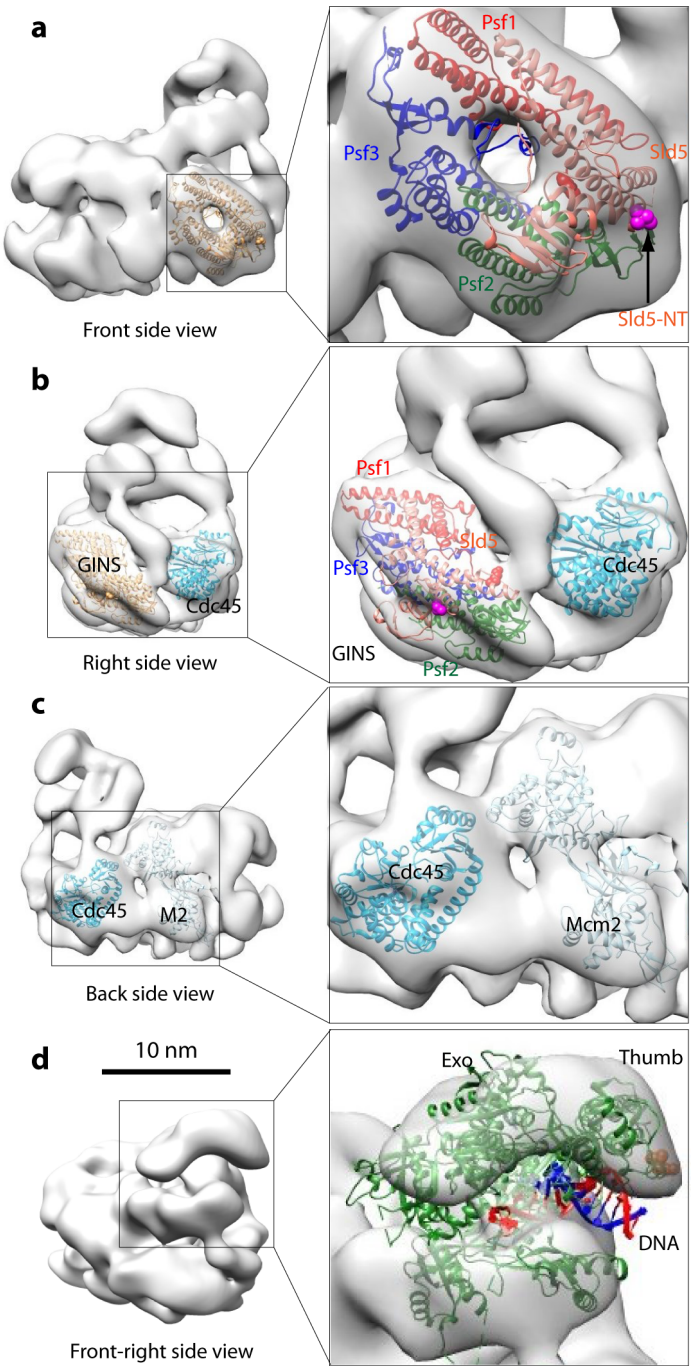

Figure 3. Rigid-body docking of CMG subunits into the CMGE density map with available crystal structures. (a) The crystal structures of human GINS complex (PDB ID 2E9X) fit well in the EM density. The four GINS subunits were colored red (Psf1), green (Psf2), blue (Psf3), and orange (Sld5) respectively. The red spheres show the last residue in the CTD-truncated Psf1 crystal structure. The magenta spheres show the N-terminal first 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 ID 1IR6, cyan). **(c)** Back side view of the catalytic core of the RecJ exonuclease homolog to Cdc45 (PDB ID 1IR6, cyan) adjacent to the *Sulfolobus solfataricus* MCM monomer crystal structure (PDB ID 3F9V) labeled "M2". **(d)** Crystal structure of the Pol2 catalytic N-terminal domain complexed with primed DNA and dNTP (PDB ID 4M8O) is docked such that DNA aligns into the large groove; this is speculative and should be regarded as tentative due to the underweight density of Pol ε in the CMGE structure. Template and primer strand DNA is in red and blue, respectively.

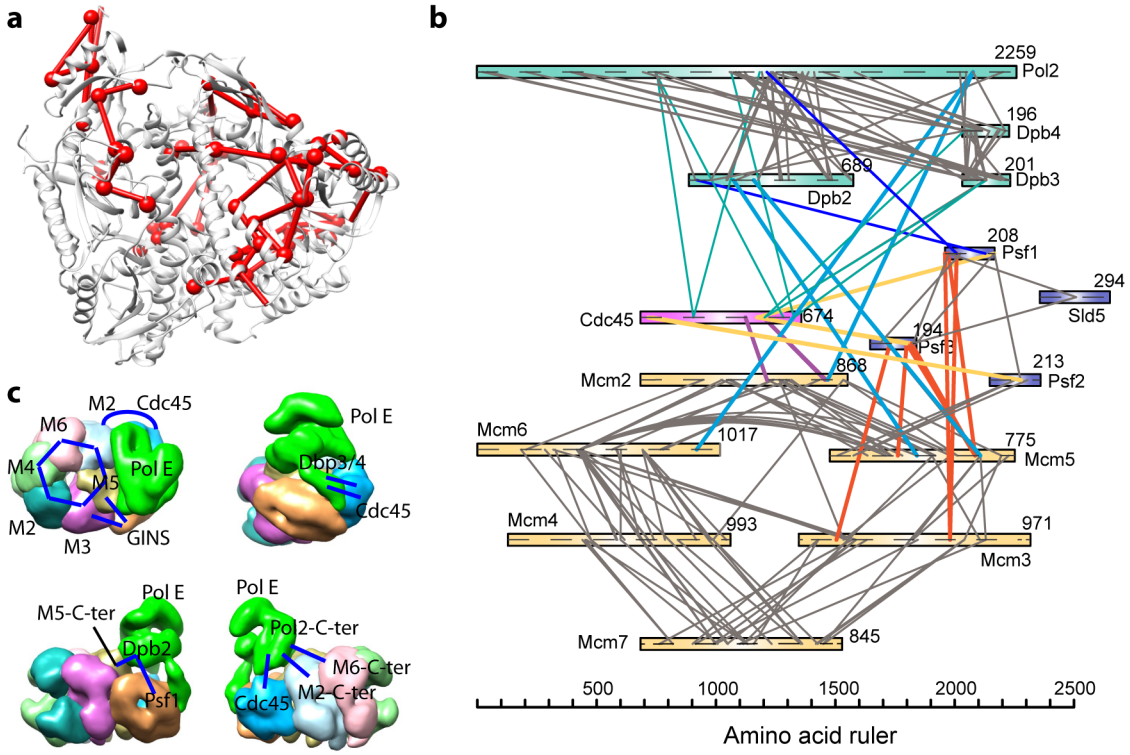

Figure 4. Subunit proximities within CMGE determined by chemical cross-linking with mass spectrometry readout (CX-MS). CMGE was cross-linked with a lysine specific bifunctional cross-linker, then fragmented by proteolysis followed by identification of cross-linked peptides by mass spectrometry. **(a)** Overview of cross-links observed within the region of Pol2 corresponding to the crystal structure (PDB 4M8O). 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: Mcms (yellow), GINS (blue), Cdc45 (purple), and Pol ε (green). **(c)** The upper left view illustrates major cross-links observed to connect GINS and Cdc45 to Mcm3/5 and Mcm2, respectively. Remaining views illustrate cross links between Pol ε and CMG, clockwise: Putative Dpb3/Dpb4 to Cdc45. Pol2 C-terminal region cross-links to the C-terminal regions of Mcm2 and Mcm6, and Cdc45. Dpb2 cross-links to the Mcm5 C-terminal region and to Psf1 of GINS.

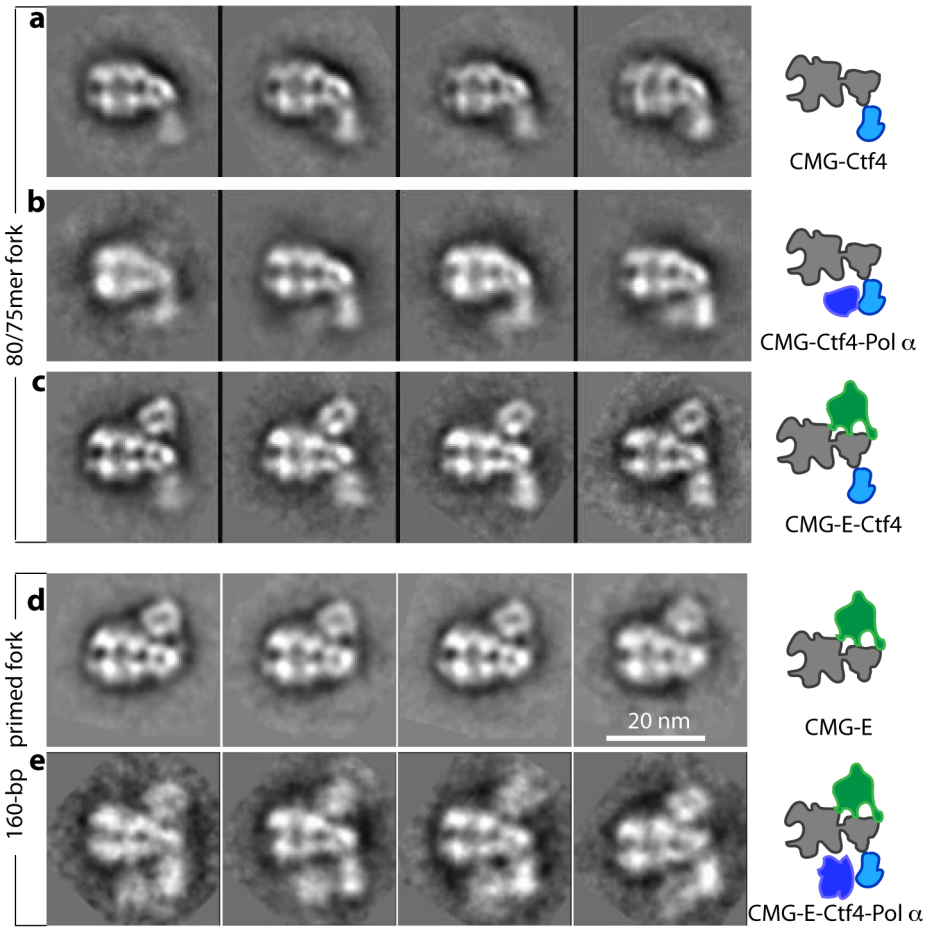

Figure 5. Staged assembly of the eukaryotic replisome. (a-c). Replisome reconstitution. Selected side views of reference-free class averages of: (**a**) CMG mixed with Ctf4 and 80/75mer, (**b**) CMG mixed with Ctf4, Pol α and 80/75mer, (**c**) CMG mixed with Pol ε, Ctf4 and 80/75mer. **(d)** CMG mixed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

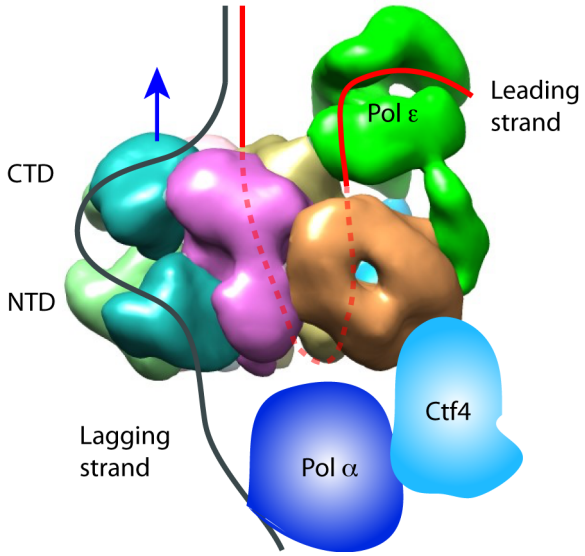

Figure 6. Architecture of the eukaryotic replisome. Replisome structure and the proposed DNA path through the replisome. Pol α is shown in blue. Ctf4 is in cyan. Red and black lines illustrate possible leading and lagging strand DNA. The blue arrow indicates the direction of replisome movement on DNA. The diagram indicates a long leading strand DNA path through the entire Mcm ring and then bending back up to Pol ε, requiring about 40 ntd of ssDNA. Leading ssDNA is illustrated as going completely through the Mcm2-7 complex, then bending up through the second "accessory" channel of CMG, but this path is speculative. Other DNA paths are possible. See text and supplementary Fig. 6 for further details.