THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

# Structural basis of Mcm2–7 replicative helicase loading by ORC–Cdc6 and Cdt1

**Citation for published version:**
Yuan, Z, Riera, A, Bai, L, Sun, J, Nandi, S, Spanos, C, Chen, ZA, Barbon, M, Rappsilber, J, Stillman, B, Speck, C & Li, H 2017, 'Structural basis of Mcm2–7 replicative helicase loading by ORC–Cdc6 and Cdt1' Nature Structural & Molecular Biology. DOI: 10.1038/nsmb.3372

**Digital Object Identifier (DOI):**
10.1038/nsmb.3372

**Link:**
Link to publication record in Edinburgh Research Explorer

**Document Version:**
Peer reviewed version

**Published In:**
Nature Structural & Molecular Biology

**General rights**
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 05. Apr. 2019

# Structural basis of MCM2-7 replicative helicase loading by ORC-Cdc6 and Cdt1

Zuanning Yuan[1,7], Alberto Riera[2,7], Lin Bai[1,7], Jingchuan Sun[1], Saikat Nandi[3], Christos Spanos[4], Zhuo Angel Chen[4], Marta Barbon[2], Juri Rappsilber[4,5,*], Bruce Stillman[3,*], Christian Speck[2,6,*], and Huilin Li[1,*]

[1] Cryo-EM Structural Biology Laboratory, Van Andel Research Institute, Grand Rapids, MI 49503, USA

[2] DNA Replication Group, Institute of Clinical Sciences (ICS), Faculty of Medicine, Imperial College London, London, W12 0NN, United Kingdom

[3] Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA

[4] Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh, United Kingdom

[5] Chair of Bioanalytics, Institute of Biotechnology, Technische Universität Berlin, Berlin, Germany

[6] MRC Clinical Sciences Centre (CSC), Du Cane Road, London W12 0NN

[7] These authors contributed equally to this work.

* Corresponding authors: juri.rappsilber@ed.ac.uk, stillman@cshl.edu, chris.speck@imperial.ac.uk, Huilin.Li@vai.org

## Abstract

To start DNA replication, the Origin Recognition Complex (ORC) and Cdc6 load a Mcm2-7 double hexamer onto DNA. Without ATP hydrolysis, ORC-Cdc6 recruits one Cdt1-bound Mcm2-7 hexamer, forming an ORC-Cdc6-Cdt1-Mcm2-7 (OCCM) helicase loading intermediate. Here we report a 3.9Å structure of the OCCM on DNA. Flexible Mcm2-7 winged-helix domains (WHD) engage ORC-Cdc6. A three-domain Cdt1configuration embraces Mcm2, Mcm4, and Mcm6, nearly half of the hexamer. The Cdt1 C-terminal domain extends to the Mcm6 WHD, which binds Orc4 WHD. DNA passes through the ORC-Cdc6 and Mcm2-7 rings. Origin DNA interaction is mediated by an α-helix in Orc4 and positively charged loops in Orc2 and Cdc6. The Mcm2-7 C-tier AAA+ ring is topologically closed by a Mcm5 loop that embraces Mcm2, but the N-tier ring Mcm2-Mcm5 interface remains open. This structure suggests loading mechanics of the first Cdt1-bound Mcm2-7 hexamer by ORC-Cdc6.

## INTRODUCTION

The *S. cerevisiae* Origin Recognition Complex (ORC) is an ATPase complex composed of Orc1-6 [1-3]. The composition and architecture of ORC is conserved in all eukaryotes [4]. Low resolution electron microscopy (EM) showed that the six subunits are arranged into a crescent in the order of Orc1-Orc4-Orc5-Orc3-Orc2, with Orc6 binding to Orc2/Orc3 [5,6]. This architecture is confirmed by a recent crystal structure of an inactive *Drosophila* ORC (DmORC) [7]. This DmORC core is a notched two-tiered ring composed of the N-tier ring of five AAA+ domains and the C-tier ring of five winged helix (WH) domains of Orc1-5. Because the DmOrc1 AAA+ domain blocks the putative central DNA binding channel in the auto-inhibited conformation [7], the configuration of an active ORC has been unknown. The budding yeast ORC binds the replication origins throughout the cell division cycle [1], but they are "licensed" during the G1 phase [8]. An early step is the binding of initiation factor Cdc6 to DNA-bound ORC to form the ORC-Cdc6-DNA complex [9]. EM has shown that Cdc6 closes a gap in the crescent-shaped ORC to form a ring [9,10] and apparently activates a molecular switch in ORC, converting it from an origin DNA binder to an active Mcm2-7 loader [6,9]. However, the physical nature of the molecular switch is currently unknown due to the lack of a high-resolution structure of ORC-Cdc6 on DNA.

The next steps involve the sequential recruitment onto the origin DNA of two Cdt1-bound hexamers of Mcm2-7 by ORC-Cdc6 to form a Mcm2-7 double-hexamer (D-H) that forms part of the pre-Replicative Complex (pre-RC) [8,11]. *In vitro* reactions using purified components have demonstrated that a high salt-stable Mcm2-7 D-H is loaded on DNA in an ATP dependent manner [12,13]. Each hexamer within the D-H is assembled in a way that their respective Mcm3 and Mcm6 subunits face each other, as revealed by EM of Maltose Binding Protein (MBP) tagged D-H as well as a 3.8 Å resolution cryo-EM structure [14,15]. The Mcm2-7 double-hexamer has a central channel that is wide enough for passage of double stranded DNA, in agreement with biochemical findings [12,13]. Because the two Mcm2-7 hexamers are twisted relative to each other, it was speculated that within the interface between the two hexamers an inflection point of the DNA path is created, possibly promoting melting of the double stranded DNA when the helicase becomes activated in S phase [14]. At the G1-S transition, the inactive D-H is converted into an active replicative helicase that consists of a Mcm2-7 hexamer bound to Cdc45 and the four-subunit GINS complex, called the CMG [16,17]. Disruption of the Mcm2-7 D-H and assembly of CMG require activation by the Dbf4-Cdc7 protein kinase (DDK) and Cyclin-Dependent Kinase (CDK; Clb5-Cdc28), which phosphorylate some of the pre-RC components, such as Mcm2-7 subunits and accessory loading proteins Sld2 and Sld3 [1,8,18-26]. Subsequently, primase and DNA polymerases load, along with many other replication factors, to form the replisome that executes DNA synthesis [27,28].

The two Mcm2-7 hexamers are loaded on DNA sequentially [15,29,30]. In the presence of ATPγS, ORC-Cdc6 loads the first Mcm2-7 hexamer on DNA, forming an ORC-Cdc6-Cdt1-Mcm2-7 intermediate (OCCM) [31]. Then ATP hydrolysis is triggered and Cdt1 is released, which leads to the formation of the ORC-Cdc6-Mcm2-7 (OCM) complex [30,32-34]. A time course analysis suggests that it is the OCM that loads the second Cdt1-bound Mcm2-7 to form an ORC-Cdc6-Mcm2-7-Mcm2-7 (OCMM) complex prior to Mcm2-7 D-H formation [15,30,35]. *In vivo* evidence

suggests that ATP hydrolysis by Cdc6 causes the separation of the D-H from ORC-Cdc6 [36]. In this work, we describe a 3.9 Å resolution cryo-EM structure of the 1.1-MDa 14-protein OCCM complex on DNA and interactions between individual proteins using mass spectrometry. The structure revealed how ORC-Cdc6 recognizes origin DNA and how this complex recruits the first Cdt1-bound Mcm2-7 hexamer, thereby illuminating a crucial step in eukaryotic DNA replication initiation.

# RESULTS

## Overall structure of the OCCM-DNA complex

We prepared the OCCM complexes in the presence of ATPγS from purified proteins on a replication origin containing plasmid attached to magnetic beads. Upon DNaseI treatment the OCCM samples were released from the beads and directly processed for cryo-EM grid preparation. We derived a 3.9 Å resolution cryo-EM 3D map of the OCCM from 304,288 particles that were selected from 7500 raw electron micrographs and nearly 1,000,000 raw particles that were recorded on a K2 camera in a Titan Krios microscope operated at a high tension of 300 kV (**Fig. 1a-c, Supplementary Figs. 1-4, Supplementary Table 1, Supplementary Video 1, online methods**). The 3D map had well-defined densities for the double strand DNA and 13 subunits of the 14-protein complex (**Fig. 1d and e**). The only protein not well resolved was Orc6; its density was visible only at a lower display threshold. Nevertheless, a conserved C-terminal α-helix of Orc6 was resolved, which is important, as it is mutated in Meier Gorlin syndrome [37,38]. To better understand the architecture of the complex, and particularly of the flexible sections involving Orc6, cross-linking/mass spectrometry (CLMS) analysis of the OCCM was performed (**Supplementary Fig. 5**), which confirmed the general architecture and also identified several interactions between Orc6 and Orc2/Mcm2 (**Fig. 2a-c, Supplementary Fig. 6, Supplementary Table 2)**. An atomic model of the OCCM was built into the EM densities guided by the published structures of the Mcm2-7 hexamer [14], the *Drosophila* ORC [7], an archaeal Cdc6 homolog [39], a homolog of the N-terminal domain (NTD) of Cdt1, the middle helical domain (MHD) and the C-terminal domain (CTD) of the human Cdt1 [40-42]. The double stranded DNA was manually built into the EM density (**Supplementary Fig. 7**). Electron densities for eight nucleotides were observed at the interface between Mcm2-Mcm6, Mcm6-Mcm4, Mcm4-Mcm7, Mcm7-Mcm3, Cdc6-Orc1, Orc1-Orc4, Orc4-Orc5, and Orc5-Orc3 (**Supplementary Fig. 8)**. The nucleotide density at the Orc4-Orc5 interface was the weakest among the eight sites. The Orc5-Orc3 interface was the smallest among the five Orc1-5 subunits, such that the complex could be divided into two sub-complexes of Orc1-Orc4-Orc5 and Orc3-Orc2. However, there was contact between Orc5 and Orc3 at the nucleotide-binding region such that the ATPγS density at the interface was clear. No nucleotide densities were resolved at interfaces between Mcm3-Mcm5 and Mcm5-Mcm2 due to the reduced resolution and increased flexibility, as they were either at or near the DNA loading gate.

In the ORC-Cdc6 top tier of the OCCM, Orc1-5 and Cdc6 formed a six-membered ring structure. The five ORC subunits were arranged in the order of Orc1:Orc4:Orc5:Orc3:Orc2, and Cdc6 bridged the gap between Orc1 and Orc2, in agreement with a previous lower resolution EM study (**Fig. 3**) [5]. Interestingly, four WH domains of Mcm3, Mcm4, Mcm6, and Mcm7 spiraled upwards from the Mcm2-7 hexamer and engaged the ORC-Cdc6 ring, each interacting with

ORC-Cdc6 subunits (**Fig. 1d and e, Supplementary Figs. 9 and 10**), while an interaction between the Mcm5 WH domain and Orc2 was seen by CLMS (**Fig. 2c**). In the bottom of the Cdt1-Mcm2-7 tier, the six Mcm subunits formed a ring structure in the order of 2-6-4-7-3-5, consistent with previous studies [14,15,22,43,44]. At the interface between the Mcm2 AAA+ domain and Mcm5 AAA+ domain, the domain swapped α-helix of Mcm5 bound to the Mcm2 AAA+ domain, thereby sealing the DNA-loading gate (**Fig. 1c, Supplementary Fig. 11**). Therefore, the C-tier AAA+ ring of the Mcm2-7 hexamer was topologically closed (**Fig. 1d**). However, the N-tier ring was still open at the Mcm2-5 interface. Hence the Mcm2-7 was in a half-open and half-closed state. This structural feature explains why the Mcm2-7 ring in the OCCM is partially salt stable on DNA, but not as salt stable as the Mcm2-7 double-hexamer [15,31,44]. The structure also explains why the DNA is intact in the C-tier AAA+ ring but invisible in the N-tier ring as DNA in this region may be less constrained or digested by the nuclease, which was used to release the DNA bound OCCM during sample preparation.

## ORC-Cdc6 forms a closed ring with a pseudo 6-fold symmetry

In the OCCM structure, ORC-Cdc6 assembled into a complete ring encircling DNA with a pseudo six-fold symmetry (**Fig. 3a-b, Fig. 4**). The N-terminal extension on Orc2 and an insertion in Orc3 give ORC a helical shape [44]. The four observed ATPγS molecules of ORC-Cdc6 define one circle on top and the four observed nucleotides in Mcm2-7 define another circle below. The top circle is larger, off center, and is tilted by 17° with respect to the lower circle (**Fig. 4a-b**). The six predicted AAA+ proteins, Orc1-5 and Cdc6, all had one AAA+ domain with an AAA-RecA-fold (the RecA fold), an alpha helical-lid domain (the lid), and a C-terminal (CT), alpha helical winged helix domain (WHD), and they were superimposable (**Fig. 3c**) [45,46]. There were variations to this general rule in that Orc2 lacked the alpha helical lid similar to the DmOrc2, and Orc3 had an insertion consisting of a helical domain. Among the six initiator AAA+ subunits, Orc4 was unique as it had one α-helix insertion and one insertion loop in the WHD. The six AAA-RecA-folds and the six WHD formed a two-tiered ring structure that surrounded the DNA within the central channel. In contrast, the six AAA-lid domains that each bridged the AAA-RecA-like domain and the WH domain formed an outer brace that spiraled around the DNA interacting domains. The bottom tier ring of the WH domains was largely responsible for interacting with the Mcm CTD domains below, as suggested [7]. The peripheral brace of the AAA-lids also interacted with the Mcm subunit WH domains.

The crystal structure of the *Drosophila* apo-ORC core revealed an auto-inhibited conformation that is not compatible with DNA binding [7] (**Fig. 3d**). DmORC structure is also incompatible with Cdc6 binding. However, we found that Orc3-4-5 were in a similar configuration in both ScORC and DmORC. By aligning the two ORC structures using the common Orc3-4-5 region as a reference, we found that the RecA-fold of DmOrc1 and the WH domain of DmOrc2 needed to move and flip by ~180° in order to match their respective yeast counterparts (**Fig. 3e, f**). These changes created a gap between Orc1-Orc2 for DNA passage as well as for Cdc6 insertion between Orc1-Orc2 following DNA binding by ORC (**Fig. 3b, d, Supplementary Video 2**). Since the Mcm2-7 hexamer has been loaded onto DNA by ORC-Cdc6, the conformation of ORC-Cdc6 in the OCCM structure is clearly in its active form, allowing DNA binding [9] and both Cdc6-Orc1 and Orc1-Orc4 ATPase activities that are required for Mcm2-7 double hexamer assembly or subsequent regulated initiation of DNA replication once per cell division cycle [9,29,30,32,47,48].

## Cdt1 forms an extended three-domain structure

In our previous low-resolution cryo-EM map of OCCM, an electron density outside of Mcm2 and Mcm6 was assigned to Cdt1 [44]. The current higher resolution EM map clearly showed that Cdt1 exists in an unusually extended three-domain structure, with density outside Mcm2 and Mcm6 assigned to the NTD and middle helical domain (MHD) of Cdt1 (**Fig. 5a-c**). Surprisingly, the Cdt1 CTD was linked to the MHD by a long loop and was 60 Å away from MHD, located between Mcm6 and Mcm4. The density of Cdt1 NTD was relatively weak, indicating a degree of flexibility. The Cdt1 NTD bound only to the Mcm2 CTD with an interface of ~ 600 □$^2$. In contrast, the Cdt1 MHD bound to both NTD and CTD of Mcm2 as well as NTD of Mcm6 with a larger interface of ~ 1000 □$^2$, which was also seen by CLMS (**Fig. 2a**). The Cdt1 CTD interacted extensively with all of the major domains of Mcm6, the Mcm4 NTD, and the Orc4 WH domain insertion loop. Consistent with this observation, a previous NMR study showed interaction between the Mcm6 WH domain and a short peptide in the CTD of Cdt1 [40]. Importantly, the Cdt1 CTD formed an arch toward the Mcm6 WH domain, which in turn interacted with the Orc4 WH domain and the Orc5 AAA-lid. The Mcm6 WH domain in the Mcm2-7 hexamer before encountering ORC-Cdc6 is likely located in the middle of the ring between Mcm2 and Mcm6, because this is where the domain is found in both active helicase CMG and in the inactive Mcm2-7 double-hexamer (**Fig. 5d**). Hence, the CT arch of Cdt1 is likely responsible for displacing the Mcm6 WHD domain by 40 Å to the periphery where the Mcm6 WHD is found in the OCCM structure. This conformational change likely explains the inhibitory role of the Mcm6 WHD, which blocks OCCM formation in the absence of Cdt1 [32]. Therefore, Cdt1 appeared to play a dual role in Mcm2-7 hexamer loading: it created the ORC-Cdc6-binding surface on the CT surface of Mcm2-7 hexamer by moving outward the obstructing Mcm6 WHD, and at the same time formed an extended 3-domain side-brace that stabilized the Mcm2-Mcm6-Mcm4 half ring, potentially allowing the other half ring Mcm5-Mcm3-Mcm7 to move (**Fig. 1d, e**). We suggest that these interactions underlie the essential roles of Cdt1 in Mcm2-7 loading on DNA [13]. Cdt1 bound in the OCCM, particularly its interaction with Mcm2 and the Mcm2-7 subunits that bind ATP (Mcm2,4,7 and 3) may keep the Mcm2-Mcm5 N-tier interface open and prevent Mcm2-7 ATP hydrolysis. ORC-Cdc6 ATPase activity, the next step after OCCM assembly, removes Cdt1 [30] and may promote Mcm2-7 ATPase activity to close the first Mcm2-7 hexamer [33,34].

To further investigate the interactions between Cdt1 and Mcm2-7, we expressed in baculoviruses each Mcm subunit and Cdt1 as a Strep-Strep-SUMO-Cdt1 (SSS-Cdt1). Each Mcm subunit alone or all six in combination were expressed and a pull down with purified SSS-Cdt1 was performed (**Supplementary Fig. 12**). Cdt1 interacted with all six Mcm proteins when they are expressed together. Individually, Mcm2, Mcm6 (most strongly) but also Mcm7 interact with Cdt1. The former two interactions were found in the OCCM model. We did not see an interaction with Mcm4, suggesting that the Cdt1-CTD interaction with Mcm4 seen in OCCM structure must depend on the prior binding to Mcm2-Mcm6 in the Mcm2-7 hexamer. The interaction with Mcm7 is not present at the stage of OCCM, but could be functional downstream, when OCM recruits the second Cdt1-bound Mcm2-7 hexamer to form the D-H [35]. A long Mcm7 alpha helix projects down toward the incoming second Mcm2-7-Cdt1 complex (**Fig. 1e**)

The Mcm2-7 hexamer in the OCCM needs to change to adopt the structure present in the Mcm2-7 D-H once the loading process is complete. By comparing the two structures, we found that the Mcm2-7 hexamer needed to undergo large conformational changes during the OCCM – D-H transition, in particular within the entire Mcm2-7 NTD ring as well as the CTDs of Mcm2 and Mcm5 (**Fig. 6, Supplementary Video 3**). Specifically, the Mcm2-7 NTD ring needed to rotate by ~25° relative to the Mcm2-7 CTD to match the MCM ring in the D-H, and the CTDs of Mcm2 and Mcm5 had to rotate by ~5° and ~15°, respectively, to form the closed interface in the D-H (**Fig. 6b,c**). Because the ATPase activity of Orc1 and Cdc6 is required during the loading reaction [29,30,32], and Orc1 and Cdc6 appear in a conformation poised to hydrolyze ATP, it is possible that the conformational changes outlined here are driven by ORC-Cdc6 ATP hydrolysis. Conceivably, the large conformational changes, which could be mediated by interactions between Orc4, Mcm6 and Cdt1 (**Fig. 5a**), would alter the Cdt1 binding surface, leading to its release from Mcm2-7. Since Cdt1 release is known to occur immediately before the recruitment of the second Cdt1-bound Mcm2-7 [30,32,35], we suggest that ATP hydrolysis by ORC-Cdc6 could facilitate Cdt1 release, completely closing the first Mcm2-7 hexamer and establishing a condition for recruitment of the next Cdt1-bound Mcm2-7 hexamer. Alternatively, MCM ATP-hydrolysis could be involved [33,34]; however the Mcm2-7 ring is broken in the OCCM, thus ATPase activity of Mcm is likely blocked at this stage [45].

## Asymmetric interaction between ORC-Cdc6 and Cdt1-Mcm2-7 enables DNA insertion

Although both the Mcm2-7 hexamer and ORC-Cdc6 form ring-like structures with a pseudo 6-fold symmetry, the interaction between the two rings is asymmetric due to the ~17° tilt of the ORC-Cdc6 ring with respect to the Mcm2-7 ring (**Fig. 1d, Fig. 4)**. As a consequence, the DNA is bent by ~20-25° at the interface between ORC-Cdc6 and Mcm2-7. Furthermore, the tilt led to a tight interface between Orc1-Orc4-Orc5 in the first half of the ORC-Cdc6 ring with the Mcm4-Mcm6-Mcm2 half ring, and an apparent "loose" interface between Orc3-Orc2-Cdc6 and Mcm5-3-7. At the tight interface, the WH domains of Orc1 and Orc4 insert into the gaps between the CT WH domains and the AAA-lid domains of Mcm4 and Mcm6; the Orc5 WH domain interacted only with Mcm2 AAA-RecA-fold because Mcm2 lacked a WH domain (**Fig. 1d-e, Supplementary Figs. 9-10**). At the "loose interface", the WH domain of Mcm5 was not visible in the EM map, but the CLMS data identified it across-the-interface partner of Orc2. Interestingly, the WH domains of Mcm3 and Mcm7 reached upwards more than 30 Å via their long loops to interact with Orc2 and Cdc6, respectively (**Fig. 1e, Supplementary Figs. 9, 10**). The loop connecting the AAA-RecA-fold and the WH domain in Mcm7 was ordered and well resolved with slightly weaker density. However, the corresponding loop in Mcm3 was disordered. Accordingly, we propose that the asymmetric interaction leaves half of the Mcm2-7 ring (Mcm5-3-7) only loosely tethered, such that the Mcm5-3-7 half ring can move away from the tightly-tethered Mcm4-6-2 half ring to open up the Mcm2/Mcm5 gate for DNA insertion and then move back to close the gate. This conformation is most likely stabilized by Cdt1.

## Protein-DNA interactions in the OCCM

We modeled 39 base pairs (bp) of DNA in the OCCM density map; 24 bp were encircled by the ORC-Cdc6 ring and the remaining 15 bp by the C-tier ring of Mcm2-7 (**Fig. 7a, Supplementary Fig. 13**). There was no apparent DNA density inside the N-tier ring of Mcm2-7. Because the N-

tier ring was open at the Mcm2-Mcm5 interface, the dsDNA there might have been digested by the DNase I nuclease that was used to cleave the loading intermediate off of the plasmid DNA. In the top ORC-Cdc6 region, Orc1, Orc3 and Orc5 had little direct interaction with DNA. DNA was held in place by interactions with the initiator-specific motif (ISM) in the AAA+ domains of Orc2, a unique Orc4-specific insertion α-helix, and by the ISM and the WH domains of Cdc6 (**Fig. 7a-e**). These four binding components spiraled around the DNA just like RFC clamp loaders spiral around the DNA [44,49,50], although the overall ORC-Cdc6 ring itself appears flat because the Orc2 and Orc3 subunits break the helical path. Archaeal AAA+ replication initiators also interact with DNA via their respective ISM [39,51]. The archaeal WH domain is known to bind DNA with both the helix-turn-helix (HTH) motif and the β-hairpin wing loop [51]. However, in the yeast ORC-Cdc6, we found that only the β-hairpin wing loops of the WH domains of Cdc6 and Orc4 bound to DNA, their respectively HTH motif did not bind to DNA but rather were engaged in subunit-subunit interactions. This different DNA binding mode was caused by a ~90° rotation of the Orc4 and Cdc6 WH domains away from the central DNA channel (**Supplementary Fig. 14**). We have previously demonstrated that ORC-Cdc6 causes a nuclease protected footprint on the origin DNA that extends to 70-78 base pairs, greater than the 44-50 base-pair footprint of ORC alone [9]. Even allowing for limited nuclease access to the DNA near edges of the ORC-Cdc6 complex, which accounts to 10 bp [52], the amount of DNA found interacting with ORC-Cdc6 by nuclease footprinting is more than twice that found interacting with ORC-Cdc6 in the OCCM structure. There are positively charged patches in the front side (Orc2,3,6) and the bottom surface of ORC-Cdc6 that is proximal to Mcm2-7 (**Supplementary Fig. 15**). It is possible that DNA bends and binds to some of these surfaces in addition to the central channel.

In the lower Cdt1-Mcm2-7 region, Mcm3 and Mcm5 did not interact with DNA. Although the Mcm2-7 ring was nearly flat, the hairpin loops in the AAA+ domains of Mcm2-Mcm4-Mcm6-Mcm7 were arranged in a spiral to grip onto the DNA (**Fig. 7a, f**). We found that Mcm2, Mcm4 and Mcm6 interacted with DNA via their respective helix-2-insert (H2I) β-hairpin loop. Mcm4, Mcm6 and Mcm7 contacted the DNA with their respective Presensor 1 (PS1) β-hairpin loops, in particular the well-conserved KA motifs (**Fig. 7f, Supplementary Figs. 13 and 16**). In contrast, the same H2I and PS1 within the Mcm2-7 D-H are involved in inter-subunit interactions [14]. In the apo form of the active CMG helicase the PS1 loops face the central DNA channel [53], and in a recent cryo-EM structure of CMG at a sub-nanometer resolution, the PS1 loops were found to interact with a 6-base ssDNA [54]. Interestingly, most of these hairpin loops, except for the Mcm4 H2I loop, interacted with the same strand of the duplex DNA (**Fig. 7f**). It is unclear if this strand functions as the leading or the lagging strand in the CMG helicase, because extensive conformation changes must occur in Mcm2-7 to form the active helicase, such that Mcm-DNA interaction in the CMG may be very different.

## DISCUSSION

Since the discovery of ORC more than two decades ago [2], it has been a key issue in the DNA replication field to identify how ORC recognizes the dsDNA and how ORC cooperates with Cdc6 to load the Mcm2-7 hexamers onto DNA. The current atomic model of the OCCM provides the first high-resolution structure of ORC-Cdc6 bound to origin DNA. Overall, the ORC-Cdc6 structure is flat, but within this structure the subunits that bind ATP, Orc1, Orc4, Orc5 and Cdc6, form a right-handed spiral around the double stranded DNA, similar to the spiral

of AAA+ subunits around primer-template DNA in RFC clamp loaders [49,50]. The human ORC has an almost identical structure (A. Tocilj, K. On, C. Yuan, J. Sun, E. Elkayam, H. Li, B. Stillman and L. Joshua-Tor, submitted). Orc4 had an unusual insertion α-helix that appears to contact the major groove of the dsDNA. This helix is unique to budding yeast Orc4 subunits, being absent in the DmORC and HsORC4 structures [7] (A. Tocilj, K. On, C. Yuan, J. Sun, E. Elkayam, H. Li, B. Stillman and L. Joshua-Tor, submitted), may help to explain the sequence specificity of ORC that is characteristic of origin binding in *Saccharomyces sp*. Cdc6 is particularly important for DNA binding with two principal DNA binding sites: the ISM and the WH domain, explaining why Cdc6 enhances ORC's specificity for DNA [9,10].

The most prominent feature of Mcm2-7 loading by ORC-Cdc6 is the extensive use of WH domains of these replication proteins. There were six resolved WH domains in ORC-Cdc6 that formed a larger WH ring in the upper tier of the OCCM, and 5 cryo-EM and CLMS resolved WH domains in Mcm2-7 that formed a second slightly smaller WH ring on the lower Mcm2-7 tier of the OCCM. This appears to be an evolutionary conserved interaction, as it was reported recently for archaeal MCM and Orc1 [55]. However, here we observed that the WHD mediated correct stacking of the two rings mediates much of the recruitment mechanism. The position of the WHD of the Mcm proteins varies widely, and can sit either right above the AAA+ domain, or move away from the main body of the protein, or even move to the side of the AAA+ domain in the case of Mcm5, as seen in the active helicase (**Supplementary Fig. 9b**). The CTD of Cdt1 plays a special role in displacing the Mcm6 WH domain to create the ORC-Cdc6 binding surface; we showed previously that the WH domain of Mcm6 is inhibitory and blocks OCCM formation in the absence of Cdt1 [32]. The attachment of the winged helix domains by flexible linker provide these domains manifold possibility to interact with ORC-Cdc6, but after helicase activation, they could also be important interaction partners with other proteins at the DNA replication fork.

The Mcm2-7 double-hexamer structure, although obtained in the absence of DNA, showed that six H2I hairpin loops are arranged in an approximately helical trajectory that was suggested to facilitate DNA translocation and unwinding. However, in the OCCM, only three H2I hairpin loops of Mcm2-Mcm4-Mcm6 contact DNA. In addition, the three PS1 hairpin loops of Mcm4, Mcm6 and Mcm7 make contact with DNA. In the D-H, the PS1 loop is involved in inter-subunit interaction, not in DNA binding, but in the CMG helicase, the PS1 loop contacts DNA [14,54]. Another interesting feature of OCCM is the partially open Mcm2-Mcm5 interface. Thus the data provide the first structural proof that the Mcm2-Mcm5 interface is the DNA entry gate in Mcm2-7 during pre-RC formation [22,31,43]. ATPγS prevents Cdt1 removal and the gate from closing [30,32] and thus it is interesting that the N-tier ring is not yet closed given the dsDNA has been loaded into the Mcm2-7 hexamer channel. We suggest that the Mcm2-7 N-tier ring closure requires ATP hydrolysis by Orc1/Cdc6 proteins and removal of Cdt1 from the Mcm2-7 hexamer. The unusually extended three-domain structure of Cdt1 most likely keeps Mcm2-7 subunits vertically aligned, preventing the left-handed spiral of the Mcm2-7 subunits in the completely assembled Mcm2-7 double hexamer (**Fig. 6c**). Thus, like the RFC-PCNA clamp loader and clamp structures, the OCCM structure provides further insight into how a AAA+ complex loads a ring-shaped, multi-subunit complex of proteins around double stranded DNA to promote DNA replication. We suggest that ORC-Cdc6 ATPase promotes complete Mcm2-7 ring closure just as

ATP hydrolysis by the clamp loader RFC locks the PCNA DNA polymerase clamp onto double stranded DNA [50].

## METHODS

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

**Accession Codes** The 3D cryo-EM map of OCCM at 3.9 Å resolution has been deposited at the EMDB database with accession code EMD-xxxx. The corresponding atomic model was deposited at RCSB PDB bank with accession code yyyy.

**Acknowledgements** Cryo-EM data was collected on a FEI Titan Krios at HHMI Janelia Farm. We also collected a cryo-EM dataset on an FEI Technai F20 equipped with a K2 detector at NRAMM in the Scripps Research Institute, which was supported by NIH grants P41 GM103310. We thank Zhiheng Yu, Chuan Hong, Rick Huang at HHMI, and Clint Porter and Bridget Carragher at Scripps for help with data collection. This work was funded by the US National Institutes of Health (GM111472 and OD12272 to H.L. and GM45436 to B.S.), the Biotechnology and Biological Sciences Research Council UK (P56061 to CS) and the Wellcome Trust (Investigator Award P56628 to CS, Senior Research Fellowship 103139 to JR, a Centre core grant 092076 to JR and an instrument grant 108504 to JR).

**Author Contributions** Y.Z., A.R., L.B., J.S., J.R., Z.A.C., B.S., C.S. and H.L. designed experiments. Z.Y., A.R., L.B., C.SPA., M.B. and J.S. performed experiments. Z.Y., A.R., L.B., J.S., Z.A.C., J.R., B.S., C.S. and H.L. analyzed the data. L.B., B.S., C.S. and H.L. wrote the manuscript.

## REFERENCES

1. Bell, S.P. & Dutta, A. DNA replication in eukaryotic cells. *Annu Rev Biochem* **71**, 333-74 (2002).
2. Bell, S.P. & Stillman, B. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* **357**, 128-34 (1992).
3. Stillman, B. Origin recognition and the chromosome cycle. *FEBS Lett* **579**, 877-84 (2005).
4. Li, H. & Stillman, B. The origin recognition complex: a biochemical and structural view. *Subcell Biochem* **62**, 37-58 (2012).
5. Chen, Z. et al. The architecture of the DNA replication origin recognition complex in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **105**, 10326-31 (2008).
6. Sun, J. et al. Cdc6-induced conformational changes in ORC bound to origin DNA revealed by cryo-electron microscopy. *Structure* **20**, 534-44 (2012).
7. Bleichert, F., Botchan, M.R. & Berger, J.M. Crystal structure of the eukaryotic origin recognition complex. *Nature* **519**, 321-6 (2015).
8. Bell, S.P. & Labib, K. Chromosome Duplication in Saccharomyces cerevisiae. *Genetics* **203**, 1027-67 (2016).
9. Speck, C., Chen, Z., Li, H. & Stillman, B. ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA. *Nat Struct Mol Biol* **12**, 965-71 (2005).

10. Speck, C. & Stillman, B. Cdc6 ATPase activity regulates ORC x Cdc6 stability and the selection of specific DNA sequences as origins of DNA replication. *J Biol Chem* **282**, 11705-14 (2007).
11. Cocker, J.H., Piatti, S., Santocanale, C., Nasmyth, K. & Diffley, J.F. An essential role for the Cdc6 protein in forming the pre-replicative complexes of budding yeast. *Nature* **379**, 180-2 (1996).
12. Evrin, C. et al. A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. *Proc Natl Acad Sci U S A* **106**, 20240-5 (2009).
13. Remus, D. et al. Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell* **139**, 719-30 (2009).
14. Li, N. et al. Structure of the eukaryotic MCM complex at 3.8 A. *Nature* **524**, 186-91 (2015).
15. Sun, J. et al. Structural and mechanistic insights into Mcm2-7 double-hexamer assembly and function. *Genes Dev* **28**, 2291-303 (2014).
16. 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).
17. Gambus, A. et al. GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat Cell Biol* **8**, 358-66 (2006).
18. Remus, D. & Diffley, J.F. Eukaryotic DNA replication control: lock and load, then fire. *Curr Opin Cell Biol* **21**, 771-7 (2009).
19. Yardimci, H., Loveland, A.B., Habuchi, S., van Oijen, A.M. & Walter, J.C. Uncoupling of sister replisomes during eukaryotic DNA replication. *Mol Cell* **40**, 834-40 (2010).
20. Botchan, M. & Berger, J. DNA replication: making two forks from one prereplication complex. *Molecular Cell* **40**, 860-1 (2010).
21. Aparicio, T., Guillou, E., Coloma, J., Montoya, G. & Mendez, J. The human GINS complex associates with Cdc45 and MCM and is essential for DNA replication. *Nucleic Acids Res* **37**, 2087-95 (2009).
22. Costa, A. et al. The structural basis for MCM2-7 helicase activation by GINS and Cdc45. *Nat Struct Mol Biol* **18**, 471-7 (2011).
23. Zegerman, P. & Diffley, J.F. Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature* **445**, 281-5 (2007).
24. Araki, H. Cyclin-dependent kinase-dependent initiation of chromosomal DNA replication. *Curr Opin Cell Biol* **22**, 766-71 (2010).
25. Sheu, Y.J. & Stillman, B. The Dbf4-Cdc7 kinase promotes S phase by alleviating an inhibitory activity in Mcm4. *Nature* **463**, 113-7 (2010).
26. Deegan, T.D., Yeeles, J.T. & Diffley, J.F. Phosphopeptide binding by Sld3 links Dbf4-dependent kinase to MCM replicative helicase activation. *EMBO J* **35**, 961-73 (2016).
27. 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).
28. Yeeles, J.T., Deegan, T.D., Janska, A., Early, A. & Diffley, J.F. Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature* **519**, 431-5 (2015).
29. Evrin, C. et al. In the absence of ATPase activity, pre-RC formation is blocked prior to MCM2-7 hexamer dimerization. *Nucleic Acids Res* **41**, 3162-72 (2013).

30. Ticau, S., Friedman, L.J., Ivica, N.A., Gelles, J. & Bell, S.P. Single-molecule studies of origin licensing reveal mechanisms ensuring bidirectional helicase loading. *Cell* **161**, 513-25 (2015).
31. Samel, S.A. et al. A unique DNA entry gate serves for regulated loading of the eukaryotic replicative helicase MCM2-7 onto DNA. *Genes Dev* **28**, 1653-66 (2014).
32. Fernandez-Cid, A. et al. An ORC/Cdc6/MCM2-7 complex is formed in a multistep reaction to serve as a platform for MCM double-hexamer assembly. *Mol Cell* **50**, 577-88 (2013).
33. Kang, S., Warner, M.D. & Bell, S.P. Multiple functions for Mcm2-7 ATPase motifs during replication initiation. *Mol Cell* **55**, 655-65 (2014).
34. Coster, G., Frigola, J., Beuron, F., Morris, E.P. & Diffley, J.F. Origin licensing requires ATP binding and hydrolysis by the MCM replicative helicase. *Mol Cell* **55**, 666-77 (2014).
35. Evrin, C. et al. The ORC/Cdc6/MCM2-7 complex facilitates MCM2-7 dimerization during prereplicative complex formation. *Nucleic Acids Res* **42**, 2257-69 (2014).
36. Chang, F. et al. Cdc6 ATPase activity disengages Cdc6 from the pre-replicative complex to promote DNA replication. *Elife* **4**(2015).
37. Bleichert, F. et al. A Meier-Gorlin syndrome mutation in a conserved C-terminal helix of Orc6 impedes origin recognition complex formation. *Elife* **2**, e00882 (2013).
38. Balasov, M., Akhmetova, K. & Chesnokov, I. Drosophila model of Meier-Gorlin syndrome based on the mutation in a conserved C-Terminal domain of Orc6. *Am J Med Genet A* **167A**, 2533-40 (2015).
39. Gaudier, M., Schuwirth, B.S., Westcott, S.L. & Wigley, D.B. Structural basis of DNA replication origin recognition by an ORC protein. *Science* **317**, 1213-1216 (2007).
40. Liu, C. et al. Structural insights into the Cdt1-mediated MCM2–7 chromatin loading. *Nucleic acids research* **40**, 3208-3217 (2012).
41. Wei, Z. et al. Characterization and structure determination of the Cdt1 binding domain of human minichromosome maintenance (Mcm) 6. *J Biol Chem* **285**, 12469-73 (2010).
42. Lee, C. et al. Structural basis for inhibition of the replication licensing factor Cdt1 by geminin. *Nature* **430**, 913-917 (2004).
43. Bochman, M.L. & Schwacha, A. The Mcm2-7 complex has in vitro helicase activity. *Mol Cell* **31**, 287-93 (2008).
44. 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).
45. Enemark, E.J. & Joshua-Tor, L. On helicases and other motor proteins. *Curr Opin Struct Biol* **18**, 243-57 (2008).
46. Neuwald, A.F., Aravind, L., Spouge, J.L. & Koonin, E.V. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res* **9**, 27-43 (1999).
47. Bowers, J.L., Randell, J.C., Chen, S. & Bell, S.P. ATP hydrolysis by ORC catalyzes reiterative Mcm2-7 assembly at a defined origin of replication. *Mol Cell* **16**, 967-78 (2004).
48. Frigola, J., Remus, D., Mehanna, A. & Diffley, J.F. ATPase-dependent quality control of DNA replication origin licensing. *Nature* **495**, 339-43 (2013).
49. O'Donnell, M. & Kuriyan, J. Clamp loaders and replication initiation. *Curr Opin Struct Biol* **16**, 35-41 (2006).

50. Kelch, B.A., Makino, D.L., O'Donnell, M. & Kuriyan, J. How a DNA polymerase clamp loader opens a sliding clamp. *Science* **334**, 1675-80 (2011).
51. Dueber, E.L.C., Corn, J.E., Bell, S.D. & Berger, J.M. Replication origin recognition and deformation by a heterodimeric archaeal Orc1 complex. *Science* **317**, 1210-1213 (2007).
52. Suck, D. & Oefner, C. Structure of DNase I at 2.0 A resolution suggests a mechanism for binding to and cutting DNA. *Nature* **321**, 620-5 (1986).
53. Yuan, Z. et al. Structure of the eukaryotic replicative CMG helicase suggests a pumpjack motion for translocation. *Nat Struct Mol Biol* **23**, 217-24 (2016).
54. Abid Ali, F. et al. Cryo-EM structures of the eukaryotic replicative helicase bound to a translocation substrate. *Nat Commun* **7**, 10708 (2016).
55. Samson, R.Y., Abeyrathne, P.D. & Bell, S.D. Mechanism of Archaeal MCM Helicase Recruitment to DNA Replication Origins. *Mol Cell* **61**, 287-96 (2016).

**Figure legends**

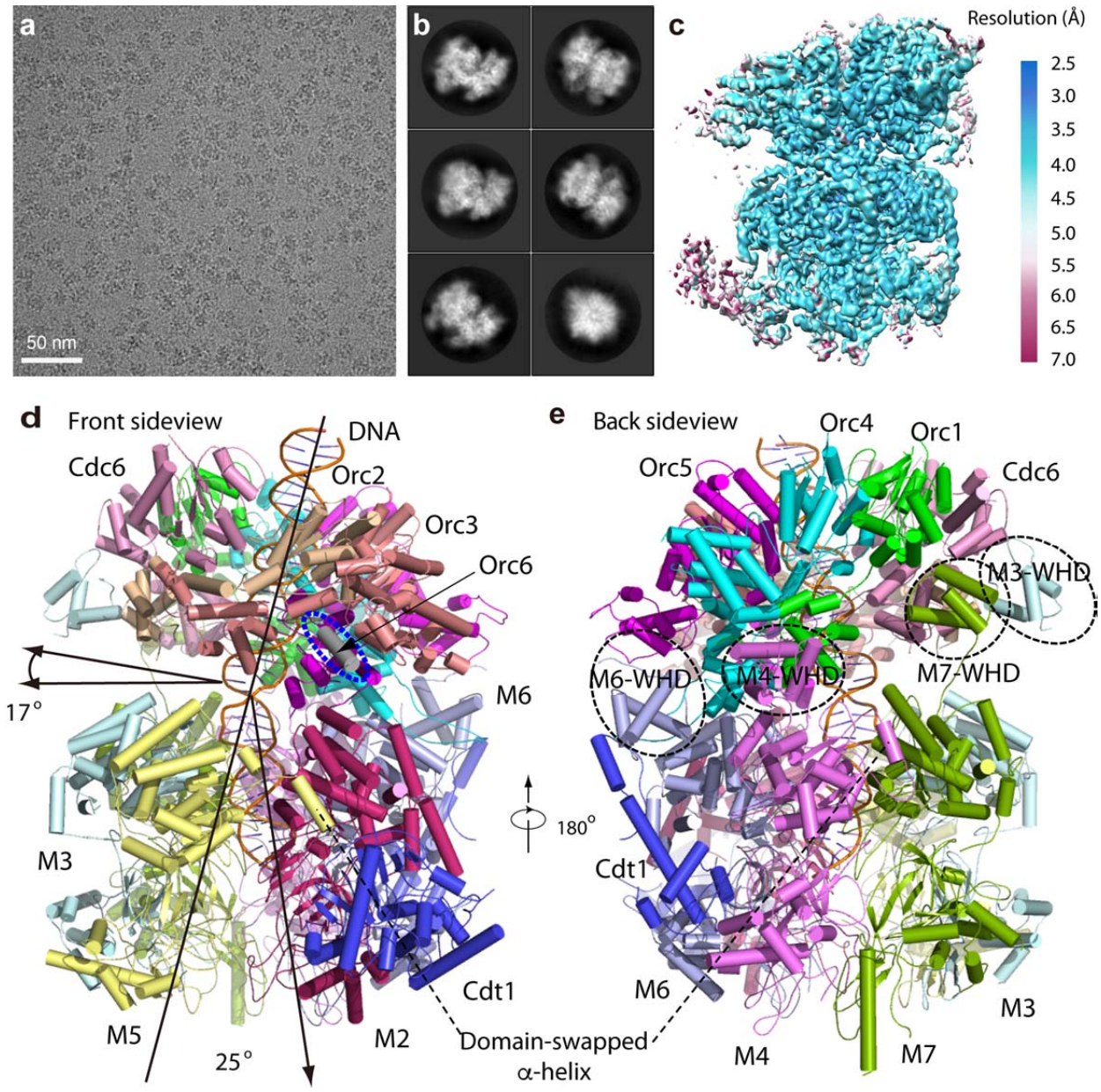


**Figure 1. Cryo-EM and overall structure of the *S. cerevisiae* OCCM complex**. (**a**) A typical motion-corrected raw image of frozen OCCM particles recorded on a direct detector. (**b**) Selected six 2D averages representing the particles in different views. (**c**) 3D cryo-EM map of OCCM color coded by local resolution. Overall resolution is 3.9 Å. (**d**) Cartoon view of the atomic model of OCCM as viewed from Front side. The two black arrows in left indicates that the ORC-Cdc6 ring lays on the MCM ring tilted by an angle of ~17°. The two black arrows in middle shows the DNA in central channel is bent by ~25°. The blue oval marks the short helix of Orc6. (**e**) Cartoon view of the OCCM model as viewed from the backside. The black circles mark the WHDs of Mcm3, Mcm4, Mcm6 and Mcm7, respectively.

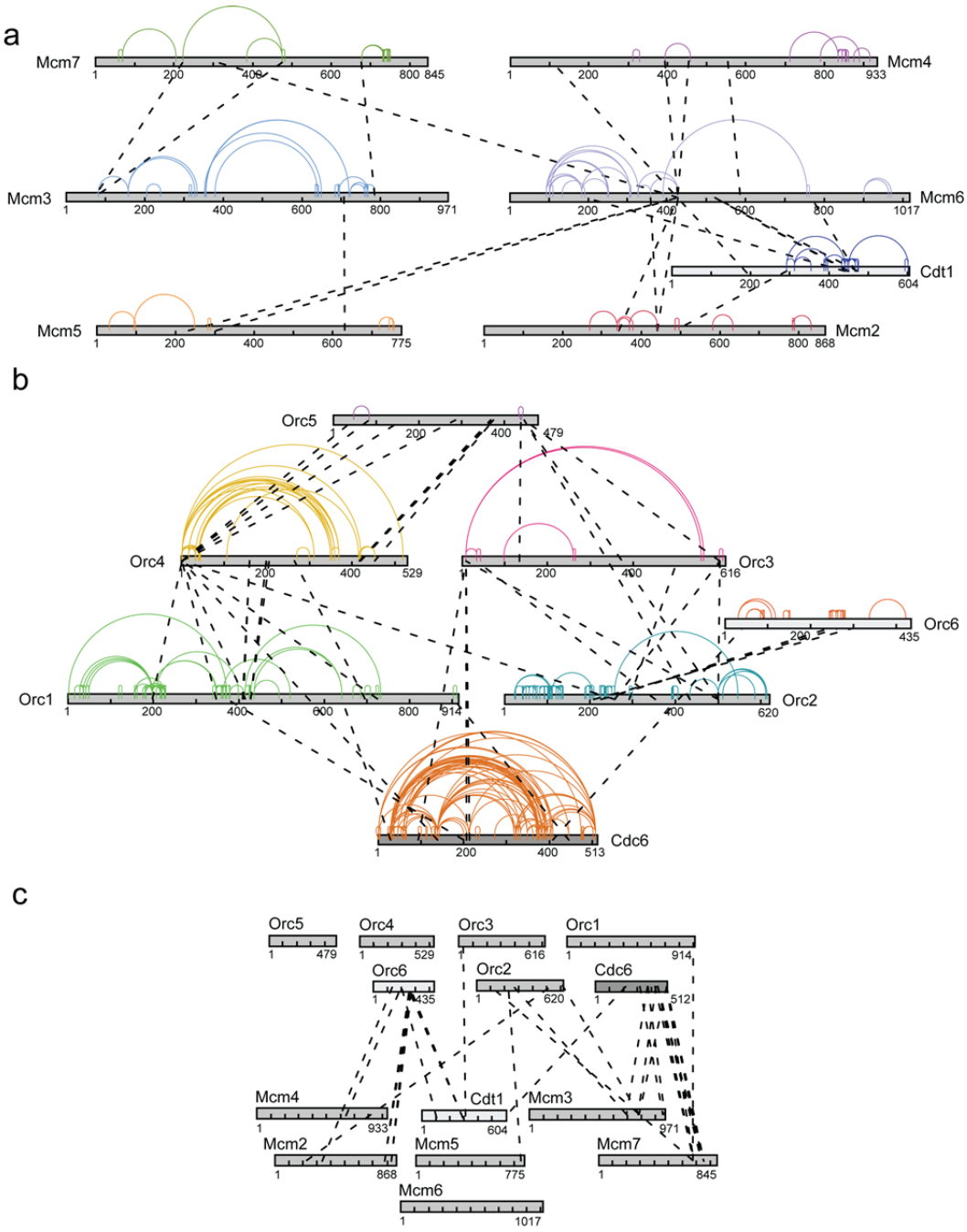


**Figure 2. Cross-linking/mass spectrometry analysis of *S. cerevisiae* OCCM complex.** (**a**) Linkage map showing the observed cross-linked residue pairs within the MCM2-7/Cdt1 complex. Intra-molecular cross-links are color coded, while inter-molecular cross-links are shown in black. (**b**) Linkage map showing the observed cross-linked residue pairs within the ORC/Cdc6 complex. Intra-molecular cross-links are color coded, while inter-molecular cross-links are shown in black. (**c**) Linkage map showing the observed cross-linked residue pairs between ORC/Cdc6 complex and Mcm2-7/Cdt1 complex. Orc6, which was only partially resolved by cryo-EM, is in close proximity to Mcm2 and Cdt1. The Winged Helix Domain of Mcm5, that was only partially resolved by cryo-EM, is in close proximity to the N-terminal region of Orc2.

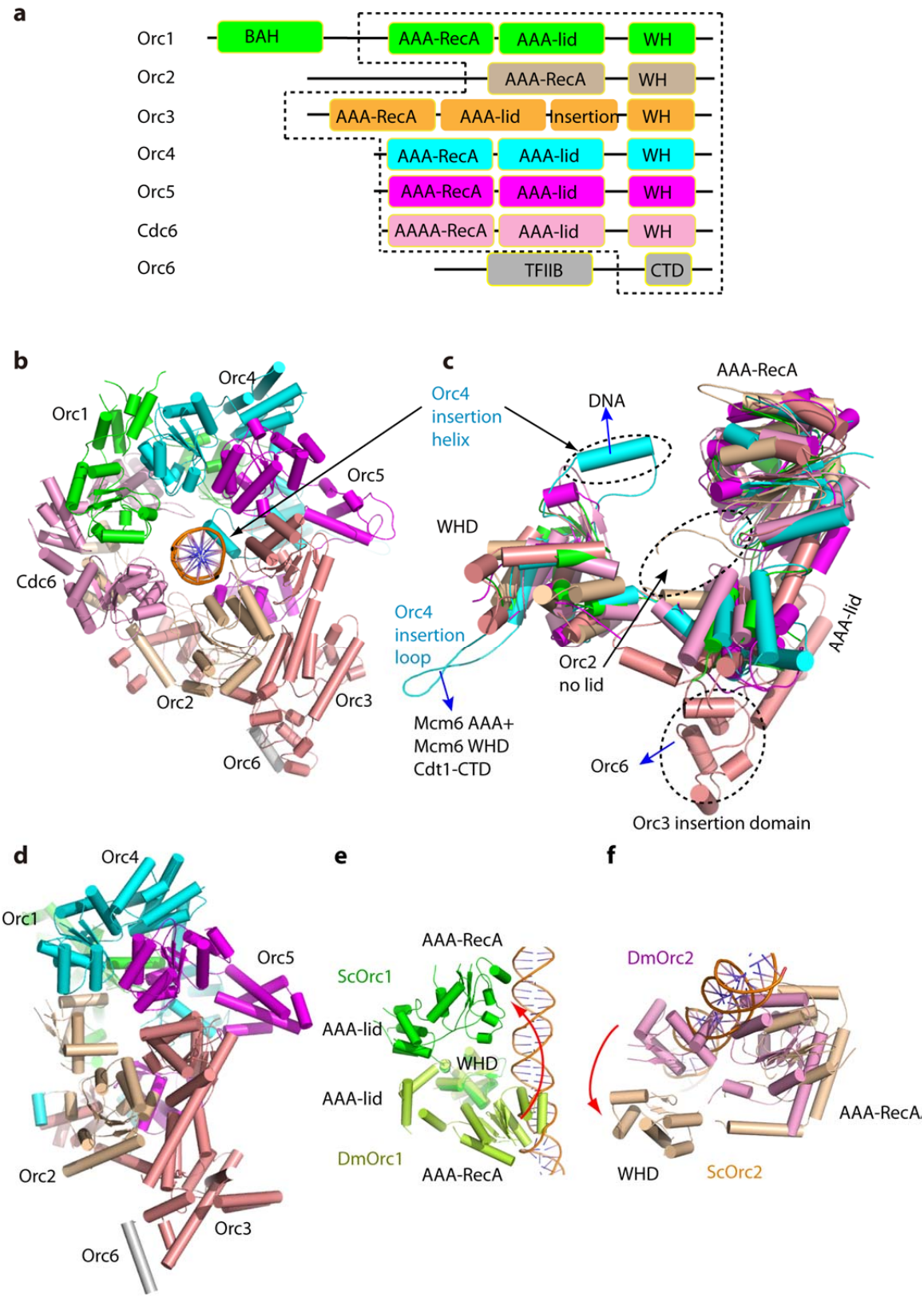


**Figure 3. ORC-Cdc6 encircles the origin DNA with the Orc4 insertion helix binding to a major groove.** (**a**) Domain organization of *S. cerevisiae* Orc1-5 subunits and Cdc6. Dashed lines

mark the ORC-Cdc6 core regions resolved in our model (TFIIB and CTD, transcription factor-II-like and C-terminal domains in Orc6; BAH, bromo-adjacent homology domain in Orc1). (**b**) The ORC-Cdc6 structure in our *S. cerevisiae* OCCM model in top view. (**c**) Superposition of Orc1-5 and Cdc6, highlighting their similar overall structures. Orc2 lacked the AAA-lid domain, resulting in a relatively open interface between Orc2-Cdc6. Orc3 had an insertion domain between the AAA-lid domain and the WHD domain that interacted with Orc6. The blue arrows point to structures with which these marked elements interact. The black arrow points to the missing lid domain in Orc2. (**d**) Crystal structure of *Drosophila* ORC complex in a similar subunit color scheme. (**e, f**) Alignment of DmORC with ScORC-Cdc6 using the most similar Orc3-5 region as a reference showed that the AAA-RecA-fold domain of DmOrc1 (**e**) and WHD of DmOrc2 (**f**) needed to move and rotate by 180° to assume their respective position in ScOCCM. See also Supplemental video 1.

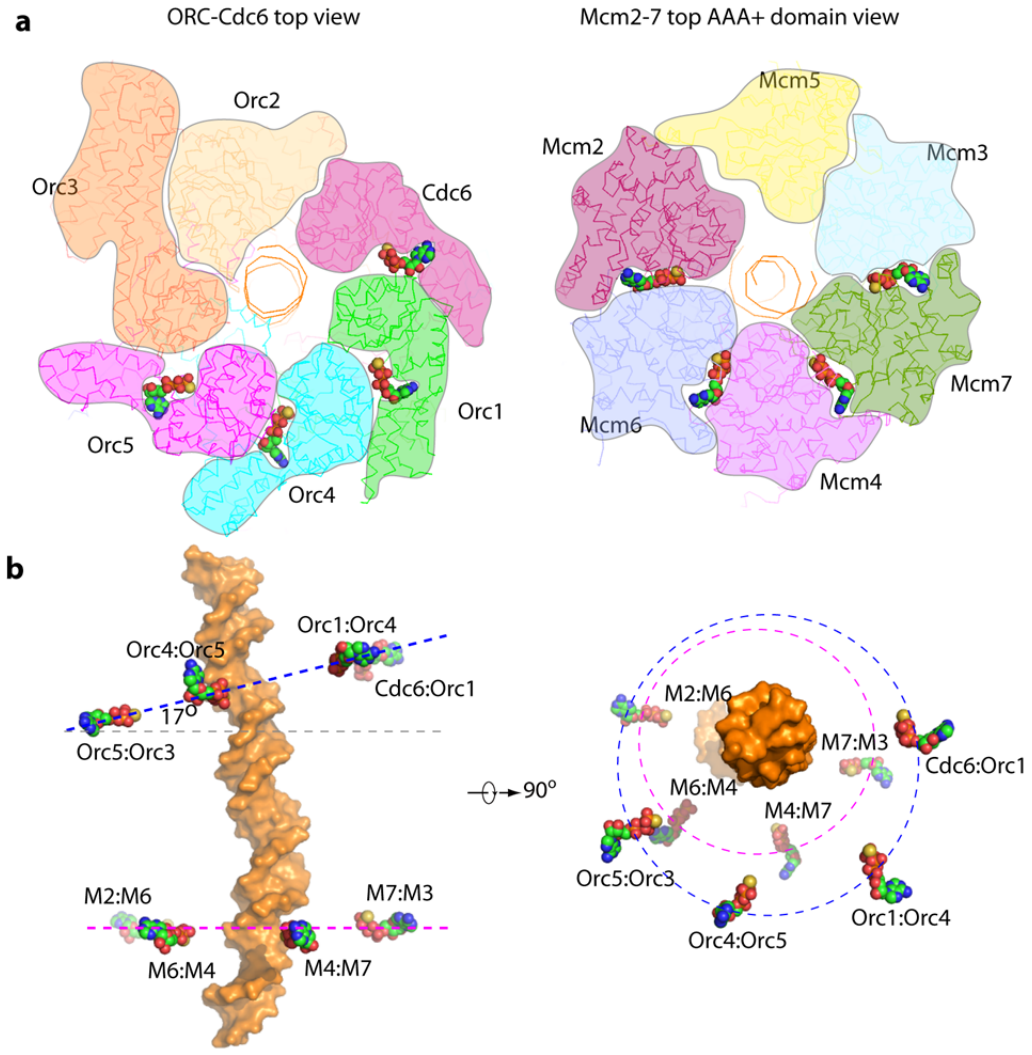


**Figure 4. Nucleotide binding sites and configuration in OCCM.** (**a**) Cut-open top view of ORC-Cdc6 and Mcm2-7 shown in surface view. The four ATPγS molecules identified in ORC-Cdc6 at the interface between Cdc6-Orc1, Orc1-Orc4, Orc4-Orc5, and Orc5-Orc3 (right), and four ATPγS molecules in Mcm2-7 at the interface between Mcm2-Mcm6, Mcm6-Mcm4, Mcm4-Mcm7, and Mcm7-Mcm3 are shown as spheres (carbon in green, oxygen in red, nitrogen in blue, sulfur in yellow). (**b**) The positons of the observed nucleotides in OCCM relative to the DNA, which is shown in orange surface. The left panel is a side view with Mcm4 in front and the right panel is a top view with ORC-Cdc6 on top, but proteins are not shown in order to highlight the nucleotides. The four ATPγS molecules in ORC-Cdc6 are co-planar, but the plane is tilted by ~17° with respect to the plane formed by the nucleotides in Mcm2-7. An imaginary circle defined by the nucleotide in ORC-Cdc6 is larger (75 Å) than the circle defined by nucleotides in Mcm2-7 (65 Å), and the two circles are acentric.

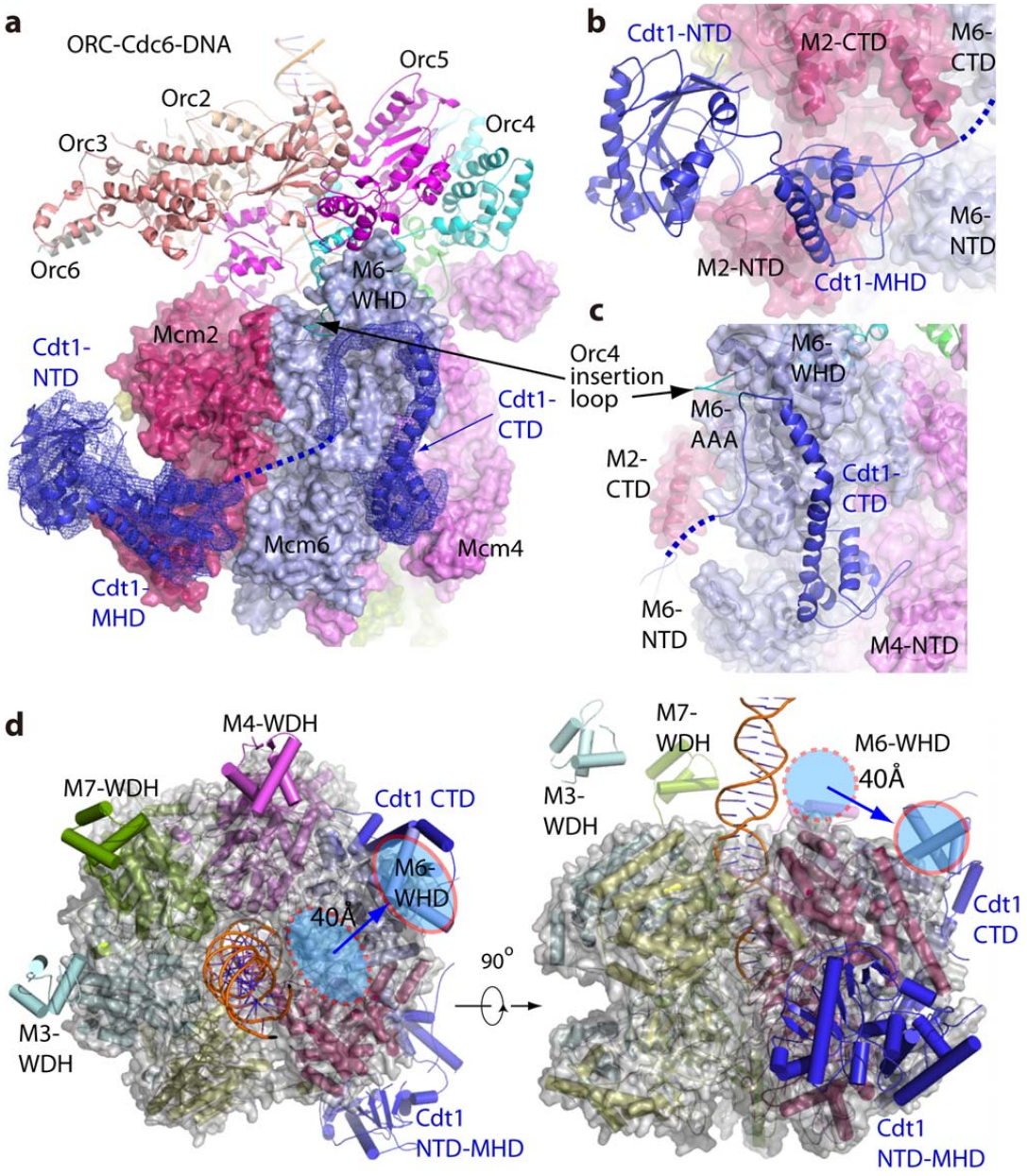


**Figure 5. Extensive interactions between Cdt1 and MCM hexamer.** (**a**) OCCM structure with Cdt1 electron density shown in blue mesh. The CTD of Cdt1 locates between Mcm6 and Mcm4, over 60 Å away from the NTD and MHD of Cdt1. (**b**) Zoomed view of the Cdt1 NTD and MHD showing their interactions with Mcm2 and Mcm6. (**c**) Zoomed view showing the Cdt1 CTD interacting with Mcm6 WHD. The dotted blue line in (**a-c**) indicates a flexible loop connecting Cdt1 MHD and CTD. (**d**) The top view (left) and front side view (right) of Mcm2-7 structure in cartoon and semi-transparent surface view. The red oval marks Mcm6-WHD in OCCM, and the dashed red oval the position of Mcm6-WHD in CMG helicase. The blue arrow shows the displacement of Mcm6 WHD in OCCM due to interaction with Cdt1 CTD. Such displacement forms an unobstructed Mcm2-7 C-terminal face for binding with ORC-Cdc6.

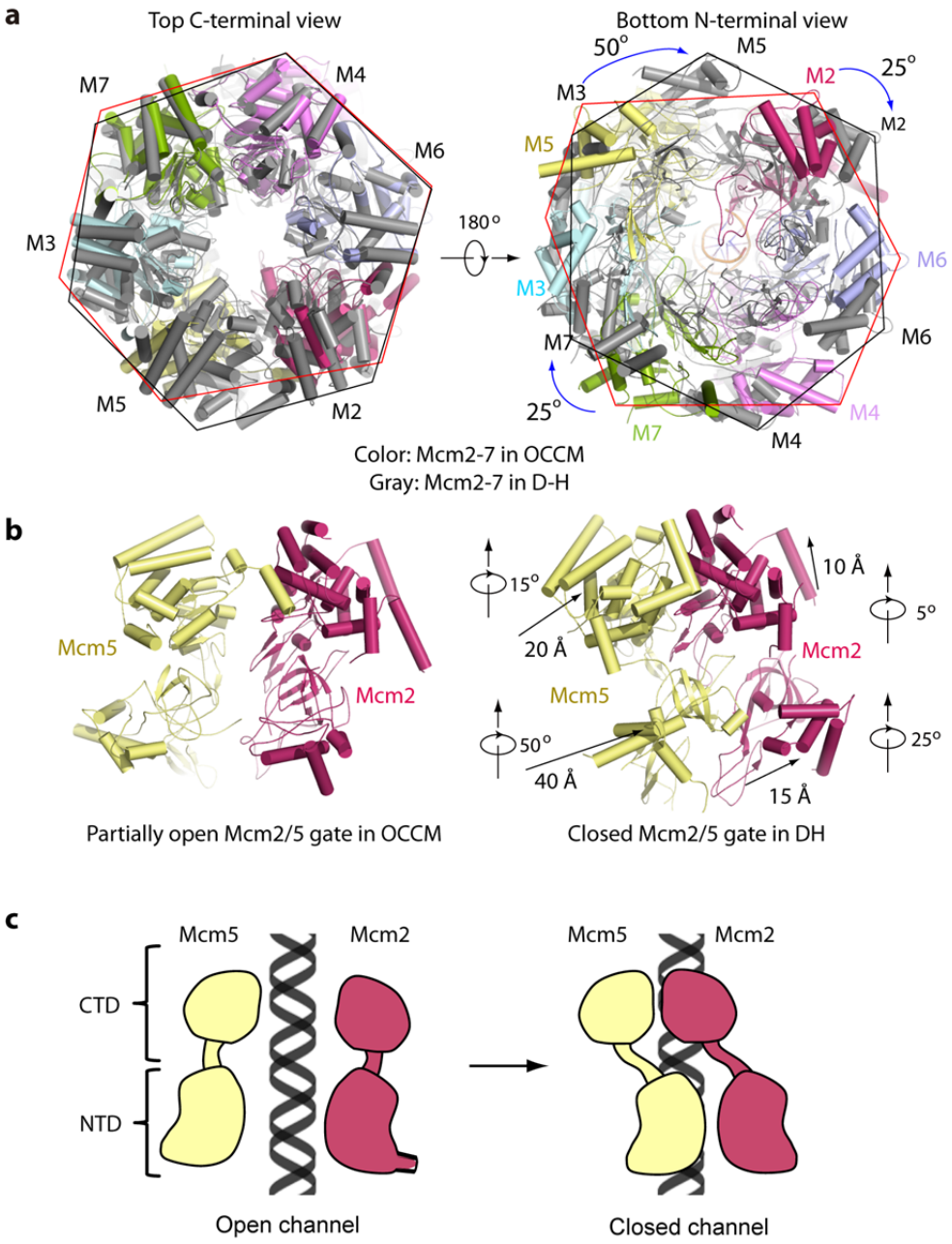


**Figure 6. Conformational changes between the Mcm2-7 in OCCM and double-hexamer (D-H).** (**a**) Comparison of the top CTD view (left) and the bottom NTD view (right) of the Mcm2-7 structure in the double-hexamer (gray cartoon) with the Mcm2-7 structure in the OCCM. The two structures were aligned using CTDs of Mcm4-6-7 as reference. Changes in the CTD ring are focused in Mcm2-5-3. The NTD ring rotated en bloc by about 25°. (**b**) Front Mcm2/5 side view of the Mcm2-7 hexamer in the OCCM structure (left) as compared to that in the D-H (right). Transitioning from OCCM to double hexamer, each CTD AAA+ domain and NTD of Mcm2 and Mcm5 undergoes a combination of rotation and translation, with the degree of rotation and translation shown as labeled. Mcm5 NTD needs to rotate by as much as 50° to close the DNA entry gate. (**c**) A sketch showing how the gate between Mcm2 and Mcm5 can be open for DNA insertion in OCCM (left) and how the gate is closed in the D-H (right).

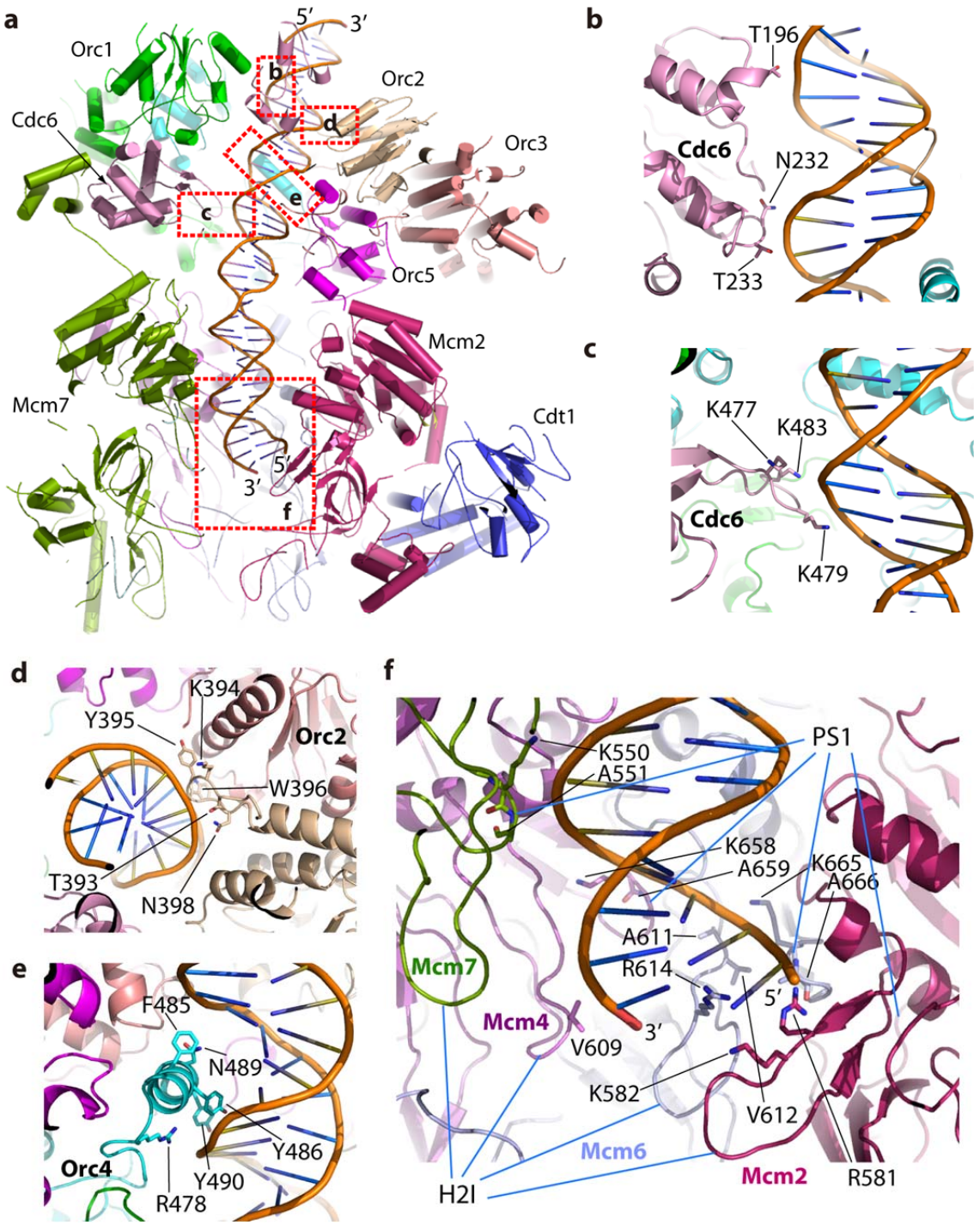


**Figure 7. Interactions between OCCM and DNA.** (**a**) An overview of OCCM-DNA structure in side view. Subunits in front of DNA including parts of Orc1 and Cdc6, and all of Mcm3 and Mcm5 are removed to show DNA. Five rectangle-marked areas are enlarged in panels (**c-f**). (**b-c**) Detailed view of the Cdc6 interaction with dsDNA. (**d**) Orc2 interaction with DNA. (**e**) Orc4 interaction with DNA. (**f**) Mcm2-6-4-7 and DNA interfaces. PS1: Pres-sensor 1 β-hairpin loop, H2I: Helix-2-insert β-hairpin loop.

## EXPERIMENTAL PROCEDURES

### Sample preparation and electron microscopy

The *Saccharomyces cerevisiae* loading intermediate OCCM was assembled *in vitro* with purified ORC, Cdc6, Cdt1 and Mcm2-7 on plasmid DNA containing the ARS1 sequence in the presence of ATPγS, and isolated by the magnetic beads pull-down approach as described previously [44] with minor modifications. 24 pre-RC reactions containing 40 nM ORC, 80 nM Cdc6, 40 nM Cdt1, 40 nM MCM2-7 and 6 nM pUC19-ARS1 beads in 50 µl buffer A (50 mM HEPES-KOH (pH 7.5), 100 mM potassium glutamate, 10 mM magnesium acetate, 50 µM zinc acetate, 3 mM ATPγS, 5 mM DTT, 0.1% Triton X-100 and 5% glycerol) were incubated for 15 min at 24 °C. After three washes with buffer B (50 mM HEPES-KOH (pH 7.5), 100 mM K acetate, 3 mM ATPγS) the complex was eluted with 1 U of DNase I in buffer B and 1 mM $CaCl_2$.

To prepare cryo-EM grids, we pooled all the elutions together and concentrated the sample to about 0.9 mg/ml in buffer B with a Microcon centrifugal filter unit (YM-100 membrane). Before EM grid preparation, we checked the sample for homogeneity by negative-stain electron microscopy. We then applied 3 µl of OCCM sample at a final concentration of 0.9 mg/ml to glow-discharged C-flat 1.2/1/3 holey carbon grids, incubated for 10 s at 6 °C and 95% humidity, blotted for 3 s then plunged into liquid ethane using an FEI Vitrobot IV. We loaded the grids into an FEI Titan Krios electron microscope operated at 300 kV high tension and collected images semi-automatically with SerialEM under low-dose mode at a magnification of ×29,000 and a pixel size of 1.01 Å per pixel. A Gatan K2 summit direct electron detector was used under super resolution mode for image recording with an under-focus range from 1.5 to 3.5 µm. The dose rate was 10 electrons per $Å^2$ per second and total exposure time was 5 seconds. The total dose was divided into a 25-frame movie and each frame was exposed for 0.2 s.

### Image processing and 3D reconstruction

Approximately 7500 raw movie micrographs were collected. The movie frames were first aligned and superimposed by the program Motioncorr [56]. Contrast transfer function parameters of each aligned micrograph were calculated using the program CTFFIND4 [57]. All the remaining steps, including particle auto selection, 2D classification, 3D classification, 3D refinement, and density map post-processing were performed using Relion-1.4 [58]. We manually picked ~10,000 particles from different views to generate 2D averages, which were used as templates for subsequent automatic particle selection. Automatic particle selection was then performed for the entire data set. 1,371,667 particles were initially selected. Then we carefully checked the particles obtained from automatically picking, removed the bad particles and re-picked the missing good ones. Particles were then sorted by similarity to the 2D references; the 10% of particles with the lowest z-scores were deleted from the particle pool. 2D classification of all remaining particles was performed and particles in unrecognizable classes by visual inspection were removed. A total of 601,095 particles were used for 3D classification. We derived six 3D models from the dataset, and found two models were similar to each other and their associated particles were combined for further refinement; the other four models were distorted and those particles were discarded, leading to a dataset size of 304,288 particles. This final dataset was used for further 3D refinement, resulting in the 3.91 Å 3D density map. The resolution of the

map was estimated by the so-called gold-standard Fourier shell correlation, at the correlation cutoff value of 0.143. The 3D density maps were corrected for the detector modulation transfer function and sharpened by applying a negative B-factor of -112 $Å^2$. The particles had some preference for end-on views but because of the large number of particles used virtually all of the angular space was sampled.

**Structural modeling, refinement, and validation**

The yeast Mcm2-7 models were directly extracted from the cryo-EM structure of the yeast Mcm2-7 double hexamer (PDB code 3JA8) [14]. For subsequent docking, each Mcm proteins was split into 2 parts: NTD and CTD. The initial models of the *S. cerevisiae* Orc1-5 subunits were generated from the crystal structure of the *Drosophila* ORC complex (PDB ID: 4XGC) [7] and Cdc6 subunit from the homologue Archaeal structure (PDB ID: 2V1U) [39] using the SWISS-MODEL server [59]. These models were first docked and fitted into the density map using COOT [60] and Chimera [61]. We found Orc1 and Orc2 in the yeast OCCM had gone through dramatic conformational changes compared to the fly ORC structure. Thus, we split each of the two proteins into 3 parts: the AAA-RecA-fold and the AAA-lid domain (where present) and the C-terminal WH domain, and fitted these domains independently into the 3D density map. Based on the structural features, the extra density outside Mcm2 and Mcm6 clearly belonged to NTD and middle helical domain (MHD) of Cdt1 [44]. We used the SWISS-MODEL server to generate the atomic models of Cdt1 NTD and MHD from their respective homologous structures (PDB ID: 5C3P and 2ZXX) [42,62]. The Cdt1 CTD model was generated from human Cdt1 crystal structure (PDB ID: 2WVR) [63], and this domain was found in the region between Mcm6 and Mcm4. Four bulky densities outside the core structure of Cdc6-ORC were identified to be the WH domains of Mcm3, Mcm4, Mcm6, and Mcm7, respectively. Their atomic models were generated from their corresponding homologue structures (PDB ID: 3NW0, 2M45, 2KLQ, and 2OD5) [41,64,65] using the SWISS-MODEL server. The Orc6 density was very weak, indicating high flexibility of this protein. However, based on the homologous *Drosophila* ORC core complex structure (PDB ID: 4XGC) [7], we identified a short α-helix at the C-terminus of Orc6 that bond to and was stabilized by Orc3. Finally, the double stranded DNA was built into the long helical density that ran through the OCCM structure in the program COOT. The entire OCCM atomic model was subsequently adjusted manually and rebuilt in COOT. Clearly resolved bulky residues such as Phe, Tyr, Trp, and Arg were used for sequence registration.

The manually built atomic model was then iteratively refined in real space by phenix.real_space_refine [66] and rebuilt in COOT. We also performed the reciprocal space refinement procedure with the application of secondary structure and stereochemical constraints in the program Phenix [67]. The structure factors (including phases) were calculated by Fourier transform of the experimental density map with the program Phenix.map_to_structure_factors. The atomic model was validated using MolProbity [68]. Structural figures were prepared in Chimera and Pymol (https://www.pymol.org).

The final model was cross validated using a method described previously [69]. Briefly, we randomly added 0.1 Å noise to the coordinates of the final model using the PDB tools in Phenix, then refined the noise-added against the first half map (Half1) that was produced from one half of the particle dataset during refinement by RELION. We performed one round of coordinate refinement, followed by a B-factor refinement. The refined model was then correlated with the

3D maps of the two half maps (Half1, Half2) in Fourier space to produce two FSC curves: $FSC_{work}$ (model versus Half1 map) and $FSC_{free}$ (model versus Half2 map), respectively. A third FSC curve was calculated between the refined model and the final 3.9 Å resolution density map produced from all particles. The general agreement of these curves was taken as an indication that the model was not over-fitted.

**Cross-linking/Mass Spectrometry analysis**

240 nM pUC19-ARS1 beads were used to assemble the OCCM complex as previously described. The complex was cross-linked in the presence of the beads with BS3 1:8100 (molar protein:cross-linker ratio) for 2 hours at 4 ºC. Then, the reaction was quenched with 50 ul of saturated ammonium bicarbonate for 45 minutes at 4 ºC and it was transferred into digestion buffer (50 mM ammonium bicarbonate, 8 M Urea) followed by reduction with DTT and alkylation with iodoacetamide. 3 ug LysC (with estimated 1:50 enzyme to protein ratio) were added and incubated at room temperature for 4 hours. The digestion buffer was then diluted with 50 mM ammonium bicarbonate to a final Urea concentration of 2M. 3 ug of trypsin were added (at estimated 1:50 enzyme to protein ratio) and incubated for 16 hours at room temperature. After digestion, the supernatant was collected and acetified using 200 ul 10% Trifluoroacetic acid (TFA). The peptide mixture was then desalted using C18-Stage-Tips for mass spectrometric analysis [70]. LC-MS/MS analysis was performed using an Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer (Thermo Scientific) applying a "high-high" acquisition strategy [71]. 2 ug peptide mixture was injected for each mass spectrometric acquisition. Peptides were separated on a 50 centimeter EASY-Spray column (Thermo Scientific) assembled in an EASY-Spray source (Thermo Scientific), operated at 50 ºC column temperature. Mobile phase A consisted of water, 0.1% v/v formic acid and 5% v/v DMSO. Mobile phase B consisted of 80% v/v acetonitrile, 0.1% v/v formic acid and 5% v/v DMSO. Peptides were loaded at a flow-rate of 0.3 µl/min and eluted at 0.2 µl/min using a linear gradient going from 2% mobile phase B to 40% mobile phase B over 139 minutes, followed by a linear increase from 45% to 95% mobile phase B in 12 minutes. The eluted peptides were directly introduced into the mass spectrometer. MS data was acquired in the data-dependent mode with the top-speed option. For each three-second acquisition cycle, the mass spectrum was recorded in the Orbitrap with a resolution of 120,000. The ions with a precursor charge state between 3+ and 8+ were isolated and fragmented. The fragmentation spectra were then recorded in the Orbitrap. Dynamic exclusion was enabled with single repeat count and 60-second exclusion duration. To improve the identification of cross-linked peptides, in total 9 acquisitions were carried out with variations on parameters mainly related to criteria for ion selection for fragmentation and fragmentation methods.

**Identification of cross-linked peptides**

The raw mass spectrometric data files were processed into peak lists using MaxQuant version 1.5.2.8 [72] with default parameters, except for "FTMS top peaks per 100 Da" was set to 20 and "FTMS de-isotoping" was disabled. The peak lists were searched against the sequences as well as the reversed sequences (as decoy) of 14 OCCM subunits using Xi software (ERI, Edinburgh) for identification of cross-linked peptides. Search parameters were as follows: MS accuracy, 6 ppm; MS2 accuracy, 20 ppm; enzyme, trypsin; specificity, fully tryptic; allowed number of missed cleavages, four; cross-linker, BS3; fixed modifications, carbamidomethylation on cysteine; variable modifications, oxidation on methionine, modifications by BS3 that are hydrolyzed or amidated on the other end. The reaction specificity for BS3 was assumed to be for

lysine, serine, threonine, tyrosine and protein N-termini. For acquisitions where CID and HCD fragmentations were applied (acquisition 1.3.4.6.7 and 9), only b- and y-ions were considered for the fragment ion matches; while for data acquired using a combined fragmentation of ETD and CID or HCD (acquisition 2,5 and 8), b-, y-, c-ions and z-ions were considered for fragment ion matches. Two independent quality control approaches have been applied for identified cross-linked peptide candidates. All cross-linked peptides with estimated 5% FDR at residue pair level were accepted for further structural interpretation. We also accepted cross-linked peptides identified with MS2 spectra that passed machine-learning based auto-validation.

**In vitro pulldown of Cdt1 with Mcm2-7 proteins**

Baculoviruses expressing Mcm2-7 subunits (the Mcm3 was tagged with HA at the N-terminus) and Strep-Strep-SUMO-Cdt1 were prepared in the Profold-C1 expression vector (AB Vector, San Diego). 2.5 x 107 Hi-Five insect cells were infected at a multiplicity of infection of 10 with each MCM subunit alone or in combination with SSS-Cdt1, and harvested at 55 hours post infection. After washing cells in cold PBS, they were placed for 10 minutes on ice in < 1 ml of hypotonic buffer (25 mM Hepes-KOH pH 7.5, 20 mM K glutamate, 1 mM Mg acetate, 1 mM DTT, 5 mM ATP, Protease Inhibitor (Roche, 1 tablet in 50 ml). Cell extracts were prepared by Dounce B homogenization and centrifugual clarification (10,000 rpm in Sorvall SS34 rotor), then 100 µl of supernatant was incubated of 2 hours on ice with 50 µl of Strep-Tactin sepharose (IBA) that had previously been washed in PBS and hypotonic buffer. Beads were washed (3 x 5 minutes) with IP buffer (25 mM Hepes-KOH pH 7.5, 300 mM KGlutamate, 10 mM MgAcetate, 0.04% NP40, 1 mM DTT, 5 mM ATP and protease inhibitor). Bead bound proteins were run on a SDS-PAGE gel and stained with silver.

## METHOD SECTION REFERENCES


56. Li, X. et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods* **10**, 584-90 (2013).
57. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-21 (2015).
58. Scheres, S.H. Semi-automated selection of cryo-EM particles in RELION-1.3. *J Struct Biol* **189**, 114-22 (2015).
59. Biasini, M. et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic acids research*, W252-258 (2014).
60. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).
61. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-12 (2004).
62. Li, W., Zhang, T. & Ding, J. Molecular basis for the substrate specificity and catalytic mechanism of thymine-7-hydroxylase in fungi. *Nucleic acids research*, 10026-10038 (2015).
63. De Marco, V. et al. Quaternary structure of the human Cdt1-Geminin complex regulates DNA replication licensing. *Proceedings of the National Academy of Sciences* **106**, 19807-19812 (2009).
64. Doyle, J.M., Gao, J., Wang, J., Yang, M. & Potts, P.R. MAGE-RING protein complexes comprise a family of E3 ubiquitin ligases. *Molecular cell* **39**, 963-974 (2010).
65. Wiedemann, C. et al. Structure and regulatory role of the C-terminal winged helix domain of the archaeal minichromosome maintenance complex. *Nucleic acids research* **43**, 2958-2967 (2015).
66. Afonine, P.V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352-67 (2012).
67. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).
68. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).
69. Amunts, A. et al. Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485-9 (2014).
70. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2**, 1896-906 (2007).
71. Chen, Z.A. et al. Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J* **29**, 717-26 (2010).
72. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-72 (2008).