Single-molecule studies of eukaryotic helicase loading

by

Simina Ticau

B.A. Molecular and Cell Biology
B.S. Bioengineering
UC Berkeley, 2010

Submitted to the Department of Biology in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy in Biology
at the
Massachusetts Institute of Technology

February 2016

© 2016 Simina Ticau. All rights reserved.

The author hereby grants MIT the permission to reproduce and distribute publicly paper
and electronic copies of this thesis in whole or in part.

Signature redacted

Department of Biology
January 4, 2016

Certified by:  

Signature redacted

Stephen P. Bell
Professor of Biology
Thesis Supervisor

Accepted by:  

Signature redacted

Amy E. Keating
Professor of Biology
Co-Chair, Biology Graduate Committee

The author hereby grants MIT permission to reproduce and to
distribute publicly paper and electronic copies of this thesis document
in whole or in part in any medium now known or hereafter created.
Single-molecule studies of eukaryotic helicase loading

by

Simina Ticau

Submitted to the Department of Biology on January 4, 2016 in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biology

ABSTRACT

Cells must duplicate their genomic content fully and accurately in each cell cycle to maintain cellular identity and ensure the viability of their progeny. The first step in eukaryotic DNA replication initiation is the loading of the heterohexameric Mcm2-7 helicase. In eukaryotes helicase loading is tightly regulated to ensure that this event only occurs during the G1 phase of the cell cycle. Because helicase activation can only occur once cells enter S phase, cells can only load and activate the helicase once per cell cycle, ensuring the genome is replicated once and only once in each cell cycle. The selection of sites of helicase-loading also marks all potential origins of replication. Once loaded, the helicases encircle dsDNA and are linked in a head-to-head double hexamer. Although the proteins involved in helicase loading are known (ORC, Cdc6, and Cdt1), the mechanism by which they load two oppositely-oriented helicases and ensure their proper architecture remains under intense investigation.

In this thesis I describe a novel single-molecule helicase-loading assay that allows monitoring of protein associations and dissociations on a one-second time scale. By labeling pairs of helicase-loading proteins simultaneously, I determined the relative time of association and dissociation for the helicase and each of the helicase-loading proteins during helicase loading. Additionally, I determined the stoichiometry of each helicase-loading protein with respect to the origin DNA. Adapting this assay to read out distance information using single-molecule FRET, I monitored formation of the final double-hexamer in real time.

These single-molecule assays uncovered that helicase loading occurs in a one-at-a-time manner and discovered novel steps in the mechanism of helicase loading. Following the initial association of ORC/Cdc6/Cdt1/Mcm2-7 with the origins of DNA replication, Cdc6 and then Cdt1 are released sequentially. A new Cdc6, and Cdt1/Mcm2-7 are subsequently recruited and the same ordered sequential release is observed, although with different kinetics. Although two Cdc6 and Cdt1 proteins are required for loading a double hexamer, a single ORC is sufficient. Additionally, double-hexamer formation is a rapid event upon association of the second helicase, suggesting a model in which the two helicases are recruited and loaded around dsDNA by distinct mechanisms.

Thesis Supervisor: Stephen P. Bell
Title: Professor of Biology
Acknowledgements

I am grateful for the many friends and colleagues I have met during my time here at MIT. Friends who have been there for me both academically and socially and with whom I have shared many celebrations. In particular members of the Bell lab who have helped create a fun but also mentally stimulating environment. It has been a privilege to be surrounded by such amazing people and brilliant scientists.

I would like to thank the members of my thesis committee: Iain Cheeseman and Bob Sauer for their support and scientific advice over the years. I would also like to thank my collaborators at Brandeis University: Larry Friedman and Jeff Gelles. They have played a crucial part in the development of my thesis as well as my growth as a scientist and have been a pleasure to work with.

I especially thank my advisor, Steve Bell, whose guidance throughout the years has been invaluable to my development as a scientist. His dedication to scientific excellence has continuously inspired and encouraged me to grow as a scientist.

Finally, I would like to thank my parents and my two wonderful grandmothers who have inspired me through their determination and hard work and always supported and encouraged me in all aspects of my life.
TABLE OF CONTENTS

Abstract ...................................................................................................................... 3
Acknowledgements .................................................................................................... 5
Table of Contents ...................................................................................................... 7

Chapter I: Introduction .............................................................................................. 9
  Overview .................................................................................................................. 10
  Origins of DNA Replication .................................................................................... 13
    E. coli .................................................................................................................. 14
    S. cerevisiae ........................................................................................................ 16
  Initiator Proteins ..................................................................................................... 18
    E. coli: DnaA ....................................................................................................... 19
    S. cerevisiae: ORC .............................................................................................. 22
  Additional Helicase Loading Proteins ................................................................... 25
    E. coli: DnaC ....................................................................................................... 25
    S. cerevisiae: Cdc6 and Cdt1 ............................................................................ 26
      Cdc6 ................................................................................................................ 27
      Cdt1 ................................................................................................................ 29
  The Replicative Helicase ......................................................................................... 30
    E. coli: DnaB ....................................................................................................... 30
    S. cerevisiae: Mcm2-7 ....................................................................................... 31
  Helicase Loading: Two Helicases, Opposite Orientations ..................................... 36
    E. coli .................................................................................................................. 37
    S. cerevisiae ........................................................................................................ 40
  Regulation of DNA Replication ............................................................................. 44
    E. coli .................................................................................................................. 45
    S. cerevisiae ........................................................................................................ 46
  Single Molecule Studies of DNA Replication ....................................................... 47
  Thesis Summary ..................................................................................................... 49
  References .............................................................................................................. 50

Chapter II: Single-molecule Studies of Origin Licensing Reveal Mechanisms
  Ensuring Bidirectional Helicase Loading ................................................................. 65
  Summary ................................................................................................................ 66
  Introduction ............................................................................................................ 67
Chapter III: Discussion

Key conclusions .................................................. 138
The Mcm2-Mcm5 gate ......................................... 139
The role of ORC and Cdc6 ATP hydrolysis in helicase loading ..................................... 141
The role of the second Cdc6 during helicase loading ...................................................... 142
Extending the single-molecule assay to replication initiation ........................................ 145
References ......................................................... 148
Chapter I

Introduction
OVERVIEW

The duplication of genetic content is crucial for all domains of life. Archaea, bacteria and eukaryotic cells all must accurately duplicate their genomes for cell/organismal survival. The process of DNA replication is carefully regulated through multiple mechanisms in eukarya and bacteria (Alberts et al., 2002). In eukaryotes, imprecise, excessive or misregulated DNA replication leads to genomic instability, developmental abnormalities, cancerous transformation and death (Blow and Gillespie, 2008; Klingseisen and Jackson, 2011).

Over 50 years ago Jacob, Brenner and Cuzin proposed the first model for DNA replication initiation. Their replicon model hypothesized that binding of a protein (the initiator) to a specific genomic region (the replicator) positively regulates DNA replication (Jacob et al., 1963). This basic outline of this model has been confirmed in all cellular organisms studied. On the other hand, it is now clear that this model describes only the first step in a complex process in which the initiator recruits multiple proteins to the replicator, eventually leading to assembly of two replisomes, the macromolecular protein assemblies that duplicate DNA.

The number of replicators on a chromosome varies between organisms. A single replicator is sufficient to replicate the 4.6 Mb genome of Escherichia coli (E. coli). In contrast, tens of thousands of replicators are spread across the 3,000 Mb human genome (reviewed in O’Donnell et al., 2013). In organisms with larger genomes, the presence of
multiple replicators enables a ‘divide and conquer’ approach to replication. By assembling many replisomes across the genome, even a huge genome can be replicated in a timely fashion.

The replicative helicase is at the heart of DNA replication. Loading of the DNA helicase is the first step in building an active replisome. In both bacteria and eukaryotic cells, the majority of replication fork proteins are recruited by interactions with the loaded helicase. The central role of the helicase continues during DNA synthesis. At this stage, interactions between the helicase and the DNA polymerases and their accessory proteins coordinate DNA unwinding with DNA synthesis (O’Donnell, 2006). For example, in *E. coli*, the helicase interacts with the primase and the sliding clamp loader, which in turn associate with the DNA polymerases.

Helicase loading is a critical step of DNA replication initiation that is highly regulated. Eukaryotic cells safeguard against re-replication by temporally separating loading of the Mcm2-7 helicase, which occurs during G1, and activating the helicase, which occurs upon entry into S phase (Diffley, 2004). Much of this regulation is focused on preventing helicase loading outside of G1. This temporal separation ensures that the genome is duplicated only once per cell cycle. In *E. coli*, the critical regulated step involves binding of the initiator to DNA. The initiator connects the replicator with helicase loading, acting as a scaffold to recruit the proteins required for loading. Initiator recruitment to the replicator is blocked by multiple mechanisms after DNA replication initiation has already occurred, to safeguard against multiple rounds of replication in a
single cell cycle (reviewed in Mott and Berger, 2007). In most organisms helicase loading is mechanistically accomplished with the help of one or more helicase loading proteins. Importantly, at chromosomal replicators two helicases must be loaded in opposite orientations to ensure formation of a bidirectional fork and, therefore, replication of the DNA on both sides of the replicator.

The initiation of DNA replication has been best studied in E. coli and eukaryotic cells. In these settings, replication initiation involves a step-wise process in which multiple proteins load the replicative helicase onto DNA and coordinate activation of the helicase with initiation of DNA synthesis. Both systems follow the basic outline of the replicator-bound initiator recruiting one or more helicase-loading proteins followed by recruitment and loading of the replicative helicase. I will focus on helicase-loading in the eukaryotic organism S. cerevisiae as well as in E. coli where the most is known about the proteins involved and their interactions.

This thesis focuses on the mechanism of eukaryotic helicase loading in S. cerevisiae. In this introduction, I review our current understanding of the molecular players involved in helicase loading and compare our knowledge of this process in two systems: the bacterial E. coli and the eukaryotic S. cerevisiae. After discussing the proteins involved, I will introduce several proposed mechanisms of how the two replicative helicases are loaded in opposite orientations in both of these domains of life.
For the entire genome of an organism to be duplicated in a timely manner, initiation of DNA replication must occur at one or more sites along a chromosome. Whereas replicators are the cis-acting sequences that direct recruitment of the replication proteins to DNA, origins of replication are the sites at which replication initiation successfully occurs. In many instances, these DNA regions are overlapping but in other cases the replicator is much larger than the origin that it encompasses (Stillman, 1993). In *E. coli*, the single replicator, oriC overlaps with the single origin of replication (von Meyenburg et al., 1977). Similarly, although there are hundreds of potential origins in *S. cerevisiae* cells, they each tightly overlap with a conserved replicator sequence (Brewer and Fangman, 1987; Huberman et al., 1987). In contrast, in human cells this correspondence is not observed. Not only has no conserved sequence associated with origins of replication been identified (Cvetic and Walter, 2005), when mapped the replicators can be much larger than the corresponding origin. For example, the β-globin locus control region is required for initiation from an origin that is 40kb away (Stillman, 1993).

Even though replicators and origins of replication overlap in *S. cerevisiae*, this overlap is not necessary for replication (Gros et al., 2014). The inactive form of the helicase can move on naked dsDNA away from the location of its original loading and still initiate replication (Gros et al., 2014). This suggests that there is no DNA sequence requirement for activation of the helicase and initiation of replication once it is loaded.
Despite these findings, there is currently no evidence that such movement occurs *in vivo*, possibly because such sliding is restricted by nucleosomes.

**E. coli**

Bacteria typically have a single origin of replication that is sufficient to duplicate the entire genome. One of the best-characterized bacterial origins of replication is the single *E. coli* chromosomal origin, *oriC* (reviewed in Mott and Berger, 2007). *oriC* is approximately 250 bp long and is composed of multiple repeats of two sequences (Figure 1A): four 9-base pair (bp) repeats that bind the *E. coli* initiator protein, DnaA; and three 13-bp repeats (13-mers), which are AT-rich regions that serve as a DNA unwinding element (DUE) and the initial site of DNA unwinding (Kowalski and Eddy, 1989; Bramhill and Kornberg, 1988; Zyskind et al., 1983). The 9-bp repeats, are called DnaA boxes because they contain consensus binding sites for DnaA. Together with lower-affinity DnaA binding sites (termed I-sites and τ-sites), the DnaA boxes recruit multiple copies of DnaA to *oriC* (McGarry et al., 2004; Speck et al., 1999). The τ-sites differ from the DnaA consensus binding sequence and have lower affinities for DnaA, mainly recruiting ATP-bound DnaA (as opposed to ADP-bound DnaA). Interestingly, the spacing between the different elements, especially relative positioning of the DnaA binding sites and the DUE, is critical for origin function (Holz et al., 1992). Although first discovered in *E. coli*, this basic replicator structure is conserved across bacterial species (Leonard and Grimwade, 2015).
Figure 1

A.

ATP-DnaA Boxes

Escherichia coli oriC

DUE

IHF

R1

R5

R6

R7

R8

R9

R10

R11

R12

R13

R14

DnaA Box

I-site

T-site

DUE

B.

Saccharomyces cerevisiae ARS1

B2

B1

A

Cdc6

ORC

Figure 1. Overview of origins of replication
The organization of (A) the *E. coli* oriC and (B) the *S. cerevisiae* ARS1 (Costa et al., 2013).
*S. cerevisiae*

*S. cerevisiae* cells contain approximately 400 functional origins of replication (Nieduszynski et al., 2006) that each include a conserved DNA sequence. The sequence-defined nature of *S. cerevisiae* origins was discovered in screens for genomic fragments that enabled plasmids to replicate alongside the genome in each cell cycle, called autonomously replicating sequences (ARSs; Hsiao and Carbon, 1979; Stinchcomb et al. 1979). These sequences share an 11 bp sequence element called the ACS (ARS consensus sequence). Subsequent studies showed the ACS is essential for origin function (Broach et al., 1983; Celniker et al., 1984; Van Houten and Newlon, 1990). Further genome-wide comparisons led to the expansion of the ACS to a sequence encompassing ~30bp (Van Houten and Newlon, 1990; Xu et al., 2006; Liachko et al., 2013). Similar to the 9-mer sites in the *E. coli* oriC that serve to recruit DnaA, the ACS is bound by the *S. cerevisiae* initiator, the origin recognition complex (ORC; Bell and Stillman, 1992). Unlike the multiple binding sites for DnaA found at *oriC*, there is only one copy of the ACS at each *S. cerevisiae* origin.

Although essential, the ACS is not sufficient to define an origin. In a comprehensive linker-substitution study of the *ARS1* replicator, three additional elements that contribute to origin function (B1, B2, B3) were identified (Marahrens and Stillman, 1992). Deletion of any one of these elements reduced *ARS1* origin function, but mutation of all three B-elements inactivated the origin. Subsequent studies found that ORC bound not only the ACS but also part of the B1 sequence (Rao et al., 1994; Rowley et al., 1995). Although the AT-rich B2 element has been proposed to function as a DUE (Lin and
Kowalski, 1997), this element is likely to have another function since other easily unwound sequences are unable to functionally substitute for B2 (Wilmes and Bell, 2002). A screen for sequences that functioned as a B2 yielded sequences similar to an inverted ACS (Wilmes and Bell, 2002), suggesting that the B2 region recruited a second ORC to origins. However, poor ORC binding sites can also act as a B2 element and a second inverted match to the ACS is not necessary for origin function (Chang et al., 2011). The B3 element, which is not found at most origins, binds a transcription factor, Abf1 (ARS binding factor 1), a protein that influences both transcription and replication by positioning local nucleosomes (Diffley and Stillman, 1998; Lipford and Bell, 2001; Miyake et al., 2002).

Determinants beyond DNA sequence are required to define the location of a S. cerevisiae origin. The S. cerevisiae genome contains over 6,000 11bp-consensus ACS sites, yet only a few hundred are bound by ORC. An important clue to this distinction was revealed when the chromatin structure surrounding ACS sites that are bound by ORC was investigated. These sites were associated with nucleosome-free regions (NFRs), with nucleosomes positioned asymmetrically relative to the ACS (Eaton et al., 2010). ORC binding influences the positioning of the nucleosomes flanking both sides of the origin (Lipford and Bell, 2001; Eaton et al, 2010). The resulting NFR is much larger than expected if ORC DNA binding alone excluded the adjacent nucleosomes. Although the mechanism responsible for the establishment of origin-overlapping NFRs is unclear, the availability of nucleosome-free DNA beyond that bound by ORC is thought to provide a site for loading of the Mcm2-7 eukaryotic helicase. ORC remains associated with origin
DNA throughout the cell cycle (Gibson et al., 2006), possibly to maintain nucleosome positioning proximal to origins.

Unlike their bacterial and budding yeast counterparts, metazoan origins of replication are not defined by specific sequences. Instead, sites of replication initiation are correlated with actively transcribed genes, DNA sequences predicted to form G-quadruplex structures, and an open-chromatin environment (Jin et al. 2009; MacAlpine et al. 2010; Stroud et al. 2012). Interestingly, the BAH domain of the largest subunit of human ORC preferentially binds a histone modification, H4K20me2 (the bromo-adjacent homology (BAH) domain), and this recognition is conserved among multiple metazoan Orc1 proteins (Kuo et al., 2012). This modification is associated with heterochromatic regions, raising the possibility that ORC is recruited to direct initiation of replication at these regions by recognizing this modification. Whether these determinants affect chromatin structure, recruitment of proteins (e.g. ORC) or facilitate initiation through other means remains unclear.

INITIATOR PROTEINS

From both structural and biochemical studies we know that all cellular initiators contain a common nucleotide-binding domain belonging to the ATPases Associated with various cellular Activities (AAA+) family (Duderstadt and Berger, 2008). Initiators are part of the initiator/helicase-loader clade, a subtype containing an additional helix insertion called an initiator-specific motif (ISM; Duderstadt and Berger, 2008). Typically,
AAA+ proteins form homo- or hetero-oligomers where the interface between two subunits serves as an ATP binding and hydrolysis domain. One subunit contributes the bulk of the ATP binding domain using distinct signature sequences, known as the Walker A and Walker B motifs (Walker et al., 1982). The neighboring subunit contributes an arginine-finger motif that is required for ATP hydrolysis but not ATP binding. The state of nucleotide bound to the AAA+ motif typically modulates the conformation of the AAA+ assembly in a manner that either performs work or regulates protein function and/or assembly.

**E. coli: DnaA**

DnaA has multiple functions as the *E. coli* initiator (reviewed in Mott and Berger, 2007). DnaA binds specific sequences at *oriC* in an ATP-regulated fashion (see above). Once DnaA•ATP is bound to the origin, it unwinds the adjacent 13-mer region of dsDNA. Finally, DnaA is required to recruit the replicative helicase, DnaB, and works with DnaC (a helicase-loading protein) to load one helicase on each of the ssDNA strands. Importantly, the two helicases each encircle the ssDNA and must be loaded in opposite orientations to ensure bidirectional initiation.

DnaA contains four distinct domains that each contribute to one or more of the functions mentioned above. First, the N-terminal domain (domain I) mediates interactions with the DnaB helicase and promotes oligomerization of DnaA (Simmons et al., 2003; Seitz et al., 2000; Sutton et al., 1998; Weigel et al., 1999). Domain II is a poorly conserved linker domain whose function is not well defined (Messer et al., 1999; Sutton...
and Kaguni, 1997). Domain III contains the AAA+ ATPase activity and is also the main determinant of DnaA oligomerization (Erzberger et al., 2002; Erzberger and Berger, 2006; Kawakami et al., 2006). The ISM insertion plays a role in rotating the DnaA subunits relative to each other upon oligomerization. This rotation prevents the formation of a closed ring that is typical of AAA+ oligomers. Instead, DnaA forms a helical assembly (Figure 2A; Duderstadt et al., 2011; Duderstadt and Berger, 2008). The AAA+ region is also thought to act as a secondary DNA binding site due to a number of conserved and positively-charged residues lining the central axis of the helical filament (Erzberger et al., 2006). The positively-charged interior channels of many DNA binding AAA+ proteins are known to interact with DNA (Duderstadt and Berger, 2008). Domain III is also thought to engage ssDNA (Bramhill and Kornberg, 1988; Erzberger et al., 2006; Speck and Messer, 2001). Lastly, the C-terminal domain (domain IV) provides sequence-specific DNA-binding function that is necessary to recognize the DnaA boxes (Messer et al., 1999; Roth and Messer, 1995).

Formation of DnaA oligomers is critical for oriC DNA unwinding (Duderstadt and Berger, 2008). Similar to other AAA+ ATPases, the interface between two DnaA monomers forms the ATP binding and hydrolysis motif and the state of the bound nucleotide controls its oligomerization (Erzberger et al., 2006; Duderstadt et al., 2011). The γ-phosphate of the bound ATP alters the conformation of DnaA to allow the interaction of this phosphate with the arginine finger of an adjacent DnaA monomer (Erzberger et al., 2006). Multimerization at oriC, unwinds the nearby DUE region (Bramhill and Kornberg, 1988; Kowalski and Eddy, 1989). Although the mechanism of
Figure 2. DnaA bound to DNA
Ribbon diagram of the DnaA filament structure (top, PDB ID 2HCB). Hypothetical model where binding and wrapping of DNA occur (bottom; Duderstadt and Berger, 2008).
unwinding is not known, one model proposes that the DNA binds around the exterior of the DnaA filament in a right-handed helix. This imparts positive superhelicity to the bound DNA, leading to compensatory negative supercoiling of the adjacent DNA (including the 13-mers) that destabilizes the duplex DNA in the 13-mers (Erzberger et al., 2006; Duderstadt and Berger 2008). Following initial unwinding, ATP-DnaA binds one strand of the ssDNA in the partially unwound 13-mer DNA using domain III (Speck and Messer, 2001; Weigel and Seitz, 2002). Binding of ssDNA involves stretching the DNA in a way that inhibits base pairing to the unbound DNA (Duderstadt et al., 2011).

**S. cerevisiae: ORC**

The origin recognition complex (ORC) was discovered in *S. cerevisiae* due to its affinity for the conserved ACS (Bell and Stillman, 1992). ORC is a six-subunit complex that acts as the initiator in all eukaryotic organisms studied (Bell and Dutta, 2002). In *S. cerevisiae*, ORC binds to a ~30 bp region that encompasses the ACS and B1 regions of the replicator (Bell and Stillman, 1992; Lee and Bell, 1997; Rowley et al., 1995). Additionally, ORC is critical for the recruitment and loading of the replicative Mcm2-7 helicase at origins of replication (see below).

Like DnaA, five of the six ORC subunits (Orc1-5) are members of the initiator clade of the AAA+ ATPase family of proteins (Erzberger et al., 2004). In addition, each of these subunits includes a winged-helix domain at their C-terminal end. A high-resolution structure of *Drosophila melanogaster* ORC indicates that ORC forms an incomplete ring-shaped complex with a domain-swapped architecture, where the N-
terminal AAA+ domain stacks on top of the neighboring subunit’s C-terminal WH domain (Figure 3A; Bleichert et al., 2015). The ISM domains of Orc1-5 line up on the inner surface, similar to DnaA, and the diameter of the ring and the location of the ISM motifs suggest that they mediate interactions with the DNA (Figure 3B; Bleichert et al., 2015).

ATP binding and hydrolysis is important for ORC function. Only Orc1 and Orc5 bind ATP, and only Orc1 hydrolyzes ATP (Klemm et al., 1997). Orc1 ATP binding is important for origin binding, and mutants that prevent Orc1 ATP binding or hydrolysis are lethal (Bell and Stillman, 1992; Klemm et al., 1997; Bowers et al., 2004). ORC ATP hydrolysis is dispensable for helicase loading but biochemical analysis of these mutant complexes suggest that ORC ATP hydrolysis is required for repeated loading (Bowers et al., 2004). This finding suggests that multiple helicase-loading events are critical for budding yeast viability. Interestingly, ORC also has affinity for ssDNA that is sequence- and ATP-independent (Lee et al., 2000), however, this function is not known to be involved in any event during DNA replication initiation.

The Orc6 subunit is the only subunit that is not related to the AAA+ family of ATPases (Erzberger and Berger, 2006). Orc6 contains a domain that is similar to transcription factor IIB, and a C-terminal domain (Liu et al., 2013; Bleichert et al., 2013). The C-terminal domain is important for associating with Orc1-5 in metazoans (Bleichert et al., 2013). Although not required for ORC to bind DNA (Lee and Bell, 1997), Orc6 is essential in vivo (Semple et al., 2006; DaSilva and Ducker, 2007). Biochemical studies
Figure 3

(A) Surface view of ORC highlighting the domain-swap architecture of the N-terminal and C-terminal tiers.

(B) Model of Drosophila ORC bound to DNA using superposition of DNA-bound archaean Orc1-1. The ISMs of Orc1–5 are highlighted. The DNA path is shown as a tan ribbon through the central channel. Orc1 has been repositioned (Bleichert et al., 2015).
showed that Orc6 is essential for helicase loading (Chen et al., 2007) though not for initial recruitment of the helicase (Frigola et al., 2013). The precise role that Orc6 plays in helicase loading remains unknown.

**ADDITIONAL HELICASE LOADING PROTEINS**

The initiator is not sufficient to recruit the replicative helicase in *S. cerevisiae* or *E. coli*. In each case, at least one additional protein is required for helicase recruitment and loading, and these proteins serve as additional points of regulation. In *E. coli*, DnaA and DnaC are sufficient for helicase loading, whereas in *S. cerevisiae*, ORC, Cdc6 and Cdt1 are required. The roles of these additional proteins are discussed below.

**E. coli: DnaC**

DnaC binds the DnaB helicase and together with DnaA loads DnaB onto ssDNA (reviewed in O’Donnell et al., 2013). DnaC is also an AAA+ ATPase protein and a paralog of DnaA, but lacks the C-terminal dsDNA-binding domain (Mott et al., 2008). DnaC binds DnaA, DnaB and ssDNA (reviewed in Costa et al., 2013). The N-terminal ends of DnaC interact with the C-terminal ends of DnaB and this interaction inhibits the helicase activity of DnaB (Davey et al., 2002). Additionally it has been proposed that the binding of DnaC to DnaB stabilizes an open-ring form of the DnaB helicase (Arias-Palomo et al., 2013; Barcena et al., 2001; Galletto et al., 2003). Although there are different models for the stoichiometry of DnaC-bound DnaB (6:6 vs 3:6; Galletto et al., 2003; Kobori and Kornberg, 1982; Makowska-Grzyska and Kaguni, 2010), it is clear that
multiple DnaC molecules bind DnaB helicase.

Although the role of ATP in the function in DnaC has not been fully elucidated, several functions of ATP binding and hydrolysis are known. ATP binding is needed for DnaC to form a DnaA-like helical oligomer and bind ssDNA and DnaA (Mott et al., 2008; Biswas et al., 2004). Additionally, ATP-DnaC extends the already unwound region at oriC (Davey et al., 2002). Although ATP binding by DnaC is not required for binding DnaB, it is required to load DnaB (Davey et al., 2002). After loading of DnaB, DnaC must be released to activate DnaB helicase activity (Davey et al., 2002).

Multiple events stimulate the release of DnaC from DnaB. DnaC is a feeble ATPase on its own, and it has been suggested that both DnaB and ssDNA (at the unwound oriC) stimulate the ATPase activity of DnaC (Davey et al., 2002; Mott et al., 2008; Biswas et al., 2004). Additionally, primase binding to DnaB is thought to stimulate release of DnaC (Makowska-Grzyska and Kaguni, 2010). Primase interacts with the amino-terminal region of DnaB (Lu et al., 1996; Tougu et al., 1994), the opposite site from where the DnaB-DnaC interaction occurs. Interestingly, primer synthesis is required for DnaC dissociation, suggesting a mechanism by which E. coli couples activation of the helicase with initiation of DNA replication (Makowska-Grzyska and Kaguni, 2010).

**S. cerevisiae: Cdc6 and Cdt1**

In *S. cerevisiae*, there are two proteins that aid the initiator in helicase loading: Cdc6 and Cdt1. Both Cdc6 and Cdt1 are required for Mcm2-7 recruitment to origins and
are essential for viability (Cocker et al., 1996; Liang and Stillman, 1997; Donovan et al., 1997; Tanaka and Diffley, 2002). Cdc6 was identified in a screen for mutants in S. cerevisiae with changes in the cell division cycle and shown to be important for DNA replication (Hartwell, 1973). Cdt1 was first identified based on its cell cycle regulated expression in Schizosaccharomyces pombe (S. pombe) (Hofmann and Beach, 1994) and was subsequently implicated in helicase loading based on its ability to cooperate with Cdc6 to induce more efficient re-replication of the genome (Nishitani et al., 2000).

**Cdc6**

Cdc6 binds ORC and is required for recruitment and loading of the Mcm2-7 helicase to origins (Cocker et al., 1996; Liang and Stillman, 1997; Donovan et al., 1997). Cdc6 likely exerts its function by inducing a conformational change in ORC that exposes an Mcm2-7-interacting domain (Sun et al., 2012). Cdc6 association with ORC is facilitated by origin DNA, ensuring localization of ORC/Cdc6 recruitment of the helicase (Mizushima et al., 2000). Upon recruitment to origins, Cdc6 binds to Orc1-5 between the Orc2 and Orc5 subunits, and together these proteins form a ring around dsDNA (Sun et al., 2013). The C-terminal region of the Orc1-5/Cdc6 ring is thought to recruit the first helicase through interactions with the C-terminal end of the first recruited Mcm2-7 helicase ring (Figure 4; Sun et al., 2013).

In addition to being required for Mcm2-7 recruitment to origins, Cdc6 increases the specificity of ORC for origin DNA (Mizushima et al., 2000; Speck and Stillman, 2007; Duzdevich et al., 2015). This was proposed to occur by a mechanism where Cdc6
Figure 4

The EM structure of ORC/Cdc6/Cdt1Mcm2-7

Different rotational views of the ORC/Cdc6/Cdt1Mcm2-7 EM structure showing the different interactions between every subunit in this macromolecular complex. The DNA is thought to pass vertically though the center of each of these views (Sun et al., 2013).
reduces the off-rate of ORC from origin fragments, but induces release of ORC at non-origin DNA (Speck and Stillman, 2007). More recently, it was shown that Cdc6 forms a transient complex with ORC off DNA that is unable to bind DNA (Duzdevich et al., 2015). This would be consistent with Cdc6 closing the Orc1-5 ring, and thereby inhibiting DNA binding after complex formation but also stabilizing ORC-DNA association. This interaction between ORC and Cdc6 away from DNA reduces the effective concentration of ORC, causing its observed increased specificity for origin DNA, and consistent with ORC/Cdc6 forming the same closed ring off DNA (Duzdevich et al., 2015).

Like Orc1-5, DnaA, and DnaC, Cdc6 is also an AAA+ ATPase, and is highly homologous to Orc1 (Neuwald et al., 1999). Mutations of its ATP hydrolysis domain (Walker B) is essential in yeast, but is bypassed by several intragenic mutations (Perkins and Diffley, 1998; Weinreich et al., 1999; Chang et al., 2015). Biochemical studies show that ATP hydrolysis is not required for Mcm2-7 helicase loading in vitro (Kang et al., 2014; Chang et al., 2015; Coster et al., 2014). Instead, ATP hydrolysis by Cdc6 is important for its own dissociation from DNA and this event is required for subsequent helicase activation (Chang et al., 2015). Cdc6 ATP hydrolysis is also important for release of Mcm2-7 helicases that fail to load although the biological importance of this activity remains unclear (Frigola et al., 2013; Kang et al., 2014; Coster et al., 2014).

Cdt1

In budding yeast, Cdt1 forms a tight complex with Mcm2-7 and is required for the
association of the helicase with the origin/ORC/Cdc6 complex (Figure 4; Tanaka and Diffley, 2002). The highly conserved C-terminus of Cdt1 is important for interacting with Mcm2-7 (Jee et al., 2010; Khayrutdinov et al., 2009). In *Xenopus* and mouse cells, Cdt1 interacts with Mcm2 and Mcm6 and this interaction has also been shown structurally in *S. cerevisiae* (Ferenbach et al., 2005; Yanagi et al., 2002; Sun et al., 2013). Cdt1 is also thought to stabilize the Mcm2-7 helicase ring (Frigola et al., 2013). In addition to its role in stabilizing the Mcm2-7 hexamer and mediating association with ORC/Cdc6, Cdt1 could play a role in opening the Mcm2-7 ring, similar to DnaC.

**THE REPLICATIVE HELICASE**

*E. coli: DnaB*

The *E. coli* replicative helicase is a homohexameric ring called DnaB (Rehaku-Krantz and Hurwitz, 1978; Arai et al., 1981). Unlike the AAA+ ATPase domains of DnaA, DnaC, ORC, Cdc6 and Mcm2-7, DnaB contains a RecA-like ATPase domain (Bailey et al. 2007). The RecA-like family of ATPases also contains conserved Walker A, Walker B and the arginine finger motif, but have additional conserved features that differentiate them the AAA+ ATPases (Erzberger and Berger, 2006). Helicase loading yields two DnaB helicases encircling the two ssDNA strands in opposite orientations. Each of these two DnaB helicases move in a 5’ to 3’ direction on the lagging strand to unwind DNA (LeBowitz and McMacken, 1986; Lee et al., 1989; Kaplan, 2000).

Extensive structural studies have analyzed the state of DnaB under different
conditions. The DnaB hexamer forms a two-lobed structure with the C-terminal ATPase domains and the N-terminal ‘collar’ forming two stacked rings (Figure 5A; Bailey et al., 2007). DnaB exists in at least two different structural states: a symmetrical hexamer with 6-fold symmetry and as a trimer of dimers (San Martin et al., 1995; Yu et al., 1996). In all of these studies, DnaB is a closed ring but in the presence of DnaC, DnaB adopts a cracked ring conformation (Arias-Palomo et al., 2013). Interestingly, DnaB has been proposed to translocate on ssDNA in its cracked-ring conformation. In this model, movement along the ssDNA occurs as individual subunits move from the bottom to the top of the crack in the ring using a hand-over-hand translocation mechanism (Figure 5B; Itsathitphaisarn et al., 2012).

*S. cerevisiae: Mcm2-7*

The heterohexameric Mcm2-7 complex is the eukaryotic replicative helicase. Like DnaB, Mcm2-7 is at the core of both replication initiation and elongation (Schwacha and Bell, 2001; Labib et al., 2000; Aparicio et al. 1997). All six subunits are members of the AAA+ ATPase family of proteins and are essential (Duderstadt and Berger, 2008). Interestingly, although the AAA+ motifs of each of the Mcm subunits are very similar, each of the six Mcm2-7 subunits has distinguishing motifs that are conserved in all eukaryotes (reviewed in Bochman and Schwacha, 2009), suggesting that each subunit has specific and unique roles.

Each Mcm subunit contains three folded domains: a globular Domain A near or at the N-terminus; an adjacent oligonucleotide/oligosaccharide-binding (OB) fold domain;
Figure 5

(A) Structure of DnaB (Bailey et al., 2007).
(B) Structure of translocating DnaB in the presence of ssDNA and GDP-AlF₄ (Itsathitphaisarn et al., 2012).
and an AAA+ ATPase domain at the C-terminus. Similar to DnaB, the structure of Mcm2-7 is made up of two stacked rings, with the AAA+ domains forming one ring, and Domain A and the OB-fold forming a second ring (Figure 6A; Remus et al., 2009). The Mcm2-7 hexamer has a channel running through the middle of both helicases that can accommodate dsDNA (Figure 6B). For this final form of the helicase to be achieved, each hexamer must be opened, dsDNA placed inside the central channel, and then closed stably around DNA. The Mcm subunits are arranged in a defined order of the Mcm subunits in the helicase assembly: Mcm2-Mcm6-Mcm4-Mcm7-Mcm3-Mcm5. Biochemical and structural studies have identified a discontinuity between Mcm2 and Mcm5 (Mcm2-5 gate) that was later shown to be the entry site for DNA into the central channel (Figure 6C; Costa et al., 2011; Bochman et al., 2008; Sun et al., 2013).

Like other AAA+ ATPases, each subunit interface of the Mcm2-7 complex forms an ATP binding and hydrolysis domain, each of which has been shown to be essential for yeast viability (Schwacha and Bell, 2001). Unlike ORC, all six ATP interfaces are functional, although each ATP binding and hydrolysis site has distinct ATPase activities (Davey et al., 2003; Bochman et al., 2008).

Mcm2-7 ATPase activity is required for multiple events during replication with the most fundamental being the unwinding of DNA. Although the Mcm2-7 complex shows very low levels on its own (Bochman and Schwacha, 2008), binding of two helicase-activating proteins, Cdc45 and GINS, dramatically stimulate Mcm2-7 ATPase and helicase activity (Moyer et al., 2006). This Cdc45-Mcm2-7-GINS (CMG) complex is
Figure 6

(A) Structure showing the architecture of each hexamer within the double hexamer where the N-terminal domains form a ring, and interact between the two helicases (Remus et al., 2009).

(B) Same EM structure as in (A) showing the central channel of the double hexamer (Remus et al., 2009).

(C) EM of the Drosophila Mcm2-7 complex shows a lock-washer form of the helicase. This view is from the AAA+ domains of the Mcm subunits (Costa et al., 2011).
the active helicase *in vivo*. ATP hydrolysis is required for unwinding, as ATPγS does not allow unwinding. Interestingly, only a subset of the Mcm2-7 ATP binding and hydrolysis domains inhibits its ability to unwind (Bochman and Schwacha, 2008), suggesting the remaining ATPase motifs perform other functions.

Mcm2-7 ATP binding and hydrolysis have also been implicated in helicase loading and helicase activation (Kang et al., 2014; Coster et al., 2014). These studies analyzed mutations in either the Walker A or arginine-finger motifs for each of the ATP hydrolysis interfaces (Kang et al., 2014; Coster et al., 2014). All the mutants showed defects in helicase loading but to differing extents (Kang et al., 2014; Coster et al., 2014). Interestingly, making a Walker A or an arginine-finger mutation at the same ATP hydrolysis interface did not yield the same defect (Kang et al., 2014; Coster et al., 2014). The subset of ATPase mutant complexes that showed substantial helicase loading, were analyzed for functions in helicase activation. Several mutant complexes showed reduced DNA retention during helicase loading and others were unable to associate with Cdc45 and/or GINS (Kang et al., 2014)

The Mcm2-7 helicase is loaded as an inactive double hexamer that encircles dsDNA (Remus et al., 2009) but the active form is a single Mcm2-7 (as part of the CMG) that encircles ssDNA (Fu et al., 2011). Therefore, the Mcm2-7 helicase must be opened both during initial loading around double-stranded origin DNA, but also during activation of the helicase to allow the lagging strand template to escape the central Mcm2-7 channel. Both of these ring-opening events are thought to occur at the discontinuity
between the Mcm2 and Mcm5 subunits (the Mcm2-5 gate). The presence of the Mcm2-5 gate was first hypothesized based on pairwise association studies of the Mcm subunits as well as the ATPase activities of each of these pairs (Davey et al., 2003). The presence of the Mcm2-5 gate was further supported by experiments that linked the ATPase activity of specific Mcm interfaces to gate opening and closing by analyzing the affinity of mutant complexes to circular dsDNA (Bochman and Schwacha, 2008; Bochman and Schwacha, 2010). Structural studies showed that the *Drosophila* Mcm2-7 hexamer contained a gap between Mcm2 and Mcm5 (Costa et al., 2011), and this same Mcm2-5 gap is important for helicase loading (Samel et al., 2014).

What triggers the opening and closing of the helicase is unknown, but the final loaded eukaryotic double hexamer contains two closed gates (Remus et al., 2009). This final state indicates that during the loading process, the gate must be closed in a regulated fashion, after dsDNA has been placed in its central channel. The state of the Mcm2-5 gate in solution has been proposed to be in an open state in *Drosophila* (Costa et al., 2011) but in a closed state in *S. cerevisiae* (Samel et al., 2014). Further investigation into the steps of helicase loading is needed to understand how gate opening and closing is regulated.

**HELICASE LOADING: TWO HELICASES, OPPOSITE ORIENTATIONS**

Hlicase loading in both *S. cerevisiae* and *E. coli* results in two helicases encircling DNA, each in an opposite orientation. In *S. cerevisiae* the two loaded helicases
form a head-to-head double hexamer that encircles dsDNA (Remus et al., 2009; Evrin et al., 2009; Li et al., 2015). This is different from E. coli, where the two DnaB hexamers are separate and encircle ssDNA. Both helicases are believed to unwind dsDNA through a strand exclusion mechanism, where one strand of DNA passes through the central channel of the helicase and the other strand is excluded from the central channel (Hacker and Johnson, 1997; Fu et al., 2011). Therefore, during helicase activation the eukaryotic Mcm2-7 double hexamer and the origin DNA must be remodeled to separate the two helicases and extrude the lagging strand ssDNA template from the central channel of each Mcm2-7 (Figure 7). In both E. coli and S. cerevisiae helicase loading is an ordered stepwise process.

**E. coli**

The steps of helicase loading in E. coli involve the coordinate action of DnaA, DnaB and DnaC (Figure 8; Yardimci and Walter, 2014). Initial DNA binding (of the dsDNA 9-mers) and oligomerization of DnaA drives the initial melting of the origin DNA at the DUE (Figure 8a). The DnaA oligomers then associate with one strand of the bubble (top strand; Funnell et al., 1987; Speck and Messer, 2001). This strand specific binding of DnaA likely plays a role in ensuring the directionality of the loaded helicases. The helicase loader, DnaC binds the C-terminal ends of DnaB stabilizing the open form of DnaB. Current models suggest that the DnaB-DnaC complex is recruited to oriC through two distinct interactions (Figure 8b,c). Subsequently, ATP hydrolysis by DnaC is required for activation of DnaB (Figure 8d; Davey et al. 2002) and is likely triggered by primer synthesis (Makowska-Grzyska and Kaguni, 2010).

Figure 7. Transition between the inactive double hexamer and the activated CMG.
The Mcm2-7 double hexamer and the dsDNA must be remodeled in two ways: (i) the Mcm2-7 helciases must be separated, and (ii) the lagging strand must be extruded from the central channel of each Mcm2-7.
Figure 8. The mechanism of helicase loading in *E. coli*.
Summarized steps for helicase loading in *E. coli* (Yardimci and Walter, 2014).
The exact mechanism of how the two DnaB helicases are loaded, each in an opposite orientation and on a different strand of DNA remains unknown. Different mechanisms of recruitment have been proposed, wherein one helicase is recruited through DnaA/DnaB interactions and the other through DnaA/DnaC interactions (Figure 8d; Mott et al., 2008). Such a proposal could explain how cells ensure helicases are loaded in the correct orientation and on the right strand of DNA. However, how the DnaA-DnaB mechanism ensures loading on the proper strand remains unknown. Additionally this model supposes that different mechanisms load each of the two helicases.

*S. cerevisiae*

Helicase loading in *S. cerevisiae* has been reconstituted *in vitro* using purified proteins in bulk (Bowers et al., 2004; Seki and Diffley, 2000). Origin-containing dsDNA is coupled to magnetic beads, then incubated with purified ORC, Cdc6, Cdt1 and Mcm2-7. The reaction requires ATP and an origin. After incubation, washes are performed to eliminate unbound proteins. A high-salt wash can be used to remove all associated proteins except for loaded Mcm2-7 helicases. Bound proteins are then run on a gel and visualized. This reaction has been used to uncover many aspects of helicase loading.

A stepwise association process leads to loading of the *S. cerevisiae* Mcm2-7 helicase (Figure 9). First, ORC binds origins of replication in the absence of other proteins, and directly recruits Cdc6 to these sites (reviewed in Yardimci and Walter, 2014). Following Cdc6 association, Mcm2-7 in a complex with Cdt1 is recruited. In the absence of Cdt1, the Mcm2-7 helicase does not associate with ORC/Cdc6 (Tanaka and...
Figure 9. Steps of helicase loading
See text for details.
Diffley, 2002). These initial association steps do not require any ATP hydrolysis and can occur in the presence of ATPγS, a slowly-hydrolyzable ATP analog. An EM structure of this stable intermediate provides insight into how the first Mcm2-7 is recruited to the origin DNA. This complex contains ORC/Cdc6 adjacent to Cdt1/Mcm2-7 (OC₆C₁M) and both complexes are thought to encircle adjacent dsDNA (Figure 4; Sun et al., 2013). In this assembly the C-terminal ends of ORC/Cdc6 interact with the C-terminal ends of the Mcm2-7 subunits (Sun et al., 2013; Bleichert et al., 2015). Subsequent steps in the helicase loading reaction require ATP hydrolysis and result in the formation of a N-terminal-to-N-terminal Mcm2-7 double hexamer (Figure 6A; Remus et al., 2009; Li et al., 2015). The events that occur between formation of the initial OC₆C₁M complex and the formation of the Mcm2-7 double hexamer were largely unknown when the studies described in this thesis were initiated.

Several mechanisms have been proposed for how cells load two Mcm2-7 hexamers in opposite orientations (Figure 10; Yardimei and Walter, 2014). I will highlight a representative subset of these and underscore the potential pitfalls/advantages of each. It is possible that the loading of the two helicases occurs in a concerted manner (Figure 10a). This would mean that both helicases are first recruited and subsequently loaded simultaneously. A concerted model would ensure that exactly two helicases are loaded, in opposite orientations. The concerted model would require prior formation of a Mcm2-7 double hexamer in solution and this has not been observed. Additionally the gates of the two helicases would need to open and close simultaneously, requiring that the two helicases be aligned and this is not true after double hexamer formation (Li et al.,
Figure 10. Possible mechanisms for helicase loading in *S. cerevisiae*. Multiple proposed models for the mechanism of helicase loading in budding yeast (Yardimci and Walter, 2014).
Another model would involve two separate ORC complexes, each bound in opposite orientations (Figure 10d). Each of these ORC complexes would recruit distinct Cdc6 and Cdt1/Mcm2-7 molecules and load each independently. Because origins do not contain two ACS regions to enable two ORC complexes to bind, the B2 region has been proposed to serve as an inverted ORC binding site. The B2 region is not essential however, and studies have shown that it does not have to resemble the ACS (Chang et al., 2011). Additionally, the spacing between the ACS and B2 region would not allow simultaneous binding of two ORC complexes (Wilmes et al., 2002; Bell and Stillman, 1992). A last model would be that, similar to the current model in E. coli, the two helicases are recruited and loaded by distinct interactions (Figure 10e). The first Mcm2-7 helicase would be loaded by its interactions with ORC/Cdc6 whereas the second Mcm2-7 could either be recruited and loaded by the interactions with the first Mcm2-7 or a different interaction with ORC.

REGULATION OF DNA REPLICATION

Initiation of DNA replication is the first and most tightly controlled step of DNA synthesis. Both prokaryotic and eukaryotic cells have mechanisms in place to ensure replication only occurs once per cell cycle. Whereas regulation in E. coli primarily occurs at the level of DnaA binding to oriC, regulation in S. cerevisiae focuses on inhibition of helicase loading outside of the G1 phase of the cell cycle.
In *E. coli*, multiple mechanisms inhibit DnaA binding to oriC. This binding is regulated by whether DnaA is bound to ADP or ATP (McGarry et al., 2004; Speck et al., 1999). Immediately prior to the initiation of DNA synthesis, levels of DnaA-ATP rise abruptly (Kurokawa et al., 1999). Whereas ATP-DnaA can bind to high- and low-affinity OriC sites, ADP-DnaA only binds the high-affinity sites (Grimwade et al., 2000; Grimwade et al., 2007; Ryan et al., 2004; Margulies and Kaguni, 1996). It has been proposed that the nucleotide state of DnaA controls initiation of replication by controlling oligomerization of DnaA (Speck et al., 1999; Speck and Messer, 2001; Erzberger et al., 2002; Erzberger et al., 2006).

Another approach used to regulate DnaA binding to oriC, is by limiting access of DnaA to oriC. Between cycles of initiation, SeqA binds hemimethylated GATC motifs that are found in many copies at oriC (Lu et al., 1994; Nievera et al., 2006). Although oriC is fully methylated prior to replication initiation, the hemimethylated post-replication DNA state recruits SeqA. This binding sequesters oriC away from DnaA (Boye et al., 1996; Nievera et al., 2006; Slater et al., 1995). SeqA binding also inhibits methylation of the bound GATC by the Dam methyltransferase (Lu et al. 1994; Boye et al. 1996; Guarne et al. 2002; Fujikawa et al. 2004). Nevertheless, as the cell cycle progresses these sites do become methylated, relieving SeqA inhibition of DnaA binding (Kang et al., 1999). Mutants in the methyltransferase system rereplicate their genomes (Lu et al., 1994; Skarstad and Løbner-Olesen, 2003).
*S. cerevisiae*

In *S. cerevisiae*, the main mechanism cells use to prevent rereplication is the temporal separation of helicase loading and helicase activation (Diffley, 2004; Arias and Walter, 2007). Helicase loading only occurs in late M/G1 phase, whereas helicase activation only occurs in S phase. The central player in this temporal separation of helicase loading and helicase activation is the S phase cyclin-dependent kinase (S-CDK; Dahmann et al., 1995). S-CDK inhibits helicase loading by phosphorylating multiple proteins involved in helicase loading (reviewed in Kelly and Brown, 2000).

S-CDK directly interacts and phosphorylates ORC, Cdc6, and Mcm2-7. Orc2 and Orc6 are phosphorylated by S-CDK and Orc6 is further bound by Clb5, inhibiting association of the Mcm2-7 complex (Nguyen et al., 2001; Wilmes et al., 2004). S-CDK phosphorylates Cdc6 and reduces its protein levels by promoting ubiquitin-mediated degradation of Cdc6 (Drury et al., 1997; Drury et al., 2000; Elsasser et al., 1999). Mcm2-7 phosphorylation by S-CDK promotes its nuclear export leading to their exclusion from the nucleus in G2 and M phases (Nguyen et al., 2000; Labib et al., 1999). Although it is tempting to think that these mechanisms are redundant, experiments in which two of the three are inhibited show evidence of localized genome re-replication (Green et al., 2006).

Although metazoans use the same temporal separation of helicase loading and activation to ensure the genome is replicated just once per cell cycle, the mechanisms involved differ. For example, in *Xenopus*, ORC is released from chromatin upon DNA replication initiation (Sun et al., 2002), whereas human ORC is disassembled upon entry
into S phase as a consequence of Orc1 degradation (Mendez et al., 2002; Tatsumi et al., 2003). Cdc6, instead of being degraded, is transported out of the nucleus upon S-CDK phosphorylation in both *Xenopus* and humans (Petersen et al., 1999; Coverley et al., 2000; Pelizon et al., 2000; Delmolino et al., 2001). Metazoan Cdt1 is regulated by multiple pathways. Chromatin-bound Cdt1 is targeted for degradation (Arias and Walter, 2005; Li and Blow, 2005), and there is a secondary pathway that promotes Cdt1 degradation throughout the cell cycle (Li et al., 2003; Nishitani et al., 2006). An additional protein, geminin, binds directly to Cdt1 and modulates both the stability and activity of Cdt1 (Wohlshlegel et al., 2000; Tada et al., 2001). The finding that metazoans also use multiple mechanisms to restrict helicase loading to G1 emphasizes the critical nature of this regulation.

**SINGLE-MOLECULE STUDIES OF DNA REPLICATION**

Single-molecule approaches offer many advantages to ensemble biochemical assays in the study of helicase loading. Ensemble loading reaction require multiple washes to remove proteins that are not associated with the DNA. Because these washes take minutes to execute, only events that last longer than this are detectable in bulk assays. In contrast, single molecule assays allow observation of both dynamic and unstable events that occur for as little as one second. Second, single-molecule assays allow determination of the number of proteins associated with a single DNA molecule. In contrast, bulk assays can only measure the average number of molecules associated per DNA molecule. The resulting measurement could result from a mixture of different
stoichiometries and could also include incomplete reactions. Finally, the single-molecule approach offers the opportunity to synchronize events post hoc. The bulk assay is asynchronous, largely preventing temporal resolution of events. In contrast, using the single-molecule assay we can synchronize events relative to one another to resolve them at the one-second time scale.

Single-molecule approaches have been used to investigate many aspects of DNA replication (reviewed in Stratmann and van Oijen, 2014). Very early single-molecule EM and autoradiography studies discovered that DNA replication was bidirectional and coordinated with quick nucleosome deposition (Huberman and Riggs, 1968). Although EM studies continue to offer valuable information (Sun et al., 2013; Li et al., 2015), newer single-molecule assays are providing information about the forces, kinetics and dynamics of replication (reviewed in Stratmann and van Oijen, 2014). Single-molecule studies of helicases during unwinding reveal multiple states of the enzymatic complex during unwinding. Additionally, fluorescent assays using tethered DNA offer insights into the dynamics and number of DNA polymerases acting at the replication fork.

In addition to DNA replication, single-molecule approaches have been used to study many fundamental cellular processes. Single-molecule approaches have been used to dissect the order and kinetics of transcription and splicing, and the mechanism of dynein motor movement on microtubules. These studies laid the groundwork for the study of the kinetics and protein-DNA stoichiometry of the helicase-loading reaction described in Chapter 2.
THESIS SUMMARY

Prior to my thesis work, the proteins required for helicase loading and their initial order of association was known. In addition, structural studies suggested a mechanism to recruit the first Mcm2-7 to the DNA. The stoichiometry of the proteins required, the dynamics of their binding and dissociation and how they coordinate loading of two helicases in a tightly interacting head-to-head conformation remained unclear. This thesis focuses on the steps and mechanism of the helicase reaction after the initial association of ORC/Cdc6/Cdt1 and Mcm2-7. Whereas these initial associations do not require ATPase activity, the subsequent steps I have analyzed require ATP hydrolysis.

In this thesis, I describe the development of a novel single-molecule approach to investigate eukaryotic helicase loading. The single-molecule assays described here revealed multiple previously unappreciated steps during helicase loading. Additionally, the number of ORC, Cdt1 and Cdc6 molecules required to load a double hexamer are investigated.
REFERENCES


Bleichert, F., Balasov, M. Chesnokov, I., Nogales, E., Botchan, M.R., and Berger, J.M.


Bochman, M.L., and Schwacha, A. (2010). The Saccharomyces cerevisiae Mcm6/2 and Mcm5/3 ATPase active sites contribute to the function of the putative Mcm2-7 'gate'. Nucleic Acids Res. 38, 6078-6088.


subgroups are essential for coordinated ATP hydrolysis and DNA replication.


oligomerization at the E. coli chromosomal origin is required for initiation and

Skarstad, K., and Løbner-Olesen, A., (2003). Stable co-existence of separate replicons in
*Escherichia coli* is dependent on once-per-cell-cycle initiation. EMBO J. 22, 140–
150.

SeqA protein binds *oriC* in two different methylmodulated reactions appropriate
to its roles in DNA replication initiation and origin sequestration. Cell 82, 927–
936.

switch in gene regulation. EMBO J. 18, 6169–6176.

DnaA to double- and single-stranded DNA. EMBO J. 20, 1469-1476.

ORC-Cdc6 stability and the selection of specific DNA sequences as origins of


Stroud, H., Otero, S., Desvoyes, B., Ramirez-Parra, E., Jacobsen, S.E., and Gutierrez, C.
(2012). Genome-wide analysis of histone H3.1 and H3.3 variants in *Arabidopsis

Sun, J., Evrin, C., Samel, S.A., Fernández-Cid, A., Riera, A., Kawakami, H., Stillman,
intermediate containing ORC-Cdc6-Cdt1-MCM2-7 bound to DNA. Nat. Struct. Mol. Biol. 20, 944–951.


Chapter II

Single-molecule Studies of Origin Licensing

Reveal Mechanisms Ensuring Bidirectional Helicase Loading

An earlier version of this chapter was published in 2015 under the same title (Cell 161:513-525). The authors were Simina Ticau, Larry J. Friedman, Nikola Ivica, Jeff Gelles and Stephen P. Bell. S.T. designed and conducted experiments with feedback from L.J.F., J.G. and S.P.B. S.T. and L.J.F. analyzed data. N.A.I developed labeling strategies and S.T., and N.A.I. generated proteins.

I would like to thank members of the Bell laboratory for useful discussions. We thank Jonathan B. Grimm and Luke Lavis for graciously providing Janelia Fluors.
SUMMARY

Loading of the ring-shaped Mcm2-7 replicative helicase around DNA licenses eukaryotic origins of replication. During loading, Cdc6, Cdt1 and the origin-recognition complex (ORC) assemble two heterohexameric Mcm2-7 complexes into a head-to-head double hexamer that facilitates bidirectional replication initiation. Using multi-wavelength single-molecule fluorescence to monitor the events of helicase loading, we demonstrate that double-hexamer formation is the result of sequential loading of individual Mcm2-7 complexes. Loading of each Mcm2-7 molecule involves the ordered association and dissociation of distinct Cdc6 and Cdt1 proteins. In contrast, one ORC molecule directs loading of both helicases in each double hexamer. Based on single-molecule FRET, arrival of the second Mcm2-7 results in rapid double-hexamer formation that anticipates Cdc6 and Cdt1 release, suggesting Mcm-Mcm interactions recruit the second helicase. Our findings reveal the complex protein dynamics that coordinate helicase loading and indicate that distinct mechanisms load the oppositely oriented helicases that are central to bidirectional replication initiation.
Eukaryotic DNA replication must occur faithfully each cell cycle to maintain genomic stability. Initiation of replication occurs at genomic sites called origins. To ensure that no origin initiates replication more than once per cell cycle, the cell restricts the DNA loading and activation of the Mcm2-7 replicative helicase to distinct cell cycle stages (Siddiqui et al., 2013). Importantly, helicase loading (also known as pre-RC formation) licenses origins of replication by establishing the correct architecture for helicase activation and bidirectional replication initiation.

Three helicase-loading proteins direct Mcm2-7 loading: the origin recognition complex (ORC), Cdc6 and Cdt1 (reviewed in Yardimci and Walter, 2014). ORC binds origins of replication and recruits Cdc6 at the M/G1 transition. Cdc6-bound ORC recruits Mcm2-7 in complex with Cdt1 to origin DNA. In an ATP-hydrolysis-dependent reaction, recruited Mcm2-7 complexes are loaded around the origin DNA (Coster et al., 2014; Kang et al., 2014). Helicase loading requires opening and closing of the toroidal Mcm2-7 ring between the Mcm2 and Mcm5 subunits (Bochman and Schwacha, 2008; Costa et al., 2011; Samel et al., 2014). The product of helicase loading is a pair of tightly interacting Mcm2-7 complexes that encircle the double-stranded origin DNA in a head-to-head conformation, with staggered Mcm2/5 gates (Costa et al., 2014; Evrin et al., 2009; Remus et al., 2009; Sun et al., 2014).

Although the structure of the double-hexamer product of helicase loading is clear,
important questions remain about how the helicase-loading proteins achieve this outcome. In particular, the mechanisms that load the first and second Mcm2-7 complex in opposite orientations are unclear (reviewed in Yardimci and Walter, 2014). Do the two Mcm2-7 complexes associate and load simultaneously or in an ordered fashion? Do the same or different ORC and Cdc6 proteins load each Mcm2-7 complex? To address these questions we have developed single-molecule assays to monitor helicase loading.

Single-molecule studies are a powerful tool to address questions of stoichiometry and dynamics during DNA replication events. Studies of this type have led to important insights including the dynamics and number of DNA polymerases acting at the replication fork (reviewed in Stratmann and van Oijen, 2014). Extending these approaches to replication initiation has the potential for additional discovery. Unlike current ensemble helicase loading assays, which can only detect events that survive multiple washes, single-molecule approaches readily detect short-lived interactions during cycles of enzymatic function. Single-molecule approaches also allow stoichiometric determinations that are difficult with ensemble helicase loading assays due to DNA-to-DNA asynchrony and heterogeneity. Finally, although multi-step reactions are frequently asynchronous, post-hoc synchronization of single-molecule data allows precise kinetic analysis of reaction pathways.

We have developed single-molecule assays that monitor the DNA association of eukaryotic helicase-loading proteins using colocalization single-molecule spectroscopy (CoSMoS) (Friedman et al., 2006; Hoskins et al., 2011). We show that the two Mcm2-7
hexamers are recruited and loaded in separate events that require distinct Cdc6 and Cdt1 molecules. In contrast, one ORC molecule directs loading of both Mcm2-7 complexes present in a double hexamer. Consistent with distinct mechanisms loading the two hexamers, we observe kinetic differences between events associated with loading the first and second helicase. By combining CoSMoS with fluorescence resonance energy transfer (FRET), we demonstrate that formation of the Mcm2-7 double-hexamer interface precedes dissociation of Cdc6 and Cdt1, suggesting interactions with the first Mcm2-7, rather than ORC, drive recruitment of the second helicase. Our observations reveal both the complex protein coordination required to assemble Mcm2-7 double hexamers and the mechanisms ensuring the two Mcm2-7 molecules are loaded in the opposite orientations required for bidirectional replication initiation.

RESULTS

A single-molecule assay for helicase loading

To develop a single-molecule assay for eukaryotic helicase loading we used CoSMoS to monitor the origin-DNA association of the proteins required for this process (ORC, Cdc6, Cdt1, Mcm2-7). First, we immobilized origin-containing DNA by coupling it to microscope slides. We determined the location of surface-attached DNA on the slide using a DNA-coupled fluorophore (Fig. 1A). We monitored associations of one or two proteins (labeled with distinguishable fluorophores) with origin DNA using colocalization of the protein- and DNA-associated fluorophores (Fig. 2A). Fluorescent labeling of ORC, Cdc6, Cdt1 and Mcm2-7 was accomplished using a SNAP-tag or
Figure 1. Mcm2-7 hexamers associate with and are loaded on DNA in a one-at-a-time manner.

(A) Schematic for the single-molecule helicase-loading assay. Alexa488-labeled (blue circle) 1.3kb origin-DNAs were coupled to microscope slides. Purified ORC, Cdc6, and Cdt1/Mcm2-7 (at least one fluorescently labeled, Mcm2-7 in this illustration) were incubated with slide-coupled DNA and colocalization of the fluorescently-labeled protein with the DNA was monitored.

(B) Mcm2-7 complexes sequentially associate with origin DNA. Plots display the Mcm2-74SNAP549 fluorescence intensity recorded at two representative DNA molecules. Insets
show fluorescence images (4 x 1 s) taken during the sequential association of first (red arrow) and second (blue arrow) Mcm2-7.

(C) Mcm2-7 dwell times on DNA have a multiexponential distribution. Mcm2-7 dwell times were plotted as a histogram. Combined data from first and second Mcm2-7 associations are included; vertical axis represents the number of dwells of the specified duration per s per DNA molecule. Red bars are results from a separate experiment using mutant origin DNA. Inset shows the distribution of Mcm2-7 dwell times on DNA molecules as a semilogarithmic cumulative survival plot; only a portion of the entire plot is shown to emphasize that the distribution has at least two exponential components.

(D) Mcm2-7 associates with DNA one at a time. The number of associations present at standard protein concentrations before a reaction- (top) or high-salt-buffer (0.5 M NaCl; bottom) wash is compared to the number of fluorophores that are detected by photobleaching immediately after the wash.

(E) Two representative traces before and after a high-salt wash and photobleaching. Reactions were washed twice with a high-salt buffer and imaged at higher laser power in the absence of an oxygen scavenging system until all fluorophores were photobleached. Traces of Mcm2-7^SNAP549 associations during the reaction (green) are plotted adjacent to photobleaching steps after a high-salt wash (purple).
Figure 2. Mcm2-7 hexamers associate with and are loaded on DNA in a one-at-a-time manner.

(A) Whole microscope field of view showing AlexaFluor488-coupled DNA (right) and the Mcm2-7SNAP549 fluorescent spots from the same experiment at time= 600s. The three red arrows represent DNA/Mcm2-7SNAP549 colocalized pairs for which the fluorescent intensity traces are shown in S1C.
(B) Activity of fluorescently labeled proteins relative to the unlabeled proteins. Bulk helicase-loading assays were performed using unlabeled or fluorescently-labeled Mcm2-7, Cdt1/Mcm2-7, Cdc6, and ORC. All samples were subjected to a high salt wash, separated by SDS-PAGE, stained with Krypton protein stain and quantified. Reactions were performed in duplicate and error bars indicate the standard deviation.

(C) Additional fluorescence intensity traces (from the experiment shown in S1A, see red arrows) for DNA-associated Mcm2-7^SNAP549 over time showing that Mcm2-7 complexes sequentially associate with origin DNA. First and second Mcm2-7 associations are marked with red and blue arrows, respectively.

(D) Mcm2-7^SNAP549 complexes were photobleached after a high-salt wash, and the number of high-salt resistant hexamers per 100 DNA molecules was calculated. We performed this experiment at standard (0.25nM ORC, 1nM Cdc6, 2.5nM Cdt1/Mcm2-7; N_{DNA}=523) and three-fold higher concentrations of each of the proteins. The increased concentration experiment was performed with distinguishably-labeled wild-type (N_{DNA}=449) and mutated ORC binding site (the ARS consensus sequence; N_{DNA}=315) DNA molecules present in the same reaction chamber.

(E) Mcm2-7 complexes are loaded one at a time. Double-labeled Mcm2-7^SNAP549^SORT649 were used in a standard single-molecule helicase-loading assay. This reaction was not imaged during the standard 20 minutes to prevent photobleaching during the reaction. Instead, after the 20-minute helicase-loading reaction, the chamber was washed with high-salt buffer. Then, multiple fields of view were imaged under photobleaching conditions (Fig. 1E) to determine the number of associated fluorophores for each DNA molecule. A total of 540 DNA molecules with one or more Mcm2-7-associated fluorophores were analyzed. The images in the top panels illustrate the combinations of labeled and unlabeled Mcm2-7 molecules that would be scored as having one molecule of either or both fluorophores (left column) or two copies of one or both fluorophores (right column). The rows indicate the type of labeled molecules expected for a model in which both Mcm2-7 single and double hexamers are loaded/high-salt resistant (top) or only double hexamers are loaded (bottom). White circles represent unlabeled or photobleached subunits. The observed data (grey bars) is separated into the two categories described above and plotted with the expected numbers for each of two Models overlayed as single points on top of the observed data. The expected data for each Model was calculated using the separately measured labeling efficiencies for the two Mcm2-7 subunits: Mcm4-SNAP: 79% and Mcm7-Sortase: 77%.
sortase-mediated coupling of fluorescent peptides (Gendreizig et al., 2003; Popp et al., 2007). In each case, the fluorescent tags did not interfere with protein function in ensemble helicase-loading reactions (Fig. 2B). After imaging the locations of slide-coupled DNA molecules, purified ORC, Cdc6, and Cdt1/Mcm2-7 were added (one or two of which were fluorescently labeled) and the location of each DNA molecule was continuously monitored for labeled protein colocalization in one-second intervals for 20 minutes.

Multiple observations indicated that Mcm2-7-DNA colocalizations represented events of helicase loading (Table 1). First, colocalizations of Mcm2-7 with the DNA was dramatically reduced in the absence of ORC or Cdc6, proteins required for helicase loading (Yardimci and Walter, 2014). Second, stable association (>30 s) of Mcm2-7 was dependent on the presence of the ORC DNA binding site (the ARS-consensus sequence, ACS). Third, ORC, Cdc6, origin DNA and ATP hydrolysis were each required to form Mcm2-7 molecules that were resistant to a high-salt wash (Table 1), a biochemical test for loaded helicases encircling dsDNA independently of helicase-loading proteins (Donovan et al., 1997; Randell et al., 2006).

**Mcm2-7 association and loading occurs in a one-at-a-time manner**

Our initial studies monitored Mcm2-7 association with origin DNA. We performed CoSMoS helicase-loading experiments using Mcm2-7 containing SNAP-tagged Mcm4 labeled with 549 fluorophore (Mcm2-7^{SNAP549}, Fig. 1) and unlabeled ORC, Cdc6 and Cdt1. Over the course of 20 minutes we observed both single- and double-
Table 1. The single-molecule helicase-loading assay, is origin-, ORC-, Cdc6-, and ATP-dependent.

<table>
<thead>
<tr>
<th></th>
<th>ATP</th>
<th>ATPγS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+origin</td>
<td>-origin</td>
</tr>
<tr>
<td></td>
<td>+Cdc6</td>
<td>+Cdc6</td>
</tr>
<tr>
<td></td>
<td>+ORC</td>
<td>-ORC</td>
</tr>
</tbody>
</table>

**Mcm2-7 binding frequency**

| Mcm2-7 with dwell times less than 30 s (per 100 DNAs per hour) | 282 ± 13 | 133 ± 11 | 0.8 ± 0.8 | 13 ± 3 | 100 ± 19 |

| Mcm2-7 with dwell times greater than 30 s (per 100 DNAs per hour) | 209 ± 11 | 2 ± 2    | 0.8 ± 0.8 | 0      | 208 ± 15 |

**Number of loaded Mcm2-7**

| Salt-resistant Mcm2-7 (per 100 DNAs) | 21 ± 1 | 0      | 0       | 0      | 1 ± 1   |

The number of Mcm2-7 associations during a 20-minute reaction are categorized into short (<30 s) and long (>30 s) dwell times and reported as a mean binding frequency ± s.e.m. The nucleotide used and the presence or absence of ORC, Cdc6 and wild-type origin DNA are indicated (-origin DNA contains point mutations in the ACS). After 20 minutes, the reactions were subjected to a high-salt wash and the associated Mcm2-7 complexes were counted by photobleaching. Numbers are reported per 100 DNA molecules, to allow comparison between experiments that have different number of DNA molecules coupled to the slide (N_{DNA} ranged from 278 to 523). Due to decreased protein-DNA associations at lower concentrations, the ATPγS experiment was performed at 25nM ORC/Cdc6/Cdt1/Mcm2-7 compared to 0.25nM ORC/1nM Cdc6/2.5nM Cdt1/Mcm2-7 for ATP-containing experiments.
stepwise increases in Mcm2-7-associated fluorescence intensity at origin DNAs (Fig. 1B, 2C). Mcm2-7 dwell time distributions were multi-exponential with many short-lived (<30 s) and fewer longer-lived (>30 s) relative increases in fluorescent intensity, suggesting at least two distinct types of Mcm2-7 association with the DNA (Fig. 1C).

There are two possible explanations for the multiple stepwise increases in DNA-colocalized Mcm2-7-coupled fluorescence. The simplest interpretation of this data is that Mcm2-7 hexamers associate with origin-DNA in a one-at-a-time manner, with multiple hexamers accumulating over time. Alternatively, it was possible that each increase in fluorescence was due to the simultaneous association of a Mcm2-7 multimer (e.g. a pre-formed dimer of two Mcm2-7 hexamers). To distinguish between these possibilities, we used photobleaching to count the number of DNA-associated Mcm2-7 hexamers. To this end, we first observed Mcm2-74SNAP549 associations with DNA and then washed the surface-tethered DNA molecules with reaction buffer, removing unbound proteins. Then, to promote photobleaching, we increased laser excitation power and removed oxygen scavengers. Comparison of the number of Mcm2-74SNAP549 photobleaching steps after the wash with the number of association steps that accumulated before the wash showed no single-step increase in fluorescence before the wash resulted in a two-step photobleaching afterward (Fig. 1D, top). We confirmed that loss of fluorescence was due to photobleaching and not dissociation of Mcm2-7 by observing previously non-illuminated microscope fields of view. These data eliminate models in which multiple Mcm2-7 complexes are recruited simultaneously. We conclude that Mcm2-7 association occurs in a one-at-a-time manner.
We next asked whether loading of salt-resistant Mcm2-7 hexamers around origin-DNA occurred sequentially or simultaneously. We used the same photobleaching assay (described above) except a high-salt wash was used to remove any incompletely-loaded Mcm2-7 complexes prior to photobleaching. If loading of both Mcm2-7 hexamers occurs simultaneously, we should observe only even numbers of high-salt resistant hexamers. In contrast, if loading occurs sequentially, we should observe even and odd numbers of high-salt resistant hexamers. At low protein concentrations we observed both one-and two-step photobleaching events (Fig. 1D, bottom, 1E). Roughly half (79/160) of all single Mcm2-7-associated fluorophores that colocalized with origin DNA before the high-salt wash were high-salt resistant, and 67% (40/60) of the double Mcm2-7-associated fluorophores were high-salt resistant. When we increased protein concentrations, we also observed DNA molecules with three and four origin-dependent, high-salt resistant Mcm2-7 complexes (Fig. 2D), indicating that more than one double-hexamer loading event occurred at a single origin.

We considered the possibility that the apparent colocalization of odd numbers of loaded Mcm2-7 complexes was due to incomplete fluorescent labeling of Mcm2-7. For example, a single salt-resistant Mcm2-7-associated fluorophore could be the result of loading two Mcm2-7 complexes, only one of which is fluorescently labeled. To address this possibility, we purified Mcm2-7 complexes that were labeled on two subunits with different fluorophores (Mcm2-7\textsuperscript{4SNAP549/7SORT649}). Because the SNAP-tag and sortase labeling approaches are independent of each other, we could use single-molecule imaging
to determine the efficiency of each labeling protocol (79% for SNAP and 77% for sortase). This labeling protocol also increased the proportion of Mcm2-7 complexes that have at least one coupled fluorophore to 95%. Using the measured labeling efficiencies, we determined the number of high-salt resistant Mcm2-7 complexes with no more than one of each fluorophore that would be expected if only double hexamers were loaded (Fig. 2E, Model II). Assays with Mcm2-7\textsuperscript{4SNAP549/SORT549} yielded single, salt-resistant fluorophores in a proportion that is inconsistent with this model. Instead our data is consistent with a model where both single and double hexamers are loaded (in a 52:48 ratio based on our data; Fig. 2E, Model I). We conclude that Mcm2-7 complexes are both recruited and loaded onto origin DNA in a sequential manner.

**Distinct Cdc6 and Cdt1 molecules load the first and second Mcm2-7**

We investigated the number of Cdt1 and Cdc6 molecules required for helicase loading and their relative times of DNA association. Both proteins are essential for loading but show little or no association with DNA in bulk assays (Coster et al., 2014; Kang et al., 2014), suggesting that their protein and/or DNA associations during helicase loading are transient. To detect these associations, we simultaneously monitored the binding of two different protein pairs labeled with distinguishable fluorophores: either Cdt1\textsuperscript{SORT549} with Mcm2-7\textsuperscript{4SNAP646} or Cdc6\textsuperscript{SORT549} with Mcm2-7\textsuperscript{4SNAP646}. The associations of both fluorophores with origin-DNA were monitored simultaneously, revealing relative times of arrival and departure for the two molecules in each pair.
Consistent with being recruited to origins as a complex, we typically observed that Cdt1 and Mcm2-7 associated with origin DNA simultaneously (Fig. 3A, 4A-C). Uncommon instances where Cdt1 or Mcm2-7 are seen associating separately (Cdt1 alone: 11.4%, Mcm2-7 alone: 18.6%) are likely caused by incomplete dye labeling of the other protein because the frequencies of these events are similar to the fractions of unlabeled Mcm2-7 or Cdt1 (14% and 20%, respectively). Like Mcm2-7, Cdt1 dwell times followed a multi-exponential distribution, indicating the presence of at least two types of Cdt1-containing complexes on the DNA (Fig. 3B). Consistent with this interpretation, we identified two classes of Mcm2-7/Cdt1 dwell-time and departure behaviors. In many instances Cdt1 and Mcm2-7 were released simultaneously (i.e. within 1 s, see Fig. 4B and 4C). This release pattern occurs whether or not the DNA molecule already had an associated Mcm2-7. These associations were typically short-lived (Fig. 3C) and represent non-productive binding events. Interestingly, these events were less frequent if the Mcm2-7/Cdt1 was the second (29%) rather than the first (53%) to arrive at the DNA. In the remaining cases, Cdt1 was typically longer-lived (Fig. 3D) and was released from origin DNA by itself, leaving behind an associated Mcm2-7. Clearly, only instances when Cdt1 is released independently of Mcm2-7 can be on the pathway for double-hexamer formation. Because Cdt1-associated fluorophore photobleaching was much slower than Cdt1 dissociation (Fig. 4D; Table 2), nearly all loss of fluorescent colocalization was due to dissociations, not photobleaching.
Figure 3. Distinct Cdt1 molecules load the first and the second Mcm2-7 hexamer. (A) Cdt1 molecules arrive with Mcm2-7 but release quickly after the complex arrives. A representative two-color recording of Mcm2-7S_NAPF646 and Cdt1SORT546 fluorescence at an origin-DNA location is shown. The baseline of the red plot (Mcm2-7) is shifted up relative to the green plot throughout the manuscript when two-color recordings are displayed together. The sequence of single-frame images of the Cdt1- and Mcm2-7-fluorescent spots illustrates the concurrent arrival of Cdt1 and Mcm2-7. Cdt1 release occurs either with (green arrow) or without (black arrows) concurrent Mcm2-7 release. (B) Cdt1 dwell times on DNA have a multiexponential distribution. Cdt1 dwell times were plotted as a histogram. Inset shows semilogarithmic cumulative survival plot as in Fig. 1C. (C-D) There are two types of Cdt1 release events.
(C) Histogram shows the duration of Cdt1 origin-DNA associations when Cdt1 releases with Mcm2-7. The mean dwell time ± standard error of the mean (s.e.m.) is reported.
(D) Histogram shows the duration of Cdt1 origin-DNA associations when Cdt1 releases before Mcm2-7. The mean dwell time ± s.e.m. is reported.
Figure 4

A. Additional fluorescence intensity records of DNA-associated Mcm2-7\(^{SNAPjF646}\) and Cdt1\(^{SORT549}\).

B. Additional fluorescence intensity records of DNA-associated Mcm2-7\(^{SNAPjF646}\) and Cdt1\(^{SORT549}\).

C. Additional fluorescence intensity records of DNA-associated Mcm2-7\(^{SNAPjF646}\) and Cdt1\(^{SORT549}\).

D. Relative exposure of 0.65 and Cdt1\(^{SORT549}\) dwell time.

Figure 4. Distinct Cdt1 molecules load the first and the second Mcm2-7 hexamer. (A-C) Additional fluorescence intensity records of DNA-associated Mcm2-7\(^{SNAPjF646}\) and Cdt1\(^{SORT549}\).
Example traces show:

(A) Two Mcm2-7 associations with DNA and associated Cdt1 binding and release events.

(B) Mcm2-7 and Cdt1 arriving and departing simultaneously on DNA molecules without an Mcm2-7 already present.

(C) Mcm2-7 and Cdt1 arriving and departing simultaneously as the second Mcm2-7 association occurring on a DNA (i.e., when an Mcm2-7 complex is already present). Second Mcm2-7/Cdt1 associations are indicated by blue arrows.

(D) Cdt1 photobleaching rate measurement. The calculated dwell times for Cdt1 were plotted for different relative laser exposures and fit either to single (649 data) or double (549 data) exponential functions. When fit to a double exponential, only the longer time constant was used to determine the photobleaching rate. Data for the experiments shown throughout the manuscript was acquired at a relative laser exposure of 1. To measure the effect of photobleaching, the relative exposure was modulated by either changing the intensity of the laser or the exposure time to the laser. We show example histograms of Cdt1SORT1649 dwell times at a relative exposure of 0.65 (left panel; N=612) and 2.0 (middle panel; N=795). The red lines indicate single exponential fits. For each dye, the relative exposure (x-axis) is plotted against the rate constant (y-axis). Photobleaching rate constants (slopes) from these fits are summarized in Table 2.
Table 2. Measured photobleaching rates (± s.e.) of the different fluorophores used.

<table>
<thead>
<tr>
<th>Fluorophore</th>
<th>549 bleaching rate</th>
<th>649 bleaching rate</th>
<th>JF646 bleaching rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORC</td>
<td>$0.000 \pm 0.001$ s$^{-1}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cdc6</td>
<td>$0.001 \pm 0.001$ s$^{-1}$</td>
<td>$0.013 \pm 0.003$ s$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>Cdtl</td>
<td>$0.000 \pm 0.007$ s$^{-1}$</td>
<td>$0.014 \pm 0.002$ s$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>Mcm2-7 (SNAP)</td>
<td>$0.004 \pm 0.001$ s$^{-1}$</td>
<td>$0.016 \pm 0.001$ s$^{-1}$</td>
<td>$0.0004 \pm 0.0003$ s$^{-1}$</td>
</tr>
</tbody>
</table>

All proteins are sortase-tagged unless otherwise indicated. The rates reported are the photobleaching rates at the 1.0 relative exposure condition of the experimental measurements. These photobleaching rates can be converted to a fluorophore half-life. For example, Mcm2-7$^{4\text{SNAP}549}$ fluorophore has a half-life of $\frac{\ln(2)}{0.004}$ s$^{-1}$ = 173 s. The ORC, Cdc6, and Cdtl 549 rates are all zero within experimental error.
Like Cdt1, Cdc6 association with the DNA is dynamic with distinct molecules acting during loading of the first and second Mcm2-7 (Fig. 5A, 6A). Simultaneous analysis of Mcm2-7 and Cdc6 DNA association showed short Cdc6-DNA associations (mean lifetime $27.8 \pm 1.5$ s; Fig. 6B), a subset of which directed Mcm2-7 recruitment (35.8%, N=514, Fig. 5A, 6A). Cdc6 consistently anticipated Mcm2-7 arrival at the DNA (>85%, Fig. 5A, 6A). The remaining cases likely reflected the action of unlabeled Cdc6. We observed distinct Cdc6 proteins direct recruitment of the first and second Mcm2-7 with a similar rate constant (Fig. 6C). The high frequency of Cdc6 DNA associations led us to test and confirm that sequential binding of Cdc6 and Mcm2-7 was not coincidental for either Mcm2-7 loading event (Fig. 5B).

**Release of Cdc6 and Cdt1 is sequential during helicase loading**

We next asked whether helicase loading led to a defined order of Cdc6 and Cdt1 release. We took two approaches to address this question: (1) we performed experiments in which Cdc6 and Cdt1 were labeled with different fluorophores, and (2) we compared the times of Cdc6 and Cdt1 release relative to the time of the corresponding Mcm2-7 association in the previously described double-labeled experiments (Mcm2-7$^{4\text{SNAPJF646}}$ with either Cdt1$^{\text{SORT549}}$ or Cdc6$^{\text{SORT549}}$).
Figure 5

A. Mcm2-7
Cdc6

B. 1st Mcm2-7
2nd Mcm2-7

Figure 5. Distinct Cdc6 molecules recruit and load the first and the second Mcm2-7 hexamer.
(A) Distinct Cdc6 molecules anticipate each Mcm2-7 association. A representative fluorescence intensity record for Mcm2-7^{SNAPF646} and Cdc6^{SORT549} at origin-DNA. Images of the Cdc6- and Mcm2-7-associated fluorescent spots show Cdc6 binds before the arrival of the first Mcm2-7 complex.

(B) Cdc6 association anticipates binding of the first and second Mcm2-7. Full histogram (top) and expanded view (bottom) of Mcm2-7 arrival time minus the closest Cdc6 arrival time on the same DNA molecule (blue bars). Data is separated into Mcm2-7 complexes arriving at the DNA first (left) or second (right). In >85% of the observations the difference was greater than zero, indicating that Cdc6 arrived before Mcm2-7; in the remaining <15%, Mcm2-7 arrived before Cdc6 (likely due to an unlabeled Cdc6 molecule). Red bars show a control analysis in which each Mcm2-7 arrival time was paired with the closest Cdc6 arrival time on a different, randomly selected DNA molecule. The randomized control does not show the prominent peak at differences between 0 and +50 s indicating the sequential association of Cdc6 and Mcm2-7 was not coincidental.
Figure 6

(A) Additional fluorescence intensity traces for DNA-associated Mcm2-7₄SNAPJF₆₄₆ and Cdc6SORT₅₄₉ over time showing double hexamer associations.

(B) Measurement of Cdc6 dissociation and photobleaching rates. Cdc6 spot disappearance rate constants at different laser exposures were measured as described in Fig. S2D. Dwell time distributions were fit with a single exponential model.

(C) First and second Mcm2-7 complexes associate with Cdc6-bound DNA with similar rates. Histograms showing the time of arrival of the first (left; N=60) or second (right; N=62) Mcm2-7 complex relative to the arrival of the associated Cdc6. Single exponential fits are plotted in red. The fits yield the following rate-constants: 0.045 s⁻¹ ± 0.006 s⁻¹ (for the first Mcm2-7) and 0.05 s⁻¹ ± 0.01 s⁻¹ (for the second Mcm2-7).

Figure 6. Distinct Cdc6 molecules recruit and load the first and the second Mcm2-7 hexamer.

(A) Additional fluorescence intensity traces for DNA-associated Mcm2-7₄SNAPJF₆₄₆ and Cdc6SORT₅₄₉ over time showing double hexamer associations.

(B) Measurement of Cdc6 dissociation and photobleaching rates. Cdc6 spot disappearance rate constants at different laser exposures were measured as described in Fig. S2D. Dwell time distributions were fit with a single exponential model.

(C) First and second Mcm2-7 complexes associate with Cdc6-bound DNA with similar rates. Histograms showing the time of arrival of the first (left; N=60) or second (right; N=62) Mcm2-7 complex relative to the arrival of the associated Cdc6. Single exponential fits are plotted in red. The fits yield the following rate-constants: 0.045 s⁻¹ ± 0.006 s⁻¹ (for the first Mcm2-7) and 0.05 s⁻¹ ± 0.01 s⁻¹ (for the second Mcm2-7).
When Cdc6 and Cdt1 were labeled in the same experiment, we consistently saw Cdc6 associating with and releasing from origin-DNA before Cdt1 (Fig. 7A, 8A). Because only non-productive Cdt1-DNA interactions had dwell times less than 6 s (see Fig. 4C), we excluded these molecules from our analysis. Cdc6$^{\text{SORT649}}$ is released before Cdt1$^{\text{SORT549}}$ in $>95\%$ of cases when Cdt1 and Cdc6 were co-localized on a DNA (Fig. 7B). When the fluorophores coupled to the proteins were swapped (Cdc6$^{\text{SORT549}}$ and Cdt1$^{\text{SORT649}}$) $>90\%$ of observations showed Cdc6 dissociates from origin-DNA before Cdt1 (Fig. 8B). This lower percentage is likely due to the higher photobleaching rate of the 649 dye (Table 2). These results suggest that Cdc6 is released prior to Cdt1 during helicase loading.

Because Mcm2-7 was unlabeled in the previous experiments, we did not know which of the Cdc6-Cdt1 DNA co-localization events directed double-hexamer formation. To address whether Cdc6 is released before Cdt1 during double-hexamer formation, we analyzed the time that each Cdc6 or Cdt1 spent on the DNA with Mcm2-7. Consistent with the Cdc6-Cdt1 double-labeling experiments, the average time between Mcm2-7 arrival and Cdc6 release is significantly shorter than the corresponding time before Cdt1 release (Fig. 7C). Both the Cdc6$^{\text{SORT549}}$ and Cdt1$^{\text{SORT549}}$ release times are $>50$-fold shorter than the fluorescent dye lifetimes calculated from photobleaching rates (Table 2), verifying that these are dissociation events and not due to photobleaching. We conclude that each Mcm2-7 loading event is associated with the ordered release of Cdc6 followed by Cdt1 from the DNA.
Figure 7. Cdc6 release occurs before Cdt1 release.

(A) Three representative fluorescence intensity records for Cdc6 and Cdt1 release. The graphs show the release and arrival times of Cdc6 and Cdt1 proteins, highlighting the sequence of events. Cdc6 release occurs before Cdt1 release.

(B) Scatter plot showing the release and arrival times of Cdc6 and Cdt1 proteins. The plot indicates that Cdc6 releases first with 95.5% confidence.

(C) Graphs showing the fraction bound of Cdc6 and Cdt1 proteins over time. Cdc6 releases first, followed by Cdt1.

(D) Graphs showing the fraction of bound Cdc6 and Cdt1 proteins over time, with inset graphs providing detailed data for the first and second releases of Cdc6 and Cdt1 proteins.

(E) Graphs showing the fraction of bound Cdt1 proteins over time, with inset graphs providing detailed data for the first and second releases of Cdt1 proteins.

(A) Three representative fluorescence intensity records for Cdc6 and Cdt1 release.
showing arrival and departure of Cdc6 before Cdt1.

(B) Release of Cdc6 anticipates Cdt1 release in a majority of cases. Time of Cdt1 release (y-axis) is plotted against time of Cdc6 release (x-axis, both times are measured from start of simultaneous presence of Cdc6 and Cdt1). The red line represents where points would fall if Cdc6 and Cdt1 released simultaneously. The fraction of measurements in which Cdc6 is released before Cdt1 is reported.

(C) Release of Cdc6 occurs before release of Cdt1 during double hexamer formation. Survival function for Cdc6$^{S0RT549}$ and Cdt1$^{S0RT549}$ dwell times after the first or second Mcm2-7 associates with origin DNA. The y-axis represents the fraction of Cdc6 or Cdt1 molecules that are still associated after the time represented on the x-axis.

(D-E) Cdc6 and Cdt1 release events are slower for the second versus the first Mcm2-7 loading events.

(D) The time of Cdc6 release after Mcm2-7 association is plotted for the first (blue) and second Mcm2-7 association (red) as a survival plot (the fraction of Cdc6 molecules that remain DNA-associated is plotted against time). Inset shows the first 40 s of the entire plot to emphasize the presence of a lag prior to DNA release. Numbers are mean release times ± s.e.m. for the first or second Mcm2-7-associated Cdc6 molecule.

(E) Cdt1 release after the first (blue) and second Mcm2-7 association (red) as a survival plot as described for 7D.
Figure 8. Cdc6 release occurs before Cdt1 release.
(A) Representative fluorescence intensity trace for DNA-associated Cdc6$^{\text{SORT549}}$ and Cdt1$^{\text{SORT649}}$ over time showing arrival and departure of Cdc6 before Cdt1.
(B) Release of Cdc6 anticipates Cdt1 release. Time of Cdc6 release (measured relative to the time when both Cdc6 and Cdt1 are associated, x-axis) is plotted against time of Cdt1 release (also relative to when both Cdc6 and Cdt1 are present, y-axis; N=142). The red line represents where points would fall if Cdc6 and Cdt1 were released together. The percent of points falling above the red line, representing cases where Cdc6 is released before Cdt1 is reported.
Kinetic evidence for distinct mechanisms loading the first and second helicase

We reasoned that if loading of the first and second helicases occurred by different mechanisms, the time that Cdc6 and Cdt1 would spend associated with the first versus the second Mcm2-7 would differ. The resulting survival curves showed delays between arrival of Mcm2-7 and release of Cdc6 or Cdt1, suggesting that the release of both proteins involves multiple steps after Mcm2-7 recruitment. Although the order of Cdc6 and Cdt1 release remained the same, we found that the release times were significantly longer for the second Mcm2-7 loading event for both Cdc6 ($p < 0.003$, Fig. 7D) and Cdt1 ($p < 0.001$, Fig. 7E). These kinetic data suggest that loading of the first and second helicase occurs through distinct mechanisms.

A single ORC directs formation of the Mcm2-7 double hexamer

There are multiple models for the stoichiometry of ORC during helicase loading (Fig. 9A). One ORC molecule could direct both helicase loading events (Model I). Alternatively, two ORC molecules could be present throughout the loading reaction (Model II). Finally, it is possible that distinct ORC molecules direct each loading event but both ORC molecules are only present for the second loading event (Model III) or, like Cdc6 and Cdt1, each ORC is only present during loading of one Mcm2-7 (Model IV). To distinguish between these models, we performed CoSMoS with simultaneous labeling of ORC and Mcm2-7.
Figure 9. One ORC molecule directs recruitment and loading of the first and second Mcm2-7 hexamers.
(A) Models for ORC function during helicase loading. The initial Mcm2-7 (shown in gray) is associated with ORC/Cdc6 via its C-terminal end (Sun et al., 2013). The second Mcm2-7 is shown in yellow. See text for additional details.

(B) ORC exhibits longer dwell times in the presence of origin-containing DNA. Dwell times for ORC<sup>1SOSRT549</sup> associations in the presence of unlabeled Cdc6, Cdtl and Mcm2-7, are plotted as histograms for either DNA with a wild-type ACS (left panel; N=949) or or a mutated ACS (right panel; N=1362). Insets show expanded y-axes.

(C) Measurement of ORC photobleaching rate. Cdc6 spot disappearance rate constants at different laser exposures were measured as described in Fig. S2D. Dwell time distributions were fit with a double exponential function and the long time constant was plotted against relative laser exposure.

(D) Additional representative traces of ORC<sup>1SOSRT549</sup> and Mcm2-7<sup>4SNAPJF646</sup> origin-DNA colocalizations plotted over time. Association of first and second Mcm2-7 are marked with red and blue arrows, respectively.

(E) ORC associates with DNA as a monomer. Number of ORC associations observed before a wash using reaction buffer, and the number of fluorophores that are bleached immediately after.
Initially, we fluorescently labeled ORC on the Orc1 subunit (ORC\textsuperscript{1SORT549}) and observed associations with DNA in the presence of unlabeled Cdc6, Cdt1 and Mcm2-7. ORC DNA binding showed a broad distribution of dwell times (Fig. 9B, left panel). Consistent with the long-lived associations reflecting ORC binding to the ACS, mutation of this element resulted in >94% of ORC DNA associations being short-lived (<10 s, Fig. 9B, right panel). The associations of ORC are shorter than the calculated fluorescent dye lifetimes confirming that we are observing dissociations and not photobleaching (Fig. 9C, Table 2).

To identify ORC molecules involved in helicase loading, we simultaneously monitored ORC and Mcm2-7 DNA associations (Fig. 10A). As expected, ORC associates with DNA and Cdt1/Mcm2-7. Unlike Cdc6 and Cdt1, we consistently observed a single increase in ORC fluorescence that remained present continuously during recruitment of the first and second Mcm2-7 complexes (Fig. 9D, 10A).

Because ORC multimers have been detected (Sun et al., 2012), we addressed whether ORC complexes dimerize in solution prior to DNA binding by counting the number of photobleaching steps associated with single increases in ORC-associated fluorescence (as was described for Mcm2-7). The large majority of cases were consistent with ORC binding as a single complex (67 of 69, Fig. 9E). These data confirmed that the single increases in ORC-associated fluorescence were due to single ORC molecules associating with origin-DNA during loading.
Figure 10

(A) Representative fluorescence intensity record for ORC\textsuperscript{1SORT549} and Mcm2-7\textsuperscript{4SNAPJF646} at an origin-DNA location. Association of first and second Mcm2-7 are marked with red and blue arrows, respectively.

(B) A single ORC complex directs recruitment of two hexamers. The fraction (± s.e.) of DNA molecules observed to have zero, one or two ORC fluorophores bound when the second Mcm2-7 was recruited, is plotted (bars) together with the predicted number of associated fluorophores (red and blue squares) of different models (see Fig. S5A).

(C) ORC is released rapidly after recruitment of the second Mcm2-7 hexamer. Histograms showing the time between the association of the second Mcm2-7 and ORC release (top), or association of the first Mcm2-7 and ORC release (bottom).

(D) Release of Cdc6\textsuperscript{SORT549} (blue), Cdt1\textsuperscript{SORT549} (red), and ORC\textsuperscript{SORT549} (black) after the association of the second Mcm2-7\textsuperscript{4SNAPJF646} complex is plotted as a survival function. There are two ORC molecules that associate for >400s (1033.8s and 709.6s) that are not shown and disproportionately affect the mean dwell time. Grey lines represent a 2.5% and 97.5% confidence interval for the ORC dataset showing that there is no significant difference between Cdt1 and ORC release time distributions. Numbers in parentheses represent the mean release times ± s.e.m.
Although the majority of observations involved a single ORC directing loading of two Mcm2-7 hexamers, occasionally we observed the presence of multiple DNA-bound ORC molecules at the time of a Mcm2-7 association. To address which models for ORC function during helicase loading were possible, we counted the number of DNA-associated ORC molecules (by counting step-wise increases in ORC fluorescence) during the second Mcm2-7 hexamer association (Fig. 10B). Models II and III predict two ORC molecules bound to DNA when the second Mcm2-7 is recruited. In contrast to these models, we observed two ORC molecules associated during loading of the second hexamer only 5% of the time (as opposed to 70% predicted by Model II or III using the measured ORC labeling efficiency; 85%; see Experimental Procedures). Instead, we observe a single ORC present during association of the second helicase 80% (96/120) of the time, very close to the percentage expected if a single ORC is responsible for loading the second Mcm2-7 (85%). To distinguish between models I and IV, we asked whether the same or different ORC molecules directed the first and second helicase-loading events. Consistent with model I, 94% (N=96) of observations showed a single ORC complex continuously present during both Mcm2-7 recruitment events. Thus, our data indicates one ORC molecule directs loading of both the first and the second Mcm2-7 hexamer (Model I).

Interestingly, in most traces where two Mcm2-7 associate with the DNA, we observed dissociation of ORC from origin DNA soon after binding of the second Mcm2-7 hexamer (see Fig. 9D, 10A). Plotting the times between the association of the second Mcm2-7 hexamer and ORC release (Fig. 10C, blue bars), we observed only one instance
where ORC released from DNA in <15 s (13.1 s), followed by a short time interval (15 s - 90 s) during which 87% of the ORC complexes were released. The shape of this distribution suggests that, like Cdc6 and Cdt1, release of ORC is a multi-step process. In contrast, a much broader distribution was observed when ORC release was measured relative to DNA association of the first Mcm2-7 hexamer (Fig. 10C, red bars), suggesting ORC release is independent of this event. To investigate the order of ORC release relative to the other helicase-loading proteins, we compared the distribution of ORC, Cdc6 and Cdt1 dwell times after binding of the second Mcm2-7 complex (Fig. 10D), using data from two-color experiments with Mcm2-745SNAPF646 and 549-labeled ORC, Cdt1 or Cdc6. Photobleaching of the 549-labeled proteins was insignificant relative to their observed dwell times (Table 2). Although there is a significant difference between release of Cdc6 and ORC (p < 0.001), we saw no significant difference in the distributions of Cdt1 and ORC release (Fig. 10D). Thus, loading of the first Mcm2-7 allows ORC retention, whereas loading of the second Mcm2-7 appears to induce the linked release of ORC and the second Cdt1.

Recruitment of a Second Mcm2-7 Results in Rapid Double Hexamer Formation

The interactions that drive recruitment of the second Mcm2-7 remain unclear (Yardimci and Walter, 2014). To gain insight into this event, we used fluorescence resonance energy transfer (FRET)-CoSMoS (Crawford et al., 2013) to detect the proximity of the Mcm7 N-terminal domains upon double-hexamer formation (Costa et al., 2014; Sun et al., 2014). To this end we labeled the Mcm7 subunit in separate preparations of Mcm2-7 with either 549 (Mcm2-778SOR1549 - donor) or 649 (Mcm2-
When mixed in an equimolar ratio, the differently labeled Mcm2-7 should be in the same double hexamer ~50% of the time, and those molecules should exhibit substantial FRET efficiency ($E_{\text{FRET}}$) because the Mcm7 N-terminal regions are in close proximity in the double hexamer (Sun et al., 2014). We alternated between 633 and 532 nm laser excitation to monitor both arrival of each Mcm2-7 and $E_{\text{FRET}}$. Importantly, when Mcm2-7$^{7\text{SORT}549}$ and Mcm2-7$^{7\text{SORT}649}$ were sequentially recruited to the origin DNA (in either order) we observed rapid development of a high $E_{\text{FRET}} \approx 0.53$ state (Fig. 11B, 12, 11C, blue bars). A second peak at $E_{\text{FRET}} \approx 0.02$ was also observed in the absence of acceptor (Fig. 11C, red bars) and thus represents state(s) with no detectable FRET. Consistent with the detected FRET signal occurring as a consequence of double-hexamer formation, the high $E_{\text{FRET}}$ state was stable for hundreds of seconds and 95% (55/58) of the complexes that exhibited $E_{\text{FRET}} \approx 0.53$ were high-salt resistant.

To determine when double-hexamer formation occurs relative to binding of the second Mcm2-7, we compared the time of FRET formation to the time of arrival of the second Mcm2-7 (Fig. 11D). We found the mean time between recruitment of the second Mcm2-7 hexamer until formation of FRET was $7.8 \pm 0.1$ s. This time is significantly shorter than release of Cdc6 after arrival of the second Mcm2-7 hexamer ($23.2 \pm 1.7$ s, $p < 0.001$), indicating that formation of the N-terminal-to-N-terminal interactions anticipates, and is therefore independent of, Cdc6 and Cdt1 release (Fig. 11D).
Figure 11

A. No FRET

Dex (532 nm) → D em (≤635 nm)

FRET

Dex (532 nm) → A em (>635 nm)

no A em

B. A ex/ A em

D ex/ D em

D ex/ A em

black = D ex/ (A em + D em)

green = D ex/ D em

red = D ex/ A em

C. N=86

N=86

D. FRET (N=106): 7.8s ± 0.1s

Cdc6 (N=72): 23s ± 12s

Cdt1 (N=74): 65s ± 5s

ORC (N=98): 77s ± 1s

Figure 11. Double hexamer formation occurs quickly upon recruitment of the second Mcm2-7 hexamer
(A) When the two fluorophores (green circle = Dy549, red circle = Dy649) are not associated, excitation of the donor (D ex) will only yield emission from the donor (D em). However when the two fluorophores are in close proximity, we observe acceptor emission (A em) upon D ex, and a weaker D em signal. Wavelengths represent laser excitation and emissions that were monitored.

(B) Representative fluorescence records for experiments using a mixture of Mcm2-7⁷SORT⁵⁴⁹ and Mcm2-7⁷SORT⁶⁴⁹ showing FRET upon arrival of the second Mcm2-7. Red squares highlight when Mcm2-1-⁷SORT⁶⁴⁹ associates with DNA (Mcm2-7⁷SORT⁵⁴⁹ is already present), and blue squares highlight when FRET occurs. Images and records of fluorescence intensity for D ex/D em (Mcm2-7⁷SORT⁵⁴⁹), A ex/A em (Mcm2-7⁷SORT⁶⁴⁹), total emission (D ex / (D em + A em), and FRET (D ex/A em) are shown together with calculated EFRET.

(C) Histogram of EFRET is plotted for times when a single Mcm2-7⁷SORT⁵⁴⁹ and a single Mcm2-7⁷SORT⁶⁴⁹ are present (blue bars) or when only Mcm2-7⁷SORT⁵⁴⁹ is associated with the DNA (empty grey bars). The histogram displays the first ten consecutive EFRET measurements after arrival of the second Mcm2-7 for 86 DNA molecules (the same number of molecules and time points were used for the control). EFRET data below -0.5 was excluded from the plot (3/860 signal points and 17/860 control points).

(D) Double-hexamer formation anticipates Cdc6, Cdt1 and ORC release. Survival after the association of the second Mcm2-7 complex of the no-FRET state (green) and of DNA-bound Cdc6SORT⁵⁴⁹ (blue), Cdt1SORT⁵⁴⁹ (red), and ORC1SORT⁵⁴⁹ (black). Mean times ± s.e.m. until FRET increase and ORC, Cdt1, Cdc6 release are reported for comparison.
Figure 12. Double hexamer formation occurs quickly upon recruitment of the second Mcm2-7 hexamer
(A-D) Additional representative records (labeled as in Fig. 6A) for experiments using a 1:1 mixture of Mcm2-7\textsuperscript{SORT1549} and Mcm2-7\textsuperscript{SORT1649} showing FRET upon arrival of the second Mcm2-7. Red squares highlight when Mcm2-7\textsuperscript{SORT1549} associates with DNA second (A-C), green square highlights when Mcm2-7\textsuperscript{SORT1649} associates with DNA second (D), and blue squares highlight when FRET occurs. In panel (D) FRET efficiency is only plotted during the time interval when Mcm2-7\textsuperscript{SORT1649} is present because $E_{\text{FRET}}$ cannot be accurately measured in the absence of donor. The loss of acceptor fluorescence at $\sim 1000$ s is likely due to bleaching of Mcm2-7\textsuperscript{SORT1649} and shows an associated increase in donor signal and decrease to background $E_{\text{FRET}}$ levels.
DISCUSSION

By determining precise protein/DNA stoichiometry and real-time dynamics, the single-molecule observations of helicase loading described here provide important insights into this event. Together, our findings strongly support a model in which the first and second helicase are loaded by distinct mechanisms and the second Mcm2-7 complex is recruited through interactions with the first. Accordingly, we propose a new model for helicase loading that is consistent with our current data and is described below (Fig. 13).

Recruitment and loading of Mcm2-7 helicases occur in a one-at-a-time manner

Monitoring associations in real-time reveals sequential recruitment and loading of Mcm2-7 helicases to origin-DNA. One-at-a-time recruitment is consistent with an initial complex containing a single Mcm2-7 associated with the three helicase-loading proteins (Sun et al., 2013) and ensemble assays that show temporal separation of Mcm2-7 recruitment (Fernández-Cid et al., 2013). Recent structural observations indicate that the Mcm2/5 gates, which must open to provide DNA access to the Mcm2-7 central channel (Samel et al., 2014), are staggered in the double hexamer (Costa et al., 2014; Sun et al., 2014). Concerted Mcm2-7 loading would require alignment of the two Mcm2/5 gates to allow simultaneous DNA entry into the central channels of both hexamers. In contrast, sequential Mcm2-7 loading can readily accommodate the formation of a staggered-gate double-hexamer structure.
Figure 13. Proposed model for helicase loading.
Proteins present are indicated adjacent to each illustration (O=ORC, C_6=Cdc6, C_1=Cdt1, M=Mcm2-7). Reversible steps that are observed are indicated. See text for details.
Although high-salt resistant single hexamers have been detected after artificially closing the Mcm2/5 gate (Samel et al., 2014), previous studies have not detected single loaded (high-salt-resistant) Mcm2-7 complexes in unperturbed helicase-loading reactions (Evrin et al., 2009; Kang et al., 2014; Remus et al., 2009). This difference may be due to the higher protein concentrations used in these ensemble reactions. Alternatively, the high-salt-resistant single hexamers may be less stable than the double hexamers resulting in their loss during sample preparation for chromatography or EM. Indeed, a higher percentage of double hexamers showed high-salt resistance relative to single hexamers (74% versus 49%; see Fig.1D). The high-salt wash is effective in the single-molecule assay setting, however, as this treatment efficiently releases incompletely loaded Mcm2-7 formed in the absence of ATP hydrolysis (Table 1, ATPγS).

**Ordered release of Cdc6 and Cdt1 molecules during double-hexamer loading**

Our studies provide insights into Cdc6 and Cdt1 function during helicase loading. Previously, robust DNA association of these proteins was only observed when helicase-loading reactions were arrested at an early ATP-dependent step. We found that the initial ORC-Cdc6-Cdt1-Mcm2-7 (OC₆C₁M) complex has two possible fates (Fig.13, left): (i) simultaneous release of Mcm2-7 and Cdt1 or (ii) sequential release of Cdc6 and Cdt1 with retention of Mcm2-7. The former is most likely the reversal of the initial Mcm2-7/Cdt1 association whereas the latter pathway leads to sequential formation of OC₁M and OM complexes and Mcm2-7 loading. Based on this distinction, we propose that release of Cdt1 independent of Mcm2-7 is coupled to successful helicase loading (illustrated as closing of the Mcm2/5 gate, Fig. 13). Consistent with this hypothesis, treatments (e.g.
ATPγS) or mutations (e.g. Mcm2-7 ATPase mutations, Coster et al., 2014; Kang et al., 2014) that lead to Cdt1 retention prevent helicase loading. We note that other times of ring closure (and opening) than those illustrated in the model are possible.

Electron microscopic (EM) and ensemble assays suggest the existence of helicase loading intermediates with ORC-Cdc6-Mcm2-7 (OC₆M) and ORC-Cdc6-Mcm2-7-Mcm2-7 (OC₆MM, Sun et al., 2014). Our findings suggest that the OC₆M complex is a short-lived intermediate formed prior to recruitment of the second Mcm2-7/Cdt1 complex rather than being formed by release of Cdt1 from the OC₆C₁M (Fernández-Cid et al., 2013). We do not see evidence of an OC₆MM complex during helicase loading and there is no direct evidence that Cdc6 is present in the 2-D class averages used in these studies (Sun et al., 2014). Given their relatively lower resolution, these studies could have detected either the OC₆C₁MM or OC₁MM complexes that we observe (Fig. 13, right).

Our previous studies found an intermediate with two Cdt1 complexes that is not detected in the current studies (Takara and Bell, 2011). During efforts to reconcile these findings, we found the Mcm2-7 protein used in the previous studies contained a non-lethal mutation in the C-terminus of Mcm3 that is predicted to inhibit Cdc6 interactions (Frigola et al., 2013). We suspect that this mutant enhanced dependence on other interactions leading to the detection of two Cdt1 associations.

**Loading of the first and second Mcm2-7 occurs by distinct mechanisms**

In addition to answering a long-standing question about ORC function, our data indicating that one ORC molecule directs Mcm2-7 double-hexamer formation strongly
suggests that different mechanisms direct loading of the first and second Mcm2-7. EM studies suggest that during helicase loading ORC interacts with the C-terminal end of the first Mcm2-7 on adjacent DNA (Sun et al., 2014; 2013). Assuming this configuration, direct recruitment of the second Mcm2-7 complex by the same ORC would load the two Mcm2-7 molecules in a head-to-tail fashion (Fig. 14, top). Even if ORC had a second binding site for Mcm2-7 on its opposite side, a similar direct interaction with Mcm2-7 could not load two Mcm2-7 complexes with adjacent N-terminal domains (Fig. 14, bottom). Further evidence in favor of distinct mechanisms loading the first and second Mcm2-7 include: (i) the two loading events show different Cdc6, Cdt1 and ORC release kinetics; (ii) Cdt1 associated with the second loading event shows an increased propensity to release without Mcm2-7.

We considered the possibility that a second ORC binds DNA in the opposite orientation and loads the second helicase by the same mechanism as the first. Several observations argue against this model. First, because we do not consistently detect a second ORC during recruitment of the second Mcm2-7, the average dwell time for this second ORC would have to be below our detection limit (~0.5 s). This limit is >10-fold shorter than the average dwell time observed for ORC on non-origin DNA (Fig. 9B). Second, in contrast to a model in which a short-lived second ORC directs loading of the second Mcm2-7, the Cdc6 protein associated with loading the second Mcm2-7 is easily detected (23.2 s average dwell time, Fig. 7D). Third, even at diffusion-limited binding rates the sequential association of Cdc6 and Mcm2-7/Cdt1 with such a short-lived ORC is improbable. Finally, experiments showing that soluble ORC is not required for helicase
Figure 14. Proposed model for helicase loading.
Three models for how a single ORC could act to load both Mcm2-7 helicases. The initial Mcm2-7 (shown in gray) is associated with ORC/Cdc6 via its C-terminal end (Sun et al., 2013). The second Mcm2-7 is shown in yellow.
(I) Sequential model. ORC recruits a second Mcm2-7 complex via the same C-terminal interactions, leading to head-to-tail double hexamers.
(II) Mcm-Mcm model. Interactions between the two N-terminal regions of the Mcm2-7 complexes recruit the second Mcm2-7 complex to yield an N-term-to-N-term double hexamer (right).
(III) Both sides of ORC model. ORC interacts with the second Mcm2-7 using its opposite face and subsequently releases. Regardless of whether recruitment of the second Mcm2-7 complex occurred via its C- or N-terminal region this model would not yield an N-term-to-N-term double hexamer (right).
loading if ORC is pre-loaded onto DNA (Bowers et al., 2004; Fernández-Cid et al., 2013; D. Duzdevich and E. Greene, personal communication) are not consistent with a role for a short-lived second ORC.

**Recruitment of the Second Mcm2-7**

Instead of ORC and Cdc6 directly recruiting the second Mcm2-7/Cdt1 complex, our findings suggest that interactions involved in stabilizing the Mcm2-7 double hexamer mediate the recruitment of the second Mcm2-7/Cdt1. We detect these interactions prior to Cdc6 or Cdt1 release (Fig. 11), suggesting that formation of double hexamer interactions anticipates loading of the second helicase. Recent EM studies of a complex between one ORC and a head-to-head Mcm2-7 double hexamer are consistent with this hypothesis (Sun et al., 2014). Because FRET is not observed immediately upon recruitment of the second Mcm2-7, an intervening event (e.g. a Mcm2-7 conformational change or ATP hydrolysis) may be required to bring the Mcm7 N-terminal domains into close proximity. We do not know which parts of the Mcm2-7 N-terminal domains drive the proposed interactions. For simplicity, the model (Fig. 13) illustrates interactions consistent with those observed in EM studies of Mcm2-7 double hexamers (Costa et al., 2014; Sun et al., 2014). One argument against a model in which Mcm2-7 N-terminal domains drive recruitment of the second Mcm2-7 is the observation that a C-terminal mutation in Mcm3 that interferes with recruitment of the first Mcm2-7 also inhibits recruitment of the second Mcm2-7 (Frigola et al., 2013). This mutant has additional defects in Mcm2-7 ATP hydrolysis, however, which could explain a loading defect for the second Mcm2-7 (Coster et al., 2014; Kang et al., 2014; Sun et al., 2014).
Because purified Mcm2-7 complexes do not show affinity for one another in solution (Evrin et al., 2009), the first Mcm2-7 must be altered to facilitate interactions with a second Mcm2-7. A likely possibility is that ORC and Cdc6 alter the conformation of the first Mcm2-7 to facilitate these interactions (shown as separation of the Mcm2/Mcm5 N-terminal regions, Sun et al, 2013). In support of a role for Cdc6, although we observe an ORC-Mcm2-7 (OM) intermediate after the first loading event, this complex is unable to recruit a second Mcm2-7 until a second Cdc6 protein associates (OC₆M).

The model for helicase loading presented here has several advantageous features. Because Cdc6 ATPase activity is required to remove incorrectly or incompletely loaded Mcm2-7 (Coster et al., 2014; Frigola et al., 2013; Kang et al., 2014), the use of a different Cdc6 proteins to load the first and second Mcm2-7 would allow each event to be evaluated separately. More importantly, the use of Mcm2-7 N-terminal domain interactions to recruit the second Mcm2-7 ensures the establishment of a head-to-head double hexamer. This conformation is the first step in the establishment of bidirectional replication initiation and could be essential for initial DNA melting. Finally, the retention of ORC after the first loading event coupled with the release of ORC after the second loading event has the advantage of promoting the formation of double hexamers while inhibiting repeated loading of single hexamers.
EXPERIMENTAL PROCEDURES

Protein Purification and Labeling

Wild-type Mcm2-7/Cdt1 and ORC complexes were purified as described previously (Kang et al., 2014). Wild-type Cdc6 was purified as described in Frigola et al., 2013. We used a variety of protein fusions to fluorescently label ORC (Ubiquitin-GGG-Flag at the N-terminus of Orc1), Cdc6 (GST-SUMO-GGG tag at the N-terminus), and Cdt1/Mcm2-7 (Ubiquitin-GGG-Flag at the N-terminus of Mcm7 or Cdt1, and/or a SNAP-tag (NEB) at the N-terminus of Mcm4). The Ubiquitin (in vivo) and GST-SUMO (using Ulp1 protease) fusions were removed to reveal three N-terminal glycines required for sortase labeling. Sortase was used to couple fluorescently labeled peptide (DY549P1- or DY649P1-CHHHHHHHHLPETGG; referred to as SORT549 and SORT649 respectively in the manuscript) to the N-terminus of these proteins. SNAP-Surface549 (NEB, SNAP549 in the manuscript) or SNAP-Janelia Fluor 646 (SNAPJF646; Grimm et al., 2015) was coupled to SNAP-tagged Mcm2-7 (see below for these purification protocols). For Sortase labeling, peptide-coupled proteins were separated from uncoupled proteins using Complete-His-Tag Resin (Roche).

Purification and Fluorescent Labeling of Mcm2-7/Cdt1

*S. cerevisiae* (W303 background) strains yST147, yST161, yST166 or yST173 (Table 3,4) were grown to OD_{600} = 1.2 in 8 liters of YEP supplemented with 2% glycerol (v/v) at 30°C. Addition of 2% galactose (w/v) and α-factor (100 ng/mL) induced Mcm2-7/Cdt1 expression and arrested cells at G1. After 6 hours cells were harvested and
<table>
<thead>
<tr>
<th>Strains</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>yST144</td>
<td>ade2-1 trp1-1 leu2-3,112 his3-11,15 ura3-1 can1-100</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>bar1::HisG lys2::HisG pep4Δ::unmarked</td>
<td></td>
</tr>
<tr>
<td></td>
<td>his3::pSKM004 (GAL1,10-MCM2, Flag-MCM3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ura::pALS1 (GAL1,10-Cdt1, GAL4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>lys::pSKM002 (GAL1,10-MCM4, MCM5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>trp::pSKM003 (GAL1,10-MCM6, MCM7)</td>
<td></td>
</tr>
<tr>
<td>ySK100</td>
<td>ade2-1 ura3-1 his3-11,15 trp1-1 leu2-3,112/LEU2::ORC1 can1-100 MATa</td>
<td>(Tsakraklides and Bell, 2010)</td>
</tr>
<tr>
<td>(aka F1ORC1)</td>
<td>orcl::hisG lys2::pLys2Gal1,10ORC2,5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pep4::KanMX Orc4-6XGly-His6 tag(hph)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>his::p403-GAL1-10-1xFlag-ORC1,6</td>
<td></td>
</tr>
<tr>
<td>yST147</td>
<td>ade2-1 trp1-1 leu2-3,112 his3-11,15 ura3-1 can1-100</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>bar1::HisG lys2::HisG pep4Δ::unmarked</td>
<td></td>
</tr>
<tr>
<td></td>
<td>his3::pSKM004 (GAL1,10-MCM2, Flag-MCM3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ura::pALS1 (GAL1,10-Cdt1, GAL4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>lys::pST022 (GAL1,10-SNAP-MCM4, MCM5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>trp::pSKM003 (GAL1,10-MCM6, MCM7)</td>
<td></td>
</tr>
<tr>
<td>yST161</td>
<td>ade2-1 trp1-1 leu2-3,112 his3-11,15 ura3-1 can1-100</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>bar1::HisG lys2::HisG pep4Δ::unmarked</td>
<td></td>
</tr>
<tr>
<td></td>
<td>his3::pSKM004 (GAL1,10-MCM2, Flag-MCM3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ura::pALS1 (GAL1,10-Cdt1, GAL4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>lys::pSKM002 (GAL1,10-MCM4, MCM5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>trp::pST030 (GAL1,10-MCM6, UbSORT-MCM7)</td>
<td></td>
</tr>
<tr>
<td>yST163</td>
<td>ade2-1 ura3-1 his3-11,15 trp1-1 leu2-3,112/LEU2::ORC1 can1-100 MATa</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>orcl::hisG lys2::pLys2Gal1,10ORC2,5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pep4::KanMX Orc4-6XGly-His6 tag(hph)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>his::pST015(GAL1,10-3xFlag-GGG-ORC1, ORC6)</td>
<td></td>
</tr>
<tr>
<td>yST166</td>
<td>ade2-1 trp1-1 leu2-3,112 his3-11,15 ura3-1 can1-100</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>bar1::HisG lys2::HisG pep4Δ::unmarked</td>
<td></td>
</tr>
<tr>
<td></td>
<td>his3::pSKM004 (GAL1,10-MCM2, Flag-MCM3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ura::pALS3 (GAL1,10-UbSORT-Cdt1, GAL4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>lys::pST022 (GAL1,10-SNAP-MCM4, MCM5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>trp::pSKM003 (GAL1,10-MCM6, MCM7)</td>
<td></td>
</tr>
<tr>
<td>yST173</td>
<td>ade2-1 trp1-1 leu2-3,112 his3-11,15 ura3-1 can1-100</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>bar1::HisG lys2::HisG pep4Δ::unmarked</td>
<td></td>
</tr>
<tr>
<td></td>
<td>his3::pSKM004 (GAL1,10-MCM2, Flag-MCM3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ura::pALS1 (GAL1,10-Cdt1, GAL4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>lys::pST022 (GAL1,10-SNAP-MCM4, MCM5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>trp::pST030 (GAL1,10-MCM6, UbSORT-MCM7)</td>
<td></td>
</tr>
</tbody>
</table>
### Table 4. Yeast and bacterial plasmids used in this study.

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pSKM002</td>
<td>pRS307 (GAL1, 10-MCM4, MCM5)</td>
<td>(Kang et al., 2014)</td>
</tr>
<tr>
<td>pSKM003</td>
<td>pRS404 (GAL1, 10-MCM6, MCM7)</td>
<td>(Kang et al., 2014)</td>
</tr>
<tr>
<td>pSKM004</td>
<td>pRS403 (GAL1, 10-MCM2, Flag-MCM3)</td>
<td>(Kang et al., 2014)</td>
</tr>
<tr>
<td>pALS1</td>
<td>pRS305 (GAL1, 10-Cdt1, GAL4)</td>
<td>This study</td>
</tr>
<tr>
<td>pALS3</td>
<td>pRS305 (GAL1, 10-UbSORT-Cdt1, GAL4)</td>
<td>This study</td>
</tr>
<tr>
<td>pST015</td>
<td>pRS403 Gal1,10 3xFlag-UbSORT-ORC1, ORC6</td>
<td>This study</td>
</tr>
<tr>
<td>pST022</td>
<td>pRS307 (GAL1, 10-SNAP-MCM4, MCM5)</td>
<td>This study</td>
</tr>
<tr>
<td>pST030</td>
<td>pRS404 (GAL1, 10-MCM6, UbSORT-MCM7)</td>
<td>This study</td>
</tr>
<tr>
<td>pET29-Sortase-5</td>
<td>pET29-Sortase-5-6xHis</td>
<td>(Chen et al., 2011)</td>
</tr>
<tr>
<td>pET24d-Ulp1</td>
<td>pET24d-Ulp1(403-621)-6xHis</td>
<td>(Malakhov et al., 2004)</td>
</tr>
<tr>
<td>pSKM033</td>
<td>pGEX- GST-PP-FLAG-Cdc6</td>
<td>(Kang et al., 2014)</td>
</tr>
<tr>
<td>pET-GSS-Cdc6</td>
<td>pET23b-GST-SUMO-GGG-Cdc6</td>
<td>This study</td>
</tr>
<tr>
<td>pUC19-ARS1</td>
<td>ARS1</td>
<td>(Heller et al., 2011)</td>
</tr>
<tr>
<td>pUC19-A'B2</td>
<td>ARS1-A'B2</td>
<td>(Heller et al., 2011)</td>
</tr>
</tbody>
</table>
sequentially washed with 50 ml of ice-cold MilliQ water with 0.2 mM PMSF and 150 ml buffer A (50 mM HEPES-KOH pH [7.6], 5 mM MgOAc, 1 mM ZnOAc, 2 mM ATP, 1 mM DTT, 10% glycerol, 0.02% NP-40) supplemented with 0.1 mM EDTA, 0.1 mM EGTA, 0.75 M potassium glutamate (KGlut) and 0.8 M Sorbitol. The washed pellet was resuspended in approximately 1/3 of packed cell volume of buffer A containing 0.1 mM EDTA, 0.1 mM EGTA, 0.75 M KGlut, 0.8 M Sorbitol, cOmplete Protease Inhibitor Cocktail Tablet (1 tablet per 15 mL total volume; Roche) and frozen dropwise in liquid nitrogen. Frozen cells were lysed in a SamplePrep freezermill (SPEX) and the lysate was clarified by ultracentrifugation in Type 70 Ti rotor at 45 krpm for 90 min at 4°C. The supernatant was applied to 2 ml Anti-M2 FLAG resin (Sigma) pre-equilibrated in buffer A containing 0.1 mM EDTA, 0.1 mM EGTA and 0.75 M KGlut and incubated with rotation for 3 hours at 4°C. The resin was collected on a column and the flow-through was discarded. The resin was washed with 20 ml of buffer A with 0.3 M KGlut. Mcm2-7/Cdt1 was eluted with buffer A containing 0.3 M KGlut and 0.15 mg/mL 3xFLAG peptide. Note that for proteins with the UbSORT tag (the N-terminal methionine is replaced with the Ubiquitin followed by three glycines) the N-terminal ubiquitin is cleaved in the cells resulting in three glycines at the N-terminus of the tagged Mcm2-7/Cdt1 subunit. Peak fractions containing Mcm2-7/Cdt1 were pooled, and the protein was concentrated to ~ 1 mg/mL using a Vivaspin 6 centrifugal concentrator (molecular weight cutoff = 100 kDa, Sartorius) and aliquoted into 0.8 mL fractions. Starting with 8 L of cells, the yield is typically 2 mg of 95% pure Sort-Mcm2-7/Cdt1, according to SDS-PAGE.
SNAP-tagged Mcm2-7/Cdtl (Mcm2-7\textsuperscript{4SNAP}, Mcm2-7\textsuperscript{4SNAP/Cdt1\textsuperscript{SORT}}, Mcm2-7\textsuperscript{4SNAP\textsuperscript{7SORT}}) was labeled with SNAP-Surface549 (NEB), SNAP-Surface649 (NEB), or SNAP-JF646 by incubating with 1 nmol of dye at room temperature for 1 hr. For SORT-tagged Mcm2-7/Cdtl (Mcm2-7\textsuperscript{7SORT}, Mcm2-7\textsuperscript{4SNAP/Cdt1\textsuperscript{SORT}}, Mcm2-7\textsuperscript{4SNAP\textsuperscript{7SORT}}), 1 mg of Mcm2-7/Cdtl was incubated with equimolar amount of Srt5\textsuperscript{5} evolved sortase ([Chen et al., 2011], purification described below) and CaCl\textsubscript{2} was added to a final concentration of 5 mM in buffer A with 0.3 M KGlu. This was mixed with 100 nmol of peptide carrying a Sort-tag and labeled with either DY549-P1 or DY649-P1 (Dyomics), dissolved in 200 \textmu L of buffer A with 0.3 M KGlu (sequence and fluorescent labeling of the peptide is described below). The reaction was incubated at room temperature for 15 min, and then quenched with 20 mM EDTA. The net result of the sortase reaction is coupling of the fluorescently-labeled peptide to the N-terminus of the target protein with the sequence NH\textsubscript{2}-CHHHHHHHHHHHLPETG followed by the remainder of the tagged protein starting at amino acid 2.

For SNAP-tagged Mcm2-7/Cdt1, after coupling the proteins to fluorophore, the reaction was applied to a Superdex 200 10/300 gel filtration column equilibrated in buffer A with 0.1 mM EDTA, 0.1 mM EGTA, and 0.3 M KGlu. Peak fractions containing Mcm2-7/Cdt1 were pooled, aliquoted and stored at -80\textdegree C.

For SORT-tagged Mcm2-7/Cdt1, after dye-coupling, the reaction was applied to a Superdex 200 10/300 gel filtration column equilibrated in buffer A with 0.1 mM EDTA, 0.1 mM EGTA, 0.3 M KGlu, and 10 mM imidazole. Peak fractions containing peptide-
coupled Mcm2-7/Cdt1 were pooled and incubated with 0.5 mL of cOmplete His-Tag Purification Resin (Roche) pre-equilibrated in buffer A with 0.1 mM EDTA, 0.1 mM EGTA, 0.3 M KGlu, 10 mM imidazole, for 1 hour with rotation at 4°C. The flow-through was discarded and the resin was washed with 5 ml buffer A with 0.1 mM EDTA, 0.1 mM EGTA, 0.3 M KGlu and 10 mM imidazole. Peptide-coupled Mcm2-7/Cdt1 was eluted using buffer A with 0.1 mM EDTA, 0.1 mM EGTA, 0.3 M KGlu and 0.3 M imidazole. Peak fractions were pooled, aliquoted, and stored at -80°C.

Special note on handling of fluorescent dyes: to prevent photobleaching, all reactions containing fluorescent molecules were covered with aluminum foil and/or done in the dark room. Light sources on all machines (AKTA FPLC, HPLC) were turned off during preparative runs. Fractions containing fluorescently-labeled peptides and proteins were determined during previous analytical runs.

**Determining percent-labeled Mcm2-7^SNAP549/SORT649**

To determine the labeling efficiency of the SNAP-tag or sortase-labeling approaches, we purified and labeled Mcm2-7^SNAP/SORT from yST173. We imaged a standard reaction containing 0.25nM ORC, 1nM Cdc6 and 2.5nM Cdt1/ Mcm2-7^SNAP/SORT using the described protocol and monitored Mcm2-7-DNA colocalization (to ensure that we were monitoring fully assembled complexes). The DNA-associations were scored for complexes that labeled with both fluorophores or just a single fluorophore. We assumed the two labeling approaches were independent of each other and used the number of Mcm2-7 complexes that were labeled with either one
fluorophore, or both to determine the percent that were labeled by each approach. For example, the percentage of visible Mcm2-74SNAP549 complexes that also displayed 649 fluorescence revealed the percent labeling of the SORT tag by the 649 dye.

**Purification and Fluorescent Labeling of Cdc6**

This protocol is based on a previously published protocol for purification of Cdc6 (Mehanna and Diffley, 2012). Plasmid harboring GST-SUMO-Sort-Cdc6 (pET23b-GST-SUMO-Sort-Cdc6) was transformed into Rosetta 2(DE3) pLysS *Escherichia coli* strain. 2 L of cells were grown in LB supplemented with 100 μg/mL ampicillin and 25 μg/mL chloramphenicol to OD<sub>600</sub> = 0.6 at 37°C. Expression of GST-SUMO-Sort-Cdc6 was induced with 0.5 mM IPTG for 5 hours at 18°C. Cells were harvested by centrifugation and washed once with 50 ml of ice-cold MilliQ water containing 0.2 mM PMSF. The washed cell pellet was resuspended in 50 mL of buffer B (50 mM K<sub>2</sub>HPO<sub>4</sub>/KH<sub>2</sub>PO<sub>4</sub> [pH 7.5], 5 mM MgCl<sub>2</sub>, 1 mM DTT, 1% Triton X-100) with 2 mM ATP, 0.15 M potassium acetate (KOAc), cOmplete Protease Inhibitor Cocktail tablet (1 tablet per 15 mL of total volume; Roche), 100 μg/mL of lysozyme and incubated for 15 min on ice. Cells were lysed by sonication and the lysate was clarified by ultracentrifugation with a 70 Ti rotor at 45 krpm for 40 min at 4°C. The clarified lysate was incubated with 2 ml bed volume of Glutathione Sepharose 4 Fast Flow (GE Healthcare) resin pre-equilibrated in buffer B with 2 mM ATP, and 0.15 M KOAc for 3 hours while rotating in the cold room. The flow-through was discarded and the resin was washed with 40 ml of buffer B with 2 mM ATP and 0.15 M KOAc. The column flow was stopped and the resin was suspended in 2 ml of buffer B with 2 mM ATP, 0.15 M KOAc and 300 μg of Ulp1 protease. The Ulp1
protease cleaves immediately after SUMO, resulting in N-terminal Sortase recognition tag (three glycines) on Cdc6. The mixture was incubated at 4°C for 10 min with occasional swirling by hand to prevent resin settling. The flow through and two 2 ml washes with buffer B with 2 mM ATP and 0.15 M KOAc were collected, pooled and the KOAc concentration was adjusted to 75 mM by adding buffer B with 2 mM ATP. The protein solution was applied to 1 ml bed volume of hydroxyapatite ceramic (BioRad, 80 μm particle size) pre-equilibrated in buffer B with 2 mM ATP and 0.15 M KOAc. The flow-through was discarded and the column was washed with 5 ml of buffer B with 2 mM ATP and 0.15 M KOAc, followed by 5 ml of buffer C (50 mM Tris-HCl [pH 7.5], 5 mM MgCl₂, 1 mM DTT, 15% glycerol, 1% Triton X-100), and 5 ml of buffer C with 0.15 KOAc. Sort-Cdc6 was eluted using buffer C with 0.4 M KOAc. Starting with 2 L of cells, the yield is typically 2 mg of 95% pure Cdc6, according to SDS-PAGE.

1 mg of GGG-Cdc6 was mixed with equimolar amount of Srt5° and CaCl₂ was added to a final concentration of 5 mM in buffer C. This was mixed with 100 nmol of peptide carrying a Sort-tag and labeled with either DY549-P1 or DY649-P1 (Dyomics), dissolved in 200 μL of buffer C. Labeling was allowed to proceed for 2 min at room temperature. The reaction was terminated by the addition of 20 mM EDTA to a final concentration of 20 mM. We found that reaction times longer than 10 min result in Cdc6 aggregation. The terminated reaction was applied to a Superdex 75 10/300 gel filtration column equilibrated with buffer D (50 mM HEPES-KOH [pH 7.6], 5 mM MgCl₂, 1 mM DTT, 0.1 mM EDTA, 1% Triton X-100, 15% glycerol) containing 0.3 M KGlue and 10 mM imidazole. Peak fractions were pooled and incubated with 0.5 mL of cOmplete His-
Tag Purification Resin (Roche) pre-equilibrated in buffer D with 0.3 M KGlue and 10 mM imidazole for 2 hours with rotation at 4°C. The flow-through was discarded and the resin was washed with 5 ml of buffer D with 0.3 M KGlue and 10 mM imidazole. Fluorescently labeled Cdc6 was eluted using buffer D with 0.3 M KGlue and 0.3 M imidazole. Peak fractions were pooled, aliquoted, and stored at -80°C.

**Purification and Fluorescent Labeling of ORC**

yST163 cells were grown to OD\(_{600}\) = 1.2 in 8 liters of YEP supplemented with 2% glycerol (v/v). Expression of UbSORT-FLAG-ORC was induced with 2% galactose (w/v) for 3 hours at 30°C. The cells were α-factor arrested (100 ng/mL) for an additional 3.5 hours, then harvested and washed once with 50 ml ice-cold MilliQ water with 0.2 mM PMSF and once with 150 ml buffer E (50 mM HEPES-KOH [pH 7.6], 10% glycerol, 5 mM MgOAc, 1 mM ZnOAc) with 0.1 mM EDTA, 0.1 mM EGTA and 1 M Sorbitol. The washed pellet was resuspended in approximately 1/3 of packed cell volume in buffer E with 500 mM KCl, 0.1 mM EDTA, 0.1 mM EGTA, 0.01% NP-40 and cOmplete Protease Inhibitor Cocktail tablets (1 tablet per 15 mL of total volume; Roche), and frozen dropwise in liquid nitrogen. The cells were lysed with a SamplePrep Freezer/Mill (SPEX) and the lysate was clarified by ultracentrifugation in Type 70 Ti rotor at 45 krpm for 90 min at 4°C. The supernatant was applied to 2 ml Anti-M2 FLAG resin (Sigma) pre-equilibrated in buffer E with 0.1 mM EDTA, 0.1 mM EGTA, 500 mM KCl and 0.01% NP-40 and incubated with rotation for 3 hours at 4°C. The flow-through was discarded and the resin was washed with 20 ml of buffer E with 0.1 mM EDTA, 0.1 mM EGTA, 200 mM KCl and 0.01% NP-40. Sort-ORC was eluted with buffer E with 0.1 mM
EDTA, 0.1 mM EGTA, 200 mM KCl, 0.01% NP-40 and 0.15 mg/mL 3xFLAG peptide. Note that N-terminal ubiquitin is cleaved off in the cells resulting in N-terminal Sortase recognition tag on Orc1. The peak fractions were pooled and applied to 0.5 ml of SP resin (GE Healthcare) pre-equilibrated with buffer E with 0.1 mM EDTA, 0.1 mM EGTA, 200 mM KCl and 0.01% NP-40. The flow-through was discarded and the resin was washed with 5 ml of buffer E with 200 mM KCl and 0.01% NP-40. Sort-ORC was eluted using buffer E with 500 mM KCl and 0.01% NP-40. Starting with 8 L of cells, the yield is 0.5 mg of 95% pure Sort-ORC, according to SDS-PAGE.

0.5 mg of Sort-ORC was combined with equimolar amount of Srt5° and and CaCl₂ was added to a final concentration of 5 mM in buffer E. This was mixed with 100 nmol of peptide carrying a Sort-tag and labeled with either DY549-P1 or DY649-P1 (Dyomics), dissolved in 200 μL of buffer E. The reaction was incubated at room temperature for 15 min, and terminated by adding EDTA to 20 mM. The reaction mix was applied to a Superdex 200 10/300 gel filtration column using buffer E with 0.1 mM EDTA, 0.1 mM EGTA, 0.3 M KGlù and 10 mM imidazole. Peak fractions were pooled and incubated with 0.5 mL of cOmplete-His-Tag Purification Resin (Roche) pre-equilibrated in buffer E with 0.1 mM EDTA, 0.1 mM EGTA, 0.3 M KGlù, and 1 mM imidazole, for 1 hour with rotation at 4°C. The flow-through was discarded and the resin was washed with 5 ml of buffer E with 0.1 mM EDTA, 0.1 mM EGTA, 0.3 M KGlù and 10 mM imidazole. Fluorescently labeled ORC was eluted using buffer E with 0.1 mM EDTA, 0.1 mM EGTA, 0.3 M KGlù and 0.3 M imidazole. Peak fractions were pooled, aliquoted, and stored at -80°C.
Determining ORC labeling fraction

To determine what fraction of ORC molecules were fluorescently labeled, 20 μL (~4.2 μg) of labeled ORC^{5SORT549} was mixed with maleimide-DY649-P1 dissolved in anhydrous DMSO, in a 1:1 molar ratio at 4°C for 10 minutes. The reaction was terminated with 2 mM DTT. We added 0.25 nM of the double-labeled ORC to a slide containing origin DNA and monitored ORC-DNA colocalization (to ensure that we were monitoring fully assembled complexes). The fraction of maleimide-DY649-P1-labeled ORC molecules that also contained DY549-P1 was determined and reported as the percent labeling by the DY549-P1 (we assume that coupling of maleimide-DY649-P1 to ORC is not influenced by the presence or absence of the 549 label).

Preparation of Fluorescently Labeled Peptide for Sortase-tagging

All peptides were synthesized using solid-phase peptide synthesis (Koch Institute) and HPLC purified before use. For N-terminal labeling of proteins (carrying the N-terminal triglycine SORT-tag) we used a peptide with the sequence: H2N-(C)HHHHHHHHHLPETGG-COOH. We coupled fluorophore to peptide using two approaches: (1) maleimide conjugation to the cysteine residue, or (2) NHS-ester linkage to the N-terminal amine of the peptide. The 10xHis-tag is used to purify proteins coupled to the peptide using a His-tag purification column. LPETGG is the Sortase recognition tag. Maleimide- and NHS-ester labeled peptides showed no difference in experiments and were used interchangeably. However, maleimide conjugation was more efficient (always 99% yield or greater) and was our preferred method of coupling fluorophores to the peptide.
The following protocol describes fluorescent labeling of peptides (H$_2$N-CH$_{10}$LPETGG-COOH) with maleimide-DY549-P or maleimide-DY649-P (Dyomics). The reaction buffer was prepared by dissolving 9.36 g/L of Potassium Phosphate Monobasic Anhydrous and 32.73 g/L of Sodium Phosphate Dibasic Heptahydrate in ddH$_2$O and if necessary the pH was adjusted to 7.0. The buffer was filtered using vacuum filtration (0.2 μm pore size) and placed in a Buchner flask with a stirring bar. The buffer was vacuum degassed for 1 hour at room temperature while stirring. After degassing, N$_2$ gas was bubbled through the buffer solution.

3 μmol of peptide was dissolved in 2 mL of degassed buffer. 2-3 fold molar excess of peptide relative to the maleimide dye was sufficient to drive the reaction to completion. To reduce disulfide bonds between peptides, 0.3 mL of ethanethiol (Sigma) was added to the dissolved peptide. Although poorly soluble in water (0.251 M solubility in water), this amount will reduce all disulfide bonds. Ethanethiol was removed by evaporation using a speedvac for 90 min. After removal of ethanethiol, the dissolved peptide was transferred into a fresh 5 mL tube and quickly mixed with 1 μmol of maleimide-linked fluorophore dissolved in 0.5 mL of anhydrous DMSO (Sigma). The tube was covered with aluminum foil and rotated at room temperature for 2 hours. The resulting mixture was aliquoted into ten 2 mL eppendorf tubes and lyophilized overnight.

Fluorescently-labeled peptide was separated from unlabeled peptide using HPLC. We added 0.1 M DTT to the peptide mixture before HPLC separation to improve the
separation of labeled peptide from unlabeled peptide and oxidized peptide dimers. The peptide mixture was separated using a Agilent Zorbax SB-C18 column at a flow rate of 1ml/min using a gradient from 0 to 80%acetonitrile containing 0.06%TFA. Fractions containing fluorescently labeled peptide were pooled and lyophilized overnight. The labeled peptide was dissolved in 200 µl of sortase reaction buffer (50 mM Hepes-KOH pH 7.6, 0.3 M potassium glutamate, 5 mM MgCl₂, 10% glycerol) immediately before the labeling reaction.

Fluorescent labeling of peptide (H₂N-H₁₀LPETGG-COOH) using NHS-ester-DY549-P1 or NHS-ester-DY649-P1 used a similar protocol with the following changes. The reaction was performed in 0.1 M sodium bicarbonate buffer pH 8.3, that was filtered before use (0.2 µm pore size). 10 µmol of peptide was dissolved in 3 mL of the reaction buffer, and mixed with 1 µmol of the NHS-ester dye dissolved in 1 mL anhydrous DMSO. The reaction was incubated at room temperature for 2 hours while rotating, aliquoted into ten 2 mL eppendorf tubes and lyophilized overnight. Fluorescently labeled peptide was isolated following the same protocol as for the maleimide-labeled peptide.

**Purification of Sortase A pentamutant**

Sortase A P94R/D160N/D165A/K190E/K196T (Srt5°) mutant bearing a C-terminal 6xHis-tag was evolved, purified and characterized previously (Chen et al., 2011). pET29-Srt5° was transformed into BL21(DE3) *Escherichia coli* strain, grown in 1 L of LB with 100 µg/mL kanamycin to OD₆₀₀= 0.6 at 37°C and the expression of Srt5° was induced with 0.4 mM IPTG for 3 hours at 30°C. Cells were pelleted by
centrifugation and washed once with 50 mL of ice-cold MilliQ water. The washed pellet was resuspended in 50 mL of buffer F (50 mM Tris-HCl pH [7.9], 300 mM NaCl, 5 mM MgCl₂, 10% glycerol, 0.1% NP-40) with 15 mM imidazole and cOmplete Protease Inhibitor Cocktail tablet (1 tablet per 15 mL of total volume; Roche) and the cells were lysed by sonication. The lysate was cleared by ultracentrifugation in a 70 Ti rotor at 35 krpm for 30 min at 4°C and the supernatant was incubated in the cold room with 2 mL bed volume of cOmplete His-tag Purification Resin (Roche) pre-equilibrated in buffer F with 15 mM imidazole for 2 hours while rotating. The flow-through was discarded and the column was washed with 40 mL of buffer F with 10 mM imidazole. Srt5⁰ was eluted using buffer F with 0.35 M imidazole. Peak fractions containing Srt5⁰ were pooled and applied to pre-equilibrated Superdex 75 16/600 gel filtration column using buffer G (50 mM HEPES-KOH pH [7.6], 200 mM KCl, 5 mM MgCl₂, 1 mM DTT, 10% glycerol, 0.02% NP-40). Peak fractions containing Srt5⁰ were pooled, aliquoted, and stored at -80°C. 1 L of cells typically yields 10 mg of Srt5⁰, 95% pure according to SDS-PAGE.

**Purification of Ulp1**

A catalytically active fragment (amino acids 403-621) of Ubl-specific Protease 1 (Ulp1) that is C-terminally tagged with 6xHis was described and characterized previously (Malakhov et al., 2004). pET24d-Ulp1(403-621)-6xHis was transformed into BL21(DE3) Escherichia coli strain. 3 L of cells were grown to OD₆₀₀ = 0.6 at 37°C and the expression of Ulp1 was induced with 0.5 mM IPTG for 4 hours at 30°C. Cells were pelleted by centrifugation and washed once with 50 mL of ice-cold MilliQ water. The washed pellet was resuspended in 50 mL of buffer H (50 mM Tris-HCl pH 8.0, 0.3 M NaCl, 2 mM...
DTT, 10% glycerol, 0.1% NP-40) with 10 mM imidazole and cOmplete Protease Inhibitor Cocktail tablet (1 tablet per 15 mL of total volume; Roche) and the cells were lysed by sonication. The lysate was cleared by ultracentrifugation in a 70 Ti rotor at 35 krpm for 30 min at 4°C. The supernatant was incubated with 2 mL bed volume of cOmplete His-tag Purification Resin (Roche) pre-equilibrated in buffer H with 10 mM imidazole for 2 hours while rotating. The flow-through was discarded and the column was washed with 40 mL of buffer H with 10 mM imidazole. Ulp1 was eluted using buffer H with 0.35 M imidazole. Peak fractions containing Ulp1 were pooled and applied to pre-equilibrated Superdex 75 16/600 gel filtration column using buffer I (50 mM HEPES-KOH pH 7.6, 100 mM KCl, 2 mM DTT, 10% glycerol, 0.02% NP-40). Peak fractions containing Ulp1 were pooled, aliquoted, and stored at -80°C. 3 L of cells typically yielded 8 mg of Ulp1, 95% pure according to SDS-PAGE.

**Single-Molecule Microscopy**

The micro-mirror total internal reflection (TIR) microscope used for multiwavelength single-molecule using excitation wavelengths 488, 532, and 633 nm has been previously described (Friedman and Gelles, 2012; Friedman et al., 2006). Biotinylated AlexFluor488-labeled 1.3kb-long DNA molecules containing an origin were coupled to the surface of a reaction chamber through streptavidin. Briefly, the chamber surface was cleaned and derivatized using a 200:1 ratio of silane-NHS-PEG and silane-NHS-PEG-biotin (see Experimental Procedures). We identified DNA molecule locations by acquiring 4-7 images with 488 nm excitation at the beginning of the experiment. Unless otherwise noted, helicase loading reactions contained 0.25nM ORC, 1nM Cdc6
and 2.5nM Cdt1/Mcm2-7. Reaction buffer was as previously described (Kang et al., 2014) except without any glycerol and with the addition of 2 mM dithiothreitol, 2 mg/ml bovine serum albumin (EMD Chemicals; La Jolla, CA), and an oxygen scavenging system (glucose oxidase/catalase) to minimize photobleaching (Friedman et al., 2006).

After addition of protein to the DNA-coupled chamber, frames of one second duration were acquired according to the following protocol: (1) a single image frame visualizing the DNA positions (488 excitation), (2) 60 frames monitoring both the 549 and 649 fluorophores (simultaneous 532 and 633 excitation) and (3) a computer-controlled focus adjustment (using a 785 nm laser). This cycle was repeated roughly 20 times in the course of an experiment (~20 mins). Chambers were then washed with either three chamber volumes of reaction buffer or two volumes of the same buffer with 0.5M NaCl in place of 300mM K-glutamate and one volume reaction buffer. For photobleaching, lasers power(s) were increased and one or multiple fluorophores were imaged simultaneously until no visible spots remained. Typically, photobleaching was also examined in a second field of view that was not imaged during the loading reaction.

**Slide Preparation**

Microscope slides for single molecule studies were prepared using methods similar to those previously described (Gandhi et al., 2010). Specifically, we cleaned glass slides and glass coverslips using sonication in four sequential solutions (1 hr in 2% Micro-90, 1hr 0.1M KOH, 1hr 100% EtOH and 15 min deionized water in a Fisher Scientific FS30H bath sonicator). The cleaned slides and coverslips were covered with 20mg/mL mPEG-Silane-2000 and 0.1 mg/ml biotin-mPEG-Silane-3400 (Laysan Bio)
(1:200 ratio) dissolved in 80% ethanol, pH 2, 70°C for 2 hours (100 μL for the coverslip, 200 μL for the slide), resulting in the PEGylation of their surfaces. After this incubation, the slides were washed with 10 mM Tris-HCl pH 7.0, dried, and incubated two more times (20 mg/ml mPEG-silane (no biotin-mPEG) in 80% ethanol, pH 2, 70°C for 2 hours. The slides were washed and dried between these incubations and at the end with 10 mM Tris-HCl pH 7.0. The PEG-treated slides were stored under N₂ gas at -80°C and used in experiments the following day. Flow chambers were constructed using high vacuum grease (Dow Corning) to separate the slides and the coverslips and to delineate flow chamber lanes. Flow chamber volumes were typically ~20μL.

**Data Analysis Procedures**

Analysis of the CoSMoS data sets was similar to (Hoskins et al., 2011). Specifically, we typically followed these four steps: (1) defining the spatial relationship between the two images created at different excitation/emission wavelengths from the single field of view by the dual-view optical system ("mapping"), (2) correcting the data set for stage drift that occurred during the experiment ("drift correction"), (3) imaging the label on origin-DNA to identify the locations of single DNA molecules on the surface, and (4) integration of fluorescence emission from small regions centered at the pre-defined locations of coupled DNA locations in each acquired image to obtain plots of fluorescence intensity vs. time. These steps were carried out using custom image-processing software implemented in MATLAB [The Mathworks (Natick, MA)].
Both the dual imaging optics and chromatic aberrations result in spatial displacement between fluorescent spot images of co-localized species that are labeled with different color dyes. Accurate co-localization of the differentially-labeled species therefore requires use of a mapping procedure. For each pair of colors a list of several hundred reference spot pairs were collected using a sample containing a surface-tethered oligonucleotide that was labeled with Alexa488, Cy3 and Cy5. Mapping the coordinates of a fluorescent spot to the equivalent location at a different color was performed using a transformation with fit parameters based on just the 15 nearest reference spots (Crawford et al., 2013).

Drift correction was carried out using fluorescent protein spots that had a duration of at least 50 continuous frames during the experiment. Multiple spots were chosen so as to cover the image sequence at least three times. Each spot image was fit to a Gaussian function to determine its center during every frame of the image sequence in which the spot was tracked. The changes in spot position between successive image frames were averaged for the visible tracked spots, and this described the stage drift over the experimental time course. That measured stage drift was smoothed and then used to correct the positions of all spot images recorded throughout the experiment.

location was manually inspected to verify that the software correctly identified the center of each spot and that no spurious signals were incorrectly identified (for example edges of the field of view or background noise) as DNA spots. Furthermore, in cases where stoichiometry of the proteins was determined, the photobleaching of the identified DNA spots was monitored to ensure that a single DNA molecule was present in each analyzed area.

Fluorescence emission from labeled complexes was integrated over a 0.37 μm² area centered at each drift corrected origin-DNA location, yielding for each DNA molecule a separate intensity time course for each color of fluorescent label being observed.

**Fitting to single and double exponential distribution**

Fitting of dwell times was performed using maximum likelihood algorithms and bootstrap methods were used to determine uncertainty estimates, similar to what has been previously described (Crawford et al., 2013; Friedman and Gelles, 2012).

**FRET experiments**

The conditions for monitoring FRET were similar to the other experiments, with a few exceptions. Typical reactions contained 0.75 nM ORC, 3nM Cdc6, 5nM Cdt1/Mcm2-7^{T54} and 5nM Cdt1/Mcm2-7^{T64}. DNA was imaged before and immediately after adding the reaction to the slide but not throughout the experiment. The imaging protocol alternated between 1 s frames with the 532 laser on and 1 s frames with
the 633 laser on over 20-30 minutes. Apparent $E_{\text{FRET}}$ was calculated as described (Crawford et al., 2013).

**FRET data analysis**

Images containing spots that were analyzed to produce a FRET time course were first mapped and drift-corrected (see above). By alternating between their laser excitation wavelengths we monitored the co-localization of both the donor and acceptor-labeled Mcm2-7 hexamers with the origin-DNA molecule. To determine the time until formation of the high-FRET state, we noted the earliest time when both of the labeled Mcm2-7 hexamers were present and subtracted that value from the time when the >635 nm emission FRET signal rose above background noise level (as determined by monitoring the signal in the >635 nm field when only the 532 laser was turned on). Spots were selected manually based on FRET signal that was above background noise and contained only one of each fluorophore.

To calculate apparent FRET efficiencies, fluorescence intensity traces were background subtracted using custom Matlab (MATHWORKS) image processing software that has been previously described (Crawford et al., 2013). FRET efficiency was calculated using $E_{\text{FRET}} = \frac{l_{\text{Acceptor}}}{l_{\text{Acceptor}} + l_{\text{Donor}}}$ where $l_{\text{Acceptor}}$ and $l_{\text{Donor}}$ are the acceptor and donor emission intensities observed during donor excitation, respectively. No gamma correction was applied because no systematic change in $(l_{\text{Acceptor}} + l_{\text{Donor}})$ was observed upon changes in $E_{\text{FRET}}$ (Fig. 6A, S6) or upon acceptor photobleaching. For each molecule, $E_{\text{FRET}}$ values were reported for only the first 10 frames captured (~27s) after
both fluorescently-labeled Mcm2-7 complexes associated with origin-DNA.
REFERENCES


Takara, T.J., and Bell, S.P. (2011). Multiple Cdt1 molecules act at each origin to load replication-competent Mcm2–7 helicases. EMBO J. 30, 4885–4896.


Chapter III

Further Discussion and Future Directions
Key Conclusions

The work described in this thesis focuses on the mechanism of eukaryotic helicase loading. In Chapter II, I described novel assays that I used to analyze the steps of helicase loading in a single-molecule setting. These assays allowed us to determine the relative stoichiometry of the various helicase-loading proteins that are required to load a head-to-head Mcm2-7 double hexamer. Whereas a single ORC complex directs the loading of both helicases, two separate Cdc6 and Cdt1 helicase-loading proteins are required. Our findings show that dissociation of each of these proteins occurs in an ordered fashion. Additionally, I presented evidence that indicates that loading of the two helicases in the final double hexamer occurs sequentially and with differing kinetics. Interestingly, I found that recruitment of the second helicase is quickly followed by formation of the interactions that mediate the tight N-terminal-to-N-terminal double hexamer. Collectively these data support a sequential loading model in which a single ORC complex loads both helicases in the double using distinct mechanisms (Figure 13, Chapter II).

The mechanism of helicase loading in S. cerevisiae displays similarities to DnaB loading in E. coli. A single, asymmetrically bound ORC is sufficient to load two Mcm2-7 helicases. Although multiple DnaA proteins form a directional filament, this oligomer exists on one side of the DUE and establishes the orientation of the loaded DnaB hexamers (see Figure 8, Chapter I; reviewed in Duderstadt and Berger, 2008). Although ORC is a single complex, it is formed of multiple AAA+ related proteins that can be considered analogous to the DnaA oligomer. My finding that one asymmetric ORC directs loading of both helicases is similar to the asymmetric location of the DnaA
The kinetic differences observed in loading of the first versus the second helicase as well as the architecture of the initial OC₆C₁M (Sun et al., 2013; Figure 4, Chapter I) lead us to propose that the two eukaryotic helicases are recruited and loaded by distinct mechanisms. DnaB is also thought to be recruited and loaded by distinct mechanisms: (i) one helicase is recruited in the opposite orientation through DnaA-DnaC-DnaB interactions and (ii) the other helicase is proposed to be recruited through DnaA-DnaB interactions.

Many questions remain about the mechanism of eukaryotic helicase loading. Although we described several release steps throughout the process of helicase loading, what triggers Cdc6, Cdt1 and ORC release remains unknown. Additionally, when each Mcm2-Mcm5 gate closes around dsDNA and what controls this event remains to be understood. If, as I propose, the second Mcm2-7 is recruited through interactions with the previously loaded Mcm2-7 then it is unclear why a second Cdc6 is required for association of this second Mcm2-7. In the following paragraphs I will focus on (a subset of) these questions and how they could be addressed.

**The Mcm2-Mcm5 gate**

The developed single-molecule assays offer a unique opportunity to analyze when during helicase loading the Mcm2-Mcm5 gate is opened/closed and when this event occurs relative to Cdc6 and Cdt1 release. We have recently developed an assay that uses FRET between Mcm2 and Mcm5 (each subunit is labeled internally using a SNAP/CLIP fusion) to monitor the status of the gate during helicase loading. We find that the gate is
in an open conformation upon initial association with ORC/Cdc6 at origins, and closes after a short delay (data not shown). Quantitative analysis shows that this delay is correlated with the release of Cdt1 in the reaction. If I tether Mcm2-7 complexes in the presence or absence of Cdt1, both populations exist in a predominantly open conformation, suggesting that the presence of Cdt1 alone is not sufficient to open the gate. This contrasts with evidence in *E. coli* that suggests that DnaC stabilizes an open form of DnaB (Arias-Palomo et al., 2013).

Although we observe a correlation between Cdt1 release and gate closing, it remains unknown whether one event causes the other to occur or whether a third event (such as release of Cdc6 or ATP hydrolysis by one of the Mcm2-7 interfaces) is responsible for both. Interestingly, mutants in the ATPase motif formed between Mcm2 and Mcm5 have been implicated in ring opening/closing. Studies of the Mcm2-7 complexes containing either a Mcm5 Walker B mutant or the corresponding Mcm2 arginine-finger mutant proposed that they are in a constitutively open and closed state, respectively (Bochman and Schwacha, 2008). Analyzing these mutants in the newly developed assay for the status of the Mcm2-5 gate during helicase loading may offer insights into the role of this Mcm2-7 ATPase site during ring opening/closing. Such studies also have the potential to explain the loading defects that were observed for these mutants in bulk (Kang et al., 2014; Coster et al., 2014).
The role of ORC and Cdc6 ATP hydrolysis in helicase loading

Although ORC and Cdc6 ATP hydrolysis mutants are lethal in vivo (Bowers et al., 2004; Perkins and Diffley, 1998; Weinreich et al., 1999; Chang et al., 2015), the mechanism behind this phenotype is largely unknown. In fully reconstituted in vitro assays, both of these mutants are capable of loading Mcm2-7 double hexamers at the same levels as wild type (Kang et al., 2014; Evrin et al., 2013). With the Cdc6 ATPase mutant, a defect in release of Cdc6 has been reported (Kang et al., 2014). It is possible that the high protein concentrations used and the asynchrony of the bulk assay obscure the defects of the mutant proteins. Additionally, the bulk assay only analyzes end products, which could hide kinetic defects.

There are several steps during helicase loading at which these mutant proteins may be defective. One possibility is that the lower protein concentrations used in the single-molecule assay will reveal kinetic defects of these mutants in the helicase loading reaction. One hint that proteins concentrations can affect the helicase loading reaction is that Cdc6 ATP hydrolysis mutants loaded significantly fewer Mcm2-7 helicases, relative to wild type, when the concentration of Cdc6 was lowered threefold (Kang et al., 2014). If concentrations are not the only reason why the ORC and Cdc6 ATP hydrolysis mutants show no defects in vitro, other steps of the reaction may be kinetically slower, for example the sequential release of Cdc6 or Cdt1. ORC ATP hydrolysis may alternatively be important for multiple rounds of double hexamer loading (Bowers et al., 2004). If the bulk conditions are such that only one round of double hexamer loading occurs, no difference in the amount of Mcm2-7 helicases loaded would be observed.
To address whether the Cdc6 or ORC ATPase mutants alter specific steps in helicase loading, it would be informative to label each of these ATP hydrolysis mutant proteins and monitor them in the single-molecule assay. Pairwise labeling of different proteins may reveal a Cdc6 or Cdt1 release defect. In fact, the release of Cdc6 and Cdt1 are likely impaired in the presence of the Cdc6 ATP hydrolysis mutant (Fernandez-Cid et al., 2013; Kang et al., 2014). The time for Cdc6 and Cdt1 release can be analyzed in both the ORC and Cdc6 ATP hydrolysis mutants. To determine whether the ORC ATP hydrolysis mutant is defective for multiple round of loading, photobleaching of high-salt resistant Mcm2-7 complexes can be used to count the number of hexamers loaded by wild type and ORC ATPase mutants. In this case, it would be informative to optimize the single-molecule assays to load three and four Mcm2-7 complexes per DNA molecule (Figure 2, Chapter II).

The role of the second Cdc6 during helicase loading

The consistent association of a second Cdc6 molecule prior to recruitment of the second helicase suggests that these events are coupled (see Chapter II, Figures 5 and 6), yet the role that the second Cdc6 molecule plays remains unknown. Although our studies do not rule out that in vivo two ORC molecules direct helicase loading, they show that a single ORC is sufficient to load a double hexamer. Cdc6 is proposed to change the conformation of ORC, making it competent to associate with Cdt1/Mcm2-7 (Sun et al., 2012). However, once the first Mcm2-7 is loaded, the second Cdc6 would either (i) have to transmit this conformational change through the first loaded Mcm2-7 (Figure 1;
Figure 1. Models for how the second Cdc6 recruits a second Mcm2-7 complex
See text for details.
conformational change) or (ii) allow a different interface to recruit the second Mcm2-7 hexamer (Figure 1; different interface).

I favor the conformational change model for multiple reasons. First, it is hard to imagine how the different interface model would lead to the N-terminal-to-N-terminal double hexamer, regardless of where this second association occurs or in what orientation. Additionally, this model would likely require release or significant remodeling of the interaction with the first Mcm2-7. For the first Mcm2-7 to recruit the second, it would need to exist in a novel conformation that allows stable interactions with a second Mcm2-7. However, the conformational change model does not explain why the C-terminus of Mcm3 is required for recruitment of both the first and second Mcm2-7 hexamer (Frigola et al., 2013). In the co-axial arrangement of (ORC/Cdc6):(Mcm2-7)$_1$:(Mcm2-7)$_2$ shown in the conformational change model, the C-terminal ends of the second Mcm2-7 are roughly 230 Å away from ORC/Cdc6 (inferred from the size of the double hexamer; Remus et al., 2011).

These two models for one ORC loading both helicases could be distinguished using a single-molecule FRET-based assay. Analyzing both where Cdc6 associates relative to ORC and relative to Mcm2-7 would inform whether the second Cdc6 associates with ORC in the same way as the first and where the second Mcm2-7 is recruited. Structural data suggests that Cdc6 associates with ORC between the Orc1 and Orc2 (Bleichert et al., 2015; Sun et al., 2013). Based on this information, one could fluorescently tag Cdc6 and Orc1 or Orc2 to detect the predicted interaction of ORC with
Cdc6 by FRET. If the second Cdc6 molecule associates with ORC in the same manner as the first, then one would expect similar FRET efficiencies between ORC and the first and second Cdc6. Structural data suggests that the C-terminus of Mcm7 is near the C-terminus of Cdc6. Since we have shown that the second Cdc6 associates before the second Mcm2-7, we can ask whether the first and second Cdc6 form the same interactions with the first helicase. If so, this would support a model in which the first loaded helicase maintains its original interactions with ORC/Cdc6. If the first and second Cdc6 do not form the same interactions with the first helicase, it would suggest that there is a different binding site for Cdc6. The abundance of structural data available (Bleichert et al., 2015; Li et al, 2015; Sun et al, 2013) will greatly facilitate the identification of sites to label ORC, Cdc6 and Mcm2-7 subunits with fluorophores for these proposed FRET studies.

**Extending the single-molecule assay to replication initiation**

The same approaches described in Chapter II can be used to analyze the downstream steps of replication initiation. Using similar approaches one could define the order of protein association and dissociation, relative protein stoichiometries and relative proximity (using FRET). Similar to the situation before my studies of helicase loading, bulk assays have determined the requirements of each protein for the association of the remaining proteins with the DNA (Heller et al., 2011; Yeeles et al., 2015), nevertheless, several questions remain open. The stoichiometries of some of these helicase activation proteins are unknown. Although it appears that the activation of both helicases is coordinated (Duzdevich et al., 2014), what enables this coordination or at what point it is
established remains unknown. Finally, the protein requirements for and timing of the major events remodeling the loaded helicases and the associated DNA that must occur during helicase activation are unknown. As with the ensemble helicase-loading assays, the current bulk assays for helicase activation are both asynchronous and occur at low efficiency, making these questions difficult to address.

The single-molecule approach taken here to dissect the mechanism of helicase loading can be adapted to analyze the proteins involved in helicase activation. The single-molecule approach will also address whether helicase activation is ordered or occurs simultaneously and the proteins that might order these events. These proteins can be labeled pair-wise to determine stoichiometry as well as order of association and dissociation. This type of analysis will likely uncover new steps during helicase activation, similar to the sequential release steps of Cdc6 and Cdt1 that were observed in this thesis. Additionally, stoichiometric information and order of release can distinguish steps that occur simultaneously for both helicases and those that do not, providing insights into the mechanism of coordinated helicase activation.

The two main transitions that remodel the inactive double hexamer to the active helicase are (i) separation of the two helicases and (ii) opening and closing of each Mcm2-7 ring to allow extrusion of the lagging strand template and closure around the leading strand template. Both of these transitions can be monitored using the FRET-assays that we have already developed. Instead of monitoring the formation of a high FRET state due to formation of the Mcm2-7 double hexamer, we can ask what helicase
activation proteins are required to drive the separation of the two helicases and when this occurs relative to the arrival/departure of each helicase activating protein. Similarly, instead of monitoring when the Mcm2-5 gate closes, we can instead monitor the loss of a high FRET state due to gate opening and the subsequent reappearance of the high FRET state due to reclosing of the helicase around ssDNA. The proteins/events required for each of these transitions, as well as their relative order of remains unknown.
REFERENCES


