Impact of DNA Accessibility and Sequence on Eukaryotic Origin Licensing

by

Megan D. Warner

B. A. Chemistry
Pomona College, 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

June, 2017

Signature of Author: Signature redacted
Department of Biology
May 26, 2017

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

Impact of DNA Accessibility and Sequence on Eukaryotic Origin Licensing

by

Megan D. Warner

Submitted to the Department of Biology
on May 26, 2017 in partial fulfillment of the
requirements for the Degree of Doctor of
Philosophy in Biology

ABSTRACT:

Eukaryotic DNA replication is a complex process that requires the coordination of many different proteins to ensure duplication of the genome occurs rapidly and only once during each cell cycle. Establishing a pair of replication forks moving apart from a single starting location is essential for replicating the genome. The first step in this process, termed origin licensing, occurs during G1. During origin licensing, two hexamers of the eukaryotic helicase, Mcm2-7, are loaded at potential origins of replication in a head-to-head orientation. They remain together, inactive, until entry into S phase. Origin licensing depends on three proteins, the origin-recognition complex (ORC), Cdc6, and Cdt1, as well as the DNA sequence present at potential origins. Although in *S. cerevisiae,* ORC binds to a known DNA sequence, where the other proteins, particularly the Mcm2-7 helicase, interact with DNA is not known. To gain additional insights into the initial placement of the Mcm2-7 complex, we used a method to create DNA-protein roadblocks at defined positions within origin DNA. Because both ORC and Mcm2-7 must encircle DNA, these roadblocks prevent stable DNA association.

When a single roadblock was used, only defects in ORC-DNA binding were observed, despite that the Mcm2-7 helicase is known to encircle DNA. We hypothesized that sliding of intermediates in the origin licensing process could explain this discrepancy. When we tested this hypothesis by placing an additional roadblock to prevent sliding, severe defects in stable Mcm2-7 loading, but not ORC recruitment, were observed. Interestingly, in the context where sliding was inhibited, the observed helicase loading was dependent on a particular sequence motif, the B2 element. This sequence is important for origin function *in vivo,* but a role in origin licensing *in vitro* has not previously been reported. Placing roadblocks around origin DNA was reminiscent of the nucleosome architecture present *in vivo.* These natural roadblocks behaved similarly to our artificial roadblocks, and origin licensing was dependent on B2 in both instances. These data suggest that origin licensing is a dynamic process and that the B2 sequence element is important when sliding of origin licensing intermediates is restricted.

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

Acknowledgements:

There are many people that contributed to this document in some form of another. I would first like to thank my thesis advisor Steve, who has taught me how to do careful science, and has been instrumental in helping to design experiments, and write this dissertation. I would also like to thank Ishara Azmi and Sukhyun Kang for their help with key experiments.

I would like to thank all the members of the Bell Lab, past and present, who make the lab a wonderful place to work.

I would like to thank my thesis committee members, Iain Cheeseman and Tania Baker, who have given advice and support as I learned what it means to be a scientist. I would also like to thank Jeff Gelles and Larry Friedman, who have also supported my work.

I would finally like to thank my wonderful family and friends for their tireless support. Especially my husband, Fowler, who made sure I actually ate lunch and dinner throughout the process of writing my thesis. I don't know what I would have done without you.

Table of Contents

Chapter I:

Introduction

1. Introduction: Why Study DNA replication?

Your body is made up of trillions of cells. Yet life begins much more simply: as a single cell. Incredibly, the genomic DNA in this cell includes all the information needed to develop into a human adult, with its many diverse tissues. As that single cell divides, all of the DNA must be propagated to each daughter cell. Copying the entire genome ensures that regardless of any cell's eventual fate within the body, it too will contain all the necessary information to carry out its function. The process of duplicating the genome is termed DNA replication.

Copying genomic DNA is no small feat. DNA replication needs to be accurate, but at the same time must occur rapidly. Genomes are relatively large: even a small genome like that of *E. coli* contains around five million base pairs of DNA. Humans, as complex, multicellular organisms, contain two complete copies the genomic DNA per cell, amounting to around six billion base pairs in total. Yet cells must divide at a relatively fast rate if a single cell is to develop into an organism, or if an *E. coli* population is to evolve.

The replicative DNA helicase is at the heart of all DNA replication events. Regardless of whether the genomic DNA is composed of a single, circular chromosome (as is observed in most bacteria) or multiple linear chromosomes (in eukaryotes), copying of the genetic material cannot occur without the DNA unwinding catalyzed by this enzyme. DNA molecules consist of two complementary strands in a double-helical structure, each of which contains the information necessary for duplication. However, for this information to be accessed, the two strands must unwind, exposing the DNA bases required for nascent strand synthesis.

Although the DNA helicase's primary job is to unwind the genome, considerable preparation is necessary before this goal can be achieved. The regulation of this process is accomplished differently in prokaryotes and eukaryotes. In prokaryotes, careful coordination occurs between cell growth and initiation of DNA unwinding to ensure replication only occurs when cells are prepared for cell division. In eukaryotic organisms, cells prepare for synthesizing new DNA during G1, which occurs prior to S phase, the cell-cycle period when actual DNA

synthesis occurs (Figure 1). During G1, the eukaryotic helicase, Mcm2-7, is placed onto specific sites along the genomic DNA. The location where these helicases are placed determines where DNA unwinding and nascent DNA synthesis will begin. For this reason, these sites are termed origin of replication and assembling the helicase at these sites is called origin licensing. This process requires careful coordination between three additional proteins, the origin recognition complex (ORC), Cdc6, and Cdt1. ORC recognizes eukaryotic origins of replication, and sequential recruitment of Cdc6 and an Mcm2-7/Cdt1 complex loads two Mcm2-7 hexamers at each origin. The end result of origin licensing is a stable yet inactive Mcm2-7 double hexamer in a head-to-head orientation. This orientation poises the two helicases to leave the origin in opposite directions, and thus represents the first marker of the bidirectional nature characteristic of replication initiation. At the G1 to S phase transition, dramatic increases in the levels of a critical class of kinases (S-CDKs/DDK) both activate the loaded helicases and inhibit placement of new Mcm2-7 complexes onto the genomic DNA. Because these kinases remain at high levels until cell division, this regulation ensures that helicases can only be loaded and activated once per cell cycle. Thus, DNA replication initiation can occur once and only once per cell division.

Disruption of the normal process of DNA replication can have serious consequences in both bacteria and eukaryotes. Inaccurate, incomplete, or excessive genome replication are all detrimental to a cell. Many heritable diseases can be traced back to a single base pair change in the genome. More large-scale changes can result when incomplete or additional DNA replication occurs within a cell. In the case of incomplete replication, or under-replication, chromosomes are still connected during mitosis. For the daughter cells to separate the interconnected DNA, the DNA must break, resulting in damage and often permanent loss of genetic information.

Although the reasoning is less straightforward, over-replication is also detrimental for the cell. Re-initiation at origins causes double-strand breaks if the entire genome is not duplicated again. Repair of these breaks can lead to gene amplification, a phenomenon frequently

Figure 1

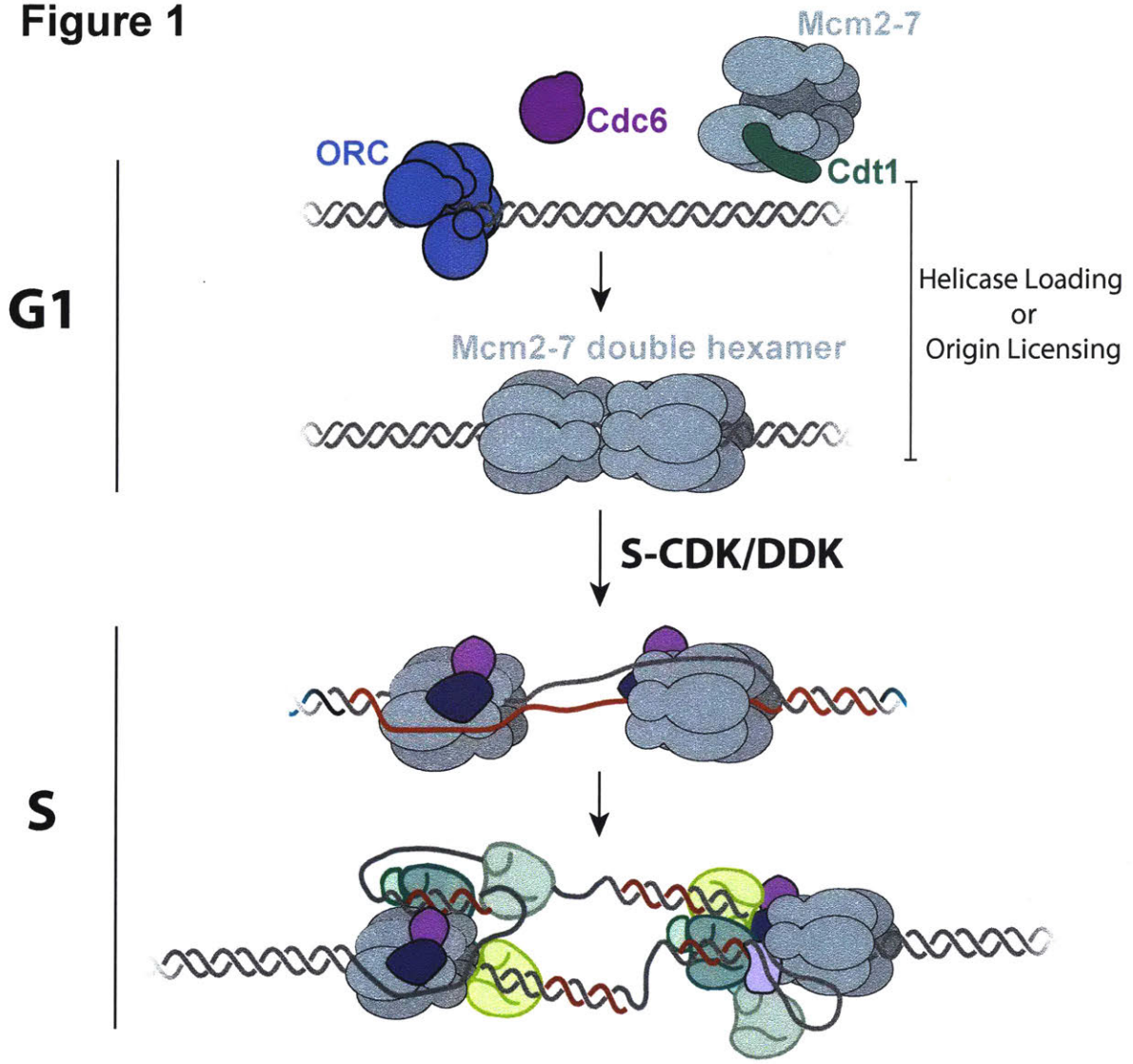


Figure 1: An overview of Eukaryotic DNA replication

Although the events of DNA unwinding and nascent DNA synthesis occur during S phase, the DNA replication process begins during the preceding G1. During G1, the eukaryotic helicase (Mcm2-7) is placed at multiple discrete locations within the genome termed origins of replication. This process, termed helicase loading or origin licensing, requires the help of three proteins – ORC, Cdc6, and Cdt1. The Mcm2-7 helicase is loaded as a double hexamer and remains inactive until changes in levels of the S-CDK and DDK kinases activate them in S phase.

associated with cancerous transformation (Green et al., 2010). Indeed, two replication initiation factors, Cdt1 and Cdc6, are both oncogenes (Arentson et al., 2002; Liontos et al., 2007; Seo et al., 2005). Extra copies of genetic information also have detrimental effects on cell biology (Torres et al., 2007). Excessive replication initiation also reduces limiting replication factors or substrates, such as RPA or dNTPs, which can result in DNA damage (Aye et al., 2015; Mantiero et al., 2011; Toledo et al., 2013).

Since mistakes in the genomic DNA are so detrimental, DNA replication is carefully regulated to minimize these errors. Although single-base-pair mistakes can often be fixed after replication via DNA repair mechanisms, large scale changes caused by under or over-replication of the genome are harder to fix post-hoc. Thus, cells carefully regulate replication initiation so helicase loading cannot after new DNA synthesis at origins. Eukaryotic cells achieve this regulation by limiting replication to once and only once per cell division. Although bacteria, such as *E. coli,* can initiate replication of their genomes more than once per cell division during periods of exponential growth, multiple mechanisms exist to regulate the rate of initiation to ensure each daughter cell gets only a single copy of the genome (for more detail, see section 6.1).

This thesis focuses on the critical first step in the DNA replication process: origin licensing. Although a detailed model for this process exists, the location where the Mcm2-7 complex loads within origin DNA remains unclear. In this introduction, I will introduce our current understanding of helicase loading in the budding yeast *S. cerevisiae*. I will begin with a discussion of replication origins sequences, the genomic site of origin licensing and replication initiation. For comparison, I will also discuss the origins of replication found in *E.coli.* Next, I will describe what is known about the mechanism of replicative helicase loading at origins of replication, both in *E. coli* and *S. cerevisiae*. Finally, I will discuss how origin licensing is regulated to ensure replication occurs only when cells are ready.

2. Where it all begins: Origins of Replication

Origins of replication are specific locations within the genome where new DNA synthesis initiates. In *E. coli* and many other prokaryotes, replication initiation occurs at a single sequence within the circular genome (termed *oriC* in *E. coli)*. In contrast, eukaryotes, with their larger genomes organized into linear chromosomes, initiate replication from hundreds or thousands of sites within the genomic DNA that are distributed across each chromosome (reviewed in (Rivera-Mulia and Gilbert, 2016)). Like *E. coli, S. cerevisae* utilizes a specific DNA sequence to specify origins. Although many matches to this sequence exist in the budding yeast genome, only a subset are licensed (Berbenetz et al., 2010; Eaton et al., 2010). Thus, additional factors, such as the chromatin context of the origin DNA, are essential to allow for successful helicase loading. In higher eukaryotes, a defined DNA sequence for origins of replication has not been discovered. Features such as DNA conformation and chromatin are believed to define origins.

2.1 The *E. coli* origin sequence: *oriC*

Most bacteria have a single circular chromosome containing one origin of replication, making them an attractive model organism in which to study origin function. The *E. coli oriC* origin of replication is particularly well understood. *oriC* encompasses an approximately 260 bp region of the *E. coli* genome. Within this region, three key repeated sequences contribute to *oriC* function (Figure 2). A series of five 9 bp sequence motifs, termed DnaA boxes, are responsible for the initial recruitment of the bacterial replication initiator protein DnaA (Fuller et al., 1984; Matsui et al., 1985). Interspersed throughout the DnaA boxes are lower affinity sites for DnaA binding termed I site (Grimwade et al., 2000; Ryan et al., 2002). These sites only bind ATP-bound DnaA (McGarry et al., 2004). The *oriC* DNA unwinding element (DUE) is located adjacent to the DnaA boxes. This region is AT-rich, contains three 13mer repeats and an AT cluster, and is the site of initial DNA unwinding (Bramhill and Kornberg, 1988; Gille and Messer, 1991). A third class of DnaA sites also exists within this region, termed ATP-DnaA sites. Like

Figure 2

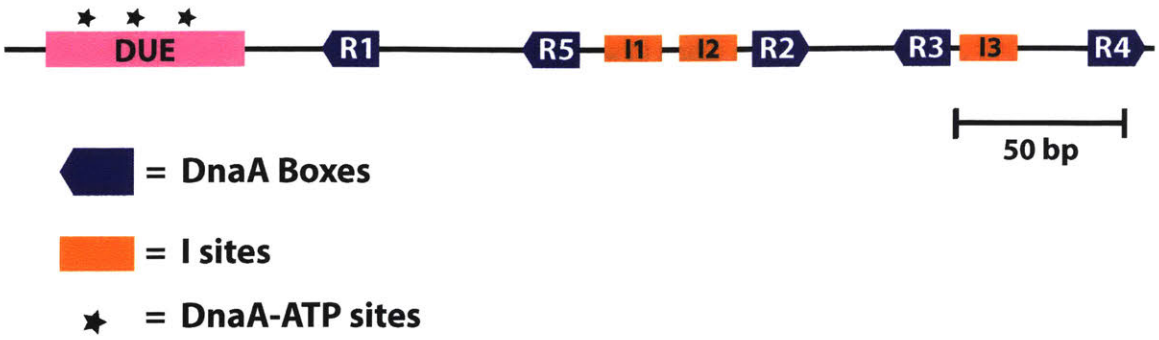


Figure 2: Architecture of *E. coli oriC.*

The five DnaA (R1-R5) boxes are shown in Blue, the I sites are shown in orange, and the DUE is marked in pink. Stars mark the ATP-DnaA sites within the DUE needed to DnaA-ssDNA binding. Modified from Mott and Berger, 2007

the I sites, only ATP-bound DNA can interact at these locations (Speck and Messer, 2001; Speck et al., 1999). How these sequences interact with DnaA to trigger DNA unwinding and helicase loading will be discussed in detail in section 3.1.

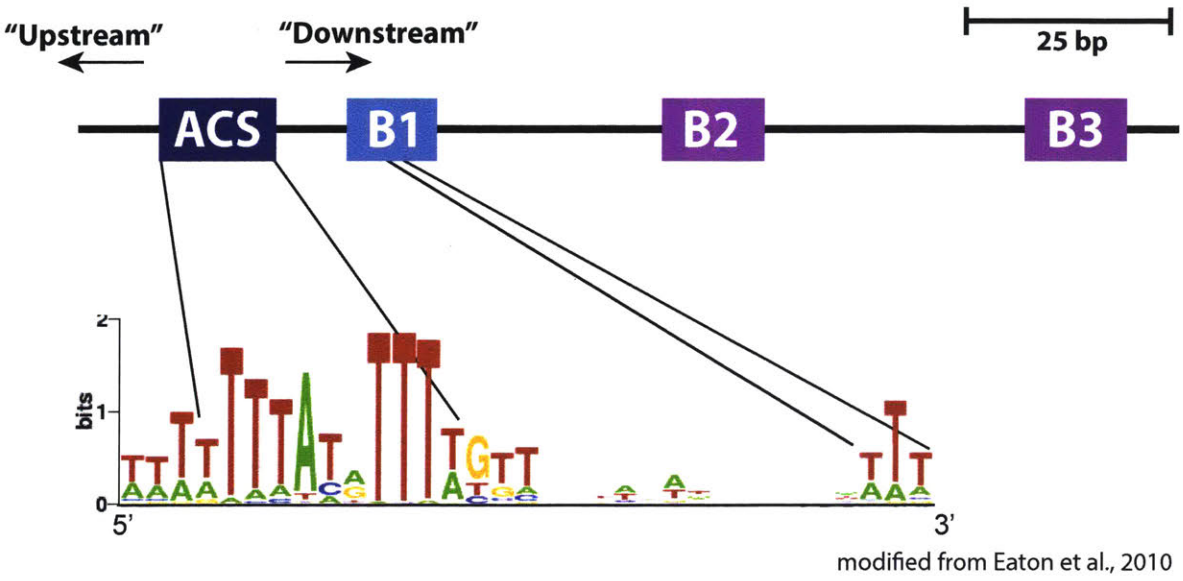
2.2 Origins of Replication in *S. cerevisiae*

2.2.1 DNA sequence elements found at *S. cerevisiae* origins

Similar to *E. coli*, potential *S. cerevisiae* origins of replication contain a number of specific sequences that are important for function (Figure 3A). All potential *S. cerevisiae* origins include an 11 bp *ARS* consensus sequence (ACS) (Broach et al., 1983). In addition to the ACS, there are also two or three additional elements that individually enhance origin function. These elements lie to one side of the ACS and are termed B elements. Unlike the ACS, the B elements have very little sequence conservation. Mutations in any individual B element reduce origin function, but when all the B elements are mutated simultaneously, origin function is lost (Marahrens and Stillman, 1992). Thus, the ACS is necessary but not sufficient for a DNA region to function as a replication origin. Both the ACS and B elements are AT-rich, each having an A-rich and a T-rich strand. Interestingly, the AT-richness of each region is inverted relative to one another (Figure 3B) (Breier et al., 2004). Thus, the T-rich strand of the ACS connects to the A-rich strand of the B elements. This T to A transition is also found at promoters, leading to the hypothesis that such a transition might be important for regions that undergo DNA unwinding (reviewed in (Rando and Chang, 2009)). For reference, and ease of conversation, all origin figures in this dissertation will display the T-rich strand in the 5' to 3' direction, and I will refer to the B elements as downstream of the ACS and the opposite side as upstream (Figure 3).

The various B elements play different roles at origins. Both B1 and B3 are binding sites for specific proteins. B1 elements contain key conserved A/T base pairs that make up part of the bipartite binding site for the first protein to bind to origins, the origin recognition complex (ORC) (Bell and Stillman, 1992; Rao and Stillman, 1995; Rowley et al., 1995). B3, a sequence motif

Figure 3



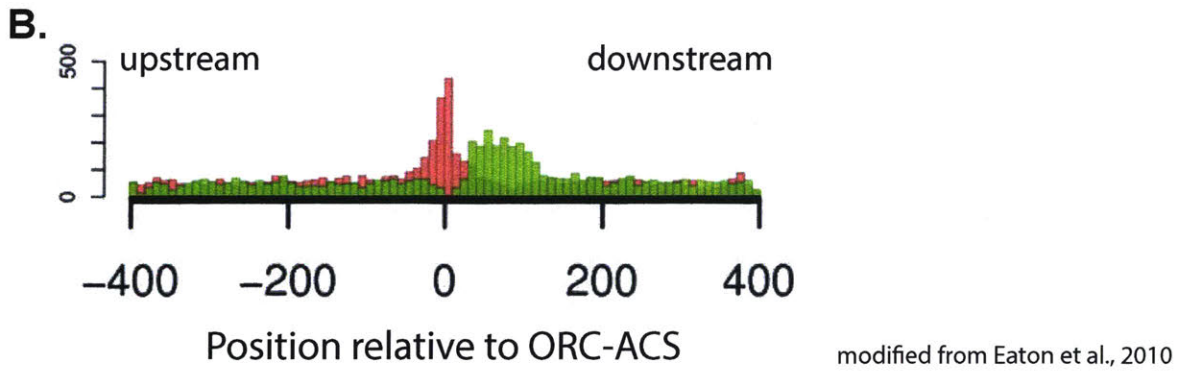


Figure 3: Architecture of a region where origin licensing occurs in S. cerevisiae.

A) The relative locations of the ACS, B1, B2 and B3 elements in a typical budding yeast origin. The LOGO motif shows the ORC-ACS. The ACS and B1 sequence elements within the ORC-ACS are highlighted (modified from Eaton et al., 2010)

B) Diagram showing the A/T disparity on a single strand of the origin DNA. Red is the abundance of T's. Green is the abundance of A's (modified Eaton et al., 2010)

only found at a subset of budding yeast origins, is a binding site for Abf1 (ARS binding factor 1) (Diffley and Stillman, 1988; Eisenberg et al., 1988). Abf1 binding, which is found at certain transcriptional promoters in addition to origins, helps establish proper chromatin organization (Fox et al., 1993; Lipford and Bell, 2001; Della Seta et al., 1990).

The final B element, B2, although important for origin function, is still not well understood. Unlike B1 and B3, which have some sequence preferences, no sequence similarity has been identified for all annotated B2 elements. All studied B2 elements are AT-rich, but can vary in length substantially (Chang et al., 2011; Huang and Kowalski, 1996, 1993; Rao et al., 1994). Despite their differences, B2 elements can be swapped, and origin function preserved, suggesting there is a conserved function for this element (Lin and Kowalski, 1997; Rao et al., 1994).

Three hypotheses have been proposed to explain the role B2 element plays in origin function. First, B2 may serve as a region of DNA unwinding during helicase activation. Annotated B2 elements are AT-rich and therefore are reminiscent of the 13-mer repeats in the DUE region of *oriC*. Indeed, initial experiments showed the B2 elements could be replaced with sequences of low helicase stability (Huang and Kowalski, 1993). However, several pieces of evidence suggest B2 does not function as an unwinding element. First, a more thorough examination of B2 sequence requirements at the *ARS1* origin showed that helical instability did not correlate with B2 element function *in vivo* (Wilmes and Bell, 2002). Second, Mcm2-7 complexes are able to initiate DNA replication even when pushed away from the origin sequence *in vitro* and *in vivo* (Gros et al., 2015). Thus, the Mcm2-7 complex, once activated, is capable of unwinding diverse DNA sequences.

A second hypothesis for the functional role of B2 at origins is that it serves as the initial binding site for the replicative helicase, Mcm2-7. B2 elements usually lie around 60 bp downstream of the ACS (Figure 3A). This location places the B2 element squarely in the region predicted to be encircled by the Mcm2-7 complex based on structural studies (for more details,

see section 3.2.5) (Sun et al., 2013, 2014; Yuan et al., 2017) . Mutating the B2 element *in vivo* effects helicase loading, but not ORC binding (Lipford and Bell, 2001; Zou and Stillman, 2000). In addition, *in vivo* footprinting at the chromosomal location of *ARS1* shows protection of the B2 region in addition to the regions bound by ORC alone (ACS and B1) (Diffley et al., 1994). This protection could reflect the location of Mcm2-7 binding, although *in vitro* footprinting with ORC and Cdc6 (but lacking Mcm2-7) yielded a very similar footprint (Speck et al., 2005). Thus, although much circumstantial data supports this hypothesis, no one has demonstrated direct Mcm2-7 binding at the B2 element.

The third hypothesis is that B2 functions as a second ORC binding site. The B2 element in *ARS1* is a close (9/11) match to the ACS sequence, although it is in the opposite orientation (Figure 4) (Wilmes and Bell, 2002). ORC can bind to the B2 element in *in vitro* footprinting assays, although binding at B2 only occurs when the ACS is mutated or Orc6 (the smallest ORC subunit) is not included (Bell and Stillman, 1992; Lee and Bell, 1997; Wilmes and Bell, 2002). Mutations in the B2 element that increase its sequence similarity to the ACS correlate with B2 function *in vivo* but not ORC binding at B2 *in vitro* (Wilmes and Bell, 2002). In addition, only a single ORC molecule is observed in a single-molecule visualization of the helicase loading process (Ticau et al., 2015). Thus, although important, how B2 contributes to origin licensing or replication initiation remains mysterious and is a question that I will address in Chapter 2.

2.2.2 The influence of the local chromatin environment on origin licensing

In budding yeast, the ACS sequence is necessary, but not sufficient, to define a region as a potential origin of replication. As mentioned above, many close matches to the ACS exist in the *S. cerevisiae* genome, yet only a subset are licensed and even fewer actually initiate DNA replication (see section 4 for more information on selection of origins to initiate). Indeed, many ACS-containing DNA fragments, although able to function as origins on plasmids, do not retain

Figure 4

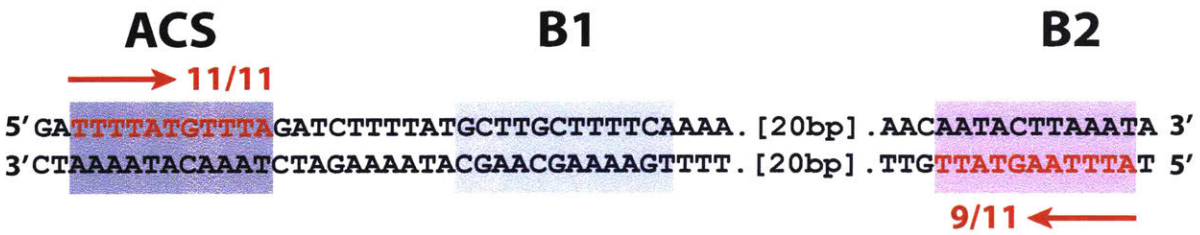


Figure 4: The B2 element of *ARS1* is an inverted 9/11 match to the ACS sequence

Diagram showing the consensus sequence matches presence at the ACS and B2 elements of *ARS1*. To orient the reader, the ACS, B1 and B2 sequences are highlighted. The T-rich strand of the ACS match within the ACS and B2 is shown in red. Arrows indicate the orientation of the match. Note that the ACS and B2 matches are inverted relative to one another.

this function at their endogenous chromosomal location (Newlon et al., 1993). Genome-wide studies of nucleosome location provide insight into the distinction between active and inactive ACS matches. These studies found that the ACS within licensed origins is asymmetrically located within a nucleosome-free region (NFR) that is surrounded by regularly spaced nucleosomes (Berbenetz et al., 2010; Eaton et al., 2010).

The NFR found at origins is typically ~125 bp in size, with the ACS very close to upstream nucleosome, and the B elements occupying the rest of the NFR (Figure 5A and 5B) (Berbenetz et al., 2010; Eaton et al., 2010). Similar NFRs are found at transcriptional start sites (TSSs) (reviewed in (Rando and Chang, 2009)). The establishment of the NFR is important for protein binding events involved in the origin licensing process. Moving the NFR-adjacent nucleosomes into the NFR negatively affects ORC and Mcm2-7 binding to the origin, impairing origin function. Restoring accessibility to origin DNA by moving the nucleosomes back to their original location restores function (Azmi et al., 2017; Simpson, 1990).

Both passive and active mechanisms establish the NFR at potential origins of replication. The ACS and B elements contain poly-A sequences, a composition that disfavors nucleosome formation (Segal and Widom, 2009). However, other sites in the genome with similar A-richness, while nucleosome free, do not have the same organized nucleosomal pattern observed at origins (and TSSs) (Berbenetz et al., 2010; Eaton et al., 2010). Both origin-licensing factors and chromatin remodeling enzymes (CREs) play an active role in organizing origin-proximal nucleosomes. ORC-DNA binding positions nucleosomes around the origin sequences (Eaton et al., 2010; Lipford and Bell, 2001). The Orc1 subunit encodes a BAH domain, a domain that can interact with histones and might aid in this process (Müller et al., 2010; Yang and Xu, 2013). In addition, origins with a B3 element utilize the Abf1 protein to help establish the location of the nucleosome on the downstream (B element) side of the ACS (Diffley and Stillman, 1988; Lipford and Bell, 2001).

Figure 5

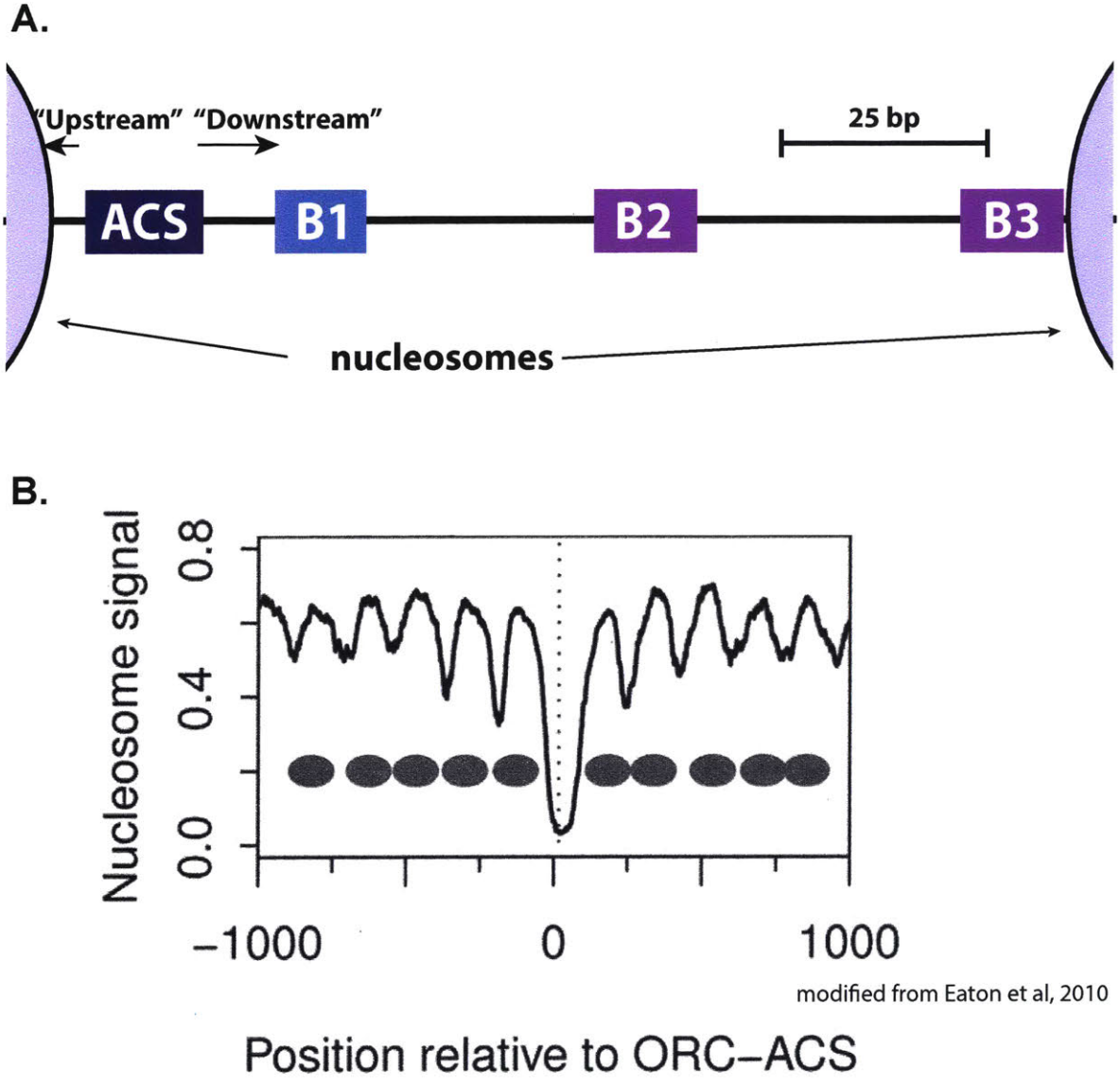


Figure 5: Nucleosomes flank origin sequences

A) Diagram from Figure 2 modified to show the predicted edges of the adjacent nucleosomes.

B) Genome-wide nucleosome positioning around origens of replication with well-defined ACSs. Peaks indicate the dyad of each nucleosome. 0 marks the first base of the ORC-ACS. (figure from Eaton et al, 2010)

Chromatin remodeling enzymes (CREs) also impact nucleosome positioning around origins genome-wide (Struhl and Segal, 2013). CREs utilize energy from ATP to place, move, or otherwise modify nucleosome positioning or composition (Clapier and Cairns, 2009; Papamichos-Chronakis and Peterson, 2012). Many of these complexes can positively or negatively regulate replication initiation *in vivo* (Biswas et al., 2008; Flanagan and Peterson, 1999). It seems likely that CREs regulate origins in part by positioning nucleosomes in or around the NFR. A recent *in vitro* study shows that different CREs establish different patterns of origin-proximal nucleosomes around *ARS1* (Azmi et al., 2017). Some CREs even placed nucleosomes within the NFR, thereby inhibiting ORC binding and helicase loading. As nucleosomes adjacent to the NFR are more dynamic that the average nucleosome (Dion et al., 2007), it is likely that interplay of many CREs *in vivo* establish certain regions as potential origins and others as not.

The establishment of the NFR may do more than simply provide access to important origin sequence motifs. NFR-adjacent nucleosomes may actively participate in the helicase loading process. Movement of the ACS-adjacent nucleosome 70bps upstream of the ACS has negative consequences for helicase loading and origin function (Lipford and Bell, 2001). This observation suggests that correct positioning of this nucleosome positively impacts the helicase loading process. Recent high-resolution MNase mapping of protein binding at origins suggests that loaded Mcm2-7 complexes interact tightly with nucleosomes (Belsky et al., 2015). It is possible that this interaction aids loading of the helicases as well. This hypothesis is further supported by the identification of a histone binding site on Mcm2 (Foltman et al., 2013; Huang et al., 2015).

3. Loading the Replicative helicase

Placement of a helicase on the DNA marks a DNA region as a potential origin of replication. Helicase loading on origin DNA is separated from helicase activation in eukaryotes

(reviewed in (Bell and Labib, 2016)). In contrast, the *E. coli* helicase is rapidly activated once placed at the origin (Makowska-Grzyska and Kaguni, 2010). In this section, I will talk about how the helicases are loaded in *E. coli,* and budding yeast cells.

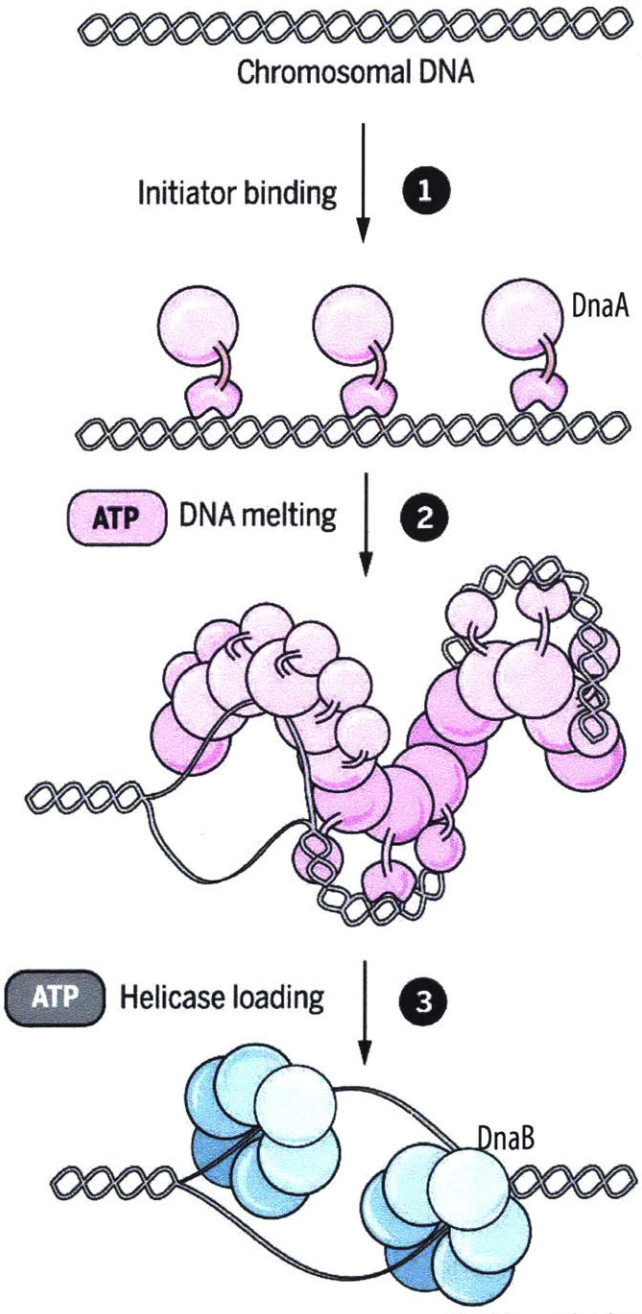
3.1: Helicase Loading in *E. coli*

In *E. coli,* two proteins, DnaA and DnaC, coordinate the placement of the helicase, DnaB, onto origin DNA (Figure 6). Most of the heavy-lifting is done by DnaA. DnaA both recognizes the origin sequence and unwinds the DUE element. DnaC binds to the DnaB helicase and breaks open the hexameric ring. Together, the DnaB/DnaC complex interacts with DnaA bound at the origin to load two DnaB helicases within the ssDNA bubble formed at the DUE. In this section, I will briefly discuss the mechanisms used by DnaA and DnaC to place DnaB at *oriC.*

3.1.1 DnaA recognizes *oriC*

DnaA recognizes and binds directly to sequence-specific motifs in *oriC.* DnaA is a AAA+ ATPase that also contains a helix-turn-helix (HTH) motif at the C-terminus (Figure 7A). Initial binding of DnaA to *oriC* occurs via the interactions between the HTH motif and the DnaA boxes (Figure 7B, right) (Fujikawa et al., 2003; Roth and Messer, 1995; Sutton and Kaguni, 1997). Although DnaA's interaction with the DnaA boxes is tight, binding to the five DnaA boxes at *oriC* alone is insufficient to direct replication initiation (McGarry et al., 2004). Once DnaA has bound the DnaA boxes, cooperative interactions via the AAA+ domain, along with some contacts at the N-terminus, recruit more DnaA molecules to the lower affinity I sites (Figure 7B) (Erzberger et al., 2006; Kawakami et al., 2005; McGarry et al., 2004; Rozgaja et al., 2011; Simmons et al., 2004; Weigel et al., 1999). Unlike the DnaA boxes, which DnaA interacts with regardless of nucleotide, the lower affinity I sites require ATP-bound DnaA (McGarry et al., 2004).

Figure 6



modified from Bleichart, Botchan and Berger, 2017

Figure 6: Diagram of helicase loading in *E. coli.*

(1) DnaA binds to the DnaA boxes. **(2)** DnaA oligomerizes, melts, and binds the DUE element. **(3)** DnaA and DnaC coordinate to load the DnaB helicase to the newly formed ssDNA. Figure modified from Bleichart, Botchan and Berger, 2017

Figure 7

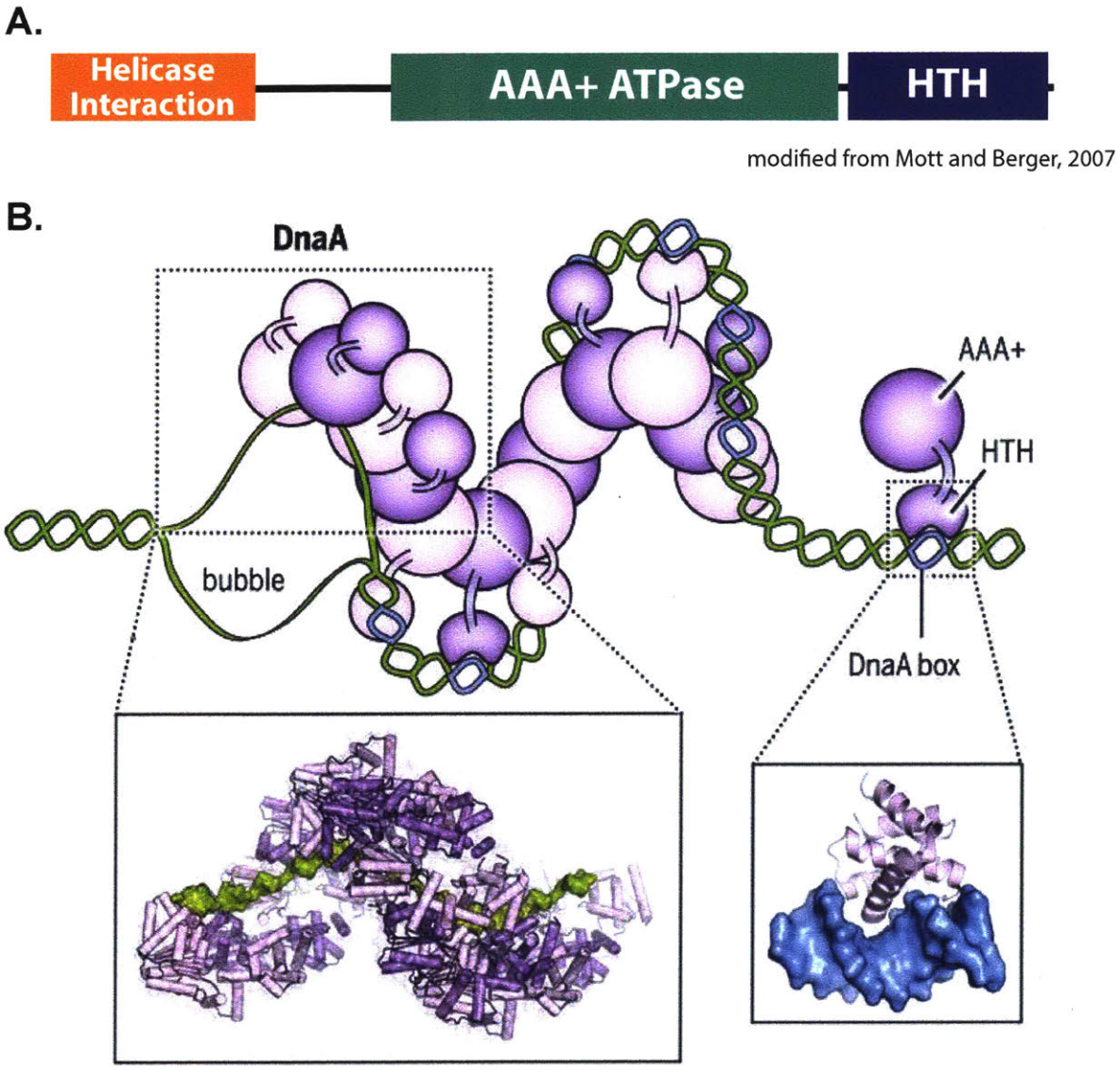


Figure 7: Current models for DnaA recognition and unwinding of *oriC*.

A) Diagram showing the important domains of DnaA.

B) A annotated model for DnaA function at *oriC*. The crystal structure of HTH motif interacting with dsDNA and ATP-DnaA bound to ssDNA are shown. Modified from Bleichart, Botchan and Berger, 2017

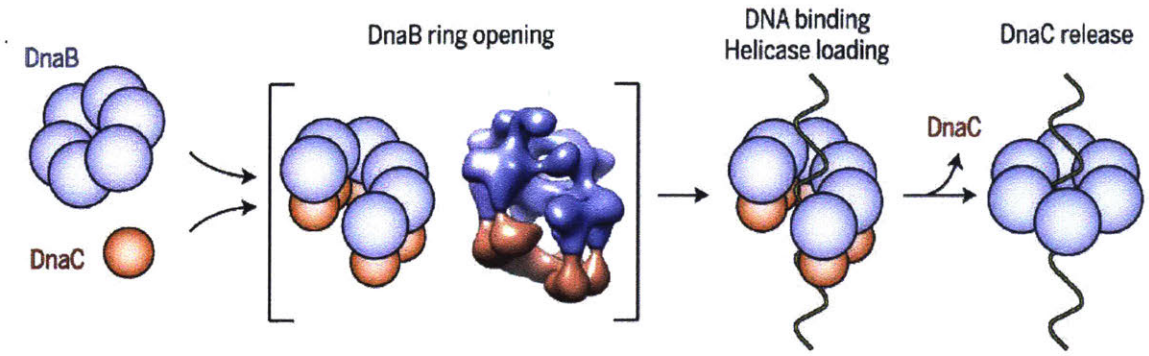
3.1.2 How does DnaA unwind the DUE?

DnaA not only recognizes origins, but also melts the DNA duplex at the DUE. Although the exact DnaA structure required for DNA unwinding is unknown, recent structural data has led to a model for DnaA-induced unwinding. A crystal structure of multiple DnaA molecules bound to ATP reveals a right-handed helical filament of DnaA (Erzberger et al., 2006). Origin DNA has been proposed to wrap around the outside of DnaA during origin recognition (Bramhill and Kornberg, 1988; Fuller et al., 1984), and wrapping around this helical structure would introduce positive supercoils into the DNA (Erzberger et al., 2006). Adjacent DNA would compensate by introducing negative supercoils, driving unwinding in a region of low helical stability like the DUE. Once the DNA is unwound, DnaA binds its third type of binding site, the ATP-DnaA sites, to stabilize the ssDNA. A recent DnaA/ssDNA co-crystal structure shows that the DnaA AAA+ motif holds ssDNA in an extended conformation, thus preventing reannealing (Figure 7B) (Duderstadt et al., 2011).

3.1.3 Placement of DnaB on DNA

Once DnaA forms and stabilizes ssDNA, the DnaB helicase must be placed onto the ssDNA for the genome to be unwound and replicated (reviewed in (Bell and Kaguni, 2013)). However, the active DnaB helicase is a closed, six-membered ring in solution (Lo et al., 2009; Wang et al., 2008) . To unwind the genome, the DnaB ring must be opened, placed around ssDNA, and closed. DnaB ring opening is accomplished by a AAA+ protein, DnaC (Arias-Palomo et al., 2013). Unlike DnaB, DnaC exists as individual subunits in solution (Kobori and Kornberg, 1982). However, when DnaB is present, the two proteins form a DnaC/DnaB hetero-dodecamer ((Arias-Palomo et al., 2013; Galletto et al., 2003; Kobori and Kornberg, 1982) (Figure 8). A low-resolution electron microscopy (EM) structure of the DnaC/DnaB complex shows a three tier structure, with the ATPase domain of DnaB interacting with the N-terminal extensions of DnaC (Arias-Palomo et al., 2013). Importantly, this double-hexamer structure is a

Figure 8



modified from Bleichart, Botchan and Berger, 2017

Figure 8: Current Model for how DnaC interacts with DnaB to load the helicase on ssDNA origin DNA.

DnaB recruits 6 DnaC monomers for form a DnaB/DnaC dodecamer. Importantly, this dodecamer forms a cracked ring structure and the DnaB helicase is inactive. Once placed on DNA, DnaC is ejected from DnaB after the first primer is synthesized, and DnaB hexamer becomes active once DnaC leaves. Figure modified from Bleichart, Botchan and Berger, 2017

cracked ring, with an opening large enough to allow ssDNA to pass into the charged DnaB central channel (Arias-Palomo et al., 2013). Although the process by which the DnaB/DnaC hetero-dodecamer is still unclear, interactions between an auxillary N-terminal domain of DnaA and the DnaB N-terminal collar are important for helicase recruitment to origins (Fang et al., 1999; Marszalek and Kaguni, 1994; Seitz et al., 2000; Sutton et al., 1998). DnaA/DnaC interactions may also help facilitate this process (Arias-Palomo et al., 2013; Mott et al., 2008).

3.2: Helicase Loading in *S. cerevisiae*

In *S. cerevisiae,* three different proteins, the origin recognition complex (ORC), Cdc6, and Cdt1, cooperate to load the core of the replicative helicase, Mcm2-7, onto origin DNA during G1 (Figure 9). ORC recognizes the origin DNA directly. Two additional helicase loaders, Cdc6 and Cdt1, coordinate with ORC and Mcm2-7 to sequentially load two Mcm2-7 hexamers. The end result of this process is placement of two Mcm2-7 hexamers in an inactive head-to-head conformation at origins of replication. Unlike *E. coli* DnaB, the Mcm2-7 double hexamer encircles dsDNA after loading and multiple additional factors are required to activate the loaded helicases. The presence of an inactive Mcm2-7 never-the-less marks a location as a potential origin of replication. In this section, I will discuss the *S. cerevisiae* helicase loading process in detail, starting with ORC binding, and ending with the formation of the stable Mcm2-7 double-hexamer, which is poised to initiate replication in S phase.

3.2.1: ORC recognizes origins of replication

The Origin Recognition Complex, or ORC, recognizes and binds origin DNA. ORC is composed of six different subunits, Orc1-Orc6 (Figure 10). ORC was first identified by its ability to bind to the ACS *in vitro* (Bell and Stillman, 1992). The Orc1-Orc5 subunits each contain two domains: an N-terminal AAA+ ATPase (or ATPase-like) domain and a C-terminal winged-helix (WH) domain (Figure 10A). Although these five subunits each adopt a classic AAA+ ATPase

Figure 9

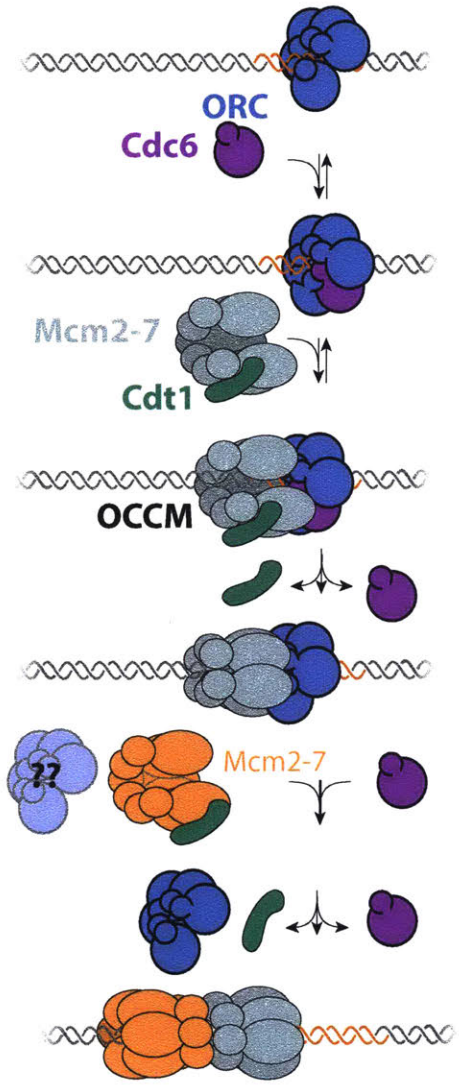


Figure 9: Current Model for Origin Licensing.

ORC recognizes the ACS and B1 elements of the origin. Once bound to origin DNA, it recruits Cdc6. ORC-Cdc6 recruits an Mcm2-7/Cdt1 complex, forming the OCCM. In a process that involves ATP hydrolysis, Cdc6 and Cdt1 depart and the Mcm2-7 ring closes, leaving a stable single hexamer adjacent to ORC. In a process that involves another Cdc6 and Mcm2-7/Cdt1 complex, and may or may not involve an additional ORC molecule, a second Mcm2-7 hexamer is loaded in a head-to-head orientation to form a stable Mcm2-7 double hexamer.

Figure 10

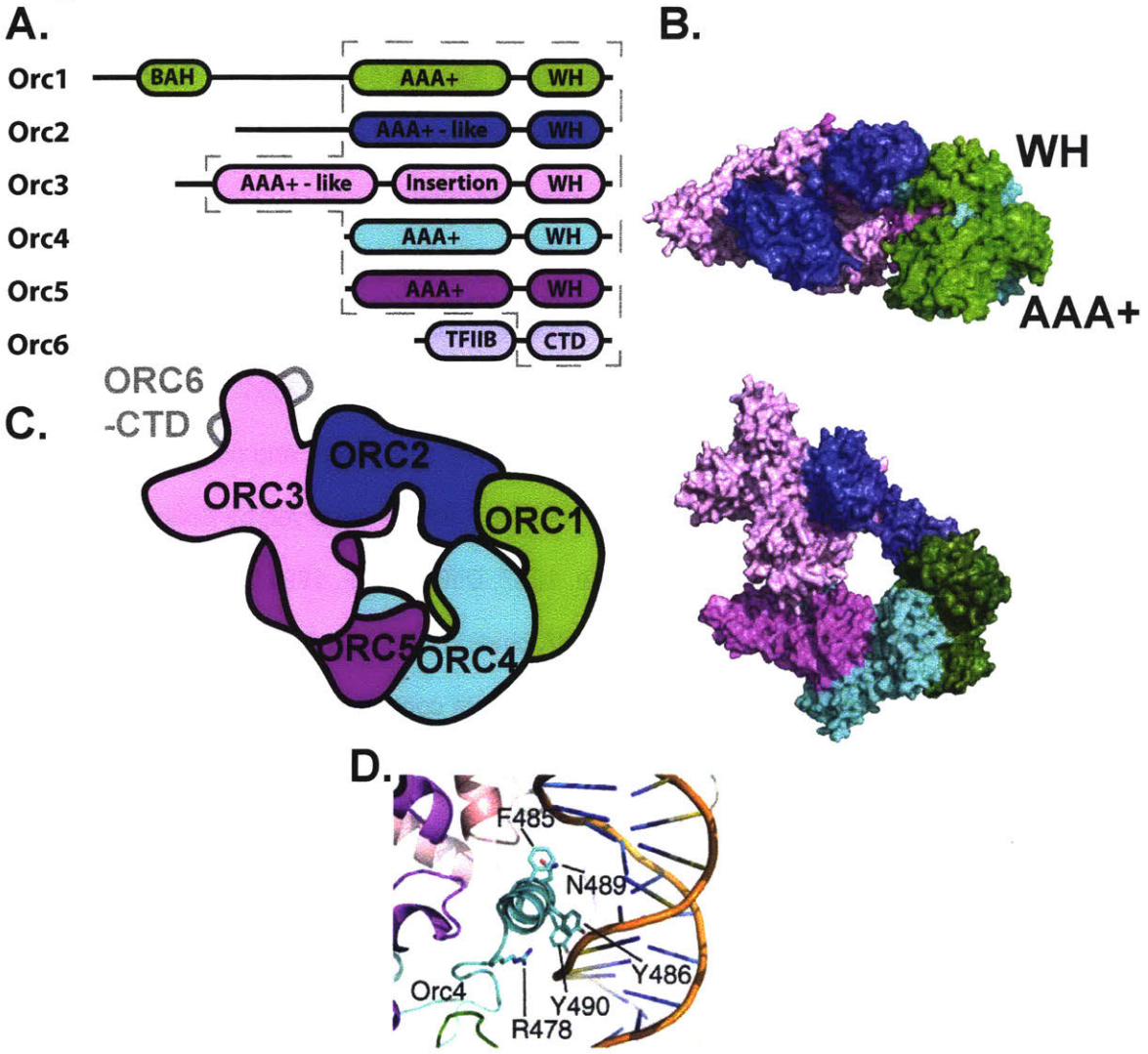


Figure 10: The Origin Recognition Complex (ORC)

A) Functional domains of the Orc1-Orc6 subunits. Domains included/observed in crystallography/EM structural studies are outlined in grey (modified from Bleichart et al, 2015).

B) Side view of the Orc1-5 ring. The winged-helix (WH) domains, which interact with Mcm2-7, are at the top, and the AAA+ domains are at the bottom. (human ORC model, Tocilj et al, 2017)

C) Diagram (left) and structural top view (right) of ORC. Although this structure does not have Orc6, the location to which the C-terminus binds is shown in the diagram (left). A structure where the CTD of Orc6 is resolved is shown in Figure 9B (human ORC model, Tocilj et al, 2017)

D) ORC binding to DNA in yeast involves a unique interaction with ORC4. (structure from Yuan et al., 2017)

fold, with ATPase-like sites at the interface between adjacent subunits (Bleichert et al., 2015), only Orc1 is an active ATPase (Chesnokov et al., 2001; Giordano-Coltart et al., 2005; Klemm et al., 1997). Residues needed for ATP binding are missing in Orc2 and Orc3, and critical residues needed for ATP hydrolysis are missing in Orc4 and Orc5 (Bleichert et al., 2015; Chesnokov et al., 2001; Giordano-Coltart et al., 2005; Klemm et al., 1997; Kong et al., 2003). Orc6, the smallest ORC subunit, is unrelated to the other subunits and includes a TFIIB-like domain and a conserved C-terminal domain (Bleichert et al., 2013; Chesnokov et al., 2001; Liu et al., 2011).

In *S. cerevisiae,* ORC recognizes and binds directly to the ACS and conserved A/T base pairs found in B1 elements at origins. Protection of these sites is observed by DNase I footprinting *in vitro* and *in vivo* (Bell and Stillman, 1992; Diffley et al., 1994). Mutating these elements decrease (B1) or inhibit (ACS) ORC-DNA binding (Bell and Stillman, 1992; Rao and Stillman, 1995; Rowley et al., 1995). This recognition motif containing both the ACS and B1 is referred to as the ORC-ACS (Berbenetz et al., 2010; Eaton et al., 2010; Xu et al., 2006). In *S. cerevisiae,* ORC binding to origin DNA only requires Orc1-5, although Orc6 is essential for the subsequent loading of the replicative helicase (Lee and Bell, 1997) (see section 3.2.2 for more on Orc6).

Recent structural studies provide insights into how ORC binds DNA (Bleichert et al., 2015; Tocilj et al., 2017; Yuan et al., 2017). The core of the Orc1-5 subunits form a cracked ring with a defined subunit order: Orc1-Orc4-Orc5-Orc3-Orc2 (Bleichert et al., 2015) (Figure 10C). The WH and ATPase domains form two distinct rings (Figure 10B) (Bleichert et al., 2015; Tocilj et al., 2017; Yuan et al., 2017). Interestingly, the WH domain of one subunit reaches over and binds the AAA+ domain of the adjacent subunit to create a domain-swap architecture (Figure 10B) (Bleichert et al., 2015; Tocilj et al., 2017). A C-terminal helix in Orc6 binds to an extension in Orc3, although the rest of Orc6 was removed or unstructured in these studies (Bleichert et al., 2015; Yuan et al., 2017).

Although no structure of ORC bound to DNA alone has been captured, the structure of a key intermediate in helicase loading, the ORC-Cdc6-Cdt1-Mcm2-7 (OCCM) complex (for more, see section 3.2.5) was determined with dsDNA (Yuan et al., 2017). Importantly, the ORC-Cdc6 ring in this structure is similar to the modeled structure of the active ORC complex alone (Bleichert et al., 2015; Tocilj et al., 2017). Therefore, the contacts between the DNA and ORC in in the OCCM may be the same as those made during initial ORC binding. In the OCCM structure, Orc1-5 encircles the DNA, with Cdc6 completing the ring (see section 3.2.3 for more on Cdc6) (Yuan et al., 2017). Although the sequence of DNA bound to ORC is not determined in the OCCM structure, Orc4 and Orc2 interact with the DNA via parts of both their AAA+ and WH domains (Yuan et al., 2017). Interestingly, one particular alpha helix in Orc4 interacts extensively with the DNA (Figure 10D) and represents an insertion present only in *Saccharomyces* species. Thus, this helix is likely important for the sequence-specific nature of ScORC binding relative to most other eukaryotes (Yuan et al., 2017). In both the OCCM and models of a DNA-bound ORC structure, 20-24 bps fit within the ORC ring (Bleichert et al., 2015; Tocilj et al., 2017; Yuan et al., 2017). In contrast DNase I footprinting shows 45-50bps of protection (Bell and Stillman, 1992). Conformational changes not observed in these static structures or bending of the DNA to interact with the outside of the ORC ring could account for this discrepancy.

3.2.2 Orc6: The odd subunit out

Although all the other ORC subunits are related to AAA+ ATPases, Orc6 is an unrelated but essential subunit that is conserved from yeast to humans (Bleichert et al., 2013). Little is known about the structure of Orc6. In all the structural studies of the ORC complex, Orc6 was either truncated to a conserved C-terminal region (Bleichert et al., 2015), not included at all (Tocilj et al., 2017), or unresolved (Bleichert et al., 2013; Yuan et al., 2017). The C-terminal domain folds into an alpha helix and bound to a conserved extension for uniquely in within Orc3

(Figure 10C) (Bleichert et al., 2013, 2015; Yuan et al., 2017). The middle domain of human Orc6 was crystalized alone and adopted a classic TFIIB-type fold as expected, but the importance of this structure in the context of the full ORC complex is unknown (Liu et al., 2011).

Although Orc6 is dispensable for sequence-specific binding in yeast, it is required for helicase loading (Bell and Stillman, 1992; Chen et al., 2007; Lee and Bell, 1997). Orc6 interacts directly with Cdt1, which is bound to Mcm2-7 prior to helicase loading (for more detail on Cdt1, see section 3.2.4) (Chen et al., 2007). This interaction may help facilitate Mcm2-7/Cdt1 recruitment to the origin (Chen et al., 2007). However, Orc6 is not essential for initial Mcm2-7 recruitment in a fully reconstituted *in vitro* assay (Frigola et al., 2013). Interestingly, Orc6 contains two distinct Cdt1 binding sites, suggesting possible mechanisms where the first and second hexamer interact with Orc6 in different manners during the helicase loading process (Chen et al., 2007). In budding yeast, the interaction between Orc6 and Mcm2-7/Cdt1 is inhibited by CDK phosphorylation of Orc6, an event that plays an important role in preventing re-initiation (Chen and Bell, 2011) (see section 6.2 for more details on mechanisms to prevent re-loading).

3.2.3 Forming the ORC-Cdc6 complex on origin DNA

After ORC recognizes and binds to origin DNA, it recruits Cdc6. Initial binding of Cdc6 to DNA-bound ORC is required for Mcm2-7 recruitment (Randell et al., 2006; Remus et al., 2009; Ticau et al., 2015). Similar to the Orc1-Orc5 subunits, Cdc6 is comprised of two domains: an N-terminal (and functional) AAA+ ATPase domain and a C-terminal WH domain (Figure 11A). In the OCCM intermediate, Cdc6 binds in the gap between the Orc1 and Orc2 subunits to complete the six membered ring (Figure 11B) (Yuan et al., 2017). Consistent with Cdc6 being structurally related to Orc1-5, Cdc6 adopts a nearly identical structure to the ORC subunits: its WH domain sits atop the ATPase domain of the adjacent Orc1 subunit (Tocilj et al., 2017; Yuan et al., 2017). Like Orc4 and Orc2, Cdc6 also makes direct contact with DNA via both its AAA+

Figure 11

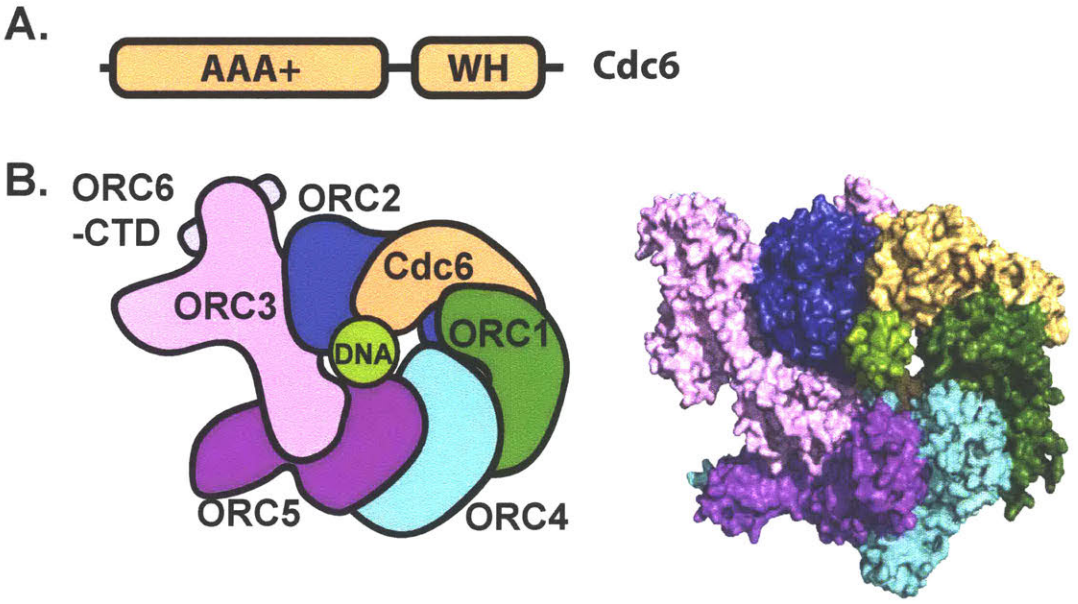


Figure 11: Cdc6 completes the Orc1-Orc6 ring.

A) Diagram of functional domains of Cdc6.

B) Diagram (left) and structure (right) of the ORC-Cdc6 ring from the AAA+ domain side of the the ORC-Cdc6-Cdt1-Mcm2-7 (OCCM) intermediate structure.

and WH domains (Yuan et al., 2017). These interactions support observations that Cdc6 makes ORC binding to the origin more sequence-specific in budding yeast (Duzdevich et al., 2015; Speck and Stillman, 2007).

Although structural data presents a simple model for ORC and Cdc6 binding to origin DNA, there is likely more to this process than the structural data suggests. DNase I footprinting of origin DNA in the presence of both ORC and Cdc6 shows a 70-80 bp footprint that is much larger than the 24 bps predicted in the OCCM structure (Speck et al., 2005; Yuan et al., 2017). This extended footprint includes the B2 element and is very similar to the *in vivo* footprint observed in G1 extracts (Diffley et al., 1994; Speck et al., 2005). Since we do not have a structure of ORC/Cdc6 alone, this complex may adopt a different conformation than that observed in the OCCM. Alternatively, DNA may interact with the outside of the ORC ring in the absence of a bound Mcm2-7/Cdt1. Further studies are necessary to resolve this discrepancy.

3.2.4 The Mcm2-7 helicase, and its helper, Cdt1:

Once the ORC-Cdc6 complex forms on DNA, it recruits the core of the replicative helicase, the Mcm2-7 complex. Unlike the *E. coli* helicase DnaB, which is composed of six identical subunits, Mcm2-7 is comprised of related but distinct subunits (Figure 12A) (reviewed in (Bochman and Schwacha, 2009)). Each subunit is a AAA+ ATPase and together they form into a ring with a defined order: Mcm2, Mcm6, Mcm4, Mcm7, Mcm3 and Mcm5 (Figure 12B) (Bochman et al., 2008; Davey et al., 2003; Li et al., 2015). All six of the Mcm2-7 subunits are active ATPases, with the ATP binding and hydrolysis domain residing at the interface of each adjacent Mcm2-7 subunit (as is typical for AAA+ ATPases) (Figure 12B) (Bochman et al., 2008; Ilves et al., 2010; Schwacha and Bell, 2001).

Each subunit can be divided into two domains: an the N-terminal domain, which contains a folded subdomain (NTD-A) and an OB fold, and the C-terminal domain, which contains the AAA+ domain (Li et al., 2015). There are also Zn-Finger motifs in N-terminal domain (between

Figure 12

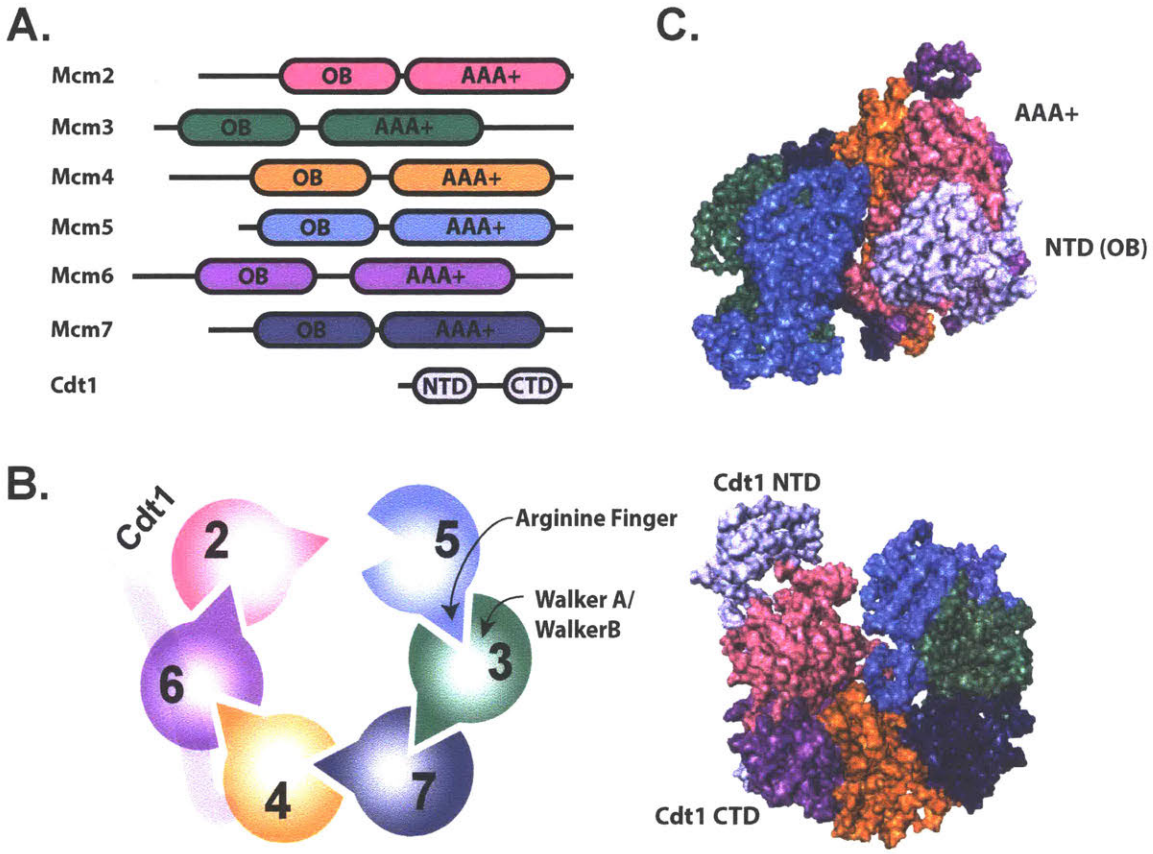


Figure 12: The structure of Mcm2-7/Cdt1

A) Domain architecture of the six Mcm2-7 subunits and Cdt1.

B) Diagram (left) and structural top view (right) of the Mcm2-7/Cdt1 complex. Diagram shows contributing arginine figures (points) and Walker A/Walker B motifs (indents) within the ATPase domains at each subunit interface. The N- and C-terminal domains of Cdt1 are labeled. A gap is visible between the Mcm2 and Mcm5 subunits. Structure from (Zhai et al., 2017).

C) Side view of the Mcm2-7/Cdt1 structure showing the cracked-ring structure. Structure from (Zhai et al, 2017). Although cracked, the break in the Mcm2-7 ring is not large enough for dsDNA to enter the central channel in this structure. However, movement in the actual molecule in solution could easily widen the gap to allow dsDNA access to the positively charged channel

the A domain and the OB fold) that help mediate interactions between the two hexamers in the stable double-hexamer form of Mcm2-7 (Li et al., 2015). Recent structural studies also identify WH domains at the very C-terminus of the Mcm2-7 subunits (Li et al., 2015; Yuan et al., 2017; Zhai et al., 2017). These domains undergo rearrangements during the helicase loading process, and appear to be particularly important for mediating interactions between ORC/Cdc6 and Mcm2-7/Cdt1 (Li et al., 2015; Yuan et al., 2017; Zhai et al., 2017) (For more details on the interactions required to recruit Mcm2-7, see sections 3.2.5 and 3.2.7 below).

Like DnaB, Mcm2-7 cannot localize to the origin DNA without assistance. It relies on interactions with ORC-Cdc6, and another helper protein, Cdt1, for recruitment to origin DNA (Randell et al., 2006; Remus et al., 2009). In budding yeast, Mcm2-7 interacts with the helicase loading protein, Cdt1, in solution (Remus et al., 2009; Tanaka and Diffley, 2002). Cdt1 interacts with Mcm2, Mcm4, and Mcm6, in the neck region where the N- and C-domains of the Mcm2-7 subunits meet each other (Figure 12B, 12C) (Fernández-Cid et al., 2013; Takara and Bell, 2011; Yuan et al., 2017; Zhai et al., 2017). Cdt1 is one of two proteins involved in helicase loading that isn't an AAA+ ATPase, or AAA+ like (Orc6 is the other). Although the exact role of Cdt1 in the process of helicase loading is still unclear, Cdt1 helps recruit Mcm2, Mcm4 and Mcm6 to origin DNA (Frigola et al., 2013). It also may have a role in closing the Mcm2-7 ring (see section 3.2.6 below for more details.)

3.2.5 The ORC-Cdc6-Cdt1-Mcm2-7 (OCCM):

ORC-Cdc6 binding to the Mcm2-7/Cdt1 complex forms the ORC-Cdc6-Cdt1-Mcm2-7 (OCCM) intermediate (Randell et al., 2006; Sun et al., 2013). Although this intermediate is short-lived in the presence of ATP (Ticau et al., 2015), inhibition of ATP hydrolysis by addition of ATPγS stabilizes this complex (Randell et al., 2006). Performing assays in the presence of ATPγS, a number of important contacts for OCCM formation have been identified biochemically. The very C-terminus of Mcm3 is important for the recruitment of the Mcm2-7 subunits 3, 5, and

7 to the ORC/Cdc6 ring (Frigola et al., 2013). A Cdt1-Orc6 interaction has been implicated in helicase recruitment (Chen et al., 2007), although this interaction was not strictly necessary *in vitro* (Frigola et al., 2013). Mutation of the Mcm6 arginine finger also had an effect on the formation of the OCCM complex, although why this defect is observed is not clear (Coster et al., 2014; Kang et al., 2014).

The stability of the OCCM complex in ATPγS has also allowed for structural studies of this intermediate (Sun et al., 2013; Yuan et al., 2017). The OCCM structures reveal that the Mcm2-7 and ORC/Cdc6 rings interact tightly with each other and encircle adjacent DNA (Figure 13). Although the ORC/Cdc6 ring is locked tightly around 24 bps of dsDNA, only the C-terminal ATPase domains of Mcm2-7 are closed. A gap is visible between the Mcm2 and Mcm5 N-terminal domains (Figure 13) (Yuan et al., 2017). The closed C-terminal region of Mcm2-7 covers 15 bps, and the N-terminal domain is expected to cover another 15 (Sun et al., 2013; Yuan et al., 2017).

The interaction between the ORC/Cdc6 and Mcm2-7 rings occurs between the C-terminal WH domains of ORC/Cdc6, and the C-terminal AAA+ ATPase domains of the Mcm2-7 ring (Figure 13). Specifically, WH domains from Mcm3, Mcm4, Mcm6 and Mcm7 reach up out of the core of the Mcm2-7 C-terminal domain and make contacts with the ORC ring (Yuan et al., 2017). Although the WH domain of Mcm5 was not resolved in the structure, cross-linking mass spectrometry (CLMS) suggested this domain also interacts with the ORC/Cdc6 ring(Yuan et al., 2017). Importantly, these WH domains adopt very different conformations in either the Mcm2-7/Cdt1 complex alone or in the Mcm2-7 double hexamer (Li et al., 2015; Yuan et al., 2017; Zhai et al., 2017). It is likely that the rearrangement of these WH domains is important in the helicase-loading process.

Despite the extensive contacts seen in the EM structure, the interaction holding the OCCM together, or interacting with the DNA itself may not be particularly stable. The Mcm2-7 double-hexamer, once loaded onto dsDNA, becomes topologically linked to the DNA, and thus

Figure 13

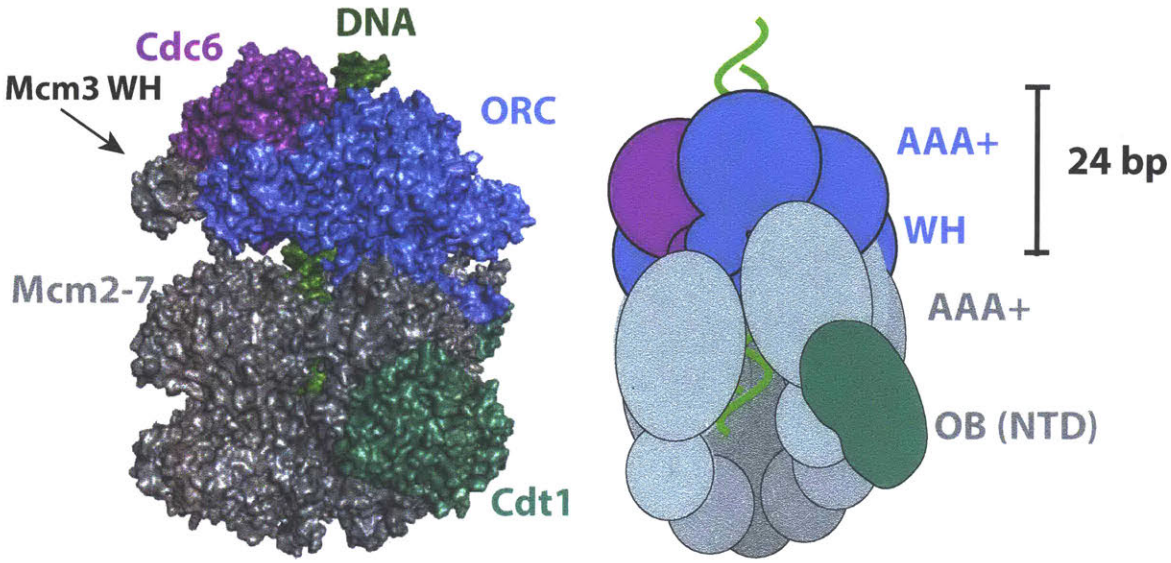


Figure 13: The structure of the OCCM complex

Diagram (left) and Structural side view (right) of the OCCM complex. ORC is shown in blue, Cdc6 is shown in purple, Cdt1 in Green and Mcm2-7 in grey. The structure shows the Mcm2/Mcm5 gate closed at the C-terminus but open at the N terminus. The ORC/Cdc6 ring covers 24 bp of DNA. Structure from Yuan et al, 2017.

is stable to a 500mM NaCl salt extraction (Donovan et al., 1997; Randell et al., 2006). Despite that the EM structures reveal an intermediate in state where the Mcm2/Mcm5 gate is either closed (Sun et al., 2013), or half-closed (Yuan et al., 2017), and ORC/Cdc6 forms a closed ring that appears to be topologically linked to the DNA, the OCCM complex is readily removed from DNA with such a high-salt wash (Randell et al., 2006). Since EM structural determination is based on averaging single particles, the state of the OCCM complex is likely to be more dynamic than these structures suggest.

3.2.6 The Mcm2/Mcm5 gate

During the process of helicase loading, the cracked Mcm2-7 ring is placed around the DNA, and then the ring is closed, resulting in a Mcm2-7 ring encircling dsDNA. Importantly, only the interface between the Mcm2 and Mcm5 subunits can allow dsDNA to enter the Mcm2-7 central channel (Samel et al., 2014). For this reason, this interface is termed the Mcm2/Mcm5 gate (Bochman et al., 2008). As with all the Mcm2-7 subunit interfaces, this is a site of ATP hydrolysis. There is some evidence that ATP binding state influences the interaction between the two subunits. Studies monitoring the binding of Mcm2-7 to a closed, circular ssDNA showed that ATP binding inhibited loading, and thus was in a closed conformation (Samel et al., 2014). A closed state for the Mcm2-7 ring was also observed in ATPγS (Sun et al., 2013). In contrast, more recent structural studies suggest the Mcm2-7 is an open ring in the presence of ATP (Zhai et al., 2017). Future studies will hopefully flesh out how this open to closed ring transition occurs.

Closing the Mcm2/5 gate does not require ORC or Cdc6 ATP hydrolysis but interestingly, ATP hydrolysis at the Mcm3/Mcm5 interface is required for gate closure. Mutation in the Mcm5 arginine finger motif (thought to impact ATP hydrolysis in AAA+ ATPases) resulted in an Mcm2-7 complex incapable of closing the Mcm2/5 gate or releasing Cdt1 (Ticau et al.,

2017). It remains to be determined if Mcm2-7 ATP hydrolysis causes ring closure, which in turn causes Cdt1 release or vice versa.

Although Cdt1 departure correlates with Mcm2/5 gate closure, its exact role in the process remains unclear. Recent single-molecule experiments utilized a FRET-based assay observed Mcm2/Mcm5 gate closure concomitant with Cdt1 release (Ticau et al., 2015, 2017). However, outside of the context of the helicase loading reaction, the Mcm2/Mcm5 gate was open regardless of the presence of Cdt1 (Ticau et al., 2017). Although early negative-stain EM images showed the Mcm2-7 complex to form a closed ring (Samel et al., 2014), a recent low (10-15Å) resolution EM structure of the budding yeast Mcm2-7 with and without Cdt1 found that both structures have an open ring (Zhai et al., 2017). Interestingly, the break between Mcm2 and Mcm5 is not wide enough to allow passage of dsDNA through the open channel (Figure 12C) (Zhai et al., 2017). Taken together, these studies argue against a simple model where Cdt1 opens the otherwise-closed Mcm2-7 ring, but the Mcm2-7 complex may be quite dynamic in solution. Cdt1 may still stabilize the ring, or help pry Mcm2-7 ring open further in order to allow dsDNA into the central channel.

3.2.7 Loading the second hexamer and releasing ORC

Origin licensing requires two Mcm2-7 complexes to be recruited to potential origins, and they must interact in a head-to-head orientation. Because the OCCM complex can be isolated, and initial Mcm2-7 binding is fairly easy to monitor, much is known about recruitment of the first hexamer. In contrast, how the second Mcm2-7 hexamer is recruited is not as well understood. Two competing models of how loading of the second hexamer occurs have been presented (Figure 14). First, the second hexamer could be loaded by a distinct mechanism from the first hexamer (Figure 14, left). Single-molecule experiments support this hypothesis. In these experiments, one ORC is sufficient to recruit both the first and the second Mcm2-7 helicase, but distinct Cdc6 and Cdt1 molecules are involved with loading the two Mcm2-7 complexes (Ticau

Figure 14

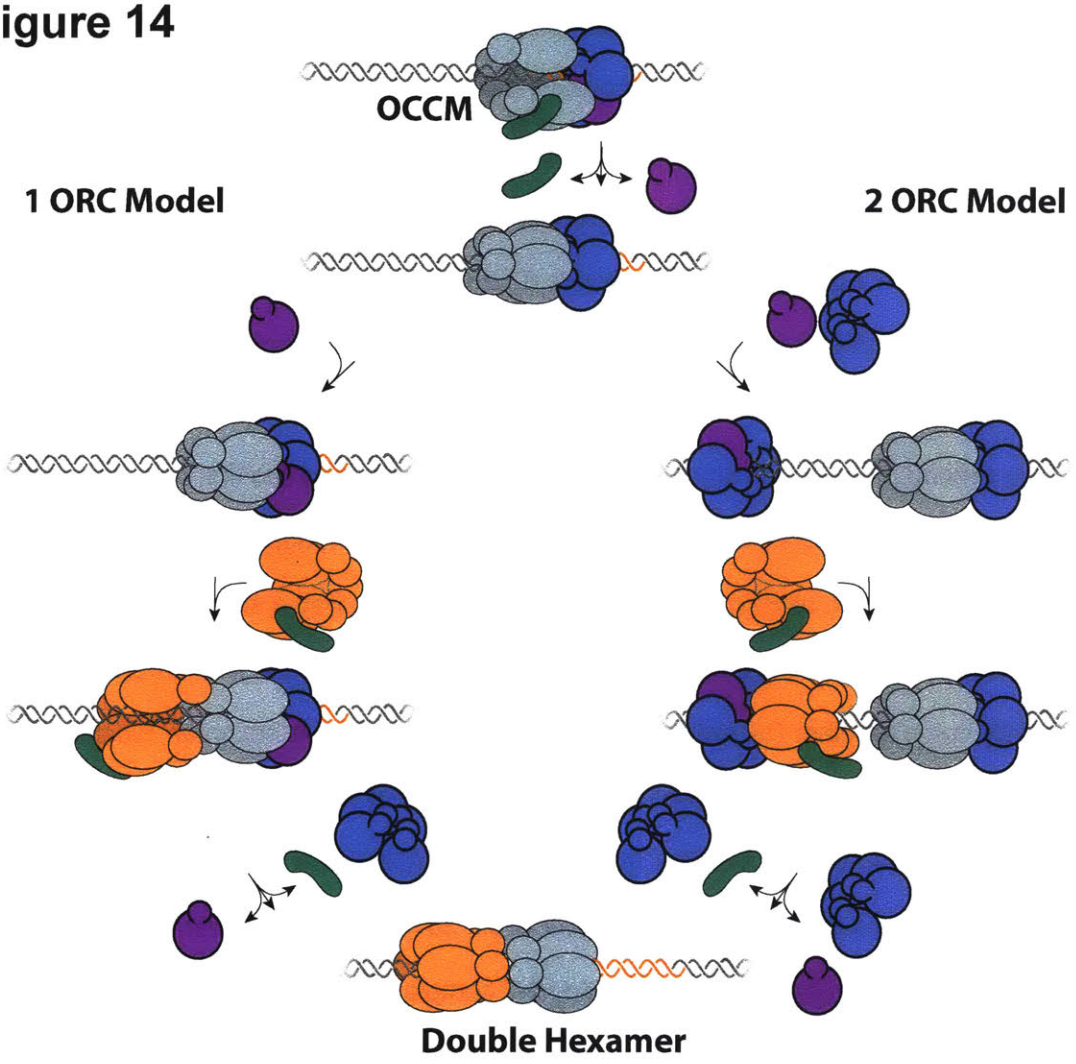


Figure 14: One ORC vs Two ORC model for loading the second Mcm2-7 helicase.

After OCCM formation, two options are proposed. The first ORC could remain behind, and recruit as second Cdc6 and Mcm2-7/Cdt1 complex, as observed via single-molecule microscopy. Interactions between the N-terminal domains of the first and second Mcm2-7 would be responsible for loading the second hexamer. Alternatively after OCCM formation, ORC could bind at a distal region of DNA and recruit a second Cdc6 and Mcm2-7/Cdt1 molecule. In this model, the second Mcm2-7 hexamer is recruited via the same mechanism as the first, yet is loaded in close enough proximity to the first Mcm2-7 complex that the proper N-terminal Mcm2-7 interactions form.

et al., 2015). Importantly, the kinetics of Cdc6 and Cdt1 release are different between first and second Mcm2-7 loadings, as would be expected if the first and second hexamers load via different mechanisms (Ticau et al., 2015). Different single-molecule experiments also support this idea. After initial ORC-DNA binding, no new ORC molecules were required to load replication-competent Mcm2-7 double hexamers (Duzdevich et al., 2015). Bulk biochemical assays also demonstrated that Cdt1 can interacts with two different domains of Orc6, which also supports the separate mechanism hypothesis (Chen et al., 2007).

The second model for second hexamer loading suggests that the first and second Mcm2-7 loading event occurs via the same mechanism (Figure 14, right). In this model, separate ORC molecules, binding in opposite orientations, each load a single Mcm2-7 complex through the same ORC-Cdc6-Mcm2-7-Cdt1 interactions. The loaded single Mcm2-7 complexes subsequently come together (by sliding on the DNA) to form a double hexamer. Biochemical data exists to support this model as well. The Mcm3 C-terminal residues, which are highly conserved and form an WH domain that interact with ORC/Cdc6, are important for loading both the first and the second double hexamer (Frigola et al., 2013). This study also found that Orc6 was not essential for recruitment of the first Mcm2-7 complex, although Orc6 was still essential for stable loading (Frigola et al., 2013). The B2 element of *ARS1,* being a close match to the ACS, has been hypothesized as a potential second ORC binding site (Frigola et al., 2013; Wilmes and Bell, 2002). Importantly, this sequence is an inverted match to the ACS, which would ensure the second ORC could load a second Mcm2-7 in the opposite orientation of the first hexamer (Wilmes and Bell, 2002). However, not all B2 elements are matches to the ACS (Chang et al., 2011) and ORC binding at the *ARS1* B2 site is not observed when the ACS is intact (Bell and Stillman, 1992). In fact, the presence of Orc6 sterically hinders two ORC complex from binding the two sites simultaneously (Lee and Bell, 1997; Wilmes and Bell, 2002).

Once the second Mcm2-7 loads at the origin, all associated loading proteins must detach from the Mcm2-7 double hexamer. In single-molecule experiments, the second Cdc6

departs first, followed by the single observed ORC and the second Cdt1 (Ticau et al., 2015). The second Mcm2/Mcm5 gate closing corresponds with ORC/Cdt1 departure (Ticau et al., 2017). ORC release is not observed in bulk biochemical experiments (although decreased amounts of ORC binding following Mcm2-7 binding have been reported occasionally (Tsakraklides and Bell, 2010)) or *in vivo,* likely due to rapid re-binding to the origin. Because of the Mcm2-7 double-hexamer's ability to slide on the DNA (Evrin et al., 2009; Remus et al., 2009), it is likely that once the Mcm2-7 hexamer is loaded, the complex can move away from the origin, allowing access for rebinding of ORC and potentially a new round of Mcm2-7 loading. Multiple rounds of Mcm2-7 loading have been reported *in vivo* and have been linked to origin timing (see section 4 below). *In vivo*, ORC binds origin DNA throughout the cell cycle, even though the complex is only able to recruit Mcm2-7 in G1 (for more on this regulation, see section 5.2). Although the reason for ORC-DNA binding outside of G1 is unknown, it may help maintain the NFR throughout the cell cycle.

3.2.8 The end result of origin licensing: The Mcm2-7 double hexamer

The result of the helicase loading process is a stable head-to-head Mcm2-7 double hexamer (Figure 15). This Mcm2-7 double hexamer is predicted to cover roughly 60 bps of dsDNA (Sun et al., 2014). A recent high-resolution EM structure of the Mcm2-7 double hexamer provides insight into both helicase loading and the rearrangements that need to occur during activation (Li et al., 2015). In this structure, the two Mcm2/Mcm5 gates are not aligned directly over on another (Figure 15A), supporting the idea that helicases are loaded one at a time (Li et al., 2015). Although structures of the Mcm2-7 single hexamer show that the N- and C-terminal domains are aligned (Zhai et al., 2017), in the double hexamer the N-terminal and C-terminal (AAA+) domains are rotated relative to one other (Figure 15B). This conformation has been proposed to store energy to be released upon activation (Zhai et al., 2017). Once in the double hexamer conformation, the Mcm2-7 double hexamer is stable, even without detectable DNA in

Figure 15

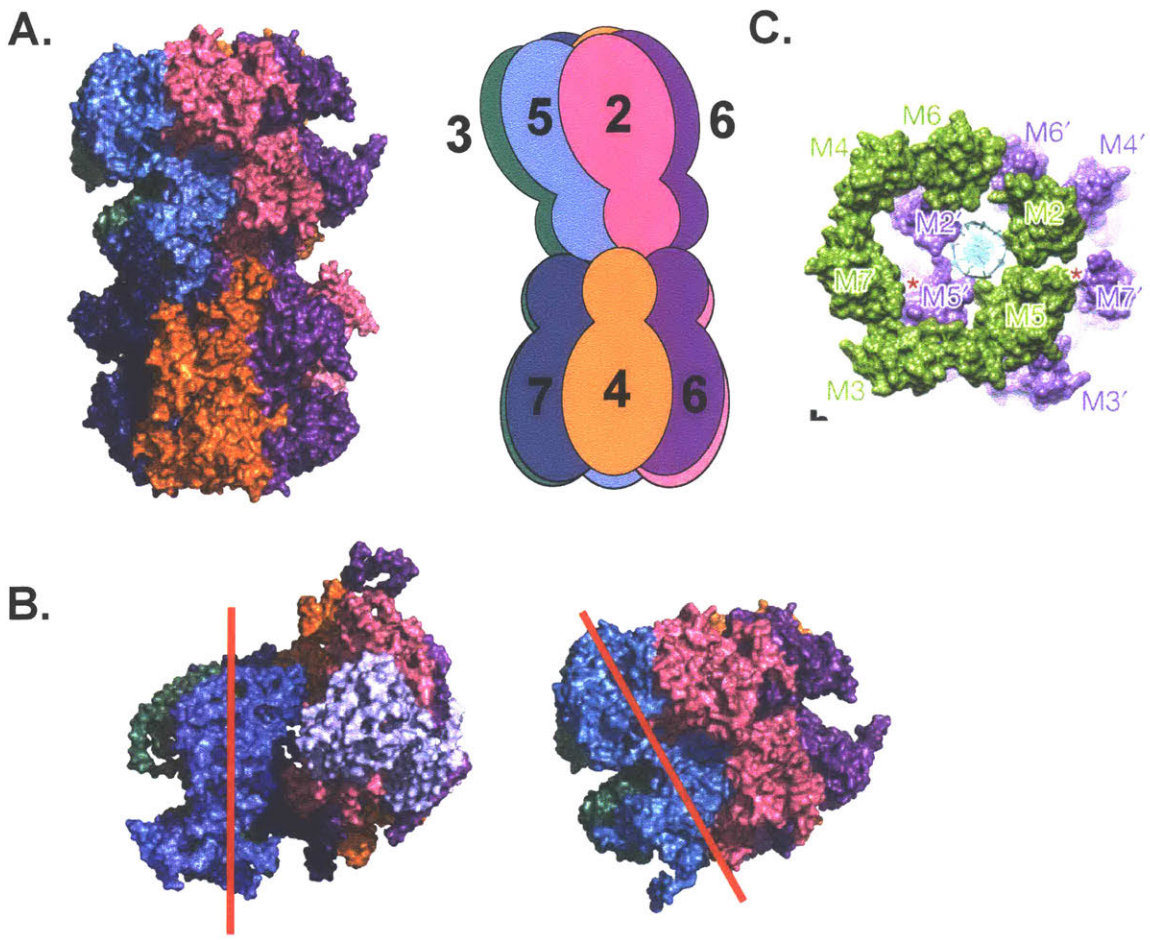


Figure 15: Structure of the stable, inactive Mcm2-7 double hexamer.

A) Structural side view (left) and diagram (right) of the Mcm2-7 double hexamer. Subunits are colored as in Figure 10, and labeled in the diagram.

B) Structural side view a single hexamer from the Mcm2-7/Cdt1 heptamer structure (left) compared to the same side view a single Mcm2-7 hexamer from the double hexamer structure (right). The red line shows the tilt of the Mcm5 C-terminus relative to the N-terminus in both structures.

C) End on view of the Mcm2-7 double hexamer highlighting the offset of the two hexamers and the relative location of the Mcm2/Mcm5 gates. Red stars show the hypothesized ssDNA capture locations upon DNA melting.

the central channel (Evrin et al., 2009). Interactions between the zinc finger domains and unique N-terminal insertions in Mcm3, Mcm5 and Mcm7 hold the two hexamers together tightly (Li et al., 2015).

The loaded double-hexamer is not an active helicase. However, the orientation of the two hexamers suggests possible mechanisms for initial activation. The central channel through the two Mcm2-7 rings is offset and the two hexamers are tilted at a 14 degree angle relative to each other (Figure 15A)(Li et al., 2015). Any dsDNA that would pass through both Mcm2-7 hexamers would therefore be in a strained conformation, aiding initial DNA melting. The Mcm2/Mcm5 gates are on opposition sides of each hexamer, and thus are positioned so that a different DNA strand could be extruded from each gate (Figure 15C). Rotation of the N and C domains back to an aligned position, a position observed in the active CMG helicase, would also be expected to provide an unwinding force (Costa et al., 2011, 2014; Yuan et al., 2016; Zhai et al., 2017).

3.2.9: The role of ATP binding and hydrolysis in the helicase loading process

Both ATP binding and ATP hydrolysis are essential for the successful loading of the Mcm2-7 complex. As discussed above, inhibition of all ATP hydrolysis by the addition of ATPγS, a slowly hydrolysable ATP analog, blocks the Mcm2-7 loading process at the OCCM complex. The ATP sites at the Orc1/Orc4, Cdc6/Orc1 and all the Mcm2-7 subunit interfaces are active ATPases, and the Orc4/Orc5 and Orc5/Orc3 interfaces can bind ATP. Much work has been done to investigate the effects of both ATP binding and hydrolysis during helicase loading. Here I summarize what is known about the role of ATP in helicase loading.

3.2.9a ORC ATP hydrolysis is important for multiple rounds of Mcm2-7 loading

Similar to its *E. coli* counterpart DnaA, ORC requires ATP binding, but not ATP hydrolysis to bind origin DNA (Bell and Stillman, 1992). The presence of the origin DNA, but not

other DNA sequences, specifically inhibits Orc1/Orc4 ATP hydrolysis (Klemm et al., 1997). Thus, ORC ATP hydrolysis may play a role in removing ORC from DNA when the complex binds at improper locations (Klemm et al., 1997). As mentioned above, although Orc1-5 are all AAA+ ATP-like, only a subset of the AAA+ sites (Orc 1/4, Orc4/5 and Orc5/3) are actually able bind ATP and only the Orc1/Orc4 ATPase site shows ATP hydrolysis activity. Hydrolysis of ATP by the ATP binding pocket at the Orc1/Orc4 interface is important for helicase loading *in vivo* (Bowers et al., 2004). However, mutating this ATPase site did not have an effect on salt-stable helicase loading *in vitro* (Bowers et al., 2004; Coster et al., 2014). Instead, ORC ATP hydrolysis may be important for release of the ORC complex from DNA after loading has been completed (Rowles et al., 1996; Tsakraklides and Bell, 2010). This observation would be consistent with other studies suggesting that ORC ATP hydrolysis is needed for multiple rounds of helicase loading (Bowers et al., 2004).

3.2.9b Cdc6 ATP hydrolysis is important for recycling improperly loaded Mcm2-7 and activation of the helicase

The role of ATP binding and hydrolysis on Cdc6 function is still not well understood. Although Cdc6 ATP binding is essential for helicase loading, (Coster et al., 2014; Perkins and Diffley, 1998; Weinreich et al., 1999) whether ATP hydrolysis is required is still unclear. Cdc6 ATP hydrolysis mutants are able to load salt-stable Mcm2-7 complexes *in vitro* and these Mcm2-7 complexes are competent for replication in extract-based *in vitro* experiments (Coster et al., 2014; Kang et al., 2014). However, helicase loading was impaired at low concentrations of Cdc6 (Kang et al., 2014). Based on these findings, Cdc6 ATP hydrolysis was proposed to serve a role in the release of incorrect or unsuccessful loading events (Coster et al., 2014; Frigola et al., 2013; Kang et al., 2014). However, these same Cdc6 mutants caused a G1 arrest *in vivo* (Chang et al., 2015). Unlike wild-type Cdc6, these Cdc6 mutants remained aberrantly associated with chromatin. Degradation of Cdc6 allowed S phase to proceed. Thus, Cdc6 ATP

hydrolysis may play a role in disassembling Cdc6 from the Mcm2-7 complex, and allowing the Mcm2-7 helicase to be activated (Chang et al., 2011).

3.2.9c The role of the Mcm2-7 ATP binding and hydrolysis in loading of the helicase.

ATP hydrolysis by Mcm2-7 is essential. Mutations in the Walker A or Walker B motifs of any of the six subunits are lethal, and only one of the arginine finger residues (the Mcm4 arginine finger) is non-essential (Bochman et al., 2008). Interestingly, ATP binding and hydrolysis by Mcm2-7 is required for helicase loading. Any mutation in an ATPase site, even the mutation of the non-lethal Mcm4 arginine finger, yielded defects in stable Mcm2-7 double-hexamer formation in *in vitro* helicase loading assays (Coster et al., 2014; Kang et al., 2014). The subset that can assemble limited amounts of Mcm2-7 double hexamers all have defects at various steps downstream of helicase loading, but prior to DNA unwinding (Kang et al., 2014). There was no correlation between which mutations (Walker A, Walker B or Arginine finger) would have a moderate versus a severe defect, and even mutations within the same interface showed defects at different steps (Coster et al., 2014; Kang et al., 2014). These data suggest a mechanism where ATP binding and hydrolysis play critical roles in multiple steps leading up to helicase activation. Recent single-molecule studies have identified ATP hydrolysis at the Mcm3/5 interface to be important for Mcm2/Mcm5 gate closure (discussed in detail in section 3.2.6) (Ticau et al., 2017). Examining other Mcm2-7 ATPase mutations in this assay may shed light on additional role ATP binding and hydrolysis plays during helicase loading and activation.

4. Helicase loading is common, helicase activation is rare: origin usage in *S. cerevisiae*

In E. coli, only two DnaB helicases are placed at the origin and they rapidly initiate unwinding. Yet, just as not all ACSs in the yeast genome are licensed, not all Mcm2-7 double hexamers initiate replication during any given cell cycle. Upon entry into S phase, increased levels of two kinases, the S-phase CDKS and Dbf4-dependent kinase (DDK) phosphorylate key

proteins, including the Mcm2-7 complex itself (reviewed in Labib and Bell, 2016). The end result of this process is the formation of the active CMG (Cdc45-Mcm2-7-GINS) helicase that unwinds the genome (Aparicio et al., 1997; Gambus et al., 2006; Ilves et al., 2010; Kanemaki and Labib, 2006; Kanemaki et al., 2003; Tercero et al., 2000). Yet, helicase activation occurs at only a subset of possible locations.

In yeast, it is estimated that there are roughly 600 unique locations where Mcm2-7 can be loaded (Siow et al., 2012) (ignoring 150 copies of the rDNA repeats on chromosome XII, which each contain a single origin). Only a quarter of these sites will initiate replication in any given cell cycle (Pasero et al., 2002). In addition, those origins that do initiate replication during S phase do not do so simultaneously. Some origins initiate early in S phase while others fire later (Raghuraman et al., 2001). Although initiation at early origins is sufficient to complete replication, late origins are thought to be important to ensure the genome completes replication under stressful conditions. If two adjacent replication forks stall prior to completing replication of the intervening DNA, a new origin must initiate between them to finish DNA synthesis (Ge et al., 2007). Because origins cannot be re-licensed in S phase (see section 5 for a discussion on how this regulation is accomplished), licensed late origins are required to accomplish this task.

Although the propensity of origins to initiate early or late during S phase could be the result of a pre-determined order of initiation, there is mounting evidence that origin timing is determined stochastically via a competition model (Bechhoefer and Rhind, 2012; Kaykov and Nurse, 2015). Single-cell studies of replication initiation show stochastic origin usage, including instances where classically early origins fire after classically late origins (Czajkowsky et al., 2008). As would be predicted by a competition-based model, increasing the amount of limiting replication factors *in vivo* shifted the time of late replicating origins to earlier in S phase (Mantiero et al., 2011; Tanaka et al., 2011). It has been proposed that one mechanism by which origins compete for these limiting factors is by increasing the number of Mcm2-7 helicases at a given origin (Das et al., 2015; Yang et al., 2010). Early origins would have more Mcm2-7 double

hexamers, and so would be better able to complete for this limiting factors. Computational modeling showed that this hypothesis could yield a pattern observed for origin timing in yeast (Yang et al., 2010), and *in vivo* experiments confirmed that reducing Mcm2-7 loading could cause a normally early origin to fire later (Das et al., 2015).

Although many origins appear to rely primarily on competition to determine their origin timing, a subset utilize additional active mechanisms to ensure they replicate at the correct time. These origins are usually within regions of the genome that serve a particular function, such as the centromere or the telomere (Natsume et al., 2013; Stevenson and Gottschling, 1999). There are also a subset of origins regulated by the forkhead box transcription factors. Forkhead appears to cluster certain origins into particular locations in the nucleus, which gives them an advantage when competing for limiting factors (Knott et al., 2012; Lõoke et al., 2013).

5. Preventing DNA re-replication

Once a helicase loads onto an origin, this binding location is the site (*E. coli*) or a potential site (eukaryotes) of new DNA synthesis. However, once DNA synthesis has initiated, it is critical that helicases are not immediately reloaded, as this would allow rapid and uncontrolled re-initiation of DNA replication. As discussed in the introduction, if done in an uncontrolled manner, re-initiation leads to double-strand breaks, gene duplication events, and potentially cancerous transformation. Here, I will briefly discuss mechanisms by which DNA replication is regulated both in *E. coli* and *S. cerevisiae*.

5.1 Preventing Re-replication in *E. coli*

Although DNA replication can reinitiate prior to cell division in *E. coli,* it is still important to ensure un-regulated re-initiation does not occur. Even under extremely fast growth conditions, bacterial cells ensure replication of the genome is properly timed and removal of this coordination is detrimental to cell fitness (reviewed in (Mott and Berger, 2007)). Although in

eukaryotes, regulation centers on separating helicase loading and helicase activation, in *E. coli,* DnaB is activated very rapidly once placed around ssDNA. Thus, regulation of *E. coli* DNA replication occurs at the earliest event in replication, the binding of DnaA to the origin, upstream of ssDNA formation and helicase loading.

Three mechanisms prevent DnaA re-association with *oriC.* First, binding of the SeqA protein to newly synthesized DNA at *oriC* blocks DnaA binding (Nievera et al., 2006). SeqA binds its DNA binding sequence in a specific methylation state this is indicative of recent replication (Brendler et al., 1995; Slater et al., 1995). SeqA slowly releases from the DNA, allowing the origin to be reset (Boye et al., 1996; von Freiesleben et al., 2000; Lu et al., 1994). Second, careful titration of DnaA levels is critical to prevent re-initiation. DnaA, in addition to being the replication initiator, is also a transcription factor and regulates its own expression. DnaA binds upstream of its own promoter and inhibits transcription of the dnaA gene (Atlung et al., 1985; Braun et al., 1985). Accumulation of a critical number of DnaA molecules at the origin is required for initiation. However, when levels of DnaA are sufficiently high for replication initiation, they are also high enough to inhibit new DnaA transcription (Speck et al., 1999). Finally, DnaA function is regulated by ATP hydrolysis. Although ADP-bound DnaA can bind the DnaA boxes, binding at these sites alone is insufficient to initiate replication (McGarry et al., 2004). ATP hydrolysis is triggered by factors involved with DNA synthesis, which helps to keep re-initiation from occurring (Katayama et al., 1998; Nishida et al., 2002).

5.2 Regulation of Eukaryotic Replication:

In eukaryotes, a key principle used to prevent re-replication is to limit origin licensing to G1 and helicase activation and new DNA synthesis to S phase. This separation ensures that replication occurs once and only once per cell division. In eukaryotes, both activation of the helicase and repression of origin licensing is regulated by changes in kinase activity. During G1, the cyclin-dependent kinase (CDK) levels are low. This condition is permissive for helicase

loading onto chromatin. At the onset of S phase, S-phase CDKs phosphorylate proteins that promote assembly of the active CMG helicase complex. At the same time, these kinases target and inactivate the helicase-loading apparatus to ensure reloading does not occur. Although the general principle of using kinase activity to simultaneously inhibiting new helicase loading while activating already loaded helicases is conserved from yeast to higher eukaryotes, the targets and mechanisms are different. In metazoans, there is also critical non-CDK-dependent mechanisms that target Cdt1 function (reviewed in (Arias and Walter, 2007)). This additional regulation is important, as there are DNA replication checkpoint mechanisms that inactivate CDK activity in metazoans. For simplicity, I will only discuss the *S. cerevisiae* mechanisms here.

S-CDK activity in *S. cerevisiae* inhibits reloading in three ways. First, phosphorylation of Mcm3 by S-CDK causes nuclear export of complexes not bound to DNA (Labib et al., 1999; Nguyen et al., 2000, 2001). Second, CDK phosphorylates Cdc6, leading to Cdc6 degradation (Drury et al., 2000). Without Cdc6, ORC cannot recruit or load interact with the Mcm2-7 complex (see section 3.2.3 above). Finally, phosphorylation of ORC inhibits the complex's ability to load Mcm2-7 helicases, at least in part due to inhibition of Mcm2-7/Cdt1 recruitment (Chen and Bell, 2011; Fernández-Cid et al., 2013; Frigola et al., 2013). Although robust re-replication occurs only when all three mechanisms are eliminated (Nguyen et al., 2001), eliminating even a single mechanism is detrimental. A few specific origins are able to re-initiate when only a subset of these mechanism are bypassed (Green and Li, 2005) and defects in long term health over many generations or within a population is observed when even a single control is bypassed (Diffley, 2010). Thus, although it was initially thought that these mechanisms were redundant, it is now clear that all these mechanism are important for ensuring re-initiation does not occur at hundreds of different origins over the long term.

To make this S-CDK-based system work properly, at no time can helicase loading and helicase activation ever be active simultaneously. Thus, Cdc6 degradation actually starts in late G1, activated initially by the G1-CDKS and maintained by the S-CDKs (Drury et al., 2000).

Because the G1-CDKs cannot activate loaded helicases, there is a period where neither origin licensing nor helicase activation can occur (Drury et al., 2000; Labib et al., 1999). Similar mechanisms are in place at the end of mitosis into the next G1. Degradation of the Dbf4 subunit of DDK (an essential kinase of helicase activation) occurs prior to the decreasing the levels of CDKs in the cell (Donaldson, 2000; Oshiro et al., 1999; Shirayama et al., 1999). In addition, all helicases loaded during the previous G1 are removed during S phase, as activated helicases are actively unloaded at replication termination, or are released from the DNA when a replication fork passes through where they are loaded (Santocanale and Diffley, 1996). These gaps where neither origin licensing nor helicase activation can occur are critical to insulate the cell from a situation where intermediate S-CDK activity leads to simultaneous helicase loading and activation.

6. Summary of thesis work

In Chapter 2, I will discuss my work dissecting the role of origin DNA in loading of the Mcm2-7 helicase. Although the site of ORC DNA binding has been defined to the ORC-ACS by DNase footprinting, gel shift assays and genome-wide immunoprecipitation experiments (Bell and Stillman, 1992; Belsky et al., 2015; Eaton et al., 2010; Xu et al., 2006), the location of initial Mcm2-7 loading remains unclear. Structural studies suggest Mcm2-7 binds adjacent to ORC. A particular sequence within origin DNA, the B2 element, has been hypothesized as the site of initial Mcm2-7. However, direct binding of the Mcm2-7 complex to B2 has never been reported.

To gain additional insights into the initial location of Mcm2-7 helicase loading, we assessed the impact of protein-DNA adducts at different sites along the origin DNA on Mcm2-7 loading. We found that many locations interfered with ORC binding (and thus subsequent Mcm2-7 association), but only one location interfered with Mcm2-7 loading alone. Interestingly, the adducts affecting ORC spanned a region much larger than anticipated by structural evidence. Creating origin DNA templates with blockages on either side of the ACS showed a

strikingly different pattern. With adducts placed both upstream and downstream of the ACS, blockages in the B2 element region to have defects in Mcm2-7 loading, but not ORC binding. Interestingly, this defect was dependent on the B2 element sequence, a dependence not observed without protein adducts. These data suggest helicase loading is a more dynamic process for Mcm2-7 helicase loading by ORC then structural evidence suggests.

7. References

Aparicio, O.M., Weinstein, D.M., and Bell, S.P. (1997). Components and Dynamics of DNA Replication Complexes in S. cerevisiae: Redistribution of MCM Proteins and Cdc45p during S Phase. Cell *91*, 59–69.

Arentson, E., Faloon, P., Seo, J., Moon, E., Studts, J.M., Fremont, D.H., and Choi, K. (2002). Oncogenic potential of the DNA replication licensing protein CDT1. Oncogene *21*, 1150–1158.

Arias, E.E., and Walter, J.C. (2007). Strength in numbers: preventing rereplication via multiple mechanisms in eukaryotic cells. Genes Dev. *21*, 497–518.

Arias-Palomo, E., O'Shea, V.L., Hood, I.V., and Berger, J.M. (2013). The Bacterial DnaC Helicase Loader Is a DnaB Ring Breaker. Cell *153*, 438–448.

Atlung, T., Clausen, E.S., and Hansen, F.G. (1985). Autoregulation of the dnaA gene of Escherichia coli K12. MGG Mol. Gen. Genet. *200*, 442–450.

Aye, Y., Li, M., Long, M.J.C., and Weiss, R.S. (2015). Ribonucleotide reductase and cancer: biological mechanisms and targeted therapies. Oncogene *34*, 2011–2021.

Azmi, I.F., Watanabe, S., Maloney, M.F., Kang, S., Belsky, J.A., MacAlpine, D.M., Peterson, C.L., and Bell, S.P. (2017). Nucleosomes influence multiple steps during replication initiation. Elife *6*, e22512.

Bechhoefer, J., and Rhind, N. (2012). Replication timing and its emergence from stochastic processes. Trends Genet. *28*, 374–381.

Bell, S.P., and Kaguni, J.M. (2013). Helicase loading at chromosomal origins of replication. Cold Spring Harb. Perspect. Biol. *5*, a010124.

Bell, S.P., and Labib, K. (2016). Chromosome Duplication in Saccharomyces cerevisiae. Genetics *203*, 1027–1067.

Bell, S.P., and Stillman, B. (1992). ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature *357*, 128–134.

Belsky, J.A., MacAlpine, H.K., Lubelsky, Y., Hartemink, A.J., and MacAlpine, D.M. (2015). Genome-wide chromatin footprinting reveals changes in replication origin architecture induced by pre-RC assembly. Genes Dev. *29*, 212–224.

Berbenetz, N.M., Nislow, C., and Brown, G.W. (2010). Diversity of Eukaryotic DNA Replication Origins Revealed by Genome-Wide Analysis of Chromatin Structure. PLoS Genet. *6*, e1001092.

Biswas, D., Takahata, S., Xin, H., Dutta-Biswas, R., Yu, Y., Formosa, T., and Stillman, D.J. (2008). A Role for Chd1 and Set2 in Negatively Regulating DNA Replication in Saccharomyces cerevisiae. Genetics *178*.

Bleichert, F., Balasov, M., Chesnokov, I., Nogales, E., Botchan, M.R., and Berger, J.M. (2013). A Meier-Gorlin syndrome mutation in a conserved C-terminal helix of Orc6 impedes origin recognition complex formation. Elife *2013*, 1–29.

Bleichert, F., Botchan, M.R., and Berger, J.M. (2015). Crystal structure of the eukaryotic origin recognition complex. Nature *519*, 321–326.

Bleichert, F., Botchan, M.R., and Berger, J.M. (2017). Mechanisms for initiating cellular DNA replication. Science 355, eaah6317.

Bochman, M.L., and Schwacha, A. (2009). The Mcm complex: unwinding the mechanism of a replicative helicase. Microbiol. Mol. Biol. Rev. *73*, 652–683.

Bochman, M.L., Bell, S.P., and Schwacha, A. (2008). Subunit organization of Mcm2-7 and the unequal role of active sites in ATP hydrolysis and viability. Mol. Cell. Biol. *28*, 5865–5873.

Bowers, J.L., Randell, J.C.W., Chen, S., and Bell, S.P. (2004). ATP hydrolysis by ORC catalyzes reiterative Mcm2-7 assembly at a defined origin of replication. Mol. Cell *16*, 967–978.

Boye, E., Stokke, T., Kleckner, N., and Skarstad, K. (1996). Coordinating DNA replication initiation with cell growth: differential roles for DnaA and SeqA proteins. Proc. Natl. Acad. Sci. U. S. A. *93*, 12206–12211.

Bramhill, D., and Kornberg, A. (1988). Duplex opening by dnaA protein at novel sequences in initiation of replication at the origin of the E. coli chromosome. Cell *52*, 743–755.

Braun, R.E., O'Day, K., and Wright, A. (1985). Autoregulation of the DNA replication gene dnaA in E. coli K-12. Cell *40*, 159–169.

Breier, A.M., Chatterji, S., and Cozzarelli, N.R. (2004). Prediction of Saccharomyces cerevisiae replication origins. Genome Biol. *5*, R22.

Brendler, T., Abeles, A., and Austin, S. (1995). A protein that binds to the P1 origin core and the oriC 13mer region in a methylation-specific fashion is the product of the host seqA gene. EMBO J. *14*, 4083–4089.

Broach, J.R., Li, Y.Y., Feldman, J., Jayaram, M., Abraham, J., Nasmyth, K.A., and Hicks, J.B. (1983). Localization and sequence analysis of yeast origins of DNA replication. Cold Spring Harb. Symp. Quant. Biol. *47 Pt 2*, 1165–1173.

Chang, F., Riera, A., Evrin, C., Sun, J., Li, H., Speck, C., Weinreich, M., Speck, C., Li, H., and Speck, C. (2015). Cdc6 ATPase activity disengages Cdc6 from the pre-replicative complex to promote DNA replication. Elife *4*, 20–25.

Chang, F.J., May, C.D., Hoggard, T., Miller, J., Fox, C. a., and Weinreich, M. (2011). High-resolution analysis of four efficient yeast replication origins reveals new insights into the ORC and putative MCM binding elements. Nucleic Acids Res. *39*, 6523–6535.

Chen, S., and Bell, S.P. (2011). CDK prevents Mcm2-7 helicase loading by inhibiting Cdt1 interaction with Orc6. Genes Dev. *25*, 363–372.

Chen, S., de Vries, M.A., and Bell, S.P. (2007). Orc6 is required for dynamic recruitment of Cdt1 during repeated Mcm2-7 loading. Genes Dev. *21*, 2897–2907.

Chesnokov, I., Remus, D., and Botchan, M. (2001). Functional analysis of mutant and wild-type Drosophila origin recognition complex. Proc. Natl. Acad. Sci. U. S. A. *98*, 11997–12002.

Clapier, C.R., and Cairns, B.R. (2009). The Biology of Chromatin Remodeling Complexes. Annu. Rev. Biochem. *78*, 273–304.

Costa, A., Ilves, I., Tamberg, N., Petojevic, T., Nogales, E., Botchan, M.R., and Berger, J.M. (2011). The structural basis for MCM2-7 helicase activation by GINS and Cdc45. Nat. Struct. Mol. Biol. *18*, 471–477.

Costa, A., Renault, L., Swuec, P., Petojevic, T., Pesavento, J.J., Ilves, I., MacLellan-Gibson, K., Fleck, R.A., Botchan, M.R., Berger, J.M., et al. (2014). DNA binding polarity, dimerization, and ATPase ring remodeling in the CMG helicase of the eukaryotic replisome. Elife *3*, e03273.

Coster, G., Frigola, J., Beuron, F., Morris, E.P., and Diffley, J.F.X. (2014). Origin Licensing Requires ATP Binding and Hydrolysis by the MCM Replicative Helicase. Mol. Cell 1–12.

Czajkowsky, D.M., Liu, J., Hamlin, J.L., and Shao, Z. (2008). DNA Combing Reveals Intrinsic Temporal Disorder in the Replication of Yeast Chromosome VI.

Das, S.P., Borrman, T., Liu, V.W.T., Yang, S.C.-H., Bechhoefer, J., and Rhind, N. (2015). Replication timing is regulated by the number of MCMs loaded at origins. Genome Res. *25*, 1886–1892.

Davey, M.J., Indiani, C., and O'Donnell, M. (2003). Reconstitution of the Mcm2-7p heterohexamer, subunit arrangement, and ATP site architecture. J. Biol. Chem. *278*, 4491–4499.

Diffley, J.F.X. (2010). The many faces of redundancy in DNA replication control. Cold Spring Harb. Symp. Quant. Biol. *75*, 135–142.

Diffley, J., and Stillman, B. (1988). Purification of a yeast protein that binds to origins of DNA replication and a transcriptional silencer. Proc. Natl. Acad. Sci. *85*, 2120–2124.

Diffley, J.F.X., Cocker, J.H., Dowell, S.J., and Rowley, A. (1994). Two steps in the assembly of complexes at yeast replication origins in vivo. Cell *78*, 303–316.

Dion, M.F., Kaplan, T., Kim, M., Buratowski, S., Friedman, N., and Rando, O.J. (2007). Dynamics of Replication-Independent Histone Turnover in Budding Yeast. Science (80-.). *315*, 1405–1408.

Donaldson, A.D. (2000). The yeast mitotic cyclin Clb2 cannot substitute for S phase cyclins in replication origin firing. EMBO Rep. *1*, 507–512.

Donovan, S., Harwood, J., Drury, L.S., and Diffley, J.F. (1997). Cdc6p-dependent loading of Mcm proteins onto pre-replicative chromatin in budding yeast. Proc. Natl. Acad. Sci. U. S. A. *94*, 5611–5616.

Drury, L.S., Perkins, G., and Diffley, J.F.X. (2000). The cyclin-dependent kinase Cdc28p regulates distinct modes of Cdc6p proteolysis during the budding yeast cell cycle. Curr. Biol. *10*, 231–240.

Duderstadt, K.E., Chuang, K., and Berger, J.M. (2011). DNA stretching by bacterial initiators promotes replication origin opening. Nature *478*, 209–213.

Duzdevich, D., Warner, M.D., Ticau, S., Ivica, N.A., Bell, S.P., and Greene, E.C. (2015). The Dynamics of Eukaryotic Replication Initiation: Origin Specificity, Licensing, and Firing at the Single-Molecule Level. Mol. Cell *58*, 483–494.

Eaton, M.L., Galani, K., Kang, S., Bell, S.P., and MacAlpine, D.M. (2010). Conserved nucleosome positioning defines replication origins. Genes Dev. *24*, 748–753.

Eisenberg, S., Civalier, C., and Tye, B.K. (1988). Specific interaction between a Saccharomyces cerevisiae protein and a DNA element associated with certain autonomously replicating sequences. Proc. Natl. Acad. Sci. U. S. A. *85*, 743–746.

Erzberger, J.P., Mott, M.L., and Berger, J.M. (2006). Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling. Nat. Struct. Mol. Biol. *13*, 676–683.

Evrin, C., Clarke, P., Zech, J., Lurz, R., Sun, J., Uhle, S., Li, H., Stillman, B., and Speck, C. (2009). A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. Proc. Natl. Acad. Sci. U. S. A. *106*, 20240–20245.

Fang, L., Davey, M.J., and O'Donnell, M. (1999). Replisome Assembly at oriC, the Replication Origin of E. coli, Reveals an Explanation for Initiation Sites outside an Origin. Mol. Cell *4*, 541–553.

Fernández-Cid, A., Riera, A., Tognetti, S., Herrera, M.C., Samel, S., Evrin, C., Winkler, C., Gardenal, E., Uhle, S., and Speck, C. (2013). An ORC/Cdc6/MCM2-7 Complex Is Formed in a Multistep Reaction to Serve as a Platform for MCM Double-Hexamer Assembly. Mol. Cell *50*, 577–588.

Flanagan, J.F., and Peterson, C.L. (1999). A role for the yeast SWI/SNF complex in DNA replication. Nucleic Acids Res. *27*, 2022–2028.

Foltman, M., Evrin, C., De Piccoli, G., Jones, R.C., Edmondson, R.D., Katou, Y., Nakato, R., Shirahige, K., and Labib, K. (2013). Eukaryotic replisome components cooperate to process histones during chromosome replication. Cell Rep. *3*, 892–904.

Fox, C.A., Loo, S., Rivier, D.H., Foss, M.A., and Rine, J. (1993). A transcriptional silencer as a specialized origin of replication that establishes functional domains of chromatin. Cold Spring Harb. Symp. Quant. Biol. *58*, 443–455.

von Freiesleben, U., Krekling, M., Hansen, F., and Lobner-Olesen, A. (2000). The eclipse period of Escherichia coli. Embo J. *19*, 6240–6248.

Frigola, J., Remus, D., Mehanna, A., and Diffley, J.F.X. (2013). ATPase-dependent quality control of DNA replication origin licensing. Nature *495*, 339–343.

Fujikawa, N., Kurumizaka, H., Nureki, O., Terada, T., Shirouzu, M., Katayama, T., and Yokoyama, S. (2003). Structural basis of replication origin recognition by the DnaA protein. Nucleic Acids Res. *31*, 2077–2086.

Fuller, R.S., Funnell, B.E., and Kornberg, A. (1984). The dnaA protein complex with the E. coli chromosomal replication origin (oriC) and other DNA sites. Cell *38*, 889–900.

Galletto, R., Jezewska, M.J., and Bujalowski, W. (2003). Interactions of the Escherichia coli DnaB Helicase Hexamer with the Replication Factor the DnaC Protein. Effect of Nucleotide Cofactors and the ssDNA on Protein–Protein Interactions and the Topology of the Complex. J. Mol. Biol. *329*, 441–465.

Gambus, A., Jones, R.C., Sanchez-Diaz, A., Kanemaki, M., van Deursen, F., Edmondson, R.D., and Labib, K. (2006). GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. Nat. Cell Biol. *8*, 358–366.

Ge, X.Q., Jackson, D.A., and Blow, J.J. (2007). Dormant origins licensed by excess Mcm2-7 are required for human cells to survive replicative stress. Genes Dev. *21*, 3331–3341.

Gille, H., and Messer, W. (1991). Localized DNA melting and structural pertubations in the origin of replication, oriC, of Escherichia coli in vitro and in vivo. EMBO J. *10*, 1579–1584.

Giordano-Coltart, J., Ying, C.Y., Gautier, J., and Hurwitz, J. (2005). Studies of the properties of human origin recognition complex and its Walker A motif mutants. Proc. Natl. Acad. Sci. U. S. A. *102*, 69–74.

Green, B.M., and Li, J.J. (2005). Loss of rereplication control in Saccharomyces cerevisiae results in extensive DNA damage. Mol. Biol. Cell *16*, 421–432.

Green, B.M., Finn, K.J., and Li, J.J. (2010). Loss of DNA Replication Control Is a Potent Inducer of Gene Amplification. Science (80-.). *329*.

Grimwade, J.E., Ryan, V.T., and Leonard, A.C. (2000). IHF redistributes bound initiator protein, DnaA, on supercoiled oriC of Escherichia coli. Mol. Microbiol. *35*, 835–844.

Gros, J., Kumar, C., Lynch, G., Yadav, T., Whitehouse, I., and Remus, D. (2015). Post-licensing Specification of Eukaryotic Replication Origins by Facilitated Mcm2-7 Sliding along DNA. Mol. Cell *60*, 797–807.

Huang, R.-Y., and Kowalski, D. (1996). Multiple DNA Elements in ARS305 Determine Replication Origin Activity in a Yeast Chromosome. Nucleic Acids Res. *24*, 816–823.

Huang, R.Y., and Kowalski, D. (1993). A DNA unwinding element and an ARS consensus comprise a replication origin within a yeast chromosome. EMBO J. *12*, 4521–4531.

Huang, H., Strømme, C.B., Saredi, G., Hödl, M., Strandsby, A., González-Aguilera, C., Chen, S., Groth, A., and Patel, D.J. (2015). A unique binding mode enables MCM2 to chaperone histones H3–H4 at replication forks. Nat. Struct. Mol. Biol. *advance on*, 618–626.

Ilves, I., Petojevic, T., Pesavento, J.J., and Botchan, M.R. (2010). Activation of the MCM2-7 Helicase by Association with Cdc45 and GINS Proteins. Mol. Cell *37*, 247–258.

Kanemaki, M., and Labib, K. (2006). Distinct roles for Sld3 and GINS during establishment and progression of eukaryotic DNA replication forks. EMBO J. *25*, 1753–1763.

Kanemaki, M., Sanchez-Diaz, A., Gambus, A., and Labib, K. (2003). Functional proteomic identification of DNA replication proteins by induced proteolysis in vivo. Nature *423*, 720–724.

Kang, S., Warner, M.D.D., and Bell, S.P.P. (2014). Multiple Functions for Mcm2-7 ATPase Motifs during Replication Initiation. Mol. Cell *55*, 655–665.

Katayama, T., Kubota, T., Kurokawa, K., Crooke, E., and Sekimizu, K. (1998). The Initiator Function of DnaA Protein Is Negatively Regulated by the Sliding Clamp of the E. coli Chromosomal Replicase. Cell *94*, 61–71.

Kawakami, H., Keyamura, K., and Katayama, T. (2005). Formation of an ATP-DnaA-specific initiation complex requires DnaA Arginine 285, a conserved motif in the AAA+ protein family. J. Biol. Chem. *280*, 27420–27430.

Kaykov, A., and Nurse, P. (2015). The spatial and temporal organization of origin firing during the S-phase of fission yeast. Genome Res. *25*, 391–401.

Klemm, R.D., Austin, R.J., and Bell, S.P. (1997). Coordinate Binding of ATP and Origin DNA Regulates the ATPase Activity of the Origin Recognition Complex. Cell *88*, 493–502.

Knott, S.R.V., Peace, J.M., Ostrow, A.Z., Gan, Y., Rex, A.E., Viggiani, C.J., Tavaré, S., and Aparicio, O.M. (2012). Forkhead Transcription Factors Establish Origin Timing and Long-Range Clustering in S. cerevisiae. Cell *148*, 99–111.

Kobori, J. a, and Kornberg, A. (1982). The Escherichia coli dnaC Gene Product. Biol. Chem. *257*, 13770–13775.

Kong, D., Coleman, T.R., DePamphilis, M.L., Abdurashidova, G., Deganuto, M., Klima, R., Riva, S., Biamonti, G., Giacca, M., Falaschi, A., et al. (2003). Xenopus origin recognition complex (ORC) initiates DNA replication preferentially at sequences targeted by Schizosaccharomyces pombe ORC. EMBO J. *22*, 3441–3450.

Labib, K., Diffley, J.F.X., and Kearsey, S.E. (1999). G1-phase and B-type cyclins exclude the DNA-replication factor Mcm4 from the nucleus. Nat. Cell Biol. *1*, 415–422.

Lee, D.G., and Bell, S.P. (1997). Architecture of the yeast origin recognition complex bound to origins of DNA replication. Mol. Cell. Biol. *17*, 7159–7168.

Li, N., Zhai, Y., Zhang, Y., Li, W., Yang, M., Lei, J., Tye, B.-K., and Gao, N. (2015). Structure of the eukaryotic MCM complex at 3.8 Å. Nature *524*, 186–191.

Lin, S., and Kowalski, D. (1997). Functional equivalency and diversity of cis-acting elements among yeast replication origins. Mol. Cell. Biol. *17*, 5473–5484.

Liontos, M., Koutsami, M., Sideridou, M., Evangelou, K., Kletsas, D., Levy, B., Kotsinas, A., Nahum, O., Zoumpourlis, V., Kouloukoussa, M., et al. (2007). Deregulated Overexpression of hCdt1 and hCdc6 Promotes Malignant Behavior. Cancer Res. *67*.

Lipford, J.R., and Bell, S.P. (2001). Nucleosomes Positioned by ORC Facilitate the Initiation of DNA Replication. Mol. Cell *7*, 21–30.

Liu, S., Balasov, M., Wang, H., Wu, L., Chesnokov, I.N., and Liu, Y. (2011). Structural analysis of human Orc6 protein reveals a homology with transcription factor TFIIB. Proc. Natl. Acad. Sci. U. S. A. *108*, 7373–7378.

Lo, Y.-H., Tsai, K.-L., Sun, Y.-J., Chen, W.-T., Huang, C.-Y., and Hsiao, C.-D. (2009). The crystal structure of a replicative hexameric helicase DnaC and its complex with single-stranded DNA. Nucleic Acids Res. *37*, 804–814.

Lõoke, M., Kristjuhan, K., Värv, S., and Kristjuhan, A. (2013). Chromatin-dependent and -independent regulation of DNA replication origin activation in budding yeast. EMBO Rep. *14*, 191–198.

Lu, M., Campbell, J.L., Boye, E., and Kleckner, N. (1994). SeqA: A negative modulator of replication initiation in E. coli. Cell *77*, 413–426.

Makowska-Grzyska, M., and Kaguni, J.M. (2010). Primase Directs the Release of DnaC from DnaB. Mol. Cell *37*, 90–101.

Mantiero, D., Mackenzie, A., Donaldson, A., and Zegerman, P. (2011). Limiting replication initiation factors execute the temporal programme of origin firing in budding yeast. EMBO J. *30*, 4805–4814.

Marahrens, Y., and Stillman, B. (1992). A yeast chromosomal origin of DNA replication defined by multiple functional elements. Science *255*, 817–823.

Marszalek, J., and Kaguni, J.M. (1994). DnaA protein directs the binding of DnaB protein in initiation of DNA replication in Escherichia coli. J. Biol. Chem. *269*, 4883–4890.

Matsui, M., Oka, A., Takanami, M., Yasuda, S., and Hirota, Y. (1985). Sites of dnaA protein-binding in the replication origin of the Escherichia coli K-12 chromosome. J. Mol. Biol. *184*, 529–533.

McGarry, K.C., Ryan, V.T., Grimwade, J.E., and Leonard, A.C. (2004). Two discriminatory binding sites in the Escherichia coli replication origin are required for DNA strand opening by initiator DnaA-ATP. Proc. Natl. Acad. Sci. U. S. A. *101*, 2811–2816.

Mott, M.L., and Berger, J.M. (2007). DNA replication initiation: mechanisms and regulation in bacteria. Nat. Rev. Microbiol. *5*, 343–354.

Mott, M.L., Erzberger, J.P., Coons, M.M., and Berger, J.M. (2008). Structural Synergy and Molecular Crosstalk between Bacterial Helicase Loaders and Replication Initiators. Cell *135*, 623–634.

Müller, P., Park, S., Shor, E., Huebert, D.J., Warren, C.L., Ansari, A.Z., Weinreich, M., Eaton, M.L., MacAlpine, D.M., and Fox, C.A. (2010). The conserved bromo-adjacent homology domain of yeast Orc1 functions in the selection of DNA replication origins within chromatin. Genes Dev. *24*, 1418–1433.

Natsume, T., Müller, C. a., Katou, Y., Retkute, R., Gierliński, M., Araki, H., Blow, J.J., Shirahige, K., Nieduszynski, C. a., and Tanaka, T.U. (2013). Kinetochores coordinate pericentromeric cohesion and early DNA replication by Cdc7-Dbf4 kinase recruitment. Mol. Cell *50*, 661–674.

Newlon, C.S., Collins, I., Dershowitz, A., Deshpande, A.M., Greenfeder, S.A., Ong, L.Y., and Theis, J.F. (1993). Analysis of replication origin function on chromosome III of Saccharomyces cerevisiae. Cold Spring Harb. Symp. Quant. Biol. *58*, 415–423.

Nguyen, V.Q., Co, C., Irie, K., and Li, J.J. (2000). Clb/Cdc28 kinases promote nuclear export of the replication initiator proteins Mcm2–7. Curr. Biol. *10*, 195–205.

Nguyen, V.Q., Co, C., and Li, J.J. (2001). Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. Nature *411*, 1068–1073.

Nievera, C., Torgue, J.J.-C., Grimwade, J.E., and Leonard, A.C. (2006). SeqA Blocking of DnaA-oriC Interactions Ensures Staged Assembly of the E. coli Pre-RC. Mol. Cell *24*, 581–592.

Nishida, S., Fujimitsu, K., Sekimizu, K., Ohmura, T., Ueda, T., and Katayama, T. (2002). A nucleotide switch in the Escherichia coli DnaA protein initiates chromosomal replication: evidnece from a mutant DnaA protein defective in regulatory ATP hydrolysis in vitro and in vivo. J. Biol. Chem. *277*, 14986–14995.

Oshiro, G., Owens, J.C., Shellman, Y., Sclafani, R.A., and Li, J.J. (1999). Cell cycle control of Cdc7p kinase activity through regulation of Dbf4p stability. Mol. Cell. Biol. *19*, 4888–4896.

Papamichos-Chronakis, M., and Peterson, C.L. (2012). Chromatin and the genome integrity network. Nat. Rev. Genet. *14*, 62–75.

Pasero, P., Bensimon, A., and Schwob, E. (2002). Single-molecule analysis reveals clustering and epigenetic regulation of replication origins at the yeast rDNA locus. Genes Dev. *16*, 2479–2484.

Perkins, G., and Diffley, J.F.X. (1998). Nucleotide-Dependent Prereplicative Complex Assembly by Cdc6p, a Homolog of Eukaryotic and Prokaryotic Clamp-Loaders. Mol. Cell *2*, 23–32.

Raghuraman, M.K., Winzeler, E.A., Collingwood, D., Hunt, S., Wodicka, L., Conway, A., Lockhart, D.J., Davis, R.W., Brewer, B.J., and Fangman, W.L. (2001). Replication Dynamics of the Yeast Genome. Science (80-.). *294*.

Randell, J.C.W.W., Bowers, J.L., Rodríguez, H.K., and Bell, S.P. (2006). Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase. Mol. Cell *21*, 29–39.

Rando, O.J., and Chang, H.Y. (2009). Genome-Wide Views of Chromatin Structure. Annu. Rev. Biochem. *78*, 245–271.

Rao, H., and Stillman, B. (1995). The origin recognition complex interacts with a bipartite DNA binding site within yeast replicators. Proc. Natl. Acad. Sci. *92*, 2224–2228.

Rao, H., Marahrens, Y., and Stillman, B. (1994). Functional conservation of multiple elements in yeast chromosomal replicators. Mol. Cell. Biol. *14*, 7643–7651.

Remus, D., Beuron, F., Tolun, G., Griffith, J.D., Morris, E.P., Diffley, J.F.X., Remus, D., and Beuron, F. (2009). Concerted Loading of Mcm2-7 Double Hexamers around DNA during DNA Replication Origin Licensing. Cell *139*, 719–730.

Rivera-Mulia, J.C., and Gilbert, D.M. (2016). Replicating Large Genomes: Divide and Conquer. Mol. Cell *62*, 756–765.

Roth, A., and Messer, W. (1995). The DNA binding domain of the initiator protein DnaA. EMBO J. *14*, 2106–2111.

Rowles, A., Chong, J.P.., Brown, L., Howell, M., Evan, G.I., and Blow, J.J. (1996). Interaction between the Origin Recognition Complex and the Replication Licensing Systemin Xenopus. Cell *87*, 287–296.

Rowley, a, Cocker, J.H., Harwood, J., and Diffley, J.F. (1995). Initiation complex assembly at budding yeast replication origins begins with the recognition of a bipartite sequence by limiting amounts of the initiator, ORC. EMBO J. *14*, 2631–2641.

Rozgaja, T.A., Grimwade, J.E., Iqbal, M., Czerwonka, C., Vora, M., and Leonard, A.C. (2011). Two oppositely oriented arrays of low-affinity recognition sites in oriC guide progressive binding of DnaA during Escherichia coli pre-RC assembly. Mol. Microbiol. *82*, 475–488.

Ryan, V.T., Grimwade, J.E., Nievera, C.J., and Leonard, A.C. (2002). IHF and HU stimulate assembly of pre-replication complexes at Escherichia coli oriC by two different mechanisms. Mol. Microbiol. *46*, 113–124.

Samel, S.A., Fernández-Cid, A., Sun, J., Riera, A., Tognetti, S., Herrera, M.C., Li, H., and Speck, C. (2014). A unique DNA entry gate serves for regulated loading of the eukaryotic replicative helicase MCM2-7 onto DNA. Genes Dev. *28*, 1653–1666.

Santocanale, C., and Diffley, J.F. (1996). ORC- and Cdc6-dependent complexes at active and inactive chromosomal replication origins in Saccharomyces cerevisiae. EMBO J. *15*, 6671–6679.

Schwacha, A., and Bell, S.P. (2001). Interactions between Two Catalytically Distinct MCM Subgroups Are Essential for Coordinated ATP Hydrolysis and DNA Replication. Mol. Cell *8*, 1093–1104.

Segal, E., and Widom, J. (2009). What controls nucleosome positions? Trends Genet. *25*, 335–343.

Seitz, H., Weigel, C., and Messer, W. (2000). The interaction domains of the DnaA and DnaB replication proteins of Escherichia coli. Mol. Microbiol. *37*, 1270–1279.

Seo, J., Chung, Y.S., Sharma, G.G., Moon, E., Burack, W.R., Pandita, T.K., and Choi, K. (2005). Cdt1 transgenic mice develop lymphoblastic lymphoma in the absence of p53. Oncogene *24*, 8176.

Della Seta, F., Ciafré, S.A., Marck, C., Santoro, B., Presutti, C., Sentenac, A., and Bozzoni, I. (1990). The ABF1 factor is the transcriptional activator of the L2 ribosomal protein genes in Saccharomyces cerevisiae. Mol. Cell. Biol. *10*, 2437–2441.

Shirayama, M., Tóth, A., Gálová, M., and Nasmyth, K. (1999). APCCdc20 promotes exit from mitosis by destroying the anaphase inhibitor Pds1 and cyclin Clb5. Nature *402*, 203–207.

Simmons, L.A., Felczak, M., and Kaguni, J.M. (2004). DnaA Protein of Escherichia coli: oligomerization at the E. coli chromosomal origin is required for initiation and involves specific N-terminal amino acids. Mol. Microbiol. *49*, 849–858.

Simpson, R.T. (1990). Nucleosome Positioning Can Affect the Function of a Cis-Acting DNA Element In Vivo. Nature *343*.

Siow, C.C., Nieduszynska, S.R., Muller, C.A., and Nieduszynski, C.A. (2012). OriDB, the DNA replication origin database updated and extended. Nucleic Acids Res. *40*, D682–D686.

Slater, S., Wold, S., Lu, M., Boye, E., Skarstad, K., and Kleckner, N. (1995). E. coli SeqA protein binds oriC in two different methyl-modulated reactions appropriate to its roles in DNA replication initiation and origin sequestration. Cell *82*, 927–936.

Speck, C., and Messer, W. (2001). Mechanism of origin unwinding: sequential binding of DnaA to double- and single-stranded DNA. EMBO J. *20*, 1469–1476.

Speck, C., and Stillman, B. (2007). Cdc6 ATPase activity regulates ORC??Cdc6 stability and the selection of specific DNA sequences as origins of DNA replication. J. Biol. Chem. *282*, 11705–11714.

Speck, C., Weigel, C., and Walter, M. (1999). ATP- and ADP-DnaA protein, a molecular switch in gene regulation. EMBO J. *18*, 6169–6176.

Speck, C., Chen, Z., Li, H., and Stillman, B. (2005). ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA. Nat. Struct. Mol. Biol. *12*, 965–971.

Stevenson, J.B., and Gottschling, D.E. (1999). Telomeric chromatin modulates replication timing near chromosome ends. Genes Dev. *13*, 146–151.

Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. Nat. Struct. Mol. Biol. *20*, 267–273.

Sun, J., Evrin, C., Samel, S.A., Fernández-Cid, A., Riera, A., Kawakami, H., Stillman, B., Speck, C., Li, H., Fernandez-Cid, A., et al. (2013). Cryo-EM structure of a helicase loading intermediate containing ORC-Cdc6-Cdt1-MCM2-7 bound to DNA. Nat. Struct. Mol. Biol. *20*, 944–951.

Sun, J., Fernandez-Cid, A., Riera, A., Sun, J., Fernandez-Cid, A., Riera, A., Tognetti, S., Yuan, Z., Stillman, B., Speck, C., et al. (2014). Structural and mechanistic insights into Mcm2-7 double-hexamer assembly and function. Genes Dev. *28*, 2291–2303.

Sutton, M.D., and Kaguni, J.M. (1997). Threonine 435 of Escherichia coli DnaA Protein Confers Sequence-specific DNA Binding Activity. J. Biol. Chem. *272*, 23017–23024.

Sutton, M.D., Carr, K.M., Vicente, M., and Kaguni, J.M. (1998). Escherichia coli DnaA protein. The N-terminal domain and loading of DnaB helicase at the E. coli chromosomal origin. J. Biol. Chem. *273*, 34255–34262.

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.

Tanaka, S., and Diffley, J.F.X. (2002). Interdependent nuclear accumulation of budding yeast Cdt1 and Mcm2–7 during G1 phase. Nat. Cell Biol. *4*.

Tanaka, S., Nakato, R., Katou, Y., Shirahige, K., and Araki, H. (2011). Origin association of Sld3, Sld7, and Cdc45 proteins is a key step for determination of origin-firing timing. Curr. Biol. *21*, 2055–2063.

Tercero, J.A., Labib, K., and Diffley, J.F. (2000). DNA synthesis at individual replication forks requires the essential initiation factor Cdc45p. EMBO J. *19*, 2082–2093.

Ticau, S., Friedman, L.J., Ivica, N.A., Gelles, J., and Bell, S.P. (2015). Single-molecule studies of origin licensing reveal mechanisms ensuring bidirectional helicase loading. Cell *161*, 513–525.

Ticau, S., Friedman, L.J., Champasa, K., Corrêa, I.R., Gelles, J., and Bell, S.P. (2017). Mechanism and timing of Mcm2–7 ring closure during DNA replication origin licensing. Nat. Struct. Mol. Biol. *24*, 309–315.

Tocilj, A., On, K.F., Yuan, Z., Sun, J., Elkayam, E., Li, H., Stillman, B., Joshua-Tor, L., Echols, N., Kuriyan, J., et al. (2017). Structure of the active form of human origin recognition complex and its ATPase motor module. Elife *6*, 30810–30821.

Toledo, L., Altmeyer, M., Rask, M., and Lukas, C. (2013). ATR prohibits replication catastrophe by preventing global exhaustion of RPA. Cell *155*, 1088–1103.

Torres, E.M., Sokolsky, T., Tucker, C.M., Chan, L.Y., Boselli, M., Dunham, M.J., and Amon, A. (2007). Effects of Aneuploidy on Cellular Physiology and Cell Division in Haploid Yeast. Science (80-.). *317*.

Tsakraklides, V., and Bell, S.P. (2010). Dynamics of pre-replicative complex assembly. J. Biol. Chem. *285*, 9437–9443.

Wang, G., Klein, M.G., Tokonzaba, E., Zhang, Y., Holden, L.G., and Chen, X.S. (2008). The structure of a DnaB-family replicative helicase and its interactions with primase. Nat. Struct. Mol. Biol. *15*, 94–100.

Weigel, C., Schmidt, A., Seitz, H., Tungler, D., Welzeck, M., and Messer, W. (1999). The N-terminus promotes oligomerization of the Escherichia coli initiator protein DnaA. Mol. Microbiol. *34*, 53–66.

Weinreich, M., Liang, C., and Stillman, B. (1999). The Cdc6p nucleotide-binding motif is required for loading mcm proteins onto chromatin. Proc. Natl. Acad. Sci. U. S. A. *96*, 441–446.

Wilmes, G.M., and Bell, S.P. (2002). The B2 element of the Saccharomyces cerevisiae ARS1 origin of replication requires specific sequences to facilitate pre-RC formation. Proc. Natl. Acad. Sci. U. S. A. *99*, 101–106.

Xu, W., Aparicio, J.G., Aparicio, O.M., and Tavaré, S. (2006). Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in S. cerevisiae. BMC Genomics *7*, 257–286.

Yang, N., and Xu, R.-M. (2013). Structure and function of the BAH domain in chromatin biology. Crit. Rev. Biochem. Mol. Biol. *48*, 211–221.

Yang, S.C.-H., Rhind, N., and Bechhoefer, J. (2010). Modeling genome-wide replication kinetics reveals a mechanism for regulation of replication timing. Mol. Syst. Biol. *6*.

Yuan, Z., Bai, L., Sun, J., Georgescu, R., Liu, J., O'Donnell, M.E., and Li, H. (2016). Structure of the eukaryotic replicative CMG helicase suggests a pumpjack motion for translocation. Nat. Struct. Mol. Biol. *23*, 217–224.

Yuan, Z., Riera, A., Bai, L., Sun, J., Nandi, S., Spanos, C., Chen, Z.A., Barbon, M., Rappsilber, J., Stillman, B., et al. (2017). Structural basis of Mcm2–7 replicative helicase loading by ORC–Cdc6 and Cdt1. Nat. Struct. Mol. Biol. *24*, 316–324.

Zhai, Y., Cheng, E., Wu, H., Li, N., Yung, P.Y.K., Gao, N., and Tye, B.-K. (2017). Open-ringed structure of the Cdt1–Mcm2–7 complex as a precursor of the MCM double hexamer. Nat. Struct. Mol. Biol. *24*, 300–308.

Zou, L., and Stillman, B. (2000). Assembly of a Complex Containing Cdc45p, Replication Protein A, and Mcm2p at Replication Origins Controlled by S-Phase Cyclin-Dependent Kinases and Cdc7p-Dbf4p Kinase. Mol. Cell. Biol. *20*, 3086–3096.

Chapter II:

Restricting access to *ARS1* origin DNA reveals helicase loading dynamics and a role for the B2 sequence element in Mcm2-7 helicase loading

Experiments contained in this chapter were performed by Megan Warner, with the exception of Figure 16, which was performed by Sukhyun Kang. The M.HpaII expression construct was a generous gift from NEB and J. Walter and the TALE expression construct was a generous gift from J. Gelles.

Introduction:

In eukaryotic organisms, DNA replication initiates from multiple locations termed origins of replication to ensure the entire genome replicates in a timely manner. These sites are selected during origin licensing, a process that loads the core enzyme of the eukaryotic replicative DNA helicase, the Mcm2-7 complex, at each origin of replication. In *S. cerevisiae,* three different proteins – the origin recognition complex (ORC), Cdc6, and Cdt1 – coordinate the licensing process. ORC recognizes the origin DNA directly. Cdc6 and Cdt1 coordinate with ORC and Mcm2-7 to sequentially load two Mcm2-7 hexamers. The end result is an inactive Mcm2-7 head-to-head double hexamer at origins of replication. Although topologically linked to the dsDNA and able passively slide, active DNA unwinding does not occur until entry into S phase. Never-the-less, the loaded Mcm2-7 double hexamers serve as the first marker of bidirectional initiation as they are poised to leave the origin in opposite directions (reviewed in (Bell and Labib, 2016)).

In *S. cerevisiae* cells, origins of replication are defined by a combination of specific sequences and local chromatin. All potential *S. cerevisiae* origins include an 11 bp match to the ARS consensus sequence (ACS) (Broach et al., 1983). In addition to the ACS, studied origins in *S. cerevisiae* also have AT-rich B elements (B1, B2 and sometimes B3) that further facilitate origin function *in vivo* (Marahrens and Stillman, 1992; Rao et al., 1994). Mutation of any of individual B elements reduces origin function, but when all B elements are mutated simultaneously, origin function is lost (Marahrens and Stillman, 1992). *In vivo*, these origin sequences reside within a nucleosome-free region (NFR) of ~125 bp that is flanked on either side by regularly spaced nucleosomes (Berbenetz et al., 2010; Eaton et al., 2010). Moving nucleosomes into the NFR negatively affects origin function and restoring accessibility restores function (Azmi et al., 2017; Simpson, 1990).

Although the functions of the B1 and B3 elements are clear, the function of B2 remains mysterious. Along with the ACS, the B1 element forms the bipartite ORC binding site found at

each origin (Bell and Stillman, 1992; Rao and Stillman, 1995; Rowley et al., 1995). B3 is only present at a subset of origins and serves as a binding site for Abf1 (Diffley and Stillman, 1988; Eisenberg et al., 1988). When present, Abf1 binding helps to establish the NFR at origins (Diffley and Stillman, 1988; Fox et al., 1993; Lipford and Bell, 2001). In contrast, many different hypotheses exist for the role of the B2 element. B2 elements, although lacking clear sequence conservation, have been found at all origins examined in detail. Despite differences in sequence and even length, identified B2 elements can be swapped between origins and their function preserved (Lin and Kowalski, 1997; Rao et al., 1994). Because of the AT-rich nature of B2, it was originally hypothesized to function as an unwinding element (Huang and Kowalski, 1993). However, helical stability does not correlate with B2 element function (Wilmes and Bell, 2002). It has also been postulated to serve as a second ORC binding site or as a site of initial Mcm2-7 association (Frigola et al., 2013; Lipford and Bell, 2001; Wilmes and Bell, 2002; Zou and Stillman, 2000).

The B2 element sequence is placed in a location within the NFR ideal for Mcm2-7 binding. The site of ORC-DNA binding includes the ACS and part of the B1 element as determined by DNase footprinting, gel shift assays and genome-wide chromatin IP experiments (Bell and Stillman, 1992; Belsky et al., 2015; Eaton et al., 2010; Rao and Stillman, 1995; Xu et al., 2006). Recent structural studies strongly suggest Mcm2-7 binding occurs directly adjacent to ORC binding. An early Mcm2-7 loading intermediate shows a complex of ORC and Cdc6 encircling DNA with an Mcm2-7/Cdt1 complex bound to ORC and encircling the adjacent DNA (Sun et al., 2013, 2014; Yuan et al., 2017). Although the orientation of this ORC-Cdc6-Cdt1-Mcm2-7 (OCCM) complex is not known with respect to the sequence elements of the origin, it is proposed that the Mcm2-7 complex would overlap and potentially bind to one or more B-elements. Indeed, there would be sufficient room to place a Mcm2-7 double hexamer in the B element region of the NFR while the double hexamer remains bound to ORC. Consistent with this idea, chromatin immunoprecipitation (ChIP) experiments reveal that disruption of the B2

element *in vivo* specifically inhibits Mcm2-7 but not ORC origin association (Lipford and Bell, 2001; Zou and Stillman, 2000). On the other hand, no role for B2 has been reported in *in vitro* loading experiments and direct binding experiments have not localized Mcm2-7 to B2 at any stage of the loading process.

The B2 element has also been proposed as a site for a second ORC binding event. *In vitro* helicase loading assays suggest that the two Mcm2-7 hexamers are loaded via the same mechanism, and would thus likely require a second ORC binding site (Frigola et al., 2013). The B2 element of *ARS1* is a close (9/11) match to the ACS and sequences that retained similarity to the ASC could substitute for B2 *in vivo* (Wilmes and Bell, 2002). However, despite their sequence similarity, ACS-similar B2 substitutions did not bind ORC *in vitro* (Wilmes and Bell, 2002). ORC can bind the B2 element in *ARS1,* but cannot bind both the ACS and B2 simultaneously without the deletion of Orc6 (Bell and Stillman, 1992; Lee and Bell, 1997; Wilmes and Bell, 2002). Thus, the role of the B2 element in the helicase loading process remains mysterious.

To gain additional insights into helicase loading, we assessed how placement of protein-DNA adducts at different sites along the origin DNA impact Mcm2-7 loading. We found that all but one of our positioned adducts interfered with both Mcm2-7 loading and ORC-DNA binding. Interestingly, the locations affecting ORC binding spanned a region larger than anticipated by current structural evidence but consistent with ORC DNase I footprinting. Although single protein blockages within the B elements did not affect helicase association or loading, creating origin DNA templates with blockages on either side of the origin showed a strikingly different pattern. Adding a second blockage upstream of the ACS simultaneously with one in the B2 element region caused severe defects in Mcm2-7 loading without altering ORC DNA binding. Interestingly, the observed helicase loading onto the space-restricted origins was dependent on the B2 element sequence, despite B2 element sequence mutations having little effect on helicase loading when the origin was unmodified. Creating a restricted origin with nucleosomes

rather than artificial protein blockages also yielded a B2-specific helicase loading defect. These data indicate that Mcm2-7 helicase loading by ORC is a more dynamic process than current structural evidence suggests and that the B2 sequence element becomes important when space around the origin is restricted.

Results:

A protein adduct within origin DNA can disrupt helicase loading

Structural studies of the ORC-Cdc6-Cdt1-Mcm2-7 (OCCM) helicase-loading intermediate reveal that ORC/Cdc6 and Mcm2-7/Cdt1 are bound to one another as interacting ring structures encircling adjacent regions of double-stranded DNA at the origin (Sun et al., 2013; Yuan et al., 2017). In this structure, ORC/Cdc6 encircles 24 bps of DNA. The ORC/Cdc6 ring sits adjacent to the Mcm2-7 ring, which suggests different regions of the origin DNA contribute to ORC/Cdc6 DNA binding versus subsequent Mcm2-7/Cdt1 recruitment. Despite its relatively high-resolution, the sequence of the associated DNA in the OCCM is not discernable. Thus, although all evidence indicates that ORC/Cdc6 is bound to the ACS, where Mcm2-7/Cdt1 is located relative to ORC/Cdc6 within the asymmetric origin sequence is still unknown.

To investigate the location of Mcm2-7/Cdt1 relative to ORC during helicase loading, we used protein-DNA adducts to interfere with protein binding. We chose this approach instead of DNA mutations for two reasons: (1) previous studies have shown that loaded Mcm2-7 proteins are able to slide on DNA, strongly suggesting that Mcm2-7 does not make sequence-specific contacts with the DNA; (2) protein-adducts are known to prevent Mcm2-7 from encircling DNA (Duxin et al., 2014; Fu et al., 2011; Long et al., 2014). To create site-specific protein-DNA adducts, we crosslinked the HpaII methyltransferase (M.HpaII) to DNA. M.HpaII can be covalently cross-linked with high efficiency to a M.HpaII consensus motif (CCGG) that includes a 5-fluoro-2'-deoxycytidine (5FdC) at an internal cytosine (Chen et al., 1991). Importantly,

Figure 1

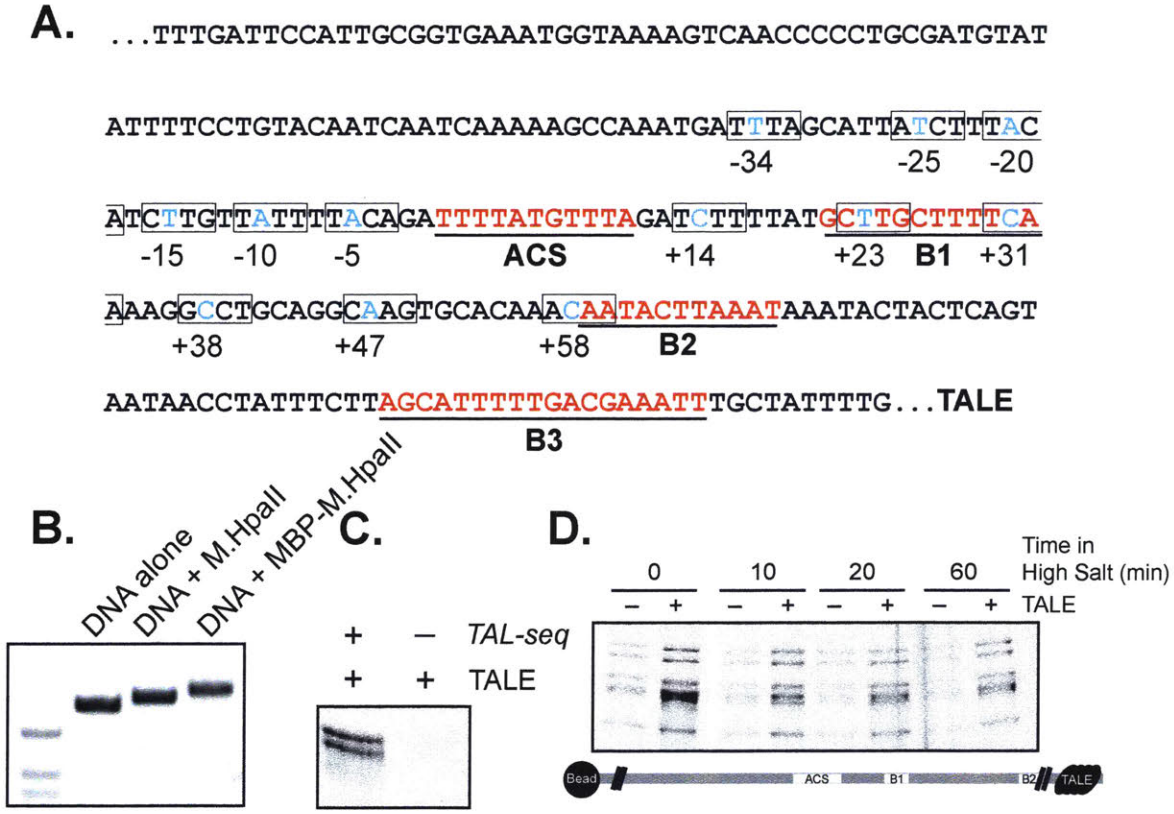


Figure 1:

A) Sequence 5' to 3' of the T-rich ARS1 DNA strand showing the locations of 5FdC-modified M.HpaII consensus sequence placement. Important features of *ARS1* are highlighted in red and labeled. Grey boxes identify sequences that were modified to the CCGG M.HpaII recognition motif. DNA residues in blue represent the exact location of the 5FdC to which the M.HpaII becomes crosslinked. The number assigned to that location is written below each box.

B) M.HpaII binding is quantitative. M.HpaII or MBP-M.HpaII was incubated with DNA containing a 5FdC-modified M.HpaII recognition sequence and resulting template run on a 1.5% agarose gel containing ethidium bromide. All detectable DNA template exhibits a shift when incubated with either M.HpaII or MBP-M.HpaII, suggesting quantitative levels of DNA-protein crosslinking, as has been previously reported (Duxin et al., 2014, Long et al., 2014)

C) TALE protein binding is sequence-specific. DNA containing or lacking a TALE binding motif were incubated with purified TALE protein for 15 minutes, followed by a high salt wash. TALE binding was only observed when the binding motif (*TAL-seq*) was included.

D) TALE binding prevents sliding of Mcm2-7 off the free end of the DNA template. The location of TALE is shown by the black cylinder. Substantially more helicases remain on the DNA when TALE is bound to the end of the template even at the end of the 30 min loading reaction (performed at 300mM potassium glutamate). These Mcm2-7 complexes are retained over a 60 minute incubation in high salt in the presence of TALE. Without the TALE protein, Mcm2-7 signal was steadily lost. This is consistent with previous reports (Evrin et al, 2009, Remus et al., 2009).

crosslinking the 45 kDa M.HpaII to the translocated ssDNA has previously been shown to stop a moving replication fork (Duxin et al., 2014; Long et al., 2014), presumably by preventing Mcm2-7 helicase from encircling the modified DNA. Thus, the attachment of M.HpaII to double-stranded DNA should inhibit the central channel of Mcm2-7 or ORC/Cdc6 from encircling the M.HpaII modified DNA (Li et al., 2015; Yuan et al., 2017).

We initially crosslinked M.HpaII to 12 different sites within the *ARS1* origin of replication (Figure 1A and Figure 2A). In each case, we substituted the 5FdC-modified M.HpaII consensus sequence for a four base pair sequence in the origin (Figure 1A). Crosslinking with M.HpaII was highly efficient, yielding quantitative shifts by agarose gel (Figure 1B). In all our experiments, the 0 position is the first base pair of the T-rich strand of the ACS. Positive numbers indicate adduct locations in the B element region (downstream of the ACS), while negative numbers indicate adducts on the other side (upstream of the ACS) (Figure 1A, Figure 2A). The OCCM structure suggests that Mcm2-7 encircles the DNA adjacent to ORC. Previous studies on the Mcm2-7 inactive double hexamer show that each Mcm2-7 should cover around 30 bps (Li et al., 2015; Sun et al., 2014). Thus, our initial modification sites spanned the region 30 bp on either side of asymmetric ORC binding site (the ACS and B1 motifs). We did not modify sequences within the most important residues in the ACS and B1 elements to avoid interfering with ORC binding and origin function in the absence of added M.HpaII (Liachko et al., 2013; Marahrens and Stillman, 1992) (Figure 1A, Figure 2A). Because Mcm2-7 can slide off the end of linear DNA (Evrin et al., 2009; Remus et al., 2009), we also introduced a binding site for an engineered sequence-specific TALE (Transcription activator-like effector) protein at the non-bead-attached end of each template (Figure 2A). TALE protein binding was dependent on the presence of the TALE sequence motif and stable to the high-salt (500mM NaCl) extraction condition utilized to assess the formation of a topologically linked Mcm2-7 complex (Donovan et al., 1997; Randell et al., 2006) (Figure 1C). Importantly, addition of TALE to these templates inhibits sliding of the loaded Mcm2-7 complex off the DNA (Figure 1D).

Figure 2

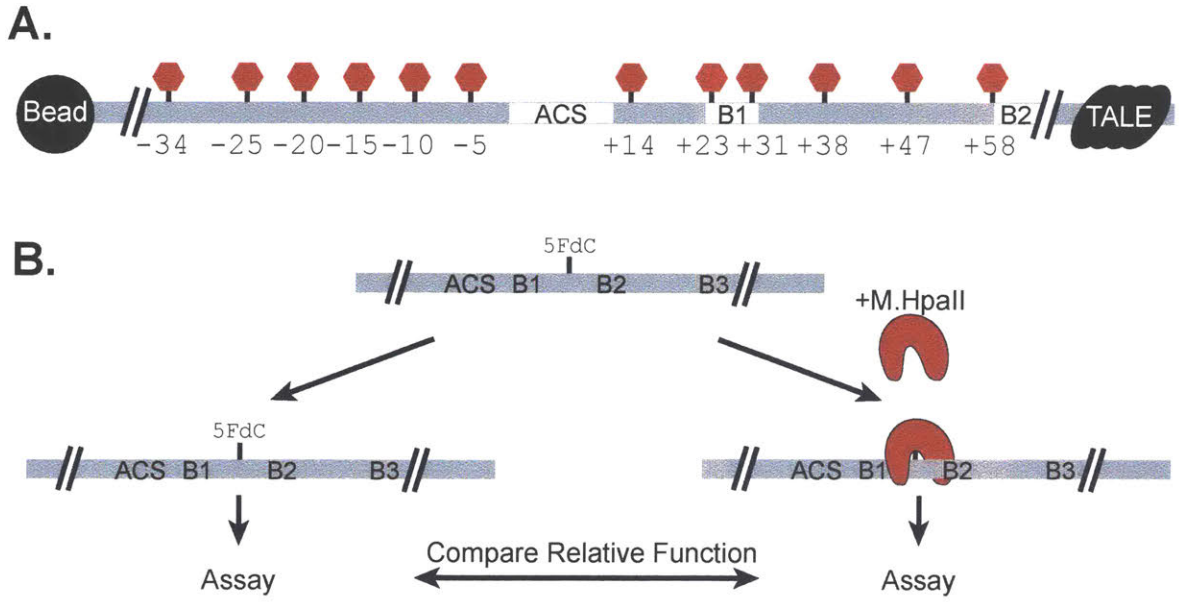


Figure 2: Using M.HpaII modification of DNA templates to assess the helicase loading process.

A)The M.HpaII consensus sequence was placed at 12 different locations on both sides of the ACS in *ARS1*. Important origin sequences are diagramed. Possible M.HpaII crosslinking locations are shown as red stop signs. Each template in the experiments below contained only one such modification. All templates contain a TALE protein bound to its consensus motif at the free end of the template to prevent helicases from sliding off the linear end.

B) To eliminate effects due to differences in DNA bead preparation or from introduction of the M.HpaII consensus sequence, relative changes between identical DNA templates either treated or untreated with M.HpaII were compared and quantified.

We tested each of the M.HpaII-crosslinked templates for helicase-loading *in vitro* (Evrin et al., 2009; Kang et al., 2014; Remus et al., 2009). Origin-containing DNA templates attached to magnetic beads were incubated with helicase-loading proteins, and the formation of stable, topologically linked Mcm2-7 complexes were assessed by high-salt extraction (Donovan et al., 1997; Randell et al., 2006). To ensure that any defects observed were due to the presence of a M.HpaII protein adduct and not insertion of the M.HpaII target sequence alone, we compared the extent of Mcm2-7 helicase loading with and without M.HpaII-crosslinking to the same DNA-bead preparation (Figure 2B, Figure 3). In each case, we report the ratio of helicase loading with and without M.HpaII modification (Figure 2B, Figure 3B). Mutations in the ACS and B2 sequence elements (A-B2-) eliminated detectable helicase loading, confirming that our assays were fully dependent on a functional origin sequence (Figure 3A compare lanes 1 and 3). Addition of the M.HpaII protein to DNA without a C(5FdC)GG motif had no effect on helicase loading (Figure 3A, compare lanes 1 and 2).

Formation of stable, topologically linked Mcm2-7 complexes was defective when M.HpaII adducts were placed between -15 and +38 bp relative to the first base of the ACS (Figure 3A and Figure 3B). The strongest M.HpaII-adduct inhibition was observed for the +14, +23 and +31 templates (Figure 3A, lanes 17-18, 21-24, quantified in Figure 3B), which are located in and around the ACS and B1 element of *ARS1* (inset diagram in Figure 3A). Approximately two-fold defects were observed for the +38, -5 and -10 protein adducts (Figure 3A, lanes 13-16, 25-26, quantified in Figure 3B) and a weak, but reproducible defect was observed when the M.HpaII adduct was placed at -15 (Figure 3A lanes 11-12, quantified in Figure 3B). Protein adducts located outside of the -15 and +38 bp region showed no defects. Doubling the size of the protein adduct by fusing MBP to M.HpaII (M.HpaII = 45 kDa; M.HpaII-MBP fusion = 80 kDa), crosslinked just as efficiently to DNA (Figure 1B) and did not expand the range of adducts that inhibited helicase loading, arguing against the hypothesis that ORC/Cdc6 or Mcm2-7 can accommodate the smaller M.HpaII within their central channels (Figure 4).

Figure 3

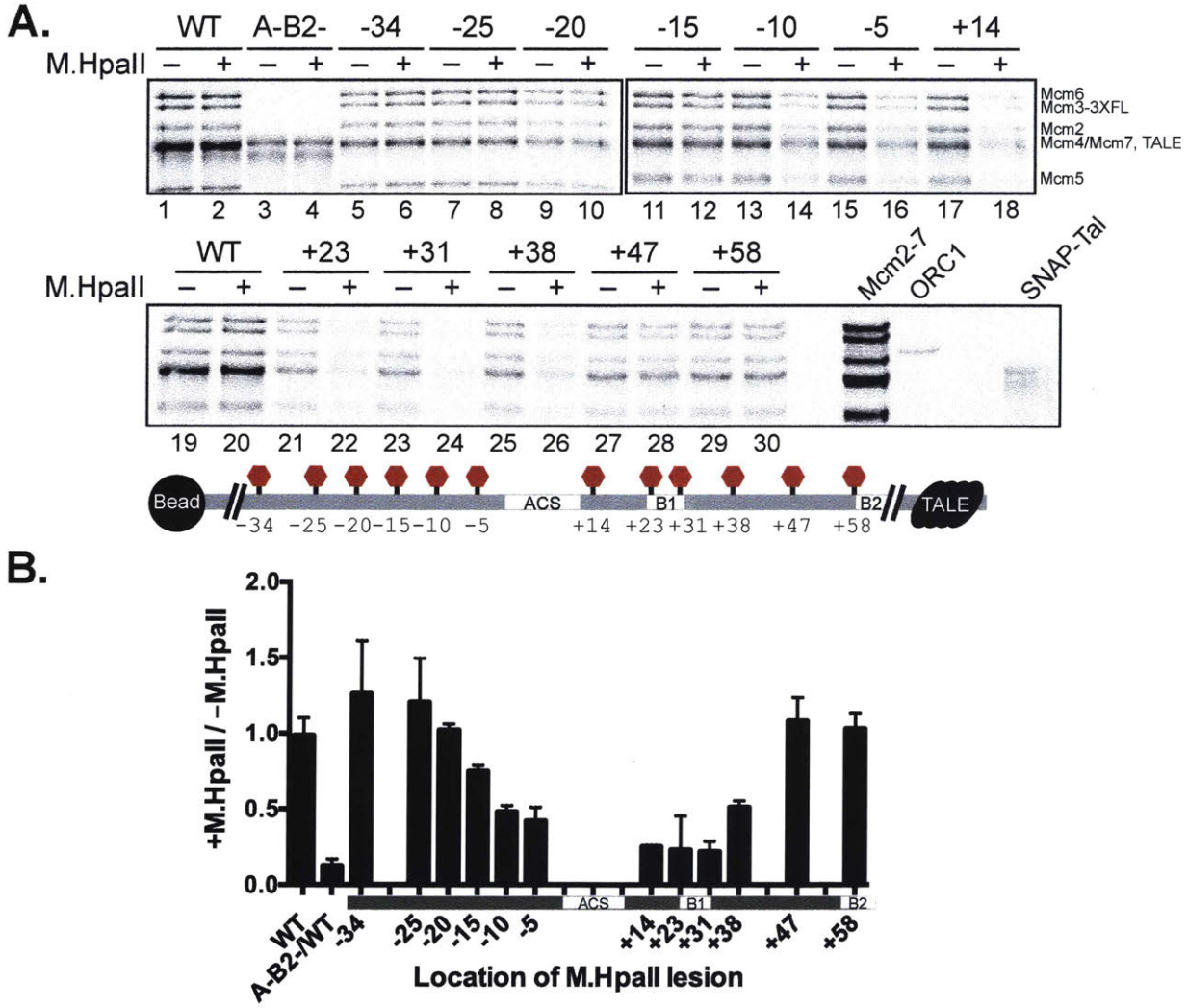


Figure 3: M.HpaII modification at a subset of locations within the *ARS1* origin sequence negatively affect Mcm2-7 helicase loading.

A) Stable helicase loading was assessed on templates containing M.HpaII adducts. Templates were either treated with M.HpaII (+) or untreated (-). The location of the 5FdC-modified M.HpaII motif is noted above each pair of lanes. Prior to helicase loading, all DNA templates were treated with a site-specific TALE protein, to retain loaded Mcm2-7 complexes. Templates were then incubated with helicase loading proteins for 30 minutes and the formation of stable Mcm2-7 loaded assessed by a high-salt wash. The stably associated proteins were released from the DNA beads with DNase and run on an SDS-PAGE gel, followed by fluorescent protein staining.

B) Comparative levels of Mcm2-7 loading across origins containing a 5FdC-modified M.HpaII recognition motif at the indicated location. Amount of helicase loading is reported as a fraction of loading when treated with M.HpaII relative to the level of loading on the same template without treatment. Error bars are the standard deviation from the mean calculated from at least two independent experiments.

Figure 4

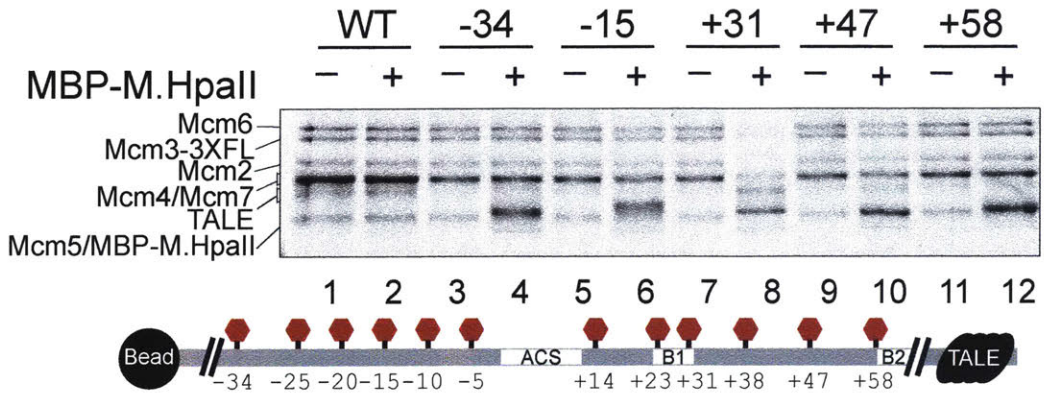


Figure 4: Increasing the size of M.HpaII does not affect helicase loading.

ARS1 origin templates containing the M.HpaII recognition motif at the indicated location were incubated with a MBP-M.HpaII (80 kDa), and a loading assay was performed as described in Figure 2C.

M.HpaII adducts near the ACS and B1 affect ORC binding

Since stable Mcm2-7 loading depends on the association of ORC and Cdc6 on the DNA prior to Mcm2-7/Cdt1 binding (Evrin et al., 2009; Randell et al., 2006; Remus et al., 2009), we sought to determine what step during origin licensing was inhibited by the different sites of M.HpaII modification. We first asked whether the OCCM complex could form in the presence of ATPγS (Randell et al., 2006). Unlike our observations after Mcm2-7 loading, no TALE-dependent Mcm2-7 retention was observed in ATPγS (Figure 5), so the TALE protein was not included in these assays (Figure 6A inset). Helicase loading was still dependent on a functional ACS in the presence of ATPγS (Figure 6A, compare lanes 1 and 3, quantified in 6B). A very similar subset of M.HpaII adducts affected OCCM formation. We observed nearly complete elimination of OCCM formation when M.HpaII was placed at the +14 and +31 locations, strong defects for the +23 and +38 adducts and a two-fold decrease at the -5 location (Figure 6A, lanes 11 to 20, quantified in 6B). Despite having a two-fold defect in helicase loading (Figure 3), placing an adduct at -10 showed only a small defect in Mcm2-7 association in ATPγS. The remaining adducts exhibited no defect.

Since ATPγS arrests helicase loading at the OCCM, these assays also monitored defects in ORC association. At all locations displaying a helicase-recruitment defect (-10, -5, +14, +23, +31 and +38), a corresponding decrease in ORC binding was also observed (Figure 6A, Lanes 11 to 20, compare ORC1 or ORC2 bands, quantified in Figure 6B). This observation suggested that the loss of stable double-hexamer formation observed in ATP at these locations (Figure 3) was caused by inhibition of ORC-DNA binding rather defects in Mcm2-7 recruitment alone. To test this hypothesis, we measured ORC binding to M.HpaII-modified templates directly using an electrophoretic-mobility-shift assay (EMSA) (Rao and Stillman, 1995). Radiolabeled *ARS1* DNA probes (~260bp) containing the same set of modified M.HpaII consensus motif insertions were incubated with various amounts of ORC with or without prior M.HpaII crosslinking. As with the previous experiments, we compared ORC-DNA binding with

Figure 5

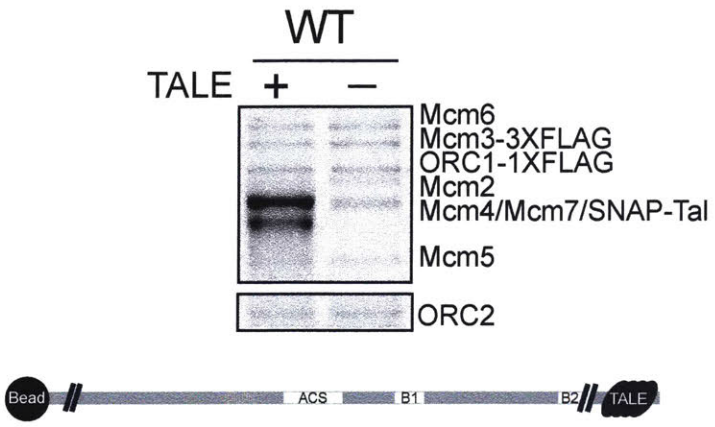


Figure 5: No sliding of the ORC-Cdc6-Cdt1-Mcm2-7 (OCCM) complex is observed in ATPγS.

An OCCM formation assay was performed on wildtype origin templates with and without a TALE protein bound to the free end of the DNA.

Figure 6

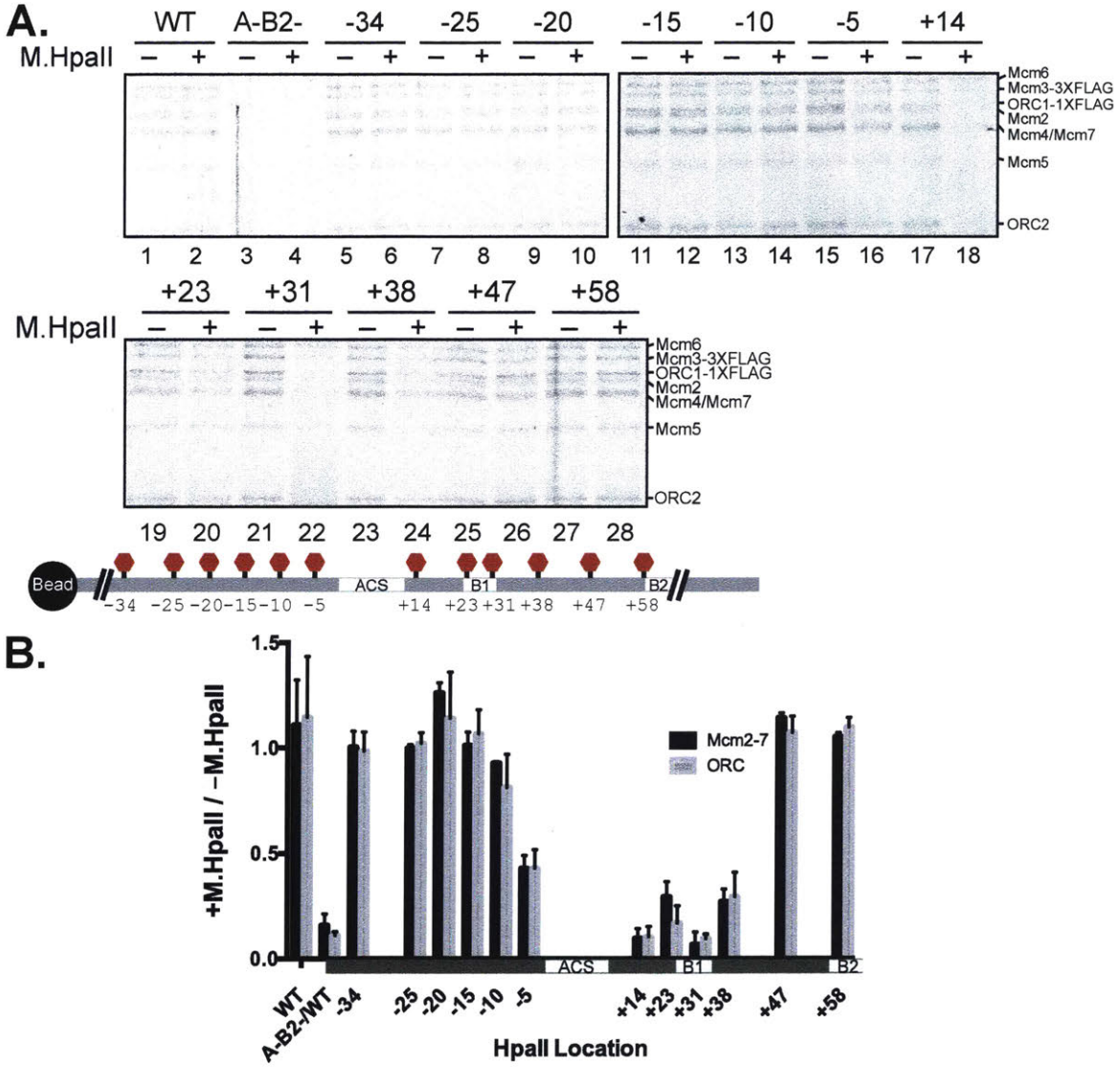


Figure 6: M.HpaII modifications affect both ORC and Mcm2-7 association in ATPγS

A) Different M.HpaII-containing templates were assessed for their ability to form the OCCM complex in the presence of ATPγS. Templates were either treated (+) or untreated (-) as indicated. The location of the M.HpaII recognition sequence is noted above each pair of lanes. Templates were incubated with helicase loading proteins for 30 minutes in the presence of ATPγS and the complex washed with low salt to remove excess proteins. The associated proteins were released from the DNA beads with DNAse and run on an SDS-PAGE gel, followed by fluorescent protein staining.

B) Comparative levels of Mcm2-7 and Orc2 association in ATPγS across templates crosslinked to M.HpaII at the indicated location. Amount of Mcm2-7 and Orc2 association are reported as a fraction of association when treated with M.HpaII relative to the amount of association on the same template without M.HpaII treatment. Error bars are the standard deviation from the mean calculated from at least two independent experiments.

Figure 7

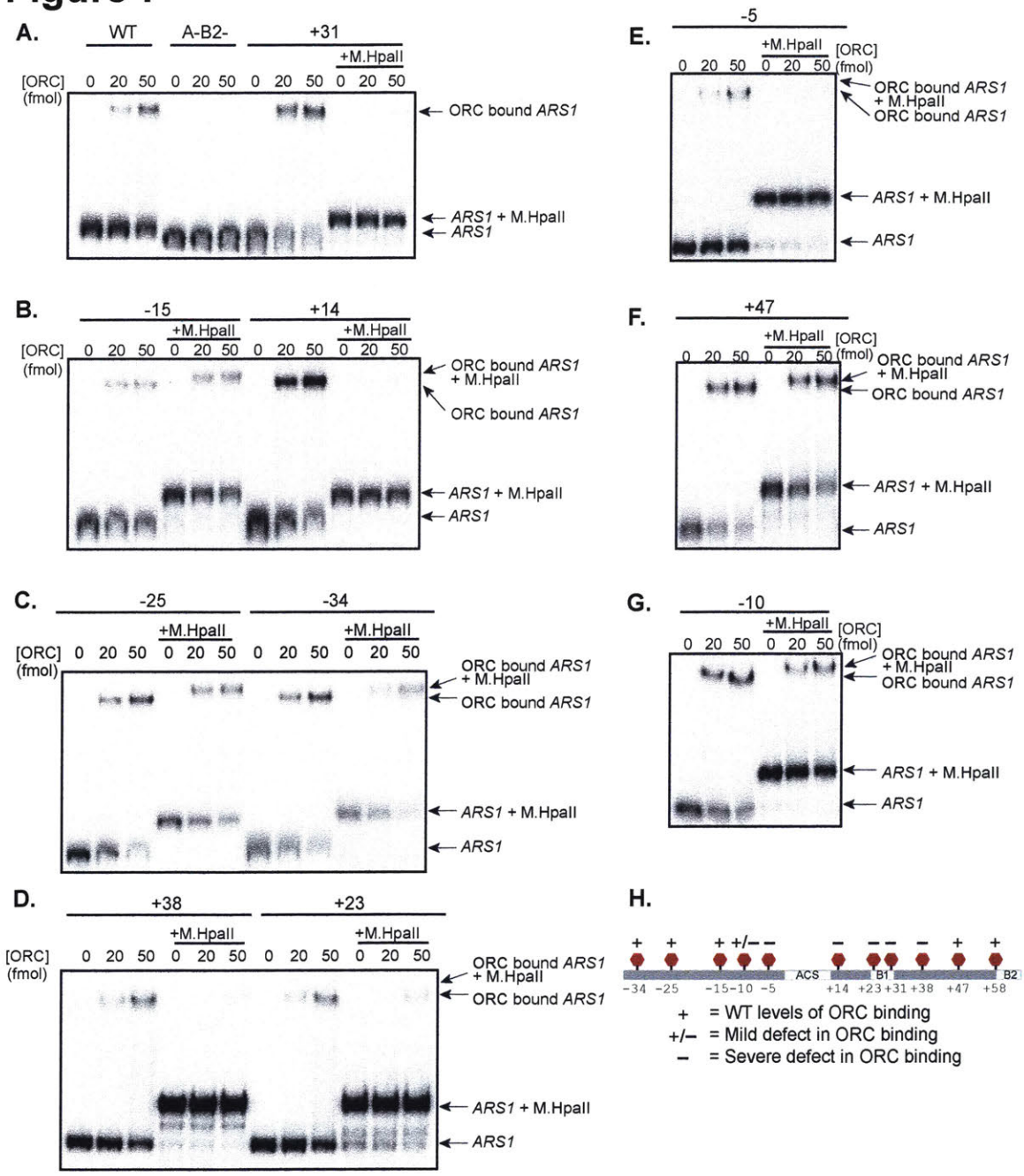


Figure 7: M.HpaII modification at the -5 through +38 locations directly affects ORC binding by EMSA

(A-G) 260 bp *ARS1* origin-containing dsDNA fragments were radiolabeled and incubated with increasing amounts of unlabeled ORC protein. The resulting ORC-DNA complexes were run on a non-denaturing PAGE gel. In cases where ORC bound to the DNA, a shift in the mobility relative to the radiolabeled DNA alone is observed. Where indicated, a 5FdC-modified M.HpaII binding site and M.HpaII has been added. Results are summarized in (**H**)

and without M.HpaII crosslinking for each DNA. As expected, mutation of the ACS (A-B2-) resulted in complete inhibition of ORC binding (Figure 7A, left).

All of our modified templates were able to bind ORC when M.HpaII was absent, confirming that inserting the M.HpaII recognition motifs at their respective locations did not affect ORC binding in the absence of crosslinked M.HpaII. In contrast, M.HpaII placement at the -10, -5, +14, +23, +31 and +38 positions affected ORC DNA binding (Figure 7A-G, summarized in Figure 7H). ORC binding was almost entirely abolished with the +14, +31, and -5 adducts. Although some ORC binding was observed with the +23 and +38 adducts, close examination suggests the binding observed is to residual DNA template without M.HpaII modification (Figure 7D). A slight defect in ORC binding was also observed at the -10 location (Figure 7G). Although the ORC binding defects appear more severe in the EMSA assay (likely due to the addition of competitor DNA), the EMSA experiments identify largely the same subset of adducts as the ATPγS and ATP experiments, suggesting that the major consequence of the M.HpaII roadblocks that inhibit helicase loading is the inhibition of ORC binding to the origin.

DNA in the B element region is important for helicase loading when protein blockages are placed surrounding the origin

The results of our M.HpaII crosslinking studies are not consistent with a model for helicase loading in which ORC remains bound to the ACS as it sequentially recruits two Mcm2-7 complexes to encircle the adjacent ~60 bp of DNA. If this model were correct, regardless of which side the Mcm2-7 proteins binds, several of the tested M.HpaII adducts should inhibit Mcm2-7 loading without affecting ORC-DNA binding. Instead, we identified an ~50 bp region where M.HpaII adducts inhibit ORC-DNA binding, and a 10 bp region between -10 and -20 that does not affect ORC-DNA binding, but shows small defects in Mcm2-7 helicase loading. To align our data with the OCCM structure, we hypothesized that at least one intermediate complex formed during helicase loading could slide along dsDNA after initial sequence-specific ORC-

Figure 8

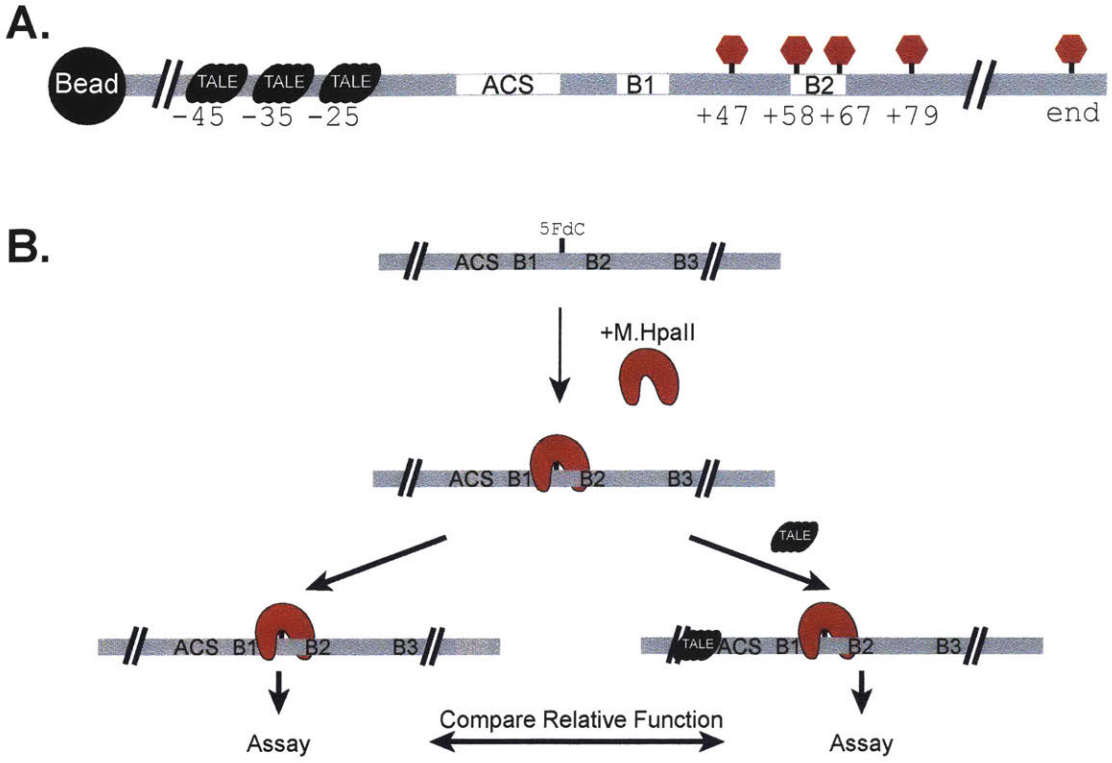


Figure 8: Using two protein blockages to assess the impact of sliding during the helicase loading process

A) Templates were constructed to contain both a TALE-binding sequence on one side of the ACS and a cross-linked M.HpaII on the other side. As before, possible locations for placement of the M.HpaII consensus motif are shown as red stop signs. Possible locations for TALE protein binding are marked by a black TALE cylinder. Each template in the experiments below contained one TALE-binding motif and one M.HpaII consensus sequence.

B) Schematic for reactions performed in **(C)**. Different M.HpaII/TALE binding templates were assessed for their ability to stably load Mcm2-7. All templates were treated with M.HpaII and then either treated (+) or not treated (-) with TALE protein as indicated above each lane. The location of the 5FdC-modified M.HpaII recognition motif (*M.HpaII-seq*) and the location of the TALE-binding sequence (*TALE-seq*) are noted.

Figure 9

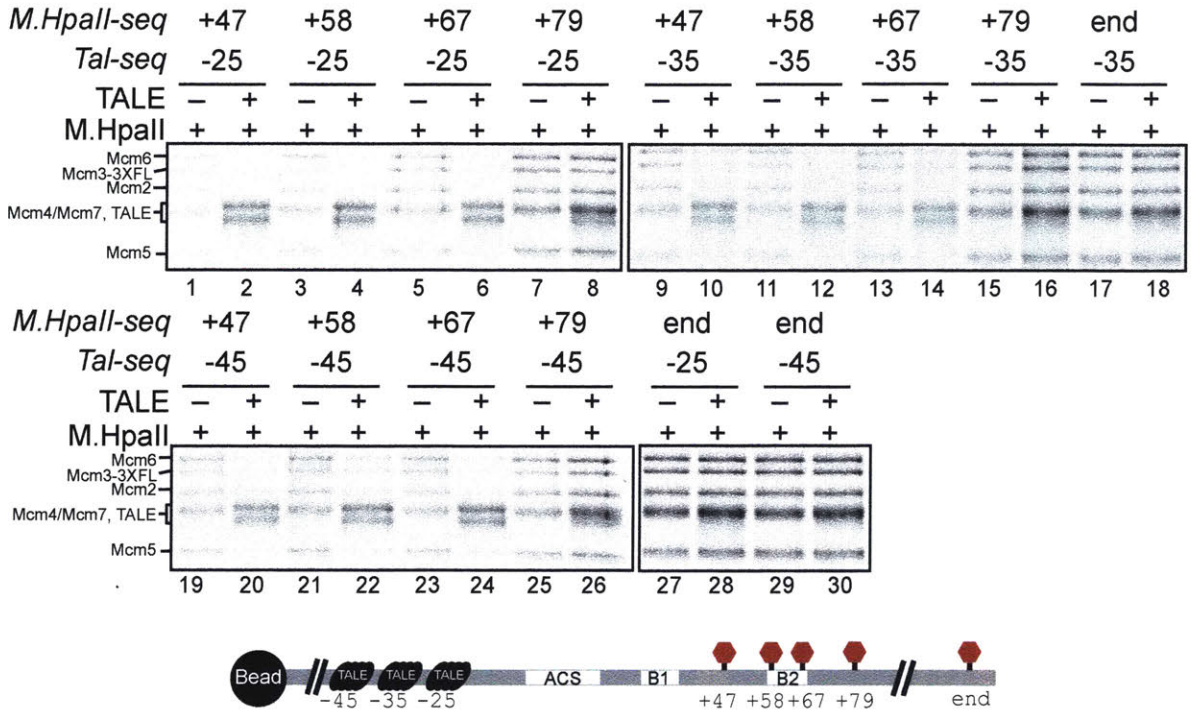


Figure 9: Placing blockages on both sides of the ACS has different effects on helicase loading than the individual blocks alone.

Protein-modified templates were incubated with helicase loading proteins for 30 minutes and the formation of stable Mcm2-7 loading assessed by a high-salt wash. The stably bound proteins were released from the DNA beads with DNAse and run on an SDS-PAGE gel, followed by fluorescent protein staining. Paired lanes can be compared.

DNA binding. If this is the case, a single protein adduct would only have an effect if it inhibited initial ORC DNA binding, because the intermediate could slide to free up the space necessary to load an Mcm2-7 helicase. However, placement of two barriers flanking the origin could restrict this movement, potentially changing the impact of the M.HpaII adducts.

To create these double-barrier templates, we added the TALE binding sequence (used at the free end of our DNA templates in previous assays) at positions -25, -35 and -45 relative to the ACS for several of the M.HpaII-modified templates (Figure 8A, Figure9, and Figure 10). Importantly, the crosslinked M.HpaII complexes can stop an Mcm2-7 complex from sliding off the free end of the DNA template (Figure 11), as was observed with the TALE complex (Figure 1D). Having two orthogonal protein-DNA adducts allowed us to test whether having a TALE blockage upstream of the ACS changed the effect of a M.HpaII adduct placed downstream of the ACS (Figure 8B and Figure 9). Importantly, all the positions of both the TALE binding sites and the M.HpaII adducts tested together were located outside of DNA regions identified as important for ORC binding and helicase loading with individual blocks (Figures 2-7).

Although none of the M.HpaII or TALE placement positions alter helicase loading or ORC binding individually in this subset of adduct locations (Figure 3 and Figure 9), when both the M.HpaII and TALE adducts were present simultaneously, we observed strong defects. With M.HpaII crosslinked at +47, +58 or +67, TALE binding to any of the three locations tested (-25, -35 or -45) inhibited helicase loading (Figure 9). These defects were not due to inhibition of initial ORC-DNA binding or Mcm2-7/Cdt1 association, because OCCM formation was observed in the presence of ATPγS (Figure 12). Increasing the spacing between the two roadblocks attenuated these defects. Moving the M.HpaII adduct to the +79 location or the end of the DNA template (~+280) showed no inhibition of helicase loading when the TALE protein bound at any location (Figure 9, compare lanes 15-16 and 17-18 for example).

Figure 10

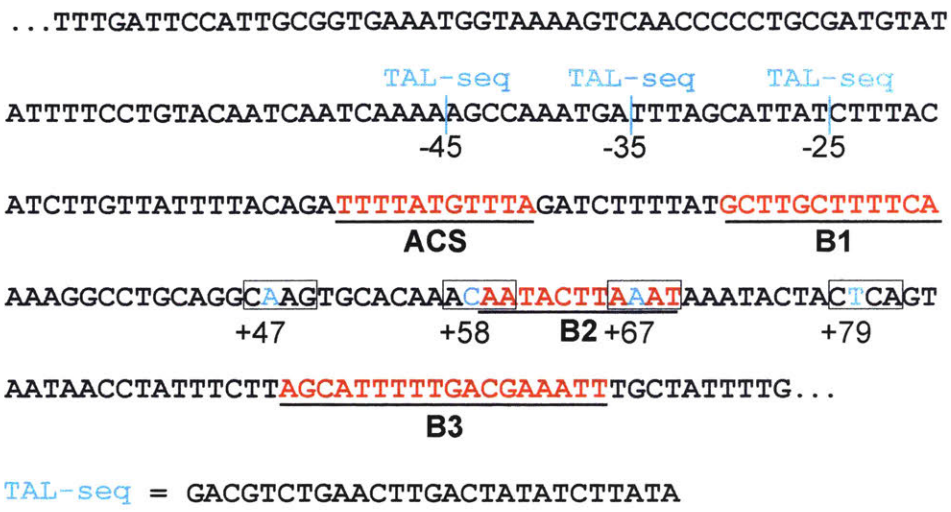


Figure 10: Location of the TALE consensus sequence insertions used for Figure 9

The sequence 5' to 3' of the T-rich strand of ARS1 is shown. Key sequence features of the origin are highlighted in red and labeled underneath. Locations that were changed to the M.HpaII sequence motif are boxed in grey. Locations where the TALE-binding sequence was inserted are shown by blue lines. The TALE-binding sequence is shown.

Figure 11

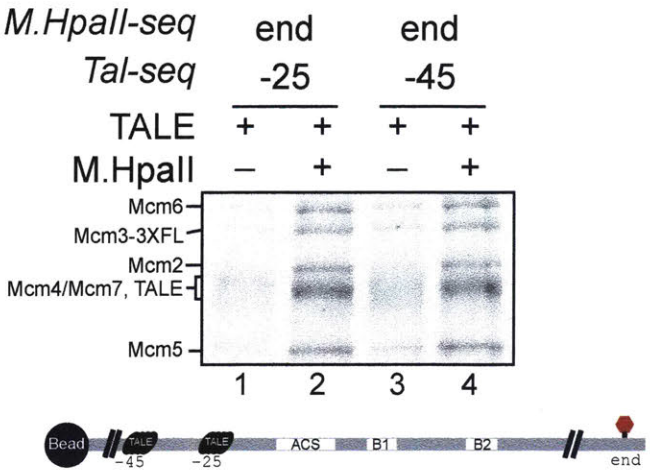


Figure 11: M.HpaII can block Mcm2-7 sliding off DNA ends

Helicase loading assays were performed in the presence of TALE protein binding with and without M.HpaII. The M.HpaII recognition sequence was placed at the end of the template. Similar to what is observed in Figure 1D, Mcm2-7 signal is retained when M.HpaII is placed at the end of the template.

Figure 12

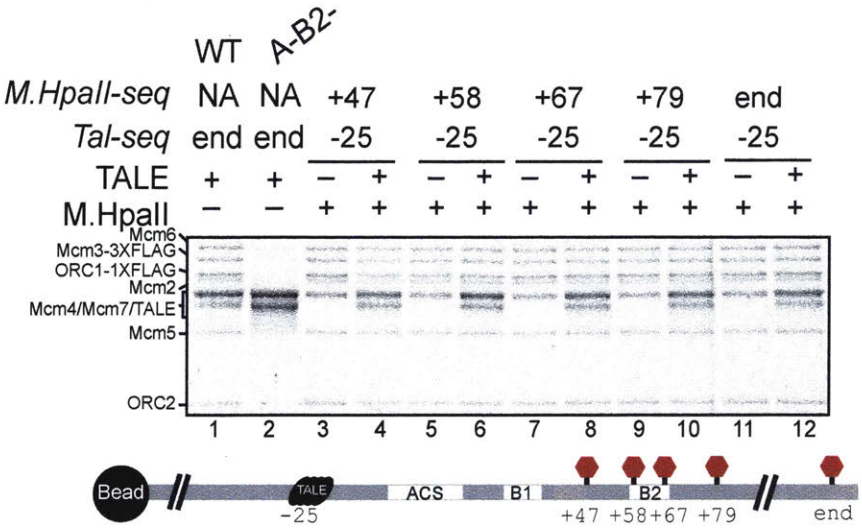


Figure 12: Placing two blockages around the origin does not affect OCCM formation in ATPγS.

Preliminary helicase loading results using a TALE placed at the -25 location and M.HpaII placed any of the possible locations tested show no defect in OCCM formation. These results are consistent with all roadblocks being outside the ORC binding region previously identified

The B2 Element of *ARS1* is important for stable helicase loading when the origin is bounded by proteins.

Although it was possible that a minimum distance between the two roadblocks was required for helicase loading, examination of the distance between the two blocks showed that this was not the sole determinant. The -25/+79, -35/+67 and -45/+58 templates have blockages spaced ~100 bp apart and include the ACS and B1 elements between the two adducts, yet only one of these templates (-25/+79) successfully loads Mcm2-7. Interestingly, of these three templates, only the -25/+79 combination included the B2 element of *ARS1* between the two adducts. Although the B2 element plays a role in helicase loading *in vivo* (Lipford and Bell, 2001; Zou and Stillman, 2000), a role for B2 has not been reported *in vitro*.

To examine the B2 element in more detail, we first tested whether B2 contributed to helicase loading under our assay conditions on DNA templates without M.HpaII adducts (Figure 13). These templates contained a TALE-consensus sequence at the end of the template, and TALE protein was added to prevent sliding (Figure 13A). Although a mutation in the ACS showed a severe defect in Mcm2-7 loading (Figure 13B, compare lanes 1 and 3, quantified in Figure 13C), mutation of the B2 element showed only a minor ~10% decrease in helicase loading (Figure 13B, compare lanes 1 and 4, quantified in C). It has previously been reported that the B2 element can act as an ORC binding site under certain conditions (Bell and Stillman, 1992; Lee and Bell, 1997; Wilmes and Bell, 2002). However, simultaneous mutation of the ACS and B2 elements was not significantly different from mutation of the ACS alone in our assays (Figure 13B, compare lanes 2 and 3, quantified in Figure 13C). Thus, under the conditions used for helicase loading in this study, the B2 element does significantly contribute to Mcm2-7 loading in unmodified templates.

Despite the minor effect of B2 mutations on unmodified DNA templates, the apparent connection between exclusion of the B2 element and loss of helicase loading in our double-blocked templates led us to ask if B2 was important when sliding of helicase loading

Figure 13

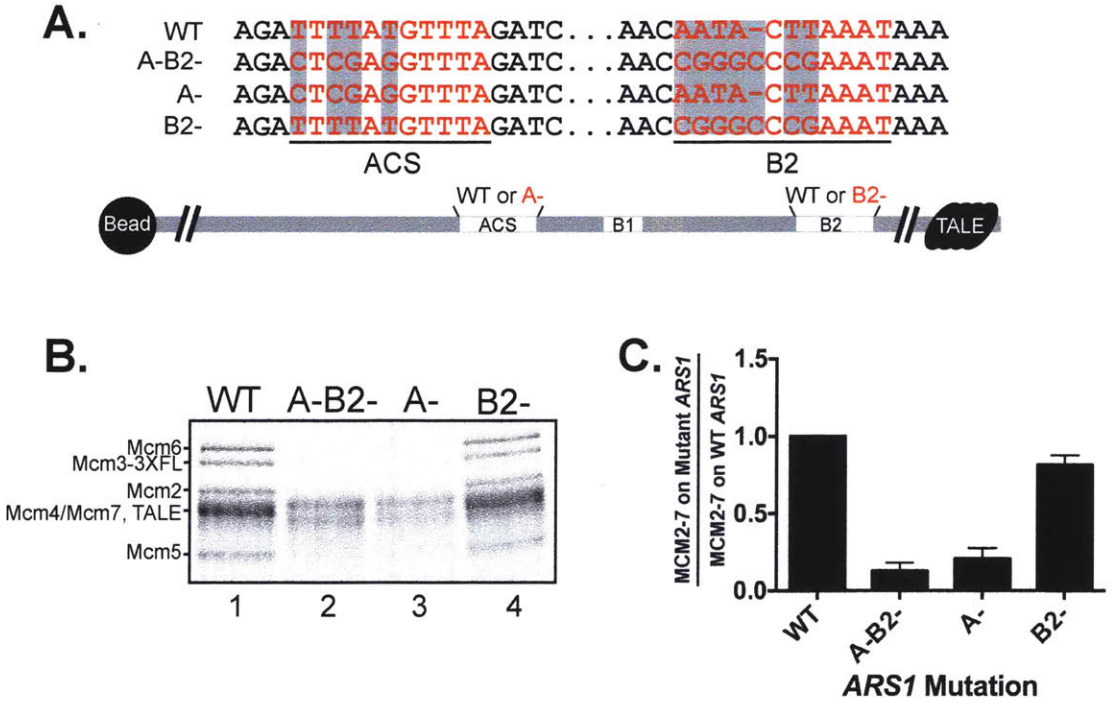


Figure 13: The B2 element has only a very minor defect in helicase loading without blockages.

A) Sequence alignment of the T rich strand of origin DNA. WT and mutant origin templates used in this study are shown. The ACS and B2 elements are highlighted in red and mutated residues are highlighted in grey.

B) Representative assay showing the amount of helicase loading following a high salt wash using various origin mutant templates.

C) Quantification of the relative amount of helicase loading. The wildtype level of loading has been set to 1 and the amount of loading in the other templates is the ratio over wildtype.

intermediates was restricted. To test this hypothesis, we mutated the B2 element in the presence of the +79 M.HpaII modification and assessed whether TALE binding at any of the three locations tested previously influenced helicase loading (Figure 14A). Strikingly, although little to no defect in stable Mcm2-7 loading was observed when the B2 element was present, mutation of the B2 element completely inhibited helicase loading in the -25/+79 and -35/+79 double-blocked templates (Figure 14B, lanes 1-8). Consistent with our sliding hypothesis, the -45/+79 template showed an approximately 50% reduction in stable helicase loading in the presence of both blockages (Figure 14B, lanes 9-12).

The observed dependence on B2 requires both adducts. In the absence of TALE addition, the M.HpaII-bound +79 showed strong helicase loading (Figure 14B, lane 1 and lane 3 for example). However, once TALE is added, there is a striking defect when B2 is mutated, which is absent when B2 is wildtype (Figure 14B, compare lanes 2 and 4 for example). The B2-dependent defects also required placement of the two adducts close to the origin. A TALE-bound template where the M.HpaII was crosslinked at the end of the template (~250 bps from the origin sequence) also showed no B2-dependent defect (Figure 14B, lanes 13-16 and data not shown). Although the B2 element can act as a secondary ORC binding site (Bell and Stillman, 1992; Lee and Bell, 1997; Wilmes and Bell, 2002), initial association of ORC and Mcm2-7 in ATPγS was unaffected by the mutation of the B2 element even with protein roadblocks (Figure 15), suggesting that the B2-dependent defect we observed occurs downstream of initial Mcm2-7 recruitment.

The B2 element is important for helicase loading in the context of nucleosomes.

Although our artificial roadblocks are not present *in vivo,* origins of replication are consistently found within nucleosome-free regions (NFRs) flanked by positioned nucleosomes (Berbenetz et al., 2010; Eaton et al., 2010). We hypothesized that these flanking nucleosomes could restrict DNA access in a similar manner to the artificial roadblocks we tested. To test this

Figure 14

A.

B.

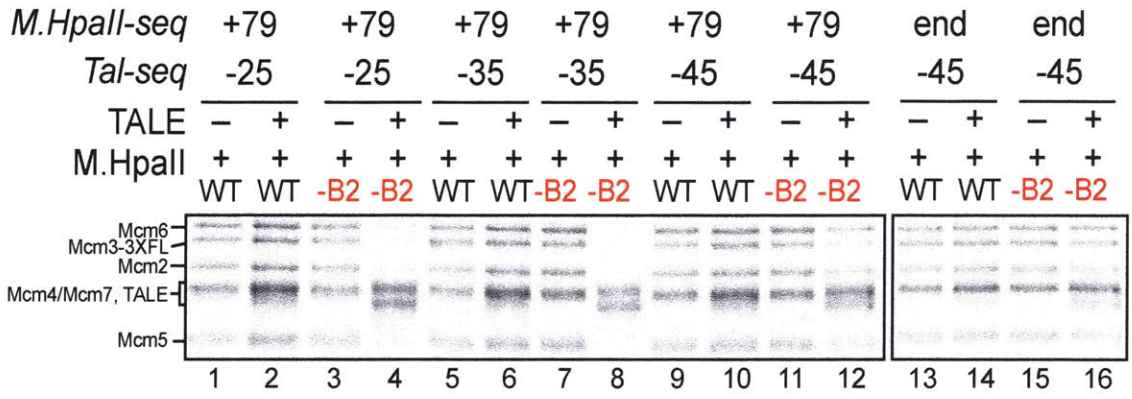


Figure 14: The B2 element is important for Mcm2-7 helicase loading when space around the origin is restricted

A) Templates constructed to contain both a TALE-binding sequence upstream of the ACS and a cross-linked M.HpaII at the +79 location were modified to mutate the conserved B2 element. Possible locations for placement of the TALE protein are marked by black TALE cylinders. Each template in the experiments below contained one TALE-binding motif and one M.HpaII consensus sequence at the +79 location or at the end of the template (~ +250 bp).

B) Different TALE-binding templates were assessed for their ability to stably load Mcm2-7 when a modified M.HpaII consensus motif was placed at the +79 location or at the end of the templateto block sliding). The B2 element of *ARS1* was mutated or wild-type as indicated. Templates were treated with M.HpaII or TALE protein as indicated above each lane. The location of the 5FdC-modified M.HpaII recognition motif (*M.HpaII-seq*) and the location of the TALE-binding sequence (*TALE-seq*) are noted above each pair of lanes. Protein-modified templates were incubated with helicase loading proteins for 30 minutes and the formation of stable Mcm2-7 loading assessed by a high-salt wash. The stably bound proteins were released from the DNA beads with DNAse and run on an SDS-PAGE gel, followed by fluorescent protein staining.

Figure 15

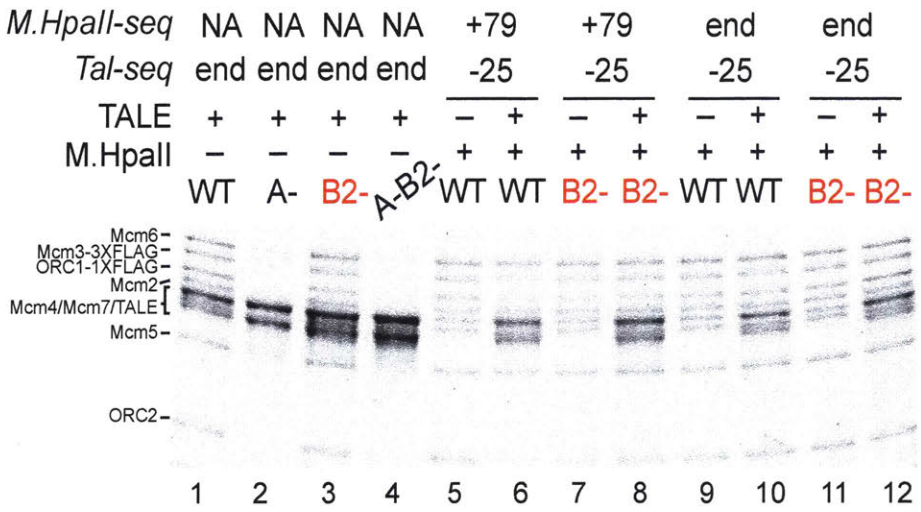


Figure 15: The B2 element is not important for OCCM formation even when space around the origin is restricted

OCCM formation assays were performed on a subset of templates used in (B). Although clear defects in OCCM formation were observed when the ACS was mutated, no detectable changes in OCCM formation were observed in either WT or B2- double-modified templates.

hypothesis, we assembled chromatinized plasmids using a previously established method (Eaton et al., 2010) and looked at helicase loading on plasmids with the ACS or B2 element mutated or wild-type. These experiments utilized a yeast extract-based helicase loading assay (Eaton et al., 2010; Seki and Diffley, 2000). In this assay, yeast are arrested in G1, lysed, and the extract (containing all proteins necessary for helicase loading) is incubated with an origin-containing DNA template. Because helicase loading was assessed on plasmids unattached to beads, high-salt extraction was not feasible. However, the amount of Mcm2-7 helicase observed after a low-salt extraction with 300mM potassium glutamate (KGlut extraction) and after a high-salt extraction with 500mM sodium chloride (NaCl extraction) was the same under the conditions used, suggesting that in our reactions, the majority of Mcm2-7 forms a salt-stable complex (Figure 16A). DNA containing mutations in the ACS, the B2 element or both were either chromatinized or left untreated, then incubated in G1 extract to load proteins. Extract-treated plasmids were isolated by gel filtration and associated proteins released with MNase (Figure 16B).

Although the B2 element had only a minor effect on helicase loading on the naked ARS1 plasmids, a much stronger (~75%) helicase loading defect is observed in the context of nucleosomes (Figure 16C, quantified in Figure 16D). As anticipated, mutating the ACS caused a severe helicase loading defect on both chromatinized and naked templates (Figure 16C, quantified in 16D). As observed in our double-blocked templates (Figure 15), no ORC binding defect was observed for the WT and B2- templates in either naked or nucleosomal templates. This observation further emphasizes that the B2 element is not important for stable ORC binding (Figure 16C).

Nucleosome movement has been implicated previously to impact helicase loading (Azmi et al., 2017; Lipford and Bell, 2001; Simpson, 1990). To ensure that nucleosome movement was not the cause of the defects we observe with the B2- templates, nucleosome positioning was determined for plasmids constructed via the method described in Figure 16B. Importantly, the

Figure 16

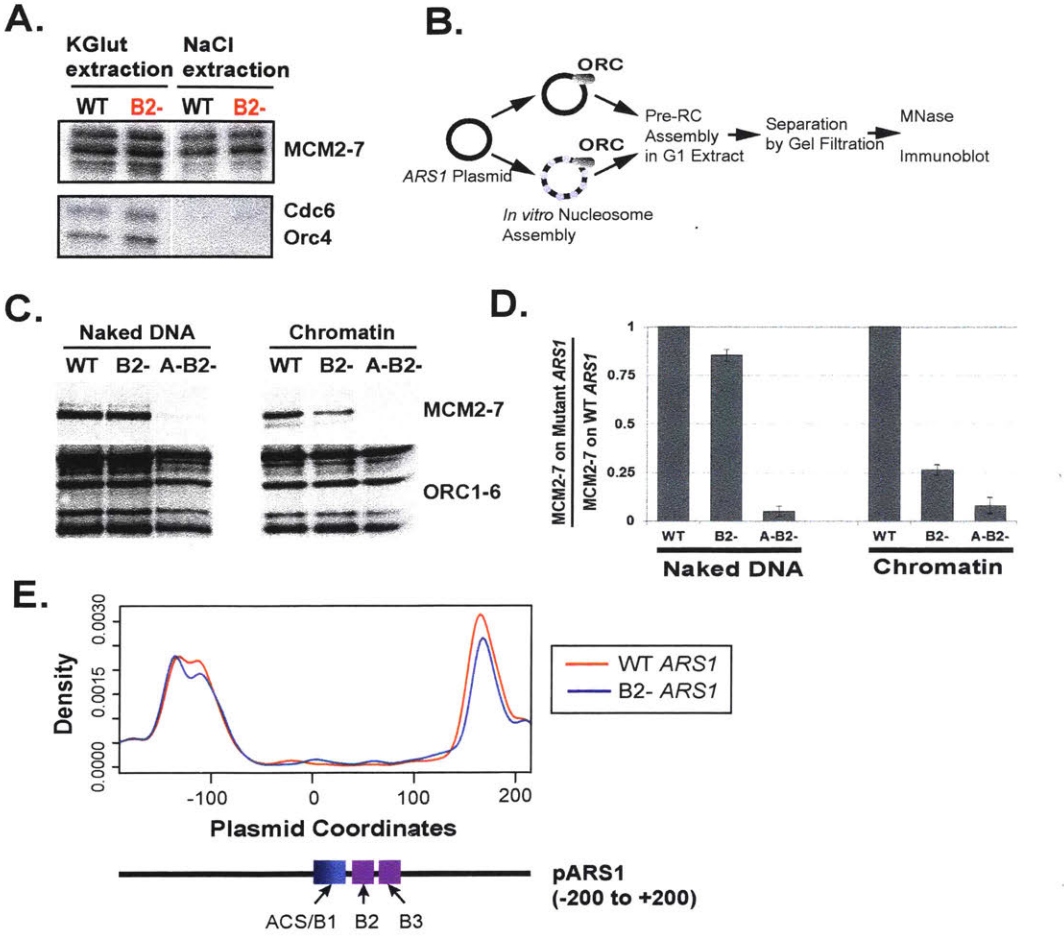


Figure 16: The B2 element is important when nucleosomes are placed around the origin.

A) Helicase loading assays were performed using G1 cellular extract. A 1 kb DNA template was incubated in a G1 extract for 15 minutes. Templates were either subjected to a low stringency wash (KGlut extraction) or a high stringency wash (NaCl extraction).

B) Reaction scheme for helicase loading assay in performed in (C). Nucleosomal templates were assembled as previously described (Eaton et al., 2010). Helicase loading in a G1 extract was performed for 15 minutes, then the plasmids isolated by gel filtration. Loaded Mcm2-7 helicase was released from the plasmid with MNase, and helicase amount determined by western blot.

C) Immunoblot showing the amount of ORC and Mcm2-7 loading onto naked and chromatinized plasmids.

D) Quantification of Mcm2-7 loading in (C). Error bars are the standard deviation of three independent experiments

E) Nucleosome positioning data from *in vitro* assembled WT-ARS1 (red) and B2- ARS1 (blue) templates.

NFR region around the ACS was unchanged between the WT and B2- plasmids, suggesting that changes in nucleosome positioning were not the cause of the helicase loading defects observed (Figure 16E). Taken together, our findings suggest that the importance of the B2 sequence is dependent on restriction of the origin DNA accessibility, regardless of the mechanism of restriction.

Discussion:

Based on structural studies, a detailed model for the process of eukaryotic helicase loading has been created (Figure 17A). First, ORC binds to the 24-30 bp ORC-ACS. ORC then recruits Cdc6 and the first Mcm2-7/Cdt1 heptamer to origin DNA to form the OCCM. A second Mcm2-7 hexamer is then loaded, directly adjacent to the previous hexamer, in a process that utilizes an additional Cdc6, Mcm2-7/Cdt1, and possibly another ORC molecule. In this model, the first ORC molecule remains bound to the ORC-ACS throughout the entire process. The two Mcm2-7 hexamers bind directly adjacent to ORC, encircling a contiguous DNA strand, and are expected to encircle approximately 60 bps of DNA. Although structural data does not tell us whether the Mcm2-7 double hexamer loads upstream or downstream of the ACS, because of the ACS is close to the upstream nucleosome within the NFR (Chapter I, Figure 5), Mcm2-7 association is hypothesized to occur within the B elements.

There are two aspects of this model that are contradicted by our data. First, we identify an ~50 bp region that effects ORC's ability to bind to origin DNA, rather than the 24 bps predicted by structural data (Yuan et al., 2017). The region we identify is consistent with both ORC DNAse footprinting and ORC-DNA cross-linking experiments (Bell and Stillman, 1992; Lee and Bell, 1997; Rowley et al., 1995). This distinction could be explained by ORC adopting a different configuration on origin DNA in the absence of Cdc6 or Mcm2-7/Cdt1 binding. Consistent with this hypothesis, crosslinking studies of ORC place Orc2, Orc4, Orc5, and Orc6 near or past the B1 element of *ARS1* (Lee and Bell, 1997). Alternatively, regions of ORC not

Figure 17

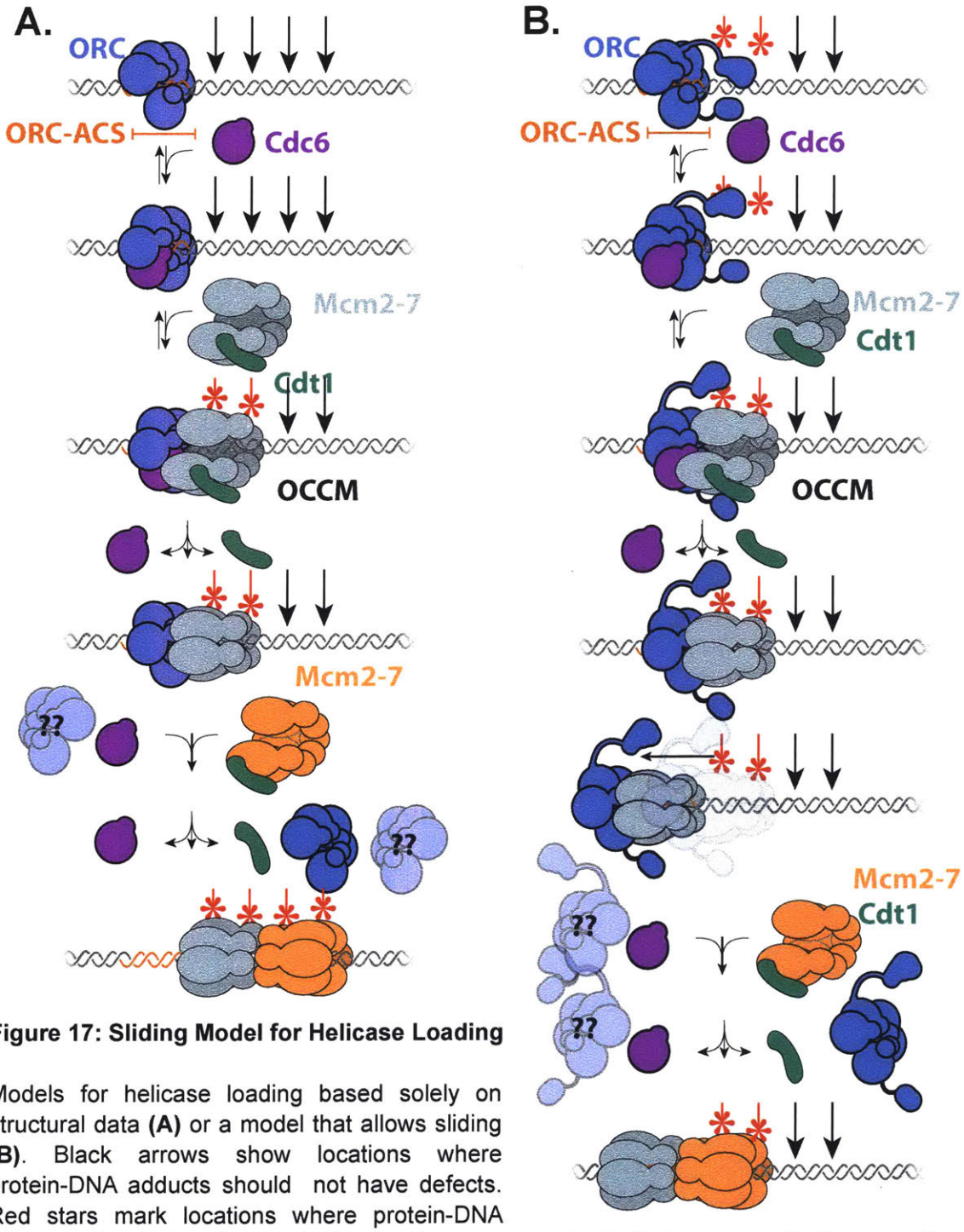


Figure 17: Sliding Model for Helicase Loading

Models for helicase loading based solely on structural data **(A)** or a model that allows sliding **(B)**. Black arrows show locations where protein-DNA adducts should not have defects. Red stars mark locations where protein-DNA adducts should interfere with the loading process. In both instances, ORC recognizes the ORC-ACS. In **(A)** it never moves from this location. In **(B)**, a helicase loading intermediate slides to allow loading of a double-hexamer. In **(A)**, placing protein-DNA adducts within the 60 bp where Mcm2-7 binds should cause helicase loading defects. In **(B),** only regions that effect ORC binding would be expected to show defects in downstream steps.

observed in the structural studies may bind to or occupy space in the vicinity of the extra ~25 bp of DNA. Placement of large protein adducts on DNA adjacent to ORC may sterically clash with regions of ORC that are not observed in ORC-DNA structures. Likely candidates include portions of Orc6 and Orc2 that are unstructured or missing from recent structural studies (Bleichert et al., 2015; Tocilj et al., 2017; Yuan et al., 2017). Regardless of the mechanism, although the locations of our DNA-protein adducts do not eliminate specific sequences required for ORC binding, they clearly destabilize sequence-specific ORC binding.

Second, we did not observe individual adducts that specifically hinder Mcm2-7 binding but not ORC binding, as would be predicted in the structural model (Figure 17A). If ORC remains bound to the ORC-ACS throughout loading, we would expect to observe a 60 bp region on one side of the ORC binding site that must be adduct-free to allow the sequential loading of two Mcm2-7 complexes. Adducts within this 60 bp window would be expected to specifically inhibit Mcm2-7 association but not ORC binding. Curiously, we only find a single adduct location (-15) that reduces Mcm2-7 loading but has no effect on ORC binding. M.HpaII crosslinking at this location showed only a mild defect in stable Mcm2-7 double hexamer formation (around 20%), and this mutation is not in the B element region. Since increasing the size of the protein adduct didn't change the pattern of defects (Figure 4), it is unlikely that Mcm2-7 complexes can enclose the protein-DNA adduct within its central channel.

To explain our data, we propose that a helicase-loading intermediate can slide after initial origin-DNA binding (Figure 17B). In this sliding model, ORC still initially binds to the ORC-ACS, although parts of it extend to interact with an additional 25 bps in the B region. We propose that ORC undergoes a structural change to accommodate loading of the first Mcm2-7 complex on adjacent dsDNA, forming the OCCM complex that covers an estimated 55 bps (Yuan et al., 2017). To allow loading of the second hexamer and not observe defects when adducts are placed adjacent to this 50 bp region (such as those at the +47, +58 and +67 locations), we propose that an intermediate after OCCM formation slides away from the ORC-

ACS. This sliding could provide the minimum 30 bps of accessible DNA needed for loading the second Mcm2-7 complex. Importantly, this model maintains the protein-protein contacts observed in structural and biochemical studies, but allows the flexibility to accommodate protein-DNA adducts within 60 bp of the ORC binding site. My studies demonstrate that the OCCM complex formed in ATPγS cannot slide (Figure 5), suggesting the mobile intermediate would occur at some point after this step. However, a complex containing these same components may be able to slide following ATP hydrolysis.

Placement of multiple protein adducts, one upstream and one downstream of the ACS, support our sliding hypothesis. Adding an additional adduct upstream of the ACS in addition to adducts in the B element region can test helicase loading when sliding is restricted. We would expect that when both adducts are present simultaneously, additional locations in the B element region would affect Mcm2-7 binding independent of ORC. Indeed, we observe a region around B2 where adducts have a defect in stable Mcm2-7 double hexamer formation but not ORC binding or OCCM formation, roughly 30 bps in size (Figure 9, Figure 12).

Although the different results we get when two blockages are placed around the ACS supports our sliding hypothesis, it also demonstrated that adequate space alone was insufficient for stable helicase loading. Under conditions where a minimum amount of space for double hexamer formation is allowed, we observed a very strong requirement for the B2 sequence during helicase loading. The dependence on B2 only occurred in the presence of two blockages, one upstream and one downstream of the ACS. Identical mutations in the B2 element on unmodified templates or in the presence of a single adduct had little to no effect on helicase loading *in vitro.* Restricting origin DNA accessibility with nucleosomes rather than artificial protein blockages yielded similar B2-specific helicase-loading defects.

Two hypotheses have been put forward for the role of the B2 element in helicase loading. It could be a second ORC binding site or an Mcm2-7 binding site. Further investigation is necessary to determine which of these two models is correct (discussed in more detail in

Chapter 3). Why is the B2 element only required when sliding is prevented? Again, this can be explained by sliding of a helicase-loading intermediate. Such a sliding intermediate could explore the remainder of the 1 kb DNA template until it finds an accessible B2-like element elsewhere to allow completion of loading.

Methods:

Purification of M.HpaII and MBP-M.HpaII:

A plasmid carrying the M.HpaII expression construct (pM15MH or pM15MH-MBP) was transformed into BL21-DE3. After 1 hour rescue in SOC, the entire aliquot of cells was transferred to 100mL of LB containing 35µg/mL chloramphenicol (CAM) and 100µg/mL ampicillin (AMP) and grown overnight. The resulting cells were transferred to 1L of LB with 35µg/mL CAM and 100µg/mL AMP to an O.D. of 0.7-0.8 and expression of M.HpaII was induced by the addition of (CONC) IPTG for 2.5 hours at 37°C. Cells were harvested at 5000RPM for 5 min at 4°C by centrifugation, the resulting packed cells resuspended in 25mL of M.HpaII Loading Buffer (20mM TRIS-Cl pH 8.5, 500mM KCl, 10% glycerol, 10mM Imidizole) with protease inhibitors (Roche) and snap-frozen in liquid nitrogen and stored at -80°C until ready to proceed.

Cells were thawed at room temperature for 20 minutes. Loading buffer containing protease inhibitors (Roche) was added to bring the volume up to 50mL. Lyzozyme was added to a final concentration of 100µg/mL and the mixture was incubated at 4°C for 30minutes while rotating. The mixture was then sonicated on ice at 35% for 2min (5sec on, 5sec off) and the resultant mixture spun at 40,000rpm for 35 minutes. The resulting supernatant was loaded onto a 2mL Ni-NTA resin column pre-equilibrated with M.HpaII loading buffer. The mixture was incubated for 2 hours, washed with 2 column volumes of M.HpaII wash buffer (20mM TRIS-Cl pH 8.5, 500mM KCl, 10% Glycerol, 30mM Imidizole) then eluted with 4x1mL fractions in M.HpaII elution buffer (20mM TRIS-Cl pH 8.5, 100mM KCl, 10% Glycerol, 1mM DTT) with

500mM Imidizole. Peak fractions were pooled, concentrated to 1mL and the resulting protein put over a Superdex 75 size exclusion column equilibrated with M.HpaII elution buffer. Peak fractions were assayed with Bradford, and the protein concentrated and exchanged into M.HpaII elution buffer containing 30% Glycerol. Protein was flash frozen in liquid nitrogen and stored at -80°C until used.

Purification of Helicase loading proteins:

ORC, Mcm2-7/Cdt1 and Cdc6 were purified as previously described (Kang et al, 2014).

Purification of TALE:

A plasmid carrying the TALE expression construct (p204) transformed into Rosetta II cells. After 1 hour rescue in SOC, the entire aliquot of cells was transferred to 50mL of LB containing 35µg/mL chloramphenicol (CAM) and 100µg/mL ampicillin (AMP) and grown overnight. The resulting cells were transferred to 1L of LB with 35µg/mL CAM and 100µg/mL AMP to an O.D. of ~0.6 and expression of TALE was induced by the addition of 500µL of 1M IPTG (final) for 3 hours at 37°C. Cells were harvested and resuspended in 25mL of TALE lysis buffer (50mM Tris-HCl, 500mM NaCl, 10mM Imidazole, 10% Glycerol, 0.5mM TCEP, EDTA free protease inhibitors (Roche)), snap frozen in liquid nitrogen and stored at -80°C until ready to proceed.

Frozen cells were thawed at room temperature for ~1 hour. TALE Lysis buffer was added to bring the volume up to 50mL. Lyzozyme was added to a final concentration of 100µg/mL and the mixture was incubated at 4°C for 30minutes while rotating. The resultant mixture was then sonicated on ice at 35% for 2min (5sec on, 5sec off) and the resultant mixture spun at 40,000rpm for 35 minutes. The resulting supernatant was loaded onto a 1mL packed volume Ni-NTA resin column prequilibrated with Buffer A (50mM Tris-HCl pH 7.5, 500mM NaCl, 10% Glycerol, 10mM Imidazole, 0.5mM TCEP). The mixture was incubated for 1 hour, washed 2 x 3mL with buffer A, then eluted with 4 x 1mL fractions with Buffer B (50mM Tris-HCl pH7.5,

500mM NaCl, 10% Glycerol, 500mM Imidazole, 0.5mM TCEP). Peak fractions were pooled, concentrated to 1mL and the resulting protein put over a Superdex 75 size exclusion column equilibrated with Buffer A lacking imidazole. Peak fractions were assayed with Bradford, and the protein concentrated and the protein diluted 1:1 with storage buffer (80% glycerol, 6mM NaCl, 50mM Tris-HCl pH7.5, 0.5mM TCEP). Protein was flash frozen in liquid nitrogen and stored at -80°C.

Construction of M.HpaII-modified DNA:

Oligonucleotides were designed to introduce the M.HpaII consensus motif (CCGG) at different locations across the *ARS1* origin sequence. Oligonuclotides containing a 5FdC-modified M.HpaII consensus motif (CCGG) and 15-20bp of complementary DNA sequence were synthesized (Biosynthesis) and subsequently used to PCR dsDNA containing part of the *ARS1* sequence at one end and a biotin on the other resulting in a ~700bp (Large) fragment. A reverse-complementary oligonucleotide to the 5FdC oligo containing the CCGG motif but lacking the 5FdC modification was used to PCR the remaining regions of the origin, creating a partial origin-containing 250bp (small) dsDNA fragment. The small fragment also contained the TALE-binding sequence. The two fragments contained an overlapping region of around 30bps.

Large and small origin-fragment DNA was subjected to T4 DNA polymerase in buffer lacking dNTPs for 30min (Large) or 20min (Small) to resect the 3' ends of the PCR fragments. The resulting 5' overhangs were phosphorylated with T4 PNK, the large and small fragments mixed in a 1:5 molar ratio and incubated for 10minutes at 80°C to ensure single-stranded regions were accessible, then slow cooled to room temperature to allow annealing of the complementary 3' regions. The annealed templated were then incubated with T4 DNA polymerase and T4 DNA ligase in buffer containing dNTPs and ATP for 2 hours at 16°C. Full length (950bp) *ARS1* DNA formation was assayed on a 1.5% agarose gel containing ethidium bromide. T4 DNA polymerase digest, annealing and fill-in were repeated until templates showed

greater than 90% full length template. Completed templates were coupled to streptavidin-coated magnetic beads (Dynabead M280, *Invitrogen*) and the amount of DNA coupled to the beads assayed by EcoRI digest.

M.HpaII coupling to 5FdC-modified origin DNA

0.75pmol of 1.3 kb origin-containing DNA coupled to magnetic beads (Dynabead M280 beads, *Invitrogen*) was incubated in M.HpaII binding buffer (50mM TRIS-HCl pH 7.5, 0.5mM β-Mercaptoethanol, 10mM EDTA) with or without a 10 fold molar excess of M.HpaII or MBP-M.HpaII for 1.5 hours at 37°C in the presence of 80μM S-adenosylmethionine (SAM, *New England Biolabs*). Excess M.HpaII was removed by three successive washes with H buffer (50mM HEPES-KOH pH 7.6, 5mM MgOAc, 1mM ZnOAc, 1mM DTT, 10% glycerol, 0.1mM EDTA, 0.1mM EGTA, 0.02% NP40) with 300mM potassium glutamate (H/300).

TALE binding to origin DNA:

0.75pmol of origin-containing DNA coupled to magnetic beads was incubated with a 10-fold molar excess of purified TALE protein at 25°C for 15 minutes. Excess protein was removed by two successive washes with H/300 buffer.

Helicase-Loading Assay:

Helicase-loading assays were performed similar to that described in Kang et al, 2014 with some slight modifications. 0.75pmols of origin-containing DNA coupled to magnetic beads (M280 beads, *Invitrogen*), either pre-treated or not with M.HpaII and/or SNAP-Tal, were incubated with 0.17pmol ORC, 3pmols Cdc6 and 3pmols of Mcm2-7/Cdt1 protein in Reaction Buffer (25mM HEPES-KOH, 12mM MgOAc, 50μM ZnOAc, 350mM Potassium glutamate, 3mM ATP, 4mg/mL BSA) for 30minutes at 25°C in a thermo-mixer at 1200RPM. Following incubation, reactions were magnetized and then wash once with H/300, once with H buffer containing 500mM NaCl,

and a final time with H/300. 1U of DNAse (*Worthington*) in DNAse Buffer (H buffer with 5mM $CaCl_2$, 5mM MgOAc, 50mM KCl) was added to each reaction and incubated at 25°C and 1200RPM for 10 minutes. The beads were magnetized and the resulting supernatant was added to 1X SDS sample buffer. Samples were run on an 8% SDS-PAGE gel, and stained with fluorescent protein stain (Krypton, *Thermo-Fisher Scientific*) according to the manufacturer's instructions.

Helicase Association in ATPγS:

Association assays were performed identically to the helicase loading assays expect for the following exceptions: 1pmol of ORC was used, 5mM ATPγS was used in place of 3mM ATP and three successive washes with H/300 were performed following the 30 minute incubation.

Insertion of TAL-sequence at a location near the ACS:

An oligonucleotide containing the TAL-binding sequence (5' TGAACTTGACTATATCTTATA 3') along with an AatII cut site was inserted into the pUC19-ARS1 plasmid via standard site-directed mutagenesis (Quikchange lightning kit, *Agilent*). The resulting colonies were tested for the TALE-sequence insertion by AatII digestion, and the correct insertion sequence was verified by Sanger sequencing. Plasmids containing this TALE-sequence insertion were then used as the template for PCR to create origin-containing DNA templates as described above.

Radiolabeled DNA templates for ORC EMSA Assays

A plasmid containing two NcoI cutsites (pUC19-Ars1-NcoI) was created using site-directed mutagenesis (Quikchange Lightning, *Agilent*). The resulting plasmid was used as a template to make 5FdC-containing 1kB DNA templates as described above lacking the biotinylated end. 1.5-2μg of full length 5FdC-containing templates were then digested with NcoI and purified by phenol/chloroform extraction and ethanol precipitation. The resulting DNA was suspended in

10uL of TE and radiolabel added by a fill-in reaction with T4 DNA polymerase and alpha-labeled radioactive dCTP. The resulting radiolabeled DNA was run on a 5% PAGE gel in 1X TBE, imaged, and the correct size band cut out of the gel. Radiolabeled probe was isolated from the gel using the standard crush and soak method. Purified Probe was diluted 50uL and the CPM of each probe determined via scintillation counting.

ORC EMSA Assays:

2000 CPM of each DNA probe was incubated with or without M.HpaII (as indicated in the figure) in M.HpaII buffer in the presence of SAM to allow M.HpaII to crosslink. Increasing amounts of ORC (as indicated in the figures) was added to the probe and incubated in H buffer with 100mM KCl and poly dG/dC competitor DNA at room temperature for 10 minutes, then incubated on ice for 10minutes. The resulting ORC-bound templates were run on a 0.5X TBE 3.5% PAGE gel containing 80µg/mL BSA in 0.5X TBE containing 80µg/mL of BSA at 4°C at 100 milliamps until a Bromophenol Blue dye standard had run 2/3 of the way down the gel. The resulting gels were dried and imaged using autoradiography.

M.HpaII and TALE-labeled DNA templates:

PCR was used to create the large PCR products with a 5FdC-modified M.HpaII methylation sequence (as described above) containing the TALE-binding sequence at the -25, -35 and -45 locations. The resulting DNA was purified using a PCR clean-up column (Clonetech) and coupled to M280 streptavidin-coated magnetic beads (Invitrogen). Templates were first incubated with M.HpaII, then with TALE protein as described above. Templates bound to both TALE and M.HpaII were then tested in various assays.

Chromatin-based helicase loading assay:

Assays were performed as described in Eaton et al., 2010.

References:

Azmi, I.F., Watanabe, S., Maloney, M.F., Kang, S., Belsky, J.A., MacAlpine, D.M., Peterson, C.L., and Bell, S.P. (2017). Nucleosomes influence multiple steps during replication initiation. Elife *6*, e22512.

Bell, S.P., and Labib, K. (2016). Chromosome Duplication in Saccharomyces cerevisiae. Genetics *203*, 1027–1067.

Bell, S.P., and Stillman, B. (1992). ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature *357*, 128–134.

Belsky, J.A., MacAlpine, H.K., Lubelsky, Y., Hartemink, A.J., and MacAlpine, D.M. (2015). Genome-wide chromatin footprinting reveals changes in replication origin architecture induced by pre-RC assembly. Genes Dev. *29*, 212–224.

Berbenetz, N.M., Nislow, C., and Brown, G.W. (2010). Diversity of Eukaryotic DNA Replication Origins Revealed by Genome-Wide Analysis of Chromatin Structure. PLoS Genet. *6*, e1001092.

Bleichert, F., Botchan, M.R., and Berger, J.M. (2015). Crystal structure of the eukaryotic origin recognition complex. Nature *519*, 321–326.

Broach, J.R., Li, Y.Y., Feldman, J., Jayaram, M., Abraham, J., Nasmyth, K.A., and Hicks, J.B. (1983). Localization and sequence analysis of yeast origins of DNA replication. Cold Spring Harb. Symp. Quant. Biol. *47 Pt 2*, 1165–1173.

Chen, L., MacMillan, a M., Chang, W., Ezaz-Nikpay, K., Lane, W.S., and Verdine, G.L. (1991). Direct identification of the active-site nucleophile in a DNA (cytosine-5)-methyltransferase. Biochemistry *30*, 11018–11025.

Diffley, J., and Stillman, B. (1988). Purification of a yeast protein that binds to origins of DNA replication and a transcriptional silencer. Proc. Natl. Acad. Sci. *85*, 2120–2124.

Donovan, S., Harwood, J., Drury, L.S., and Diffley, J.F. (1997). Cdc6p-dependent loading of Mcm proteins onto pre-replicative chromatin in budding yeast. Proc. Natl. Acad. Sci. U. S. A. *94*, 5611–5616.

Duxin, J.P., Dewar, J.M., Yardimci, H., and Walter, J.C. (2014). Repair of a DNA-Protein Crosslink by Replication-Coupled Proteolysis. Cell *159*, 346–357.

Eaton, M.L., Galani, K., Kang, S., Bell, S.P., and MacAlpine, D.M. (2010). Conserved nucleosome positioning defines replication origins. Genes Dev. *24*, 748–753.

Eisenberg, S., Civalier, C., and Tye, B.K. (1988). Specific interaction between a Saccharomyces cerevisiae protein and a DNA element associated with certain autonomously replicating sequences. Proc. Natl. Acad. Sci. U. S. A. *85*, 743–746.

Evrin, C., Clarke, P., Zech, J., Lurz, R., Sun, J., Uhle, S., Li, H., Stillman, B., and Speck, C. (2009). A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. Proc. Natl. Acad. Sci. U. S. A. *106*, 20240–20245.

Fox, C.A., Loo, S., Rivier, D.H., Foss, M.A., and Rine, J. (1993). A transcriptional silencer as a specialized origin of replication that establishes functional domains of chromatin. Cold Spring Harb. Symp. Quant. Biol. *58*, 443–455.

Frigola, J., Remus, D., Mehanna, A., and Diffley, J.F.X. (2013). ATPase-dependent quality control of DNA replication origin licensing. Nature *495*, 339–343.

Fu, Y.V., Yardimci, H., Long, D.T., Guainazzi, A., Bermudez, V.P., Hurwitz, J., van Oijen, A., Schärer, O.D., and Walter, J.C. (2011). Selective Bypass of a Lagging Strand Roadblock by the Eukaryotic Replicative DNA Helicase. Cell *146*, 931–941.

Huang, R.Y., and Kowalski, D. (1993). A DNA unwinding element and an ARS consensus comprise a replication origin within a yeast chromosome. EMBO J. *12*, 4521–4531.

Kang, S., Warner, M.D.D., and Bell, S.P.P. (2014). Multiple Functions for Mcm2-7 ATPase Motifs during Replication Initiation. Mol. Cell *55*, 655–665.

Lee, D.G., and Bell, S.P. (1997). Architecture of the yeast origin recognition complex bound to origins of DNA replication. Mol. Cell. Biol. *17*, 7159–7168.

Li, N., Zhai, Y., Zhang, Y., Li, W., Yang, M., Lei, J., Tye, B.-K., and Gao, N. (2015). Structure of the eukaryotic MCM complex at 3.8 Å. Nature *524*, 186–191.

Liachko, I., Youngblood, R. a., Keich, U., and Dunham, M.J. (2013). High-resolution mapping, characterization, and optimization of autonomously replicating sequences in yeast. Genome Res. *23*, 698–704.

Lin, S., and Kowalski, D. (1997). Functional equivalency and diversity of cis-acting elements among yeast replication origins. Mol. Cell. Biol. *17*, 5473–5484.

Lipford, J.R., and Bell, S.P. (2001). Nucleosomes Positioned by ORC Facilitate the Initiation of DNA Replication. Mol. Cell *7*, 21–30.

Long, D.T., Joukov, V., Budzowska, M., and Walter, J.C. (2014). BRCA1 promotes unloading of the CMG helicase from a stalled DNA replication fork. Mol. Cell *56*, 174–185.

Marahrens, Y., and Stillman, B. (1992). A yeast chromosomal origin of DNA replication defined by multiple functional elements. Science *255*, 817–823.

Randell, J.C.W.W., Bowers, J.L., Rodríguez, H.K., and Bell, S.P. (2006). Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase. Mol. Cell *21*, 29–39.

Rao, H., and Stillman, B. (1995). The origin recognition complex interacts with a bipartite DNA binding site within yeast replicators. Proc. Natl. Acad. Sci. *92*, 2224–2228.

Rao, H., Marahrens, Y., and Stillman, B. (1994). Functional conservation of multiple elements in yeast chromosomal replicators. Mol. Cell. Biol. *14*, 7643–7651.

Remus, D., Beuron, F., Tolun, G., Griffith, J.D., Morris, E.P., Diffley, J.F.X., Remus, D., and Beuron, F. (2009). Concerted Loading of Mcm2-7 Double Hexamers around DNA during DNA Replication Origin Licensing. Cell *139*, 719–730.

Rowley, a, Cocker, J.H., Harwood, J., and Diffley, J.F. (1995). Initiation complex assembly at budding yeast replication origins begins with the recognition of a bipartite sequence by limiting amounts of the initiator, ORC. EMBO J. *14*, 2631–2641.

Seki, T., and Diffley, J.F. (2000). Stepwise assembly of initiation proteins at budding yeast replication origins in vitro. Proc. Natl. Acad. Sci. U. S. A. *97*, 14115–14120.

Simpson, R.T. (1990). Nucleosome Positioning Can Affect the Function of a Cis-Acting DNA Element In Vivo. Nature *343*.

Sun, J., Evrin, C., Samel, S.A., Fernández-Cid, A., Riera, A., Kawakami, H., Stillman, B., Speck, C., Li, H., Fernandez-Cid, A., et al. (2013). Cryo-EM structure of a helicase loading intermediate

containing ORC-Cdc6-Cdt1-MCM2-7 bound to DNA. Nat. Struct. Mol. Biol. *20*, 944–951.

Sun, J., Fernandez-Cid, A., Riera, A., Sun, J., Fernandez-Cid, A., Riera, A., Tognetti, S., Yuan, Z., Stillman, B., Speck, C., et al. (2014). Structural and mechanistic insights into Mcm2-7 double-hexamer assembly and function. Genes Dev. *28*, 2291–2303.

Tocilj, A., On, K.F., Yuan, Z., Sun, J., Elkayam, E., Li, H., Stillman, B., Joshua-Tor, L., Echols, N., Kuriyan, J., et al. (2017). Structure of the active form of human origin recognition complex and its ATPase motor module. Elife *6*, 30810–30821.

Wilmes, G.M., and Bell, S.P. (2002). The B2 element of the Saccharomyces cerevisiae ARS1 origin of replication requires specific sequences to facilitate pre-RC formation. Proc. Natl. Acad. Sci. U. S. A. *99*, 101–106.

Xu, W., Aparicio, J.G., Aparicio, O.M., and Tavaré, S. (2006). Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in S. cerevisiae. BMC Genomics *7*, 257–286.

Yuan, Z., Riera, A., Bai, L., Sun, J., Nandi, S., Spanos, C., Chen, Z.A., Barbon, M., Rappsilber, J., Stillman, B., et al. (2017). Structural basis of Mcm2–7 replicative helicase loading by ORC–Cdc6 and Cdt1. Nat. Struct. Mol. Biol. *24*, 316–324.

Zou, L., and Stillman, B. (2000). Assembly of a Complex Containing Cdc45p, Replication Protein A, and Mcm2p at Replication Origins Controlled by S-Phase Cyclin-Dependent Kinases and Cdc7p-Dbf4p Kinase. Mol. Cell. Biol. *20*, 3086–3096.

Chapter III:

Further Discussion and Future Directions

Key Conclusions:

During G1, ORC, Cdc6, and Cdt1 coordinate to load an inactive Mcm2-7 double-hexamer onto origin DNA. Despite the increasingly detailed understanding of the protein interactions during this process, the contribution of the origin DNA sequence outside of initial ORC binding remains unclear. In Chapter II, we use protein-DNA adducts to investigate where the Mcm2-7 complex initially loads on origin DNA. Placement of an adduct within the initial ORC/Cdc6 or Mcm2-7 binding site stops ORC/Cdc6 and Mcm2-7 from stably encircling dsDNA. Using this technique, we found a large region of origin DNA which needs to be adduct-free for initial ORC binding. The size and location of this region is consistent with ORC footprinting results (Bell and Stillman, 1992), but inconsistent with the ORC-DNA binding region predicted in recent structures (Bleichert et al., 2015; Tocilj et al., 2017; Yuan et al., 2017). Also inconsistent with the current structure-based helicase loading model, we did not observe a 60 bp window of DNA that needed to be adduct-free for loading of the Mcm2-7 double hexamer (Li et al., 2015; Sun et al., 2014). To accommodate our data, we proposed a model where a helicase-loading intermediate slides along DNA, allowing the helicase loading process to occur away from origin DNA (Chapter II, Figure 17B).

Restricting access to DNA by placing adducts upstream and downstream of the ACS had two key effects on helicase loading. First, we observe an ~30 bp region downstream of the area needed for stable ORC binding that affects Mcm2-7 hexamer association specifically. Second, under conditions where two protein adducts flank the ACS, the B2 element sequence is essential for stable helicase loading. A B2-dependent helicase loading defect was not observed in unobstructed templates, or templates with a single adduct. Helicase loading onto nucleosomal templates, which also restrict the DNA that can be explored during the helicase loading process, was also dependent on B2. In this chapter, I will discuss questions raised by these data and propose future experiments to address them.

How is does Mcm2-7 load when a protein adduct is on the DNA?

In Chapter II, I present a simple model in which an Mcm2-7 intermediate can slide along DNA to finish loading the Mcm2-7 double hexamer if a protein-DNA adduct blocks the site where the helicase might otherwise bind (Chapter II, Figure 14B). In support of this model, placing a protein adduct both upstream and downstream of the ACS allowed us to identify an ~30bp region where helicase loading, but not ORC binding, was defective. Why don't we observe a 60 bp region, the predicted amount of DNA covered by the Mcm2-7 double hexamer? We hypothesize that the first Mcm2-7 hexamer binds to a region where protein-DNA adducts also destabilize ORC binding. We propose that the ~30 bp we observe when we restrict sliding corresponds to loading of the second Mcm2-7 hexamer. However, this model cannot be tested with the assays described in Chapter II. Our assays can only identify the earliest defect in the helicase loading process. If an adduct affects an early step, downstream steps never occur. Recognition of the ORC-ACS by ORC is the first step in the origin licensing process. Thus, if placement of an adduct at a certain location would affect first ORC binding and then Mcm2-7 loading, our assays would only be able to detect the first defect. In support of a model where DNA that first interacts with ORC later interacts with the Mcm2-7 complex, the structure of the OCCM loading intermediate covers roughly 55 bps of DNA. Why might ORC cover the region where the first Mcm2-7 hexamer binds? It is possible ORC protects the region needed for Mcm2-7 binding, to ensure once it recruits Cdc6 and Mcm2-7/Cdt1, it can rapidly convert them to stable single hexamer.

The model presented in Chapter II proposes that an intermediate in the helicase loading process can slide along dsDNA (Chapter II, Figure 14B). Which intermediate(s) in the helicase loading process slide(s)? If the first Mcm2-7 is loaded within the region needed for stable ORC binding (as our sliding model predicts), sliding would first be required at the ORC-Mcm2-7 stage. The ORC/Mcm2-7 complex is a good candidate for an intermediate that can slide. Single-molecule assays show that once Cdc6/Cdt1 depart, the first hexamer is topologically linked to

the DNA and the Mcm2-7 single hexamer is stable through high salt extraction (Ticau et al., 2015, 2017). Despite the appeal of this intermediate, the only complex we can eliminate as a sliding intermediate is the OCCM complex. We demonstrate that the OCCM complex in ATPγS does not slide (Chapter II, Figure 4). However, a complex containing all the same components may be able to slide following an ATP hydrolysis event. Indeed, modeling of the helicase loading process on a single-molecule level suggests that a number of distinct complexes may exist that are comprised on these same protein components (data not shown).

To determine which intermediates can slide, experiments need to be undertaken where the DNA sequence that interacts helicase loading components can be isolated directly. With recent increases in our knowledge of the Mcm2-7 double-hexamer structure (Li et al., 2015) and the OCCM complex (Yuan et al., 2017), it might be possible to introduce a DNA-crosslinking agent at a specific amino acid within the Mcm2-7 or ORC/Cdc6 central channels. Samples could be isolated at various times during the helicase loading process and the DNA within the channels sequenced. However, such an experiment would not be trivial. One might also be able to monitor sliding in a single-molecule experiments. Sliding of DNA-associated labeled Mcm2-7 or ORC complex can theoretically be monitored by changes in the intensity of the fluorophore as it moves within the evanescent field during TIRF microscopy (Figure 1A). However, these intensity changes would likely be quite subtle.

Can Mcm2-7 helicases be loaded in different regions of the origin *in vivo*?

It is unclear whether Mcm2-7 helicases are consistently loaded at the same location *in vivo* or whether shifting to accommodate a protein-DNA lesion can only occur *in vitro.* Both ORC and Mcm2-7 usually give a single peak of signal at origins by chromatin immunoprecipitation (ChIP) (Eaton et al., 2010; Wyrick et al., 2001; Xu et al., 2006) but small shifts in the location of Mcm2-7 or ORC would likely be within error of these experiments. Flexibility in where Mcm2-7 loads within origin sequences makes sense from an evolutionary standpoint. The organization

of nucleosomes around the NFR are known to change during the cell cycle (Belsky et al., 2015) and likely change due to outside stimuli as well. Because DNA replication is essential for continued cell proliferation, it would be advantageous to the cell to be able to accommodate these small changes. A recent study suggests that the Mcm2-7 double hexamer is capable of unwinding from many different locations within the genome (Gros et al., 2015). Thus, replication initiation would likely take place regardless of where loading occurs.

What role does the B2 element sequence play in helicase loading?

The data presented in Chapter II suggests that having enough space to accommodate Mcm2-7 binding is insufficient for salt-stable helicase loading. When sliding is minimized, either by addition of two protein adducts on either side of the ACS or by nucleosomes, the B2 element sequence becomes important for stable helicase loading. The experiments in Chapter II were unable to identify for what step in the helicase-loading process the B2 element is important. Since B2 has sequence specificity, it is likely to be a protein binding site.

The B2 element has previously been proposed as either an ORC binding site or an Mcm2-7 binding site. *In vitro* and *in vivo* data support and contradict both of these hypotheses. In support of the B2 element as an ORC binding site, B2 in *ARS1* is a 9/11 match to the ACS motif and ORC can bind B2 (Bell and Stillman, 1992; Wilmes and Bell, 2002). Importantly, the match exists in the opposite orientation to the ACS (Chapter 1, Figure 3) (Wilmes and Bell, 2002). The N-terminal domains of the Mcm2-7 hexamers must interact to form the double hexamer (Li et al., 2015). Assuming the second Mcm2-7 hexamer is loaded via the same mechanism as the first (which is supported experimentally (Frigola et al., 2013)), the second ORC must bind origin DNA in the opposite orientation to maintain its own contacts with Mcm2-7 and align the two Mcm2-7 N-termini (Chapter 1, Figure 13).

In support of the Mcm2-7 binding site hypothesis, mutations in the B2 element show no difference in ORC binding by ChIP *in vivo*, but a defect in Mcm2-7 binding is observed (Lipford

and Bell, 2001; Zou and Stillman, 2000). The experiments described in Chapter II also do not reveal any ORC binding defects when adducts are placed within the B2 element, or when the B2 element is mutated. Single-molecule experiments definitively show that helicase loading is capable of occurring with only a single ORC present (Ticau et al., 2015). However, a different mechanism may be used *in vivo*, especially under conditions where sliding is restricted.

Single-molecule analysis could be used to determine whether B2 acts as a second ORC binding site or aids in Mcm2-7 recruitment directly when sliding is restricted. The conventional assay utilizes a labeled 1.3 kb DNA template coupled to a glass coverslip, and allows the user to monitor co-localization events of fluorescently-labeled proteins of interest (Figure 1A)(Ticau et al., 2015). We can easily replace the DNA templates used in this conventional assay and use instead versions with both a M.HpaII and a TALE protein bound to prevent sliding (Figure 1B). Association of Mcm2-7, ORC, and Cdc6 can then be monitored in this context, and changes in both stoichiometry of the complexes required for helicase loading and association/dissociation rates observed.

Figure 1

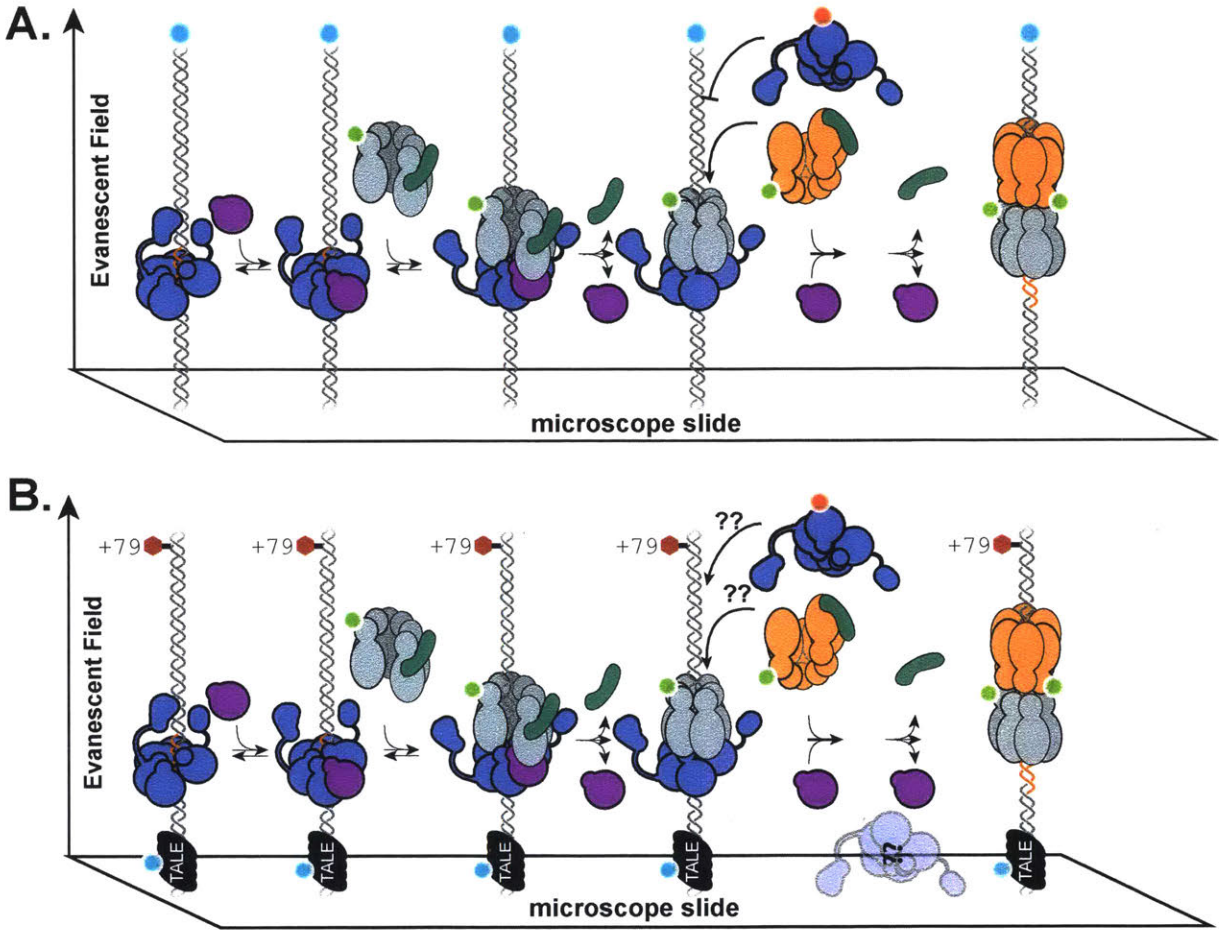


Figure 1: Single-Molecule Experiments to address helicase loading mechanism with two protein-DNA adducts.

A) Set-up used for experiments conducted in Ticau et al, 2015. A blue flurophore was added to the end of the origin DNA template, and the association of labelled proteins (ORC and Mcm2-7 diagrammed here) at labeled DNA is monitored. In these experiments, only a single ORC molecule was required to load an Mcm2-7 double hexamer.

B) Modified Set-up to examine the helicase loading reaction when adducts are placed around the origin. A blue flurophore is attached to the TALE protein, and M.HpaII is added. Association of proteins can be monitored as done previously to see if stochiometry or kinetics changes in this context. A B2 element mutation can be introduced to see how the reaction changes.

References:

Bell, S.P., and Stillman, B. (1992). ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature *357*, 128–134.

Belsky, J.A., MacAlpine, H.K., Lubelsky, Y., Hartemink, A.J., and MacAlpine, D.M. (2015). Genome-wide chromatin footprinting reveals changes in replication origin architecture induced by pre-RC assembly. Genes Dev. *29*, 212–224.

Bleichert, F., Botchan, M.R., and Berger, J.M. (2015). Crystal structure of the eukaryotic origin recognition complex. Nature *519*, 321–326.

Eaton, M.L., Galani, K., Kang, S., Bell, S.P., and MacAlpine, D.M. (2010). Conserved nucleosome positioning defines replication origins. Genes Dev. *24*, 748–753.

Frigola, J., Remus, D., Mehanna, A., and Diffley, J.F.X. (2013). ATPase-dependent quality control of DNA replication origin licensing. Nature *495*, 339–343.

Gros, J., Kumar, C., Lynch, G., Yadav, T., Whitehouse, I., and Remus, D. (2015). Post-licensing Specification of Eukaryotic Replication Origins by Facilitated Mcm2-7 Sliding along DNA. Mol. Cell *60*, 797–807.

Li, N., Zhai, Y., Zhang, Y., Li, W., Yang, M., Lei, J., Tye, B.-K., and Gao, N. (2015). Structure of the eukaryotic MCM complex at 3.8 Å. Nature *524*, 186–191.

Lipford, J.R., and Bell, S.P. (2001). Nucleosomes Positioned by ORC Facilitate the Initiation of DNA Replication. Mol. Cell *7*, 21–30.

Sun, J., Fernandez-cid, A., Riera, A., Sun, J., Fernandez-cid, A., Riera, A., Tognetti, S., and Yuan, Z. (2014). Structural and mechanistic insights into Mcm2 – 7 double-hexamer assembly and function. 2291–2303.

Ticau, S., Friedman, L.J., Ivica, N.A., Gelles, J., and Bell, S.P. (2015). Single-molecule studies of origin licensing reveal mechanisms ensuring bidirectional helicase loading. Cell *161*, 513–525.

Ticau, S., Friedman, L.J., Champasa, K., Corrêa, I.R., Gelles, J., and Bell, S.P. (2017). Mechanism and timing of Mcm2–7 ring closure during DNA replication origin licensing. Nat. Struct. Mol. Biol. *24*, 309–315.

Tocilj, A., On, K.F., Yuan, Z., Sun, J., Elkayam, E., Li, H., Stillman, B., Joshua-Tor, L., Echols, N., Kuriyan, J., et al. (2017). Structure of the active form of human origin recognition complex and its ATPase motor module. Elife *6*, 30810–30821.

Wilmes, G.M., and Bell, S.P. (2002). The B2 element of the Saccharomyces cerevisiae ARS1 origin of replication requires specific sequences to facilitate pre-RC formation. Proc. Natl. Acad. Sci. U. S. A. *99*, 101–106.

Wyrick, J.J., Aparicio, J.G., Chen, T., Barnett, J.D., Jennings, E.G., Young, R.A., Bell, S.P., and Aparicio, O.M. (2001). Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: high-resolution mapping of replication origins. Science *294*, 2357–2360.

Xu, W., Aparicio, J.G., Aparicio, O.M., and Tavaré, S. (2006). Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in S. cerevisiae. BMC Genomics *7*, 257–286.

Yuan, Z., Riera, A., Bai, L., Sun, J., Nandi, S., Spanos, C., Chen, Z.A., Barbon, M., Rappsilber, J., Stillman, B., et al. (2017). Structural basis of Mcm2–7 replicative helicase loading by ORC–Cdc6 and Cdt1. Nat. Struct. Mol. Biol. *24*, 316–324.

Zou, L., and Stillman, B. (2000). Assembly of a Complex Containing Cdc45p, Replication Protein A, and Mcm2p at Replication Origins Controlled by S-Phase Cyclin-Dependent Kinases and Cdc7p-Dbf4p Kinase. Mol. Cell. Biol. *20*, 3086–3096.