The role of DNA sequence during helicase loading at *S. cerevisiae* origins of replication

by

Wendy M. Lam

B.A. Molecular and Cell Biology
University of California, Berkeley, 2003

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

September 2010

Signature of Author: _____
Department of Biology
June 28, 2010

Certified by: _____
Stephen P. Bell
Professor of Biology
Thesis Supervisor

Accepted by: _____
Stephen P. Bell
Professor of Biology
Chair, Committee for Graduate Students

The role of DNA sequence during helicase loading at *S. cerevisiae* origins of replication

By

Wendy M. Lam

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

ABSTRACT

Eukaryotic chromosomal DNA replication is a tightly regulated process that initiates at multiple origins of replication throughout the genome. As cells enter the G1 phase of the cell cycle, the Mcm2-7 replicative helicase is loaded at all potential origins of replication but remains inactive until S phase. Although the proteins involved in helicase loading have been elucidated, little is understood about the role of origin DNA sequence in helicase loading beyond its role in recruiting the initiator ORC.

Of the eukaryotic origins of replication studied, those derived from the budding yeast *Saccharomyces cerevisiae* are most defined. These origins contain multiple functional DNA elements: a conserved and essential autonomously replicating sequence (ARS) consensus sequence (ACS) and multiple non-conserved B-elements that collectively are essential for origin function. Although the ACS and one of the B elements (B1) have been demonstrated to bind ORC, the functions of the other B-elements are poorly understood. *In vivo* studies indicate that the B2-element functions during helicase loading, but whether B2 is important for the initial recruitment or for the stable loading of the helicase (i.e. topological linkage around DNA) had not been determined.

To understand how origin DNA facilitates helicase loading, I identified the specific DNA length and sequence requirements for helicase loading. I found that Mcm2-7 helicases are predominantly located at the region containing the B-elements, termed the B-side, at origins *in vivo*. This region coincides with a nucleosome-free region associated with each origin. Deletion of sequences within this region had no effect on the initial association of Mcm2-7 complexes to the origin, but either reduced or eliminated stable Mcm2-7 loading. *In vitro* assays that restricted the accessibility of ORC-adjacent DNA, either by the presence of nucleosomes or by truncation, revealed a function for B2 during helicase loading. Further analysis showed that B2 functions after the initial association of Mcm2-7 to facilitate helicase loading. This work provides insights into how origin sequences facilitate the specific protein/DNA and nucleosomal architecture necessary for helicase loading at eukaryotic origins of replication.

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

Acknowledgements

First and foremost, I would like to thank my thesis advisor, Steve Bell. His guidance, wisdom, and support have been invaluable to my growth as a scientist over the years. I would also like to thank my colleagues in the Bell Lab for creating a stimulating and enjoyable environment to work in. It has truly been a pleasure working with a cohort of immensely talented scientists and wonderful friends. I would also like to acknowledge my past mentors, Paul Kaufman and Alexa Franco, for teaching me by example to love the science. I am especially indebted to Paul for his friendship, candor, and inspiration even after so many years – and don't worry, Paul, I am "all set."

I would like to thank the members of my thesis committee -- Angelika Amon, Tania Baker, David Jeruzalmi and Frank Solomon -- for their scientific advice and critical feedback over the years. I would especially like to thank Frank, who truly went above and beyond the call of duty by imparting his consistently sage advice whenever I knocked on his door. Science could use some more kids like you, Frank. Don't ever change.

I am greatly indebted to my family and friends for their love and support. I would like to acknowledge a few special individuals for making my years at MIT some of the best of my life: Jane Kim, Renuka Sastry, Michelle Sander, Nan Gu, Sudeep Agarwala, and Robert Wang. Words cannot express how grateful I am to have these remarkable people in my life. They have opened my eyes up to so many new things in the past six years, and to this day their utter and extreme awesomeness never ceases to amaze me.

Last but not least, I would like to thank my mother, who was truly an extraordinary woman in so many ways. Without her inspiration, encouragement and support (not to mention her 23 chromosomes), I would not be here today. This thesis is dedicated in loving memory of her.

TABLE OF CONTENTS

Chapter I

Introduction

OVERVIEW

During the eukaryotic cell cycle, genome duplication occurs in a highly accurate and timely fashion prior to chromosome segregation and cell division. To ensure genomic stability, chromosomal replication is tightly coupled to the progression of the eukaryotic cell cycle, occurring exactly once during each S phase. The events that lead to the initiation of DNA replication begin at the end of mitosis and continue during the G1 phase of the cell cycle. Although no DNA synthesis occurs during this time, the genomic sequences at which replication initiates, called origins of replication, are selected and primed for activation in S phase (Bell and Dutta 2002). This process has been given several names, including origin licensing, origin selection and pre-RC formation, but is most simply and accurately described as the loading of the replicative helicase onto origin DNA.

The initiation of DNA replication in eukaryotes is strictly divided into two phases: origin selection and origin activation (Fig. 1). The temporal separation of these two phases is controlled by cell cycle-dependent changes in the activity of two kinases: Cyclin Dependent Kinase (CDK) and Dbf4-dependent Kinase (DDK) (Bell and Dutta 2002). In the origin selection phase, the Mcm2-7 helicase is assembled onto origins of replication throughout the genome. Helicase loading is mediated by the pre-Replicative Complex (pre-RC), which, along with the Mcm2-7 complex, consists of the six-subunit Origin Recognition Complex (ORC), Cdc6, and Cdt1 (Bell and Dutta 2002). For the

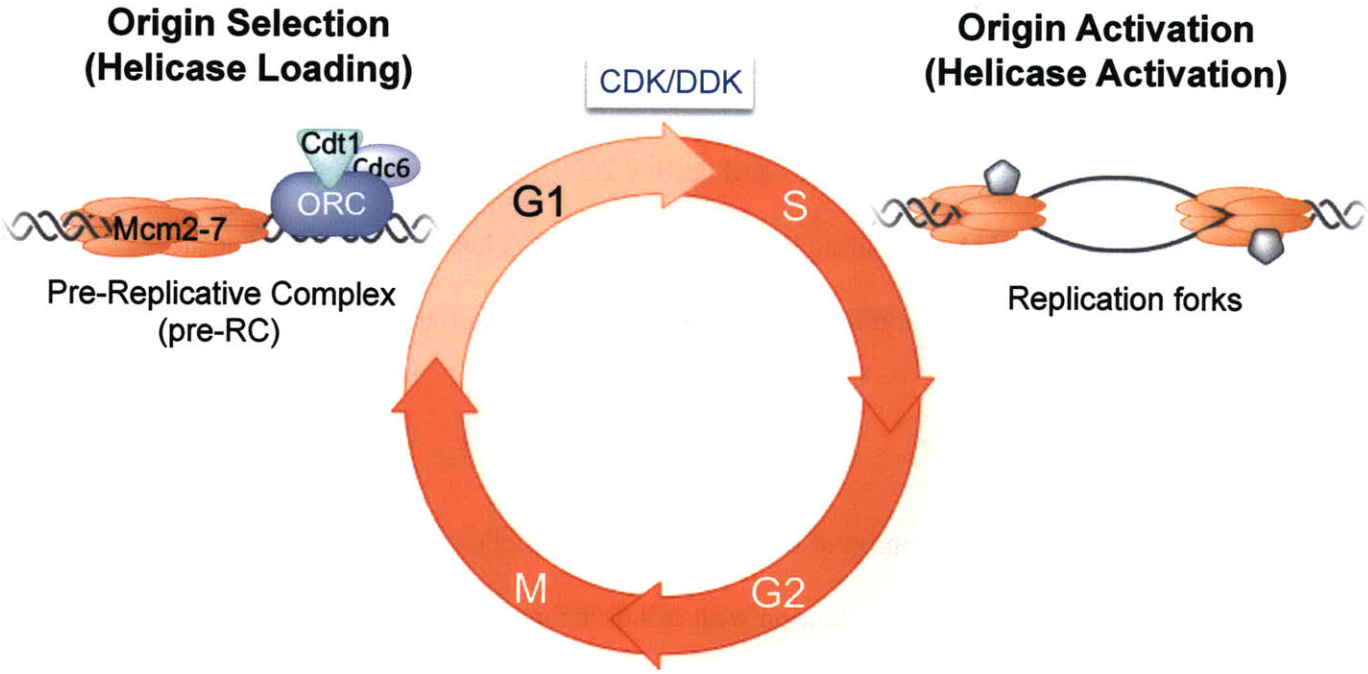


Figure 1. Cell cycle regulation of DNA replication initiation in eukaryotes.

In late mitosis and in G1 phase, when CDK activity is low, pre-Replicative Complexes (pre-RCs) assemble at all potential origins of replication, resulting in the loading of the replicative Mcm2-7 helicase onto DNA. During the G1-to-S-phase transition, an increase in CDK/DDK activity leads to the recruitment of additional replication factors and polymerases to origins (grey pentagons). The Mcm2-7 helicase is activated to unwind DNA, and DNA synthesis initiates. The high level of CDK activity during the S, G2, and M phases (dark pink) inhibits the re-assembly of pre-RC complexes at origins, thus preventing the re-replication of genomic DNA within a single cell cycle.

duration of G1, when CDK and DDK activity is low, the Mcm2-7 helicase remains at origins and is inactive for DNA unwinding.

Origin activation is triggered by an increase in CDK and DDK activity during the G1 to S transition. These kinases promote the binding of additional initiation factors, including Cdc45, Sld2, Sld3, Dpb11 and the hetero-tetrameric GINS complex, to the proteins already at the origin (Remus and Diffley 2009). The association of these proteins activates DNA unwinding by the replicative DNA helicase, allowing for the subsequent recruitment of multiple DNA polymerases to generate a pair of replication forks.

The temporal separation of helicase loading and helicase activation is crucial for the cell to coordinate DNA replication with cell division and to ensure that the genome is replicated exactly once per cell cycle. Pre-RCs assembled in G1 are dismantled after replication, and the reassembly of pre-RCs outside of G1 phase is prevented by various means (Blow and Dutta 2004). The coordination of initiation events with cell cycle progression ensures that helicase loading at origins happens exactly once per cell cycle, and that helicases are activated only once per cell cycle.

The general events of replication initiation described above, as well as the initiator proteins involved, are highly conserved from yeast to man. Much of our understanding of how eukaryotic initiator proteins function at a molecular level comes from studies of the budding yeast *S. cerevisiae*. For example, the mechanisms by which

ORC and Cdc6 act to load the Mcm 2-7 helicase onto origins have been most extensively studied in budding yeast (Bowers et al. 2004; Randell et al. 2006). Subsequent studies of their homologues in *S. pombe*, *X. laevis*, *D. melanogaster* and mammalian cells revealed functional similarities.

Despite the high level of conservation of initiator proteins, the origin sequences at which initiation takes place are highly divergent among eukaryotic organisms (Cvetic and Walter 2005). *S. cerevisiae* origins, called Autonomously Replicating Sequences (ARSs), are relatively short in length (100-150 bp) and contain varying combinations of short (10-12 bp) DNA sequence elements that contribute to origin function. One of these sequence elements, termed the ARS consensus sequence (ACS), is essential for origin function and is found in every *S. cerevisiae* origin (Celniker et al. 1984; Van Houten and Newlon 1990). In contrast, origins in all other eukaryotes studied are far more ambiguous in terms of sequence. Although it has been demonstrated that origins in several of these organisms exhibit some general characteristics in common (such as AT-richness), no conserved sequence elements analogous to the ACS in *S. cerevisiae* has been identified in any other eukaryote. In the most extreme case, in *Xenopus* and *Drosophila* embryonic systems, any DNA sequence can be used as an origin of replication. Despite this lack of sequence specificity in other eukaryotes, it is clear that replication initiation events in these systems are not random, and even in the embryonic systems in which any DNA sequence is competent for initiation, the initiator protein ORC preferentially binds AT-rich DNA (Kong et al. 2003).

Why is there such a significant discrepancy between the high conservation of initiator proteins and lack of conservation of origin sequences among eukaryotes? The fact that the initiator proteins are so highly conserved suggests that the basic mechanism by which they assemble and function at an origin must also be conserved. Thus it is reasonable to hypothesize that the nature of the interaction of those proteins with origin DNA, whether it is in budding yeast or in a mouse cell, should also be conserved. Although origin sequences are not conserved among eukaryotes, it is likely that other features of origins, such as flexibility, helical instability, or chromatin structure and modifications, are conserved among organisms. Thus it is important to not only identify which sequences are important for origin function in *S. cerevisiae* but also how they are functionally important, as this may provide insights into what characteristics define an origin in more complicated systems.

Despite the fact that *S. cerevisiae* origin sequences have been extensively characterized relative to those of other model organisms, many details of how these DNA sequences facilitate helicase loading remain unclear. Up to this point, most studies of pre-RC assembly have focused primarily on identifying all the protein components necessary and sufficient for helicase loading *in vitro* and characterizing their biochemical functions. However, less is understood about the origin DNA at which these proteins are assembled, and the exact nature of their interactions. Previous studies have demonstrated that several origin sequence elements in *S. cerevisiae* are important for pre-RC assembly, but the function of these elements remain poorly characterized. For which steps in the mechanism of pre-RC assembly are these sequence elements necessary, and how do they

facilitate pre-RC assembly at a molecular level? In addition, with the exception of ORC, virtually nothing is known about how the remaining pre-RC components are arranged spatially at the origin. Of particular interest is where the Mcm2-7 helicase is loaded at the origin. Does DNA sequence play a role in dictating where helicase loading occurs at the origin, and if so, how?

This thesis will focus on the functional contribution of origin DNA to the stepwise mechanism of pre-RC assembly. In this introduction, I will review our current understanding of origin sequences in different eukaryotic organisms, with a focus on *S. cerevisiae*. Additionally, I will review the stepwise mechanism of helicase loading, as well as what is understood about the architecture of the pre-RC bound to origin DNA. Elucidating how origin DNA facilitates the mechanism of helicase loading in the simplest model eukaryotic organism will not only add to our understanding of how this initiation event establishes the foundation for subsequent events in DNA replication, but also provide us further insights into how it could work in more complex eukaryotic systems.

ORIGINS OF REPLICATION

In 1963, Jacob, Brenner and Cuzin proposed the replicon model to describe the regulation of genome duplication in *E. coli* (Jacob and Brenner 1963). By this model, a positive *trans*-acting factor (the initiator) binds to a *cis*-acting sequence (the replicator) in the bacterial chromosome to activate DNA replication. Over forty years later, the key aspects of the replicon paradigm still hold true, not only for *E. coli* but for other prokaryotes and eukaryotes as well. However, just as the operon model failed to capture the complexity of transcriptional regulation, the replicon model is too simplistic to fully explain the intricate nature of chromosomal replication control in eukaryotes.

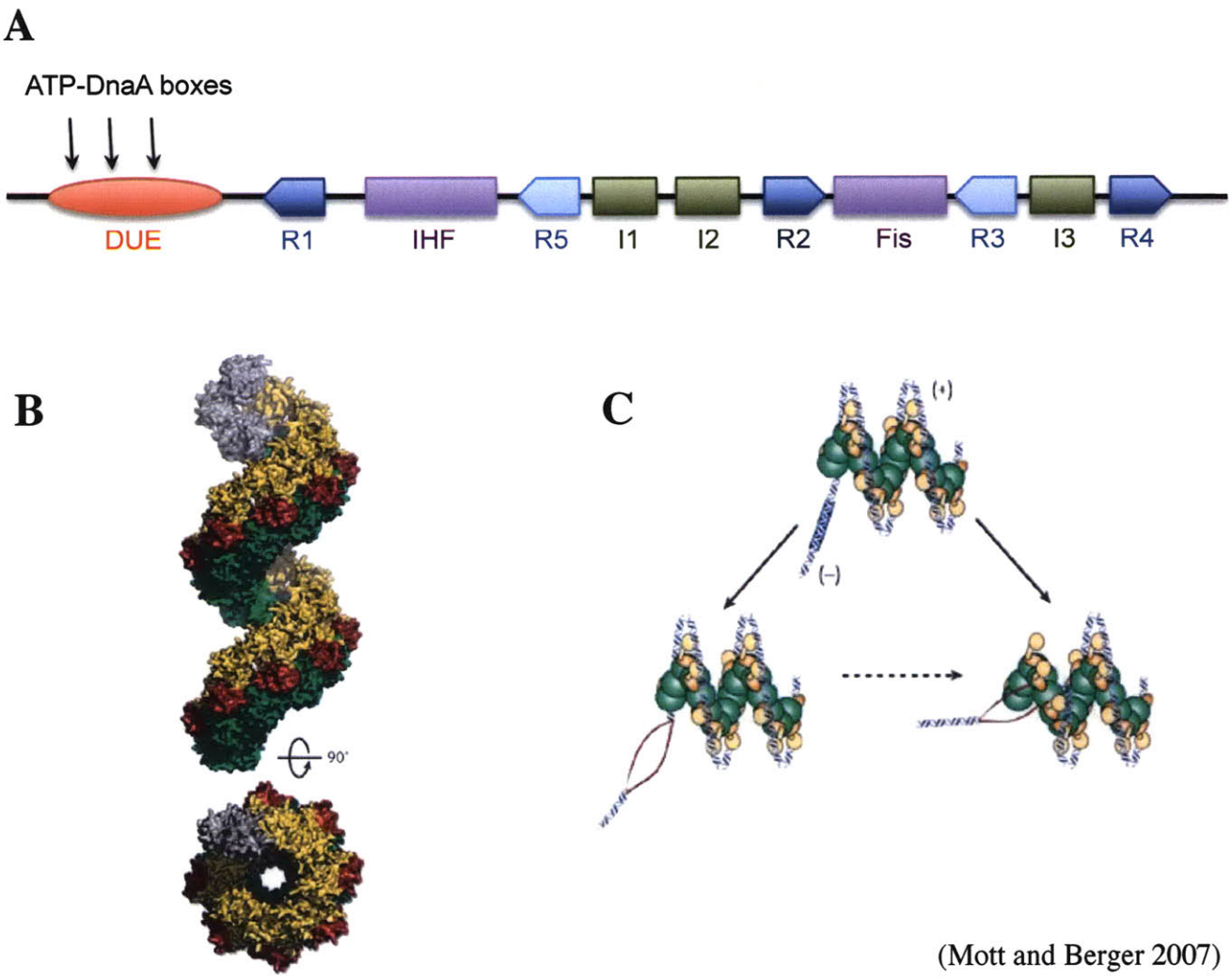
To gain a more thorough understanding of how chromosomal replication is initiated in eukaryotes, we must expand the replicon paradigm to take such complexities into account. First, we need to understand what defines a replicator, not only in terms of DNA sequence but other physical features as well, since many organisms lack conserved origin sequences. We need to understand how specific features within the replicator contribute functionally (as protein binding sites, helically unstable regions, etc.) to facilitate replication initiation. Second, we need to understand the nature by which the initiator interacts with the replicator. Specifically, what physical contacts are made between the replicator and the initiator? What are the factors (conformational changes, DNA topology, protein modifications, to name a few) that modulate and regulate these interactions?

In this section, I will begin by reviewing and comparing the replicator sequences and initiator proteins of prokaryotic and eukaryotic organisms, with a focus on *S. cerevisiae*.

Escherichia coli

The bi-directional replication of the entire circular *E. coli* genome begins at a single origin, *oriC*, which was first cloned and mapped in a screen for genomic sequences that conferred replication competence to a plasmid (Yasuda and Hirota 1977; Mott and Berger 2007). *oriC* is approximately 250 bp in length and contains multiple 9mer repeats called DnaA boxes that serve as sequence-specific binding sites for the initiator protein DnaA (Fig. 2A). In addition to the array of DnaA boxes, *oriC* also contains a second conserved element, three AT-rich 13mer repeats (Bramhill and Kornberg 1988) that were found to act as a DNA Unwinding Element (DUE) (Kowalski and Eddy 1989). A conserved component of origins of replication across all organisms, a DUE is generally AT-rich in sequence and forms a helical structure that is easily melted, which facilitates the initial unwinding of duplex DNA by the initiator. Lastly, *oriC* also contains binding sites for accessory proteins, such as the bacterial histone-like proteins HU, IHF, and Fis (Drlica and Rouviere-Yaniv 1987).

The DnaA boxes within the array at *oriC* vary slightly in sequence. A single monomer of DnaA binds most strongly to the DnaA box consensus sequence 5'-TTATCCACA-3' and its affinity for other DnaA box sequences depends on the degree of similarity to the consensus sequence. DnaA, which is a member of the AAA+ family of



(Mott and Berger 2007)

Figure 2. *E. coli*: replicator and initiator.

(A) Diagram of the sequence features of the *oriC* origin of replication in *E. coli*. The DNA Unwinding Element (DUE) is indicated in red; three ATP-DnaA boxes are located within the DUE region. High-affinity DnaA boxes (R1, R2, and R4) are shown in dark blue, and weaker affinity DnaA boxes (R3 and R5) are shown in light blue. I sites (another class of sequences bound by DnaA that differ slightly from the DnaA box consensus sequence) are indicated in green. Binding sites for histone-like factors IHF and Fis are shown in purple.
(B) ATP-DnaA forms a right-handed helical filament. Side and axial view of four DnaA tetramers. The AAA+ ATPase is indicated in green, the DNA binding domain is indicated in yellow, and the lid domain (which plays a role in DnaA oligomerization) is incidated in red.
(C) Model for DNA unwinding by DnaA. The formation of a right-handed, toroidal DNA wrap around the DnaA core destabilizes the origin by introducing strain into the DUE through compensatory negative supercoiling (middle panel). Once the DUE is melted, ATP-bound DnaA directly engages the single-stranded region to maintain its unwound state (right panel).

ATPases, binds to the consensus DnaA box with similar affinities in its ATP- and ADP-bound form (Speck et al., 1999). However, ATP binding by DnaA allows it to recognize and bind an additional 6-bp element called the 'ATP-DnaA box' (Speck et al. 1999; Speck and Messer 2001). ATP-DnaA boxes are low-affinity binding sites, and binding to this 'weak' site is facilitated by DnaA binding to an adjacent 9mer high-affinity DnaA box. DnaA remains bound to three of its strong DnaA boxes throughout most of the cell cycle, and binds to the weak boxes only at the start of initiation, when it is in its ATP-bound form. The binding and oligomerization of multiple DnaA molecules at *oriC* leads to melting of the DUE, generating the single-stranded DNA necessary for loading the replicative helicase DnaB.

Like all prokaryotic and eukaryotic initiators identified to date, DnaA has a C-terminal DNA-binding domain adjacent to the AAA+ core. The domain that contains the conserved AAA+ nucleotide-binding region also houses the oligomerization interface (Erzberger et al. 2002). Structural studies of the DnaA homologue in the thermophile *Aquifex aeolicus* suggest a mechanism for the oligomerization of DnaA.
When bound to ATP, the AAA+ module adopts an open conformation that permits the binding of the DnaA monomer to a second monomer (Erzberger et al. 2006). The second monomer docks an arginine finger into the active site of the first, forming an interaction with the γ-phosphate of ATP and stabilizing the oligomeric state. By triggering a switch between the monomeric and oligomeric states of DnaA, ATP binding modulates DnaA's ability to bind weaker affinity sites during initiation.

Interestingly, ATP-binding by *A. aeolicus* DnaA facilitates the formation of a right-handed helical filament (Fig. 2B). The formation of this non-planar spiral filament is mediated by an α-helical 'steric wedge' that protrudes away from the AAA+ core. Although the crystal structure of this filament lacks origin DNA, it is believed that a right-handed DNA wrap is formed around the DnaA core, which is consistent with the fact that the DNA binding-domain is located on the outside of the surface of the filament. These observations suggest a possible mechanism for DNA unwinding by DnaA through generating positive supercoils at the origin and focusing the compensatory negative superhelical stress into the DUE (Fig. 2C) (Erzberger et al. 2006).

ATP hydrolysis by DnaA is very slow, and is stimulated by the β-sliding clamp of polymerase III holoenzyme and the AAA+ regulatory protein Hda (Katayama et al. 1998; Kato and Katayama 2001). In a process called RIDA (regulatory inactivation of DnaA), ATP hydrolysis by DnaA inactivates the initiator, preventing reinitiation of recently replicated origins. This inactivation by RIDA depends on ongoing DNA synthesis, and is one of several independent mechanisms that ensure that DNA replication occurs exactly once per cell cycle (Kaguni 2006).

Saccharomyces cerevisiae

The larger size of eukaryotic genomes, relative to that of bacteria, presents a much greater challenge for the accurate and timely duplication of the genome. In eukaryotes, each chromosome is replicated from multiple origins within a single S phase. Coordination and regulation of the firing of these origins requires a highly complex network of *cis*- and *trans*-acting regulators. The first eukaryotic replicator and initiator were identified in the budding yeast *S. cereviasiae*, and since then, our understanding of the molecular details of replication initiation in *S. cerevisiae* has served as a paradigm for other eukaryotes.

S. cerevisiae origins of replication were originally identified in a screen for DNA sequences that confer autonomous replication upon plasmids in yeast cells (Hsiao and Carbon 1979; Stinchcomb et al. 1979). These sequences were termed autonomously replicating sequences (ARSs) and were subsequently shown to act as origins of replication in their normal chromosomal position (Brewer and Fangman 1987; Huberman 1987; Dubey et al. 1991). To map origins genome-wide, a variety of genomic methods were developed, including density transfer to map initial sites of replication, and chromatin immunoprecipitation coupled with microarrays (ChIP-chip) or high-throughput sequencing (ChIP-seq) to map ORC and Mcm2-7 binding sites (Raghuraman et al. 2001; Wyrick et al. 2001; Xu et al. 2006; Eaton et al. 2010). Together these studies have identified ~350 origins across the 16 *S. cerevisiae* chromsomes.

Several of these origins have been analyzed in detail, revealing a similar modular structure in each case (Fig. 3 and 4). Each origin contains a T-rich 11-bp element called the ARS consensus sequence (ACS) that is essential for origin function. The ACS constitutes half of the bipartite binding site for the yeast initiator ORC (Bell and Stillman 1992; Rao and Stillman 1995), and although it is essential for ORC binding and origin function, the ACS alone is not sufficient for either. Genome-wide approaches used to map origins demonstrated that the number of potential matches to the ACS in the genome (estimated to be 6,000-40,000 by various methods) far exceeds the number of ORC binding sites in the genome (between 250-400) (Xu et al. 2006; Eaton et al. 2010). Following the original characterization of the 11-bp consensus sequence, subsequent genomic and computational analyses identified a conserved 32-bp expanded ACS motif that is a more accurate indicator of a *bona fide* ARS and ORC binding site (Theis and Newlon 1994; Xu et al. 2006; Eaton et al. 2010).

In addition to the ACS, origins include several additional elements, termed B-elements, that contribute to origin function (Celniker et al. 1984). Mutating an individual B-element reduces but does not eliminate origin function. However, mutation of all B-elements within a single origin completely eliminates function. Unlike the ACS, B-elements exhibit very little or no sequence conservation among origins. Of the origins examined, B-elements are located on one side of the origin with respect to the asymmetric ACS. This region is referred to as the 'B-side' of the origin and is located 3' of the T-rich strand of the ACS.

Comparison of Eukaryotic Origin Structure

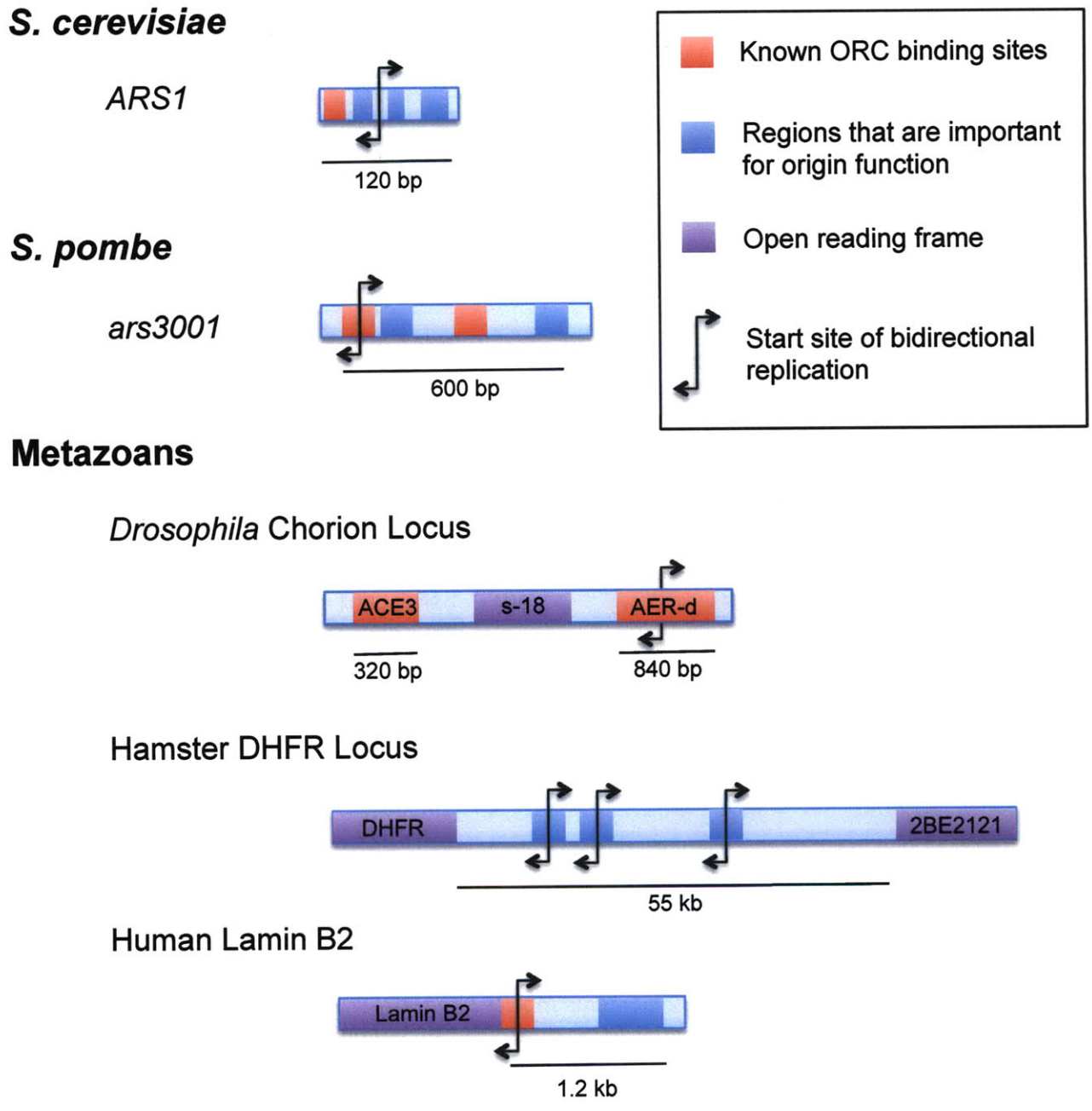


Figure 3. Comparison of replication origin structure in different eukaryotic organisms. Note that for the hamster DHFR locus, multiple replication initiation sites have been identified in the 55-kb zone; only three are depicted here.

Comparison of Origin Structure in *S. cerevisiae*

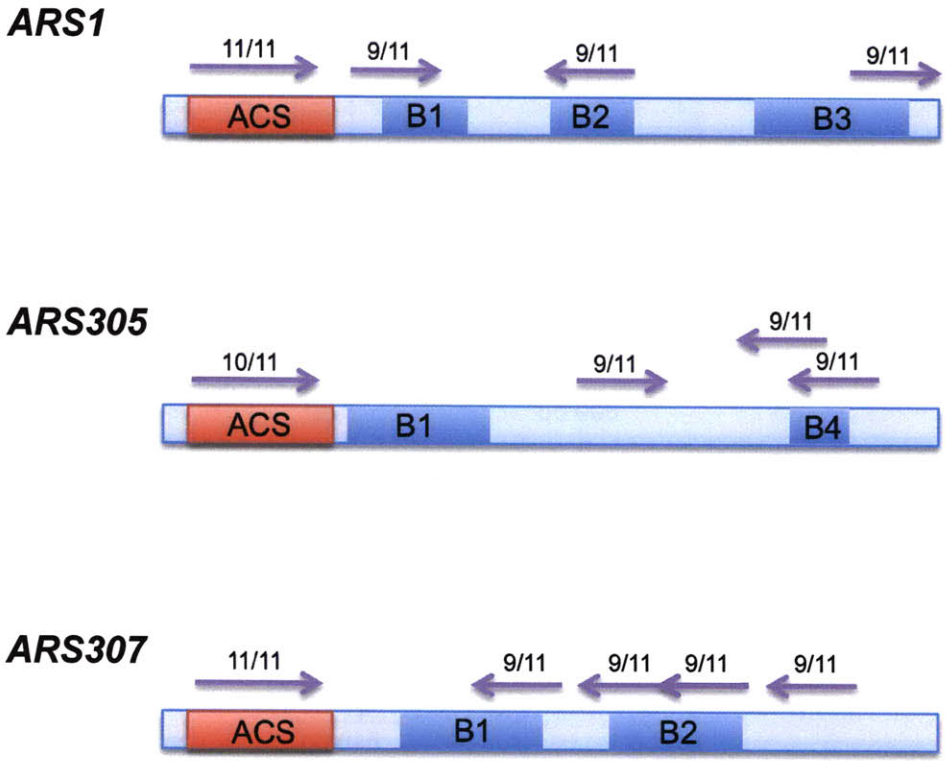


Figure 4. Comparison of structure of three *S. cerevisiae* origins of replication. Shown here are three representative origins – *ARS1*, *ARS305*, and *ARS307* – aligned at the essential ACS (red). Diagrams are drawn to scale. Note that the size of sequence elements and the spacing between elements vary among origins. B-elements (blue) are named to reflect their functional equivalency; i.e. the B2 elements in *ARS1* and *ARS307* are functionally interchangeable. Unlike the essential ACS, sequences are not conserved among functionally equivalent B-elements. Purple arrows indicate perfect and partial matches to the 11-bp ACS.

ARS1, the most well characterized origin in *S. cerevisiae*, contains three such B-elements in the region located downstream (3') of the T-rich strand of the ACS (Marahrens and Stillman 2001). These elements were named B1, B2, and B3. Importantly, following their discovery in *ARS1*, similar B-elements were identified in other origins and were named based on their ability to functionally substitute for the original *ARS1* elements (Rao et al. 1994; Huang and Kowalski 1996).

The B1 element comprises the second half of the ORC binding site (Rao et al. 1994; Rowley et al. 1995). However, some mutations of the B1 sequence in *ARS1* reduce origin function but have no effect on ORC binding, suggesting that B1 plays an additional role in the replicator other than binding ORC (Rao and Stillman 1995).

The B3 element is a binding site for the Abf1 transcription factor and prevents adjacent nucleosomes from encroaching into the origin DNA (Diffley and Stillman 1988; Venditti et al. 1994). B3 is present at *ARS1* but many origins lack a similar motif. Despite this, a common property of origins is the presence of a nucleosome free region (NFR) corresponding to the B-side of the origin. Subsequent studies have shown that ORC and the A-rich nature of the B-side of origins both contribute to this property.

B2 is the least understood of the B-elements. Nevertheless, functionally related elements have been identified in other origins (Rao et al. 1994; Theis and Newlon 1994). Although mutations in B2 result in a clear defect in Mcm2-7 loading *in vivo* (Zou and

Stillman 2000; Wilmes and Bell 2002), how the B2 element exerts this effect remains unclear.

Two hypotheses for how B2 functions have been proposed. Initially, several lines of evidence suggested that B2 functions as a DNA unwinding element (DUE). Of the origins examined, the B2 element tends to overlap with a broad region of helical instability at the origin. In *ARS1*, the B2 element is located immediately adjacent to the initiation site for bidirectional replication (Fig. 3) (Bielinsky and Gerbi 1998). In contrast to B1, single point mutations in B2 elements had no effect on origin function, suggesting that B2 plays a structural role rather than acting as a protein binding site (Natale et al. 1992; Rao et al. 1994). Also consistent with a DUE role, mutational analyses of *ARS1* and *ARS305* demonstrated that an exogenous, easily unwound sequence could functionally replace their B2 elements (Huang and Kowalski 1996; Lin and Kowalski 1997). However, subsequent studies confirmed that B2 is not a DUE. The B2 element of *ARS1* cannot be inverted, and replacement of B2 with a randomly generated helically unstable sequence did not always substitute for B2 function (Lin and Kowalski 1997; Wilmes and Bell 2002).

Alternatively, other evidence suggests that the B2 element could act as a binding site for a pre-RC component. What seems to be the most critical property of B2 in *ARS1* is how close in sequence it is to the ACS (*ARS1* B2 is a 9/11 match to the ACS in an inverse orientation relative to the ACS) (Fig. 3; Wilmes and Bell 2002). Although this does not rule out the possibility that B2 acts as a structural element (other than a DUE),

the simplest interpretation of all of these observations is that B2 could serve as a second binding site for ORC, resulting in two ORC complexes bound to the origin simultaneously. However, this is not the case *in vitro* (Bell and Stillman 1992) and it is unlikely that this is the case *in vivo*, as recent ChIP-seq mapping of ORC at origins showed that the peak of ORC enrichment is centered at the ACS and does not extend into the B-side region (Eaton et al. 2010).

Another possibility is that B2 acts as a binding site for other protein components of the pre-RC, such as the Mcm2-7 helicase or Cdc6. The Cdc6 protein is a close homolog of several of the ORC subunits, raising the possibility that it could interact with B2. Purified ORC and Cdc6 together form an extended DNase I footprint that covers the B2 element (Speck et al. 2005). Interestingly, overexpression of Cdc6p rescues the replication defect of B2- origin plasmids *in vivo*. Similarly, the Mcm2-7 complex is another candidate for binding the B2 element. As mentioned above, introduction of B2 mutant sequences into the chromosome at *ARS1* led to a decrease in MCM association as assayed by chromatin immunoprecipitation (Rao et al. 1995; Wilmes and Bell 2002). Despite these observations, neither Cdc6 nor the Mcm2-7 complex has been shown to bind the B2 sequence directly.

The very specific origin sequences in *S. cerevisiae* allowed for the first identification of a eukaryotic initiator, the origin recognition complex (ORC). ORC was first identified in *S. cerevisiae* as a heterohexameric complex that binds specifically to origin DNA *in vitro* in an ATP-dependent manner (Bell and Stillman 1992). All six

subunits of ORC, Orc1-6 (named in order of decreasing mass), are essential for viability and replication in yeast. Subsequent studies identified ORC homologues in numerous other eukaryotes, and ORC was shown to be essential for the initiation of DNA replication in all eukaryotes studied (Bell and Dutta 2002).

The interaction of ORC with DNA has been most extensively characterized in budding yeast. *In S. cerevisiae,* all six subunits of ORC remain bound to origins throughout the cell cycle (Cocker et al. 1996), whereas some evidence in mammalian cells suggests that Orc1 is released from chromatin and degraded in S phase (Diffley 2004). DNase I footprinting revealed a protected region over *ARS1* spanning roughly 30 bp, which includes the ACS and B1 element (Bell and Stillman 1992). Subsequent DNA modification-interference studies suggested that ORC interacts preferentially with the A-rich strand of the ACS at *ARS1* (Lee and Bell 1997). Like DnaA, ORC is responsible for the recruiting and loading the replicative helicase onto DNA, but in contrast to DnaA, there is no evidence that ORC melts the DNA prior to helicase loading.

Similar to that of DnaA, ORC's origin-binding activity is modulated by ATP binding. Three subunits of ORC, Orc1, Orc4, and Orc5, include consensus nucleotide binding motifs and are members of the AAA+ family. Two other subunits, Orc2 and Orc3, are more distantly related to the AAA+ family (Erzberger and Berger 2006). The smallest subunit, Orc6, is the only subunit not related to the AAA+ family, and is believed to be involved in the recruitment of the Cdt1/Mcm2-7 complex to the origin via a direct physical interaction with Cdt1 (Chen et al. 2007).

Mutational analyses of ORC subunits suggest that only two subunits, Orc1 and Orc5, bind ATP. ATP binding, but not hydrolysis, by Orc1 is essential for ORC's sequence-specific DNA binding activity *in vitro*, and mutations in the ATP binding site of ORC1 are lethal (Klemm et al. 1997). Consistent with these observations, DNA binding inhibits ATP hydrolysis by ORC in a sequence-specific manner. Together, these observations suggest that when bound to origin DNA throughout the cell cycle *in vivo*, ORC is primarily in its ATP-bound form, until ATP hydrolysis by ORC is necessary for pre-RC assembly in G1 (see below).

Although a single molecule of ORC is required to load the helicase at each origin, the six ORC subunits are arranged within the complex in a manner that is believed to be analogous to oligomerized DnaA. Structural studies of the closely related *D. melanogaster* ORC revealed that all five AAA+ domains of ORC contain the α-helical insert that is believed to mediate formation of the right-handed helical filament in DnaA (Clarey et al. 2006). Consistent with this observation, the structure of the heteropentameric core of DmORC (Orc1-Orc5) is very similar to that of a helical ATP-DnaA pentamer. Like DnaA and other oligomeric AAA+ complexes, ORC's ATP binding sites are located at the interface between subunits. EM reconstructions of *S. cerevisiae* ORC suggest a subunit arrangement in which Orc1 interacts directly with Orc4, which is consistent with the observation that an arginine finger residue in Orc4 is required to stimulate ATP hydrolysis by Orc1 (Bowers et al. 2004). Based on both

studies of *S. cerevisiae* and *D. melanogaster* ORC, it has been proposed that Cdc6 binds to an unoccupied AAA+ surface at the base of the ORC filament.

As previously mentioned, many potential ACS matches exist in the *S. cerevisiae* genome, yet ORC is still capable of binding a small subset of these ACS matches with remarkable specificity. This suggests that other physical characteristics, beyond DNA sequence alone, define the locations of ORC-bound origins. In particular, nucleosome positioning is a critical determinant of origin function. Nucleosome mapping at *ARS1* demonstrated that a ~180 bp region overlapping the core origin elements was nucleosome-free, and that ORC binding is required to position nucleosomes adjacent to the origin (Lipford and Bell 2001). Consistent with these observations, genome-wide mapping of nucleosomes at all potential ACS matches revealed a similar pattern of nucleosome occupancy surrounding 219 ORC-bound ACS sites. At these *bona fide* origins, the ACS is located asymmetrically within a 125-bp nuclesome-free region (NFR), which includes and extends 3' of the T-rich strand of the ACS (Eaton et al. 2010). For all of these origins, an A-rich island (the B-side) is located adjacent to the T-rich ACS, suggesting that this pattern acts as the major sequence determinant of the NFR. This pattern is unique to ORC-bound ACS sites and is not observed at ACS sites that are not bound by ORC, although weak and symmetric nucleosome depletion is detected at these sites. In addition, ORC is necessary and sufficient for positioning of the origin-adjacent nucleosomes. Together these data suggest that the nucleosomal pattern at origins is established in two steps: 1) a nucleosome-free region is determined at the level of DNA sequence, giving ORC access to an unobstructed, nucleosome-free ACS, and 2) once

bound to the origin, ORC positions the nucleosomes immediately flanking the origin, establishing the precise nucleosomal pattern observed at all potential origins.

Schizosaccharomyces pombe

Studies in the fission yeast *S. pombe* and in metazoans suggested that although the original replicon paradigm generally holds true for budding yeast, the model is less applicable for other eukaryotic organisms. As in *S. cerevisiae*, *S. pombe* ARS elements were originally identified through similar plasmid transformation studies (reviewed in Clyne and Kelly 1997). However, in contrast to *S. cerevisiae* origins, *S. pombe* origins are much larger in size (0.5-1 kb) and no conserved or essential sequences elements analogous to the budding yeast ACS have been identified (Clyne and Kelly 1995; Dubey et al. 1996). Instead, *S. pombe* origins are composed of AT-rich stretches (Fig. 3). Roughly 300-500 potential origins with this extended AT-rich structure have been mapped throughout the genome (Segurado et al. 2003; Feng et al. 2006; Heichinger et al. 2006; Hayashi et al. 2007).

The structure of *S. pombe* ORC explains its distinct DNA binding specificity. SpORC4, contains an N-terminal extension that is not found in *S. cerevisiae* and metazoan ORC homologs. This extension contains nine AT-hook DNA binding motifs that have previously been shown to bind to the minor groove of short AT tracts in a non-sequence-specific manner (Reeves and Beckerbauer 2001). Consistent with this domain mediating its DNA binding specificity, *S. pombe* ORC preferentially binds AT-rich DNA,

rather than a specific sequence and this binding is dependent on the AT-hook domain of SpORC4. Comparison of the binding affinity of SpORC to an *S. pombe* origin depends largely on its length and AT content (Chuang et al. 2002). However, recent studies suggest that other structural features, such as negative supercoiling, contribute positively to SpORC's affinity to origin DNA (Houchens et al. 2008).

Metazoans

Studies of replication origins in *Xenopus laevis* and *Drosophila melanogaster* embryonic systems as well as mammalian cells suggest that specific DNA sequences are not required for pre-RC assembly and origin function (Remus and Diffley 2009). For this reason, identification and mapping of replicators in higher eukaryotes has lagged behind their single cell counterparts. Early studies in *Xenopus* egg extracts demonstrated that any DNA sequence, including those from completely exogenous sources, could replicate efficiently (Cvetic and Walter 2005). Subsequent two-dimensional gel analysis revealed that in the cell-free system, replication initiation events occurred randomly throughout the DNA sequences tested (Hyrien and Mechali 1992; Mahbubani et al. 1992). However, the high concentration of ORC in *Xenopus* egg extracts could be driving non-specific ORC DNA binding. In diluted *Xenopus* egg extracts, ORC preferentially binds asymmetric AT-rich sequences -- for example, poly (dAdT/dTdA) rather than poly (dA/dT) -- suggesting that some level of sequence specificity is masked by high ORC concentrations in undiluted extracts (Kong et al. 2003). Interestingly, there is evidence that there are

significantly fewer active origins later in *Xenopus* development when ORC levels are lower (Hyrien et al. 1995).

One unique example of sequence-specific replication initiation in metazoans comes from the scheme used to amplify the chorion gene loci in *Drosophila* follicle cells. The most well characterized follicle cell amplicon located on the third chromosome contains two sequence elements, *ACE3* and oriβ, that are necessary and sufficient for amplification. Insertion of these two sequence elements into an exogenous genomic location is sufficient to direct amplification (Lu et al. 2001). DmORC binds to both *ACE3* and oriβ (which is part of the larger *AER-d* element), and although replication can initiate from multiple sites within this replication zone, the majority of replication initiates from oriβ, suggesting that *ACE3* is able to activate initiation from a distal, downstream origin (Austin et al. 1999; Lu et al. 2001).

Investigators have been unable to isolate an autonomously replicating sequence in mammalian cells. In an attempt to identify replicators in human cells using a plasmid transformation screen, it was found that any sufficiently large fragment of DNA could confer replication competence to a plasmid (Heinzel et al. 1991). However, replication in human cells does exhibit some, albeit low, sequence specificity, as large DNA fragments from *E. coli* inserted into plasmids replicated less efficiently in human cells. Despite this low level of sequence specificity, replication does localize to some specific loci in higher eukaryotes, and a small number of origins have been indentified in mammalian cells. Most of these origins fall under the category of initiation zones, in which replication can

initiate from one of many potential sizes within a large stretch of DNA. The most well characterized initiation zone is the Chinese hamster dihydrofolate reductase (DHFR) locus, in which replication initiates from many potential sites spanning a 55 kb intergenic region between the DHFR and 2BE2121 loci (Fig. 3) (Hamlin et al. 2010). These initiation sites are highly inefficient and redundant, and mutation of these sites had no effect on the overall efficiency of initiation of the DHFR locus. Interestingly, deletion of the 3' end of the DHFR gene completely abrogated replication initiation at this zone, and a subsequent study revealed that the integrity of the promoter of the DHFR gene was absolutely necessary for downstream initiation events (Mesner et al. 2003; Saha et al. 2004). These data suggest a role for transcription in regulating the initiation activity of origins.

In an alternative class of mammalian replication origins, initiation is restricted to a much smaller, discrete site. One such example is the human lamin B2 origin, a 1.2-kb region in which a discrete start site of bidirectional DNA synthesis has been mapped to a single nucleotide (Fig. 3) (Abdurashidova et al. 2000). In a manner analogous to that of *S. cerevisiae* origins, *in vivo* DMS footprinting analysis demonstrated an extended footprint at the lamin B2 locus in G1 that shrank upon entry into S phase (Abdurashidova et al. 1998). This pattern of protection was most likely caused by the cell cycle mediated binding of the pre-RC to the lamin B2 origin (Abdurashidova et al. 2003).

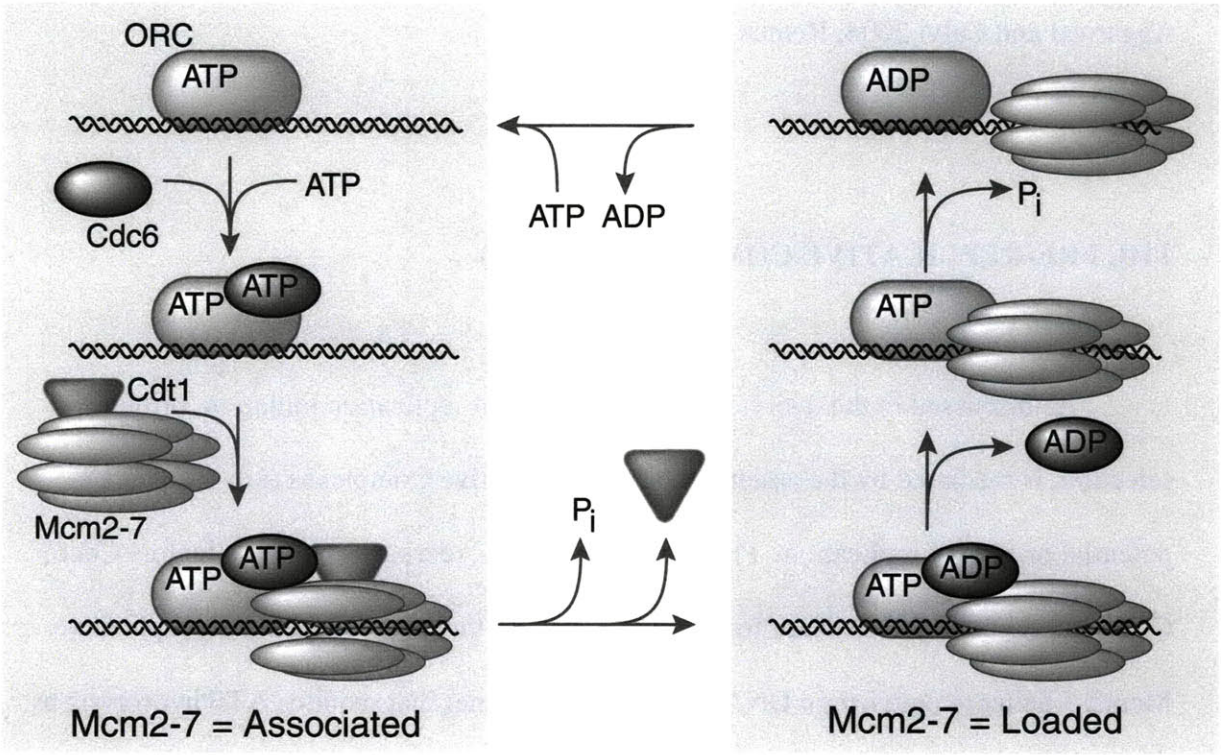
Although ORC is conserved in metazoans, and is clearly essential for replication, the molecular basis by which ORC selects origins is poorly understood. ORC in these

systems does not appear to exhibit any sequence-specific DNA binding activity. However, ORC binding appears to be directed by transcriptional activity, histone methylation and acetylation, AT-richness, and DNA supercoiling (Vashee et al. 2003; Aggarwal and Calvi 2004; Remus et al. 2004).

THE PRE-REPLICATIVE COMPLEX (PRE-RC)

As discussed in the overview, the first phase of replication initiation, origin selection, is mediated by the assembly of pre-Replicative Complexes (pre-RCs) at all potential origins of replication. Prior to S phase, ORC recruits replication factors Cdc6, Cdt1, and the Mcm2-7 helicase to origins. Together, these proteins load the replicative Mcm2-7 helicase onto origin DNA in a stepwise manner that requires ATP hydrolysis by ORC and Cdc6 (Perkins and Diffley 1998; Bowers et al. 2004; Randell et al. 2006). It is thought that ORC and Cdc6 act as a coordinated molecular machine in a manner analogous to that of sliding clamp loaders (Seybert et al. 2002; Johnson and O'Donnell 2003; Randell et al. 2006).

Recent studies in *S. cerevisiae* have partially elucidated the stepwise mechanism of pre-RC assembly at origins (Fig. 5). First, Mcm2-7 and Cdt1, which are believed to be in a complex prior to pre-RC assembly, are recruited to the origin by ORC and Cdc6 independently of ATP hydrolysis (Bowers et al. 2004; Remus et al. 2009). Second, ATP hydrolysis by Cdc6 stimulates the stable loading of the Mcm2-7 helicase onto DNA



(Randell et al. 2006)

Figure 5. A model for the stepwise mechanism of helicase loading in *S. cerevisiae*.
ATP-bound ORC binds to origin DNA and recruits Cdc6. Cdc6 binds ATP in an ORC-dependent manner. ORC and Cdc6 recruit a complex of Mcm2-7/Cdt1 to the origin; at this point the helicase is associated but not topologically linked around origin DNA. ATP hydrolysis by Cdc6 leads to the loading of the Mcm2-7 helicase and the release of Cdt1. ATP hydrolysis by ORC completes the cycle, allowing for additional rounds of Mcm2-7 loading at the same origin.

(Randell et al. 2006). Finally, ATP hydrolysis by ORC resets the cycle, allowing for additional Mcm2-7 helicases to be loaded per origin. These findings in budding yeast have provided the basic framework for our understanding of how helicases are loaded at origins in eukaryotes. However, many of the molecular details of pre-RC assembly are poorly understood. Little is known about the architecture of the final complex and the intermediates leading up to it. How the protein components of the pre-RC interact with one another and the origin DNA as well as how these interactions change during pre-RC formation is largely unclear.

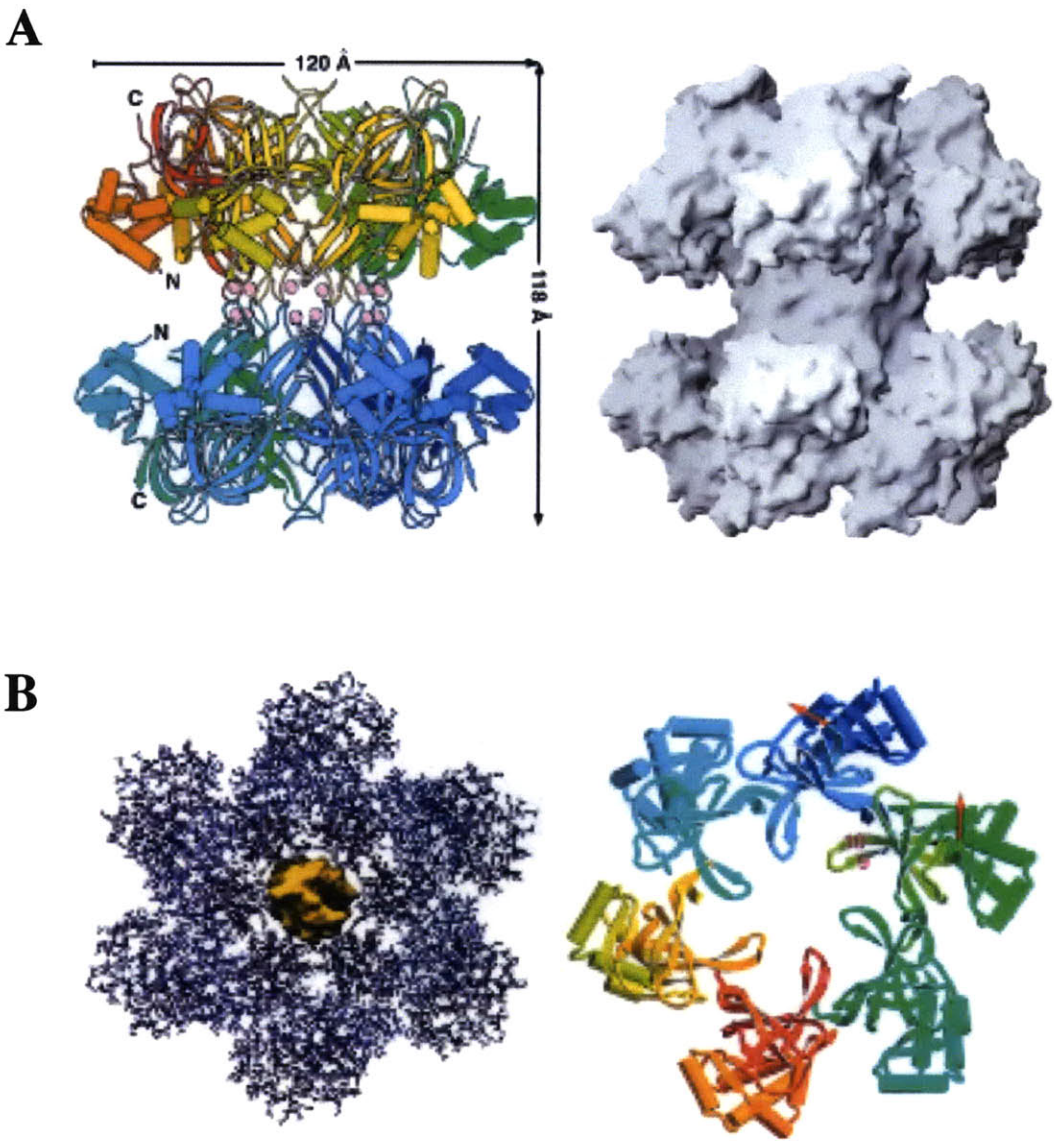
Prior to discussing the stepwise mechanism of helicase loading, I will provide background information on the replicative Mcm2-7 helicase, with a focus on our current structural and functional understanding of the complex. I will then review the three steps involved in pre-RC assembly, as summarized above: helicase association, helicase loading, and resetting of the cycle. The bulk of our mechanistic knowledge of pre-RC assembly comes from studies in *S. cerevisiae*, although relevant details gathered from other model organisms will be discussed as well. Finally, this section will conclude with a discussion of the architecture of the pre-RC bound to origin DNA.

Mcm2-7: The Replicative Helicase

The Mcm2-7 (Mini-Chromosome Maintenance) complex is the replicative helicase in eukaryotes. This heterohexameric complex contains one copy each of six related polypeptides, each of which is essential for replication (Bochman and Schwacha 2009). The sequence similarity among the Mcm subunits is most prominent within the

shared C-terminal AAA+ ATPase domain; the N-terminal domains are more distinct in terms of sequence but are structurally very similar. Crystallographic studies of the N-terminal domain of a related archaeal homohexameric MCM complex in the *Methanobacterium thermoautotrophicum* revealed an oligonucleotide/oligosaccharide-binding (OB)-like fold that is characteristic of many ssDNA binding proteins (Fletcher et al. 2003). In addition, most Mcm proteins contain a zinc finger motif near their N-termini that is thought to stabilize folding of their N-terminal domains. In *S. cerevisiae,* both Mcm2 and Mcm5 contain the zinc finger motif, and Mcm4, 6, and 7 each has a loose variation of the motif; Mcm3 completely lacks this motif (Tye 1999). Mutation of the Mcm2 zinc finger motif in *S. cerevisiae* leads to lethality, and mutation of the Mcm5 zinc finger motif leads to temperature-sensitivity (Yan et al. 1991).

EM studies of both the *S. cerevisiae* Mcm2-7 complex and the related *M. thermoautotrophicum* MCM complex indicate a head-to-head double hexameric structure with a central channel large enough for dsDNA to pass through (Fig. 6 and 7; Fletcher et al., 2003; Remus et al., 2009). In *S. cerevisiae*, double hexamers are only observed once loaded onto DNA; unbound Mcm2-7 complexes exist exclusively as single hexamers (Sato et al. 2000; Remus et al. 2009). Biochemical studies demonstrated that following pre-RC assembly, the Mcm2-7 complex encircles double-stranded DNA and, as an inactive helicase, translocates along dsDNA in a non-directional, energy-independent manner *in vitro* (Remus et al. 2009). Together, these observations provide important insights into how the Mcm2-7 helicase, when activated, unwinds DNA. Several models have been proposed, many of which are based on current mechanistic knowledge of



(Fletcher et al. 2003)

Figure 6. Structure of the *M. thermoautotrophicum* N-MCM double hexamer.
(A) A 3.1 Å crystal structure of the N-terminal half of mtMCM (N-mtMCM). Side view (*left*) and surface representation (*right*) of the double-hexamer structure. (B) Central channel of the N-mtMCM double hexamer. *Left* – dsDNA (yellow) modeled in the central channel at its narrowest point (23 Å). *Right* – View of the six β-hairpin fingers at the narrowest point of the central channel. These β-hairpin fingers are believed to be important for gripping the DNA passing through the narrowest point in the channel.

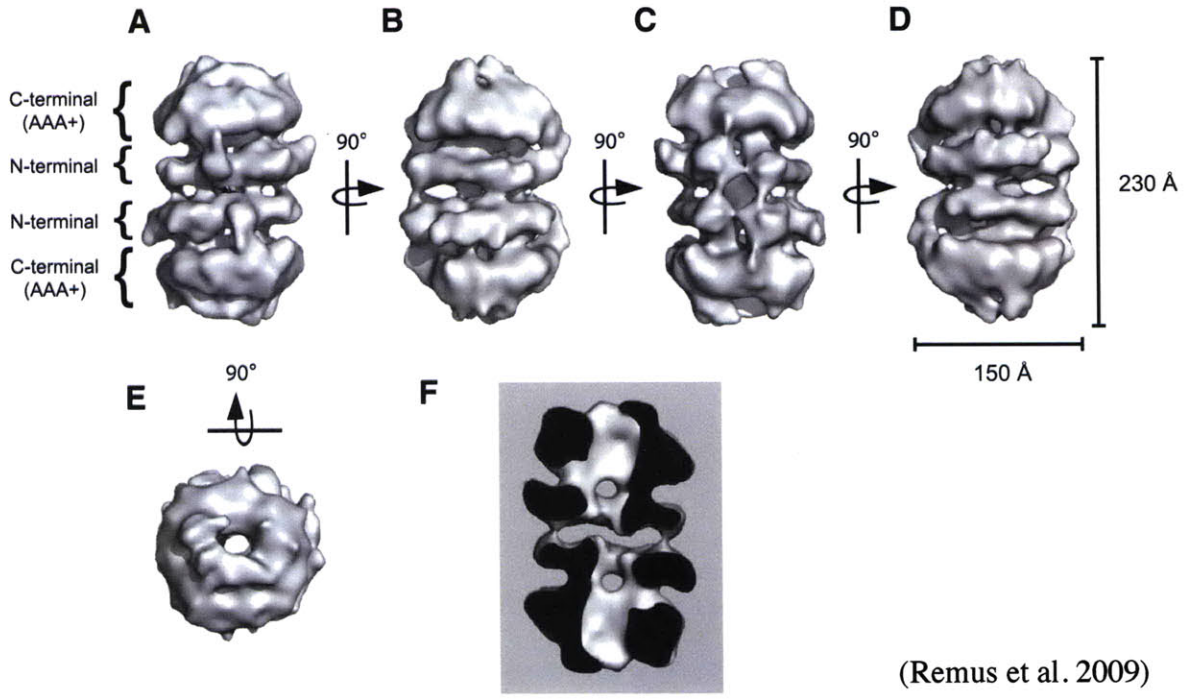


Figure 7. Three-dimensional reconstruction of the *S. cerevisiae* Mcm2-7 double hexamer. ***(A-E)*** Surface representations of a 3D reconstruction of the Mcm2-7 double hexamer at a 30 Å resolution. The data set used to calculate the reconstruction was obtained by transmission electron microscopy of negatively stained purified Mcm2-7 complexes bound to DNA. ***(B)*** Cut-open side view of the double hexamer.

bacterial and archaeal helicases. However, the mechanism by which Mcm2-7 complexes unwind DNA remains controversial. Two major issues that have yet to be resolved are (1) whether the active Mcm2-7 helicase encircles one DNA strand or both strands to unwind DNA, and (2) whether the helicase in its active form operates as a double hexamer or a single hexamer. The answers to these two questions would have significant implications for the different models proposed (Bochman and Schwacha 2009; Takara and Bell 2009).

The Mcm2-7 complex is the only component of the pre-RC that is required for both replication initiation and elongation (Labib et al. 2000). During G1 phase, the Mcm2-7 complex is loaded onto origin DNA in its inactive form. Upon entry into S phase, an increase in CDK and DDK activity leads to phosphorylation of Mcm2-7 by DDK as well as the recruitment of additional replication factors, including Cdc45 and the GINS complex. These events lead to the activation of DNA unwinding by Mcm2-7. In *Drosophila* embryo extracts, Mcm2-7 copurifies with Cdc45 and GINS in a complex (termed the CMG complex) that exhibits helicase activity (Moyer et al. 2006). Biochemical analysis of this complex *in vitro* demonstrated that the association of Cdc45 and GINS with the Mcm2-7 complex significantly increased its ATPase activity as well as its affinity for both forked substrates and ssDNA (Ilves et al. 2010). Recent biochemical studies in *S. cerevisiae* demonstrated that purified recombinant Mcm2-7 unwinds DNA *in vitro* (Bochman and Schwacha 2008). Although it has not yet been determined whether *S. cerevisiae* Mcm2-7 complex acts alone in or in a functional complex with Cdc45 and GINS *in vivo*, it is interesting to note that Mcm2-7, Cdc45, and

GINS all have been shown to travel with the replication fork *in vivo* (Aparicio et al. 1997; Calzada et al. 2005; Kanemaki and Labib 2006). This suggests that Cdc45 and GINS are not only required for initiation but their functions are also required at the replication fork.

The Stepwise Mechanism of Pre-RC Assembly

Step 1: Helicase association with origin DNA

Assembly of the pre-Replicative Complex (pre-RC) begins at the M- to G1-phase transition, when CDK levels are low enough to permit assembly. In *S. cerevisiae*, ORC remains bound to all potential origins throughout the cell cycle (Aparicio et al. 1997; Liang and Stillman 1997). As discussed earlier, ORC binds to the roughly 325 origins in an ATP-dependent manner. As cells enter G1 phase, ORC recruits Cdc6 to the origin. Because ORC on its own has no detectable affinity for Mcm2-7 or Cdt1, Cdc6 must first bind to ORC to recruit the remaining two pre-RC components to the origin.

Like ORC, Cdc6 belongs to the AAA+ family of ATPases, and exhibits a strong amino acid sequence similarity with Orc1 (Bell et al. 1995). However, Cdc6 on its own does not bind or hydrolyze ATP (Randell et al. 2006). Once bound to ORC, Cdc6 binds ATP in an ORC- and DNA-dependent manner (Randell et al. 2006). ATP-bound Cdc6 then facilitates the recruitment of Cdt1 and the Mcm2-7 helicase to the origin. It is unclear whether Cdc6 recruits Cdt1/Mcm2-7 through a direct interaction or by inducing a conformational change in ORC that facilitates Mcm2-7/Cdt1 binding, or both. This initial association of the Cdt1/Mcm2-7 complex occurs independently of ATP hydrolysis

by ORC or Cdc6. Indeed, this complex can only be observed when ATP hydrolysis is prevented.

Biochemical and structural analyses of pre-RC assembly suggest that the recruitment of the Mcm2-7 complex and Cdt1 to the origin is coupled, and that the two exist in a complex prior to recruitment. Cdt1 and Mcm2-7 associate with the ORC- and Cdc6-bound origin with similar kinetics *in vitro* (Randell et al. 2006). When purified recombinant Cdt1 and Mcm2-7 are added separately to an *in vitro* pre-RC assembly assay, Mcm2-7 loading does not occur, whereas allowing Cdt1 and Mcm2-7 to pre-bind prior to addition to a reconstituted assembly assay results in Mcm2-7 loading (Tsakraklides and Bell 2010). Recent EM studies revealed that prior to loading onto DNA, Mcm2-7 single hexamers exist in complex with a single Cdt1 molecule, and the loaded form of this complex contains two Mcm2-7 hexamers in a head-to-head orientation (Remus et al. 2009). This suggests that Cdt1 may be involved in recruitment of the Mcm2-7 hexamer to the pre-RC, possibly acting as a physical bridge between Orc6 and Mcm2-7, as discussed in the previous section. Alternatively, Cdt1 could also play a role in maintaining the Mcm2-7 helicase in a state that is competent for loading. At this intermediate stage in pre-RC assembly, Mcm2-7 still depends on the rest of the pre-RC components for association with the origin.

Step 2: Helicase loading

Following helicase recruitment to the origin, ATP hydrolysis by Cdc6 is required for the first round of stable loading of the Mcm2-7 helicase onto DNA as well as the

release of Cdt1 from the complex (Randell et al. 2006). Once loaded, Mcm2-7 complexes are topologically linked around double-stranded DNA. For many years, it was only assumed but not confirmed experimentally that the Mcm2-7 helicase encircled DNA. This original assumption was based on observations that after assembly onto DNA, Mcm2-7 complexes are highly resistant to salt extraction *in vivo* and *in vitro* (Donovan et al. 1997; Rowles and Blow 1997; Edwards et al. 2002; Bowers et al. 2004). Recent EM analysis suggested that Mcm2-7 helicases encircled double stranded DNA, which was confirmed biochemically by the observation that Mcm2-7 complexes loaded onto circular DNA had a significantly greater half life (>60 minutes) than Mcm2-7 complexes loaded onto linear DNA (10 minutes) (Remus et al. 2009). These data suggest that not only do Mcm2-7 complexes encircle and bind DNA in the central channel, but that they are also mobile, capable of sliding on DNA independent of ATP-hydrolysis and presumably sliding off the ends of linear DNA templates.

It is unclear how many Mcm2-7 complexes are initially recruited and loaded onto a single origin following Cdc6 ATP hydrolysis. Following reconstituted pre-RC assembly on circular DNA, single Mcm2-7 hexamers were not observed on the DNA by electron microscopy. Instead, the majority of DNA molecules contained a single dimer of Mcm2-7 hexamers oriented in a head-to-head fashion, and in less than 5% of the molecules observed, two or three double hexamers were observed (Remus et al. 2009). This suggests that stable Mcm2-7 loading requires dimer formation. However, because these EM studies did not analyze intermediate steps in pre-RC assembly, it remains possible that hexamers are loaded one at a time.

Step 3: Resetting the cycle

ATP hydrolysis by ORC catalyzes a subsequent step in pre-RC formation. Genetic studies of an *ORC1* allele that can bind but not hydrolyze ATP demonstrate that this event is essential and important for pre-RC formation (Klemm et al. 1997). Analysis of this mutant *in vitro* suggests that the mutation allows an intermediate level of Mcm2-7 loading that increases when ORC ATP hydrolysis occurs (Bowers et al. 2004). This observation suggests that ORC ATP hydrolysis is required for multiple rounds of Mcm2-7 loading. The molecular details of how ORC ATP hydrolysis facilitates additional loading remain unclear. One hypothesis is that ATP hydrolysis by ORC causes the dissociation of the remaining pre-RC components, allowing for a second round of helicase loading. An alternative hypothesis is that the remaining pre-RC components dissociate from the complex on their own following the first round of helicase loading, and that ATP hydrolysis by ORC resets the conformation of ORC to allow for the subsequent recruitment of Cdc6, Cdt1 and Mcm2-7.

It has been observed that an excess of Mcm2-7 complexes are loaded in almost every organism studied. In *S. cerevisiae*, it is estimated that 2-4 Mcm2-7 hexamers are loaded per origin. However, in metazoans, it has been estimated that the number of DNA-bound Mcm2-7 complexes exceeds the number of initiation events by a factor of 10-40 (Mahbubani et al. 1997; Walter and Newport 1997). Immunofluorescence studies in *Xenopus* egg extracts suggest that the majority of Mcm2-7 complexes do not preferentially co-localize with sites of DNA replication (Romanowski et al. 1996). The reason for this excess of Mcm2-7 complexes in higher eukaryotes is unclear, although

recent evidence suggests that excess Mcm2-7 complexes license dormant origins under conditions of replicative stress in *Xenopus* (Woodward et al. 2006). Despite this excess, only a subset of these complexes is activated for initiation, and DNA replication is still highly efficient even when the number of Mcm2-7 complexes is severely reduced, suggesting that this excess is not essential (Lucas et al. 2000; Edwards et al. 2002).

Architecture of the Pre-RC

Genomic footprinting studies of several origins in *S. cerevisiae* cells revealed two distinct patterns of protection of these sites that correlated with specific phases of the cell cycle. In G1, the pre-RC protects a region that extends from the ACS to the B2 element in *ARS1*. This G1-dependent footprint is Cdc6-dependent, suggesting that protection is indeed due to pre-RC assembly at those origins (Santocanale and Diffley 1996; Perkins and Diffley 1998). Unexpectedly, this footprint closely resembles the DNase I footprint of purified ORC and Cdc6 bound to *ARS1 in vivo*, suggesting that this extended footprint in G1 is not dependent on Mcm2-7 loading (Speck et al. 2005). Throughout the remainder of the cell cycle, the post-replicative footprint at these origins *in vivo* most closely resembles that of purified ORC bound to origins *in vitro* (Bell and Stillman 1992; Cocker et al. 1996). The simplest interpretation of these data is that the footprint extension in G1 is due to Cdc6 binding to origin DNA in an ORC-dependent manner. However, these data cannot exclude the possibility that ORC undergoes a conformational change upon pre-RC assembly, leading to an extended ORC footprint in G1.

Of all the components of the pre-RC, ORC is the only one whose DNA binding activity has been extensively characterized. At *ARS1*, ORC interacts specifically with the ACS and B1 in an ATP-dependent manner, and interacts predominantly with the A-rich strand of the ACS (Lee and Bell 1997). Interestingly, DNase I footprinting of purified ORC bound to *ARS1* revealed three distinct sites of enhanced DNase I cleavage – one within the B1 element, and two between the B1 and B2 elements – that are 10 bp apart from one another, suggesting that DNA might wrap around ORC during binding (Bell and Stillman 1992). Consistent with this observation, single particle EM studies of *D. melanogaster* ORC bound to origin DNA results in the shortening of the DNA by ~130 bp (Clarey et al. 2008). Additionally, ORC-dependent DNA bending has been observed at the *S. cerevisiae ARS1* origin *in vitro* (Lee and Bell 1997).

Crosslinking studies of *S. cerevisiae* ORC bound to several yeast origins have provided clues into the spatial arrangement of ORC subunits when bound to DNA (Fig. 8) (Lee and Bell 1997). Different ORC subunits were distributed over different regions of the origin, with a significant clustering of subunits over the ACS. Interestingly, crosslinking of three subunits, Orc2, Orc4, and Orc6, was observed in the downstream region beyond the ORC binding site (as defined by ORC's DNase I footprint), extending into the region between the B2 and B3 elements. EM reconstructions of *S. cerevisiae* ORC provided yet another model for the architecture of ORC (albeit not bound to DNA) that was consistent with previous observations (Chen et al. 2008). In the EM reconstruction, the termini of the five largest subunits of ORC were clustered to each other in opposite ends of the ORC complex. It is important to note that the C-termini of

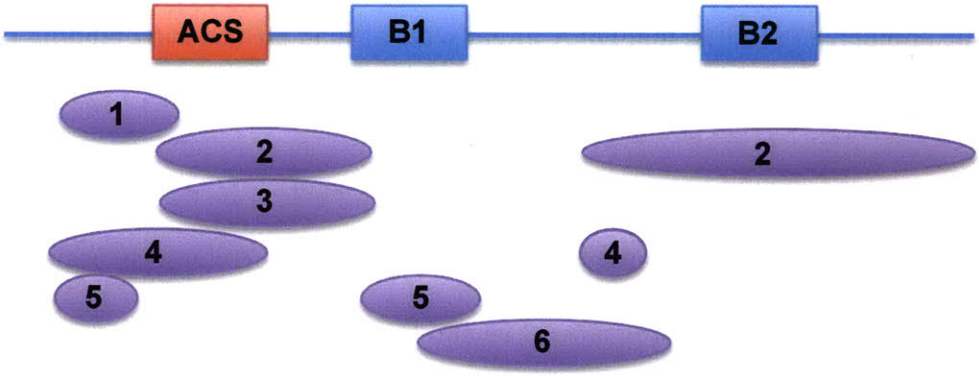


Figure 8. Summary of ORC subunit arrangement at *ARS1*.
Summary of data obtained from 4'AZPB crosslinking of ORC subunits to *ARS1* (Lee and Bell 1997). The majority of ORC subunits crosslinked to the core binding site (ACS and B1). Orc2, Orc4, and Orc6 were crosslinked to the flanking region containing the B2 element. (Lee and Bell 1997)

Orc1-5 each contain a DNA-binding winged helix domain (WHD) next to their AAA+ folds (Speck et al. 2005). Orc6, the only non-AAA+ related subunit in the complex, has no affinity for DNA and is not necessary for ORC binding (Chen et al. 2007; Lee and Bell 1997), although Orc6 crosslinks with the region between B1 and B2 *in vitro*.

Whether and how Cdc6 interacts with origin DNA is less clear. Like Orc1-5, Cdc6 also contains a WHD near its C-terminus. *In vitro,* recombinant *S. cerevisiae* Cdc6 was found to bind double-stranded DNA indiscriminately (Feng et al. 2000), but a later study did not detect any specific DNA binding activity by Cdc6 at origins (Speck et al., 2005). Consistent with the genomic footprint at *ARS1*, purified ORC and Cdc6 together produced an extended footprint that was strikingly similar to the pre-RC genomic footprint observed *in vivo* (Speck et al. 2005). Interestingly, a point mutation in B1 (B1 A838G), which was shown to have no effect on ORC binding to *ARS1* but caused a defect in plasmid stability, showed a considerable reduction in the formation of the ORC-Cdc6 extended footprint. Based on a combination of biochemical and structural data from both *S. cerevisiae* and *D. melanogaster*, it has been proposed that the binding of Cdc6 to ORC produces a ring-like structure analogous to other hexameric AAA+ machines (Speck et al. 2005; Clarey et al. 2008).

The spatial arrangement of the remaining components of the pre-RC, Cdt1 and the Mcm2-7 complex at the origin is unknown. Cdt1 is not thought to interact with origin DNA directly and associates only transiently with the origin DNA (Randell et al. 2006; Chen et al. 2007). The Mcm2-7 complex has been shown to encircle DNA, but because

Mcm2-7 binding does not produce a detectable DNase I footprint, it is unclear where the complex binds within the origin sequence. It has been proposed that B2 acts as a dedicated binding/loading site for the Mcm2-7 complex, since mutation of the B2 sequence *in vivo* leads to a decrease in Mcm2-7 association with the origin by chromatin immunoprecipitation (Zou and Stillman 2000; Wilmes and Bell 2002). However, by this result alone, it is equally likely that B2 is a binding site for another pre-RC component such as Cdc6.

Despite the lack of mapping data, several lines of evidence are consistent with the possibility that Mcm2-7 complexes preferentially occupy the B-side region of origins. In *S. cerevisiae* cells, the nucleosome-free region of origins encompasses the B-side region downstream of the site of ORC binding (Eaton et al. 2010). This unobstructed nucleosome-free space would be ideal for Mcm2-7 binding, but current Mcm2-7 mapping studies are not sufficient to resolve this issue. However, the fact that the Mcm2-7 complex cannot produce a detectable DNase I footprint *in vivo* or *in vitro* suggests that the complex is relatively dynamic, and that even if Mcm2-7 loading or binding is restricted to a discrete region, it is possible that the complex does not remain strictly associated with one specific sequence following loading.

THESIS SUMMARY

In this thesis, I describe the DNA requirements for the stepwise process of pre-RC assembly, as well as identify the step at which B2 is important in the mechanism of helicase loading. First, I determined that Mcm2-7 complexes are predominantly loaded on the B-side of origins throughout the yeast genome, coinciding with the nucleosome free region (NFR) of those origins. Second, I demonstrated that the amount of B-side DNA adjacent to the ORC binding site is a critical determinant of how many Mcm2-7 hexamers can be loaded at the origin. However, this is not simply a matter of having sufficient space near ORC to accommodate Mcm2-7 binding. The sequence of the region adjacent to ORC, specifically the B2 element, is also important for facilitating Mcm2-7 loading at the origin. I found that mutating B2 element causes a defect in Mcm2-7 loading *in vitro* under specific conditions: 1) when nucleosomes are assembled onto the DNA prior to pre-RC assembly, and 2) when the amount of ORC-adjacent B-side DNA is limited. In addition, the B2 element is important at the step of helicase loading but not the initial association to the origin. Together, these data suggest a role for B2 in facilitating the stable loading of Mcm2-7 hexamers when the accessibility of ORC-adjacent DNA is restricted.

REFERENCES

Abdurashidova, G., Danailov, M.B., Ochem, A., Triolo, G., Djeliova, V., Radulescu, S., Vindigni, A., Riva, S., and Falaschi, A. 2003. Localization of proteins bound to a replication origin of human DNA along the cell cycle. *Embo J* **22**(16): 4294-4303.

Abdurashidova, G., Deganuto, M., Klima, R., Riva, S., Biamonti, G., Giacca, M., and Falaschi, A. 2000. Start sites of bidirectional DNA synthesis at the human lamin B2 origin. *Science* **287**(5460): 2023-2026.

Abdurashidova, G., Riva, S., Biamonti, G., Giacca, M., and Falaschi, A. 1998. Cell cycle modulation of protein-DNA interactions at a human replication origin. *Embo J* **17**(10): 2961-2969.

Aggarwal, B.D. and Calvi, B.R. 2004. Chromatin regulates origin activity in Drosophila follicle cells. *Nature* **430**(6997): 372-376.

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**(1): 59-69.

Austin, R.J., Orr-Weaver, T.L., and Bell, S.P. 1999. Drosophila ORC specifically binds to ACE3, an origin of DNA replication control element. *Genes Dev* **13**(20): 2639-2649.

Bell, S.P. and Dutta, A. 2002. DNA replication in eukaryotic cells. *Annu Rev Biochem* **71**: 333-374.

Bell, S.P., Mitchell, J., Leber, J., Kobayashi, R., and Stillman, B. 1995. The multidomain structure of Orc1p reveals similarity to regulators of DNA replication and transcriptional silencing. *Cell* **83**(4): 563-568.

Bell, S.P. and Stillman, B. 1992. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* **357**(6374): 128-134.

Bielinsky, A.K. and Gerbi, S.A. 1998. Discrete start sites for DNA synthesis in the yeast ARS1 origin. *Science* **279**(5347): 95-98.

Blow, J.J. and Dutta, A. 2005. Preventing re-replication of chromosomal DNA. *Nat Rev Mol Cell Biol* **6**(6): 476-86.

Bochman, M.L. and Schwacha, A. 2008. The Mcm2-7 complex has in vitro helicase activity. *Mol Cell* **31**(2): 287-293.

Bochman, M. L. and Schwacha, A. 2009. The Mcm complex: unwinding the mechanism of a replicative helicase. *Microbiol Mol Biol Rev* **73**(4): 652-683.

Bowers, J.L., Randell, J.C., 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**(6): 967-978.

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**(5): 743-755.

Brewer, B.J. and Fangman, W.L. 1987. The localization of replication origins on ARS plasmids in S. cerevisiae. *Cell* **51**(3): 463-471.

Calzada, A., Hodgson, B., Kanemaki, M., Bueno, A., and Labib, K. 2005. Molecular anatomy and regulation of a stable replisome at a paused eukaryotic DNA replication fork. *Genes Dev* **19**(16): 1905-1919.

Celniker, S.E., Sweder, K., Srienc, F., Bailey, J.E., and Campbell, J.L. 1984. Deletion mutations affecting autonomously replicating sequence ARS1 of Saccharomyces cerevisiae. *Mol Cell Biol* **4**(11): 2455-2466.

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**(22): 2897-2907.

Chen, Z., Speck, C., Wendel, P., Tang, C., Stillman, B., and Li, H. 2008. The architecture of the DNA replication origin recognition complex in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **105**(30): 10326-10331.

Chuang, R.Y., Chretien, L., Dai, J., and Kelly, T.J. 2002. Purification and characterization of the Schizosaccharomyces pombe origin recognition complex: interaction with origin DNA and Cdc18 protein. *J Biol Chem* **277**(19): 16920-16927.

Clarey, M.G., Botchan, M., and Nogales, E. 2008. Single particle EM studies of the Drosophila melanogaster origin recognition complex and evidence for DNA wrapping. *J Struct Biol* **164**(3): 241-249.

Clarey, M.G., Erzberger, J.P., Grob, P., Leschziner, A.E., Berger, J.M., Nogales, E., and Botchan, M. 2006. Nucleotide-dependent conformational changes in the DnaA-like core of the origin recognition complex. *Nat Struct Mol Biol* **13**(8): 684-690.

Clyne, R.K. and Kelly, T.J. 1995. Genetic analysis of an ARS element from the fission yeast Schizosaccharomyces pombe. *Embo J* **14**(24): 6348-6357.

Cocker, J.H., Piatti, S., Santocanale, C., Nasmyth, K., and Diffley, J.F. 1996. An essential role for the Cdc6 protein in forming the pre-replicative complexes of budding yeast. *Nature* **379**(6561): 180-182.

Cvetic, C. and Walter, J.C. 2005. Eukaryotic origins of DNA replication: could you please be more specific? *Semin Cell Dev Biol* **16**(3): 343-353.

Diffley, J.F. 2004. Regulation of early events in chromosome replication. *Curr Biol* **14**(18): R778-786.

Diffley, J.F. and Stillman, B. 1988. Purification of a yeast protein that binds to origins of DNA replication and a transcriptional silencer. *Proc Natl Acad Sci U S A* **85**(7): 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**(11): 5611-5616.

Drlica, K. and Rouviere-Yaniv, J. 1987. Histonelike proteins of bacteria. *Microbiol Rev* **51**(3): 301-319.

Dubey, D.D., Davis, L.R., Greenfeder, S.A., Ong, L.Y., Zhu, J.G., Broach, J.R., Newlon, C.S., and Huberman, J.A. 1991. Evidence suggesting that the ARS elements associated with silencers of the yeast mating-type locus HML do not function as chromosomal DNA replication origins. *Mol Cell Biol* **11**(10): 5346-5355.

Dubey, D.D., Kim, S.M., Todorov, I.T., and Huberman, J.A. 1996. Large, complex modular structure of a fission yeast DNA replication origin. *Curr Biol* **6**(4): 467-473.

Eaton, M.L., Galani, K., Kang, S., Bell, S.P., and MacAlpine, D.M. 2010. Conserved nucleosome positioning defines replication origins. *Genes Dev* **24**(8): 748-753.

Edwards, M.C., Tutter, A.V., Cvetic, C., Gilbert, C.H., Prokhorova, T.A., and Walter, J.C. 2002. MCM2-7 complexes bind chromatin in a distributed pattern surrounding the origin recognition complex in Xenopus egg extracts. *J Biol Chem* **277**(36): 33049-33057.

Erzberger, J.P. and Berger, J.M. 2006. Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu Rev Biophys Biomol Struct* **35**: 93-114.

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**(8): 676-683.

Feng, L., Wang, B., Driscoll, B., and Jong, A. 2000. Identification and characterization of Saccharomyces cerevisiae Cdc6 DNA-binding properties. *Mol Biol Cell* **11**(5): 1673-1685.

Feng, W., Collingwood, D., Boeck, M.E., Fox, L.A., Alvino, G.M., Fangman, W.L., Raghuraman, M.K., and Brewer, B.J. 2006. Genomic mapping of single-stranded DNA in hydroxyurea-challenged yeasts identifies origins of replication. *Nat Cell Biol* **8**(2): 148-155.

Fletcher, R.J., Bishop, B.E., Leon, R.P., Sclafani, R.A., Ogata, C.M., and Chen, X.S. 2003. The structure and function of MCM from archaeal M. Thermoautotrophicum. *Nat Struct Biol* **10**(3): 160-167.

Hamlin, J.L., Mesner, L.D., and Dijkwel, P.A. 2010. A winding road to origin discovery. *Chromosome Res* **18**(1): 45-61.

Hayashi, M., Katou, Y., Itoh, T., Tazumi, A., Yamada, Y., Takahashi, T., Nakagawa, T., Shirahige, K., and Masukata, H. 2007. Genome-wide localization of pre-RC sites and identification of replication origins in fission yeast. *Embo J* **26**(5): 1327-1339.

Heichinger, C., Penkett, C.J., Bahler, J., and Nurse, P. 2006. Genome-wide characterization of fission yeast DNA replication origins. *Embo J* **25**(21): 5171-5179.

Heinzel, S.S., Krysan, P.J., Tran, C.T., and Calos, M.P. 1991. Autonomous DNA replication in human cells is affected by the size and the source of the DNA. *Mol Cell Biol* **11**(4): 2263-2272.

Houchens, C.R., Lu, W., Chuang, R.Y., Frattini, M.G., Fuller, A., Simancek, P., and Kelly, T.J. 2008. Multiple mechanisms contribute to Schizosaccharomyces pombe origin recognition complex-DNA interactions. *J Biol Chem* **283**(44): 30216-30224.

Hsiao, C.L. and Carbon, J. 1979. High-frequency transformation of yeast by plasmids containing the cloned yeast ARG4 gene. *Proc Natl Acad Sci U S A* **76**(8): 3829-3833.

Huang, R.Y. and Kowalski, D. 1996. Multiple DNA elements in ARS305 determine replication origin activity in a yeast chromosome. *Nucleic Acids Res* **24**(5): 816-823.

Huberman, J.A. 1987. Eukaryotic DNA replication: a complex picture partially clarified. *Cell* **48**(1): 7-8.

Hyrien, O., Maric, C., and Mechali, M. 1995. Transition in specification of embryonic metazoan DNA replication origins. *Science* **270**(5238): 994-997.

Hyrien, O. and Mechali, M. 1992. Plasmid replication in Xenopus eggs and egg extracts: a 2D gel electrophoretic analysis. *Nucleic Acids Res* **20**(7): 1463-1469.

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**(2): 247-258.

Jacob, F. and Brenner, S. 1963. [On the regulation of DNA synthesis in bacteria: the hypothesis of the replicon.]. *C R Hebd Seances Acad Sci* **256**: 298-300.

Johnson, A. and O'Donnell, M. 2003. Ordered ATP hydrolysis in the gamma complex clamp loader AAA+ machine. *J Biol Chem* **278**(16): 14406-14413.

Kaguni, J.M. 2006. DnaA: controlling the initiation of bacterial DNA replication and more. *Annu Rev Microbiol* **60**: 351-375.

Kanemaki, M. and Labib, K. 2006. Distinct roles for Sld3 and GINS during establishment and progression of eukaryotic DNA replication forks. *Embo J* **25**(8): 1753-1763.

Katayama, T., Ozaki, S., Keyamura, K., and Fujimitsu, K. 1998. Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and oriC. *Nat Rev Microbiol* **8**(3): 163-170.

Kato, J. and Katayama, T. 2001. Hda, a novel DnaA-related protein, regulates the replication cycle in Escherichia coli. *Embo J* **20**(15): 4253-4262.

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**(4): 493-502.

Kong, D., Coleman, T.R., and DePamphilis, M.L. 2003. Xenopus origin recognition complex (ORC) initiates DNA replication preferentially at sequences targeted by Schizosaccharomyces pombe ORC. *Embo J* **22**(13): 3441-3450.

Kowalski, D. and Eddy, M.J. 1989. The DNA unwinding element: a novel, cis-acting component that facilitates opening of the Escherichia coli replication origin. *Embo J* **8**(13): 4335-4344.

Labib, K., Tercero, J.A., and Diffley, J.F. 2000. Uninterrupted MCM2-7 function required for DNA replication fork progression. *Science* **288**(5471): 1643-1647.

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**(12): 7159-7168.

Liang, C. and Stillman, B. 1997. Persistent initiation of DNA replication and chromatin-bound MCM proteins during the cell cycle in cdc6 mutants. *Genes Dev* **11**(24): 3375-3386.

Lin, S. and Kowalski, D. 1997. Functional equivalency and diversity of cis-acting elements among yeast replication origins. *Mol Cell Biol* **17**(9): 5473-5484.

Lipford, J.R. and Bell, S.P. 2001. Nucleosomes positioned by ORC facilitate the initiation of DNA replication. *Mol Cell* **7**(1): 21-30.

Lu, L., Zhang, H., and Tower, J. 2001. Functionally distinct, sequence-specific replicator and origin elements are required for Drosophila chorion gene amplification. *Genes Dev* **15**(2): 134-146.

Lucas, I., Chevrier-Miller, M., Sogo, J.M., and Hyrien, O. 2000. Mechanisms ensuring rapid and complete DNA replication despite random initiation in Xenopus early embryos. *J Mol Biol* **296**(3): 769-786.

Marahrens, Y. and Stillman, B. 1992. A yeast chromosomal origin of DNA replication defined by multiple functional elements. *Science* **255**(5046): 817-823.

Mahbubani, H.M., Chong, J.P., Chevalier, S., Thommes, P., and Blow, J.J. 1997. Cell cycle regulation of the replication licensing system: involvement of a Cdk-dependent inhibitor. *J Cell Biol* **136**(1): 125-135.

Mahbubani, H.M., Paull, T., Elder, J.K., and Blow, J.J. 1992. DNA replication initiates at multiple sites on plasmid DNA in Xenopus egg extracts. *Nucleic Acids Res* **20**(7): 1457-1462.

Mesner, L.D., Li, X., Dijkwel, P.A., and Hamlin, J.L. 2003. The dihydrofolate reductase origin of replication does not contain any nonredundant genetic elements required for origin activity. *Mol Cell Biol* **23**(3): 804-814.

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

Moyer, S.E., Lewis, P.W., and Botchan, M.R. 2006. Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc Natl Acad Sci U S A* **103**(27): 10236-10241.

Natale, D.A., Schubert, A.E., and Kowalski, D. 1992. DNA helical stability accounts for mutational defects in a yeast replication origin. *Proc Natl Acad Sci U S A* **89**(7): 2654-2658.

Perkins, G. and Diffley, J.F. 1998. Nucleotide-dependent prereplicative complex assembly by Cdc6p, a homolog of eukaryotic and prokaryotic clamp-loaders. *Mol Cell* **2**(1): 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* **294**(5540): 115-121.

Randell, J.C., Bowers, J.L., Rodriguez, H.K., and Bell, S.P. 2006. Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase. *Mol Cell* **21**(1): 29-39.

Rao, H., Marahrens, Y., and Stillman, B. 1994. Functional conservation of multiple elements in yeast chromosomal replicators. *Mol Cell Biol* **14**(11): 7643-7651.

Rao, H. and Stillman, B. 1995. The origin recognition complex interacts with a bipartite DNA binding site within yeast replicators. *Proc Natl Acad Sci U S A* **92**(6): 2224-2228.

Reeves, R. and Beckerbauer, L. 2001. HMGI/Y proteins: flexible regulators of transcription and chromatin structure. *Biochim Biophys Acta* **1519**(1-2): 13-29.

Remus, D., Beall, E.L., and Botchan, M.R. 2004. DNA topology, not DNA sequence, is a critical determinant for Drosophila ORC-DNA binding. *Embo J* **23**(4): 897-907.

Remus, D., Beuron, F., Tolun, G., Griffith, J.D., Morris, E.P., and Diffley, J.F. 2009. Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell* **139**(4): 719-730.

Remus, D. and Diffley, J.F. 2009. Eukaryotic DNA replication control: lock and load, then fire. *Curr Opin Cell Biol* **21**(6): 771-777.

Romanowski, P., Madine, M.A., Rowles, A., Blow, J.J., and Laskey, R.A. 1996. The Xenopus origin recognition complex is essential for DNA replication and MCM binding to chromatin. *Curr Biol* **6**(11): 1416-1425.

Rowles, A. and Blow, J.J. 1997. Chromatin proteins involved in the initiation of DNA replication. *Curr Opin Genet Dev* **7**(2): 152-157.

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**(11): 2631-2641.

Saha, S., Shan, Y., Mesner, L.D., and Hamlin, J.L. 2004. The promoter of the Chinese hamster ovary dihydrofolate reductase gene regulates the activity of the local origin and helps define its boundaries. *Genes Dev* **18**(4): 397-410.

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**(23): 6671-6679.

Sato, M., Gotow, T., You, Z., Komamura-Kohno, Y., Uchiyama, Y., Yabuta, N., Nojima, H., and Ishimi, Y. 2000. Electron microscopic observation and single-stranded DNA binding activity of the Mcm4,6,7 complex. *J Mol Biol* **300**(3): 421-431.

Segurado, M., de Luis, A., and Antequera, F. 2003. Genome-wide distribution of DNA replication origins at A+T-rich islands in Schizosaccharomyces pombe. *EMBO Rep* **4**(11): 1048-1053.

Seybert, A., Scott, D.J., Scaife, S., Singleton, M.R., and Wigley, D.B. 2002. Biochemical characterisation of the clamp/clamp loader proteins from the euryarchaeon Archaeoglobus fulgidus. *Nucleic Acids Res* **30**(20): 4329-4338.

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.*

Speck, C. and Messer, W. 2001. Mechanism of origin unwinding: sequential binding of DnaA to double- and single-stranded DNA. *Embo J* **20**(6): 1469-1476.

Speck, C., Weigel, C., and Messer, W. 1999. ATP- and ADP-dnaA protein, a molecular switch in gene regulation. *Embo J* **18**(21): 6169-6176.

Stinchcomb, D.T., Struhl, K., and Davis, R.W. 1979. Isolation and characterisation of a yeast chromosomal replicator. *Nature* **282**(5734): 39-43.

Takara, T.J. and Bell, S.P. 2009. Putting two heads together to unwind DNA. *Cell* **139**(4): 652-654.

Theis, J.F. and Newlon, C.S. 1994. Domain B of ARS307 contains two functional elements and contributes to chromosomal replication origin function. *Mol Cell Biol* **14**(11): 7652-7659.

Tsakraklides, V. and Bell, S.P. 2010. Dynamics of pre-replicative complex assembly. *J Biol Chem* **285**(13): 9437-9443.

Tye, B.K. 1999. MCM proteins in DNA replication. *Annu Rev Biochem* **68**: 649-686.

Van Houten, J.V. and Newlon, C.S. 1990. Mutational analysis of the consensus sequence of a replication origin from yeast chromosome III. *Mol Cell Biol* **10**(8): 3917-3925.

Vashee, S., Cvetic, C., Lu, W., Simancek, P., Kelly, T.J., and Walter, J.C. 2003. Sequence-independent DNA binding and replication initiation by the human origin recognition complex. *Genes Dev* **17**(15): 1894-1908.

Venditti, P., Costanzo, G., Negri, R., and Camilloni, G. 1994. ABFI contributes to the chromatin organization of Saccharomyces cerevisiae ARS1 B-domain. *Biochim Biophys Acta* **1219**(3): 677-689.

Walter, J. and Newport, J.W. 1997. Regulation of replicon size in Xenopus egg extracts. *Science* **275**(5302): 993-995.

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**(1): 101-106.

Woodward, A.M., Gohler, T., Luciani, M.G., Oehlmann, M., Ge, X., Gartner, A., Jackson, D.A., and Blow, J.J. 2006. Excess Mcm2-7 license dormant origins of replication that can be used under conditions of replicative stress. *J Cell Biol* **173**(5): 673-683.

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**(5550): 2357-2360.

Xu, W., Aparicio, J.G., Aparicio, O.M., and Tavare, 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**: 276.

Yan, H., Gibson, S., and Tye, B.K. 1991. Mcm2 and Mcm3, two proteins important for ARS activity, are related in structure and function. *Genes Dev* **5**(6): 944-957.

Yasuda, S. and Hirota, Y. 1977. Cloning and mapping of the replication origin of Escherichia coli. *Proc Natl Acad Sci U S A* **74**(12): 5458-5462.

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**(9): 3086-3096.

Chapter II

DNA length and sequence requirements for helicase loading at *S. cerevisiae* origins of replication

This work was done in collaboration with Sukhyun Kang and Matthew L. Eaton. WML performed the chromatin immunoprecipitation experiment in Figure 1, and MLE performed the data analysis. SK performed the pre-RC assembly assays on chromatinized DNA shown in Figures 6 and 7. SK performed the nucleosome mapping experiment in Figure 8, and MLE performed the data analysis.

SUMMARY

S. cerevisiae origins of replication include multiple functional DNA elements: an essential ARS (autonomously replicating sequence) consensus sequence (ACS) and multiple, adjacent B-elements that collectively are also essential for origin function. In addition to recruiting the initiator ORC, these sequences contribute to subsequent events in replication initiation by mechanisms that are yet to be understood. In this study, we show that the Mcm2-7 replicative helicase is preferentially loaded within the nucleosome-free region (which includes the B-elements) downstream of the ORC binding site at origins *in vivo*. Reducing the extent of these flanking sequences *in vitro* does not alter the initial recruitment of ORC or Mcm2-7 to the origin, but reduces or eliminates Mcm2-7 loading. Restricting the accessibility of ORC-adjacent DNA, either by incorporation of the template into nucleosomes or by truncation, revealed a function for the B2 element during Mcm2-7 loading. We propose that the B2 element acts in multiple ways to facilitate helicase loading under conditions that restrict the accessibility of ORC-adjacent DNA sequence.

INTRODUCTION

In the budding yeast *S. cerevisiae*, chromosomal replication initiates at ~250-350 origins of replication throughout the genome (Raghuraman et al. 2001; Wyrick et al. 2001; Xu et al. 2006; Eaton et al. 2010). Detailed studies of a subset of these origins of replication revealed a common modular structure that is reminiscent of eukaryotic transcriptional promoters. These origins are on average 100-150-bp in size, and contain a conserved, T-rich 11-bp element called the ARS (autonomously replicating sequence) consensus sequence (ACS) that is essential for origin function (Celniker et al. 1984; Van Houten and Newlon 1990). In addition to the ACS, each origin contains 2-3 additional sequences called B-elements that are located 3' of the T-rich strand of the ACS. These B-elements are individually important and collectively essential for origin function (Marahrens and Stillman 1992). Unlike the ACS, the B-elements are not conserved in sequence although substitution experiments have shown that these elements are functionally interchangeable among different origins (Rao et al. 1994; Theis and Newlon 1994; Huang and Kowalski 1996; Lin and Kowalski 1997).

ARS1, the most well characterized *S. cerevisiae* origin, contains three B elements: B1, B2, and B3 (Marahrens and Stillman 1992). The B1 element and the ACS comprise the bipartite recognition sequence for the origin recognition complex (ORC), the eukaryotic initiator (Bell and Stillman 1992; Rao and Stillman 1995; Rowley et al. 1995). Mutational analysis of the B1 element suggests that it plays another role in replication independent of its function in binding ORC; a single point mutation in B1 that led to a plasmid loss defect *in vivo* had no effect on ORC binding *in vitro* (Rao and Stillman

1995). However, the exact nature of the additional function of B1 is unclear. The B3 element is a binding site for the Abf1 transcription factor and prevents adjacent nucleosomes from encroaching into the origin DNA (Diffley and Stillman 1988; Venditti et al. 1994), but is not found at most origins. In comparison to the B1 and B3 elements, the role of the B2 element is less clear, although functionally related elements have been identified in other origins (Rao et al. 1994; Theis and Newlon 1994; Huang and Kowalski 1996). *In vivo* studies have implicated a role for the B2 element in pre-Replicative Complex (pre-RC) formation at origins of replication (Zou and Stillman 2000; Wilmes and Bell 2002). However, *in vitro* studies of this event have revealed no role for this element.

Pre-RC formation is the process by which the Mcm2-7 replicative helicase is loaded onto origin DNA in eukaryotes, marking all potential origins of replication prior to their activation in S phase. The assembly of the pre-RC at origins occurs in a stepwise manner. ORC binds to origin sequences throughout the cell cycle (Aparicio et al. 1997). Starting in late mitosis and in G1 phase, ORC recruits Cdc6 to origins, and ORC and Cdc6 coordinately recruit a complex of Cdt1 and the Mcm2-7 replicative helicase. At this stage, all of the pre-RC proteins are associated with the origin DNA but the Mcm2-7 helicase is not yet independently engaged with the origin DNA. Finally, in a process that requires ATP hydrolysis by Cdc6, the Mcm2-7 helicase is loaded onto the origin DNA (Randell et al. 2006). The resulting loaded helicase encircles both strands of the origin DNA and is inactive for DNA unwinding (Remus et al. 2009). The helicase remains inactive until cells enter S phase, in which an increase in CDK and DDK activity leads to

the recruitment of additional replication factors, activating the helicase to unwind origin DNA.

Although the functions of *trans*-acting factors in helicase loading have been elucidated, the function of origin DNA sequence and chromatin structure at origins remain poorly understood. First, the location of the Mcm2-7 complexes after origin loading has not been defined, and whether specific DNA sequences act as Mcm2-7 binding sites remains unclear. Second, although B2 functions in pre-RC assembly *in vivo* (Zou et al. 2000; Wilmes and Bell 2002), whether B2 functions in the initial recruitment, the subsequent loading of the Mcm2-7 complex, or both has not been determined. Finally, how chromatin structure influences these events remains unclear.

In this study, we investigated the DNA requirements for the stepwise process of pre-RC assembly. We found that Mcm2-7 complexes are predominantly loaded at the B-side region adjacent to ORC binding throughout the yeast genome, coinciding with the nucleosome-free region (NFR) associated with each origin (Eaton et al. 2010). Reducing the extent of these flanking sequences *in vitro* does not alter the initial recruitment of ORC or the Mcm2-7 helicase to the origin, but reduces or completely eliminates Mcm2-7 loading. Finally, we assessed the DNA sequence requirements for helicase loading *in vitro* and show that the B2 element is important for facilitating Mcm2-7 loading in a chromatin context. These findings provide important insights into how origin DNA sequences establish the specific protein/DNA and nucleosomal architecture necessary to load the replicative helicase at origins of replication.

RESULTS

Mcm2-7 complexes are preferentially loaded downstream but not upstream of the ORC binding site

To examine the location of Mcm2-7 loading relative to known origin sequence elements *in vivo*, we analyzed Mcm2-7 binding sites throughout the genome prior to replication initiation. We arrested yeast cells in G1 phase with α-factor and used chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-Seq) to map sites of Mcm2-7 chromatin association with high resolution. Comparison of Mcm2-7 ChIP-Seq data to our previously published ORC ChIP-Seq data revealed extensive overlap of Mcm2-7 and ORC binding sites on all 16 chromosomes (Eaton et al. 2010) (Fig. 1A, black histogram; Fig. 1B), consistent with observations from previous genome-wide studies (Wyrick et al. 2001; Xu et al. 2006).

To determine the location of Mcm2-7 and ORC enrichment relative to known origin sequences, we analyzed the 219 ORC-bound ACS sites that were identified in a previous study (Eaton et al. 2010). We created a composite profile of ORC and Mcm2-7 binding at these ORC-ACS sites by aligning the sequences of the 219 ORC-ACS sites using the T-rich strand of the ACS, and plotting the average enrichment at these sites relative to the center of a 33-bp consensus ORC binding site as defined in Eaton et al. (position 0) (Fig. 1C). For future reference, sequences 5' of the T-rich strand of the ACS will be referred to as "upstream" and sequences 3' of the T-rich strand will be referred to

Figure 1

A

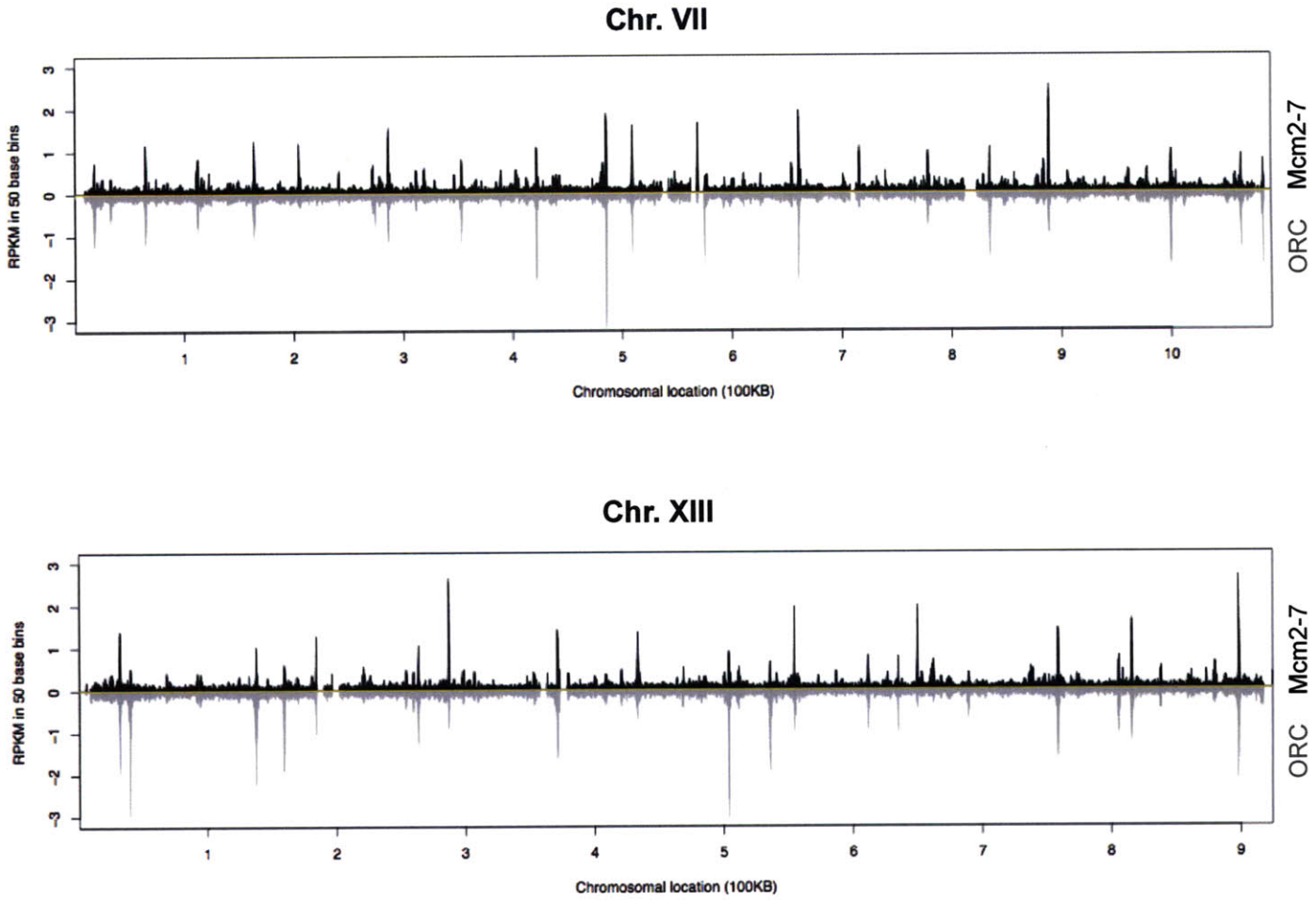


Figure 1. ChIP-Seq mapping of ORC and Mcm2-7 occupancy at origins.

(A) Mcm2-7 (*up*) and ORC (*down*) occupancy on two representative chromosomes, Chr. VII and XIII.

Figure 1

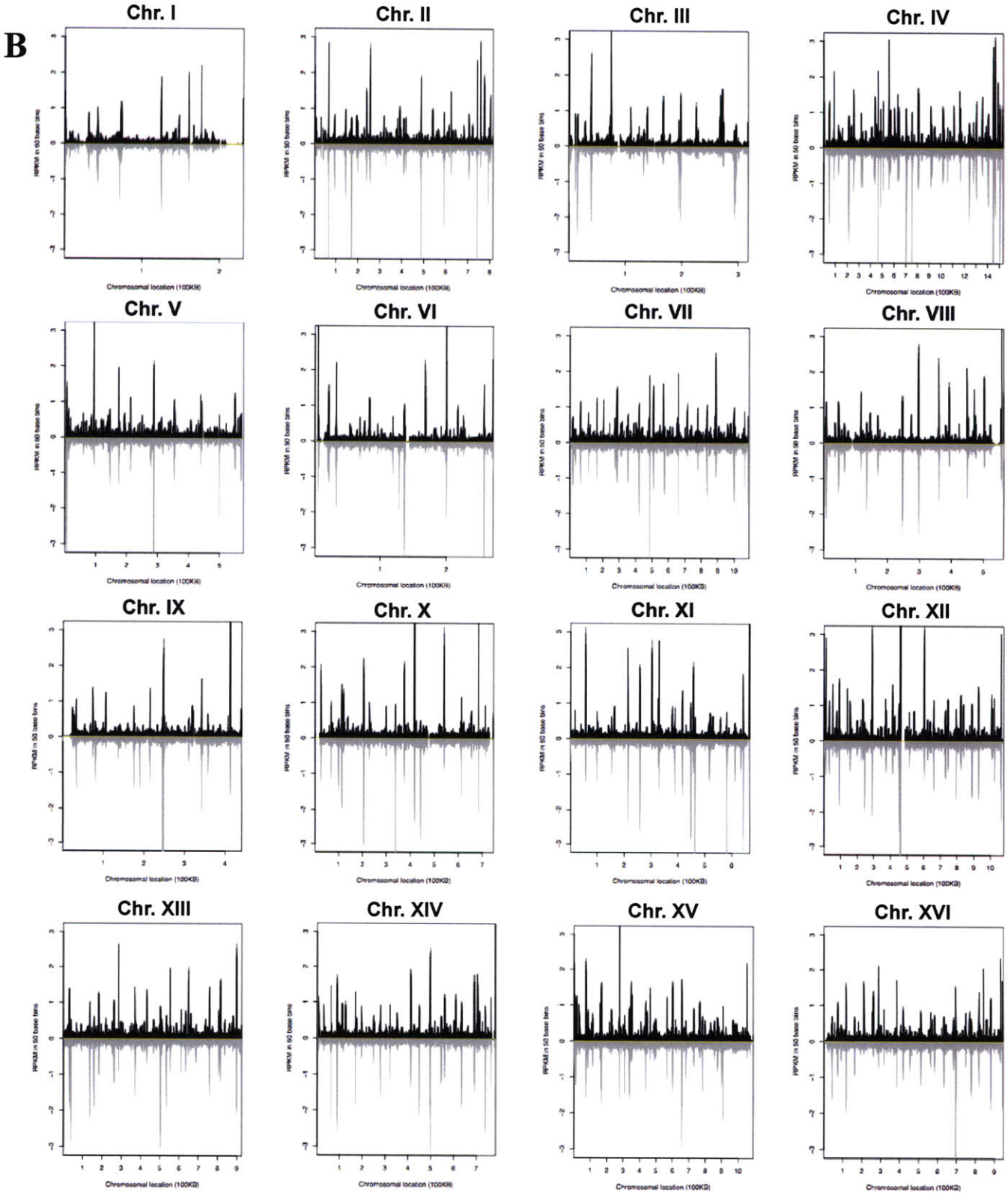


Figure 1. ChIP-Seq mapping of ORC and Mcm2-7 occupancy at origins.

(B) ChIP-Seq mapping of Mcm2-7 *(up)* and ORC *(down)* binding sites across *S. cerevisiae* chromosomes I-XVI.

Figure 1

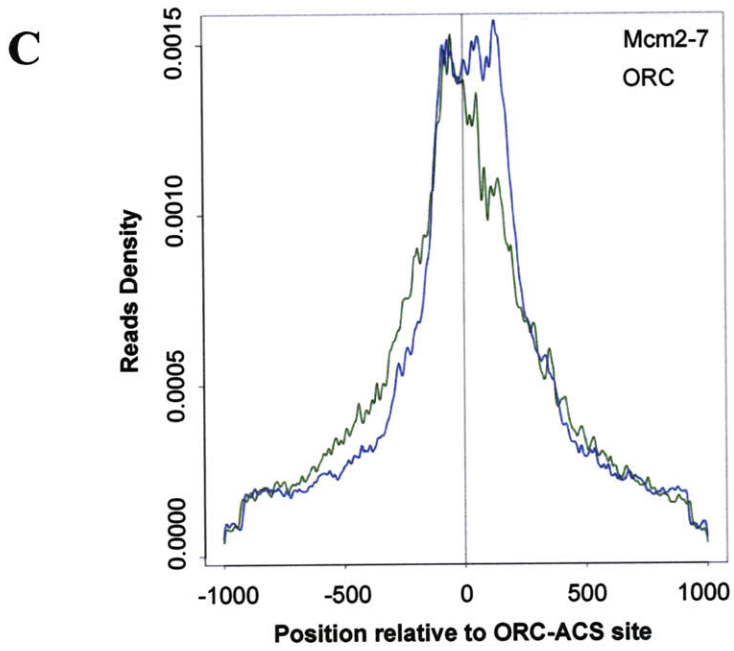


Figure 1. ChIP-Seq mapping of ORC and Mcm2-7 occupancy at origins.

(C) Composite profile of average Mcm2-7 (*blue*) and ORC (*green*) enrichment at 219 ORC-ACS sites. All ORC-ACS sites were aligned relative to the center of the 32-bp consensus ORC binding sequence as defined in Eaton et al. (position 0) (Eaton et al., 2010).

as "downstream." Although the ORC and Mcm2-7 peaks overlapped at position 0, the location of Mcm2-7 enrichment extended downstream, but not upstream, of the location of ORC enrichment, coinciding with the B-side region of the origin. Mcm2-7 enrichment in this downstream region appears to be uniformly distributed across the nucleosome-free region (NFR) (Eaton et al. 2010). These data suggest that in G1 phase, Mcm2-7 complexes are preferentially loaded onto the downstream nucleosome-free region adjacent to ORC binding at the origin.

Length of ORC-adjacent ARS1 DNA modulates efficiency of MCM loading

Because the ChIP-seq data suggested that Mcm2-7 complexes are preferentially located downstream but not upstream of ORC at origins *in vivo* (G1), we asked whether DNA in this downstream region facilitates Mcm2-7 loading *in vitro*. To this end, we used an *in vitro* pre-RC assembly assay (Seki and Diffley 2000; Bowers et al. 2004) in which extracts derived from G1-arrested yeast cells are used to assemble pre-RCs onto immobilized linear DNA templates containing a single copy of the *ARS1* origin of replication located near the free end of the DNA (Fig. 2A). We will refer to the ACS and B1 as the ORC binding site, and flanking sequences downstream of the 3' end of the B1 element will be referred to as ORC-adjacent DNA.

To identify regions of ORC-adjacent DNA that are important for pre-RC assembly, we used a truncation approach. We made deletions of the free end of the DNA

Figure 2

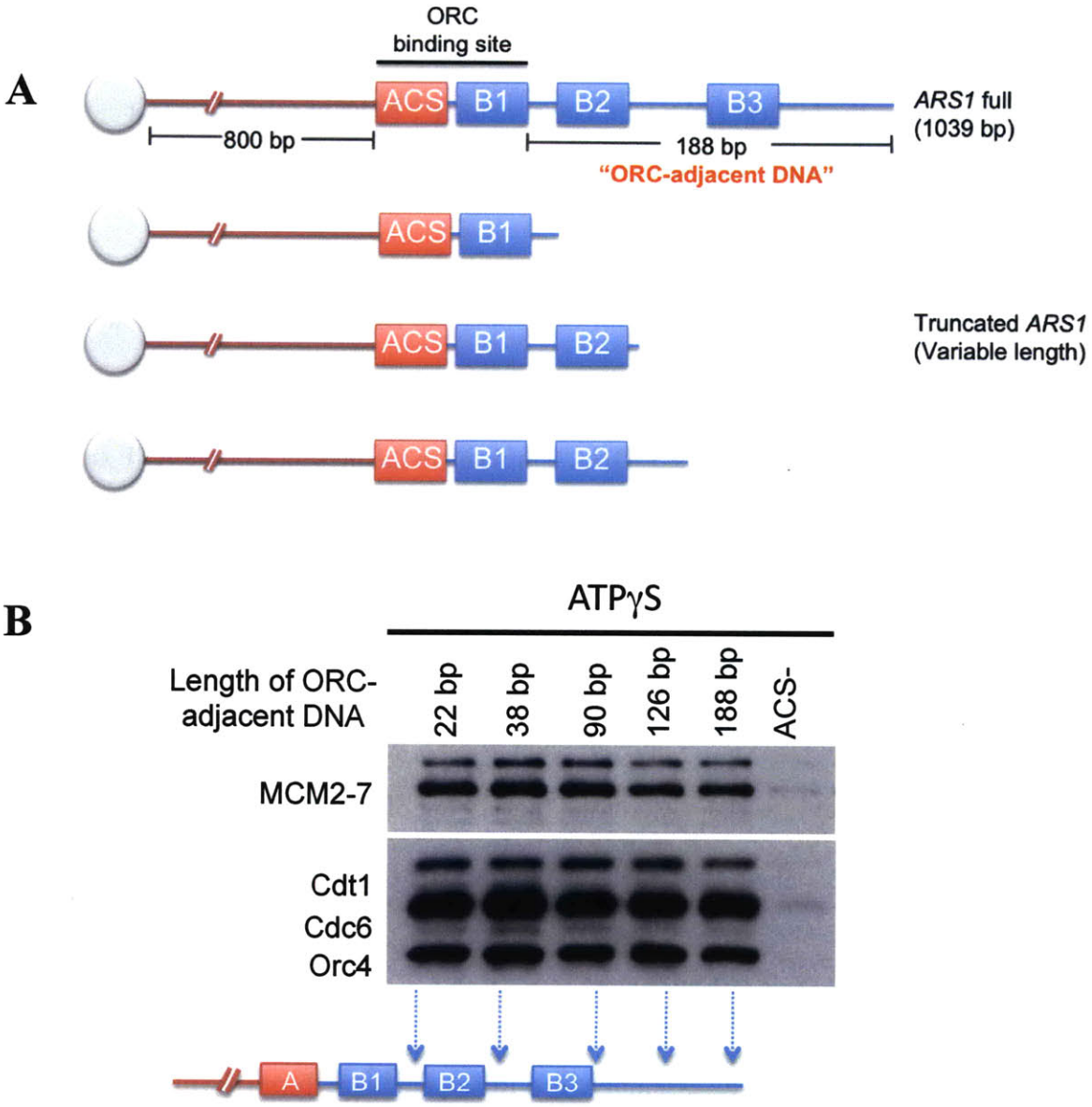


Figure 2. ORC-adjacent B-side sequence is not important for initial Mcm2-7 association.

(A) Diagram of DNA templates used to assess the B-side region of *ARS1* in the *in vitro* pre-RC assembly assay. *Top*: standard 1039 bp (full-length) *ARS1* DNA template. *ARS1* sequence is oriented such that the ACS is closest to the bead-DNA attachment and B2 element is closest to the free end of the DNA. (Note that B3 is omitted from this template; refer to text for explanation.) This full-length template contains 188 bp ORC-adjacent sequence flanking the 3' end of the B1 element. Following fragments contain truncations in the ORC-adjacent sequence. ***(B)*** Initial Mcm2-7 association at *ARS1* templates. Pre-RC assembly was performed with the series of deletion templates depicted in Fig. 2A in the presence of ATPγS. A 1-kb *ARS1* ACS- template was included as a negative control.

while keeping the length of DNA between the ORC binding site and the bead constant in length (Fig. 2A). The shortest DNA template contained only 22 bp of ORC-adjacent sequence, whereas the longest DNA template contained 320 bp of ORC-adjacent sequence.

We first determined the amount of ORC-adjacent DNA required for the initial association of pre-RC factors with the DNA. To this end, we performed pre-RC assembly reactions using the series of DNA templates in the presence of the ATP analog ATP-γ-S. Previous studies showed that under these conditions, all four of the proteins involved in helicase loading associate stably with the origin DNA but Mcm2-7 loading is inhibited (Randell et al. 2006). We observed equal amounts of Mcm2-7 association with each of the DNA templates tested, including the shortest template that retained only 22 bp of B-side sequence (Fig. 2B). Therefore, the amount of ORC-adjacent DNA does not influence the initial association of Mcm2-7 complexes at the origin.

In contrast, we found that the amount of ORC-adjacent DNA was critical for the helicase loading step of pre-RC assembly. Pre-RC assembly performed with the same series of DNA templates in the presence of ATP (to allow for complete helicase loading) showed dramatic changes in the extent of Mcm2-7 binding. Three levels of ATP-dependent Mcm2-7 binding were observed (Fig. 3A). The shortest DNA template with 22 bp of ORC-adjacent DNA showed background levels of ATP-dependent Mcm2-7 binding but ORC and Cdc6 association were unaffected. Thus, there is a minimal flanking DNA requirement for Mcm2-7 complexes to remain associated with origin

Figure 3

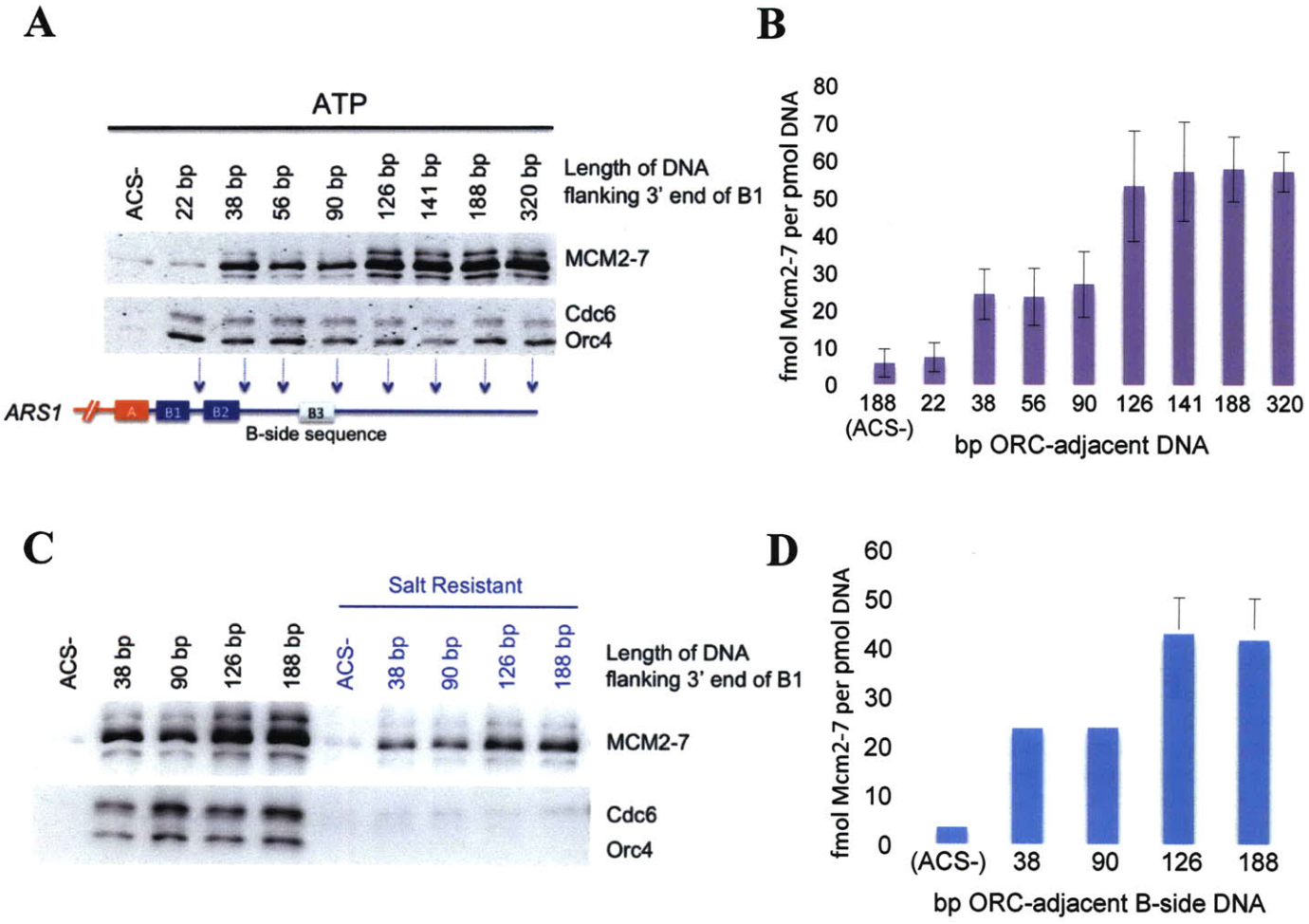


Figure 3. Length of ORC-adjacent B-side DNA is important for Mcm2-7 loading.

(A) Pre-RCs were assembled onto a series of *ARS1* DNA templates in the presence of ATP. Bound proteins were assessed by immunblotting (antibodies: polyclonal Mcm2-7 (UM174), monoclonal Orc4 (SB6) and monoclonal 12CA5 (Cdc6-HA)). A full length (1039 bp) *ARS1* ACS- template was included as a negative control. ***(B)*** Quantification of Mcm2-7 association from experiment described in 3A. Averages were taken from four independent experiments. ***(C)*** Pre-RCs were assembled onto full-length and truncated DNA fragments as described in 3A. Reactions were split in half; one half was subjected to a high salt (500 mM NaCl) extraction. *Lanes 1-5:* Pre-RCs assembled onto templates as indicated. *Lanes 6-10:* Pre-RCs assembled onto DNA templates and washed once with H^+ buffer (50 mM HEPES [pH 7.6], 5 mM Mg-acetate, 1 mM EDTA, 1 mM EGTA, 10% glycerol, and 1mM DTT) plus 500 mM NaCl. ***(D)*** Quantification of salt-resistant Mcm2-7 loading in Fig. 3C (lanes 6-10 only). Averages were taken from three independent experiments.

Figure 3

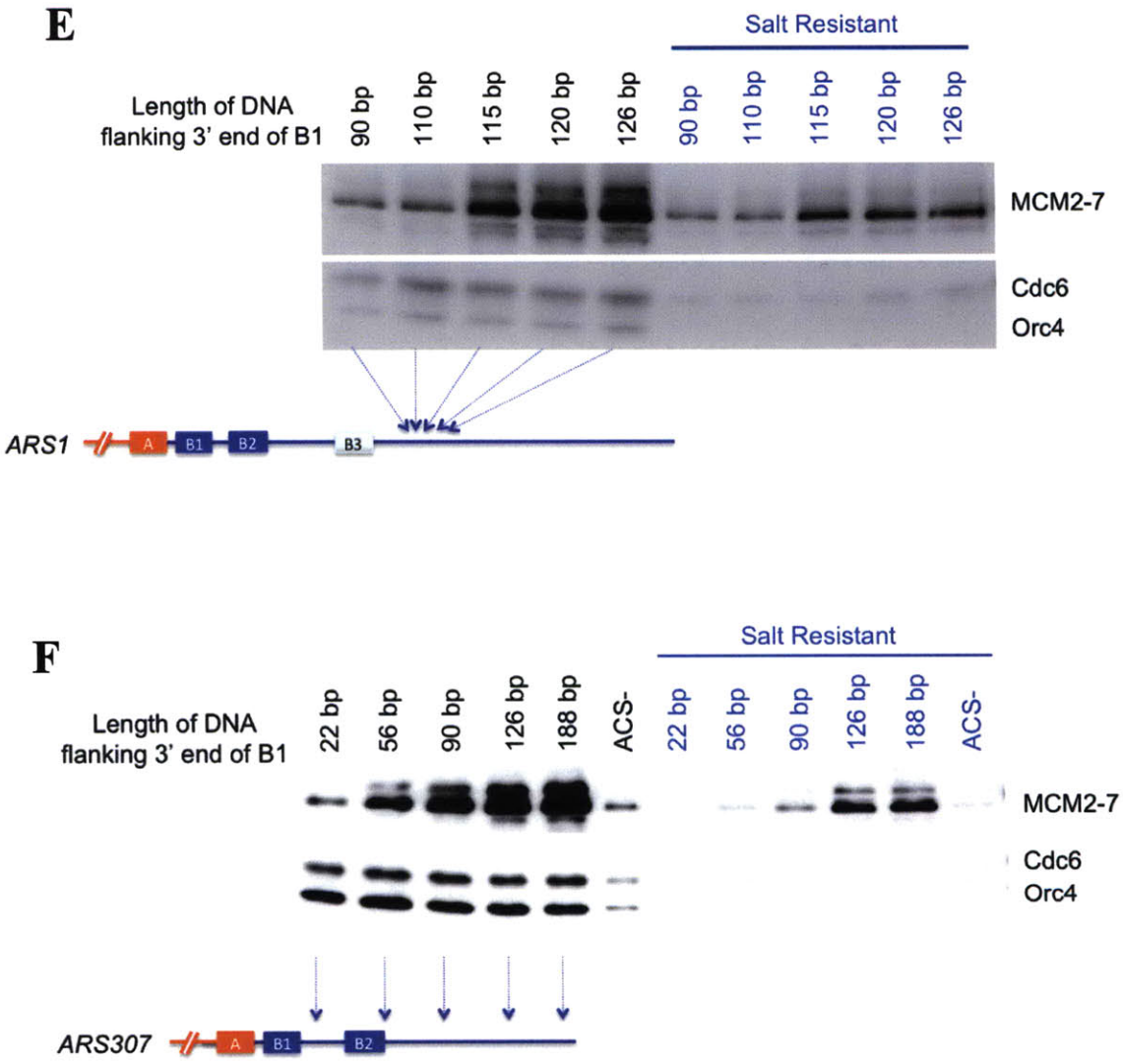


Figure 3. Length of ORC-adjacent B-side DNA is important for Mcm2-7 loading.

(E) Finer mapping of the stepwise increase in Mcm2-7 loading on *ARS1* DNA. Pre-RCs were assembled onto DNA templates containing between 90 and 126 bp of ORC-adjacent (*lanes 1-5*). Salt extractions were performed on assembled pre-RCs (*lanes 6-10*). ***(F)*** Mcm2-7 loading on the *ARS307* origin of replication. Pre-RCs were assembled onto *ARS307* DNA templates of the lengths indicated (lanes 1-6), and salt extracted *(lanes 7-12)*.

DNA. Low levels of Mcm2-7 association were observed for templates containing 38, 56 and 90 bp of ORC-adjacent DNA. Interestingly, we consistently observed a 2-fold increase in ATP-dependent Mcm2-7 binding for a template containing 126 bp of ORC-adjacent DNA sequence (Fig. 3B). Further increases in the length of ORC-adjacent DNA did not result in increased levels of ATP-dependent Mcm2-7 binding. To assess whether Mcm2-7 proteins were stably loaded onto the truncated fragments, we washed assembled pre-RCs with a high-salt buffer (Bowers et al. 2004). Salt-extractions of the pre-RCs assembled onto DNA fragments indicate that all of the DNA templates tested facilitated a similar percentage of stable Mcm2-7 loading (approximately 50%, Fig. 3C).

The 2-fold increase in Mcm2-7 loading in the presence of ATP occurred at a precise length of flanking B-side DNA. We tested an additional set of deletion mutations in the region between the B1+90 and B1+126 mutations (Fig. 3C). This region does not include any of the previously defined elements in *ARS1* (the 3' end of the B3 element is 83 bp downstream of B1). Interestingly, the increase in Mcm2-7 loading upon extension of this region was not gradual but was instead sharp, occurring within a 5-bp region. Pre-RC assembly onto the finer deletion mutations revealed that when the length of ORC-adjacent DNA was extended from 110 to 115 bp, the same 2-fold increase in Mcm2-7 loading was observed (Fig. 3E).

The dependence on the length of ORC-adjacent DNA is not unique to *ARS1*. We examined the *ARS307* origin of replication, which differs from *ARS1* in both the overall sequence of the origin and the spacing between the ACS, B1, and B2 elements (*ARS307*

lacks a B3 sequence) (Rao et al. 1994). Nevertheless, when pre-RCs were assembled onto *ARS307*-containing DNA fragments of different lengths, we observed the same increase in Mcm2-7 loading at roughly the same location (126 bp of ORC-adjacent DNA adjacent to B1) (Fig. 3F). The lack of sequence similarity between the two origins strongly suggests that this effect is not due to a particular sequence but is instead mediated by the length of ORC-adjacent sequence.

In addition to the B-side, the region upstream of the T-rich strand of the ACS, which is referred to as the C-side, has also been shown to contribute to *ARS1* function (Celniker et al. 1984; Srienc et al. 1985; Strich et al. 1986). However, in contrast to the B-side, the C-side contributes to a lesser extent to origin function, and is not known to contain any discrete sequence elements (Celniker et al. 1984). Using an approach similar to the one used to assess the importance of ORC-adjacent DNA in pre-RC assembly, we examined the C-side in the assay and observed Mcm2-7 binding on all DNA templates tested (Fig. 4). Increasing the amount of C-side DNA led to a gradual increase in MCM binding, rather than the stepwise increase observed on the B-side of the origin. Unlike the B-side deletions, even the smallest deletion retained the ability to bind Mcm2-7. These observations suggested that the ORC-adjacent region within the B-side contains unique features that are especially important in modulating levels of MCM loading at the origin and is consistent with this region being the site of Mcm2-7 loading. Thus, we focused our studies on the ORC-adjacent region within the B-side of the origin.

Figure 4

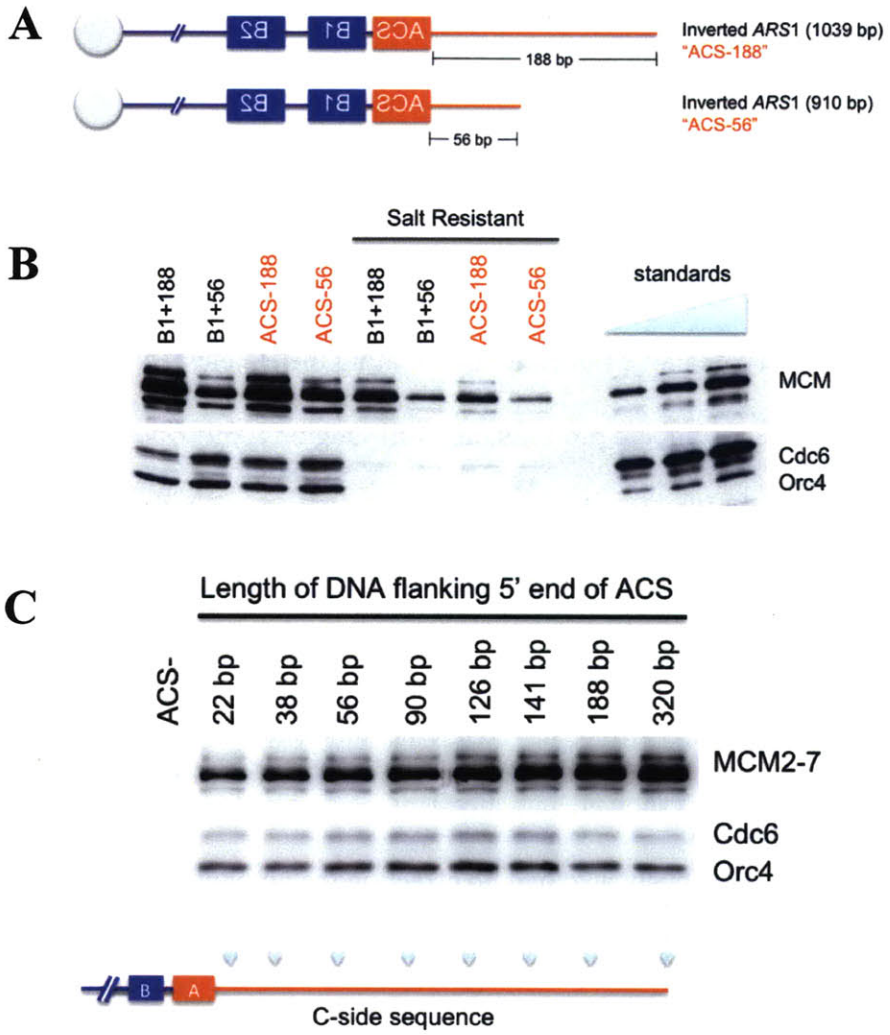


Figure 4. Role of upstream C-side DNA in helicase loading at *ARS1*.

(A) Diagram of DNA templates used to assess C-side Mcm2-7 loading. The orientation of *ARS1* is inverted such that the B2 element is closest to the bead and the ACS is closest to the free end of the DNA. Truncations were made on the free end of the DNA. ***(B)*** Mcm2-7 loading on truncated DNA templates is salt-resistant. Pre-RCs were assembled onto full length and truncated templates as indicated (*lanes 1-4*). Pre-RCs assembled onto DNAs were then washed with 500 mM NaCl (*lanes 5-8*). ***(C)*** Pre-RCs were assembled onto the series of DNA templates analogous to that described in Figure 2A, except that the orientation of the ARS1 sequence is inverted. A full length (ACS-188) DNA template lacking the ACS sequence (ACS-) was included as a negative control.

The role of origin DNA sequences in modulating pre-RC assembly

One interpretation of the minimum length requirement for 38 bp of B-side DNA was that the B2 element (which is included in the 38 bp region beyond the end of B1, but not in the 22 bp region) is required for Mcm2-7 loading onto the truncated DNA template. Although this would be consistent with the previous observation that the B2 element is important for pre-RC assembly *in vivo*, this role for B2 has not yet been confirmed *in vitro*. To examine the effect of mutating the B2 element on pre-RC assembly *in vitro*, we assembled pre-RCs onto full-length (1039) DNA templates containing either a wild-type B2 element or a linker scanning mutation of the B2 sequence that was previously shown to exhibit decreased ARS activity *in vivo* (Marahrens and Stillman 1992). Equal levels of Mcm2-7 loading were observed on both templates (Fig. 5). Similarly, we observed a template containing a linker scanning mutation of mutation in B3 showed an equivalent amount of Mcm2-7 loading. Mutating both B2 and B3 simultaneously also had no effect on the amount of Mcm2-7 loading. Thus, these *in vitro* conditions were not able to recapitulate the events that require B2 function *in vivo*.

The B2 element is important for pre-RC assembly onto chromatinized DNA

A recent study of nucleosome positioning around origins of replication suggested that primary DNA sequence, specifically the asymmetric arrangement of a T-rich ACS

Figure 5

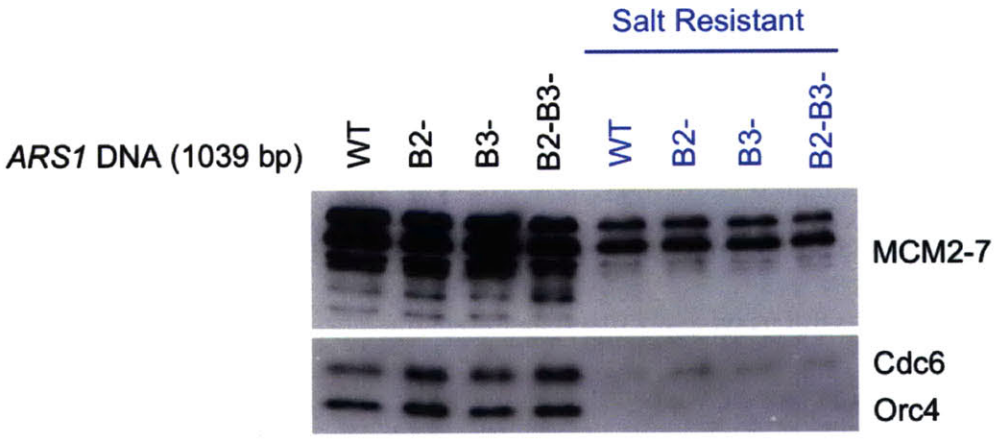


Figure 5. ARS1 B2 and B3 elements are not important for MCM loading onto full-length DNA templates *in vitro*.

Pre-RCs were assembled onto full-length (1039 bp) DNA templates with or without B2 and B3 linker scanning mutations as indicated (*lanes 1-4*). Salt extractions were performed on assembled pre-RCs (*lanes 4-8*).

followed by a downstream A-rich region, plays a critical role in establishing the nucleosome-free region at origins (Eaton et al. 2010). Because the B2 element is A-rich in sequence, we suspected that the need for B2 *in vivo* might be due to a need to overcome chromatin structure to allow for helicase loading onto an unobstructed B-side region. We suspected that the discrepancy between our *in vitro* observations and the *in vivo* data was due to the lack of chromatinized DNA in our *in vitro* pre-RC assembly assay. Attempts to assemble nucleosomes on bead-attached, linear *ARS1* DNA used above were unsuccessful as were efforts to assemble nucleosomes on plasmid DNA attached to beads. In contrast, we could assemble nucleosomes with the characteristic pattern of positioning around the origin on free plasmid DNA (Eaton et al. 2010). To use this template to measure pre-RC formation, however, we needed a new method to separate DNA-associated pre-RC proteins from free proteins. To this end, we assembled pre-RCs onto single-copy *ARS1* plasmids by incubation in G1-arrested yeast extracts, and separated free and DNA-bound proteins by gel filtration (Fig. 6B). We first tested this approach using non-nucleosomal DNA templates, and immunoblotting of DNA-containing fractions confirmed that Mcm2-7 binding occurred on the circular templates in a Cdc6- and *ARS1*-dependent manner (Fig. 6C).

To test the effect of nucleosomes on pre-RC formation, we assembled nucleosomes onto *ARS1* plasmid DNA using purified *S. cerevisiae* histones and the ISWI nucleosome remodeling complex (Fig. 7A) (Vary et al. 2004; Eaton et al. 2010). MNase digestion of *ARS1* plasmid DNA following nucleosome assembly confirmed the formation of regularly-spaced nucleosomes (Fig. 7B). Assembly of pre-RCs onto naked

Figure 6

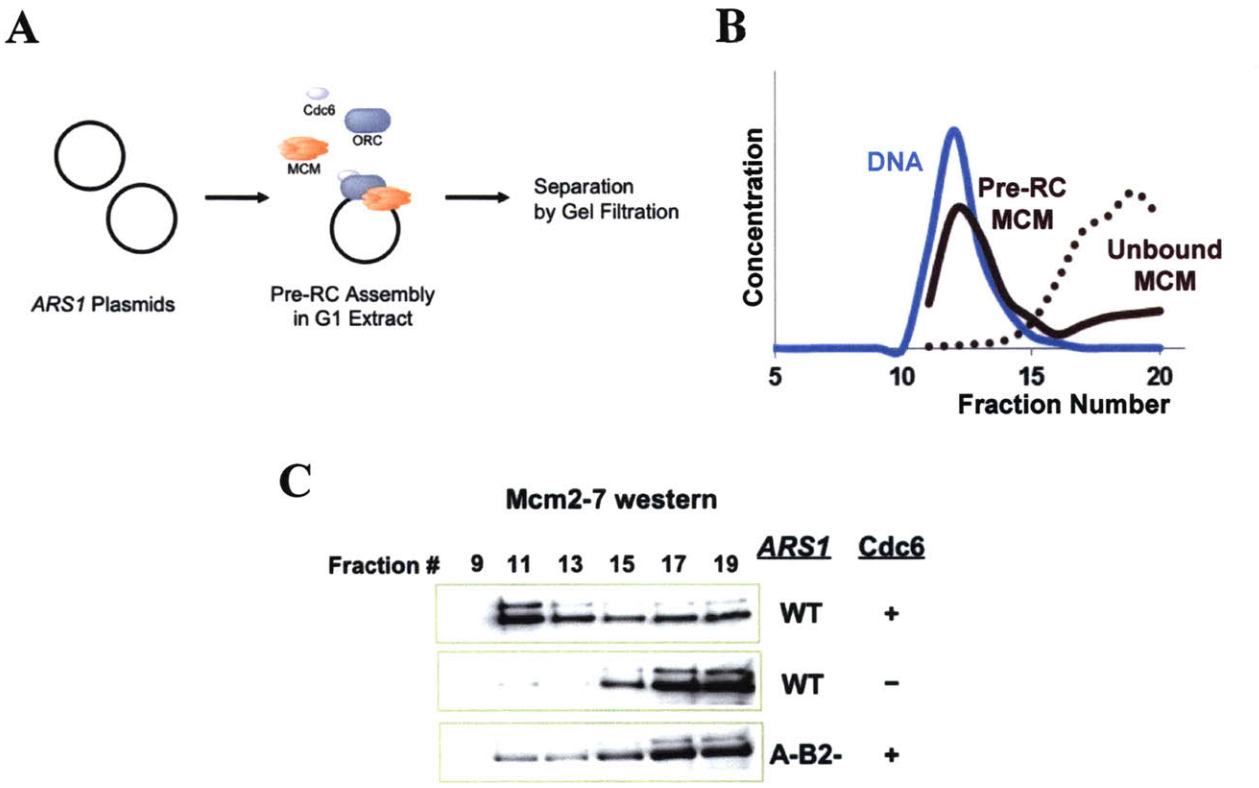


Figure 6. Pre-RC assembly onto *ARS1* plasmids *in vitro*.

(*A*) Experimental outline for assembly assay using circular *ARS1* templates. Pre-RCs were assembled onto wild type or A-B2- pUC19 ARS1 1Kb plasmid DNA by incubation in G1-arrested ySC15 extract with or without Cdc6 protein. Reaction mixtures were then loaded onto a 5 ml Sephacryl S500 gel filtration column (as described in *Materials and Methods*). **(*B*)** Elution pattern of pre-RC-bound plasmid DNA (blue), pre-RC bound Mcm2-7 (solid brown line), and free mcm2-7 (dotted brown line), separated on a Sephacryl S500 column. The plasmid DNA, whose molecular weight is over the separation range of Sephacryl S500, eluted at the void volume (peak at fraction 11). **(*C*)** Pre-RC assembly on *ARS1* plasmids. Mcm2-7 immunoblot of void volume fractions. MCM co-elution with plasmid DNA was dependent on *ARS1* and Cdc6.

Figure 7

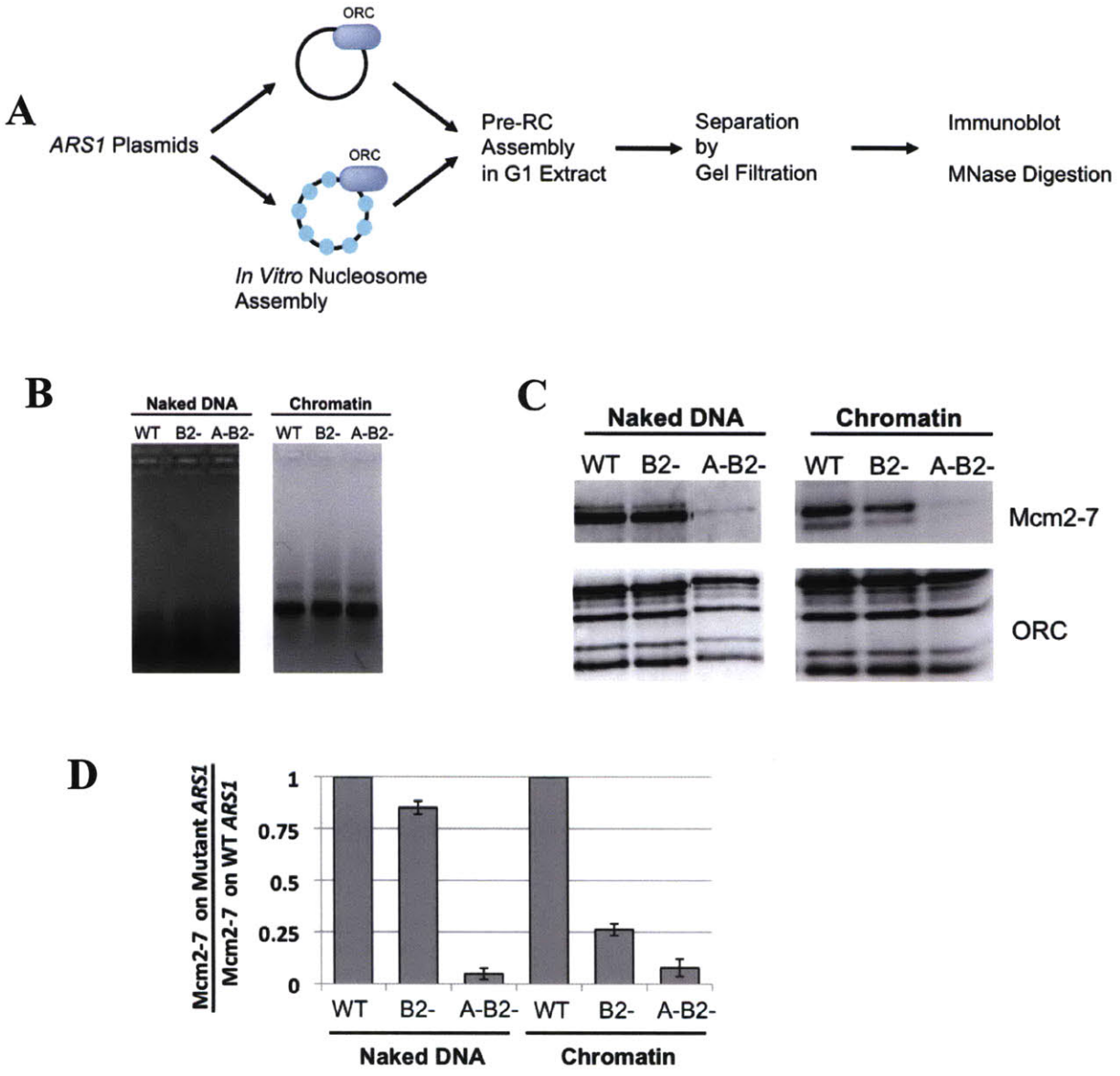


Figure 7. The B2 element is important for pre-RC assembly on chromatinized *ARS1* plasmids *in vitro*.

(*A*) Experimental outline for pre-RC assembly assay on chromatinized DNA. Nucleosomes were assembled onto wild type, B2-, or A-B2- pUC19 ARS1 1Kb plasmid (as described in *Materials and Methods*) prior to pre-RC assembly, separated by gel filtration, and analyzed by immunoblotting and MNase digestion. ***(B)*** MNase digestion of naked and chromatinized plasmid DNA. Plasmids were digested for 10 min, and after proteinase K digestion, DNA samples were run on a 1.5% agarose gel and stained with ethidium bromide. **(*C*)** Pre-RC assembly onto naked and chromatinized *ARS1* plasmids. Pre-RCs were assembled onto either naked plasmids or chromatinized plasmids containing a wild type, B2-, or A-B2- *ARS1* origin. After separation by gel filtration, fraction number 11 was analyzed by immunoblotting for Mcm2-7. **(*D*)** Quantification of Mcm2-7 loading onto *ARS1* plasmids from 7C. Mcm2-7 loading on each mutant plasmid was expressed as a ratio relative to wild-type.

ARS1 plasmids showed that, as before, in the absence of nucleosomes, mutation of B2 had no effect on pre-RC formation. In contrast, in the presence of nucleosomes, mutation of B2 results in a consistent 4-fold reduction in pre-RC formation (Fig. 7C and 7D). The extent of this defect is directly comparable to the reduction in Mcm2-7 binding caused by the same B2 mutation *in vivo* (Wilmes and Bell 2002).

The B2 element is not an essential determinant of nucleosome positioning at the origin

As mentioned earlier, one explanation for the decrease in Mcm2-7 loading in the absence of a functional B2 element is that this element prevents nucleosomes from encroaching into the origin. To test this hypothesis, we assembled nucleosomes in the presence or absence of ORC onto *ARS1* plasmids containing either a wild-type or mutant B2 sequence. We then mapped the position of nucleosomes by micrococcal nuclease digestion followed by high-throughput sequencing of the resulting mononucleosome-associated DNA (Fig. 8). Black and red lines indicate the number of 5' and 3' sequence reads (respectively) of the MNase-treated DNAs along the *ARS1* origin region. Separation of ~147 bp between 5' and 3' sequence read peaks is indicative of the presence of a nucleosome.

We observed that mutation of B2 changed the pattern of MNase sensitivity within the origin. In the absence of ORC, neither positioned nucleosomes nor a nucleosome-free region was observed for either template (Fig. 8). For wild-type origin DNA, addition

Figure 8

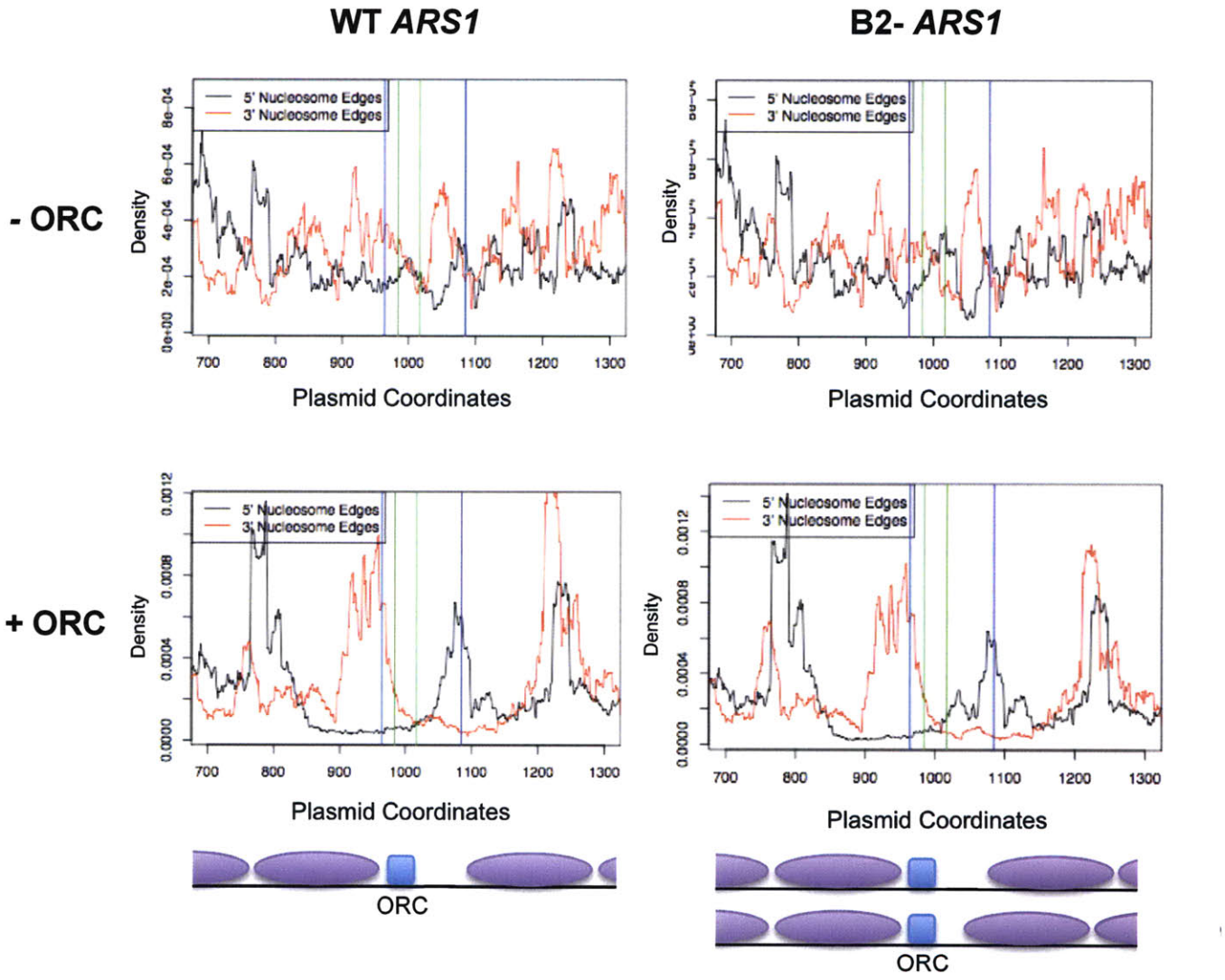


Figure 8. The B2 element is not an essential determinant of nucleosome positioning at the *ARS1* origin *in vitro*.

Nucleosomes were assembled on wild type, B2- or A-B2- *ARS1* plasmids in the presence or absence of ORC as described in *Materials and Methods.* Pre-RCs were assembled onto chromatinized plasmids, and nucleosome positions were mapped as described. Black lines indicate the 5' edges of nucleosomes and red lines indicate the 3' edges of nucleosomes. Blue vertical lines indicate the boundaries of the average nucleosome free region (NFR) relative to the ORC binding site *in vivo* (Eaton et al., 2010). Green vertical lines indicate the boundaries of the ORC binding site in *ARS1*.

of ORC led to the formation of a nucleosome-free region encompassing the origin with positioned nucleosomes on either side. Mutation of the B2 element did not change the position of the nucleosome adjacent to ORC on the 5'side of the origin. In contrast, although the majority of nucleosomes adjacent to 3' end of the NFR were unchanged, we observed a new peak of 5' end sequence reads adjacent to the 3' end of the ORC binding site. Unlike the nucleosomes flanking the origin, this peak did not have a corresponding peak of 3' end DNAs 147 bp away as expected for nucleosome-derived DNAs. It is possible, however, that such a peak could be hidden under the larger peak of 3' end DNAs observed for the majority of 3' flanking nucleosomes. We are currently using paired-end sequencing to address this possibility. However, even if the minor peak were due to nucleosome occupancy in the B-side region, the modest amount of nucleosome encroachment seems unlikely to explain fully the 4-fold decrease in Mcm2-7 loading observed for the B2 mutant origin.

The B2 element is important for facilitating accessibility to DNA for helicase loading

An alternative explanation for the B2 effect observed on the nucleosomal DNA template is that the presence of nucleosomes flanking the origin restricts the amount of DNA that is accessible for helicase loading. If this were the case, then restricting access to this DNA by other means should also reveal an increased importance for B2. To test whether B2 facilitates DNA accessibility *in vitro*, we restricted the amount of ORC-adjacent DNA using the same truncation approach as in Figure 2. We tested the naked

linear template containing 38 bp of ORC-adjacent sequence, since this was the shortest DNA length at which we observed stable helicase loading. We compared pre-RC assembly onto truncated templates containing either the wild-type or mutant B2 sequence. In the presence of ATPγS, the level of initial Mcm2-7 association on either template was equivalent (Fig. 9A, lanes 1 and 4). However, when pre-RCs were assembled onto the DNA fragments in the presence of ATP, 4-fold less Mcm2-7 binding was observed on the DNA fragment lacking B2 (Fig. 9A, lanes 2 and 5). Salt extraction of pre-RCs assembled onto both DNA fragments revealed that in the absence of B2, no stable Mcm2-7 loading is observed on the truncated template (Fig. 9A, lanes 3 and 6). In the presence of ATP, the total amount of ATP-dependent Mcm2-7 binding as well as the amount of stable Mcm2-7 loading onto the truncated B2- template was approximately 4-fold less than the levels observed on the wild-type templates, consistent with the decrease observed in Figure 6 and *in vivo*.

DISCUSSION

In this study, we addressed several longstanding questions regarding how DNA sequence and chromatin structure influence helicase loading at origins of replication. We found that Mcm2-7 complexes are predominantly loaded downstream but not upstream of the ORC binding site throughout the yeast genome, coinciding with the nucleosome-free region (NFR) associated with origin sequences. *In vitro* helicase loading studies found that deletion of sequences within this region had no effect on the initial association of

Figure 9

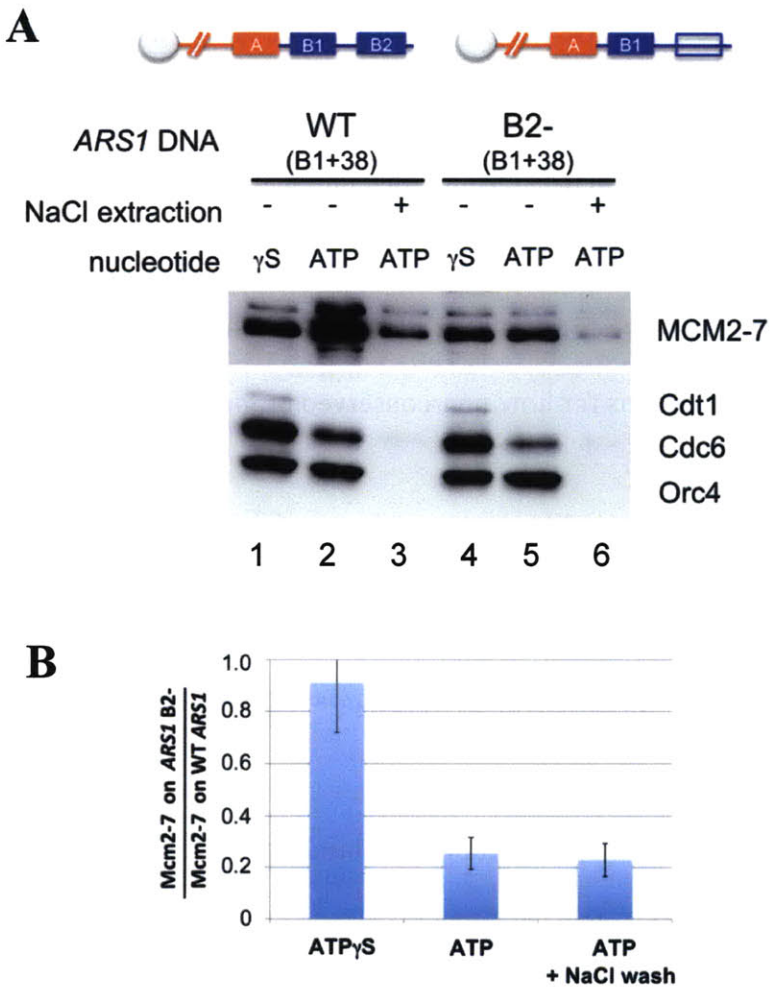


Figure 9. B2 is important for pre-RC assembly on truncated ARS1 templates in *vitro*.

(A) Pre-RCs assembly on truncated DNA fragments containing 38 bp of ORC-adjacent B-side DNA ("B1+38"). *ARS1* DNA fragments contained either a wild-type B2 sequence (*lanes 1-3*) or a mutant B2 sequence (*lanes 4-6*). Pre-RCs were assembled onto truncated DNA fragments in the presence of either ATPγS (*lanes 1 and 4*) or ATP (*lanes 2, 3, 5, 6*). Lanes 3 and 6 indicate pre-RC reactions subjected to a 500 mM NaCl extraction following assembly. ***(B)*** Quantification of defect in Mcm2-7 association on truncated (B1+38) B2- *ARS1* templates as compared to WT *ARS1*. For each condition (ATPγS, ATP, ATP/NaCl), the amount of Mcm2-7 association on the truncated B2- template was expressed as a ratio relative to wild-type. Averages were taken from three independent experiments.

Mcm2-7 complexes with the origin, but either reduced or completely eliminated Mcm2-7 loading. A key sequence element within this region, B2, is important for the stable loading of the Mcm2-7 helicase under conditions that restrict the accessibility of adjacent DNA. Together, these data provide insights into how origin sequences establish the specific protein/DNA and nucleosomal architecture necessary to facilitate helicase loading at origins in *S. cerevisiae*. Because the sequence of the B2 element is not conserved (Rao et al., 1994; Lin and Kowalski 1997; Wilmes and Bell 2002) these data have important implications for how non-conserved origin sequences contribute to helicase loading.

The Mcm2-7 helicase is preferentially loaded downstream but not upstream of ORC

Using a previously defined set of ORC binding sites for which the ORC-bound ACS is known (ORC-ACSs), we mapped the relative location of Mcm2-7 binding during G1 by ChIP-Seq. Analysis of these data revealed a distinct, but partially overlapping, localization of ORC and Mcm2-7 at origins of replication. Specifically, Mcm2-7 enrichment extends from the site of ORC binding to the downstream (but not upstream) flanking DNA. Intriguingly, this region largely coincides with the nucleosome-free region (NFR) at origins that was defined in a recent study (Eaton et al. 2010) and includes the B-side region flanking the ORC binding site.

What ensures that Mcm2-7 complexes primarily occupy the region downstream of ORC? We posit that the asymmetric orientation of ORC bound to the origin is the primary cause for preferential loading of Mcm2-7 helicases downstream but not upstream of the ORC binding site. Previous crosslinking studies of ORC bound to DNA demonstrated that the heterohexameric complex interacts with the asymmetric origin sequence with a specific directionality (Lee and Bell 1997). An attractive hypothesis based on these observations is that ORC loads Mcm2-7 complexes in a unidirectional manner, and that the specific orientation of ORC bound to the origin, as determined by the orientation of the ACS, allows for preferential helicase loading at the nucleosome-free region upstream of ORC. This model is consistent with our observation that deleting the majority of ORC-adjacent B-side DNA completely eliminated Mcm2-7 loading *in vitro*, whereas deleting an equivalent amount of C-side sequence upstream of ORC reduced but did not completely eliminate Mcm2-7 loading (Fig. 3A; Fig. 4).

In addition to unidirectional helicase loading by ORC, the nucleosomal landscape at origins of replication also is likely to restrict Mcm2-7 loading to the DNA downstream of ORC. The NFR at origins is asymmetric with respect to the ACS, extending downstream but not upstream of ORC (Eaton et al. 2010). The first nucleosome downstream of ORC is located ~60-90 bp away from the ORC binding site, whereas the first nucleosome upstream of ORC is located immediately adjacent to ORC. Given that a single Mcm2-7 hexamer is proposed to encircle approximately 35 bp of DNA, the nucleosomes positioned immediately upstream of ORC would be expected to obstruct

Mcm2-7 loading whereas the downstream nucleosome-free region would provide an ideal loading zone.

The overlap of ORC and Mcm2-7 binding at origins is unlikely to reflect simultaneous binding of both proteins to the origin DNA. Our composite data of the average occupancy of Mcm2-7 and ORC at the 219 ORC-ACS sites show a significant overlap of the two peaks over the ACS. Due to the large size of ORC and the Mcm2-7 complex and the likelihood that the Mcm2-7 complex encircles origin DNA after loading (Remus et al. 2009), we do not believe that the two proteins ever simultaneously bind to the same DNA sequence. Two properties of pre-RC formation are likely to contribute to this apparent overlap. First, as described previously, prior to loading, the Mcm2-7 proteins associate with the origin in an ORC-dependent manner. Indeed, both *in vivo* and *in vitro* studies show that approximately 50% of Mcm2-7 helicases associated with the origin are sensitive to salt-extraction (Donovan et al. 1997; Bowers et al. 2004), suggesting that these helicases are interacting with the origin via ORC. Such Mcm2-7 molecules would be expected to co-localize with ORC by ChIP. Second, recent observations suggest that the pre-RC is a dynamic complex with ORC releasing after Mcm2-7 is loaded (Tsakraklides and Bell 2010). Following ORC dissociation from the origin, the inactive Mcm2-7 helicase could then occupy the region previously bound by ORC. Such a movement of the Mcm2-7 complex after loading is consistent with the *in vitro* observation that inactive Mcm2-7 complexes are capable of sliding on double stranded DNA in a non-directional, energy-independent manner (Remus et al. 2009).

DNA length requirements for Mcm2-7 association and loading at origins

Our studies indicate that the only requirement for the initial recruitment of Mcm2-7 to origins is the presence of a functional ORC binding site (Fig. 10), and that neither the length nor the sequence of the DNA flanking the ORC binding site influences the initial association of the helicase with the origin-bound pre-RC (Fig. 2B and 8A, Fig. 10). The shortest DNA template tested, which contained 22 bp of ORC-adjacent sequence, was fully competent for the initial association of the Mcm2-7 helicase (Fig. 2B). Furthermore, increasing the amount of upstream flanking DNA did not increase the level of initial Mcm2-7 association. Our interpretation of these data is that the initial association of the Mcm2-7/Cdt1 complex to the origin is mediated primarily through protein-protein interactions with other proteins bound to the A and B1 regions (e.g. ORC). This is at least partially mediated by Cdt1's interaction with Orc6 but is likely to involve other interactions as well (Chen et al., 2007). Our observation that ORC-adjacent DNA is dispensable for the first step of helicase association does not support the existence of robust interactions between the Mcm2-7 helicase and the DNA at this intermediate step in helicase loading. The ORC binding site alone (ACS and B1) is sufficient to recruit all four pre-RC components to DNA (Fig. 10).

In contrast to its lack of impact on the initial association of the Mcm2-7 complex, the amount of ORC-adjacent downstream DNA (which we refer to as ORC-adjacent DNA hereafter) is crucial for Mcm2-7 loading. Although helicase association was observed on the shortest template containing 22 bp of flanking sequence, no Mcm2-7

Figure 10

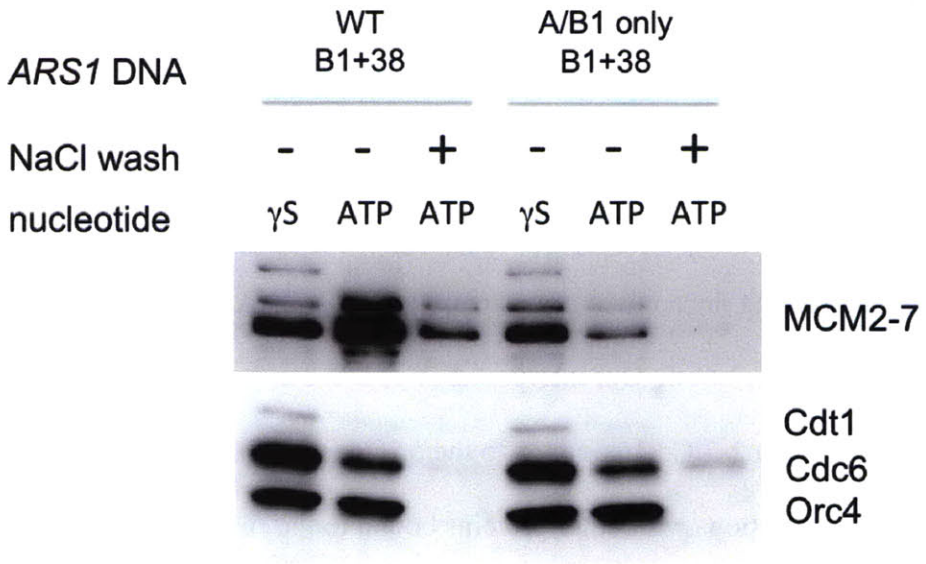


Figure 10. The ORC binding site (A and B1) is sufficient for initial helicase association to *ARS1* but not for ATP-dependent helicase loading.

A 49-bp region fragment containing the ARS1 ACS and B1 elements was subcloned into the MCS of the pBluescript vector (Stratagene), and pre-RC assembly templates were amplified by PCR. Pre-RCs were assembled onto truncated (B1+38) templates containing either the wild type *ARS1* origin *(lanes 1-3)* or the ORC binding site (A and B1) alone *(lanes 4-6)*. Assembly reactions were performed in the presence of either ATPγS (*lanes 1 and 4*) or ATP (*lanes 2, 3, 5, 6*). Lanes 3 and 6 indicate pre-RC reactions subjected to a 500 mM NaCl extraction following assembly.

binding or loading was observed at this length, suggesting that ATP hydrolysis by ORC or Cdc6 releases Mcm2-7 complexes into solution in the absence of adjacent DNA available for loading. Increasing the amount of flanking DNA allowed resulted in two levels of ATP hydrolysis-dependent Mcm2-7 loading. Our data indicate a minimum length requirement of 38 bp of ORC-adjacent DNA to allow for stable loading of Mcm2-7 complexes at the origin. Gradual extension of this ORC-adjacent DNA did not lead to any additional Mcm2-7 loading until an additional 77 bp was added back to the linear template (reaching a total of 115 bp of ORC-adjacent sequence). Intriguingly, this 2-fold increase in Mcm2-7 loading was not gradual but was instead sharp, occurring within a 5 bp window. This observation has interesting implications for the mechanism of helicase loading. If increasing the length of ORC-adjacent DNA only increased the accessibility of this region for Mcm2-7 loading, one would predict that the corresponding increase in Mcm2-7 loading would be gradual upon extension of the adjacent region. This is not the case, however, and the observation that a region as little as 5 bp in length is critical for helicase loading suggests that a precise number of basepairs of DNA is required for the helicase to interact stably with origin DNA.

Our observation that a minimum of 38 bp of ORC-adjacent DNA for the first level of Mcm2-7 loading and an additional 77 bp for the second level of Mcm2-7 association raises questions about the number and location of Mcm2-7 hexamers loaded per origin. A recent EM study indicated that a single Mcm2-7 hexamer is about 115 angstroms in length, which would occupy ~35 bp of DNA when bound (Remus et al. 2009). Thus, the simplest interpretation of our data is that 38 bp of ORC-adjacent

sequence provides sufficient space for a single Mcm2-7 hexamer to load immediately downstream of ORC. However, the same EM study also suggested that following pre-RC assembly, DNA-bound Mcm2-7 complexes exist exclusively as head-to-head double hexamers encircling double-stranded DNA. If, after loading, the second hexamer were located immediately adjacent to the first, why would an additional 77 bp, and not ~35 bp, be needed to observe the 2-fold increase in Mcm2-7 loading?

There are at least two possible models to explain the requirement for additional ORC-adjacent DNA to observe increased Mcm2-7 loading. One possibility is that a Mcm2-7 double-hexamer is loaded onto the ORC-adjacent DNA in a concerted fashion, and that 38 bp of DNA adjacent to ORC is sufficient for double-hexamer loading. If this model is correct, then the adjacent DNA would only be sufficient to pass through the channel of the first Mcm2-7 hexamer and the second hexamer would primarily be associated with the origin through its N-terminal contacts with the first hexamer. Increasing the flanking DNA length by another 77 bp would then allow for the loading of a second double hexamer, which again would remain associated with roughly ~35 bp of DNA passing through the central channel of one hexamer (Fig. 11A). The fact that the increase occurs upon addition of ~70 bp also supports this model as this is the size of a Mcm2-7 double-hexamer. An alternative model is that a single Mcm2-7 hexamer is associated with the 38 bp region immediately adjacent to ORC, and that with sufficient space, a maximum of one double hexamer is loaded per origin. In this model, there are two dedicated Mcm2-7 loading sites at the origin: the first located immediately downstream and adjacent to ORC, and the other located distal from the first loading site

Figure 11

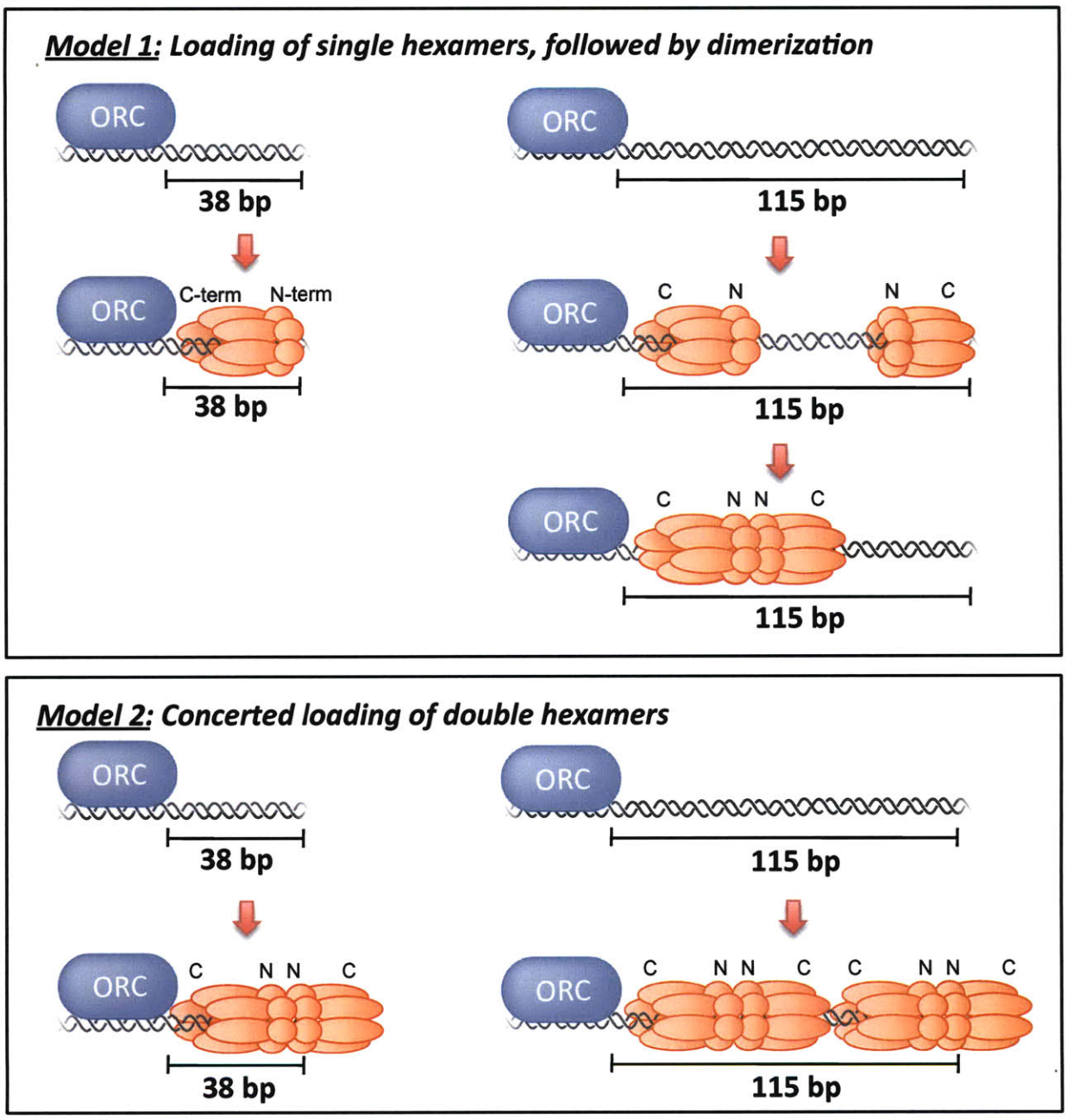


Figure 11. Two models for Mcm2-7 loading onto truncated DNA templates in the *in vitro* pre-RC assembly assay.

(A) Model 1: Loading of single Mcm2-7 hexamers, followed by translocation and dimerization of hexamers. A single Mcm2-7 hexamer binds to 38 bp of ORC-adjacent B-side DNA. Extension of this region to 115 bp allows for loading of a second Mcm2-7 hexamer at a distal site. Translocation of one or both hexamers on dsDNA is required to form a head-to-head double hexamer. ***(B)*** Model 2: Concerted loading of double Mcm2-7 hexamers. 38 bp of ORC-adjacent B-side DNA (on the truncated DNA templates) is sufficient for the loading of one double hexamer. The double hexamer remains associated with ~35 bp passing through the central channel of one hexamer. Extension of the ORC-adjacent DNA to 115 bp allows for the loading of a second double hexamer. Again, the second double hexamer remains associated to ~35 bp of DNA.

(Fig. 11B). In contrast to the first model, this model would suggest that individual hexamers can be loaded at the origin. The two hexamers would be loaded separately but in an opposite orientation to one another, and the subsequent translocation of one or both hexamers would be required to form the double hexamer.

Both models are consistent with the 2-fold increase in Mcm2-7 association on the longer DNA templates. Although quantification of the number of Mcm2-7 complexes bound per DNA molecule would distinguish between the two models, the inefficiency of Mcm2-7 loading in our population-based *in vitro* assay makes accurate quantification difficult. Single-molecule experiments would be needed to distinguish between these two possibilities.

The stepwise increase in Mcm2-7 loading is reminiscent of a previous observation in *Xenopus* egg extracts that incremental increases in DNA length led to a similar stepwise increase in Mcm2-7 loading (Edwards et al., 2002). However, in contrast to the *Xenopus* system, the increase in the amount of Mcm2-7 loading in the *S. cerevisiae* extract-based assay is not proportional to the DNA fragment size; that is, in *S. cerevisiae*, increasing the DNA length beyond 115 bp did not lead to additional Mcm2-7 loading. This could be due to inherent differences in the mechanism of helicase loading between the two organisms. For example, it is possible that the yeast proteins are only able to load Mcm2-7 hexamers locally whereas the *Xenopus* proteins are able to load Mcm2-7 hexamers at more distant sites (e.g. by displacing nucleosomes). Consistent with this possibility, it is estimated that a single molecule of ORC loads 20-40 Mcm2-7 hexamers

per origin in the cell-free *Xenopus* system (Mahbubani et al. 1997; Walter and Newport 1997), whereas the number is believed to be much lower in *S. cerevisiae* (Bowers et al. 2004; Remus et al. 2009). Alternatively, the much higher efficiency of the *Xenopus* helicase loading reaction could also be responsible for the difference.

A role for the B2 element in the mechanism of helicase loading

Unlike previous *in vitro* studies that suggested that an ORC binding site is the only origin sequence important for helicase loading, we identified two conditions that recapitulate the dependence of Mcm2-7 origin association on the B2 element that is observed *in vivo*. Mutating the B2 element in a longer DNA template had no effect on helicase loading, whereas truncation of the B-side region revealed a dependence on the B2 element for maximal Mcm2-7 loading on the shorter template. Similarly, assembling nucleosomes onto origin DNA prior to pre-RC assembly revealed a dependence on the B2 element for maximal helicase loading at the origin. In both cases, the B2 element was important under conditions that restrict the amount of ORC-adjacent DNA accessible for Mcm2-7 loading, suggesting a role for B2 in facilitating the accessibility of this DNA.

How does the B2 element function to facilitate the accessibility of adjacent DNA for helicase loading? Based on observations from previous studies, two general functions for the B2 element have been proposed: (1) the B2 element could perform a simple structural function, such as DNA bending, that is necessary for the mechanism of helicase

loading; (2) the B2 element could serve as a binding site for a protein component of the pre-RC. The two possibilities are not mutually exclusive, and several lines of evidence exist that support both models. For instance, single point mutations have no effect on B2 function (Rao et al. 1995), consistent with the model that B2 acts structurally rather than as a protein binding site. In support of a protein-binding model, an analysis of sequences that can replace B2 at *ARS1* found that the most critical determinant of B2 function is a similarity to the ACS (Wilmes and Bell 2002). In addition, inversion of the *ARS1* B2 sequence (Lin and Kowalski 1997) or replacement with a random AT-rich sequence (Wilmes and Bell 2002) is not sufficient to confer origin function.

We propose a model in which B2 serves a dual function as both a structural element and a protein binding site by facilitating the wrapping of origin DNA around ORC. Several observations from previous studies support the existence of a DNA wrap around ORC. Purified *S. cerevisiae* ORC has been shown to induce a bend in *ARS1* DNA *in vitro* (Lee et al. 1997). DNase I footprinting of purified ORC bound to *ARS1* revealed three distinct sites of enhanced DNase I cleavage – one within the B1 element, and two between the B1 and B2 elements – with a 10-bp periodicity, consistent with a DNA wrapping around ORC (Bell and Stillman 1992). Furthermore, single particle EM studies of *D. melanogaster* ORC bound to origin DNA results in the apparent shortening of the ORC-bound DNA by ~130 bp (Clarey et al. 2008), which has been interpreted as indicative of an extensive region of DNA wrapping around ORC. We hypothesize that the B2 element facilitates ORC's interaction with the downstream B-side region to facilitate DNA wrapping. Consistent with this hypothesis, crosslinking studies of

purified *S. cerevisiae* ORC bound to *ARS1* DNA showed that at least three ORC subunits – Orc2, Orc4, and Orc6 – interact with the region downstream of B1 (Lee and Bell 1997). Interestingly, Orc2 crosslinked specifically to the B2 element in *ARS1*. In addition, DNase I footprinting of purified ORC and Cdc6 bound to *ARS1* revealed an extended footprint that protected the B2 element (Speck et al. 2005). It may be the case that the extended footprint is caused by Cdc6 binding to B2, but it is also equally likely that Cdc6 binding to ORC induces a conformational change that extends ORC's interactions with downstream DNA and the B2 element. However, it is important to note that mutation of the B2 sequence does not affect ORC's affinity for *ARS1* DNA, and pattern of DNase I protection by ORC is unchanged on a B2- *ARS1* sequence (Bell and Stillman 1992). Unlike the ACS and B1 elements, the B2 element is clearly not required for ORC to recognize and bind to origin sequences. Instead, we predict that this potential interaction of ORC with the B2 element is important in the specific context of helicase loading.

There are at least two ways to envision how the formation of a DNA wrap around ORC can facilitate Mcm2-7 loading at the origin. The simplest explanation, as supported by our nucleosome mapping data, is that the B2 element assists ORC in establishing the nucleosome-free region. Based on the known ORC interaction with the A and B1 element, it is hard to understand how ORC determines the positioning of nucleosomes flanking the 3' end of the NFR, which is on average ~60-90 bp from the 3' edge of the ORC footprint. Instead, if ORC interacts with the downstream sequences directly, possibly through the B2 element, this could exclude this region from nucleosome binding and position the downstream nucleosome. It is clear, however, that

this is not the only means by which B2 facilitates helicase loading onto the B-side region. We observe a strong effect of B2 on naked DNA when the downstream flanking sequences are restricted (Fig. 9). Based on this observation, we propose that B2-dependent DNA wrapping by ORC not only keeps the region nucleosome-free but also brings that DNA region within close proximity of ORC/Cdc6, facilitating Mcm2-7 loading at the adjacent DNA. Thus, by influencing how ORC interacts with downstream B-side DNA, the B2 element can serve a dual function to facilitate helicase loading at the origin.

Implications for origin function in higher eukaryotes

Of all eukaryotic origin sequences studied to date, *S. cerevisiae* and its relatives are the only organisms known to contain a well-defined consensus sequence that is essential for origin function. The fact that no other eukaryotic origins exhibit any sequence conservation has raised questions about whether the budding yeast paradigm is applicable to all eukaryotes. However, it is important to keep in mind that although ACS is conserved in sequence among yeast origins, the B elements that also contribute to origin function are highly degenerate in sequence. The observation that the B2 element, a non-conserved *cis*-acting sequence element, is important for helicase loading is reminiscent of origins function in more complex eukaryotic systems.

Several examples of non-conserved sequence elements have been shown to contribute to origin function in other eukaryotic organisms. *S. pombe* origins contain several 30-50 bp AT-rich islands that exhibit no sequence similarity, and collectively mutating these sequences abrogates origin function (Dubey et al. 1996; Takahashi et al. 2003). Similarly, sequence elements that contribute to origin function have been identified at a locus that directs chorion gene amplification in *Drosophila* follicle cells (Lu et al. 2001) as well as in the human β-globin locus (Aladjem et al. 1998). However, these rare examples of metazoan origin sequence elements are usually not conserved in other origins within the same organism. Although some of these sequence elements have been shown to bind ORC (Austin et al. 1999; Lu et al. 2001; Takahashi et al. 2003), it is clear that some of these sequences act in *cis* to activate initiation rather than simply act as ORC binding sites. Whether these sequence elements are important at the level of pre-RC assembly, perhaps to establish a specific protein/DNA architecture to facilitate helicase loading in a manner analogous to the *S. cerevisiae* B2 element or to establish the nucleosome-free region at origins, remains to be determined.

MATERIALS AND METHODS

Yeast strains and growth conditions

Yeast strains used in this study are listed in Supplemental Table 1. For ChIP-Seq experiments, wild-type (W303 BLa) yeast cells were grown in rich medium. Cells were arrested either at G1 with 10 μg/ml α-factor, or at metaphase with 10 μg/ml nocodazole.

Protein purification

Yeast histone octamers were prepared as described previously (Luger et al. 1999). Nap1 protein and ISW1 complex were purified as described (Vary et al. 2004). Purified ORC was prepared from yeast strain F1ORC1 as described previously (Tsakraklides and Bell 2010).

Plasmids

The circular DNA substrate for *in vitro* chromatin assembly, pUC19 ARS1 1kb, was prepared by cloning a 1.1-kb fragment of yeast genomic DNA, which includes the *ARS1* replication origin oriented at the center, into the EcoRI/SphI sites of pUC19.

ChIP-Seq

Mcm2-7 chromatin immunoprecipitation was performed in extracts prepared from G1-arrested and G2/M arrested cells as described (Aparicio et al. 1997). Immunoprecipitated DNA fragments were prepared for sequencing using the genomic DNA sample preparation kit (Illumina) according to the manufacturer's protocol. The library was

sequenced using the Illumina 1G Genome Analyzer and analyzed as described previously (Eaton et al. 2010).

Preparation of whole-cell extracts (WCEs)

Yeast whole-cell extracts were prepared from strains ySC15 (Gal1,10-ORC1-6) and ySC17 (Gal1,10-ORC1-6 and Gal1,10-Cdc6) as described previously (Bowers et al., 2004), with minor modifications. Cells were grown in YP + 2% glycerol to OD_{600}=0.6. Galactose was added to a final concentration of 2% to induce overexpression of ORC and Cdc6. After 2 hours of induction, cells were arrested in μg/ml α-factor for 3 hours, and collected by centrifugation. Cells were frozen in liquid nitrogen and then lysed by grinding in a Retsch mortar grinder RM 200 for 20 minutes. The lysate was spun down by ultracentrifugation to clarify the extract.

Preparation of ARS1-DNA-coupled beads

Biotinylated *ARS1* DNA templates for pre-RC assembly were amplified by PCR using p*ARS1* or mutant versions of p*ARS1* (p*ARS1*/802-810 (B2-), p*ARS1*/757-764 (B3-), p*ARS1*/858-865 (ACS-)) as a template (Marahrens and Stillman 1992). Primer sets are listed in Supplemental Table 1. DNA was purified and coupled to streptavidin-magnetic beads (Dynal) by incubation overnight at room temperature. Beads were washed in Bind and Wash buffer [10 mM Tris-HCl pH 7.6, 1mM EDTA, 2mM NaCl] and resuspended in buffer H/300 mM potassium glutamate.

Pre-RC assembly assays

Pre-RC assembly assays were performed with 1 pmol linear *ARS1* DNA coupled to magnetic beads and 15 µl of ySC17 whole cell extract in 40 µl reactions. Reactions were incubated for 20 minutes at room temperature and washed three times with buffer H (50 mM HEPES [pH 7.6], 5 mM Mg-acetate, 1 mM EDTA, 1 mM EGTA, 10% glycerol, and 1mM DTT) with 300 mM potassium glutamate (buffer H/300). For salt extraction experiments, reactions were washed with buffer H/500 mM NaCl for 1 min. Pre-RC proteins were analyzed by immunoblotting using antiserum UM185 (MCM2-7), HM5352 (Cdt1), and monoclonal antibodies SB6 (Orc4) and 12CA5 (HA tag).

Reconstituted chromatin assembly onto pARS1 plasmids

Nucleosomes were assembled onto the pUC19 ARS1 1Kb plasmid essentially as described (Vary et al. 2004; Eaton et al. 2010). Briefly, 0.5 pmol pUC19 ARS1 1Kb plasmid DNA was incubated with 20 pmol histone octamers, 80 pmol Nap1 protein, 2.5 pmol ORC and ISW1 complex in the chromatin assembly buffer (10 mM HEPES-KOH [pH7.6], 50 mM KCl, 5mM $MgCl_2$, 0.5 mM EDTA, 10 % glycerol, 0.1 mg/ml BSA, 3 mM ATP, 30 mM phosphocreatine, 50 μg/ml creatine kinase) for 4.5 hours at 30 °C. Micrococcal nuclease (MNase) digestion was used to confirm the formation of regularly spaced nucleosomes as described previously (Vary et al. 2004).

Pre-RC assembly onto circular DNA templates

Nucleosomes were assembled onto *pARS1* plasmids as described above. For the assembly reaction with naked plasmid DNA, histones, Nap1 and ISW1 complexes were

omitted from the reaction. After the chromatin assembly reaction, the full reaction mixture (40 μl) was mixed with 85 μl pre-RC assembly reaction buffer (25 mM HEPES-KOH [pH 7.6], 10 mM Mg-acetate, 50 μM zinc acetate, 3 mM ATP, 125 μM dNTPs, 125 μM UTP/GTP/CTP, 1 mM DTT, 0.1 mM EDTA, 20 mM creatine phosphate, and 40 μg/ml creatine kinase), plus 40 μl YSC17 whole cell extract, and incubated for 15 min at room temperature. The reaction mixture was then loaded onto a 5 ml sephacryl S500 gel filtration column equilibrated with buffer H/100 mM potassium glutamate. Fractions (240 μl) were analyzed for pre-RC formation by immunoblotting and by MNase digestion to check chromatin formation.

Nucleosome Mapping

Nucleosomes were assembled onto *ARS1* plasmids in the presence as described previously (Vary et al. 2004; Eaton et al. 2010). Mononucleosomal DNA fragments were isolated after MNase digestion and sequenced by high-throughput multiplex sequencing.

Supplemental Table 1. Yeast strains and oligonucleotides

Strain Name	Genotype	Reference
W303	ade2-1, ura3-1, his3-11,15, trp1-1, leu2-3, 112 can1-100 lys2::hisG MATa	
ySC15	W303 bar1::hisG pep4::KanMX trp1::p404-GAL1-10-ORC3,4 lys2::plys2-GAL1-10-ORC2,5 his3::p403-GAL1-10-ORC1,6	Bowers et al. 2004
ySC17	W303 bar1::hisG pep4::KanMX trp1::p404-GAL1-10-ORC3,4 lys2::plys2-GAL1-10-ORC2,5 his3::p403-GAL1-10-ORC1,6 ura3::pSF322CDC6-HA	Bowers et al. 2004
Oligonucleotides		
ARS1-3bio	5' -biotin-CTGTTTTGTCCTTGGAAAAAA AGCACTACC- 3'	Bowers et al., 2004
B1+22	5'-TGTGCACTTGCCTGCAGGCCTTTTGAA AAGC-3'	This study
B1+38	5'-TTATTTAAGTATTGTTTGTGCACTTGCC TGC-3'	This study
B1+56	5'- GGTTATTACTGAGTAGTATTTATTTAA GTATTGTTTGTGC-3'	This study
B1+90	5'- AATAGCAAATTTCGTCAAAAATGC TAAGAA-3'	This study
B1+100	5'- GAGGATCCCCACTCTAACAAAATAGC-3'	This study
B1+110	5'- GAGGATCCCCACTCTAACAAAATAGC-3'	This study
B1+115	5'- CTCTAGAGGATCCCCACTCTAACAA-3'	This study
B1+120	5'- GTCGACTCTAGAGGATCCCCACTCTAA-3'	This study
B1+126	5'- CTGCAGGTCGACTCTAGAGGATCCCC ACTC-3'	This study
B1+141	5'- GCCAAGCTTGCATGCCTGCAGGTCGA CTCTAG-3'	This study
B1+188	5'-AGCGGATAACAATTTCACACAGG- 3'	This study
B1+320	5'-AGCTCTGGCACGACAGGTTTC'3'	This study
B1-813bio	5'-biotin-TCTTCGAAGAGTAAAAAATTGTA CTTGGC-3'	This study
ACS-22	5'- TCTTTACATCTTGTTATTTTACAGAT-3'	This study
ACS-38	5'- AAATGATTTAGCATTATCTTTACATC TTG-3'	This study

ACS-56	5'- CACAATCAATCAAAAAGCCAAATGA TTTAGC-3'	This study
ACS-90	5'- AAAACGACGGCCAGTGAATTCG-3'	This study
ACS-126	5'- TGGGTAACGCCAGGGTTTTCCCAG TCAC-3'	This study
ACS-141	5'- GCAAGGCGATTAAGTTGGGTAA-3'	This study
ACS-188		This study
ACS-320	5'- CCTTCAGTAGCTGCCCCTTTAAAGT CAGC-3'	This study
307 ACS-813bio	5'-biotin-AAAATAGGCATTATAGATCAGTT CGAG-3'	This study
307 B1+22	5'-ATATTGATCCTCTCTCTTCTTTATTT TCTGCCAG-3'	This study
307 B1+56	5'-TTCACACAGGAAACAGCTATGACC ATGATTA-3'	This study
307 B1+90	5'-ATGTTGAGTTGCTAACTGCCTGA-3'	This study
307 B1+126	5'-CTTCTTCTCATGCACTAACAAGTG -3'	This study
307 B1+188	5'-GGTTTTCCCAGTCACGACGTTGTAA AACGAC-3'	This study
A/B1 ACS-816bio	5'-biotin- GTTACCGGATAAGGCGCAGCG GTCG-3'	This study
A/B1 B1+38	5'- GGGCCCCCCCTCGAGGTCGACGGTA TCGATAA-3'	This study

REFERENCES

Aladjem, M.I., Rodewald, L.W., Kolman, J.L., and Wahl, G.M. 1998. Genetic dissection of a mammalian replicator in the human beta-globin locus. *Science* **281**(5379): 1005-1009.

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**(1): 59-69.

Austin, R.J., Orr-Weaver, T.L., and Bell, S.P. 1999. Drosophila ORC specifically binds to ACE3, an origin of DNA replication control element. *Genes Dev* **13**(20): 2639-2649.

Bell, S.P. and Stillman, B. 1992. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* **357**(6374): 128-134.

Bowers, J.L., Randell, J.C., 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**(6): 967-978.

Celniker, S.E., Sweder, K., Srienc, F., Bailey, J.E., and Campbell, J.L. 1984. Deletion mutations affecting autonomously replicating sequence ARS1 of Saccharomyces cerevisiae. *Mol Cell Biol* **4**(11): 2455-2466.

Clarey, M.G., Botchan, M., and Nogales, E. 2008. Single particle EM studies of the Drosophila melanogaster origin recognition complex and evidence for DNA wrapping. *J Struct Biol* **164**(3): 241-249.

Diffley, J.F. and Stillman, B. 1988. Purification of a yeast protein that binds to origins of DNA replication and a transcriptional silencer. *Proc Natl Acad Sci U S A* **85**(7): 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**(11): 5611-5616.

Dubey, D.D., Kim, S.M., Todorov, I.T., and Huberman, J.A. 1996. Large, complex modular structure of a fission yeast DNA replication origin. *Curr Biol* **6**(4): 467-473.

Eaton, M.L., Galani, K., Kang, S., Bell, S.P., and MacAlpine, D.M. 2010. Conserved nucleosome positioning defines replication origins. *Genes Dev* **24**(8): 748-753.

Huang, R.Y. and Kowalski, D. 1996. Multiple DNA elements in ARS305 determine replication origin activity in a yeast chromosome. *Nucleic Acids Res* **24**(5): 816-823.

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**(12): 7159-7168.

Lin, S. and Kowalski, D. 1997. Functional equivalency and diversity of cis-acting elements among yeast replication origins. *Mol Cell Biol* **17**(9): 5473-5484.

Lu, L., Zhang, H., and Tower, J. 2001. Functionally distinct, sequence-specific replicator and origin elements are required for Drosophila chorion gene amplification. *Genes Dev* **15**(2): 134-146.

Luger, K., Rechsteiner, T.J., and Richmond, T.J. 1999. Expression and purification of recombinant histones and nucleosome reconstitution. *Methods Mol Biol* **119**: 1-16.

Mahbubani, H.M., Chong, J.P., Chevalier, S., Thommes, P., and Blow, J.J. 1997. Cell cycle regulation of the replication licensing system: involvement of a Cdk-dependent inhibitor. *J Cell Biol* **136**(1): 125-135.

Marahrens, Y. and Stillman, B. 1992. A yeast chromosomal origin of DNA replication defined by multiple functional elements. *Science* **255**(5046): 817-823.

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* **294**(5540): 115-121.

Randell, J.C., Bowers, J.L., Rodriguez, H.K., and Bell, S.P. 2006. Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase. *Mol Cell* **21**(1): 29-39.

Rao, H., Marahrens, Y., and Stillman, B. 1994. Functional conservation of multiple elements in yeast chromosomal replicators. *Mol Cell Biol* **14**(11): 7643-7651.

Rao, H. and Stillman, B. 1995. The origin recognition complex interacts with a bipartite DNA binding site within yeast replicators. *Proc Natl Acad Sci U S A* **92**(6): 2224-2228.

Remus, D., Beuron, F., Tolun, G., Griffith, J.D., Morris, E.P., and Diffley, J.F. 2009. Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell* **139**(4): 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**(11): 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**(26): 14115-14120.

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.*

Srienc, F., Bailey, J.E., and Campbell, J.L. 1985. Effect of ARS1 mutations on chromosome stability in Saccharomyces cerevisiae. *Mol Cell Biol* **5**(7): 1676-1684.

Strich, R., Woontner, M., and Scott, J.F. 1986. Mutations in ARS1 increase the rate of simple loss of plasmids in Saccharomyces cerevisiae. *Yeast* **2**(3): 169-178.

Takahashi, T., Ohara, E., Nishitani, H., and Masukata, H. 2003. Multiple ORC-binding sites are required for efficient MCM loading and origin firing in fission yeast. *Embo J* **22**(4): 964-974.

Theis, J.F. and Newlon, C.S. 1994. Domain B of ARS307 contains two functional elements and contributes to chromosomal replication origin function. *Mol Cell Biol* **14**(11): 7652-7659.

Tsakraklides, V. and Bell, S.P. 2010. Dynamics of pre-replicative complex assembly. *J Biol Chem* **285**(13): 9437-9443.

Van Houten, J.V. and Newlon, C.S. 1990. Mutational analysis of the consensus sequence of a replication origin from yeast chromosome III. *Mol Cell Biol* **10**(8): 3917-3925.

Vary, J.C., Jr., Fazzio, T.G., and Tsukiyama, T. 2004. Assembly of yeast chromatin using ISWI complexes. *Methods Enzymol* **375**: 88-102.

Venditti, P., Costanzo, G., Negri, R., and Camilloni, G. 1994. ABFI contributes to the chromatin organization of Saccharomyces cerevisiae ARS1 B-domain. *Biochim Biophys Acta* **1219**(3): 677-689.

Walter, J. and Newport, J.W. 1997. Regulation of replicon size in Xenopus egg extracts. *Science* **275**(5302): 993-995.

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**(1): 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**(5550): 2357-2360.

Xu, W., Aparicio, J.G., Aparicio, O.M., and Tavare, 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**: 276.

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**(9): 3086-3096.

Chapter III

Discussion

DISCUSSION

Key Conclusions

The studies described in this thesis address several longstanding questions regarding how origin DNA sequence and chromatin structure influence pre-RC assembly at origins of replication in *S. cerevisiae*. In chapter two, we identified the DNA length and sequence requirements for the stepwise process of pre-RC assembly. Genome-wide mapping of Mcm2-7 occupancy at origins revealed that that Mcm2-7 complexes are predominantly loaded downstream but not upstream of the ORC binding site *in vivo*. This region coincides with the nucleosome-free region (NFR) associated with each origin (Eaton et al. 2010). By examining a series of deletions in this ORC-adjacent downstream region, we determined the lengths necessary for the steps involved in helicase loading. The amount of ORC-adjacent DNA was dispensable for the initial recruitment of the Mcm2-7 helicase to origin DNA. In contrast, the amount of ORC-adjacent DNA was crucial for the subsequent helicase loading step.

We showed that a key sequence element in this ORC-adjacent region, B2, is important for facilitating Mcm2-7 loading at the origin. Previous studies had implicated a role for the B2 element in pre-RC assembly *in vivo* (Zou and Stillman 2000; Wilmes and Bell 2002), but whether it functions in the initial recruitment of the helicase to the origin or in the stable loading of the helicase onto DNA had not been determined. We found that the B2 element was important specifically for the helicase loading step but not for the initial association of the helicase to the origin. Interestingly, the B2 element was

important for helicase loading only under conditions that restricted the accessibility of ORC-adjacent DNA. The assembly of nucleosomes onto origin DNA as well as truncation of the ORC-adjacent region revealed a dependence on the B2 element for maximal helicase loading. Together, these data suggest a role for the B2 element in facilitating the stable loading of Mcm2-7 hexamers by increasing the accessibility of ORC-adjacent DNA.

Architecture of the Pre-RC

One of the major questions regarding pre-RC assembly that remains, for the most part, unanswered to date is the spatial arrangement of the pre-RC components on origin DNA. Elucidating the architecture of the pre-RC is a prerequisite to understanding the mechanistic details of helicase loading. The development of an *in vitro* assay for pre-RC assembly enabled the identification of the two major steps involved in pre-RC assembly (ATP hydrolysis-independent helicase association and ATP-dependent helicase loading) (Bowers et al. 2004; Randell et al. 2006). The recent observation that *S. cerevisiae* Mcm2-7 complexes form a head-to-head double hexamer on origin DNA *in vitro* has shed some light on the architecture of the final complex once pre-RC formation is complete (Remus et al. 2009). Little is understood, however, about how the protein components of the pre-RC interact with each other and with the origin DNA prior to and during the steps of pre-RC assembly.

Several previous studies have provided some clues into how ORC and Cdc6 are engaged with origin DNA. Although it is clear that ORC interacts with the ACS and B1 elements within origins (Bell and Stillman 1992), the evidence for ORC making significant contacts with regions beyond these two elements is less definitive. The spatial arrangement of ORC subunits over the origin remains controversial, even though EM and crosslinking studies have provided some clues. Crosslinking studies of ORC bound to origin DNA suggest an asymmetric clustering of ORC subunits over the ACS and B1, as well as possible contacts of at least three ORC subunits (Orc2, Orc4, and Orc6) with the downstream adjacent region that includes the B2 element (Lee and Bell 1997). An EM study examining the multisubunit arrangement of ORC proposed a model in which the Orc1-5 and Cdc6 form a ring-shaped structure to facilitate interactions with the Mcm2-7 helicase ring, and that all six WHD-containing subunits contact the DNA (Chen et al. 2008), but the evidence for this spatial arrangement is not substantive. Whether Cdc6 interacts directly with DNA and how it contacts the remaining protein components of the pre-RC remain to be determined.

Our study has shed light on where the Mcm2-7 helicase interacts with origin DNA during G1 phase prior to origin activation. We observed an enrichment of Mcm2-7 binding primarily in the region downstream of ORC-binding *in vivo* (Fig. 1B), but the resolution of our ChIP-seq method was not sufficiently high to determine the precise sequences within the region that are bound by Mcm2-7 complexes. In addition, the ChIP-seq data cannot distinguish between the different ways that Mcm2-7 hexamers can associate with the origin (i.e. salt-sensitive association vs. salt-resistant topological

linkage), which makes it difficult to determine the precise location of stably loaded complexes by this method alone. Our data strongly suggest that at least one complex is loaded immediately adjacent to ORC (Fig. 3A), but where the remaining Mcm2-7 complexes are loaded could not be determined by our assays alone. The most straightforward way to resolve this problem would be to map the location of Mcm2-7 binding sites by DNase I footprinting. However, this has proven to be problematic for at least two reasons. One, loaded Mcm2-7 complexes are relatively dynamic *in vitro* (Remus et al. 2009). Two, our method for *in vitro* pre-RC assembly is not efficient enough to achieve the stoichiometric binding of Mcm2-7 complexes to DNA required to observe a footprint by DNase I mapping. Alternative methods to map Mcm2-7 binding sites include Exo III footprinting and chromatin endogenous cleavage (ChEC) (Schmid et al. 2004). Although the resolution of both methods is lower than that of DNase I footprinting, neither method requires stoichiometric binding of the protein of interest to DNA.

In addition to our limited understanding of how pre-RC components are engaged with origin DNA, the stoichiometry of the different pre-RC components bound to the origin remains poorly understood. Although the predominant thought is that one molecule of ORC is bound per origin, there is evidence supporting a model with two molecules of ORC bound at the origin. DNase I protection assays have demonstrated that when the ACS in *ARS1* is mutated, ORC binds to the B2 element (Bell and Stillman 1992). Furthermore, ORC will bind to the ACS and B2 simultaneously if the smallest ORC subunit, Orc6, is deleted (Wilmes and Bell 2002). In both cases, ORC bound to B2

is in the opposite orientation relative to the origin compared to when it is bound to the ACS. Although this has yet to be demonstrated to be the case *in vivo*, we cannot exclude the possibility that two molecules of ORC are bound to each origin. Such an arrangement could provide an interesting explanation for how two Mcm2-7 hexamers are loaded in a concerted fashion at each origin.

Despite previous attempts at quantifying the number of Mcm2-7 complexes loaded *in vitro*, it is unclear how many Mcm2-7 hexamers are bound per molecule of ORC. Quantification of Mcm2-7 complexes loaded per DNA molecule by quantitative immunoblotting is difficult in our population-based assay because of the inefficiency of pre-RC assembly. Although every bead-bound DNA fragment is bound by ORC, it is clear that not all fragments are bound by Mcm2-7 complexes; the efficiency of Mcm2-7 loading in our assay is typically between 10-20%. The current estimate is 2-4 Mcm2-7 hexamers loaded per origin *in vitro*, but in rare cases, the loading of six hexamers per origin has been observed (Remus et al. 2009).

One way to address the questions regarding the number and location of Mcm2-7 complexes at the origin would be to use a single-molecule approach in combination with our current method for *in vitro* pre-RC assembly. Recent single-molecule EM analysis of reconstituted pre-RC assembly on DNA strongly suggested that only one head-to-head dimer of Mcm2-7 hexamers is loaded per origin (Remus et al. 2009). However, the main pitfall of this experiment was that ORC and Cdc6 could not be visualized bound to origin DNA. Thus it was unclear where the Mcm2-7 double hexamer was bound relative to the

ORC binding site. The use of nanoscale biopointers (short DNA fragments with a single streptavidin tetramer bound to one end; the DNA fragment is used for visualization purposes only) targeting biotinylated ORC could be used to resolve the problem of locating ORC on origin DNA (Chastain et al. 2004).

Role of the B2 element in helicase loading

In our study, we demonstrated a role for the B2 element in the helicase loading step of pre-RC assembly *in vitro*. Now that we have recapitulated the B2 effect in our *in vitro* assay, we can use this tool to probe more deeply into the precise function of B2 in the mechanism of helicase loading. As discussed in chapter two, we were only able to observe the importance of B2 when we limited the amount of ORC-adjacent DNA accessible for Mcm2-7 loading, either by assembling nucleosomes onto the DNA prior to pre-RC assembly or by simple truncation of the DNA. We hypothesize that B2 facilitates the accessibility of this ORC-adjacent region by making additional contacts with ORC and/or Cdc6 to generate a structure, such as a DNA wrap around ORC, that is conducive to helicase loading.

The most definitive way to determine whether DNA wrapping about ORC occurs and whether the B2 element facilitates DNA wrapping would be to use a structural approach. This is extremely ambitious, however, especially considering the fact that visualizing DNA-bound ScORC by EM has been problematic in previous attempts (Evrin

et al. 2009; Remus et al. 2009). Alternatively, several well-established biochemical approaches could be used to address the DNA wrapping issue. In a previous study, purified *S. cerevisiae* ORC added to ~200 bp *ARS1* linear DNA fragments in the presence of ligase was shown to circularize DNA, suggesting that ORC generates a bend in origin DNA *in vitro* (Lee and Bell 1997). Experiments to test whether the B2 element is necessary for this DNA bending activity by ORC are currently underway. Previous DNase I protection assays of ORC bound to different origins showed a series of hypersensitive cleavage sites with a 10-bp periodicity, consistent with DNA wrapping around ORC (Bell and Stillman 1992; Rao and Stillman 1995). An interesting experiment would be to test whether ORC bound to an *ARS1* fragment containing a mutation in B2 would lose the 10-bp periodicity of hypersensitive sites spanning the B1 and B2 elements (Bell and Stillman 1992). It has been demonstrated that mutating B2 does not affect ORC binding or the pattern of DNase I hypersensitive sites observed; however, the DNA fragments used in these experiments contained ~100 bp of B1-adjacent B-side sequence. Our observation that B2 is important when the amount of ORC-adjacent DNA is restricted suggests that limiting the amount of ORC-adjacent B-side DNA in the DNA fragments subjected to DNase I cleavage could affect this pattern of hypersensitive cleavage at B1 and B2.

Two approaches that have been used to examine DNA wrapping by proteins in other organisms could also be adapted to examine ScORC. First, an adaptation of a previous single-molecule EM analysis of *D. melanogaster* ORC bound to a DNA fragment containing the *ACE3* element could be used to examine *S. cerevisiae* ORC. In

the original study, ORC binding to a DNA fragment led to shortening of the fragment by ~130 bp (Clarey et al. 2008), consistent with the DNA wrapping hypothesis. The use of nanoscale biopointers targeting ScORC could be used to resolve whether ORC is bound to the origin, assuming that visualization of DNA-bound ORC will be problematic as it has been in the past (Chastain et al. 2004). Second, a FRET approach similar to one developed to examine the kinetics of DNA bending by *E. coli* IHF could also be adapted to study ORC (Kuznetsov et al. 2006; Sugimura and Crothers 2006). It is important to keep in mind, however, that the severe DNA bend induced by IHF greatly facilitated the detection of DNA bending by FRET, and thus the success of this method will depend on whether ORC can dramatically bend origin DNA.

If ORC is indeed interacting with downstream DNA and the B2 element, which subunits of ORC are responsible for mediating this interaction? Orc2, Orc4, and Orc6 are three potential candidates. Crosslinking studies of ORC bound to *ARS1* showed crosslinking of Orc2, Orc4, and Orc6 to the region spanning the B1 and B2 elements (Lee and Bell 1997). Interestingly, Orc6 is not thought to interact with DNA, and is the only ORC subunit lacking homology to a DNA-binding winged helix domain (Speck et al. 2005). Furthermore, an ORC complex lacking Orc6 is fully capable of binding *ARS1* DNA (Wilmes and Bell 2002). It is possible, however, that Orc6 is dispensable for ORC binding to the ACS and B1 but is important for ORC's interaction with downstream DNA. As mentioned earlier, in the absence of Orc6, two complexes of Orc1-5 can bind simultaneously to the ACS and B2 in *ARS1 in vitro* (Wilmes and Bell 2002). It is unclear whether one or two molecules of ORC are bound per origin *in vivo*, but one interesting

way to interpret this observation is that Orc6 is important for the wrapping of downstream DNA around ORC, and that this wrapping prevents the binding of both nucleosomes and a second molecule of ORC to this downstream region.

Based on this study alone, we cannot rule out the possibility that B2 simply acts as a binding site for Cdc6 or the Mcm2-7 complex. Both possibilities are consistent with the previous observation introducing a mutant B2 sequence into the *ARS1* origin *in vivo* reduces the amount of Mcm2-7 association (Zou and Stillman 2000; Wilmes and Bell 2002), and that the plasmid loss defect caused by mutating the B2 element can be rescued by overexpression of Cdc6 (Wilmes and Bell 2002). Consistent with this hypothesis, the C-terminus of Cdc6 contains a putative winged-helix domain but has not been shown to have any affinity for DNA on its own. Also, purified Cdc6 and ORC together produce an expanded region of DNase I protection that extends into the B2 element (Speck et al. 2005). One interpretation of this observation is that Cdc6 binds directly the B2 element. However, it is equally likely that Cdc6 binding induces a conformational change in ORC that allows it to make additional contacts with the B2 region.

Origin function in higher eukaryotes

Detailed analysis of the function of origin sequences in other eukaryotes besides *S. cerevisiae* has been challenging for multiple reasons. The lack of well-defined origins, and in some extreme cases the lack of any sequence specificity of origins, has made it

difficult to identify functional sequence elements analogous to the *S. cerevisiae* ACS and B-elements in other eukaryotes. In most cases, chromosomal replication in metazoans begins in large zones of initiation, in which replication can initiate from one of many potential sites, or in some cases, at apparently random locations (Bogan et al. 2000). This redundancy of replication initiation sites has made it difficult to examine the effect of *cis*-acting sequences on initiation, as deletion of sequences within known origins often has no effect on the overall efficiency of initiation (Cvetic and Walter 2005). A few rare examples of *cis*-acting sequence elements have been identified in metazoan origins. The origins in which these sequence elements were identified, however, are atypical metazoan origins in that they tend to be smaller in size and contain a single discrete start site for DNA synthesis (Kitsberg et al. 1993; Lu et al. 2001), reminiscent of *S. cerevisiae* origin structure.

The development of a method to recruit ORC to plasmid DNA to create an artificial origin of replication offers a promising tool to examine the role of mammalian *cis*-acting elements on replication initiation. In this assay, a GAL4-ORC fusion protein is recruited to plasmid DNA containing a tandem array of GAL4-binding sites, creating a functional origin of replication (Takeda et al. 2005). Although this assay was used originally to characterize the effect of point mutations and deletions in various pre-RC components, the plasmid could be manipulated to test the effect of *cis*-acting sequences (such as ORC-adjacent DNA sequences) and chromatin structure, on pre-RC assembly and origin activity.

REFERENCES

Bell, S.P. and Stillman, B. 1992. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* **357**(6374): 128-134.

Bogan, J.A., Natale, D.A., and Depamphilis, M.L. 2000. Initiation of eukaryotic DNA replication: conservative or liberal? *J Cell Physiol* **184**(2): 139-150.

Bowers, J.L., Randell, J.C., 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**(6): 967-978.

Chastain, P.D., 2nd, Bowers, J.L., Lee, D.G., Bell, S.P., and Griffith, J.D. 2004. Mapping subunit location on the Saccharomyces cerevisiae origin recognition complex free and bound to DNA using a novel nanoscale biopointer. *J Biol Chem* **279**(35): 36354-36362.

Chen, Z., Speck, C., Wendel, P., Tang, C., Stillman, B., and Li, H. 2008. The architecture of the DNA replication origin recognition complex in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **105**(30): 10326-10331.

Clarey, M.G., Botchan, M., and Nogales, E. 2008. Single particle EM studies of the Drosophila melanogaster origin recognition complex and evidence for DNA wrapping. *J Struct Biol* **164**(3): 241-249.

Cvetic, C. and Walter, J.C. 2005. Eukaryotic origins of DNA replication: could you please be more specific? *Semin Cell Dev Biol* **16**(3): 343-353.

Eaton, M.L., Galani, K., Kang, S., Bell, S.P., and MacAlpine, D.M. 2010. Conserved nucleosome positioning defines replication origins. *Genes Dev* **24**(8): 748-753.

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**(48): 20240-20245.

Kitsberg, D., Selig, S., Keshet, I., and Cedar, H. 1993. Replication structure of the human beta-globin gene domain. *Nature* **366**(6455): 588-590.

Kuznetsov, S.V., Sugimura, S., Vivas, P., Crothers, D.M., and Ansari, A. 2006. Direct observation of DNA bending/unbending kinetics in complex with DNA-bending protein IHF. *Proc Natl Acad Sci U S A* **103**(49): 18515-18520.

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**(12): 7159-7168.

Lu, L., Zhang, H., and Tower, J. 2001. Functionally distinct, sequence-specific replicator and origin elements are required for Drosophila chorion gene amplification. *Genes Dev* **15**(2): 134-146.

Randell, J.C., Bowers, J.L., Rodriguez, H.K., and Bell, S.P. 2006. Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase. *Mol Cell* **21**(1): 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 U S A* **92**(6): 2224-2228.

Remus, D., Beuron, F., Tolun, G., Griffith, J.D., Morris, E.P., and Diffley, J.F. 2009. Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell* **139**(4): 719-730.

Schmid, M., Durussel, T., and Laemmli, U.K. 2004. ChIC and ChEC; genomic mapping of chromatin proteins. *Mol Cell* **16**(1): 147-157.

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.*

Sugimura, S. and Crothers, D.M. 2006. Stepwise binding and bending of DNA by Escherichia coli integration host factor. *Proc Natl Acad Sci U S A* **103**(49): 18510-18514.

Takeda, D.Y., Shibata, Y., Parvin, J.D., and Dutta, A. 2005. Recruitment of ORC or CDC6 to DNA is sufficient to create an artificial origin of replication in mammalian cells. *Genes Dev* **19**(23): 2827-2836.

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**(1): 101-106.

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**(9): 3086-3096.