The Roles of the Helicase Double-Hexamer Complex and the ssDNA-Binding
Protein RPA During Eukaryotic DNA Replication

by

Caitlin M. Friend

B.A. Biology
Amherst College, 2014

Submitted to the Microbiology Graduate Program in partial fulfillment
of the requirement for the degree of

Doctor of Philosophy
at the
Massachusetts Institute of Technology

May 2020

© 2020 Caitlin M. Friend. All right reserved.

The author hereby grants MIT the permission to reproduce and to distribute publicly
paper and electronic copies of this thesis document in whole or in part in any medium
now known or hereafter created.

Signature of Author: ____________________________________________

Caitlin M. Friend
Department of Biology
May 20, 2020

Certified by: ________________________________________________

Stephen P. Bell
Uncas and Helen Whitaker Professor of Biology
Investigator, Howard Hughes Medical Institute
Thesis Supervisor

Accepted by: ________________________________________________

Jaccquin C. Niles
Professor of Biological Engineering
Co-Director, Microbiology Graduate Program
The Roles of the Helicase Double-Hexamer Complex and the ssDNA-Binding Protein RPA During Eukaryotic DNA Replication

by

Caitlin M. Friend

Submitted to the Microbiology Graduate Program on May 20, 2020 in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biology

ABSTRACT

Eukaryotic DNA replication is a complex process that must occur accurately, completely, and only once per cell cycle. To accomplish these goals, the events of DNA replication are tightly coupled to cell-cycle progression. Origins of replication are licensed by loading of the Mcm2-7 replicative DNA helicase during G1. Two Mcm2-7 hexamers load onto each origin as a double hexamer encircling dsDNA. At this stage, the helicases are inactive. Upon entry into S phase, loaded Mcm2-7 complexes then recruit a number of other replication proteins that activate the helicase. Helicase activation results in separation of the double hexamer, a transition to encircling ssDNA, and initiation of DNA unwinding. Once activated, the helicase produces the ssDNA that acts as template for new DNA synthesis. Helicase activation is the committed step of DNA replication after which the cell must complete genome duplication before it can segregate its chromosomes and divide.

The work described in this thesis focuses on mechanisms that are essential for eukaryotic DNA replication with a focus on DNA unwinding and DNA synthesis. In Chapter II, I explore the essential functions and purpose of the double-hexamer conformation of the loaded helicases. Using a helicase mutant that loads as two single hexamers, I show that initial origin DNA melting can occur in the context of a single-hexamer helicase. Importantly, the amount of unwinding that occurs within a single helicase is not sufficient to allow the transition onto ssDNA. Further DNA unwinding and subsequent DNA synthesis requires robust double-hexamer helicase interactions. Together, my findings strongly suggest that the double-hexamer conformation is essential to complete helicase activation.

In Chapter III, I explore the role and specificity of ssDNA-binding proteins (SSBs) in eukaryotic DNA replication. To this end, I substituted the eukaryotic SSB RPA with SSBs from other systems: E. coli SSB (EcSSB) and T4 bacteriophage Gp32. I find that DNA unwinding is supported by RPA and EcSSB but not Gp32, suggesting that eukaryotic DNA unwinding requires at least one SSB function beyond ssDNA binding. Although both RPA and EcSSB support DNA synthesis, we only observed robust lagging-strand synthesis in the presence of RPA. My studies indicate that RPA must perform multiple functions beyond ssDNA binding to facilitate eukaryotic DNA replication.

Thesis Supervisor: Stephen P. Bell
Title: Uncas and Helen Whitaker Professor of Biology
ACKNOWLEDGEMENTS

To Chris, thank you for your endless faith in me. You have been my rock. You deserve your own PhD for how many practice talks and biology ramblings you've listened to over the years.

I would like to thank my friends and family. Even when you didn't understand my research, you always supported my dreams. And yes, the yeast are doing well.

Bell Lab, I would like to thank the lab members past and present who have helped me develop as a scientist. You have all made our lab into the kindest, most supportive one on campus.

Kate, you helped me survive the toughest years of my life. Thank you for always being the calm in my stormy days.

Jill, thank you for nurturing my love of biology and pushing me to be a better scientist since day one. So many of the skills you taught me as an undergraduate researcher have helped me to succeed in my PhD.

I would especially like to thank Steve, my thesis advisor, for taking a chance on a genetic-loving, microbiologist. In addition to all you've taught me about being a good scientist, I've learned to love the simple elegance of biochemistry. Thank you for your guidance and support over the years.

Thank you to my committee, Alan Grossman and Mike Laub. You were my first introduction to MIT way back during my graduate school interviews, and I am grateful that you've been a constant presence throughout my PhD. Thank you for all your advice about graduate school in general and your invaluable feedback on my research specifically.
# TABLE OF CONTENTS

Abstract .................................................................................................................................................. 3  
Acknowledgements ................................................................................................................................. 5  
Table of Contents ................................................................................................................................. 6  

**CHAPTER I: INTRODUCTION** .......................................................................................................... 9  
Overview ............................................................................................................................................... 10  
The Eukaryotic Cell Cycle ..................................................................................................................... 12  
Mcm2-7 Helicase .................................................................................................................................. 18  
Origin Structure and Helicase Loading .................................................................................................. 21  
Helicase Activation ............................................................................................................................... 24  
DNA Synthesis ...................................................................................................................................... 30  
RPA and Other ssDNA-Binding Proteins ............................................................................................... 35  
Biochemical Studies of DNA Replication ............................................................................................... 40  
SV-40 Studies of Replication ............................................................................................................... 44  
Thesis Summary ................................................................................................................................... 45  
References ............................................................................................................................................ 47  

**CHAPTER II: A CONSERVED MCM4 MOTIF IS REQUIRED FOR MCM2-7 DOUBLE-HEXAMER FORMATION AND ORIGIN DNA UNWINDING** ........................................................................ 57  
Abstract ............................................................................................................................................... 59  
Introduction .......................................................................................................................................... 59  
Results .................................................................................................................................................. 63  
  N-terminal deletions of Mcm2 and Mcm4 inhibit Mcm2-7 loading ...................................................... 63  
  A conserved Mcm4 N-terminal motif is required for stable double-hexamer formation ................ 67  
  The DoHM is required for Mcm2-7 double-hexamer stability ......................................................... 71  
  Mcm2-7-Mcm2-7 interactions occur rapidly or not at all ................................................................. 80  
  Double-hexamer formation is required for later steps of replication initiation ............................ 87  
Discussion ............................................................................................................................................. 88  
  Double-hexamer formation is a facilitated event ............................................................................. 92  
  Structure of double-hexamer and the DoHM interactions ............................................................ 96
The double hexamer is required for extensive DNA unwinding

Materials and Methods

References

CHAPTER III: THE ROLE OF ssDNA-BINDING PROTEINS IN EUKARYOTIC DNA REPLICATION

Abstract

Introduction

Results

RPA, Gp32, and EcSSB ssDNA binding

RPA is necessary for DNA unwinding and DNA synthesis

Gp32 does not support DNA unwinding or DNA synthesis

EcSSB supports DNA unwinding and altered DNA synthesis

Discussion

ssDNA binding is important for eukaryotic DNA unwinding

How do SSBs facilitate eukaryotic DNA unwinding?

Eukaryotic DNA replication requires RPA for lagging-strand synthesis

RPA specificity during lagging-strand synthesis: Priming inhibition

RPA specificity during lagging-strand synthesis: DNA synthesis inhibition

Methods

References

CHAPTER IV: EXPANDED DISCUSSION AND FUTURE DIRECTIONS

Key conclusions

Requirements for DNA unwinding: The Mcm2-7 helicase complex

Requirements for DNA unwinding: ssDNA-binding protein (SSB)

An SSB requirement for DNA synthesis

References
Chapter I: Introduction
Overview

DNA replication is a crucial process for the survival of all life on Earth. The genomic DNA provides a complete set of instructions for how a cell should function under all conditions. Thus, each time a cell divides, it must make a complete copy of its genome so both daughter cells can perform all the appropriate cellular processes to survive and reproduce. Numerous proteins and mechanisms are required to ensure that the genome is completely duplicated with minimal errors and that each daughter cell receives exactly one copy of the genome.

DNA replication is a complex operation that presents numerous opportunities for damaging or fatal mistakes (reviewed in Jackson et al., 2014). Replication errors can result in DNA damage and/or genome instability, which can lead to cell death or mutations in the DNA that cause disease or uncontrolled cell growth (i.e. cancer, reviewed in Blow & Gillespie, 2008). To ensure that this process occurs completely, accurately, and in a timely fashion, DNA replication is highly regulated and tightly linked to the transitions of the cell cycle (reviewed in Boyer et al., 2016). In addition, a complex system of surveillance mechanisms and checkpoints verifies that DNA replication proceeds properly and finishes completely before moving on to the next steps in the cell cycle: chromosome segregation and cell division. Remarkably, DNA replication occurs exactly once per cell cycle and with few errors the vast majority of the time (Ganai & Johansson, 2016; Preston et al., 2010).

Broadly, eukaryotic DNA replication begins with loading of two DNA helicase onto origins
Origins are the specific genomic sites where DNA replication initiates. The loaded helicases are activated to initiate DNA unwinding, which produces the single-stranded DNA (ssDNA) used as template to synthesize new DNA strands. The full replisome machinery is then assembled around the activated helicases to duplicate the DNA flanking the origin. Eukaryotes have multiple origins per chromosome to allow many replisomes to act in parallel to duplicate the entire chromosome (Callan, 1973). Each replisome synthesizes DNA at a rate of ~ 1600 base pairs (bp) per minute (Sekedat et al., 2010). The longest budding yeast chromosome (IV) has 1.5 million bp (Jacq et al., 1997), so if this chromosome had only one, centrally-located origin (two replisomes) it would take at least 8 hours for DNA replication to finish. Instead, yeast origins are spaced every 20 – 40 kb allowing replication to finish in as little as 20 minutes (reviewed in Barberis et al., 2010).

My thesis work has explored the mechanism of initial DNA unwinding during eukaryotic replication initiation. Initial DNA unwinding is a crucial step in helicase activation. Once helicase activation begins, the cell is committed to both completing genome duplication and cell division. In this introduction, I will first provide an overview of the eukaryotic cell cycle and how DNA replication is integrated into and regulated by these events. Then, I describe our current understanding of how eukaryotic cells first load inactive helicases onto DNA, activate these loaded helicases to initiate DNA unwinding, and finally recruit the remaining replisome machinery to synthesize DNA. In addition, I will present what is currently known about the role of ssDNA-binding proteins during DNA replication, as these proteins and their function in replication is a major focus of my work. Finally, I
discuss the biochemical assays for eukaryotic DNA replication that I used in my work.

**The Eukaryotic Cell Cycle**

Eukaryotic cellular duplication is accomplished via a sequence of highly-regulated and coordinated events called the mitotic cell cycle. During this cycle, the cellular components of the parental cell must be duplicated and divided such that both daughter cells contain all of the necessary components for growth and survival. Genomic DNA replication and segregation are at the heart of this process. The mitotic cell cycle has a Synthesis (S) phase when genome duplication occurs and a Mitotic (M) phase when the duplicated chromosome are segregated such that each cell receives one of the two copies. The S and M phases are separated by Gap (G) phases that provide time to accumulate the necessary cellular machinery for the next step and ensure that the previous phase has completed.

The stages of the cell cycle can be defined by chromosomal events. Cells prepare for DNA synthesis in G1 phase by loading replicative helicases onto origin DNA (Figure 1). The helicases are activated and DNA synthesis begins during S phase. In G2 phase, cells begin preparations for chromosome segregation by condensing and separating the chromosomes. The G2 period also allows cells to check whether DNA synthesis has occurred correctly and completely before chromosome segregation. Cells detect incomplete DNA synthesis and DNA damage via cell cycle checkpoints during S and G2 phase. Checkpoint proteins recognize aberrations in DNA synthesis and activate pathways to arrest the cell cycle and trigger repair responses. Once replication is deemed
Figure 1. Eukaryotic cell cycle with a focus on the temporal separation of helicase loading and activation. In all eukaryotic cells, helicases can only be loaded during G1 phase and are active during S, G2, and M phase. This regulation ensures that DNA replicates only once per cell cycle. (Modified from Bell and Labib, 2016).
complete and accurate, cells progress into M phase to undergo sister-chromatid segregation, nuclear division, and cytokinesis. The result of this process is the formation of two daughter cells, each with a complete copy of the parental genome.

Progression through the cell cycle is regulated primarily by Cyclin-Dependent Kinase (CDK) activity. CDK levels are low during G1 phase and increase at the G1/S phase transition. CDK activity remains high throughout S, G2, and M phase. In addition to these changes in activity levels, CDK specificity changes throughout the cell cycle. Both the activity and specificity of CDK is controlled by regulatory subunits, called cyclins, which are associated with the core CDK kinase enzyme (Bloom & Cross, 2007). Different cyclins are produced in each phase of the cell cycle (see Figure 2 for budding yeast cyclins). During their corresponding phase, the cyclins bind to the core CDK. Importantly, the combination of changes in cyclin-CDK specificity and activity drives the transitions through the cell cycle (see below).

The events of DNA replication are separated during the cell cycle to ensure replication occurs once and only once per cell division. All eukaryotic organisms regulate the cell cycle to avoid re-replication, though their regulatory mechanisms vary. In mitotically-dividing eukaryotic cells, the cell cycle program avoids re-replication by temporally separating Mcm2-7 helicase loading and activation (Figure 1; Arias & Walter, 2007). Re-replication could hypothetically occur if, after a given DNA region is replicated, another helicase was loaded on the duplicated DNA and activated during the same cell cycle. In budding yeast, this control is entirely mediated by CDK levels and specificity (Diffley,
Figure 2. CDK-cyclin regulation of the cell cycle in *S. cerevisiae*. (A) Cyclin expression levels across the cell cycle. (B) Specific cyclin-CDK complexes have specific activities and act at different times. Note Cdc28 is the budding yeast CDK protein. (Modified from Bloom and Cross 2007).
Mcm2-7 loading can only occur during G1 phase when CDK levels are low. As cells transition into S phase, increased CDK levels prevents further helicase loading. This same high CDK activity, as well as altered specificity due to new cyclin expression, drives helicase activation and subsequent DNA synthesis.

Metazoan cells also prevent re-replication by separating helicase loading and activation, but they use different mechanisms to achieve this regulation (reviewed in Arias & Walter, 2007 and Siddiqui et al., 2013). The helicase-loading protein Cdt1 is the primary target of regulation, and it is inhibited outside of G1 phase by the protein geminin and by regulated ubiquitin-mediated proteolysis. In addition, CDK targets other helicase-loading proteins to prevent re-replication. In all organisms, multiple overlapping levels of regulation ensure that re-replication is inhibited throughout the cell cycle.

If CDK levels were the only mechanism of control, the transition between G1 and S phase when CDK levels are increasing could present an opportunity for helicase loading and activation to occur simultaneously. For example, local fluctuations around a threshold between inhibiting helicase loading and activation of the helicase could lead to repeated activation and loading. To prevent this possibility, yeast cells have a small window between G1 and S phase when neither helicase loading nor helicase activation can occur (Remus & Diffley, 2009). This safeguard is achieved by the different specificity conferred on CDK through cyclins (Figure 2B). Near the end of G1 phase, the G1 cyclins (Cln1 and Cln2) accumulate leading to high levels of G1-CDK activity (Figure 2A; Reed, 1991). Phosphorylation by G1-CDK causes nuclear export of unloaded (but not loaded) Mcm2-
and degradation or inhibition of other proteins required for helicase loading (Drury et al., 2000; Labib et al., 1999). Importantly, G1-CDK cannot activate loaded helicases. This elevated level of G1-CDK stimulates two additional events: 1) synthesis of S-phase cyclins (Clb5 and Clb6); and 2) degradation of an inhibitor of S-phase CDK activity, Sic1 (Verma et al., 1997). The combination of these two events leads to a cooperative and sharp increase in S-CDK activity. High levels of S-CDK both continue to inhibit helicase loading and are required to activate loaded helicases, something that G1-CDK cannot do at any level. During G2 and M phase, a combination of S- and M-CDKs continue to inhibit helicase loading.

Upon completion of chromosome segregation, the M-phase cyclins (Clb1-4) are degraded and Sic1 is activated to inhibit S-CDKs (Visintin et al., 1998). The cells then transition into G1 phase with very low CDK levels of any type. Just as the G1-S transition offers the opportunity to have simultaneous helicase loading and activation as CDK levels increase, the same is true during the M-G1 transition when CDK levels decrease. Importantly, a similar safeguard against re-replication exists during the M/G1 transition. This mechanism involves a second kinase called DDK (Dbf4-dependent kinase). In conjunction with S-CDK, DDK is required for helicase activation (see Helicase Activation below; Sheu & Stillman, 2010). During M phase, Dbf4 is degraded to stop helicase activation, but M-CDK continues to inhibit helicase loading. Thus, the ability to activate helicases is lost prior to removal of the inhibition of helicase loading.

In summary, kinase oscillations are responsible for both driving cell cycle transitions and
preventing re-replication. In all eukaryotes, re-replication is avoided by ensuring that helicase loading and helicase activation are separated during the cell cycle. In budding yeast, oscillations in CDK and cyclin levels create this separation. Short periods when neither event is possible ensure the transition between the helicase-loading and helicase-activation state do not lead to an intermediate state when both events can occur. Although I will not discuss it in detail, other eukaryotic cells use both CDK-dependent and CDK-independent mechanisms to accomplish the same separation (Malumbres & Barbacid, 2009).

**Mcm2-7 Helicase**

The core enzyme of the eukaryotic replicative DNA helicase is a ring-shaped heterohexameric protein called the Mcm2-7 (Mini-chromosome maintenance 2-7) complex (reviewed in Bochman & Schwacha, 2009). This complex is the core of the replisome and is the only factor involved in all steps of DNA replication. In addition to its essential DNA unwinding activity, the helicase acts as a scaffold to bring together many other proteins involved in both DNA synthesis and other replication events (e.g. chromatin disassembly and assembly) to form the replisome (Alabert et al., 2017)

Mcm2-7 is made of six related but heterologous subunits. Structural studies show that the Mcm2-7 subunits are arranged in a defined order: Mcm2-Mcm6-Mcm4-Mcm7-Mcm3-Mcm5 (Figure 3A; Costa et al., 2011). In solution, Mcm2-7 complexes form an “open” ring structure with a gap between Mcm2 and Mcm5. This “Mcm2-5 gate” allows origin DNA access to the central channel of the helicase during helicase loading (Samel et al., 2014).
**Figure 3.** Mcm2-7 structure. (A) Mcm2-7 (subunits indicated by numbers) in complex with Cdt1 (from Frigola et al. 2017). (B) Diagram of domains in each Mcm2-7 subunits. The NTD and CTD are highly conserved among the six subunits. The NTD includes an OB-fold-like motif, and the CTD includes AAA+ motif and WHD. The NTEs are only present in Mcm2, Mcm4, and Mcm6. (Modified from Randell et al. 2010).
Successful helicase loading closes this gate, resulting in the Mcm2-7 complex encircling the origin DNA (dsDNA) (see Helicase Loading below). The subunits all share a similar overall structure consisting of an N-terminal domain (NTD) and C-terminal domain (CTD) (Figure 3B). The CTD is characterized by ATPase and winged-helix domains (WHDs). The NTD includes an oligonucleotide/oligosaccharide-binding (OB) fold and, in most subunits, a Zn-finger domain. Of these domains, the ATPase domains are the most highly conserved. In addition, to these regions of similarity between all the subunits, the evenly numbered subunits (Mcm2, 4, and 6) also have long, unstructured N-terminal extensions. These regions are unrelated to one another and are involved in regulation of helicase function by phosphorylation and protein-protein interactions.

The ATPase domains in the Mcm2-7 complex are all part of the ATPases Associated with various cellular Activities (AAA+) family (J. M. Miller & Enemark, 2016). Six ATP hydrolysis active sites are formed at the interfaces between each adjacent pair of subunits and are characterized by several motifs. The Walker A and Walker B motifs form the core of the ATP binding and hydrolysis motif and are provided by one of the subunits at an interface. The adjacent subunit completes the active site by providing an essential arginine-finger motif required for ATP hydrolysis. The helicase harnesses the energy of ATP binding and hydrolysis to induce conformational changes that drive DNA unwinding activity (Hanson & Whiteheart, 2005). The exact mechanism by which DNA unwinding occurs as a result of ATP-dependent helicase conformational changes is a subject of active investigation (Abid Ali et al., 2017; Eickhoff et al., 2019; H. Li & O'Donnell, 2018; Yuan et al., 2020).
Origin Structure and Helicase Loading

Unlike many other eukaryotic cells, budding yeast origins of replication are defined by a specific DNA sequence, allowing detailed studies investigating the events of replication initiation both in vivo and in vitro. Yeast origins were originally identified as autonomous replicating sequences (ARSs) that allowed plasmids containing these sequences to be stably replicated independent of the rest of the yeast genome (Leonard & Méchali, 2013). These elements are typically 100-200 base pairs in length and are A/T rich. Comparison of ARS sequences led to the identification of a conserved element found at all origins called the ARS consensus sequence (ACS) (Broach et al., 1983). Importantly, alterations to this sequence completely eliminate origin function in budding yeast cells (Van Houten & Newlon, 1990). Origins in most other eukaryotes cannot be identified or defined by their sequence. Instead, DNA accessibility, chromatin state, and potentially other unknown determinants are thought to define these origins, although the exact determinants and whether they are the same across eukaryotes is unclear (reviewed in Ganier et al., 2019).

Helicase loading marks all potential origins of replication and results in two copies of the Mcm2-7 helicase being placed around the origin DNA. This process occurs exclusively during G1 phase and is accomplished by four proteins (Figure 4). I will describe this process for budding yeast, where we have the most detailed understanding. Although the initial origin recognition step is different, the four proteins involved are all highly conserved suggesting that the remainder of the reaction is similar in all eukaryotes (Bell & Kaguni, 2013). Loading begins with binding of the sequence-specific DNA binding protein, the Origin-Recognition Complex (ORC), to the conserved ACS sequence found at each
Figure 4. Schematic of Mcm2-7 helicase loading to form the double-hexamer helicase complex in *S. cerevisiae*. Note that after the first helicase is loaded, the same ORC complex binds to a second site on the DNA on the opposite side of the loaded helicase to facilitate loading of the second Mcm2-7.
origin. ORC then recruits Cdc6 (Cell division cycle 6) followed by a complex of Mcm2-7 and Cdt1 (CDC10-dependent transcript 1) to form a short-lived intermediate called the OCCM (ORC-Cdc6-Cdt1-Mcm2-7) complex. Cdc6 and then Cdt1 depart from the origin and the first Mcm2-7 ring closes around the origin to encircle double-stranded DNA (dsDNA) (Evrin et al., 2009; Remus & Diffley, 2009). ORC remains bound to the DNA along with the first loaded Mcm2-7. This ORC/Mcm2-7 complex is capable of recruiting a second Cdc6 and Mcm2-7/Cdt1 complex to load a second Mcm2-7 at the same origin (Ticau et al., 2015). Recent evidence suggest that the second loading event occurs when ORC binds to a second, weaker ORC binding site at the origin located in the opposite orientation to the first (Coster & Diffley, 2017; T. C. R. Miller et al., 2019). This reorientation appears to be facilitated by a second interaction between ORC and the N-terminal region of Mcm2-7. By the end of helicase loading, Cdc6, Cdt1, and ORC all depart and leave behind a complex of two Mcm2-7 at the origin.

Importantly, the two Mcm2-7 helicases are loaded in opposite orientations and interact with one another in a head-to-head manner via their N-terminal domains (see Figure 4). We call this two-helicase complex the double hexamer. The double hexamer is a tight and stable complex that persists for long periods in vitro even in the absence of DNA. Although the Mcm2-7 helicase is inactive at this stage, the head-to-head conformation positions the two helicases to depart the origin in opposite directions. Thus, formation of the Mcm2-7 double hexamer is critical to establish the bidirectional nature of the subsequent events of replication initiation and elongation. Despite its likely importance during initiation, the double-hexamer interface is disrupted later in initiation to allow the
two helicases to act independently and depart the origin in opposite directions (Yardimci et al., 2010). Given the necessity of separating the double-hexamer complex, why does the cell form it in the first place? There are two possibilities: (1) the double hexamer is formed to ensure bidirectional replication occurs by setting up each helicase in the correct orientation; or (2) the double-hexamer complex is necessary to complete a downstream step in replication, for example, helicase activation. I will explore this question in Chapter II.

**Helicase Activation**

Once cells enter S phase, helicases are activated to initiate DNA unwinding. Helicase activation requires a number of events that remodel the helicase and the associated DNA to prepare it for active DNA unwinding. The active helicase is a single Mcm2-7 in complex with other activation proteins, and this complex encircles and travels along ssDNA with its N-terminal domain leading (Fu et al., 2011). This transition requires: binding of key activator proteins, separation of the double-hexamer complex, origin DNA unwinding, ssDNA strand extrusion, and helicase passing (Figure 5). Despite their importance, the order and the mechanism of these events remains unknown.

During helicase activation, the proteins Cdc45 and GINS are recruited to the Mcm2-7 to form the CMG (Cdc45/Mcm2-7/GINS) complex (Figure 6). This complex is the active form of the replicative DNA helicase. Cdc45 and GINS activate both the ATPase and helicase activity of the Mcm2-7 core (Gambus et al., 2006; Heller et al., 2011; Ilves et al., 2010). CMG assembly is accomplished by a series of events coordinated by the S-CDK and
Figure 5. Helicase remodelling during helicase activation results in CMG formation, double-hexamer separation, origin DNA melting, lagging-strand extrusion, and the two helicases passing one another to establish separate replication forks. For simplicity, these events are depicted as occurring in one potential order, but the relative timing for these events is unknown.
DDK kinases. The first step in helicase activation is DDK phosphorylation of Mcm2-7 to promote association of Sld3/7 (synthetic lethal with $dpb11^{-13}$ and 7) and Cdc45 with the loaded double hexamer. S-CDK then phosphorylates Sld3/7 and Sld2 (Tanaka et al., 2007; Zegerman & Diffley, 2007). Phosphorylated Sld2 forms a complex with Dpb11 (DNA polymerase B possible subunit 11), GINS (from the Japanese Go-Ichi-Ni-San meaning 5-1-2-3), and Pol $\varepsilon$ (polymerase epsilon) (Muramatsu et al., 2010). This complex associates with the loaded helicase via interactions with phosphorylated Sld3/7. At some point in this process, Cdc45 and GINS form direct interactions with the core of Mcm2-7 to make a stable CMG complex that still encircles dsDNA (Douglas et al., 2018; Langston & O’Donnell, 2019). Of the proteins involved in helicase activation, only the CMG complex and DNA Pol $\varepsilon$ are associated with the final replisome. The remaining proteins involved in this process must dissociate, but when this happens and whether these events are coupled to successful activation is unknown.

An additional protein, Mcm10, acts after CMG formation to fully activate the helicase (van Deursen et al., 2012). In the absence of Mcm10, CMG formation results in a small amount of origin DNA being unwound (estimated to be less than one turn of the DNA helix per CMG) (Douglas et al., 2018). This small amount of DNA unwinding is thought to correspond to “untwisting”/“melting” of the dsDNA enclosed within the helicase central channel (discussed below). Additional DNA unwinding only occurs when Mcm10 is present (Douglas et al., 2018). How Mcm10 activates DNA unwinding is not known, but many models have been proposed (reviewed in Thu & Bielinsky, 2013). Double-hexamer separation could be required for extensive DNA unwinding activity, but there is mixed
Figure 6. Schematic of CMG formation and helicase activation. The loaded Mcm2-7 double hexamer is first phosphorylated by DDK. This phosphorylation recruits Cdc45 and Sld3/7. CDK then phosphorylates Sld3/7 and Sld2. In solution, phosphorylated Sld2 forms a complex with GINS, Polymerase ε, and Dpb11. This complex is then recruited to the helicase via interactions between Dpb11 and Sld3/7. At some point, Sld3/7, Sld2, and Dpb11 dissociate from the helicase complex, and Cdc45 and GINS remain tightly bound to Mcm2-7 to form the CMG helicase complex (Cdc45-Mcm2-7-GINS). Pol ε also remains associated with the CMG complex. Mcm10 acts later to initiate DNA unwinding which produces ssDNA that is bound by RPA.
evidence about the role of Mcm10 in hexamer separation. Some studies have shown defects in double-hexamer separation with mutant Mcm10 (Quan et al., 2015), whereas others have demonstrated hexamer separation in the absence of Mcm10 (Douglas et al. 2018). Alternatively, Mcm10 has been implicated in DNA strand extrusion when the helicase transitions from encircling dsDNA to encircling ssDNA (Fu et al., 2011; Wasserman et al., 2019). A transition onto ssDNA is necessary for the full helicase activity and is the form of the enzyme that acts at the replication fork (Costa et al., 2014). Given that each Mcm2-7 complex encircles ~3 turns of the DNA helix (Noguchi et al., 2017), it is clear that more DNA unwinding than is observed as a consequence of CMG formation would be required to create enough ssDNA to allow extrusion from the Mcm2-7 central channel (Figure 5).

The initial DNA unwinding that occurs when dsDNA is “untwisted” within the helicase central channel can also be called “origin melting”. There are a number of models for how origin melting is initiated. By comparing structures of the double-hexamer Mcm2-7 to the CMG complex on DNA, it has been suggested that a lateral shift and tilt of the NTD-tier rings of Mcm2-7 stimulates dsDNA separation and extrusion from the Mcm2-7 central channel (Figure 7A, Noguchi et al., 2017). Others have proposed that the Mcm2-7 NTD-tier moves apart from the CTD-tier during helicase activation to stretch the DNA and promote DNA untwisting and melting (Figure 7B, Abid Ali et al., 2017). Still other models are possible. Origin melting is required for ssDNA extrusion, but it is unknown when this event occurs relative to double-hexamer separation. In Chapter II, I will show that an intact double-hexamer is not required for origin-melting activity.
Figure 7. Two models for how DNA unwinding initiates during helicase activation. (A) Lateral shift and tilt of NTD-tier rings might stimulate lagging strand separation and extrusion from Mcm2-7 central channel (from Noguchi et al. 2017) (B) NTD-tier rings engage with the DNA. During helicase activation, the N-tiers then move away from each other, which could pull the DNA and cause DNA unwinding (from Abid Ali et al. 2017).
Helicase passing is likely the final step in helicase activation because it requires the other events to have occurred. The two helicases in the double-hexamer are initially loaded with their N-terminal domains facing one another and surrounding dsDNA. Because the active helicase moves with the N-terminal domain at the front and the CTD trailing, the helicases must move toward one another then pass to act independently (Figure 5). On dsDNA this is not possible, however, strand exclusion results in the two helicases encircling the opposite strands of origin DNA. The strand-excluded helicases can easily pass one another and begin extensive DNA unwinding to produce ssDNA template for DNA synthesis.

DNA Synthesis

The CMG helicase generates two strands of ssDNA that acts as the templates for DNA synthesis. Only one of the DNA strands can be synthesized continuously at the replication fork because the DNA strands are antiparallel and DNA synthesis only occurs in one direction. The strand that is synthesized continuously is known as the leading strand. The other strand is synthesized in a discontinuous fashion and is called the lagging strand (Figure 8). Lagging strand synthesis requires multiple RNA priming events resulting in 150-250 base pair units of newly replicated DNA known as Okazaki fragments (Georgescu et al., 2015; Okazaki et al., 1968). The leading strand requires only a single RNA priming event and synthesis continues until the replisome meets another replisome, the end of a chromosome, or another obstacle is encountered.

The Pol α/primase holoenzyme synthesizes the RNA-DNA primers required to initiate
Figure 8. The eukaryotic replication fork (from Zhang and O'Donnell 2016). The core of the replication fork is the CMG helicase, which unwinds DNA at the fork. RPA binds ssDNA produced by helicase activity. The DNA strand where synthesis occurs continuously is called the leading strand, and the strand synthesized discontinuously is called the lagging strand. On both strands, an RNA-DNA primer must be synthesized by the Pol α/primase complex before further DNA synthesis can occur. Leading strand DNA synthesis is taken over by Pol ε and lagging strand synthesis by Pol δ. PCNA is loaded onto DNA by RFC and increases DNA synthesis processivity by interacting with the polymerases.
DNA synthesis. The primase enzyme synthesizes RNA primers that are immediately extended by DNA polymerase α. The resulting RNA-DNA units are only 20-25 nucleotides in length. Pol α/primase forms a number of interactions with the replisome that contribute to its recruitment to the replication fork. In particular, Pol α/primase is localized to the CMG complex by the accessory protein Ctf4 (chromosome transmission fidelity 4), which in turn interacts with a subunit of GINS. However, Ctf4 is not essential, suggesting Pol α/primase can be recruited to the replication fork by other mechanisms. For example, in mammalian cells it has been shown that the primase subunit interacts with the ssDNA-binding protein RPA (Braun et al., 1997; Dornreiter et al., 1992), which could recruit Pol α/primase to the ssDNA behind the helicase. The other two DNA polymerases that act during replication can only initiate DNA synthesis from RNA-DNA primers created by Pol α/primase.

Pol ε takes over DNA synthesis from the RNA-DNA primer on the leading strand and Pol δ takes over on the lagging strand (Figure 8). Pol ε is essential for helicase activation and interacts directly with GINS and Mcm2-7 (Goswami et al., 2018; Sengupta et al., 2013). Pol ε binds on the same side of the CMG where the leading strand emerges following DNA unwinding, thus positioning the polymerase to synthesize the leading strand as soon as ssDNA template is generated by the CMG helicase (Georgescu et al., 2017; Goswami et al., 2018). In contrast, Pol δ is not known to associate with the helicase. Instead, Pol δ interacts with the sliding processivity clamp, PCNA (Proliferating Cell Nuclear Antigen), which is loaded onto DNA by the sliding clamp loader, RFC (Replication Factor C). PCNA enhances the processivity of Pol δ nearly 100-fold (Chilkova et al., 2007).
Pol δ is also involved in repairing the Okazaki fragments generated during lagging-strand synthesis. The RNA portion of the primers required to start new DNA strands must be removed to complete DNA synthesis. In addition, it is also preferable to remove the DNA synthesized by DNA Pol α because, unlike DNA polymerases δ and ε, DNA Pol α does not have a proofreading exonuclease making it more prone to errors. The Okazaki repair process involves removing the RNA-DNA primers, synthesizing new DNA to fill in the gap, and ligating the fragments together to form a continuous lagging strand DNA product (Figure 9; Giannattasio & Branzei, 2019). Pol δ begins the repair process when it encounters an RNA-DNA primer. Due to the discontinuous nature of lagging-strand synthesis, the polymerase will run into the primer end of the previous Okazaki fragment as it is completing the next Okazaki fragment. When Pol δ runs into an RNA primer, it continues synthesizing DNA, resulting in primer displacement. Nucleases then act to remove the RNA and DNA that was generated by Pol α/primase activity. A number of pathways utilizing different endo- and exo-nucleases exist for this cleavage process (Liu et al., 2017; Zaher et al., 2018), but the frequency each of these pathway is used is unknown. DNA ligase 1 repairs the resulting nicks between Okazaki fragments by ligating the Pol δ generated DNA. Additional proteins are recruited to the replisome to aid in various functions during DNA replication. Topoisomerases I and II (Top1 and Top2) the topological strain induced by DNA unwinding activity. Top1 and Top2 are generally redundant, except that Top2 is essential for decatenation of chromosomes at replication termination so that chromosome segregation can occur properly (Baxter & Difffley, 2008).

The histone chaperone, FACT, disassembles nucleosomes ahead of the helicase to
Figure 9

Figure 9. Okazaki fragment repair on the lagging strand during DNA synthesis. Pol δ displaces the RNA-DNA primer synthesized by Pol α/primase. Various nuclease digest the displaced RNA and DNA. Once the entire RNA-DNA primer is removed and no flap remains, DNA ligase will ligate the Okazaki fragments together to make one continuous lagging strand.
facilitate replisome progression on chromatin (Bondarenko et al., 2015; Kurat et al., 2017). Mrc1, Tof1, and Csm3 form the MTC complex which is important for optimal replication speed (Lewis et al., 2017; Yeeles et al., 2017). The MTC complex is also involved in activation of the DNA damage checkpoint.

**RPA and Other ssDNA-Binding Proteins**

DNA unwinding activity by the CMG helicase produces ssDNA that must be protected from cleavage (which would lead to chromosome breaks) and constrained from forming dsDNA prior to being used as a template for DNA synthesis. The rapid use of the leading-strand template by Pol ε means that the primary concern for dsDNA formation is the lagging-strand template self-annealing (Figure 8). During eukaryotic replication, these protections are achieved by RPA (Replication Protein A). RPA is a ssDNA-binding protein (SSB) that is essential in eukaryotes (Wold, 1997). RPA binds ssDNA in a non-sequence-specific manner. By binding and coating ssDNA, RPA blocks nucleases from accessing the DNA and inhibits the bound-ssDNA from base pairing with the complementary strand to form dsDNA or from base pairing with itself to form hairpin structures, which would slow DNA synthesis.

RPA is a heterotrimeric complex composed of RFA1, 2, and 3 (Replication Factor A 1, 2, and 3) (Figure 10A). RFA1 is the largest subunit at 70 kDa. It contains four oligonucleotide-binding (OB) folds, three of which act as DNA-binding domains (DBDs) whereas the fourth is important for protein-protein interactions with numerous DNA-damage and chromatin-remodeling proteins (Bhat & Cortez, 2018). RFA2 is 32kDa and
Figure 10. Structures of ssDNA-binding proteins (A) yeast RPA (Rfa1 in green, Rfa2 in cyan, and Rfa3 in blue) (Yates et al., 2018), (B) T4 Gp32 (Shamoo et al., 1995), and (C) E. coli SSB (EcSSB) tetramer (Raghunathan et al., 2000). When available in the structure, ssDNA is shown in yellow.
has one central DBD, an N-terminal flexible domain, and a C-terminal winged-helix domain (WHD) that is also involved in protein-protein interactions. The last subunit RFA3 is only 14kDa but is essential for viability and RPA function. RFA3 contains a single OB fold that is important for complex trimerization.

An individual RPA complex binds approximately 30 nucleotides of ssDNA with high affinity ($K_a \sim 10^9$ M; Bastin-Shanower & Brill, 2001). The binding dynamics and affinity of RPA vary depending on DNA length due to the ability of RPA to adopt several different ssDNA-binding conformations (reviewed in Fanning et al., 2006). Two of the DBDs on RFA1 can bind as little as 8 nucleotides with a $K_a$ of $\sim 10^7$ M. All four DBDs can bind as little as 20 nucleotides, but RPA exhibits the highest ssDNA affinity when binding 28-30 nucleotides (Kim et al., 1994). RPA is thought to traverse through these different binding modes to facilitate its loading onto and its removal from DNA (Chen & Wold, 2014). At this time, RPA binding to DNA during replication is believed to have low to no cooperativity. Cooperative binding was been observed in older studies, but more recent evidence suggests cooperativity only occurs when RPA is phosphorylated by a DNA-damage response kinase, Mec1 (Yates et al., 2018).

Most of what we know about protein-protein interactions between RPA and other DNA replication proteins comes from studies of simian virus 40 (SV40) DNA replication. At this viral replication fork, RPA interacts with the polymerases $\alpha$ and $\delta$. These RPA interactions are thought to facilitate Pol $\alpha$/primase to Pol $\delta$ “polymerase switching” during lagging-strand synthesis through interactions with Pol $\alpha$/primase, RFC, and Pol $\delta$ (Figure 11;
**Figure 11.** Polymerase switching during DNA replication. DNA polymerase $\alpha$ extends the RNA primer made by primase until it is displaced by DNA polymerase $\delta$ and the processivity clamp PCNA. (Modified from Garg & Burgers, 2005).
Dornreiter et al., 1992; Waga & Stillman, 1998; Yuzhakov et al., 1999). In addition, RPA interactions increase the processivity of Pol α/primase (Braun et al., 1997; Erdile et al., 1991; Kenny et al., 1989; Tsurimoto & Stillman, 1991). Although SV40 replication was a useful tool in early studies of DNA replication, the SV40 replication fork is distinct from the eukaryotic counterpart (see SV40 Studies of Replication below). In particular, the DNA helicase used during SV40 DNA replication, Large T-antigen (LTag), binds human RPA (Melendy & Stillman, 1993), but a similar interaction has not observed between RPA and the eukaryotic CMG helicase. Indeed, SV40 DNA synthesis specifically requires human RPA; other SSBs, including yeast RPA, cannot substitute (Brill & Stillman, 1989; Kenny et al., 1989). The direct interactions between LTag and human RPA are presumed to explain why only human RPA can support SV40 DNA replication. Further studies looking at RPA function in the context of eukaryotic replication are important to clarify RPA’s role in enhancing DNA replication.

RPA also plays a role in DNA repair pathways. Generally, long stretches of ssDNA bound by RPA function as a signal for replication stress. ATR (Mec1 in budding yeast) in complex with its regulatory partner ATRIP (Ddc2 in budding yeast) is responsible for initiating the DNA damage checkpoint. RPA is essential for the recruitment of ATR-ATRIP to ssDNA through an interaction between ATRIP and RPA (Zou & Elledge, 2003). This interaction is specific to RPA as other ssDNA-binding proteins cannot recruit ATRIP to ssDNA. At stalled replication forks, RPA helps to recruit the fork-reversal protein SMARCAL1 and stimulates its activity leading to repair of the stalled fork (Ciccia et al., 2009). RPA is also implicated in DNA recombination through interactions with the helicases Ssg1 and Dna2.
DNA replication SSBs have been found in all domains of life as well as many viruses, but these proteins display a wide variety of structures and ssDNA-binding affinities (Figure 10; reviewed in Marceau, 2012). The first SSB characterized was Gp32 (Gene 32 protein), which acts during T4 bacteriophage replication (Figure 10B). The canonical bacterial SSB is simply called SSB and was first identified in *E. coli* (EcSSB, Figure 10C). Gp32 is a monomeric 34 kDa protein. EcSSB is a homotetramer with a total molecular weight of 75 kDa. Both Gp32 and EcSSB bind ssDNA with high cooperativity (Marceau, 2012). The DNA binding-site size of Gp32 is 7-10 nucleotides, whereas a single EcSSB tetramer can bind as many as 65 nucleotides. All SSBs have strong affinity for binding ssDNA. Gp32 binding to ssDNA has an association constant around $10^8$ M (S. C. Kowalczykowski et al., 1981; Rouzina et al., 2005), but EcSSB has a tighter binding affinity in the $10^{10}$ M range (Naufer et al., 2019). Like RPA, EcSSB has a variety of ssDNA-binding modes that are distinguished by their engagement of different sets of DBDs. Each EcSSB subunit contains one DBD such that the tetrameric protein has four DBDs in total. Gp32 has only one DBD. Despite these differences in structure, all three SSBs perform the same function of binding ssDNA during DNA replication. Given this similarity, I will test whether Gp32 or EcSSB can substitute for RPA during eukaryotic DNA replication in Chapter III.

**Biochemical Studies of DNA Replication**

The development of biochemical assays to study eukaryotic DNA replication has been
vital in determining most of the mechanistic details of replication described above. Many of these assays must mimic the G1-S cell-cycle transition to separate the incompatible events of helicase loading and helicase activation. Over time, the information we’ve learned from these biochemical assays has led to the development of additional assays focused on specific events and have opened the door to further discoveries. The recent development of assays that use only purified proteins has both fully defined the minimal set of proteins required and allowed the substitution of modified or mutant proteins into these assays.

Biochemical assays for DNA replication have evolved with our understanding of the process. One of the first biochemical DNA replication studies injected purified DNA and radioactive deoxythymidine into unfertilized *Xenopus laevis* frog eggs to detect replication of the DNA (Gurdon et al., 1969). A cell-free system was later developed by making extracts from *Simian Virus 40* (SV40)-infected monkey cells. However, the key SV40 protein Large T-antigen (LTag) bypasses the function of the eukaryotic replicative helicase and the helicase-loading and -activation proteins (Figure 12). Thus, although this assay resulted in the identification of a number of elongation proteins, it could not be used to study the events of initiation or the function of the eukaryotic helicase. The development of nucleus-free extract-based replication assays opened a new avenue for exploring eukaryotic replication (Walter et al., 1998). These systems synthesized DNA by sequentially incubating DNA with extracts from *Xenopus* egg or budding yeast that mimicked the G1 and S phase states (Heller et al., 2011; Walter et al., 1998).
Figure 12. Initiation of SV40 DNA replication by large T antigen (LTag or Tag). (From Lodish et al. 2000).
Most recent studies of DNA replication have moved beyond extract-based assays and use only purified proteins to conduct fully-reconstituted assays. These assays allow greater freedom to study the consequences of elimination of specific proteins or the impact of mutant proteins. The first fully-reconstituted assay recapitulated loading of the replicative helicase Mcm2-7 onto origin DNA (Evrin et al., 2009; Remus et al., 2009). This assay established that helicases are loaded as double hexamers that encircle dsDNA. The reconstitution of helicase loading has also enabled structural studies of intermediates in the helicase-loading process (Sun et al., 2013; Yuan et al., 2017). This reconstituted helicase-loading assay was then used in single-molecule studies to uncover further details about protein stoichiometry, association dynamics, and conformational changes that were not easily attainable from bulk experiments (Duzdevich et al., 2015; Ticau et al., 2015, 2017).

Recently, helicase activation and replication elongation were reconstituted with a minimal set of purified proteins (Yeeles et al., 2015). Two events in helicase activation can be monitored using the minimal set of helicase-activation proteins: CMG-helicase formation and DNA-unwinding activity. Both assays sequentially load and activate the helicases on origin DNA using purified proteins. CMG formation is monitored using a template-association assay where stable recruitment of Cdc45, GINS, and Mcm10 to the Mcm2-7 on DNA can be confirmed. DNA unwinding is assayed by using a topologically-constrained, circular DNA template. Changes in DNA topology following helicase activation indicate that DNA unwinding has occurred (Douglas et al., 2018).

Remarkably, in vitro DNA synthesis with purified proteins is also possible (Yeeles et al., 2015).
In addition to the helicase-activation proteins, purified DNA polymerases and replication-fork proteins are added to loaded Mcm2-7 helicases to initiate DNA replication. Radiolabeled dNTPs are included in the reaction to specifically detect newly synthesized DNA. DNA products are then separated on a denaturing gel such that leading and lagging strand products are distinguishable. Thus, differences in leading- and lagging-strand synthesis can be observed as well as changes in replication-fork speed (Aria & Yeeles, 2018; Lõoke et al., 2017; Yeeles et al., 2015, 2017). I will use reconstituted helicase-activation and DNA-synthesis assays to evaluate whether these activities still occur in the absence of double hexamers (Chapter II) or with alternative SSBs (Chapter III).

**SV-40 Studies of Replication**

As mentioned above, early *in vitro* studies of eukaryotic DNA replication were conducted using the simian virus SV40 as a model. Soluble extracts prepared from SV40-infected monkey cells can be used to initiate and complete replication of a plasmid DNA containing an SV40-specific origin (J. J. Li & Kelly, 1984). In addition, uninfected cell extract can support replication when purified SV40 large T antigen (LTag) is added to the reaction. SV40 primarily uses the host cell’s replication machinery to duplicate its DNA, including host DNA polymerases, RPA, and topoisomerase. LTag is the only virally-encoded protein required for DNA replication. Remarkably, this one protein functions as both the origin-recognition protein and the replicative DNA helicase (reviewed in Fanning & Knippers, 1992). In addition, SV40 replication is not cell cycle regulated, unlike eukaryotic DNA replication.
Although SV40 DNA replication has many similarities to eukaryotic replication, we now know that there are also significant differences. The biggest difference between eukaryotic and SV40 replication is that the LTag DNA helicase does not require any other proteins to load it onto SV40 origin DNA (Figure 12), unlike the eukaryotic helicase Mcm2-7, which requires ORC, Cdc6, and Cdt1 for origin loading. Similarly, once loaded, it requires no other proteins to become activated. The functional differences between the LTag and Mcm2-7 helicases suggest that they may interact differently with other replication proteins. Although there is evidence that LTag interacts with RPA, there is no evidence that Mcm2-7 or the CMG has a similar interaction with RPA. Furthermore, if there are no direct interactions between the Mcm2-7 helicase and RPA, then it is possible that other SSBs can substitute for RPA during eukaryotic replication. In Chapter III, I will explore the question of whether RPA is specifically required for eukaryotic DNA unwinding and DNA replication or if other SSBs can support these activities.

**Thesis Summary**

My thesis covers two separate projects studying eukaryotic DNA replication. These two projects can be tied together by the question: “what is required for initial origin DNA unwinding?” In the first project (Chapter II), I investigate the function of a mutant Mcm2-7 helicase that is competent for loading but defective in Mcm2-7 double-hexamer formation (see Helicase Loading above). In particular, I asked what role, if any, the double-hexamer conformation of the helicase plays during replication initiation. This project was a joint effort between myself and another student in my lab, Kanokwan (Paggard) Champasa. Paggard created the mutant helicase and assessed its helicase-loading phenotypes.
Using reconstituted assays, I used the mutant protein to show that the double-hexamer form of the helicase is not required for recruiting the helicase-activation proteins to form the CMG helicase and melt origin DNA (see Helicase Activation) but is required for subsequent DNA unwinding and DNA synthesis activity.

In the second project (Chapter III), I explore the role of the ssDNA-binding (SSB) protein in eukaryotic DNA replication. RPA is the SSB used in eukaryotic replication, but it is unknown whether RPA's function is simply to bind ssDNA following DNA unwinding or if it also has more specific functions. I compare RPA to other SSBs, specifically EcSSB and GP32, to see if other SSBs can support eukaryotic DNA replication activities which otherwise require RPA. I found that the SSBs have different phenotypes. In reconstituted assays for eukaryotic DNA replication, Gp32 cannot support DNA unwinding or DNA synthesis activity. In contrast, EcSSB allows for DNA unwinding but shows strong defects during DNA synthesis. Specifically, DNA replication with EcSSB produces only long DNA products, suggesting a defect in lagging-strand synthesis that results in either reduced synthesis activity or infrequent priming events that lead to long Okazaki fragments.
References


dimerization, and ATPase ring remodeling in the CMG helicase of the eukaryotic replisome. *Elife, 3*, e03273. https://doi.org/10.7554/eLife.03273


Chapter II: A Conserved Mcm4 Motif Is Required for Mcm2-7 Double-Hexamer Formation and Origin DNA Unwinding

An earlier version of this chapter was published as referenced below.


*indicates equal contribution. I performed and analyzed the CMG-formation, DNA-unwinding, and DNA-replication experiments with feedback from S.P.B. K.C. performed all other experiments and analyzed the data with feedback from J.G., L.J.F. and S.P.B.
ABSTRACT

Licensing of eukaryotic origins of replication requires DNA loading of two copies of the Mcm2-7 replicative helicase to form a head-to-head double-hexamer, ensuring activated helicases depart the origin bidirectionally. To understand the formation and importance of this double-hexamer, we identified mutations in a conserved and essential Mcm4 motif that permit loading of two Mcm2-7 complexes but are defective for double-hexamer formation. Single-molecule studies show mutant Mcm2-7 forms initial hexamer-hexamer interactions, however, the resulting complex is unstable. Kinetic analyses of wild-type and mutant Mcm2-7 reveal a limited time window for double-hexamer formation following second Mcm2-7 association, suggesting that this process is facilitated. Double-hexamer formation is required for extensive origin DNA unwinding but not initial DNA melting or recruitment of helicase-activation proteins (Cdc45, GINS, Mcm10). Our findings elucidate dynamic mechanisms of origin licensing, and identify the transition between initial DNA melting and extensive unwinding as the first initiation event requiring double-hexamer formation.

INTRODUCTION

Initiation of DNA replication occurs at genomic sites called origins of replication. Each eukaryotic origin is licensed during G1 phase through loading of the replicative helicase. The ring-shaped heterohexameric Mcm2-7 helicase is loaded around origin DNA by three loading proteins - ORC, Cdc6, and Cdt1 (Bell and Labib, 2016). The resulting licensed origin has two Mcm2-7 complexes encircling the DNA in a head-to-head orientation with their N-terminal domains interacting to form a “double hexamer” (Abid Ali et al., 2016;
Evrin et al., 2009; Li et al., 2015; Noguchi et al., 2017; Remus et al., 2009). Because they are poised to leave the origin in opposite directions, the head-to-head orientation of the loaded double hexamer is the first step in the establishment of bidirectional replication initiation. In the subsequent S phase, a subset of these helicases interact with two helicase-activating proteins, Cdc45 and GINS to form the active replicative helicase known as Cdc45-Mcm2-7-GINS (CMG) complex. This complex is then activated by Mcm10, leading to DNA unwinding, replication fork assembly and the initiation of DNA synthesis (Douglas et al., 2018; Lõoke et al., 2017; Yeeles et al., 2015). Separation of helicase loading and activation into distinct cell-cycle stages ensures that no origin can initiate replication more than once per cell cycle (Bell and Labib, 2016).

Helicase loading requires a series of ordered association and dissociation events. ORC-DNA binding is followed by Cdc6 association and recruitment of a Mcm2-7-Cdt1 complex to form a transient four-protein intermediate called the ORC-Cdc6-Cdt1-Mcm2-7 (OCCM) complex (Randell et al., 2006; Sun et al., 2013). The ordered release of Cdc6 and Cdt1 along with ATP hydrolysis by the Mcm2-7 complex results in closing of the first Mcm2-7 ring around the origin DNA (Ticau et al., 2017; 2015). Two models have been proposed to explain the recruitment and loading of the second Mcm2-7 (Yardimci and Walter, 2014). A two-ORC model suggests that there are two independent rounds of Mcm2-7 loading involving two ORC-Cdc6 complexes. These two loaded Mcm2-7 complexes are proposed to subsequently come together through translocation on the double-stranded origin DNA (Coster and Diffley, 2017; Frigola et al., 2013). This model is supported by the in vivo requirement for two ORC-binding sites at origins of replication and mutations that suggest
that both the first and second Mcm2-7 complexes interact with ORC during loading (Coster and Diffley, 2017; Frigola et al., 2013). A second model proposes a single ORC molecule loads both the first and second Mcm2-7 (Ticau et al., 2015), with interactions between the first and second Mcm2-7 complex playing an important role during loading of the second Mcm2-7. This model is supported by single-molecule analysis of helicase loading, which observed that one ORC molecule is sufficient for Mcm2-7 double-hexamer formation (Ticau et al., 2015).

The Mcm2-7 helicase is a two-tiered, ring-shaped complex comprised of six distinct but related subunits (Abid Ali and Costa, 2016). The C-terminal domain (CTD) of each subunit is composed of AAA+ ATPase and winged-helix domains. The folded N-terminal domain (NTD) of each subunit includes an oligonucleotide/oligosaccharide-binding- (OB-) and, in some cases, a Zn-finger fold. In addition, Mcm2, Mcm4 and Mcm6 include extensive unstructured N-terminal extensions (NTEs). Although unrelated to one another, these NTEs are a conserved feature of these Mcm subunits in all eukaryotic species examined (Miller and Enemark, 2015). Once loaded, the tight interaction between the two Mcm2-7 complexes in the double hexamer is mediated by the NTDs (Li et al., 2015; Noguchi et al., 2017). It is clear that the Mcm4 and Mcm6 NTEs are critical targets of DDK during helicase activation (Deegan et al., 2016; Randell et al., 2010; Sheu and Stillman, 2010). The NTEs are not observed in structures of the Mcm2-7 double hexamer on or off the DNA (Abid Ali et al., 2017; Li et al., 2015; Noguchi et al., 2017) suggesting that the NTEs do not form a well-ordered structure at the interface holding the two hexamers together. Despite this, the NTEs are well positioned to facilitate initial interactions between the
Mcm2-7 complexes and their role during helicase loading has not been examined in detail.

Although the formation of the Mcm2-7 double hexamer is critical to ensure bidirectional replication, it is unclear when or whether this form of the helicase is required for initiation. A key tool to address the role of double-hexamer formation would be separation-of-function mutations that allow loading of individual Mcm2-7 helicases but interfere with formation of the double hexamer. Such mutations would allow an investigation of which steps, if any, during replication initiation require the double-hexamer form of Mcm2-7.

To investigate the formation and function of the Mcm2-7 double hexamer, we initially focused on the function of the MCM NTEs. We made Mcm2-7 complexes lacking individual NTEs and found that constructs with a deletion of the Mcm2 or Mcm4 NTE were defective in helicase loading. During these studies, we identified a highly-conserved and essential motif within the Mcm4 NTD that is required for stable Mcm2-7 double-hexamer formation. Using single-molecule FRET analysis, we showed that mutation of this motif allows Mcm2-7 DNA loading and initial double-hexamer interactions, but the resulting complexes quickly dissociate. Although the separated Mcm2-7 complexes remain on the DNA, they do not form a double hexamer again. For both the wild-type and mutant Mcm2-7 complexes, all double hexamers form in a limited period of time after the DNA association of the second Mcm2-7. Mcm2-7 with a mutation in the Mcm4 motif cannot support replication initiation in vitro. Interestingly, these complexes can associate with helicase-activating proteins and perform initial origin DNA melting but fail to transition to
extensive DNA unwinding. Our observations identify a key motif required for stable Mcm2-7-Mcm2-7 interactions and identify the transition to extensive origin DNA unwinding as the first step in replication initiation that requires the double-hexamer form of the helicase.

RESULTS

N-terminal deletions of Mcm2 and Mcm4 inhibit Mcm2-7 loading

To study the role of Mcm2, Mcm4 and Mcm6 N-terminal extensions (NTEs) in helicase loading, we created deletion mutations lacking these elements (Figure 1A). We used Quick2D (Dosztányi et al., 2005; Jones, 1999; Zimmermann et al., 2017) and the location of phosphorylation sites (Randell et al., 2010) to identify the N-terminal unstructured regions for each subunit (Figure 2). Based on these criteria, we purified Mcm2-7 complexes containing Mcm2Δ2-177, Mcm4Δ2-181, or Mcm6Δ2-105. Subsequent structural studies (Abid Ali et al., 2017; Li et al., 2015; Noguchi et al., 2017) suggested that the first structured amino acids of Mcm2, Mcm4, and Mcm6 are amino acids 201, 177, and 103, respectively. Thus, we deleted the unstructured regions of Mcm2 and Mcm6, but included a small structured region in the Mcm4 NTE deletion. In addition, we purified Mcm2-7 containing Mcm4Δ2-174, which bypasses the requirement of DDK in vivo (Sheu and Stillman, 2010) but had not been characterized biochemically. Importantly, upon expression with the other five wild-type subunits and Cdt1, each mutant Mcm subunit assembled into a stable Mcm2-7 heterohexamer that bound to Cdt1 (Figure 3). We will refer to the mutant Mcm2-7 complexes by the mutated subunit and the site of the mutation (e.g. Mcm2-7 that contains Mcm4Δ2-181 is Mcm2-74Δ181).
Figure 1. N-terminal deletions of Mcm2 and Mcm4 NTEs are defective for Mcm2-7 loading and essential for viability (A) Diagram of Mcm2, Mcm4, and Mcm6 deletions. (B) Mcm2-7Δ177 and Mcm2-7Δ181 are defective in helicase loading onto origin-containing DNA. Helicase loading of the indicated Mcm2-7 N-terminal-deletions was monitored by incubating with purified helicase-loading proteins and bead-attached origin DNA followed by a high-salt wash (0.5M NaCl). All loading was dependent on Cdc6. (C) Mcm2Δ177 and Mcm4Δ181 deletions are lethal. The indicated genes were tested for complementation of a deletion of the corresponding wild-type gene before (YPD) and after (FOA) selecting against a plasmid with the corresponding wild-type gene. Endogenous MCM2, MCM4 or MCM6 gene is deleted, and a copy of the corresponding wild-type MCM is present on a URA3-containing plasmid. Indicated MCM mutants were integrated into the genome at the corresponding MCM locus. Five-fold serial dilutions of cells were spotted on the indicated media.
**Figure 2.** Secondary structure prediction of Mcm2, 4, and 6 NTE by Quick2D, using the indicated analysis tools (PSIPRED, Protein secondary structure prediction based on position-specific scoring matrices (Jones 1999); IUPRED, the pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins (Dosztanyi et al., 2005); SS, alpha-helix/beta-strand; DO, Disorder). The resulting secondary structure predictions are labeled as following; H, alpha-helix; E, beta-strand; D, Disorder. Red arrows indicate the last amino acid of the NTE deletion regions.
Figure 3. Mcm2-7 complexes containing all wild-type or one mutant subunit (indicated) assembled into stable Mcm2-7 heterohexamers that bound to Cdt1. After protein purification and gel filtration column (see Materials and Methods), Mcm2-7 complexes were separated on 8% acrylamide gel and stained with Coomassie Blue.
To determine the impact of each deletion on Mcm2-7 loading, we tested the mutant Mcm2-7 complexes in an in vitro helicase-loading assay (Remus et al., 2009). After incubating the Mcm2-7 complex with purified ORC, Cdc6, Cdt1 and DNA, we used a high-salt wash to release helicase-loading proteins and Mcm2-7 complexes that had not completed loading (Donovan et al., 1997; Randell et al., 2006). Comparison with wild-type Mcm2-7 shows that Mcm2-72Δ177 and Mcm2-74Δ181 have loading defects (Figure 1B). In contrast, Mcm2-74Δ174 and Mcm2-76Δ105 are loaded at similar levels as the wild-type protein. Thus, the NTEs of Mcm2 and Mcm4 contribute to stable helicase loading. To determine whether the NTE deletions impact MCM function in vivo, we assessed the ability of each of the mutant MCM genes to complement a deletion of the corresponding wild-type gene. Although Mcm2-72Δ177 and Mcm2-74Δ181 show only modest helicase-loading defects in vitro, each of these mutants is lethal in vivo (Figure 1C). In contrast, Mcm2-76Δ105 supports normal cell growth, and Mcm2-74Δ174 cells are viable but grow at a much slower rate. Thus, the Mcm2 and Mcm4 NTE mutants impact both helicase loading and cell viability.

A conserved Mcm4 N-terminal motif is required for stable double-hexamer formation

Because the NTEs are located at the double-hexamer interface, we asked if the NTE-deletion mutations impacted the formation of this complex. Previous studies have shown that double-hexamer formation is not required for salt-stable helicase loading detected in the previous assays (Ticau et al., 2015). Thus, we tested the ability of each of the mutant complexes to form Mcm2-7 double hexamers using a previously described gel-filtration
assay (Figure 4A, Evrin et al., 2009). For wild-type Mcm2-7, Mcm2-7^{2Δ177}, Mcm2-7^{4Δ174} and Mcm2-7^{6Δ105}, the majority of salt-resistant DNA associated Mcm2-7 complexes eluted as double hexamers (Figure 4B). In contrast, loaded Mcm2-7^{4Δ181} is entirely in the form of single hexamers (Figure 4B), suggesting that this mutation inhibits double-hexamer formation or stability.

The different functionality of the two Mcm4 mutants (Mcm2-7^{4Δ174}, no loading defect, viable; Mcm2-7^{4Δ181}, loading defect and lethal) focused our attention on the region of Mcm4 only present in Mcm2-7^{4Δ174} (amino acids 175-181). To test the importance of this region in the context of otherwise full-length Mcm4, we constructed both substitution and deletion mutations (Figure 5A). Deletion or alanine substitution of these amino acids causes a two-fold loading defect compared to wild type Mcm2-7 (Figure 3A). In the double-hexamer assay, loaded Mcm2-7^{4Δ175-181} and Mcm2-7^{4-7A} eluted almost exclusively in the form of single hexamers (Figure 5B), implicating this region of Mcm4 in double-hexamer formation or stability. Importantly, both mcm4^{Δ175-181} and mcm4-7A are unable to complement a MCM4 deletion (Figure 7).

Sequence analysis of the 175-181 region of Mcm4 identified a highly-conserved nine-amino-acid motif that overlapped with this region (Figure 6A). To test whether this motif is required for double-hexamer formation, we constructed a series of substitution mutations within this motif (Figure 6B), incorporated them into otherwise wild-type Mcm2-7 complexes and tested them for helicase loading and double-hexamer formation. Mcm2-7 complexes with mutations that disrupt the conserved motif (Mcm2-74-178A, Mcm2-74-181FA, Mcm2-74-182A and Mcm2-74-185A) show two-fold loading defects compared to
Figure 4. Deletion of the Mcm4 N-terminal domain inhibits Mcm2-7 double-hexamer formation. (A) Double-hexamer assay scheme. After performing a helicase-loading reaction with purified proteins and high-salt wash, loaded Mcm2-7 complexes were released from the bead-bound DNA by DNase I treatment. Released Mcm2-7 complexes were separated by gel filtration to separate double from single hexamers. (B) Mcm2-7Δ181 is defective for double-hexamer formation. Mcm2-7 complexes including the indicated N-terminal deletions were tested for their ability to form double hexamers as described in (A). The Flag-Mcm3 subunit of the Mcm2-7 complexes in the indicated fractions was detected by immunoblot. Wild-type Mcm2-7 double hexamers eluted early (fraction 22–26) whereas purified Mcm2-7 eluted later (fraction 30–34). Viability of a strain containing the mutation (right) was tested as in Figure 1C.
Figure 5

Figure 5. Amino acids 175-181 of Mcm4 are important for Mcm2-7 loading and double-hexamer formation. (A) Top, diagram of Mcm4 amino acids 175-181 deletion and alanine-substitution mutations. Bottom, Mcm2-7Δ175-181 and Mcm2-7AΔ175-181 are defective for helicase loading. Helicase loading of wild-type or Mcm2-7 complexes including the indicated Mcm4 mutant was monitored as described in Figure 1. The associated graph shows the relative loading of the Mcm4 mutants compared to wild-type Mcm2-7 based on three independent experiments. Error bars indicate the SD. (B) Mcm2-7Δ175-181 and Mcm2-7AΔ175-181 are defective in double-hexamer formation. Double-hexamer formation by Mcm2-7 complexes containing the indicated Mcm4 mutation was tested as described in Figure 4. Mutant viability data are from Figure 7.
wild-type, whereas mutant complexes that do not alter the conserved regions (Mcm2-74-175A and Mcm2-74-188A) show near-wild-type levels of helicase loading (Figure 6C). Testing these mutants for the formation of stable double hexamers shows that Mcm2-74-178A, Mcm2-74-181FA, Mcm2-74-182A were primarily in single-hexamer form and Mcm2-74-185A is split between single- and double-hexamer forms (Figure 6D). In contrast, the 175A and 188A mutants that are outside of the conserved motif are primarily in the double-hexamer form (Figure 6D). These results demonstrate that this conserved Mcm4 motif is important for the formation or stability of the Mcm2-7 double hexamer. Therefore, we named this region the Double-Hexamer Motif (DoHM). Consistent with double-hexamer formation being essential, we observed a close correlation between the ability of a mutant Mcm4 to form stable double hexamers and the ability of the corresponding mutant gene to complement a MCM4 deletion (Figure 6D and 7B).

The DoHM is required for Mcm2-7 double-hexamer stability

The DoHM mutants could result in reduced double-hexamer detection for two possible reasons. The first possibility is that these Mcm4 mutants allow double-hexamer formation, but the resulting assemblies fall apart due to a weakened interface. Alternatively, the DoHM could be required to initiate double-hexamer formation. In the experiments described thus far, we used an ensemble assay to monitor double-hexamer formation. This assay requires the two Mcm2-7 complexes to remain associated during size-exclusion chromatography to be detected. Thus, defects either in the rate of stable double-hexamer formation or in the lifetime of the double-hexamer complex would not be distinguished by this assay.
Figure 6. The DoHM is an essential and conserved region within Mcm4 that is required for Mcm2-7 double-hexamer formation. (A) Protein-sequence comparison of amino acids (175 to 194) surrounding the of *S. cerevisiae* DoHM (boxed) with Mcm4 of the indicated species (sp, *S. pombe*; dm, *D. melanogaster*; dr, *D. rerio*; ce, *C. elegans*; xl, *X. laevis*; hs, *H. Sapien*; at, *A. thaliana*). Identical amino acids are shaded dark green, and conserved amino acids are shaded light green. (B) Diagram of alanine mutations spanning the DoHM and the flanking amino acids. (C) Mcm4 mutations within the DoHM show helicase-loading defects. Helicase loading of Mcm2-7 complexes containing indicated Mcm4 mutants was monitored as described in Figure 1. The associated graph shows the relative loading of the Mcm4 mutants compared to wild-type Mcm2-7, based on three independent loading experiments. Error bars indicate the SD. (D) Mcm4 mutations within the DoHM are defective for double-hexamer formation. Double-hexamer formation by Mcm2-7 complexes containing indicated Mcm4 mutant was tested as described in Figure 4. Mutant viability data are from Figure 7.
Figure 7. Complementation tests for Mcm2-7 containing Mcm4 mutants. (A) Deletion (Mcm4Δ175-181) and alanine substitution (Mcm4-7A) of Mcm4 amino acid 175-181 region are lethal. (B) Mcm4 alanine mutations that disrupted the DoHM (Mcm2-74-178A, Mcm2-74-181FA, Mcm2-74-182A and Mcm2-74-185A) are lethal, whereas the mutants that did not alter the conserved regions (Mcm2-74-175A and Mcm2-74-188A) are viable. Mcm2-74-185A, which has partial defect in double hexamer formation, shows growth defect in this test. In all strains, endogenous MCM4 gene is deleted, and a copy of wild-type MCM4 is present on a URA3-containing plasmid. Indicated MCM4 mutants were integrated into the TRP1 locus. Five-fold serial dilutions of cells were spotted on the indicated media.
To monitor double-hexamer interactions in real time, we used a single-molecule FRET assay for double-hexamer formation (Ticau et al., 2015). In this experiment, the N-terminus of the Mcm7 subunit in two separate Mcm2-7/Cdt1 preparations was labeled with either a donor or an acceptor fluorophore. We incubated surface-tethered fluorescent origin DNA with an equimolar mixture of the two differentially-labeled Mcm2-7 complexes together with the three helicase-loading proteins (ORC, Cdc6 and Cdt1, Figure 8A). We alternately excited the acceptor and donor fluorophores to detect the colocalization of donor- or acceptor-labeled Mcm2-7 with the fluorescently-labeled DNA, a proxy for DNA binding. Importantly, when the donor fluorophore is excited, we are also able to detect hexamer-hexamer interactions through an increase in apparent FRET efficiency ($E_{\text{FRET}}$, Ticau et al., 2015). To eliminate helicase-loading events during which we cannot monitor these interactions, we only analyzed events in which we observed DNA association of one donor-fluorophore-labeled and one acceptor-fluorophore-labeled Mcm2-7 complex, and measured $E_{\text{FRET}}$ only after arrival of the second Mcm2-7.

Single-molecule analyses of double-hexamer formation supported a role for the DoHM in double-hexamer formation. Both wild-type Mcm2-7 and a DoHM mutant (Mcm2-7<sup>4-178A</sup>, referred to as the DoHM mutant hereafter) exhibited an increase in $E_{\text{FRET}}$ shortly (<10 s) after the arrival of the second Mcm2-7 (Figure 8B,C and 9). Analysis of wild-type helicase-loading events (N=89) revealed that immediately after arrival of the second Mcm2-7, the two-Mcm2-7 complexes were primarily in state that exhibited zero or very low FRET ($E_{\text{FRET}} = 0.029 \pm 0.003$, Figure 8D-i and 8D-ii, Table 1). However, shortly after the second Mcm2-7 arrived, many (~70%; Table 1) of the DNAs associated with wild-type Mcm2-7
Figure 8. The DoHM is required for double-hexamer formation and stability. (A) Schematic of the single-molecule helicase-loading assay. Alexa-Fluor-488-labeled (blue circle) 1.3 kb origin DNAs were coupled to a passivated microscope slide. Mcm2-7 was fluorescently labeled with donor (Dyomic-549 green circle) or acceptor (Dyomic-649, red circle) fluorophores. Purified ORC, Cdc6, and Cdt1/Mcm2–7 were incubated with slide-coupled DNA. Colocalization of the fluorescently labeled protein with the DNA and any associated FRET signal were monitored. (B) Wild-type Mcm2-7 forms long-lasting FRET signals. Representative fluorescence records for experiments using a 1:1 mixture of wild-type donor- and acceptor-labeled Mcm2-7 showed FRET after arrival of the second Mcm2–7. Records of fluorescence intensity for (i) acceptor excitation; acceptor emission (Dyomic-649-labeled Mcm2-7, red arrow marks arrival of acceptor-labeled Mcm2-7), (ii) donor excitation; donor emission (Dyomic-549-labeled Mcm2-7, green arrow marks arrival of donor-labeled Mcm2-7) and FRET (donor excitation; acceptor emission, blue arrow marks initiation of FRET), (iii) total emission (donor excitation; donor emission + acceptor emission), and (iv) calculated $E_{\text{FRET}}$ are shown. Black arrows indicate both donor and acceptor release due to the double hexamer sliding off the end of the DNA. (C) Representative fluorescence records for experiments using Mcm2–74-178A (labeled and arrows as in B) show a short-lived FRET signal. (D) (i) Time evolution of the $E_{\text{FRET}}$ distribution for 89 wild-type Mcm2-7 complexes. Only complexes with one donor-labeled and one acceptor-labeled Mcm2-7 were selected. $E_{\text{FRET}}$ values were measured only after arrival of the second Mcm2–7, which was taken to be time zero in each record. The plot is a two-dimensional histogram (see Materials and Methods) with $N_t = 2,688$ measurements within the time and $E_{\text{FRET}}$ range shown. (ii-iv) Histograms (probability density ± S.E.) of $E_{\text{FRET}}$ values recorded during the indicated time intervals after association of the second Mcm2–7 with origin DNA (black bar indicates possible intermediate $E_{\text{FRET}}$ state). $E_{\text{FRET}}$ values were globally fit to the sum (dashed cyan curves) of two Gaussians (red curves) constrained to have the same peak positions and widths at all times. (E) Analyses analogous to (D) for the DoHM mutant (114 complexes; $N_t = 2,584$ in range shown in (i). For (ii) - (iv) in (D, E), fit parameters and numbers of observations are reported in Table 1.
transitioned to a state exhibiting increased FRET ($E_{\text{FRET}} = 0.606 \pm 0.002$; Figure 8D-iii and 8D-iv, Table 1). Importantly, all of the complexes that transitioned to high $E_{\text{FRET}}$ remained in that state (Figure 8D-i and 8D-iv) until the observation was terminated by the end of the recording or by the disappearance of donor and/or acceptor fluorescence (either due to photobleaching or sliding of the double hexamer off the DNA; Figure 8B, black arrow). These observations are consistent with our previous studies showing that wild-type Mcm2-7 rapidly forms stable double hexamers upon recruitment of the second Mcm2-7 (Ticau et al., 2015). In the data collected shortly (~0-10 s) after arrival of the second Mcm2-7, there is some evidence for the presence of a transient state with $E_{\text{FRET}}$ intermediate between zero and the high value (Figure 8D-ii, black bar) but the additional state, if present, was not clearly resolved in the current experiments.

After arrival of the second Mcm2-7, the DoHM mutant also exhibited an initial state with similarly low FRET followed by a rapid transition on many of the DNA molecules to a state with high FRET (Figure 8C). Thus, like wild-type Mcm2-7, the DoHM mutant can form a structure in which the two hexamers are so closely apposed that the donor and acceptor fluorophores are separated by only a few nanometers. However, there are two noticeable differences in the $E_{\text{FRET}}$ time courses with the DoHM mutant helicase (Figure 8E). First, the elevated $E_{\text{FRET}}$ value observed for the DoHM mutant was distinct (0.468 ± 0.010) and significantly lower than that observed for wild-type Mcm2-7 (0.606 ± 0.002; Table 1). This suggests that the high $E_{\text{FRET}}$ state of the DoHM mutant has a three-dimensional structure and/or dynamics that differ from those of the wild-type double hexamers. Second, the DoHM mutant complexes that transition to the elevated $E_{\text{FRET}}$ state do not remain there.
Table 1. Fit parameters for $E_{\text{FRET}}$ distributions during double-hexamer formation$^\dagger$

<table>
<thead>
<tr>
<th>Protein</th>
<th>Parameter$^*$</th>
<th>$N$$^\ddagger$</th>
<th>Value</th>
<th>90% C.I.$^\triangle$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type Mcm2-7 (Fig. 8D) $N_{\text{molec}}=89$</td>
<td>$\mu_1$</td>
<td>5,300</td>
<td>0.029</td>
<td>[0.024, 0.034]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td></td>
<td>0.606</td>
<td>[0.603, 0.609]</td>
</tr>
<tr>
<td></td>
<td>$\sigma$</td>
<td></td>
<td>0.123</td>
<td>[0.121, 0.125]</td>
</tr>
<tr>
<td></td>
<td>$A_1$ (0–10 s)</td>
<td>254</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$A_1$ (10–30 s)</td>
<td>704</td>
<td>0.42</td>
<td>[0.38, 0.45]</td>
</tr>
<tr>
<td></td>
<td>$A_1$ (30–200 s)</td>
<td>4,342</td>
<td>0.30</td>
<td>[0.29, 0.31]</td>
</tr>
<tr>
<td>Mcm2-7$^{4-178A}$ (Fig. 8E) $N_{\text{molec}}=114$</td>
<td>$\mu_1$</td>
<td>4,908</td>
<td>0.038</td>
<td>[0.034, 0.042]</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td></td>
<td>0.468</td>
<td>[0.452, 0.482]</td>
</tr>
<tr>
<td></td>
<td>$\sigma$</td>
<td></td>
<td>0.143</td>
<td>[0.140, 0.146]</td>
</tr>
<tr>
<td></td>
<td>$A_1$ (0–10 s)</td>
<td>314</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$A_1$ (10–30 s)</td>
<td>866</td>
<td>0.54</td>
<td>[0.51, 0.58]</td>
</tr>
<tr>
<td></td>
<td>$A_1$ (30–200 s)</td>
<td>3,728</td>
<td>0.96</td>
<td>[0.96, 0.97]</td>
</tr>
</tbody>
</table>

$^\dagger$ $N_{\text{molec}}$: Number of complexes analyzed. Measurements with outlier $E_{\text{FRET}} > 1$ or $E_{\text{FRET}} < -0.35$ (1% for Mcm2-7 and 4.5% for Mcm2-7$^{4-178A}$ were excluded from fits. $^*$ $E_{\text{FRET}}$ data measured between 0 and 200 s after binding of the second Mcm2-7 were fit to a two-Gaussian mixture model that assumed time-invariant values for the standard deviation ($\sigma$) and positions of the low- and high-$E_{\text{FRET}}$ peaks ($\mu_1$, $\mu_2$), and fractional amplitudes of the two peaks ($A_1$ and $1 - A_1$) that can differ between the three time intervals. $^\ddagger$ Number of data points contributing to determination of each parameter value. $^\triangle$ 90% confidence intervals were determined by bootstrapping (250 samples).
Figure 9. Additional fluorescence records for experiments using a 1:1 mixture of donor- and acceptor-labeled wild type Mcm2-7 (A) or the DoHM mutant (B) showed FRET after arrival of the second Mcm2–7. Panels i-iv are as described in Figure 8B and 8C.
Instead, the helicases quickly (in ~20 s) transition back to a state in which both hexamers are still bound to the DNA but which display low FRET (Fig. 8E-i and iv). Together, these data indicate that the double-hexamer interactions formed by the DoHM mutant are incomplete and unstable.

**Mcm2-7-Mcm2-7 interactions occur rapidly or not at all**

To examine the formation and stability of the double-hexamer complexes, we analyzed all DNA molecules associated with one donor- and one acceptor-labeled Mcm2-7 to determine the timing of transitions between the low-$E_{\text{FRET}}$ and the high-$E_{\text{FRET}}$ states observed in the wild-type and the DoHM mutant FRET records. Both of these data sets contained significant subpopulations of DNA molecules that have bound both donor- and acceptor-labeled Mcm2-7 but which do not exhibit FRET at any point in the recording. This suggests that some complexes exist in a configuration that is refractory to double-hexamer formation, a configuration that we will refer to as non-double hexamer (non-DH). Therefore, a minimal kinetic scheme to explain the wild-type Mcm2-7 data contains at least three molecular states: the high $E_{\text{FRET}} = \sim0.45$-0.6 state that we interpret as double-hexamer (DH) and two states with indistinguishable $E_{\text{FRET}} = \sim0$ values, but that can be distinguished by their lifetimes: the long-lived non-DH state and the transient pre-double-hexamer (pre-DH) state that exists in the short time interval between binding of the second Mcm2-7 and the formation of a DH (see Figure 8D). We globally fit our set of wild-type Mcm2-7 FRET data to a kinetic model containing these three states (Figure 10A). The fitting algorithm (see Methods) yielded estimates for all six possible first-order rate constants connecting the three states (Figures 10A and 11A) and also assigns one
of the three states to each molecule at every time point in the record (Figure 10C) based on state duration and $E_{\text{FRET}}$.

The kinetic modeling for wild-type Mcm2-7 reveals that 63% (52 of 82) of second Mcm2-7 binding events resulted in double-hexamer formation, all of which occur within thirty seconds of the second Mcm2-7 arrival. In most cases (41 of 52), these molecules start in the pre-DH state and rapidly convert to the DH state with a rate constant of $0.150 \pm 0.075$ s$^{-1}$ (Figure 10A). The rate constant for the reverse transition from DH to pre-DH is not significantly different from zero (Figure 11A), suggesting that the pre-DH to DH transition is essentially irreversible. These data are consistent with the interpretation that the wild-type Mcm2-7 double hexamer is stable once formed (Figure 8B, 9 and 10A, Abid Ali et al., 2017; Li et al., 2015; Noguchi et al., 2017). Fourteen percent (11 of 82) of molecules were assigned to already be in DH state at the first time point. These events are likely to represent molecules whose sojourn in pre-DH state was too short to be detected in an experiment with this time resolution (~2.67 s). In contrast to the transition between pre-DH and DH, we do not detect significant non-zero rate constants for entering or leaving the non-DH state, consistent with the interpretation that the 35% of molecules (29 of 82) assigned to non-DH at the first time point are a refractory population trapped in a dead-end state (see Discussion). Together, these data indicate that once two wild-type Mcm2-7 complexes associate with DNA they either rapidly transition to a double hexamer or remain as a pair of single hexamers.
Figure 10

A  WT Mcm2-7

\[
\text{pre-DH} \quad \text{Low FRET} \quad \text{IF: 50%} \\
\text{non-DH} \quad \text{Low FRET} \quad \text{IF: 35%} \\
\text{DH} \quad \text{High FRET} \quad \text{IF: 14%} \\
\]

B  DoHM mutant

\[
\text{pre-DH} \quad \text{Low FRET} \quad \text{IF: 63%} \\
\text{non-DH/post-DH} \quad \text{Low FRET} \quad \text{IF: 25%} \\
\text{pseudo-DH} \quad \text{High FRET} \quad \text{IF: 13%} \\
\]

C  WT Mcm2-7

D  The DoHM mutant
Figure 10. Wild-type Mcm2-7 and the DoHM mutant attempt double-hexamer formation only once. (A-B) Kinetic models of transition among the three indicated $E_{\text{FRET}}$ states. Pooled single-molecule fluorescence records from helicase loading experiments using (A) Mcm2-7 or (B) Mcm2-74-178A were globally fit (see Materials and Methods) to generalized three-state kinetic models (i.e., all possible inter-state transitions were allowed). Black arrows indicate kinetically significant reaction steps; gray arrows indicate steps with rate constants that are not significantly greater than zero (see Figure 11). Fits also yielded the fraction (IF) of molecules in each state at the time of second hexamer binding. In the wild-type Mcm2-7 model (A), the transition from pre-DH to DH is the only process that occurs at an appreciable rate, while in the DoHM mutant model (B) both pre-DH to pseudo-DH and pseudo-DH to post-DH steps are significant. (C-D) State rastergrams from the kinetic models. Each trajectory (horizontal line) indicates the state assigned by the model to a single-molecule record with Mcm2-7 (C) or the DOHM mutant (D), starting from the time of second hexamer binding. Trajectories are sorted by the onset time of the DH or pseudo-DH state, then by record length. The blank region at the bottom of each rastergram represents low signal-to-noise records excluded from analysis (see Materials and Methods). Rare transitions (e.g., occasional transitions into or out of non-DH in wild-type Mcm2-7) which do not occur at statistically significant rates may reflect ambiguities in assigning time segments, particularly when the transitions are between two states with the same (zero) $E_{\text{FRET}}$. 
We also analyzed the Mcm2-7^4-178A \( E_{\text{FRET}} \) data by fitting it to a three-state model analogous to that used for the wild-type data. Three characteristics of the resulting analysis are similar between wild-type and the DoHM mutant (Figure. 10B). First, in most cases, the arrival of the second Mcm2-7 was followed by a rapid transition from an initial very low-FRET state (pre-DH) to a higher \( E_{\text{FRET}} \) state (62 of 100), or the high-FRET state was present at the first time point (13 of 100). Second, these transitions occurred rapidly and with a similar rate constant (0.13 ± 0.04 s\(^{-1}\)) to wild-type (0.15± 0.08). Third, the remaining second Mcm2-7 binding events (25 of 100) never exhibited the high-FRET state and were categorized as in the non-DH state. Thus, like the wild-type helicase, the DoHM mutant either rapidly exhibited a high-FRET state or remained in the low-FRET state throughout the observations.

There were two significant differences in the kinetics of double-hexamer interactions exhibited by the DoHM mutant helicase. As mentioned earlier, the high \( E_{\text{FRET}} \) state of the mutant is different from that of the DH state of the wild type (\( E_{\text{FRET}}^{\text{DoHM}} = 0.468 \pm 0.010; E_{\text{FRET}}^{\text{WT}} = 0.606 \pm 0.002; \) Figures 8D, E and 11B). Thus, we refer to this state as the pseudo-double hexamer (pseudo-DH). A second distinction is the short-lived nature of this state. The DoHM mutant rapidly (0.052 ± 0.007 s\(^{-1}\)) transitioned from the pseudo-DH state back to a low-FRET state (Figure 10B). Interestingly, the state that follows the pseudo-DH is a long-lived low-FRET state, which we termed post-double-hexamer (post-DH). This state is kinetically distinct from the short-lived low-FRET pre-DH state (Figure 10D) but is indistinguishable from the non-DH state; both are long-lived and exhibit similar low-FRET values. Thus, these states are grouped together into a single non-DH/post-DH
**Figure 11.** Complete set rate constants (s\(^{-1}\)) and \(E_{FRET}\) values derived from the kinetic model fits (Figure 10). Asterisks indicate significantly non-zero rate constants.
state for the kinetic analysis (Figure 10B,D). Although the mutant pre-DH molecules rapidly transitioned to pseudo-DH and then to post-DH, the rates of the reverse reactions (pseudo-DH to pre-DH and post-DH to pseudo-DH) were both calculated to be >100 times slower and not significantly non-zero (Figure 11B). These findings suggest that both of these transitions are essentially irreversible and that once the two mutant Mcm2-7 complexes that form a pseudo-DH separate they will not form subsequent stable interactions. Records that were interpreted as showing a transition from non-DH/post-DH back to the high $E_{\text{FRET}}$ pseudo-DH state were rare (3 of 100) and sometimes showed only a single time point in the pseudo-DH state (Figure 10D, e.g. trajectory 67), possibly arising from infrequent errors in state assignment. Even though most or all post-DH molecules never returned to a high $E_{\text{FRET}}$ state, we find that both the DoHM mutant complexes remain associated with the DNA for extended times (Figures 10D and 9), consistent with both helicases having successfully encircled the DNA. These findings support a model in which there is a limited window of opportunity to form double-hexamer interactions during helicase loading after which two loaded Mcm2-7 complexes cannot interact.

We also noted that the lifetime of two single hexamers associated with the DNA is noticeably shorter than that of a double hexamer (Figures 10C and D, non-DH and post-DH states, blue lines, note that release of either of the two Mcm2-7 present results in the end of the record as we are looking for double-hexamer formation). Two wild-type Mcm2-7 complexes in the context of a double-hexamer can remain on the DNA for over 600 s, with many lasting to the end of observation (Figure 10D, yellow lines). In contrast, the retention of both of two single hexamers with the DNA whether they are wild-type or DoHM mutants is much shorter (~300 s, Figures 10C and D, blue lines). The average
duration of DNA association for two wild-type Mcm2-7 complexes is 288 ± 23 s compared to 140 ± 9 s for the DoHM mutant. This observation suggests that the single Mcm2-7 complexes are less stably associated with the DNA relative to Mcm2-7 complexes in the form of double-hexamers. This difference is consistent with the reduced loading observed in the bulk experiments with the DoHM mutant.

**Double-hexamer formation is required for later steps of replication initiation**

Our analysis of the Mcm2-74-178A mutant indicates that the DoHM is important for double-hexamer formation and stability. Despite these defects, many of the mutant single hexamers remain stably bound to DNA. This allowed us to ask which, if any, of the subsequent steps in helicase activation and replication initiation can occur with only single-hexamers loaded onto DNA. To this end, we assessed the ability of the DoHM mutant to associate with helicase-activating proteins, unwind origin DNA and replicate DNA. Consistent with the lethality of this mutant, the DoHM mutant did not support DNA synthesis in an *in vitro* replication assay using purified replication proteins (Figure 12A; Yeeles et al., 2015).

To identify the step during replication initiation that was defective, we first assayed association of helicase-activation proteins with wild-type and the DoHM mutant proteins. We used bead-attached templates to monitor association of Cdc45, GINS and Mcm10 with origin DNA associated with the indicated Mcm2-7 (Löoke et al., 2017). Although there is less association of these proteins with the DoHM mutant, the reduction corresponds with the reduced loading of this version of Mcm2-7 (Figures 6C and 12C). Thus, DDK-
dependent association of Cdc45, GINS and Mcm10 with Mcm2-7-associated DNA was not compromised in the context of DoHM Mcm2-7 single hexamers (Figure 12B,C).

We also addressed the impact of the DoHM mutant on origin DNA unwinding using an assay that detects the formation of supercoiled DNA as a consequence of the unwinding process (Douglas et al., 2018). We initially used a 3.8 kb circular template that allows detection of extensive DNA unwinding. Consistent with the lack of replication, the DoHM mutant showed no evidence of extensive DNA unwinding (Figure 12D) in this assay, as indicated by the lack of formation of supercoiled DNA. We performed the same assay with a ~600 bp circle that allows detection of small topological changes that are formed after CMG formation but do not require Mcm10 action (Douglas et al., 2018). Consistent with productive Cdc45 and GINS association with the DoHM mutant, we observed equivalent amounts of these initial intermediates for wild-type and DoHM proteins (Figure 12E). Both the small topological changes observed with the DoHM mutant and the finding that these intermediates are formed in the presence or absence of Mcm10 indicated that they are the result of initial DNA melting. Together, these findings indicate that CMG formation and initial DNA melting/distortion are independent of double-hexamer formation but more extensive origin DNA unwinding requires these interactions.

**DISCUSSION**

Our findings identified an essential and highly conserved motif in Mcm4 that is required for stable double-hexamer formation but not initial loading of Mcm2-7. Using single-molecule studies, we show that this mutant does not prevent initial interactions between
Figure 12. Double-hexamer formation is essential for DNA replication initiation. (A) The DoHM mutant is defective in DNA replication. The Mcm2-7 complexes were tested in a reconstituted replication assay. The helicase was loaded onto DNA template and subsequently phosphorylated by DDK. Replication proteins, dNTPs and \([\alpha^{32\text{P}}]dCTP\) were added to initiate DNA replication, and the replication product is detected by phosphor imaging. Initiation of DNA replication is dependent on DDK phosphorylation. (B) The DoHM mutant can recruit Cdc45, GINS, and Mcm10. Mcm2-7 complexes were tested for CMG formation. The helicase was loaded onto bead-coupled DNA template and subsequently phosphorylated by DDK. The previous reaction mix was removed prior to addition of replication proteins. Bead-associated proteins were washed with high-salt buffer and detected by immunoblot. Omission of DDK was used as a control for non-specific DNA binding. (Mcm2-7 loading is DDK-independent). (C) The DoHM mutant defect for Cdc45 and GINS recruitment is correlated with the reduced helicase loading observed for this form of Mcm2-7. Data from three independent CMG-formation experiments (including that shown in B) were quantified using ImageJ. Error bars indicate the SD. For each Cdc45 and GINS bar, the -DDK signal was subtracted.
from the corresponding +DDK signal. (D) The DoHM mutant is defective for extensive DNA unwinding. Mcm2-7 was loaded onto Topo I-relaxed 3.8 kb plasmids followed by addition of S-CDK, DDK and helicase-activating proteins to stimulate CMG formation and activation. The reaction was quenched by SDS, separated on an agarose gel, and detected by ethidium bromide staining. DNA unwinding is detected by the formation of supercoiled DNA and is dependent on DDK phosphorylation. (E) The DoHM mutant allows initial topological changes that occur after CMG formation. Reactions were performed as in (D) except Mcm2-7 complexes were loaded onto a 616 bp circle and unwinding products were detected by SYBR Gold staining. Control reactions lacking DDK or Mcm10 are indicated. Identity of topoisomers is indicated on the right (see Figure 13).
Figure 13. Assignment of relative supercoiling states for Figure 12. Nicked, 616-bp DNA circles were ligated in the presence of the indicated concentrations of ethidium bromide (EtBr). DNA circles are increasingly negatively supercoiled as the concentration of EtBr increases. SCC = supercoiled closed-circle before nickase treatment. This state captures the $\alpha$ ground state and +1 supercoiling state. NC = nicked circle.
the two Mcm2-7 hexamers but that the resulting interactions are incomplete and unstable. Interestingly, these analyses reveal evidence consistent with a limited window of opportunity for double-hexamer formation after recruitment of the second Mcm2-7. Finally, using a DoHM mutant that prevents stable double-hexamer formation, we demonstrate that the Mcm2-7 double hexamer is required for extensive origin DNA unwinding but not initial recruitment of helicase-activating proteins or origin melting.

**Double-hexamer formation is a facilitated event**

There are two general models to explain how two Mcm2-7 complexes form a double hexamer. The first model proposes that the two Mcm2-7 helicases are loaded independently, then slide along DNA, and form a double hexamer through interactions between their N-terminal domains independent of other proteins (Coster and Diffley, 2017). An alternative model is that helicase-loading proteins are required to facilitate both loading of the Mcm2-7 complexes and double-hexamer formation. Consistent with the latter model, kinetic analyses of the single-molecule studies presented here provide evidence that double-hexamer formation is facilitated. For both wild-type and the DoHM mutant, the formation of FRET between Mcm7 subunits after arrival of the second Mcm2-7 is rapid ($k_{DHex} = 0.13-0.15 \text{ s}^{-1}$; Figures 8D-i, 8E-i, 10A and 10B). Indeed, all double-hexamer formation observed occurs within 30 seconds after arrival of the second Mcm2-7. In addition, double hexamer interaction is detected significantly sooner ($1/0.15 \text{ s}^{-1} = \sim 7 \text{ s}$) after second Mcm2-7 arrival than second Mcm2-7 ring closure (\sim 57 s, Ticau et al., 2017), indicating that interactions between the hexamers anticipates completion of loading of the second Mcm2-7. These comparisons argue against a model in which
helicase loading for both hexamers is complete at the time hexamer-hexamer interactions first occur. Instead, our findings strongly suggest that loading of the second hexamer and double-hexamer formation are coordinated.

We propose that double-hexamer formation requires one or more helicase-loading proteins and that the limited dwell times of these proteins on DNA results in a short window of opportunity for double-hexamer formation. This hypothesis is consistent with the observation that, in both wild-type Mcm2-7 and the DoHM mutant cases, two Mcm2-7 complexes either rapidly transition to a high-FRET state after the second Mcm2-7 arrives or do not visit the high-FRET state at all (Figure 10). Similarly, neither the rare wild-type nor frequent DoHM mutant instances that show a high-FRET to low-FRET transition ever displayed a second interval of high FRET (0 of 3, Figure 10C and 0 of 75, Figure 10D). These observations suggest that the establishment of the hexamer-hexamer interactions resulting in high FRET are associated with an irreversible step that prevents subsequent high-FRET transitions if that initial complex is unstable. If two Mcm2-7 complexes could form a double-hexamer unassisted, we would expect subsequent transitions into the high-FRET state in these cases, as the two non-interacting Mcm2-7 complexes frequently remain on the DNA for >100 seconds. The long dwell times of the single hexamers also argues against models in which the instability of single Mcm2-7 complexes on DNA prevents a second attempt to form the double hexamer. The hypothesis that helicase-loading proteins facilitate double-hexamer formation is consistent with previous single-molecule studies showing that Cdc6, Cdt1 and ORC are all released from the DNA shortly after arrival of the second Mcm2-7 (Ticau et al., 2015).
Finally, such a mechanism would have the advantage of only allowing double-hexamer formation during loading and preventing such interactions from occurring at other times of the cell cycle (e.g. as replication forks converge).

Although release of any of the helicase-loading factors could prevent second attempts to form a double hexamer, the average release time of the second Cdc6 after arrival of the second Mcm2-7 (~23 s) would best fit the window of opportunity that we observe (i.e. DH are not formed longer than 30 s after second Mcm2-7 arrival). Another possibility is that closure of the Mcm2-7 ring terminates the ability of Mcm2-7 to form a double hexamer. However, the average time for this event relative to arrival of the second Mcm2-7 (~57 s) is much longer than the average time for double-hexamer formation after the second Mcm2-7 arrives, making release of Cdc6 a more likely candidate for the process that prevents subsequent double-hexamer formation.

We did not see a dramatic difference in the number of second Mcm2-7 association events for the DoHM mutant, suggesting that the interactions interrupted by this mutant are not required to recruit the second Mcm2-7 to the DNA. This is consistent with models in which the recruitment of two Mcm2-7 complexes is independent of Mcm-Mcm interactions (Coster and Diffley, 2017). We note, however, that, under the conditions of our reactions, only one ORC molecule is required for these two events (Ticau et al., 2015). It remains possible that the initial recruitment of the second Mcm2-7 involves Mcm-Mcm interactions that are not detected by the FRET probe or inhibited by the DoHM mutant. Development of additional modified Mcm2-7 complexes to detect interactions at other sites may help to
address this possibility.

Although there is no difference in the number of second Mcm2-7 association for the DoHM mutant, we observed a shorter lifetime of the association of two DoHM mutant complexes with DNA (average 140 ± 9 s, Figure 10D) compared to wild-type Mcm2-7 (average 288 ± 23 s, Figure 10C). The most likely explanation of this difference is that Mcm2-7 single hexamers have a reduced stability on DNA relative to a Mcm2-7 double hexamer. Although previous studies showed that single Mcm2-7 hexamers can be loaded onto the DNA in a salt-stable manner (Ticau et al., 2015), it is possible that, the single DoHM mutant complexes are more prone to dissociate from the DNA. In support of this hypothesis, we see the same shorter dwell times for pairs of wild-type Mcm2-7 that fail to form a double-hexamer (Figure 10C, blue lines). This is consistent with the decreased helicase loading of the DoHM mutant in the ensemble loading assay (Figure 6C). Although it is also possible that the decreased loading is due to single hexamers sliding off the of the linear DNA substrate used in the initial bulk assays (e.g Figures 4-6), this is unlikely because we see the same reduction in DoHM loading when a circular DNA template is used (Figure 12B,C).

We note that in both the wild-type and DoHM mutant experiments there are frequent instances when two Mcm2-7 complexes associate with the DNA for long periods of time but show no transitions to high FRET (Figure 10C and D, non-DH traces). As discussed above, one possibility is that these molecules have lost a helicase-loading protein(s) required for double-hexamer formation before establishing this complex. Alternatively, the
lack of double-hexamer formation in these cases could be because the two Mcm2-7 complexes are loaded on the DNA in an incorrect orientation. For example, two sequential “first” loading events on the same DNA would result in two Mcm2-7 complexes in the same orientation. We note that there are many instances (primarily for the DoHM mutant) in which initial hexamer-hexamer interactions are unstable, but the separated hexamers do not make a second attempt to form a stable double hexamer. In these cases, the second hypothesis cannot explain the lack of a second attempt to form the double hexamer as the two Mcm2-7 complexes must have been in the appropriate orientation to form the initial high-FRET (pseudo-DH) interaction.

**Structure of double-hexamer and the DoHM interactions**

The lower $E_{\text{FRET}}$ value of the mutant pseudo-DH state suggests that the DoHM mutation interferes with proper interactions at the double-hexamer interface. This hypothesis is consistent with the cryo-EM structures of the Mcm2-7 double hexamer (Abid Ali et al., 2017; Li et al., 2015; Noguchi et al., 2017). In each of the structures, the DoHM is part of a key double-hexamer interface. The DoHM forms a loop that interacts with Mcm5 N-terminal domain from the opposite hexamer. In addition, the DoHM-interacts with one end of an extended Mcm7 alpha helix from the same hexamer. Importantly, the other end of this Mcm7 helix interacts with Mcm5 of the opposite hexamer (Figure 14), raising the possibility that the DoHM positions this helix to appropriately interact with the other hexamer. These interactions are present in the Mcm2-7 double-hexamer structures both with and without DNA (Abid Ali et al., 2017; Li et al., 2015; Noguchi et al., 2017). There are two sets of these three-subunit interactions (Mcm4-Mcm5-Mcm7) located on opposite
Figure 14. Location of the Mcm4 DoHM in the context of the Mcm2-7 double hexamer bound to DNA (Abid ali and Costa, 2017). The double hexamer interface is in the center of the image. The DoHM of Mcm4 (green), the region of Mcm7 that interacts with the DoHM (light blue), and the region of Mcm5 from the opposite hexamer that interacts with the DoHM (yellow) are all shown in space filling representation. The remainder of the structure is shown as a ribbon structure with Mcm6 in orange, Mcm2 in magenta, and Mcm3 in dark blue.
Figure 15. There are two DoHM-involved interfaces that are located on opposite sides of the Mcm2-7 double hexamer (Abid Ali and Costa, 2017). The DoHM of Mcm4 is represented in dark green space filling. The remainder of the structure is shown as a ribbon structure with Mcm4 in green, Mcm6 in orange, Mcm2 in magenta, Mcm5 in yellow, Mcm3 in dark blue, and Mcm7 in light blue.
sides of the double-hexamer interface (Figure 15). Thus, although the DoHM-is a small region compared to the large double-hexamer interface, disrupting this motif has the potential to interrupt the interactions involving 6 of the 12 subunits participating in the double-hexamer interface.

The double hexamer is required for extensive DNA unwinding

In this study, we show that a double hexamer is not required to recruit helicase-activation proteins or perform initial DNA unwinding. Although it was assumed that double hexamers are an essential part of DNA replication initiation, the replication initiation step that required the double hexamer was unknown. The DoHM mutant allowed us to evaluate the role of the double-hexamer complex. We found that the DoHM mutant can assemble CMG complexes that perform limited initial DNA unwinding (Figure 12). This degree of topological change corresponds to the modest unwinding of less than one turn of DNA per helicase previously observed as a consequence of CMG formation (Douglas et al., 2018). The observation of this initial DNA unwinding indicates that the interactions of Cdc45 and GINS with the DoHM mutant are productive. Since the DoHM mutant does not form stable double hexamers, this initial unwinding is likely to arise from independent helicases. We note that based on the measured lifetimes of the DoHM mutant pseudo-double-hexamers, helicase-activation proteins are not in a position to stabilize the pseudo-double hexamers before they dissociate. In these experiments, DDK phosphorylation is performed for 20 min prior to addition of the helicase-activation proteins, which is much longer than the ~20 s (1 / 0.052 s⁻¹; Figure 10B) on average that the DoHM mutant is in the pseudo-DH state. We also note that once separated, we
observe no evidence of subsequent hexamer-hexamer interactions. Thus, our findings strongly suggest that this initial unwinding occurs by manipulation of the DNA within a single CMG complex rather than through the coordinated action of two opposing helicases.

In contrast to the ability of the DoHM mutant to catalyze initial DNA unwinding, this form of the helicase cannot perform more extensive DNA unwinding that requires Mcm10. This result strongly suggests that double-hexamer interactions are required to drive this DNA unwinding. If true, this hypothesis would suggest that additional DNA unwinding outside the central channel of the helicase requires a tight interface between the two helicases. One way this tight interface could be necessary is if opposing action of the two helicases drives DNA unwinding and formation of the replication bubble. A requirement for the double-hexamer could occur before the two hexamers separate or in a second event that involves the two hexamers coming back together after they separate during CMG formation. The latter hypothesis is consistent with recent studies suggesting that CMG formation causes double-hexamer separation followed by an Mcm10-dependent transition involving two CMG complexes (Douglas et al., 2018; Langston and O'Donnell., 2019). Taken together with previous studies, our studies support a model in which DNA unwinding during DNA replication initiation occurs in at least three stages: 1) initial limited DNA melting upon CMG formation, 2) further DNA unwinding that results in ssDNA-strand extrusion that requires double-hexamer interactions and Mcm10, and 3) extensive DNA unwinding that is mediated single CMG complexes translocating on ssDNA at replication forks.
We also considered whether the DNA unwinding defects observed for the DoHM mutant were due to Mcm10-binding or ATPase defects, however, other data argue against these explanations. First, Mcm2-7 ATPase activity is required for DNA unwinding beyond that detectable upon CMG formation. We note, however, that the DoHM mutant is functional for helicase loading, which also requires Mcm2-7 ATPase activity (Coster et al., 2014; Kang et al., 2014) making an ATPase defect unlikely. Second, although Mcm10 is required for the transition from initial to more extensive unwinding, multiple findings argue against a model in which the DoHM mutant has a defect in Mcm10 binding. Although the N-terminal domain of Mcm4 has been observed to interact with Mcm10 in the absence of the rest of Mcm2-7 (Quan et al., 2015), in the context of the full complex, Mcm10’s interactions with Mcm2 and Mcm6 are more important (Douglas and Diffley, 2016; Lõoke et al., 2017; Mayle et al., 2019). Additionally, crosslinking studies of the CMG-Mcm10 structure showed only three Mcm10 crosslinking sites with Mcm4, all of which are distant from the DoHM mutation (Mayle et al., 2019). Most importantly, we found that the DoHM mutant maintains the ability to form high-salt-resistant interactions with Mcm10 (Fig. 12).

The identification of mutations that prevent double-hexamer formation provides an important tool to further explore the role of this complex in replication initiation. Although we have provided a first look of the role of the double hexamer in initiation, there is more to be understood about how this intermediate facilitates protein associations and origin unwinding. In addition, although our studies provide strong evidence that the loading of the second Mcm2-7 is coordinated with formation of the double hexamer, the mechanism of this coordination is an important area for future investigation.
MATERIALS AND METHODS

Protein purifications
Wild-type Mcm2–7/Cdt1 and ORC complexes were purified as described previously (Kang et al., 2014). Wild-type Cdc6 was purified as described in Frigola et al. (2013). DDK, S-CDK, Sld3/7, Cdc45, Sld2, Dpb11, GINS, Mcm10, Polymerase epsilon, Polymerase alpha/primase, Polymerase delta, RPA, Ctf4, RFC, PCNA, Mrc1, Csm3-Tof1, and Topo II were purified as described in Lõoke et al. (2017). Mutant Mcm2-7/Cdt1 complexes were purified as described in Kang et al., (2014) with the following modifications. For each Mcm2-7 mutant complex, the corresponding wild-type proteins were epitope-tagged with either c-Myc or V5. In the strains expressing the Mcm2 Δ2-177 and Mcm6 Δ2-105, the wild-type MCM2 and MCM6 genes were tagged with c-Myc, respectively to allow the endogenous 13Myc-tagged Mcm2 or 6 subunits to be depleted by incubating with anti-c-Myc agarose (Sigma) before applying the Mcm2-7 mutant complex to a Superdex 200 gel filtration column. In strains expressing mutant Mcm4 protein, the wild-type MCM4 gene was tagged with V5. This Mcm2-7 complexes containing the endogenous V5-tagged Mcm4 subunits to be depleted by incubating with anti-V5 agarose (Sigma) before application to a Superdex 200 gel filtration column. Yeast strains and plasmids used are listed in Table 2 and 3, respectively.

Helicase-loading and double-hexamer formation assays
Helicase-loading and double-hexamer formation assays were performed as described in Kang et al., (2014).
**Fluorescent labeling of wild-type Mcm2-7/Cdt1 and Mcm2-7^{4-178A}/Cdt1**

SORT-tagged wild-type Mcm2-7/Cdt1 was purified and labeled with either DY549-P1 or DY649-P1 (Dyomics) as described in Ticau et al., (2015). SORT-tagged Mcm2-7^{4-178A}/Cdt1 was purified and labeled using the same protocol with the following modifications. In the strain expressing the Mcm4-178A, the wild-type MCM4 gene was tagged with V5 to allow the endogenous V5-tagged Mcm4 subunits to be depleted by incubating with anti-V5 agarose (Sigma) before dye coupling as described above.

**Determining fractional labeling of Mcm2-7**

To determine what fraction of Mcm2-7 molecules were fluorescently labeled, 20 ml of DY549-SORT labeled Mcm2-7 was mixed with maleimide-DY-649P1 dissolved in anhydrous DMSO, in a 1:1 molar ratio at 4°C for 10 min. The reaction was terminated with 2 mM DTT. The double-labeled Mcm2-7 (10 nM) was added to a slide coupled to origin DNA, 0.5 nM of purified ORC, 2 nM of purified Cdc6 and monitored Mcm2-7-DNA colocalization (to ensure that we were monitoring fully assembled complexes). The fraction of maleimide-DY-649P1-labeled Mcm2-7 molecules that also contained DY-549P1 was determined and reported as the percent labeling by the DY-549P1 (we assume that coupling of maleimide-DY-649P1 to Mcm2-7 is not influenced by the presence or absence of the 549 label). The same protocol was used for DY649-SORT labeled Mcm2-7, but this complex was double-labeled with maleimide-DY-549P1.

**Single-molecule microscopy and FRET data analysis procedure**

Single molecule experiments were performed as described in Ticau et al., (2015) except
that fluorescent-beads were used for drift correction. Spots with DNA colocalization of one donor-fluorophore and one acceptor-fluorophore labeled were manually selected for $E_{FRET}$ calculation. To calculate apparent FRET efficiencies, each fluorescence intensity trace was background subtracted using custom Matlab (MATHWORKS) image processing software that has been previously described (Friedman and Gelles, 2012). For each trace, baseline segments (trace intervals lacking any spot) were joined after smoothing with a low-pass filter, and that smoothed baseline was subtracted from the initial fluorescence trace. FRET efficiency was calculated using $E_{FRET} = \frac{I_{\text{Acceptor}}}{I_{\text{Acceptor}} + I_{\text{Donor}}}$ where $I_{\text{Acceptor}}$ and $I_{\text{Donor}}$ are the acceptor and donor emission intensities observed during donor excitation, respectively. The two-dimensional Gaussian kernel histograms and the one-dimensional histogram fits in Figure 8 were generated using code from https://github.com/gelles-brandeis/jganalyze. Two-dimensional histograms used bandwidths 5 s and 0.05 on the time and $E_{FRET}$ axes, respectively and were normalized so that the probability density in each 2.67 s time slice integrated to one.

**Kinetic analysis**

Pooled single-molecule fluorescence records from helicase loading experiments from Mcm2-7 or Mcm2-74-178A were globally fit to generalized three-state kinetic models (i.e. all possible inter-state transitions were allowed). Fitting was performed using coupled hidden Markov models using an empirical Bayesian approach to estimate priors (van de Meent et al. 2014) as implemented in program ebfret-gui (https://github.com/ebfret/ebfret-gui/commit/28e548ace84190c91c4ca354f41efa5952a7895f). Outlier points with $E_{FRET}$ outside the range [-0.25, 0.85] (3.5% of Mcm2-7 data and 7% of Mcm2-74-178A data) were
excluded from the analysis as were individual DNA molecule records (7 Mcm2-7 and 11 Mcm2-7-4-178A) with anomalously low signal-to-noise as judged by their containing more than 10 outliers each. Default priors were used except that the state $E_{FRET}$ values were strongly constrained to those from independent fits (in Fig. 8 and Table 1) by setting the (hyper)parameters $\mu$ and $\beta$ (van de Meent et al. 2014) to (0.06, 0.06, 0.57) and (1,000, 1,000, 1,000) for Mcm2-7 and (0.08, 0.08, 0.47) and (10,000, 10,000, 10,000) for Mcm-4-178A.

**CMG-Formation Assay**

The DNA plasmid template pUC19-ARS1 was randomly biotinylated and coupled to streptavidin-coated magnetic beads as described previously (Heller et al. 2011). Each incubation step was performed in a thermomixer (Eppendorf) with shaking at 1250 rpm at 25°C. Mcm2-7 loading was performed by incubating 0.48 pmol of ORC, 0.52 pmol of Cdc6, and 1.14 pmol of Mcm2-7/Cdt1 with 0.125 pmol template DNA in 25 mM HEPES-KOH (pH 7.6), 10 mM magnesium acetate, 225 mM potassium glutamate, 2 mM DTT, 0.02% NP-40, 5% glycerol, 5 mM ATP, 20 mM phosphocreatine, and 0.2 μg of creatine kinase for a total volume of 10 μL. Reactions were incubated for 20 minutes, at which point 1.3 pmol of DDK was added and incubation was continued for a further 20 minutes. The supernatant was then removed by applying the reaction to a DynaMag-2 magnet (ThermoFisher Scientific) to isolate the DNA coupled to magnetic streptavidin beads from the supernatant. CMG formation was then initiated by adding 20 μL of 0.6 pmol CDK, 1 pmol Sld3/7, 1 pmol Cdc45, 1.24 pmol Sld2, 0.8 pmol Dpb11, 5 pmol GINS, 0.06 pmol Mcm10, 1.05 pmol RPA, and 0.6 pmol Pol ε in 25 mM HEPES-KOH (pH 7.6), 10 mM
magnesium acetate, 250 mM potassium glutamate, 1 mM DTT, 0.02% NP-40, 8% glycerol, 5 mM ATP, and 0.4 mg/ml BSA and incubated for 30 minutes. Reactions were washed with 300 mM potassium chloride, 25mM HEPES-KOH (pH 7.6), 5 mM magnesium acetate, 10% glycerol, and 0.01% NP-40 three times. Proteins were released from the DNA by incubating with 5 U of DNase I (Worthington) in 10 µL of 25 mM HEPES-KOH (pH 7.6), 5 mM magnesium acetate, 200 mM sodium chloride, 5% glycerol, 0.02% NP-40, and 2 mM calcium chloride for 20 min at 25°C before immunoblotting.

**Soluble DNA-Replication Assay**

Each incubation step was performed in a thermomixer (Eppendorf) with shaking at 1250 rpm at 25°C. Mcm2-7 loading and DDK phosphorylation was performed in the same conditions as the CMG formation assay except with a soluble 11.9 kb pUC19-ARS1 plasmid template. After DDK phosphorylation, replication was initiated by adding 20 µL of 0.6 pmol CDK, 1 pmol Sld3/7, 2.6 pmol Cdc45, 1.24 pmol Sld2, 0.8 pmol Dpb11, 5 pmol GiNS, 0.02 pmol Mcm10, 0.6 pmol Pol ε, 1.5 pmol Pol α, 0.5 pmol Topo II, 0.6 pmol Ct4, 2.32 pmol RPA, 0.5 pmol RFC, 0.4 pmol PCNA, 0.5 pmol Mrc1, 0.6 pmol Csm3/Tof1, 0.6 pmol Pol δ in 12.5 mM HEPES-KOH (pH 7.6), 5 mM magnesium acetate, 125 mM potassium glutamate, 1 mM DTT, 0.01% NP-40, 4% glycerol, 1.5 mM ATP, 10 mM phosphocreatine, 3 µg of creatine kinase, 0.2 mg/ml BSA, 100 µM rNTP, 10 µM dNTP, and 10 µCi [α-P32]dCTP directly to the reaction. Following 60 minutes of incubation, reactions were quenched with 30 µL of 50 mM EDTA. Unincorporated nucleotides were removed with illustra MicroSpin G-50 columns (GE Healthcare), and samples were separated on a 0.6% alkaline agarose gel in 30 mM sodium hydroxide, 2
mM EDTA. Gels were dried and autoradiographed with Amersham Hybond-XL (GE Healthcare) and imaged using a phosphor screen. Gels were scanned using a Typhoon phosphorimager (GE Healthcare).

**DNA-Unwinding Assays**

For DNA unwinding assays using the 3.8 kb template, 25 fmol soluble 3.8 kb pUC19-ARS1 plasmid template was relaxed with 0.4 pmol Topo I for 30 minutes. Each incubation step was performed in a thermomixer (Eppendorf) with shaking at 1250 rpm at 25°C unless otherwise indicated. Twenty-five minutes of Mcm2-7 loading and 30 minutes of DDK phosphorylation were performed in the same conditions as the CMG formation assay. DNA unwinding was then initiated using the same protein concentrations and final buffer concentrations as the CMG formation assay with the addition of 0.4 pmol of Topo I. This mix was added directly to the reaction. After 40 minutes, the reaction was quenched with 13 mM EDTA, 0.3% SDS, and 0.1 mg/ml Proteinase K and incubated for 20 minutes at 42°C with shaking at 1250 rpm. Samples were extracted with phenol:chloroform:isoamylalcohol (25:24:1), ethanol precipitated, and the DNA pellet was resuspended in 1x Tris-EDTA. Samples were run on native 1.5% agarose TAE gels at 1.5 V/cm for 17 hours. Gels were stained with ethidium bromide for 30 minutes and destained with milliQ-purified water (Millipore Sigma) for 1 hour before imaging.

DNA unwinding of 616-bp template was performed similarly to DNA unwinding with the 3.8 kb template but with the following modifications. 616-bp pUC19-ARS1 closed-circular template was made following the protocol in Douglas et al. 2018 except the DNA was not
radiolabeled. At the end of the reaction, samples were run on native 3.5% 29:1 acrylamide:bis-acrylamide 1x TBE gels at 3 V/cm at 4°C for 20 hours. Gels were stained with SYBR Gold (Invitrogen) for 30 minutes before imaging.

**Assigning relative supercoiling states**

Experiment was performed similar to Douglas et. al (2018). Briefly, 6 fmol/ul 616-bp DNA was nicked with 0.25 U/ul Nb.BsrDI enzyme (NEB) for 1 hour at 65°C. DNA was then extracted with phenol:chloroform:isoamylalcohol (25:24:1), ethanol precipitated, and the DNA pellet was resuspended in water. 25 fmol of the nicked DNA was incubated in the indicated ethidium bromide concentrations for 1 hour at room temperature then ligated at 18°C overnight with 10 u/uL T4 DNA ligase (NEB). DNA was phenol:chloroform extracted, ethanol precipitated, and resuspended in 1× Tris-EDTA before analysis by electrophoresis, as described above. See Figure 13.
<table>
<thead>
<tr>
<th>Name</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASY1055.1</td>
<td>MATa ade2-1 ura3-11 his3-11,15 leu2-3,12 can-100 trp1-1 mcm2::hisG [pMCM5-MCM2::URA3]</td>
<td>Schwacha and Bell, 2001</td>
</tr>
<tr>
<td>ASY1059.1</td>
<td>MATa ade2-1 ura3-11 his3-11,15 leu2-3,12 can-100 trp1-1 mcm4::hisG [pMCM5-MCM4::URA3]</td>
<td>Schwacha and Bell, 2001</td>
</tr>
<tr>
<td>ASY2157</td>
<td>MATa ade2-1 ura3-11 his3-11,15 leu2-3,12 can-100 trp1-1 mcm6::HISMX6 [pMCM5-MCM6::URA3]</td>
<td>Schwacha and Bell, 2001</td>
</tr>
<tr>
<td>yST144</td>
<td>ade2-1 trp1-1 leu2-3,112 his3-11,15 ura3-1 can1-100 bar1::HisG lys2::HisG pep4Δ::unmarked his3::pSKM004 (GAL1,10-MCM2, Flag-MCM3) ura::pALS1 (GAL1,10-Cdt1, GAL4) lys::pSKM002 (GAL1,10-MCM4, MCM5) trp::pSKM003 (GAL1,10-MCM6, MCM7)</td>
<td>Ticau et al., 2015</td>
</tr>
<tr>
<td>yST161</td>
<td>ade2-1 trp1-1 leu2-3,112 his3-11,15 ura3-1 can1-100 bar1::HisG lys2::HisG pep4Δ::unmarked his3::pSKM004 (GAL1,10-MCM2, Flag-MCM3) ura::pALS1 (GAL1,10-Cdt1, GAL4) lys::pSKM002 (GAL1,10-MCM4, MCM5) trp::pST030 (GAL1,10-MCM6, UbSORT-MCM7)</td>
<td>Ticau et al., 2015</td>
</tr>
<tr>
<td>yKC03</td>
<td>ade2-1 trp1-1 leu2-3,112 his3-11,15 ura3-1 can1-100 bar1::HisG lys2::HisG pep4::KanMX Mmc2-13xmyc(NatMX4) his3::pKC16 (GAL1,10-mcm2Δ-2-177, Flag-MCM3) ura::pALS1 (GAL1,10-Cdt1, GAL4) lys::pSKM002 (GAL1,10-MCM4, MCM5) trp::pSKM003 (GAL1,10-MCM6, MCM7)</td>
<td>This study</td>
</tr>
<tr>
<td>yKC05</td>
<td>ade2-1 trp1-1 leu2-3,112 his3-11,15 ura3-1 can1-100 bar1::HisG lys2::HisG pep4Δ::KanMX6 Mmc7-V5 (NatMX4) Mmc4-V5 (HphMX4) his3::pSKM004 (GAL1,10-MCM2, Flag-MCM3) ura::pALS1 (GAL1,10-Cdt1, GAL4) lys::pKC17 (GAL1,10-mcm4Δ-2-181, MCM5) trp::pSKM003 (GAL1,10-MCM6, MCM7)</td>
<td>This study</td>
</tr>
<tr>
<td>yKC06</td>
<td>ade2-1 trp1-1 leu2-3,112 his3-11,15 ura3-1 can1-100 bar1::HisG lys2::HisG pep4Δ::unmarked Mmc7-V5 (NatMX4) Mmc4-V5 (HphMX4) his3::pSKM004 (GAL1,10-MCM2, Flag-MCM3) ura::pALS1 (GAL1,10-Cdt1, GAL4) lys::pKC18 (GAL1,10-mcm4Δ-2-174, MCM5) trp::pSKM003 (GAL1,10-MCM6, MCM7)</td>
<td>This study</td>
</tr>
<tr>
<td>yKC08</td>
<td>ade2-1 trp1-1 leu2-3,112 his3-11,15 ura3-1 can1-100 bar1::HisG lys2::HisG pep4::KanMX Mcm6-13xmyc(NatMX4) his3::pSKM004 (GAL1,10-MCM2, Flag-MCM3) ura::pALS1 (GAL1,10-Cdt1, GAL4) lys::pSKM002 (GAL1,10-MCM4, MCM5) trp::pKC19 (GAL1,10-mcm6-Δ2-105, MCM7)</td>
<td>This study</td>
</tr>
<tr>
<td>yKC09</td>
<td>ade2-1 trp1-1 leu2-3,112 his3-11,15 ura3-1 can1-100 bar1::HisG lys2::HisG pep4Δ::KanMX6 Mcm7-V5 (NatMX4) Mcm4-V5 (HphMX4) his3::pSKM004 (GAL1,10-MCM2, Flag-MCM3) ura::pALS1 (GAL1,10-Cdt1, GAL4) lys::pKC20 (GAL1,10-mcm4-7A, MCM5) trp::pSKM003 (GAL1,10-MCM6, MCM7)</td>
<td>This study</td>
</tr>
<tr>
<td>yKC10</td>
<td>ade2-1 trp1-1 leu2-3,112 his3-11,15 ura3-1 can1-100 bar1::HisG lys2::HisG pep4Δ::KanMX6 Mcm7-V5 (NatMX4) Mcm4-V5 (HphMX4) his3::pSKM004 (GAL1,10-MCM2, Flag-MCM3) ura::pALS1 (GAL1,10-Cdt1, GAL4) lys::pKC21 (GAL1,10-mcm4-Δ175-181, MCM5) trp::pSKM003 (GAL1,10-MCM6, MCM7)</td>
<td>This study</td>
</tr>
<tr>
<td>yKC11</td>
<td>ade2-1 trp1-1 leu2-3,112 his3-11,15 ura3-1 can1-100 bar1::HisG lys2::HisG pep4Δ::KanMX6 Mcm7-V5 (NatMX4) Mcm4-V5 (HphMX4) his3::pSKM004 (GAL1,10-MCM2, Flag-MCM3) ura::pALS1 (GAL1,10-Cdt1, GAL4) lys::pKC22 (GAL1,10-mcm4-175A, MCM5) trp::pSKM003 (GAL1,10-MCM6, MCM7)</td>
<td>This study</td>
</tr>
<tr>
<td>yKC12</td>
<td>ade2-1 trp1-1 leu2-3,112 his3-11,15 ura3-1 can1-100 bar1::HisG lys2::HisG pep4Δ::KanMX6 Mcm7-V5 (NatMX4) Mcm4-V5 (HphMX4) his3::pSKM004 (GAL1,10-MCM2, Flag-MCM3) ura::pALS1 (GAL1,10-Cdt1, GAL4) lys::pKC23 (GAL1,10-mcm4-178A, MCM5) trp::pSKM003 (GAL1,10-MCM6, MCM7)</td>
<td>This study</td>
</tr>
<tr>
<td>yKC13</td>
<td>ade2-1 trp1-1 leu2-3,112 his3-11,15 ura3-1 can1-100 bar1::HisG lys2::HisG pep4Δ::KanMX6 Mcm7-V5 (NatMX4) Mcm4-V5 (HphMX4) his3::pSKM004 (GAL1,10-MCM2, Flag-MCM3) ura::pALS1 (GAL1,10-Cdt1, GAL4) lys::pKC24 (GAL1,10-mcm4-181A, MCM5) trp::pSKM003 (GAL1,10-MCM6, MCM7)</td>
<td>This study</td>
</tr>
<tr>
<td>yKC14</td>
<td>ade2-1 trp1-1 leu2-3,112 his3-11,15 ura3-1 can1-100 bar1::HisG lys2::HisG pep4Δ::KanMX6 Mcm7-V5 (NatMX4) Mcm4-V5 (HphMX4) his3::pSKM004 (GAL1,10-MCM2, Flag-MCM3)</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ura::pALS1</strong> (GAL1,10-Cdt1, GAL4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>lys::pKC25</strong> (GAL1,10-mcm4-182A, MCM5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>trp::pSKM003</strong> (GAL1,10-MCM6, MCM7)</td>
<td><strong>This study</strong></td>
<td></td>
</tr>
<tr>
<td>yKC15</td>
<td>ade2-1 trp1-1 leu2-3,112 his3-11,15 ura3-1 can1-100</td>
<td></td>
</tr>
<tr>
<td></td>
<td>bar1::HisG lys2::HisG pep4::KanMX6 Mm7-V5 (NatMX4) Mcm4-V5 (HphMX4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>his3::pSKM004 (GAL1,10-MCM2, Flag-MCM3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ura::pALS1 (GAL1,10-Cdt1, GAL4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>lys::pKC26 (GAL1,10-mcm4-185A, MCM5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>trp::pSKM003 (GAL1,10-MCM6, MCM7)</td>
<td><strong>This study</strong></td>
</tr>
<tr>
<td>yKC16</td>
<td>ade2-1 trp1-1 leu2-3,112 his3-11,15 ura3-1 can1-100</td>
<td></td>
</tr>
<tr>
<td></td>
<td>bar1::HisG lys2::HisG pep4::KanMX6 Mm7-V5 (NatMX4) Mcm4-V5 (HphMX4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>his3::pSKM004 (GAL1,10-MCM2, Flag-MCM3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ura::pALS1 (GAL1,10-Cdt1, GAL4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>lys::pKC27 (GAL1,10-mcm4-188A, MCM5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>trp::pSKM003 (GAL1,10-MCM6, MCM7)</td>
<td><strong>This study</strong></td>
</tr>
<tr>
<td>yKC20</td>
<td>ade2-1 trp1-1 leu2-3,112 his3-11,15 ura3-1 can1-100</td>
<td></td>
</tr>
<tr>
<td></td>
<td>bar1::HisG lys2::HisG pep4::KanMX6 Mm7-V5 (NatMX4) Mcm4-V5 (HphMX4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>his3::pSKM004 (GAL1,10-MCM2, Flag-MCM3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ura::pALS1 (GAL1,10-Cdt1, GAL4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>lys::pKC27 (GAL1,10-mcm4-178A, MCM5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>trp::pST030 (GAL1,10-MCM6, UbSORT-MCM7)</td>
<td><strong>This study</strong></td>
</tr>
<tr>
<td>Name</td>
<td>description</td>
<td>Source</td>
</tr>
<tr>
<td>--------</td>
<td>-------------------------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>pSKM002</td>
<td>pRS307 (GAL1,10-MCM4, MCM5)</td>
<td>Kang et al., 2014</td>
</tr>
<tr>
<td>pSKM003</td>
<td>pRS404 (GAL1,10-MCM6, MCM7)</td>
<td>Kang et al., 2014</td>
</tr>
<tr>
<td>pSKM004</td>
<td>pRS403 (GAL1,10-MCM2, Flag-MCM3)</td>
<td>Kang et al., 2014</td>
</tr>
<tr>
<td>pALS1</td>
<td>pRS305 (GAL1,10-Cdt1, GAL4)</td>
<td>Ticau et al., 2015</td>
</tr>
<tr>
<td>pST030</td>
<td>pRS404 (GAL1,10-MCM6, UbSORT-MCM7)</td>
<td>Ticau et al., 2015</td>
</tr>
<tr>
<td>pKC01</td>
<td>pJR164 (TRP1-MCM5promoter-MCM2)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC02</td>
<td>pJR164 (TRP1-MCM5promoter-mcm2-Δ2-177)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC03</td>
<td>pJR164 (NatMX-MCM5promoter-MCM4)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC04</td>
<td>pJR164 (NatMX-MCM5promoter-mcm4-Δ2-181)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC04</td>
<td>pJR164 (NatMX-MCM5promoter-mcm4-Δ2-174)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC06</td>
<td>pJR165 (LEU2-MCM5promoter-MCM6)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC07</td>
<td>pJR165 (LEU2-MCM5promoter-mcm6-Δ2-105)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC08</td>
<td>pJR164 (NatMX-MCM5promoter-mcm4-7A)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC09</td>
<td>pJR164 (NatMX-MCM5promoter-mcm4-Δ175-181)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC10</td>
<td>pJR164 (NatMX-MCM5promoter-mcm4-175A)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC11</td>
<td>pJR164 (NatMX-MCM5promoter-mcm4-178A)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC12</td>
<td>pJR164 (NatMX-MCM5promoter-mcm4-181FA)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC13</td>
<td>pJR164 (NatMX-MCM5promoter-mcm4-182A)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC14</td>
<td>pJR164 (NatMX-MCM5promoter-mcm4-185A)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC15</td>
<td>pJR164 (NatMX-MCM5promoter-mcm4-188A)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC16</td>
<td>pRS403 (GAL1,10-mcm2-Δ2-177, Flag-MCM3)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC17</td>
<td>pRS307 (GAL1,10-mcm4-Δ2-181, MCM5)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC18</td>
<td>pRS307 (GAL1,10-mcm4-Δ2-174, MCM5)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC19</td>
<td>pRS404 (GAL1,10-mcm6-Δ2-105, MCM7)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC20</td>
<td>pRS307 (GAL1,10-mcm4-7A, MCM5)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC21</td>
<td>pRS307 (GAL1,10-mcm4-Δ175-181, MCM5)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC22</td>
<td>pRS307 (GAL1,10-mcm4-175A, MCM5)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC23</td>
<td>pRS307 (GAL1,10-mcm4-178A, MCM5)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC24</td>
<td>pRS307 (GAL1,10-mcm4-181FA, MCM5)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC25</td>
<td>pRS307 (GAL1,10-mcm4-182A, MCM5)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC26</td>
<td>pRS307 (GAL1,10-mcm4-185A, MCM5)</td>
<td>This study</td>
</tr>
<tr>
<td>pKC27</td>
<td>pRS307 (GAL1,10-mcm4-188A, MCM5)</td>
<td>This study</td>
</tr>
</tbody>
</table>
REFERENCES


Chapter III: The Role of ssDNA-Binding Proteins in Eukaryotic DNA Replication
ABSTRACT

DNA replication requires an ssDNA-binding protein (SSB) to bind and stabilize the ssDNA produced by the helicase for use as template for new DNA synthesis. Although replicative SSBS are essential proteins found in all domains of life, they vary significantly in their structure and subunit composition. Here we sought to understand the importance of these differences and investigate the role of the eukaryotic SSB RPA during replication initiation and elongation. To this end, we asked how substituting RPA with either the bacterial *E. coli* SSB (EcSSB) or T4 bacteriophage SSB (Gp32) impacts eukaryotic origin DNA unwinding and DNA synthesis. We found that RPA and EcSSB both supported origin DNA unwinding but Gp32 did not, suggesting that eukaryotic DNA unwinding requires at least one SSB function beyond ssDNA binding. Consistent with DNA unwinding being a critical precursor to DNA synthesis, only RPA and EcSSB support DNA synthesis, though to varying degrees. Interestingly, robust lagging-strand synthesis was only supported by RPA, indicating that RPA has at least one function required for efficient lagging-strand synthesis that cannot be performed by other SSBS.

INTRODUCTION

Eukaryotic DNA replication requires the progressive assembly of protein complexes at origins of replication. Origins are licensed during G1 phase of the cell cycle by loading of the ring-shaped, six-subunit, replicative helicase called Mcm2-7. Two Mcm2-7 hexamers are loaded onto each origin as a head-to-head double hexamer encircling double-stranded DNA (dsDNA, Evrin et al., 2009; Gambus et al., 2011; Remus & Difflley, 2009). At this stage, the helicases are inactive. Upon S-phase entry, loaded Mcm2-7 complexes
recruit a number of other replication proteins that activate the helicase for DNA unwinding. Once activated, the helicases act separately to bidirectionally unwind DNA. The activated helicases and the single-stranded DNA (ssDNA) they generate recruit the remainder of the DNA synthesis machinery to form individual replication forks.

Helicase activation is the committed step of replication initiation and results in recruitment of key proteins to the helicase and initiation of DNA unwinding. Activation is mediated by nine proteins which I will refer to as the “activation factors”. These proteins are sufficient for activation of loaded helicases in vitro (Yeeles et al., 2015). Eight of the activation factors coordinate the recruitment of Cdc45 and GINS to Mcm2-7 to form the CMG (Cdc45/Mcm2-7/GINS) complex (reviewed in Tanaka & Araki, 2013). Cdc45 and GINS directly activate both the ATPase and helicase activity of the Mcm2-7 core (Ilves et al., 2010), making the CMG complex the active form of the replicative DNA helicase. The final activation factor, Mcm10, acts after CMG formation to fully activate the helicase for DNA unwinding (van Deursen et al., 2012). In the absence of Mcm10, CMG formation results in a small amount of origin DNA being unwound (“origin DNA melting”) but more extensive DNA unwinding only occurs when Mcm10 is present (Douglas et al., 2018).

The ssDNA generated by CMG DNA unwinding must be protected from degradation/breaks and constrained from forming dsDNA before being used as a template for DNA synthesis. During eukaryotic replication, these functions are achieved by Replication Protein A (RPA). RPA is a ssDNA-binding protein (SSB) that is essential in eukaryotes and binds ssDNA in a non-sequence-specific manner (Brill & Stillman,
By binding and coating ssDNA, RPA inhibits nuclease from accessing the DNA. RPA binding also prevents the ssDNA from base pairing with the complementary strand to form dsDNA or from base pairing with itself to form hairpin structures, both of which would slow DNA synthesis (reviewed in Deng et al., 2015).

RPA is a heterotrimeric complex composed of Rfa1, 2, and 3 (Replication Factor A1, 2, and 3). The RPA complex is 116 kDa and contains four oligonucleotide-binding (OB) folds that function as DNA-binding domains (DBDs). The larger subunits, Rfa1 and Rfa2, contribute to ssDNA binding. The smaller Rfa3 subunit is essential for complex trimerization. The binding dynamics and affinity of RPA vary depending on DNA length due to the ability of RPA to adopt several different ssDNA-binding conformations (reviewed in Fanning et al., 2006). RPA can bind as little as 8 nucleotides but exhibits the highest affinity when binding 28 – 30 nucleotides (Ka ~ 10^9, Bastin-Shanower & Brill, 2001). Although cooperative ssDNA binding was been observed previously (Alani et al., 1992), more recent evidence suggests RPA cooperativity only occurs during the DNA-damage response (Yates et al., 2018).

DNA replication SSBs have been found in all domains of life as well as a subset of viruses, but these proteins display a wide variety of structures and ssDNA-binding affinities (reviewed in Marceau, 2012). The first SSB characterized was Gp32 (Gene 32 protein), which acts during T4 bacteriophage replication. The canonical bacterial SSB is simply called SSB and was initially identified in E. coli cells (EcSSB). Gp32 is a monomeric 34 kDa protein, whereas EcSSB is a homotetramer with a total molecular weight of 75 kDa.
Both Gp32 and EcSSB bind ssDNA with high cooperativity (Marceau, 2012). The DNA binding-site size of Gp32 is 8-10 nucleotides, whereas a single EcSSB tetramer can bind as many as 65 nucleotides. Gp32 binding to ssDNA has an association constant around 10^8 M (Kowalczykowski et al., 1981; Rouzina et al., 2005), whereas EcSSB has a stronger binding affinity in the 10^10 M range (Naufer et al., 2019). Despite their different ssDNA-binding properties, all three SSBs perform the same function of binding ssDNA during DNA replication (Marceau, 2012).

Much of what we know about the role of RPA in eukaryotic DNA replication comes from studies of simian virus 40 (SV40) DNA replication in human cells. At this viral replication fork, the SV40 large T-antigen (LTag) acts as the replicative helicase and RPA interacts with LTag and the DNA polymerases α and δ (Pol α and δ). These RPA interactions increase the processivity of the Pol α/primase holoenzyme (Braun et al., 1997; Kenny et al., 1989; Weisshart et al., 1998). In addition, RPA is proposed to facilitate Pol α/primase to Pol δ “polymerase switching” during lagging-strand synthesis through interactions with both polymerases and the sliding-clamp loader, RFC (Dornreiter et al., 1992; Waga & Stillman, 1998; Yuzhakov et al., 1999). Studies of SV40 DNA replication in vitro have shown that DNA unwinding and DNA synthesis do not occur in the absence of RPA (Ishimi et al., 1994; Tsurimoto & Stillman, 1991; Wobbe et al., 1987). Human RPA has a unique function during SV40 DNA replication, but not DNA unwinding, that cannot be performed by other SSBs, including yeast RPA (Brill & Stillman, 1989; Kenny et al., 1989). The specificity of this function is presumed to be mediated by specific interactions of human RPA with the SV40 replication machinery. For example, human but not yeast RPA can
bind LTag (Melendy & Stillman, 1993). However, the importance and role of replication-protein interactions with RPA have not been addressed at the eukaryotic DNA replication fork.

Here, we ask whether RPA is specifically required for eukaryotic DNA replication. Unlike in SV40 DNA replication, interactions between the eukaryotic CMG helicase and RPA have not been observed. If there are no direct interactions between the helicase and RPA, RPA may not be specifically required for eukaryotic DNA replication. That is, it is possible that other SSBs can substitute for RPA during eukaryotic replication. To test if RPA is specifically required, we substituted either Gp32 or EcSSB for RPA in in vitro eukaryotic DNA-unwinding and DNA-synthesis assays. Interestingly, the alternative SSBs showed different phenotypes in these assays. Gp32 could not support either DNA unwinding or DNA synthesis. In contrast, EcSSB allowed for DNA unwinding, indicating that this SSB has a property required for DNA unwinding that Gp32 lacks. Substitution of EcSSB for RPA in DNA replication assays resulted in clear defects in lagging-strand synthesis, suggesting RPA facilitates replication elongation using specific contacts with the lagging-strand machinery.

**RESULTS**

**RPA, Gp32, and EcSSB ssDNA binding**

To study the role of SSBs during eukaryotic DNA replication, we used purified yeast RPA, \textit{E. coli} SSB (EcSSB), and T4 bacteriophage Gp32. We verified that all three purified proteins were competent for ssDNA-binding activity using electrophoretic mobility-shift
assays (EMSAs). We tested ssDNA binding using 3 picomoles (pmol) of radiolabeled dT$_{30}$ ssDNA and a 3-fold titration series of each protein ranging between 0.11 to 27 pmol. All three SSBs fully shifted the labeled DNA, demonstrating that they are each functional for ssDNA binding (Figure 1).

Consistent with previous independent measurements, examination of the EMSA studies showed that RPA, EcSSB, and Gp32 have different ssDNA-binding affinities. For all three proteins, the shift from primarily unbound to primarily bound ssDNA occurred between 1 and 9 picomoles (Figure 1). Gp32 required the highest concentration before detectable ssDNA binding was detected (1 pmol GP32, Figure 1B), whereas EcSSB required the lowest concentrations to detect bound ssDNA (0.1 pmol EcSSB, Figure 1C). ssDNA binding by RPA occurred at an intermediate concentration (0.3 pmol RPA, Figure 1A). The concentration of protein at which the ssDNA is fully bound can be determined by looking for the loss of “unbound” DNA signal. Similar to the first-detectable ssDNA binding, full ssDNA binding required the lowest concentration of EcSSB (1 pmol, Figure 1C), an intermediate concentration of RPA (3 pmol, Figure 1A), and the highest concentration of Gp32 (27 pmol, Figure 1B). The differences in binding-site size, cooperativity, and affinity of each protein for ssDNA all likely contribute to the differences we see in binding dynamics in this assay. Importantly, despite these differences, all three proteins bind ssDNA with high affinity.

**RPA is necessary for DNA unwinding and DNA synthesis**

RPA has previously been shown to be required for *in vitro* DNA unwinding and DNA synthesis in budding yeast (Douglas et al., 2018; Yeeles et al., 2015). To confirm this
Figure 1: (A) RPA, (B) Gp32, and (C) EcSSB bind ssDNA binding but with different affinities. Each SSB was tested in electrophoretic mobility-shift assays using 3 pmol of radiolabeled dT_{30} ssDNA. Each protein was titrated 3-fold from 0.1 pmol to 27 pmol. Unbound and SSB bound states are indicated for each SSB. In addition, a poorly populated RPA-ssDNA intermediate is indicated. RPA shows two shifted RPA-ssDNA complexes suggesting two types of DNA binding ("Bound" and "Intermediate"). The intermediate state is weakly populated and only apparent at lower RPA concentrations (0.3 – 3 pmol RPA).
requirement is observed using our purified proteins, we performed DNA-unwinding and DNA-replication assays in the presence of different concentrations of RPA (Figure 2). DNA unwinding was detected using a topologically-constrained circular DNA template and by monitoring the formation of supercoiled DNA as a consequence of DNA unwinding in the presence of a topoisomerase (Douglas et al., 2018). Consistent with these topological changes being a product of CMG helicase activity, elimination of DDK, an essential kinase required for CMG assembly, prevented formation of the supercoiled DNA (Figure 2A, lane 3). DNA replication initiation and elongation was monitored through incorporation of radiolabeled deoxynucleotide triphosphates (dNTPs) to detect newly synthesized DNA after separation on a denaturing agarose gel (Yeeles et al., 2015). As with DNA unwinding, nucleotide incorporation in this assay is DDK-dependent.

Titration of RPA in the DNA unwinding assay showed a clear correlation between RPA levels and the extent of DNA unwinding. Maximum supercoiled DNA product was observed with 70 nM RPA without detectable increases at 140 nM (Figure 2A). These data support the fact that efficient DNA unwinding requires RPA. Consistent with this conclusion, omission of RPA led to very little DNA unwinding (see “-SSB” lanes in Figures 3A and 4A).

Efficient DNA synthesis is also dependent on RPA. DNA synthesis required a specific concentration range of RPA between 75-150 nM (Figure 2B). Titrating RPA levels above or below these levels led to strong decreases in DNA synthesis activity. Omission of RPA resulted in very low background levels of DNA synthesis (see “-SSB” in Figures 3B and
Figure 2: Eukaryotic DNA unwinding and DNA synthesis are dependent on RPA. (A) DNA unwinding was detected by formation of supercoiled DNA and is dependent on DDK phosphorylation DNA. Maximum DNA unwinding occurred at 70 nM RPA. Mcm2-7 was loaded onto relaxed plasmid DNA then activated to stimulate DNA unwinding. The reaction was quenched, separated on an agarose gel, and DNA was visualized by ethidium bromide staining. DNA -Topo reaction is supercoiled DNA before relaxation step. DNA +Topo lane is plasmid DNA relaxed with Topoisomerase I. (B) DNA synthesis products were separated on a denaturing gel and detected by phosphor-imaging. Maximum DNA synthesis occurred between 75 – 150 nM RPA. Helicases were loaded onto plasmid DNA and then activated. Replication proteins, dNTPs, and [α-32P]-dCTP were added to initiate DNA replication. DNA replication was dependent on DDK phosphorylation.
4B), similar to that observed in the absence of DDK. This finding indicates that the limited unwinding observed in the minus RPA reactions is not able to support replication initiation and elongation.

**Gp32 does not support DNA unwinding or DNA synthesis**

After confirming that RPA is required for DNA unwinding and DNA synthesis, we asked if other SSBs could substitute for RPA in these processes. If the only function of RPA during a particular assay is to bind ssDNA then other SSBs should support eukaryotic DNA replication. If additional RPA activities are required, then other SSBs should not be able to substitute for RPA. To this end, we replaced RPA with either Gp32 or EcSSB in our reconstituted DNA-unwinding and DNA-synthesis assays.

We first tested if the T4 bacteriophage SSB Gp32 could substitute for RPA in eukaryotic DNA replication (Figure 3). When Gp32 was used in place of RPA in the DNA-unwinding assay, we never observed levels of DNA unwinding above that detected in the absence of any SSB (Figure 3A). Increasing the concentration of Gp32 32-fold higher (2240 nM) than the RPA concentration that showed robust DNA unwinding (70 nm) did not restore DNA unwinding activity, eliminating the possibility that a difference in ssDNA-binding affinity is responsible for the lack of DNA unwinding. Indeed, very high concentrations of Gp32 reduced the levels of supercoiled DNA detected to below that detected with no SSB. Consistent with this defect in DNA unwinding, substituting Gp32 for RPA inhibited DNA synthesis activity at all concentrations (Figure 3B). Together, these results indicate that Gp32 is unable to perform at least one RPA function that is essential for eukaryotic
Figure 3: Gp32 cannot replace RPA during (A) DNA unwinding or (B) DNA synthesis. The DNA unwinding and DNA synthesis signals when Gp32 was present are equivalent to or less than the signals observed when no SSB was added. (A) Gp32 was tested in the DNA-unwinding assay. Assays were performed as in Figure 2A except that Gp32 was substituted for RPA as indicated. (B) Gp32 was tested in DNA-synthesis assays. Assays were performed as in Figure 2B except that Gp32 was substituted for RPA as indicated. In –SSB reactions, neither Gp32 nor RPA were included.
DNA unwinding and, as a result, also is unable to support DNA synthesis.

EcSSB supports DNA unwinding and altered DNA synthesis

We next assessed the ability of the bacterial EcSSB to substitute for RPA. Unlike Gp32, EcSSB facilitated DNA unwinding (Figure 4A). At concentrations of EcSSB equal to or greater than the RPA concentration that showed robust DNA unwinding (70 – 280 nM), we observed DNA unwinding levels that are clearly above the SSB control. We note, however, DNA unwinding with EcSSB was reduced compared to the levels observed with RPA. The amount of DNA unwinding we observed did not vary across these EcSSB concentrations, suggesting that we have saturated EcSSB in the reaction. Decreasing EcSSB to 35 nM led to lower DNA-unwinding signal, indicating that we have reached an optimal level of EcSSB in the assay (Figure 5). Thus, the lower levels of DNA unwinding we observe relative to RPA are not due to excessive levels of EcSSB.

Because EcSSB supports DNA unwinding, we asked if this SSB would also support DNA synthesis. Using the reconstituted DNA-synthesis assay, we compared DNA replication with EcSSB to replication in the presence of RPA. In a DDK-dependent manner, RPA supported synthesis of both leading (~6000 bp) and lagging (~500 bp) strand DNA (Figure 4B). In contrast, we observed a distinctly different pattern when EcSSB was substituted for RPA in the assay. When present at the equivalent concentrations that result in maximum DNA synthesis with RPA (75 nM), EcSSB supports weak DNA-synthesis activity (Figure 4B). Further increasing EcSSB concentration 4-fold resulted in levels of long DNA products similar to that observed with RPA. In contrast, increasing EcSSB
Figure 4: EcSSB supports reduced (A) DNA unwinding and (B) DNA synthesis compared to RPA. (A) DNA unwinding with EcSSB was above background (−DDK) but was reduced compared to RPA. Increasing EcSSB concentration did not result in more DNA unwinding. DNA-unwinding assays were performed as in Figure 2A except that EcSSB was substituted for RPA as indicated. (B) DNA synthesis with EcSSB produced more long DNA products than short DNA products. The signal for short DNA products did not increase with increasing EcSSB concentration. DNA-synthesis assays were performed as in Figure 2B except that EcSSB was substituted for RPA as indicated. In −SSB reactions, neither EcSSB nor RPA were included.
Figure 5: Further titration of EcSSB in the DNA-unwinding assay. Decreasing EcSSB concentration results in a decrease in DNA unwinding. DNA-unwinding assays were performed as in Figure 2A except EcSSB was substituted for RPA.
concentration resulted in either no change or a slight reduction in the amount of short DNA products. Overall, DNA replication with EcSSB resulted in primarily long DNA products (~6000 bp) but limited short products (~500 bp), consistent with EcSSB supporting DNA unwinding but having a defect in lagging-strand synthesis.

**DISCUSSION**

Our findings indicate that different SSB characteristics are essential for eukaryotic DNA-unwinding versus DNA-synthesis activity. The alternative SSBs we tested exhibited different phenotypes for each eukaryotic DNA replication event. Gp32 is unable to support DNA unwinding and DNA synthesis in the eukaryotic system (Figure 3). In contrast, EcSSB is able to support DNA unwinding, although at a reduced level compared to RPA (Figure 4A). EcSSB also supports DNA-synthesis activity, but the distribution of the resulting DNA products was distinctly different (Figure 4B). This study indicates that RPA has one or more specific functions during eukaryotic DNA replication that cannot be performed by other SSBs.

**ssDNA binding is important for eukaryotic DNA unwinding**

The ssDNA-binding affinities we observed in the EMSA experiments are consistent with previous data on the ssDNA-binding kinetics of these proteins. Gp32 requires the highest concentration for ssDNA binding (Figure 1B), and it has the lowest association constant (Ka ~ 10^8 M; Kowalczykowski et al., 1981; Rouzina et al., 2005). EcSSB has the highest association constant (Ka ~ 10^10 M; Naufer et al., 2019) and binds ssDNA at the lowest concentration in our assay (Figure 1C). RPA binding to ssDNA occurs at a middle
concentration compared to Gp32 and EcSSB (Figure 1A), which is consistent with it having an association constant that falls between the other SSBs (Ka ~ 10⁹ M; Bastin-Shanower & Brill, 2001).

DNA unwinding in the absence of any SSB showed limited but detectable signal (lane 3 Figures 3A and 4A). This signal does not result from contaminating DNA unwinding activities because leaving out DDK, which is essential for CMG-helicase activity, resulted in no DNA-unwinding signal (lane 3 in Figure 2, lane 5 in Figures 3 and 4). This result is consistent with a recent study that showed faint signal when RPA was omitted from the same assay (Figure 1D in Douglas et al., 2018). Importantly, the type or amount of DNA unwinding that occurs without any SSB is not sufficient to initiate DNA replication. DNA synthesis without SSB resulted in the same background signal seen in the absence of DDK (Figure 3B and 4B).

An alternative hypothesis to explain the low levels of DNA unwinding we see without SSB is that our assays could have contaminating RPA. Such contaminating RPA would have to be present in one of the other helicase-activation protein preparations. To address this, we tested all of the proteins used in the DNA-unwinding assay by immunoblotting for RPA. One protein, DDK, showed low levels of RPA contamination (Figure 6A). However, the amount of RPA contamination was below the amount of RPA needed for DNA unwinding (70 nM RPA Figure 2A). We re-purified DDK using an alternative method and saw a significant reduction in RPA signal (see Methods; Figure 6B). This new DDK was used in all experiments presented here.
Figure 6: Testing for and removing contaminating RPA. (A) All proteins used in the DNA-unwinding assay were tested for contaminating RPA by immunoblotting with Rfa1 antibody. DDK was the only protein that showed detectable levels of RPA. Note: amount of each protein loaded in A is equal to or greater than the amounts used in DNA-unwinding assays, with the exception of RPA. The RPA control shows the amount of RPA that would result in a final concentration of 35 nM in our reactions, which only weakly supports DNA unwinding (Figure 2A). (B) Comparison of DDK-containing fractions after the FLAG (lane 1) and Heparin columns (lanes 2 – 8, see Methods). The FLAG elution shows readily detectable RPA (this is the same stage of purification shown in 6A). After running the FLAG-purified DDK over a Heparin column, the RPA signal is dramatically reduced. Note that the amount of DDK loaded in each lane of B is ~1/3 the amount used in the DNA-unwinding assay. The fraction that eluted at 0.61 M potassium acetate (KOAc) was used in all assays presented in this study.
How do SSBs facilitate eukaryotic DNA unwinding?

All three SSBs can bind ssDNA but only two support DNA unwinding, indicating that ssDNA binding is not the only essential function of SSBs during DNA unwinding. Gp32 is unable to support DNA unwinding above the level seen in the absence of any SSB (Figure 3A). In contrast, RPA and EcSSB both support DNA unwinding. This is consistent with the SV40 DNA replication system where DNA unwinding by LTag is supported by RPA and EcSSB but not Gp32 (Kenny et al., 1989). This indicates that DNA unwinding requires some function that is shared between RPA and EcSSB but is not supported by Gp32.

Why are SSBs required for DNA unwinding? Structural studies of the CMG complex found that the excluded strand (i.e. the strand that does not pass through the central channel of the CMG) interacts with the helicase central pore (Eickhoff et al., 2019; R. Georgescu et al., 2017; Goswami et al., 2018). Instead of being an interaction that facilitates DNA unwinding, recent evidence suggests that it leads to CMG stalling (Figure 6; Kose, Xie, et al., 2019). In addition, RPA binding to the excluded strand stimulates CMG helicase activity (Kose, Xie, et al., 2019). These findings have led to the hypothesis that SSB binding to the excluded ssDNA stimulates helicase activity by preventing this interaction. This hypothesis is supported by studies of the *E. coli* replicative helicase, DnaB, that showed applying tension to the excluded strand, but not the encircled strand, stimulates helicase activity (Ribeck et al., 2010). In contrast, a similar study using the T4 bacteriophage replicative helicase, Gp41, demonstrated that tension on the excluded strand inhibited DNA unwinding (Ribeck & Saleh, 2013).
Figure 7: RPA stimulates CMG-helicase activity by preventing helicase engagement with the excluded strand of ssDNA. When the excluded strand is not bound by RPA, DNA unwinding stalls. (Modified from Kose, Xie, et al., 2019).
There are several differences between Gp32 and the other SSBs tested that could impact DNA unwinding. It is possible that Gp32 is unable to support DNA unwinding because its binding to ssDNA is not sufficiently tight to prevent the excluded ssDNA from interfering with helicase activity. Although Gp32 is able to bind to ssDNA, it has the lowest affinity for ssDNA of the three SSBs tested. This lower affinity could lead to the CMG complex outcompeting Gp32 for interactions with the excluded strand. Alternatively, DNA interaction with these multiple sites leads to extensive bending and wrapping of the bound ssDNA (Figure 10 in Chapter I; Raghunathan et al., 2000; Yates et al., 2018). Although ssDNA has not been resolved in complex with Gp32, the small DNA-footprint size of Gp32 (as little as 7 nucleotides; Jensen et al., 1976) suggests that the DNA is not extensively bent or wrapped around Gp32 (Treuner et al., 1996). Thus, it is possible that SSBs that do not bend ssDNA may be unable to support CMG helicase activity. Finally, it is possible that the different sizes of the SSBs lead to their different activities. RPA and EcSSB are both large, multimeric proteins (116 and 75 kDa respectively). On the other hand, Gp32 is a much smaller (34 kDa), monomeric protein. Again, it is possible that the larger size of RPA and EcSSB prevents the helicase from accessing the excluded ssDNA in a manner that Gp32 cannot.

It is intriguing that eukaryotic CMG and SV40 LTag helicase activity both rely on a function of RPA (and EcSSB) that is not present in Gp32. The eukaryotic helicase and LTag are both replicative DNA helicases with similar hexameric, ring-shaped structures (Li et al., 2003; Noguchi et al., 2017). SV40 DNA replication uses only eukaryotic proteins except for LTag. However, LTag helicase loading and activation differ significantly from the
equivalent processes in the eukaryotic system. LTag does not require any other proteins to load it and only RPA to activate it on SV40 origin DNA. In contrast, the eukaryotic helicase requires a total of thirteen accessory proteins, including RPA, for these activities. The eukaryotic helicase unwinds DNA by translocating on ssDNA and excluding the opposite strand (Georgescu et al., 2017; Kose, Larsen, et al., 2019). The details of how LTag unwinds DNA are controversial, but recent data demonstrated that LTag is capable of unwinding DNA in a mode similar to the CMG helicase (Yardimci et al., 2012). If the CMG helicase and LTag both require RPA because of its ability to prevent the excluded strand from interfering with the helicase, it would support the model that CMG helicase and LTag unwind DNA by similar strand exclusion mechanisms.

**Eukaryotic DNA replication requires RPA for lagging-strand synthesis**

DNA replication requires SSB because of its essential function in DNA unwinding. However, an open question is whether SSBs have another specific function during DNA replication outside of their role in DNA unwinding. In the simian virus SV40, DNA unwinding is supported by various SSBs, including yeast RPA and EcSSB, but DNA replication requires human RPA specifically (Brill & Stillman, 1989; Kenny et al., 1989; Melendy & Stillman, 1993). In contrast, we saw DNA replication in the yeast system with both yeast RPA and EcSSB. Our results suggest that there is less specificity in the role of RPA in eukaryotic DNA replication compared to SV40 replication.

However, comparing DNA synthesis with EcSSB versus RPA indicates that DNA replication does require RPA for an additional reason beyond its ability to promote DNA
unwinding. Overall DNA synthesis with EcSSB is reduced when included at the same concentration that shows robust DNA synthesis with RPA (75 nM, Figure 4B). This is consistent with the DNA unwinding defect seen at the same concentration of EcSSB (Figure 4A). Increasing EcSSB concentration in the DNA-synthesis assay increases the signal for long DNA products but not for the short products made during lagging-strand synthesis (150 – 300 nM EcSSB, Figure 4B). The short DNA products show no change or a slight reduction in signal at increasing concentrations of EcSSB. The lack of short DNA products could be due to two possibilities: a reduced frequency of lagging-strand priming events or a defect in overall lagging-strand synthesis. These results suggests that eukaryotic DNA replication requires RPA specifically for lagging-strand synthesis.

**RPA specificity during lagging-strand synthesis: Priming inhibition**

Inhibition of priming would cause long lagging-strand products. Priming activity must not be completely eliminated by EcSSB because some DNA synthesis still occurs. However, priming inhibition would have a disproportionate effect on the lagging strand because priming occurs on the this strand every 150 – 250 base pairs (Georgescu et al., 2015; Okazaki et al., 1968). Priming inhibition would increase the time between priming events and thus would increase the length of lagging-strand products. In contrast, the leading strand is synthesized from a single priming event, so the length of the leading strand product would be unaffected by reduced priming activity.

One potential reason priming could be inhibited is if Pol α/primase is not recruited to the ssDNA template. Previous biochemical studies showed that Pol α/primase is defective at
binding to DNA in the absence of RPA (Yuzhakov et al., 1999). However, it is not known if other SSBs could restore Pol α/primase recruitment. Both human RPA and EcSSB are capable of binding human Pol α/primase, although EcSSB binding is weaker (Dornreiter et al., 1992). If SSB binding to Pol α/primase is all that is required for its recruitment, then both RPA and EcSSB should support Pol α/primase recruitment. Conversely, Gp32 is not able to bind this polymerase. It is possible that although EcSSB can bind to Pol α/primase, these interactions are either too weak or are otherwise not competent to promote Pol α/primase activity. If EcSSB cannot support Pol α/primase activity, it would suggest that Pol α/primase recruitment requires an SSB that forms a specific interaction with the polymerase complex.

Alternatively, priming would be reduced if Pol α/primase is successfully recruited but its synthesis activity is inhibited. RPA stimulates Pol α/primase synthesis activity in human cells, and this effect appears to be species specific (Erdile et al., 1991; Kenny et al., 1989; Tsurimoto & Stillman, 1991). Only human RPA, but not yeast RPA, EcSSB, or Gp32, support DNA synthesis by human Pol α/primase (Erdile et al., 1991; Kenny et al., 1989). It may also be the case in our yeast system that only yeast RPA can support yeast Pol α/primase. Future studies could address whether EcSSB can support yeast Pol α/primase recruitment and synthesis activity.

**RPA specificity during lagging-strand synthesis: DNA synthesis inhibition**

If lagging-strand priming does not specifically require RPA, it suggests instead that lagging-strand synthesis requires RPA. Pol δ is responsible for lagging-strand synthesis.
Lagging-strand DNA synthesis is greatly reduced without Pol δ, indicating that neither the leading-strand DNA polymerase (Pol ε) nor Pol α/primase can effectively synthesize this strand in the absence of Pol δ (Yeeles et al., 2017). Like Pol α/primase, Pol δ function could be inhibited either due to reduced recruitment to template DNA or DNA synthesis activity. In this case, we do not think that Pol δ synthesis activity is inhibited because previous studies showed that human RPA and EcSSB supported comparable levels of DNA synthesis by human Pol δ (Yuzhakov et al., 1999).

Pol δ recruitment and binding to DNA could be defective without RPA because of the need to displace Pol α before Pol δ can take over DNA synthesis. This event is referred to as “polymerase switching”. Yuzhakov et al. (1999) showed that Pol δ can load onto primed DNA with either RPA or EcSSB present. However, this experiment did not include Pol α, so it was only tested for Pol δ activity but not for “polymerase switching” (Pol δ loading and Pol α displacement). It has been suggested that RPA plays a role during polymerase switching because of its interactions with Pol α/primase, Pol δ, and the sliding clamp loader RFC (Mossi et al., 2000; Yuzhakov et al., 1999). Yuzhakov et al. (1999) demonstrated that RPA can bind Pol δ, Pol α/primase, and RFC. They argue that interactions with RPA lead to RFC displacing Pol α/primase and the subsequent recruitment of Pol δ. Consistent with this model, other studies have shown that RFC plays a role in Pol α/primase displacement (Mossi et al., 2000). Yet, Pol α and RFC do not interact (Yuzhakov et al., 1999), so it is feasible that another protein (such as RPA) facilitates their exchange. Unlike RPA, EcSSB cannot bind to RFC, suggesting that
EcSSB could not facilitate polymerase switching by this mechanism (Yuzhakov et al., 1999). It is possible that RPA is required for Pol δ recruitment because of its ability to facilitate Pol α displacement. In our study, RFC recruitment must be partially active in the absence of RPA because we do see leading-strand synthesis. Omission of RFC from DNA replication strongly inhibits both leading- and lagging-strand synthesis (Yeeles et al., 2017). Similar to priming activity, partial RFC inhibition would have a greater effect on the lagging strand than the leading strand due to the discontinuous nature of lagging-strand synthesis.

Our study suggests that lagging-strand DNA replication is inhibited without RPA. This inhibition could target either lagging-strand synthesis (by Pol δ) or lagging-strand priming (by Pol α/primase). Future experiments could address the mechanism of inhibition by examining lagging-strand DNA replication specifically. Inhibition of lagging-strand synthesis, through Pol δ, would result in few lagging-strand products. Inhibition of lagging-strand priming, through Pol α/primase, would result in long lagging-strand products. Further studies are needed to determine why eukaryotic lagging-strand synthesis has a specific requirement for RPA.
METHODS

Protein purifications

Purified *Escherichia coli* SSB and T4 bacteriophage Gp32 were purchased from Sigma-Aldrich. Mcm2–7/Cdt1 and ORC complexes were purified as described in Kang et al. (2014). Cdc6 was purified as described in Frigola et al. (2013). RPA, S-CDK, Sld3/7, Cdc45, Sld2, Dpb11, GINS, Mcm10, Polymerase epsilon, Polymerase alpha/primase, Polymerase delta, Ctf4, RFC, PCNA, Mrc1, Csm3-Tof1, and Topo II were purified as described in Lõoke et al. (2017).

DDK was purified as described in Lõoke et al. (2017) but with the following modifications. After elution for the FLAG column, DDK containing fractions were applied to a HiTrap Heparin column in 300mM KOAc and buffer A (50 mM HEPES-KOH at pH 7.6, 1 mM EDTA, 5 mM MgOAc, 10% glycerol, and 0.01% NP-40). The column was then washed in 10-CV of 650 mM KOAc and buffer A. DDK was eluted with a 20-CV gradient of 0.3–1 M KOAc in buffer A. Peak fractions of DDK eluted around 650 mM KOAc and buffer A. These fractions were tested for RPA contamination by immunoblotting with anti-Rfa1.

Electrophoretic mobility-shift assays (EMSA)

dT$_{30}$ ssDNA was radiolabeled with [γ-P$_{32}$]ATP using T4 PNK (NEB) following the manufacturer’s protocol.

Electrophoretic mobility-shift assays (EMSAs, also known as gel mobility shift assays) were performed as described in Binz et al., (2006) with 3 pmol labeled dT30 and the indicated SSB concentrations. Briefly, ssDNA and the indicated SSB were incubated for 20 min at 25 °C in filter binding buffer (30 mM HEPES-KOH (pH 7.6), 100 mM sodium chloride, 5 mM magnesium chloride, and 1 mM DTT) and 750 ng BSA. Reaction mixtures were separated on a 6% native polyacrylamide gel in 1x Tris borate-EDTA running buffer at 45 V for 2 hours. Gels were dried onto Amersham Hybond-XL (GE Healthcare) and imaged using a phosphor screen. Gels were scanned using a Typhoon phosphorimager (GE Healthcare).
**DNA-unwinding assay**

25 fmol soluble 3.8 kb pUC19-ARS1 plasmid template was relaxed with 0.4 pmol Topo I for 30 min at 30 °C. Each incubation step was performed in a thermomixer (Eppendorf) with shaking at 1250 rpm at 25 °C unless otherwise indicated. Mcm2-7 loading was performed by incubating 0.48 pmol of ORC, 0.52 pmol of Cdc6, and 1.14 pmol of Mcm2–7/Cdt1 with 0.125 pmol template DNA in 25 mM HEPES-KOH (pH 7.6), 10 mM magnesium acetate, 225 mM potassium glutamate, 2 mM DTT, 0.02% NP-40, 5% glycerol, 5 mM ATP, 20 mM phosphocreatine, and 0.2 μg of creatine kinase for a total volume of 10 μL. Reactions were incubated for 25 min, at which point 1.3 pmol of DDK was added and incubation was continued for a further 30 min. DNA unwinding was then initiated by adding 20 μL of 0.6 pmol CDK, 1 pmol Sld3/7, 1 pmol Cdc45, 1.24 pmol Sld2, 0.8 pmol Dpb11, 5 pmol GINS, 0.06 pmol Mcm10, 0.6 pmol Pol ε, 0.4 pmol of Topo I, and the indicated SSB concentration in 25 mM HEPES-KOH (pH 7.6), 10 mM magnesium acetate, 250 mM potassium glutamate, 1 mM DTT, 0.02% NP-40, 8% glycerol, 5 mM ATP, and 0.4 mg/ml BSA directly to the reaction and incubating for 40 min. The reaction was then quenched with 13 mM EDTA, 0.3% SDS, and 0.1 mg/ml Proteinase K and incubated for 20 min at 42 °C with shaking at 1250 rpm. Samples were extracted with phenol:chloroform:isoamylalcohol (25:24:1), ethanol precipitated, and the DNA pellet was resuspended in 1x Tris-EDTA. Samples were run on native 1.5% agarose TAE gels at 1.5 V/cm for 17 hr. Gels were stained with ethidium bromide for 30 min and destained with milliQ-purified water (Millipore Sigma) for 1 hr before imaging.

**DNA-synthesis assay**

Each incubation step was performed in a thermomixer (Eppendorf) with shaking at 1250 rpm at 25 °C. Mcm2-7 loading and DDK phosphorylation was performed in the same conditions as the DNA-unwinding assay except with a soluble, supercoiled 11.9 kb pUC19-ARS1 plasmid template. After DDK phosphorylation, replication was initiated by adding 20 μL of 1 pmol CDK, 1 pmol Sld3/7, 2.6 pmol Cdc45, 1.24 pmol Sld2, 0.8 pmol Dpb11, 5 pmol GINS, 0.02 pmol Mcm10, 0.6 pmol Pol ε, 2 pmol Pol α, 0.5 pmol Topo II, 0.6 pmol Ctf4, 0.5 pmol RFC, 0.4 pmol PCNA, 0.5 pmol Mrc1, 0.6 pmol Csm3/Tof1, 0.6
pmol Pol δ, and the indicated SSB concentration in 12.5 mM HEPES-KOH (pH 7.6), 5 mM magnesium acetate, 125 mM potassium glutamate, 1 mM DTT, 0.01% NP-40, 4% glycerol, 1.5 mM ATP, 10 mM phosphocreatine, 3 μg of creatine kinase, 0.2 mg/ml BSA, 100 μM rNTP, 10 μM dNTP, and 10 μCi [α-P\(^{32}\)]dCTP directly to the reaction. Following 60 min of incubation, reactions were quenched with 30 μL of 50 mM EDTA. Unincorporated nucleotides were removed with Illustra MicroSpin G-50 columns (GE Healthcare), and samples were separated on a 0.6% alkaline agarose gel in 30 mM sodium hydroxide, 2 mM EDTA. Gels were dried onto Amersham Hybond-XL (GE Healthcare) and imaged using a phosphor screen. Gels were scanned using a Typhoon phosphorimager (GE Healthcare).
REFERENCES


Noguchi, Y., Yuan, Z., Bai, L., Schneider, S., Zhao, G., Stillman, B., Speck, C., & Li, H. (2017). Cryo-EM structure of Mcm2-7 double hexamer on DNA suggests a lagging-


Chapter IV: Expanded Discussion and Future Directions
Key conclusions

The work described in this thesis focuses on mechanisms that are essential for eukaryotic DNA unwinding and DNA synthesis. In Chapter II, I showed initial origin DNA melting requires only a single-hexamer helicase. In contrast, I found that further DNA unwinding requires robust double-hexamer helicase interactions. These findings support the notion that each hexamer’s individual interactions with the DNA drive initial origin melting but interactions between the hexamers drive the more extensive DNA unwinding required for the initiation of DNA replication. In Chapter III, I showed that DNA unwinding and DNA synthesis have specific but different requirements for ssDNA-binding proteins (SSBs). The ability to bind ssDNA is necessary but not sufficient for origin DNA unwinding. This event requires at least one additional SSB function that RPA and *E. coli* SSB possess but T4 bacteriophage Gp32 does not. DNA synthesis requires DNA unwinding, but DNA unwinding alone is only sufficient to support leading-strand synthesis. Proper lagging-strand synthesis requires another function that only RPA can perform. The exact natures of these functions are currently unclear, and further research will be needed to identify these functions and to understand why they are necessary to drive DNA unwinding and lagging-strand DNA synthesis. In this chapter, I will synthesize what these two studies tell us about the requirements of DNA unwinding and DNA synthesis. In addition, I will explore future directions that could be taken to elucidate why these requirements are necessary.

Requirements for DNA unwinding: The Mcm2-7 helicase complex

In Chapter II, I showed that initial DNA unwinding does not require the double-hexamer conformation of the helicase. DNA unwinding during DNA replication can be separated
into two steps: 1) initial DNA melting that must occur inside the Mcm2-7 central channel to allow the transition from encircling dsDNA to ssDNA, and 2) further DNA unwinding that involves the helicase translocating on one strand and occluding the opposite strand. Initial DNA unwinding (“origin DNA melting”) occurs within the confines of the helicase central channel. Before this study, it was unclear whether the double-hexamer complex was necessary for origin DNA melting or if single CMGs could drive this activity. Using a mutant Mcm2-7 that is loaded onto DNA as two single hexamers instead of a double-hexamer complex, I demonstrated that single Mcm2-7 hexamers can be converted into CMGs and initiate origin DNA melting (Figure 12B and E, Chapter II). This suggests that initial unwinding occurs by manipulation of the DNA within a single CMG complex rather than through the coordinated action of two opposing helicases. However, the amount of DNA unwinding we see with single hexamers (~ 0.7 DNA turns) is not enough to support extrusion of the occluded strand to allow the helicase to encircle ssDNA (2 – 3 DNA turns; Noguchi et al., 2017).

In contrast to what we saw for initial DNA unwinding, our findings indicate that more extensive DNA unwinding requires the helicases to be able to form robust double-hexamer interactions (Figure 12D, Chapter II). This suggests that additional DNA unwinding beyond what occurs during initial origin DNA melting requires interactions between the two hexamers at the origin. Specifically, a tight interface between the two helicases is necessary to drive further DNA unwinding. One way this tight interface could be necessary is if opposing action of the two helicases causes DNA unwinding and formation of the replication bubble. A requirement for the double-hexamer could occur
either before the two helicases separate or later when the two helicases come back together to pass one another during helicase activation (see Chapter I). The latter hypothesis is supported by recent studies suggesting CMG formation causes double-hexamer separation and origin DNA melting (Douglas et al., 2018; Langston & O'Donnell, 2019). Since we see CMG formation and origin DNA melting in the absence of double hexamers, it would suggest that double hexamers are not necessary at a step before separation. Our results combined with other data in the field support a model in which DNA unwinding occurs in at least three stages (Figure 1): 1) initial origin DNA melting upon CMG formation that can be independently mediated by each CMG (this study; Douglas et al., 2018), 2) further DNA unwinding that results in ssDNA-strand extrusion and requires double-hexamer interactions and Mcm10 (this study; Douglas et al., 2018; Langston & O'Donnell, 2019), and 3) extensive DNA unwinding that is mediated by CMG complexes translocating on ssDNA at replication forks (Fu et al., 2011; Georgescu et al., 2017).

One outstanding question that arises from our study is whether helicases that do not form double hexamers are competent for DNA unwinding at a replication fork. DNA unwinding at a replication fork occurs once the helicase encircles ssDNA and is positioned at a ssDNA-dsDNA junction. This question could be addressed by testing our mutant helicase on an artificial replication fork, effectively bypassing initial DNA unwinding. Previous studies have demonstrated that purified CMG complexes can be loaded onto the ssDNA end of a Y-forked DNA templates and subsequently activated for DNA unwinding (Ilves et al., 2010; Kose, Larsen, et al., 2019; Petojevic et al., 2015). Our mutant Mcm2-7
Figure 1: Model for the steps of DNA unwinding. Helicases are initially loaded as double hexamers. Once the CMG helicase forms, each CMG independently unwinds origin DNA (“initial origin DNA melting”) inside the helicase central channel. Next, helicase-helicase interactions, along with Mcm10, drive additional DNA unwinding that allows for ssDNA-strand extrusion. Once on ssDNA, the helicases can then pass one another to unwind DNA at independent replication forks.
construct could be inserted into the CMG purification strain and tested in a DNA unwinding assay using a Y-forked template. If our mutant CMG complex can unwind DNA in this assay, it would suggest that double-hexamer interactions are only required for the DNA unwinding that occurs during initiation and the mutant CMG is otherwise competent for DNA unwinding. Importantly, if there interactions are required it must be for an event after the ~ 0.7 helical turns of origin melting. If the mutant CMG complex is unable to unwind this template, it would indicate that, in addition to disrupting double-hexamer formation, the mutation disrupts the helicase’s ability to unwind DNA. Such a disruption in activity could result from an inability to engage properly with ssDNA either for passive binding or for active translocation. These possibilities could be distinguished by confirming whether the mutant CMG complex is able to bind and load onto Y-forked DNA in a DNA-binding assay. If the mutant CMG can bind this template but not unwind it, it would support that the mutant disrupts the helicase’s ability to translocate on ssDNA at a replication fork.

**Requirements for DNA unwinding: ssDNA-binding protein (SSB)**

In Chapter III, I explored what characteristics of SSBs are essential for DNA unwinding. Eukaryotic DNA unwinding requires an SSB (Douglas et al., 2018). However, it has not been explored whether the CMG helicase specifically requires its native SSB, RPA. We compared CMG DNA unwinding with RPA to two other SSBs that act during DNA replication in their respective systems, *E. coli* SSB and T4 bacteriophage Gp32. All three SSBs were competent to bind ssDNA, but only RPA and EcSSB supported extensive DNA unwinding by the CMG (Figures 1, 2A, 3A, and 4A, Chapter III). This result indicates that ssDNA binding is necessary but not sufficient for CMG DNA unwinding. Instead, there
must be some function essential for eukaryotic DNA unwinding that is shared by RPA and EcSSB but not Gp32. What this function is, however, remains to be answered.

One hypothesis is that eukaryotic DNA unwinding requires an SSB to prevent the newly unwound ssDNA from interfering with helicase function. Recent structural studies of the CMG helicase showed that the excluded strand (i.e. the strand that is not encircled by the CMG complex) interacts with the helicase central pore (Eickhoff et al., 2019; Georgescu et al., 2017; Goswami et al., 2018). Instead of being an interaction that facilitates DNA unwinding, Kose, Xie, et al., (2019) suggests that it leads to CMG stalling (Figure 6). They showed that RPA binding to the excluded strand strongly stimulates CMG helicase activity, which they suggest prevents the helicase from stalling. It follows that DNA unwinding requires an SSB that not only binds ssDNA but that binds it in a way that prevents the excluded strand from interfering with the helicase. This hypothesis is supported by studies of the E. coli replicative helicase, DnaB, that showed applying tension to the excluded strand, but not the encircled strand, stimulates helicase activity (Ribeck et al., 2010). In contrast, a similar study using the T4 bacteriophage replicative helicase, Gp41, demonstrated that tension on the excluded strand inhibited DNA unwinding (Ribeck & Saleh, 2013). These results fit the model that CMG DNA unwinding requires tension on the excluded strand that RPA and EcSSB provide but Gp32 does not.

Alternatively, the inability of Gp32 to support DNA unwinding may result from active inhibition of eukaryotic DNA unwinding. That is, Gp32 may directly interfere with CMG helicase activity rather than passively preventing DNA unwinding because it cannot
perform an essential function. This idea is consistent with elevated levels of Gp32 resulting in further reduction of unwound DNA (Chapter III, Figure 3A). Testing DNA unwinding by combining limited Gp32 with the optimal amount of RPA would demonstrate if Gp32 actively inhibits CMG helicase activity. If this mixed reaction shows reduced DNA unwinding compared to RPA alone, it would demonstrate direct inhibition by Gp32. If the mixed reaction instead looks the same as RPA alone, it would support that Gp32 lacks a necessary function that RPA possesses.

**An SSB requirement for DNA synthesis**

DNA replication requires DNA unwinding to separate dsDNA to expose the individual DNA strands that act as template for new DNA synthesis. Thus, at a minimum DNA replication requires SSB because of its essential function in DNA unwinding. However, before this study it was unclear whether RPA had another specific eukaryotic replisome function. Our finding that DNA synthesis was supported by either yeast RPA or (to a reduced extent) EcSSB suggests that eukaryotic DNA synthesis is possible with a non-native SSB (Figures 2B and 4B, Chapter III). This result was intriguing in light of the fact that simian virus SV40 DNA replication, the system where the details of eukaryotic DNA replication were first elucidated, specifically requires its native SSB for any detectable DNA synthesis.

In SV40 DNA replication, DNA unwinding is supported by various SSBs, but DNA synthesis requires human RPA specifically, indicating that human RPA has a specific function after DNA unwinding in this system (Brill & Stillman, 1989; Kenny et al., 1989;
Melendy & Stillman, 1993). The SV40 replicative DNA helicase, LTag, interacts directly with human RPA but not yeast RPA (Melendy & Stillman, 1993). The specificity of the interaction between LTag and human RPA is presumed to explain why only human RPA can support SV40 DNA replication. In contrast, the eukaryotic CMG complex has never been shown to directly interact with RPA. Our findings suggest that there is less specificity in the role of RPA in eukaryotic DNA replication and that the CMG does not require a direct, specific interaction with RPA. In addition, it supports that SV40 DNA replication specifically requires human RPA because of its ability to interact directly with LTag.

Although DNA synthesis occurred with both RPA and EcSSB, our findings suggest that lagging-strand synthesis has a specific requirement for RPA. DNA replication with EcSSB produced primarily long DNA products but limited short products (Figure 4B, Chapter III). These results suggest that RPA is required either for efficient priming or DNA synthesis of the lagging-strand (Figure 2). Priming inhibition would result in long lagging-strand products (Figure 2A), whereas synthesis inhibition would result in limited (if any) lagging-strand products (Figure 2B).

Closer examination of the DNA products made in DNA-synthesis reactions with EcSSB could be used to distinguish between the priming-inhibition and synthesis-inhibition models (Figure 2). These models could be distinguished by independently monitoring leading- and lagging-strand synthesis. For example, the strand(s) these products correspond to could be identified using a spot-blot assay with ssDNA-probes complementary to each strand. These probes would be applied to a membrane and tested
**Figure 2**

**A** Priming-inhibition model

In the first model, RPA is required for efficient priming by Pol α/primase. Without RPA, priming occurs infrequently resulting in long lagging-strand products.

**B** Synthesis-inhibition model

In the second model, RPA is required for lagging-strand synthesis by Pol δ (e.g. for Pol α – Pol δ switching). Without RPA, Pol δ activity is inhibited leading to few or no lagging-strand products. Throughout the figure, red represents RNA-DNA primers and grey represents DNA.

**Figure 2:** Two models for lagging-strand DNA synthesis in the absence of RPA. (A) In the first model, RPA is required for efficient priming by Pol α/primase. Without RPA, priming occurs infrequently resulting in long lagging-strand products. (B) In the second model, RPA is required for lagging-strand synthesis by Pol δ (e.g. for Pol α – Pol δ switching). Without RPA, Pol δ activity is inhibited leading to few or no lagging-strand products. Throughout the figure, red represents RNA-DNA primers and grey represents DNA.
with the radioactive DNA products made in EcSSB and RPA (as a control) DNA-synthesis assays. After washing away unbound DNA, this assay would determine which strands the DNA products were synthesized from by evaluating which probes showed radioactive signal. A positive signal for the leading-strand probe would confirm that the long strands observed included leading-strand products and that leading-strand synthesis does not specifically require RPA (i.e. EcSSB can substitute). If the lagging-strand probe is bound by the replication products at the same level as the leading-strand, it would support the priming inhibition hypothesis that the long DNA products we see are a mixture of leading-strand and long lagging-strand products. On the other hand, if the lagging-strand probe shows low or no binding to our DNA products, it would support that overall lagging-strand synthesis is inhibited resulting in no or few lagging-strand products being made with EcSSB. As reviewed in the Chapter III Discussion, inhibition of priming would suggest a specific interaction between RPA and Pol α/primase whereas inhibition of synthesis would support a specific interaction between RPA and Pol δ that is required for efficient lagging-strand DNA replication. Based on the results of this experiment, the specific interaction between RPA and either Pol α/primase or Pol δ could be further evaluated in the future.
REFERENCES


