STRUCTURAL BIOLOGY OF THE C-TERMINAL DOMAIN OF EUKARYOTIC REPLICATION FACTOR MCM10

By

Patrick David Robertson

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Biological Sciences

August, 2010

Nashville, Tennessee

Approved:

Brandt F. Eichman

Walter J. Chazin

James G. Patton

Hassane Mchaourab

To my wife Sabrina,

thank you for your enduring love and support

ACKNOWLEDGMENTS

I would like to begin by expressing my sincerest gratitude to my mentor and Ph.D. advisor, Dr. Brandt Eichman. In addition to your excellent guidance and training, your passion for science has been a source of encouragement and inspiration over the past five years. I consider working with you to be a great privilege and I am grateful for the opportunity. I would also like to thank the members of my thesis committee: Drs. Walter Chazin, Ellen Fanning, James Patton, and Hassane Mchaourab. My research and training would not have been possible without your insight, advice and intellectual contributions.

I would like to acknowledge all of the members of the Eichman laboratory, past and present, for their technical assistance and camaraderie over the years. I would especially like to thank Dr. Eric Warren for his contributions to the research presented here, as well for his friendship of the years. I would also like to thank Drs. Benjamin Chagot and Sivaraja Vaithiyalingam from the Chazin laboratory for their expert assistance and NMR training.

I would especially like to thank my parents, Joyce and David, my brother Jeff, and the rest of my family for their love and encouragement throughout my life. None of this would have been possible without all of your support. I would like to thank all my friends who helped make my time at Vanderbilt enjoyable. Happy hours, intramural sports and Friday lunches were a much needed distraction over of the years and have left me with many great memories. Last, but certainly not least, I would like to acknowledge my beautiful and brilliant wife, Sabrina. Your kindness, compassion, and fortitude have no boundaries. I could not have made it without you.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

LIST OF ABBREVIATIONS

aa	Amino acids
AAA($^{+}$)	ATPases associated with diverse cellular activities
ARS	Autonomously replicating sequence
AT	Adenine and thymine
ATP	Adenosine triphosphate
BME	2-mercaptoethanol
bp	Base pair
C	Celsius
CCCC	Cysteine-Cysteine-Cysteine-Cysteine
CCCH	Cysteine-Cysteine-Cysteine-Histidine
CD	Circular Dichroism
Cdc	Cell division cycle
CDK	Cyclin dependent kinase
Cdt1	Cdc10-dependent transcript 1
CMG	Cdc45/Mcm2-7/GINS
COSY	Correlation spectroscopy
CTD	Carboxyl-terminal domain
Ctf4	Chromosome transmission fidelity
DDK	Dbf-4 dependent kinase
Dbf4	Dumb bell forming 4
DNA	Deoxyribonucleic acid

DNA43	DNA synthesis defective 43
Dpb11	DNA polymerase B subunit 11
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
EDTA	Ethylenediaminotetraacetic acid
EM	Electron Microscopy
g	Gram
G1/2	Gap 1/2 phase
GFAA	Gas Furnace Atomic Absorption
GINS	Go ichi ni san
GST	Glutathione-S-transferase
h	Hour
HCl	Hydrochloric acid
HSQC	Heteronuclear single quantum coherence
IC	Internal + C-terminal domain
ID	Internal domain
IP	Immunoprecipitation
IPTG	Isopropyl thio-beta-D-galactopyranoside
ITC	Isothermal titration calorimetry
kb	Kilobase pair
k_d	Dissociation constant
kDa	Kilo Dalton
L	Liter

LB	Luria broth
M	Molar
MALDI-TOF	Matrix assisted laser desorption ionization – time of flight
Mcm/MCM	Mini-chromosome maintenance
CeMcm10	*Caenorhabditis elegans* Mcm10
DmMcm10	*Drosophila melanogaster* Mcm10
HsMcm10	*Homo sapiens* Mcm10
MmMcm10	*Mus Musculus* Mcm10
MthMCM	*Methanobacterium thermoautotrophicum* MCM
ScMcm10	*Saccharomyces cerevisiae* Mcm10
SpMcm10	*Schizosaccharomyces pombe* Mcm10
SsoMCM	*Sulfolobus solfataricus* MCM
XlMcm10/XMcm10	*Xenopus laevis* Mcm10
μg	Mirogram
μM	Miromolar
mg	Milligram
min	Minute
mL	Milliliter
mM	Millimolar
MS	Mass Spectrometry
MTS	Methanethiosulfonate
Ni-NTA	Nickel-nitrilotriacetic acid
ng	Nanogram

NLS	Nuclear localization signal
nm	Nanometers
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
nt	Nucleotide
NTD	Amino-terminal domain
OD	Optical density
OB-fold	Oligonucleotide/oligosaccharide binding-fold
ORC	Origin recognition complex
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PDB	Protein databank
PEG	Polyethylene glycol
PMSF	Phenylmethylsulfonyl fluoride
Pol	DNA polymerase
Pol α	DNA polymerase α-primase
Pre-IC	Pre-initiation complex
Pre-RC	Pre-replication complex
PVDF	Polyvinylidene difluoride
RFC	Replication factor C
RPA	Replication protein A
rpm	rotations per minute

rmsd	Root-mean-square Deviation
RNA	Ribonucleic acid
RNAi	Ribonucleic acid interference
S	Synthesis phase
s.d.	Standard deviation
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
sec	Second
Sld	Synthetic lethality with dpb11
ssDNA	Single-stranded DNA
Tris	Tris-(hydroxymethyl-)aminomethane
TOCSY	Total correlation spectroscopy
TopBP1	Topoisomerase II binding protein 1
WT	Wild-type

CHAPTER I

INTRODUCTION

DNA replication is arguably the most critical and fundamental process in every living cell. The ability to replicate one's genome for subsequent generations is the foundation for processes such as speciation and evolution. This process of copying and passing on genetic information to a daughter cell is referred to as the cell cycle (Figure 1). Non-dividing, or quiescent, cells remain in Gap 0 or resting phase until they enter back into active division. Cells in G1 phase possess elevated biosynthetic activities, which includes increased production of enzymes required for S phase. In addition, the majority of cell growth occurs during this phase of the cell cycle. As cells transition into synthesis (S) phase, DNA replication is initiated and continues until all of the chromosomes have been duplicated. Once DNA synthesis is complete, cells enter into a short G2 phase where they undergo a checkpoint to determine if they are prepared for cell division. Cells ready for mitosis enter into M phase, where the mother cell divides into two genetically identical daughter cells. At this point, cells reenter into G2 and prepare for a new cell cycle.

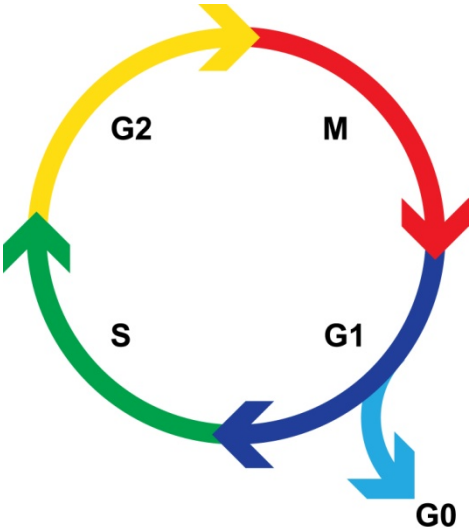


Figure 1. Schematic of the cell cycle. M represents Mitosis; G0, G1, and G2 are gap phases; S represents synthesis phase. Cycle lengths are not to scale.

Overview of DNA Replication

The overall mechanism of DNA replication in S-phase is conserved across the three different kingdoms of life. From a simplistic point of view, replication can be broken down into three core stages: initiation, elongation and termination. Initiation refers to the time in which a cell recruits and assembles multiple proteins into a macromolecular machine referred to as the replisome. This replisome complex includes a helicase to separate the two strands of duplex DNA and expose the nucleotide bases, and polymerases to create complementary strands. While the replisome is assembled during initiation, it becomes activated during elongation. During this stage, bulk DNA synthesis is initiated from origins of replication and two replication forks are formed which then proceed bidirectionally away from the origin. In addition to the helicase and polymerases, numerous other proteins are required for appropriate elongation (Figure 2). For example, single-strand DNA binding proteins bind and stabilize the exposed strands, while ligases seal the nicks between adjacent Okazaki fragments.

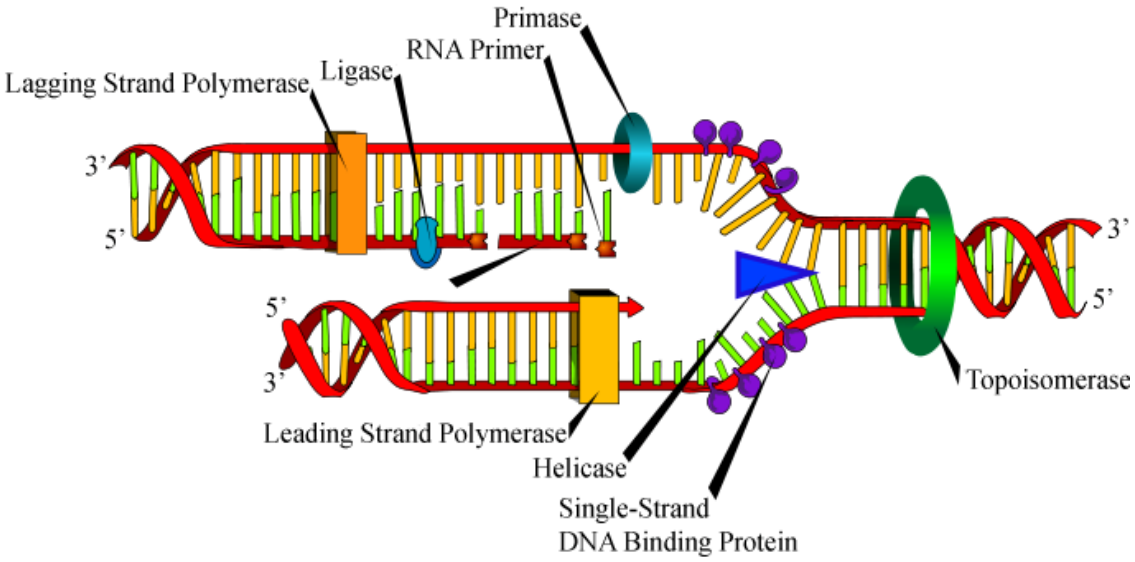


Figure 2. Model DNA replication fork.

Additionally, topoisomerases are critical for relieving the tension from positive supercoiling which occurs during the unwinding process. When replication forks travelling in opposite directions meet one another, indicating that all DNA in between the origins of these forks has been successfully copied, the elongation process is terminated. While all forms of life employ a variation of these three fundamental steps, these cellular processes become increasingly complex with each step up the evolutionary ladder.

DNA Replication in Prokaryotes

While most core components of DNA replication are shared across all organisms, striking differences exist between the various kingdoms of life. Prokaryotes employ the simplest mechanism to copy their genome. Replication begins at a single origin when DnaA binds to repetitive, non-palindromic sequences called DnaA boxes within the OriC. Upon binding, DnaA hydrolyzes ATP and separates the duplex AT-rich regions flanking the origin (Speck et al. 2001). These melted regions of DNA allow the helicase DnaB to be loaded onto the exposed single-strand DNA (ssDNA) (reviewed in (Caruthers et al. 2002)). As unwinding is initiated, single-strand DNA binding protein (SSB) binds ssDNA as it becomes exposed on the lagging strands. The elongation phase begins when the DnaG primase synthesizes RNA primers of 10-12 nucleotides in length which are later extended by DNA polymerase III (reviewed in (Corn et al. 2006)). During elongation, the leading and lagging strand polymerases remain physically tethered to the helicase. While the leading strand is able to synthesize complementary bases in the 5′ to 3′ direction with a single polymerase, the lagging strand must copy smaller regions successively due to the unidirectionality of the polymerase. As a result, a lagging strand

loop extends out of the replisome as it is being replicated until the polymerase reaches the 3′ end of a previously replicated Okazaki fragment. At this point, the polymerase uncouples from the DNA and the loop is released, and the process begins anew (reviewed in (Langston et al. 2009)). As the replication fork progresses, unwinding by the helicase causes positive supercoiling upstream of the fork that must be relieved by a topoisomerase/gyrase which breaks the DNA to relieve tension and then reanneals the lesion (reviewed in (Cozzarelli 1980)). Replication can be terminated in two ways. Either the two opposing replication forks physically collide opposite the origin on the circular chromosome, or the forks collide with a protein named Tus which binds to *cis*-acting Ter elements at specific termination sites on the chromosome (reviewed in (Dalgaard et al. 2009)).

Eukaryotic DNA Replication

DNA replication in prokaryotic organisms is remarkably simple compared to replication in eukaryotic organisms. For example, eukaryotes initiate replication from multiple origins, as opposed to just one, to accommodate their larger genome (reviewed in (Machida et al. 2005)). However, the most striking difference between the two systems is the complexity of the protein machinery and complexes involved. For example, prokaryotes have a single protein (DnaA) that binds to, unwinds and initiates origins. In contrast, eukaryotes require intricate coordination of multiple protein complexes to perform comparable tasks. The assembly and activation of these complexes require specific recruitment and loading pathways, as well as a series of unique phosphorylation events.

Eukaryotic DNA replication initiation begins in the G1 phase of the cell cycle and is arguably the most critical stage of replication. During this step, all of the protein components required for DNA synthesis are recruited and assembled into the replisome (Figure 3). Errors during this stage of replication can be detrimental to cell viability and/or the stability of its genome. Indeed, a number of human diseases, including cancer, have been shown to result from errors during the initiation process (reviewed in (Lau et al. 2007)). As such, a better understanding of the mechanisms driving this process is necessary to aid in the development of new therapeutics for such diseases. The first step of DNA replication initiation is marked by the ATP-dependent binding of the origin recognition complex (ORC) to predefined regions of chromatin known as origins of replication. In budding yeast, these origins are known to contain 11 bp autonomously replicating sequences

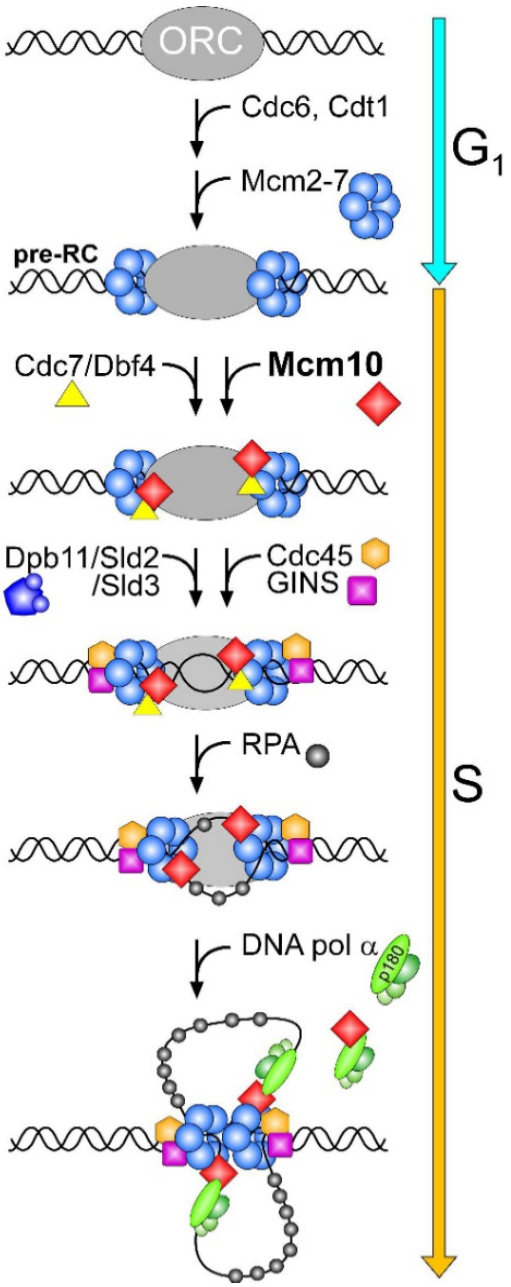


Figure 3. Overview of eukaryotic DNA replication initiation. This is a general order of recruitment of initiation proteins to the replisome determined from multiple studies in several different eukaryotic systems.

(ARS) and most fire with similar efficiency (Stinchcomb et al. 1979; Broach et al. 1983). However, origins in fission yeast and higher eukaryotes lack a consensus sequence and fire with variable efficiency (reviewed in (Aladjem 2007)). The mechanisms that dictate origin selection and firing efficiency are currently unclear, although the differential recruitment of ORC has been implicated in both of these processes (Wu et al. 2009).

Once origins are appropriately identified and loaded with ORC, replication factors Cdc6 and Cdt1 facilitate the loading of mini-chromosome maintenance proteins 2-7 (Mcm2-7) onto these ORC bound regions of chromatin (Blow et al. 2005). Mcm2-7, which forms a heterohexameric complex, has been shown to possess all of the hallmark activities of a helicase through an extensive series of *in vitro* studies (reviewed in (Bochman et al. 2009)). In addition, Mcm2-7 is also a component of the replicative helicase complex, and is believed to act as the ATPase motor (Gambus et al. 2006; Moyer et al. 2006). Although the precise mechanism for how this complex unwinds double-strand DNA is still unclear, recent studies indicate that Mcm2-7 assembles into a head-to-head double hexamer when loaded onto DNA (Remus et al. 2009). This double hexamer structure is similar to the archaeal helicase MCM, which is thought to be an evolutionary predecessor of the eukaryotic Mcm2-7 (Fletcher et al. 2003; Liu et al. 2009). Despite loading as a double hexamer, only one copy of Mcm4 was identified in a purified replisome complex, therefore implying that the helicase functions as single hexamer *in vivo* (Gambus et al. 2006). After Mcm2-7 loading, the origin becomes functional and is considered to be 'licensed.' Once cells begin the transition into S phase, cyclin levels prevent these origins from re-licensing, and thus prevent any region of chromatin from being replicated more than once (reviewed in (Blow et al. 2005)).

At the onset of S phase, two important phosphorylation events occur. First, cyclin-dependent kinase (CDK) phosphorylates Sld2 and Sld3, which facilitates their binding to Dpb11 (Tanaka et al. 2007; Zegerman et al. 2007). Secondly, a complex of Cdc7 and Dbf4, called DDK, directly phosphorylates the Mcm2 and Mcm4 subunits of the Mcm2-7 complex (Lei et al. 1997; Sheu et al. 2006). This particular phosphorylation event is dependent on an additional factor, Mcm10, which is thought to recruit DDK to the pre-RC (Lee et al. 2003). These two phosphorylation events are followed by the binding of two helicase cofactors, Cdc45 and GINS, to Mcm2-7 which forms the CMG complex (Wohlschlegel et al. 2002; Moyer et al. 2006; Pacek et al. 2006). This CMG complex is thought to function as the active replicative helicase *in vivo*, due to the fact that Mcm2-7 has limited helicase activity alone, yet possesses enhanced activity in the presence of Cdc45 and GINS (Bochman et al. 2008; Ilves et al. 2010). In addition to phosphorylation of Mcm2-7, it has recently been shown that incorporation of Cdc45 and GINS into the CMG complex is also dependent on Mcm10, Ctf4 and RecQL4 (homolog of Sld2 in yeast) (Im et al. 2009).

At this point during the initiation process, origin duplex DNA begins to be melted, presumably by the CMG helicase, and RPA begins to bind the exposed single stranded DNA (Tanaka et al. 1998; Zou et al. 2000). Soon after RPA loads, DNA polymerase α-primase (pol α) is recruited to the replisome in a Mcm10 dependent manner (Walter et al. 2000; Ricke et al. 2004; Yang et al. 2005). Both the localization and stability pol α have been shown to be affected by Mcm10, implicating it as a molecular chaperone for pol α *in vivo* (Ricke et al. 2004; Yang et al. 2005; Ricke et al. 2006). The pol α complex is composed of four subunits, each of which is essential for viability

(reviewed in (Sugino 1995)). The p48 and p58 subunits comprise the primase activity of the complex and are responsible for synthesizing short RNA primers, 7-10 nt in length (reviewed in (Arezi et al. 2000)). The p180 subunit, which has catalytic DNA polymerase activity, extends the primer to ~30 nt (Conaway et al. 1982). The fourth subunit, p68, has no known enzymatic activity, but is thought to perform a regulatory role within the complex (reviewed in (Mizuno et al. 1998)). After primers have been synthesized and extended by pol α, the processive DNA polymerases δ and ϵ are recruited to begin bulk DNA synthesis (reviewed in (Burgers 2009)). However, prior to pol δ and ϵ recruitment, PCNA must be loaded by its clamp loader RFC. Once loaded, the ring-shaped PCNA protein acts as a processivity factor for DNA pol δ and ϵ, helping to maintain their association with DNA (Garg et al. 2005). Specifically pol δ has been shown to act primarily as the leading strand polymerase, while pol ϵ synthesizes primarily on the lagging strand (Nick McElhinny et al. 2008). Once the processive DNA polymerases are loaded, the replisome is prepared for elongation phase where bulk synthesis of complementary DNA will occur.

During the elongation step of replication, thousands of replisome complexes distributed across each chromosome will begin to copy the genome. In addition to the components described above, a number of other proteins are required for this stage of replication. Once a region of DNA has been replicated, the RNA primer must then be removed and replaced by DNA. While this process occurs on both strands of DNA, the lagging strand has many more instances due to synthesis of short Okazaki fragments unlike the more continuous synthesis on the leading strand. These RNA regions are excised by RNase H1 and then further processed by FEN1, allowing DNA ligase to ligate

the two strands together (reviewed in (Burgers 2009)). The end product of this mechanism is a continuous strand of DNA, complementary to the template strand. In addition to Okazaki fragment processing, super-coil tension induced upstream of the replication fork poses another problem during elongation. This tension is relieved by topoisomerases, which introduce single-strand breaks in the DNA, thereby allowing the DNA to remain relaxed as the replicative helicase continuously unwinds the duplex DNA(reviewed in (Leppard et al. 2005)).

The final stage of DNA replication is termination. This occurs when two replication forks meet one another or when replication reaches the telomeric region at the end of a chromosome. Both situations result in the dissociation of the replisome and signal to the cell that replication of all genetic material is complete. Upon termination, cells contain two complete and identical copies of their genome, and are now able to enter into G2 phase and prepare further for mitosis.

The Mcm Family of Proteins

Mcm2-7

The first mini-chromosome maintenance genes were indentified in a genetic screen performed in *Saccharomyces cerevisiae* (*S. cerevisiae*) searching for genes involved in initiating DNA replication. The study found that mutations in *MCM2*, *MCM3* and *MCM5* genes resulted in defective plasmid segregation (Maine et al. 1984). Additional MCM mutants were also identified in this screen, including *MCM1*, *MCM17*, *MCM21* and *MCM22*, however these genes products have since been shown to function as transcription factors or participate in aspects of chromosome segregation (Hayes et al.

1988; Passmore et al. 1988; Passmore et al. 1989; Roy et al. 1997; Poddar et al. 1999). These proteins are not related or similar to classical MCM proteins, nor are they involved in DNA replication initiation; as such, they will not be discussed in any further detail. Additionally, *MCM4* and *MCM7* genes were identified in later studies as being important for cell cycle progression (Moir et al. 1982; Hennessy et al. 1991) while *MCM6* was found to be necessary for chromosome segregation (Takahashi et al. 1994). Initially, the protein products of these six genes were shown to form a heterohexameric complex which appeared to have no enzymatic activity (Coleman et al. 1996; Romanowski et al. 1996; Adachi et al. 1997). Despite the lack of *in vitro* function, subsequent mutational analysis of the ATP hydrolysis activity of these proteins revealed it to be essential for viability (Schwacha et al. 2001; Gomez et al. 2002). More recent investigations have successfully shown this complex to possess *in vitro* helicase activity (Bochman et al. 2008), as well function as a component of the CMG replicative helicase complex (Moyer et al. 2006; Ilves et al. 2010).

The Mcm2-7 proteins are classified as AAA^+ proteins and contain a well-conserved ATPase domain that has been shown to couple ATP hydrolysis with unwinding duplex DNA (Koonin 1993; Bochman et al. 2007; Bochman et al. 2008). While these proteins form a ring shaped hexamer, each subunit appears to possess distinct biochemical characteristics *in vitro*, likely due to the divergent sequences outside of the ATPase domain (Bochman et al. 2007; Bochman et al. 2008). As an example, Mcm4, Mcm6 and Mcm7 are required for ATP hydrolysis, while Mcm2, Mcm3 and Mcm5 appear to be involved with DNA unwinding activity (reviewed in (Bochman et al. 2009)). In support of this hypothesis, Walker A active site mutations in Mcm2 and Mcm5 failed

to disrupt ATP hydrolysis of the complex. However, DNA unwinding activity was nearly abolished in these mutant complexes (Bochman et al. 2008). This observation has led to proposal of a gated model for Mcm2-7 activity, in which an ATP-dependent conformational change between Mcm2 and Mcm5, dictates on and off states for the unwinding activity of the entire complex (Bochman et al. 2008). Significant progress has been made in understanding the precise biochemical function of Mcm2-7 (reviewed in (Bochman et al. 2009). However, the molecular mechanism of this complex within the replisome remains to be determined.

Archaeal MCM

Since the completion of sequencing for several archaea genomes, it has become apparent that DNA replication in archaea is more closely related to eukaryotes than prokaryotes. Indeed, many proteins involved in eukaryotic replication have newly identified homologs that have been characterized in archaeal systems. One example is the protein MCM, which functions as the replicative helicase in archaea (reviewed in (Tye 2000)). MCM forms a homohexamer and has high sequence homology to the heterohexameric eukaryotic helicase, Mcm2-7. Each of the proteins contain a well-conserved ATPase domain, and some of the proteins (MCM, Mcm4, Mcm6 and Mcm7) were shown to possess a conserved ($CXXCX_nCXXC$) zinc motif (Poplawski et al. 2001). This high sequence homology led to an evolutionary investigation into the MCM family of proteins which confirmed archaeal MCM to be an evolutionary predecessor of eukaryotic Mcm2-9 proteins (Liu et al. 2009). In support of this conclusion, an array of biochemical studies, including helicase activity, ATP hydrolysis, DNA binding and

protein-protein interactions, showed that MCM *in vitro* function is comparable to Mcm2-7 (Kelman et al. 1999; Poplawski et al. 2001; Carpentieri et al. 2002). In summary, archaeal MCM proteins exhibit robust 3′ to 5′ helicase activity that is dependent on ATP and Mg^{2+}, and are capable of unwinding a number of different DNA substrates (reviewed in (Bochman et al. 2009)).

In addition to biochemical analysis, archaeal MCM has also been extensively studied using structural techniques. These techniques have included analysis by electron microscopy (EM) (Gomez-Llorente et al. 2005; Costa et al. 2008; Bae et al. 2009)}, as well as crystal structures of MthMCM and SsoMCM (Fletcher et al. 2003; Brewster et al. 2008; Liu et al. 2008). Despite the extensive structural information for MCM, there is currently no high resolution data for a full-length hexameric complex. However, structures of individual domains or monomeric MCM have been successfully determined. The first high resolution structure was of the N-terminal domain of MthMCM which revealed a head-to-head double hexamer assembly (Fletcher et al. 2003). The hexamer-hexamer interface between the two NTD was comprised of zinc motifs from one hexamer intercalating with those of the opposing hexamer. In contrast, a later crystal structure of the NTD of SsoMCM showed a hexamer formation (Liu et al. 2008), however there was no double hexamer formation as seen with MthMCM. The oligomeric state observed in both structures was supported by studies using gel filtration chromatography, which revealed a preference for SsoMCM to form a hexamer and MthMCM to form double hexamers (Chong et al. 2000; Carpentieri et al. 2002). The only structure of a near full-length MCM protein showed the protein to be monomeric (Brewster et al. 2008), however imposing six-fold symmetry on the structure enabled docking it into a low

resolution EM reconstruction (Bae et al. 2009). While the structural data obtained from these archaeal MCM studies has advanced our understanding of MCM function, new structural studies targeting a full-length complex are critical.

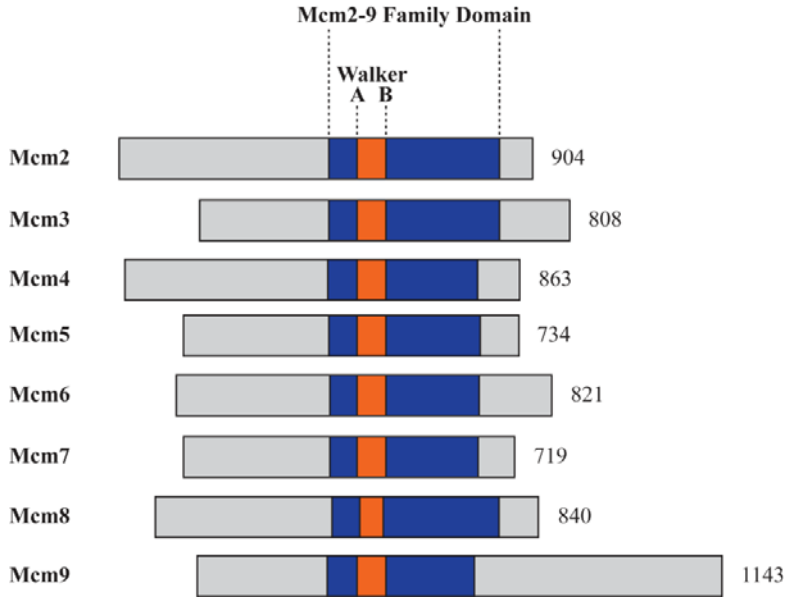


Figure 4. Schematic alignment of human Mcm2-9 proteins. The region containing the ATPase domain, which is highly conserved amongst Mcm2-9, is shown in blue. The area corresponding to the Walker A and B motifs is shown in orange. Numbers on the left indicate protein length. This alignment is adapted from (Maiorano et al. 2006).

Mcm8

One of the latest additions to the MCM family is Mcm8, originally identified in a screen to isolate cancer-related genes using Hepatitis B integration sites in hepatocellular carcinomas. In addition to sequence homology with Mcm2-7 proteins, the nuclear localization and cell cycle regulation of this protein suggested a role in DNA replication (Gozuacik et al. 2003). Further studies showed that Mcm8 colocalizes with RPA at replication foci, hydrolyzes ATP, and possesses *in vitro* helicase activity. Depletion of Mcm8 in *Xenopus* extracts had no effect on licensing, however it did reduce the rate of

DNA synthesis, implying a role during elongation rather than initiation (Maiorano et al. 2005). Contrary to this, investigations by other groups have shown that Mcm8 interacts with ORC and Cdc6 in G1, and RNAi knock-down of Mcm8 slowed S phase and reduced loading of Mcm2-7 onto chromatin, both of which are more indicative of a role in pre-RC formation (Volkening et al. 2005; Kinoshita et al. 2008). While the precise function of Mcm8 remains elusive, its involvement in DNA replication is indisputable, as is its membership in the MCM family.

Mcm9

The newest member of the MCM family is Mcm9, which was identified by studies using BLAST searches of different genomes for protein sequences similar to Mcm2-8. Mcm9 showed considerable sequence conservation to Mcm2-8, and it contained the classic ATPase domain and putative zinc motif. In addition, the authors also noted that Mcm9 was most similar to Mcm8, more so than to Mcm2-7, suggesting that these two proteins may be a sub-class of MCM (Lutzmann et al. 2005). These findings were confirmed by a separate group using a bioinformatics approach looking for proteins similar to Mcm2-8 (Yoshida 2005). Subsequent investigation of Mcm9 function revealed that it binds to chromatin in an ORC-dependent manner, is required for Mcm2-7 loading, and that it forms a stable complex with Cdt1. From these results the authors propose that Mcm2-7 loading is not only regulated by geminin, but also by Mcm9 in an opposite mechanism (Lutzmann et al. 2008). While the molecular mechanism of Mcm2-7 loading is unclear, these findings indicate that Mcm9 is an integral participant in this process.

Mcm10

The first Mcm10 gene, initially referred to as *Cdc23*, was originally indentified in a genetic screen searching for cell division cycle mutants in *Schizosaccharomyces pombe* (*S. pombe*). In this study, temperature sensitive alleles of *Cdc23* disrupted bulk DNA synthesis, and thus blocked DNA replication and mitosis (Nasmyth et al. 1981). Two subsequent similar screens in *Saccharomyces cerevisiae* (*S. cerevisiae*) also identified this gene, which was referred to as *DNA43* (Dumas et al. 1982) and *MCM10* (Maine et al. 1984). These findings were further extended by a study showing *DNA43* was essential for entry into S phase and maintaining cell viability (Solomon et al. 1992). Another interesting observation was that Mcm10 protein associates with individual subunits of the Mcm2-7 complex, suggestive of a role at early replication origins (Merchant et al. 1997). Soon thereafter, *Cdc23* and *DNA43* were determined to be homologs of one another, indicating that Mcm10 is a conserved eukaryotic replication factor. Importantly, this work also revealed putative CCCH zinc motifs in Mcm10 proteins from both *S. pombe* and *S. cerevisiae* (Aves et al. 1998). Mcm10 was also shown to associate with chromatin at known replication origins in a manner dependent on ORC and Mcm2-7 (Homesley et al. 2000). This localization to origins was later explained by identifying a physical interaction between Mcm10 and ORC (Izumi et al. 2000; Kawasaki et al. 2000). All of these results confirmed that Mcm10 was an important eukaryotic initiation factor and sparked a series of investigations to better understand its function.

The original identification and characterization of Mcm10 was completed in yeast systems. However, subsequent studies identified Mcm10 homologs in higher eukaryotes, including *Drosophila*, *Xenopus* and humans (Izumi et al. 2000; Wohlschlegel et al. 2002;

Christensen et al. 2003). Human Mcm10 (HsMcm10) binds chromatin at the G1/S boundary, contrary to yeast Mcm10 which stays constitutively bound to DNA throughout the cell cycle (Izumi et al. 2001; Liang et al. 2001). This recruitment at early S phase also eliminated Mcm10 as a possible component of the pre-RC which is assembled in G1. Importantly, these studies also revealed ubiquitinated (mono and di) and phosphorylated states of HsMcm10, and showed it to be regulated by proteosomal degradation as opposed to being controlled at the transcriptional level (Izumi et al. 2001). The next insight into Mcm10 function came using an *in vitro* Xenopus oocyte assay, which allows for depletion of individual protein factors during different stages of DNA replication. Using this technique, *Xenopus* Mcm10 (XMcm10) was shown to bind chromatin at the onset of S phase, similar to the human protein. Additionally, recruitment of XMcm10 to DNA required an already assembled pre-RC, yet preceded pre-IC formation, marked by Cdc45 and RPA recruitment. This study also revealed that XMcm10 protein levels were roughly one Mcm10 per 5000 bp of DNA or around 2 per active origin (Wohlschlegel et al. 2002). An interesting tie between Mcm10 function in yeast and higher eukaryotic systems was that *Drosophila* Mcm10 (DmMcm10) is capable of rescuing temperature sensitive depletion of Mcm10 in *S. pombe* (Christensen et al. 2003). Thus, despite subtle functional differences between yeast and vertebrate Mcm10, these homologs appeared to be extremely well conserved. In addition, the sequence XMcm10 differs only slightly from HsMcm10, suggesting these two homologs possess similar structure and activity.

As a result of these studies, it was clear Mcm10 was functioning somewhere between pre-RC and pre-IC formation, yet little was known about the mechanism during this transition. It was shown that Mcm10 is necessary for phosphorylation of the Mcm2-7

complex by DDK, a step which is currently thought to play a role in helicase activation or aid in the binding of helicase cofactors (Lee et al. 2003). Soon after this observation, Mcm10 was shown to interact with pol α, specifically the p180 subunit (Fien et al. 2004). In addition, this binding was shown to increase the polymerase activity of pol α and stabilized it *in vivo* by preventing proteosomal degradation (Fien et al. 2004; Ricke et al. 2004; Ricke et al. 2006). In addition to intimate interactions with pol α, Mcm10 has also been shown to interact genetically with pol ε and δ (Kawasaki et al. 2000). These findings contradicted a single role for Mcm10 during early S phase due to the fact that polymerases aren't recruited to the replisome until after pre-IC assembly. Instead, these results suggested that Mcm10 plays two important roles during replication initiation. The first occurs early in initiation and involves some aspect of activating the pre-RC, either by inducing a phosphorylation event or recruiting additional factors. The second role of Mcm10 occurs after the pre-IC is assembled and involves the recruitment of polymerases, the final step prior to elongation.

Despite decades of research, there remain a number of unanswered questions pertaining to Mcm10 function. Most recently, the role of Mcm10 in facilitating CMG formation was suggested, however it is not clear in what regard Mcm10 is required. Furthermore, two other proteins, Ctf4 and RecQL4, implicated in this mechanism have interactions with Mcm10 that have not been investigated yet. Similarly, Mcm10 has been shown to interact with and stabilize pol α however, it is not known whether this interaction is a means of recruitment or simply a tether for connecting the polymerase to the replisome. Additionally, the consequences of Mcm10 post-translational modifications are not clearly understood. While they have been shown to affect its interaction with

PCNA (Das-Bradoo et al. 2006), the implications of this modification have not been dissected. One final aspect of Mcm10 that remains a mystery is its oligomeric state. Mcm10 has been reported to exist in monomeric, dimeric, and hexameric states throughout the literature, making it difficult to fully understand the structural potential of Mcm10 in the replisome. Much of the work presented in subsequent chapters was designed with the intention of answering some of these fundamental questions, and thus further our understanding of Mcm10 function.

Scope of this Work

This dissertation encompasses the progress made towards elucidating the role of Mcm10 in eukaryotic DNA replication initiation by a structural and biochemical characterization of Mcm10 protein from *Xenopus laevis*. Chapter 2 introduces the domain architecture of vertebrate Mcm10, and a series of biochemical studies further address the functions of each domain individually. In Chapter 3, the structure of the C-terminal zinc-coordinating region of Mcm10 is presented and biochemical analyses provide insight into the biological role of this domain. In addition, the spatial relationship of the DNA binding domains is investigated in order to better understand how Mcm10 coordinates this function *in vivo*. Chapter 4 describes preliminary studies on the full-length Mcm10 protein, including DNA binding and oligomerization. Finally, Chapter 5 discusses how the overall conclusions from the work presented here have expanded our knowledge of Mcm10 function. In addition, future directions for this research are described.

CHAPTER II

DOMAIN ARCHITECTURE AND BIOCHEMICAL CHARACTERIZATION OF VERTEBRATE MCM10*

Abstract

Mcm10 plays a key role in initiation and elongation of eukaryotic chromosomal DNA replication. As a first step to better understand the structure and function of vertebrate Mcm10, we have determined the structural architecture of Xenopus laevis Mcm10 (XMcm10) and characterized each domain biochemically. Limited proteolytic digestion of the full-length protein revealed N-terminal (NTD), internal (ID), and C-terminal (CTD) structured domains. Analytical ultracentrifugation revealed that XMcm10 self-associates and that the N-terminal domain forms homodimeric assemblies. DNA binding activity of XMcm10 was mapped to the ID and CTD, each of which binds to single- and double-stranded DNA with low micromolar affinity. The structural integrity of XMcm10-ID and CTD is dependent on the presence of bound zinc, which was experimentally verified by atomic absorption spectroscopy and proteolysis protection assays. The ID and CTD also bind independently to the N-terminal 323 residues of the p180 subunit of DNA polymerase α-primase. We propose that the modularity of the protein architecture, with discrete domains for dimerization and for binding to DNA and DNA polymerase α-primase, provides an effective means for coordinating the biochemical activities of Mcm10 within the replisome.

* The work presented in this chapter was published in Robertson, P. D.[1], E. M. Warren[1], H. Zhang[1], D. B. Friedman, J. W. Lary, J. L. Cole, A. V. Tutter, J. C. Walter, E. Fanning and B. F. Eichman (2008). J Biol Chem 283(6):3338-48. ([1] designates co-authorship)

Introduction

Eukaryotic DNA replication is carried out by large multiprotein machines that coordinate DNA unwinding and synthesis at the replication fork. Initiation of replication involves ordered assembly of the replisome and local denaturation of duplex DNA at the origin followed by replisome activation. Screens for mutants defective in minichromosome maintenance (Mcm) and DNA replication in yeast identified a number of factors essential for replication (Nasmyth et al. 1981; Maine et al. 1984; Solomon et al. 1992; Merchant et al. 1997). Pre-replicative complexes composed of the origin recognition complex, Cdc6, Cdt1, and the hexameric Mcm2-7 helicase are assembled in G_1 (Blow et al. 2005) and converted into active replication forks at the onset of S phase. Mcm10 loads onto chromatin after pre-replicative complex assembly (Wohlschlegel et al. 2002; Ricke et al. 2004) and stimulates phosphorylation of Mcm2-7 by Dbf4-Cdc7 kinase (Lee et al. 2003). Once Mcm10 is present, Cdc45 and GINS are loaded onto chromatin (Walter et al. 2000; Wohlschlegel et al. 2002; Takayama et al. 2003) and form a Cdc45/Mcm2-7/GINS helicase complex (Pacek et al. 2004; Gambus et al. 2006; Moyer et al. 2006; Pacek et al. 2006). Cyclin- and Dbf4-dependent kinases together with Sld2, Sld3, and Dpb11 in budding yeast (Tanaka et al. 2007; Zegerman et al. 2007) stimulate origin unwinding, which is signified by recruitment of replication protein A to single stranded DNA (Tanaka et al. 1998; Zou et al. 2000). Mcm10, Cdc45, and replication protein A facilitate subsequent loading of DNA polymerase α-primase (pol α) onto chromatin (Mimura et al. 1998; Walter et al. 2000; Ricke et al. 2004; Yang et al. 2005). The association of proliferating cell nuclear antigen, RFC, and replicative DNA polymerases δ and ϵ with the origin completes the replisome (Garg et al. 2005).

A number of interactions have been observed between Mcm10 and proteins found in the pre-replicative complexes and at the replication fork. Mcm10 is a component of active replication complexes in *Xenopus* and budding yeast (Gambus et al. 2006; Pacek et al. 2006) and is associated with chromatin throughout S-phase (Ricke et al. 2004). Mcm10 interacts genetically with Mcm2-7, DNA pol δ and ϵ, origin recognition complex, and Dpb11 (Merchant et al. 1997; Homesley et al. 2000; Izumi et al. 2000; Kawasaki et al. 2000). *In vitro*, interactions of Mcm10 with initiation factor origin recognition complex, Mcm2-7, Cdc45, and Cdc7/Dbf4 have been observed by co-immunoprecipitation from cell extracts (Homesley et al. 2000; Kawasaki et al. 2000; Christensen et al. 2003; Lee et al. 2003). Importantly, Cdc45 and replication protein A cannot load onto chromatin in Mcm10-depleted *Xenopus* egg extracts, preventing DNA unwinding (Wohlschlegel et al. 2002). Thus, the essential role of Mcm10 in initiation links the pre-replicative complexes with origin unwinding.

Several lines of evidence suggest that Mcm10 migrates with the elongating replication fork through association with DNA polymerases and DNA. *Schizosaccharomyces pombe* Mcm10 (SpMcm10) affects chromatin binding and subnuclear distribution of pol α (Gregan et al. 2003; Yang et al. 2005), and *Saccharomyces cerevisiae* Mcm10 (ScMcm10) has been shown to interact with and stabilize the catalytic subunit of pol α *in vivo* (Ricke et al. 2004; Ricke et al. 2006). *In vitro*, SpMcm10 interacts with and stimulates the activity of the catalytic (polymerase) subunit of pol α (Fien et al. 2004) and has been shown to contain primase activity (Fien et al. 2006). Additionally, an interaction between diubiquitinated ScMcm10 and proliferating cell nuclear antigen is essential for replication in budding yeast (Das-Bradoo

et al. 2006). Finally, SpMcm10 binds to single (ss)- and double stranded (ds) DNA *in vitro*, and DNA binding activity is localized in the N-terminal 300 residues of the protein (Fien et al. 2004). The interactions between Mcm10, DNA, and pol α have led to the suggestion that Mcm10 helps to recruit pol α to the replisome and may regulate its activity. Studies in *Xenopus* extracts have demonstrated that when an elongating fork stalls, Mcm10 and DNA polymerases α, δ, and ϵ are uncoupled from the Cdc45/Mcm2-7/GINS helicase (Pacek et al. 2006).

Sequence alignments of Mcm10 from divergent eukaryotes show stretches of consecutive residues that are phylogenetically conserved (Figure 5A), suggesting that these regions may be important to the structure and function of the protein. Mcm10 from Metazoa contains ~100-300 residues not present in the yeast proteins, and conservation from yeast to human is limited to ~200-amino acids in the middle of the protein. Consistent with Mcm10 DNA binding activity, the conserved central region contains an invariant CCCH zinc binding motif (Homesley et al. 2000; Izumi et al. 2000; Cook et al. 2003) and a putative oligonucleotide/oligosaccharide binding fold (Ricke et al. 2006).

The lack of sequence similarity outside of the central region raises a question of whether the function of Mcm10 is conserved from yeast to Metazoa. In the present study we report the first structure-function analysis of vertebrate Mcm10 using the *Xenopus laevis* protein (XMcm10). Limited proteolytic digestion of XMcm10 revealed the protein to be composed of at least three structural domains, an N-terminal domain (NTD) that forms homodimers in solution and highly conserved internal (ID) and C-terminal domains (CTD) that bind to ssDNA, dsDNA, and to the p180 subunit of pol α. Our

results confirm and extend previous work from yeast and suggest that vertebrate Mcm10 contains a CTD not present in the yeast orthologs.

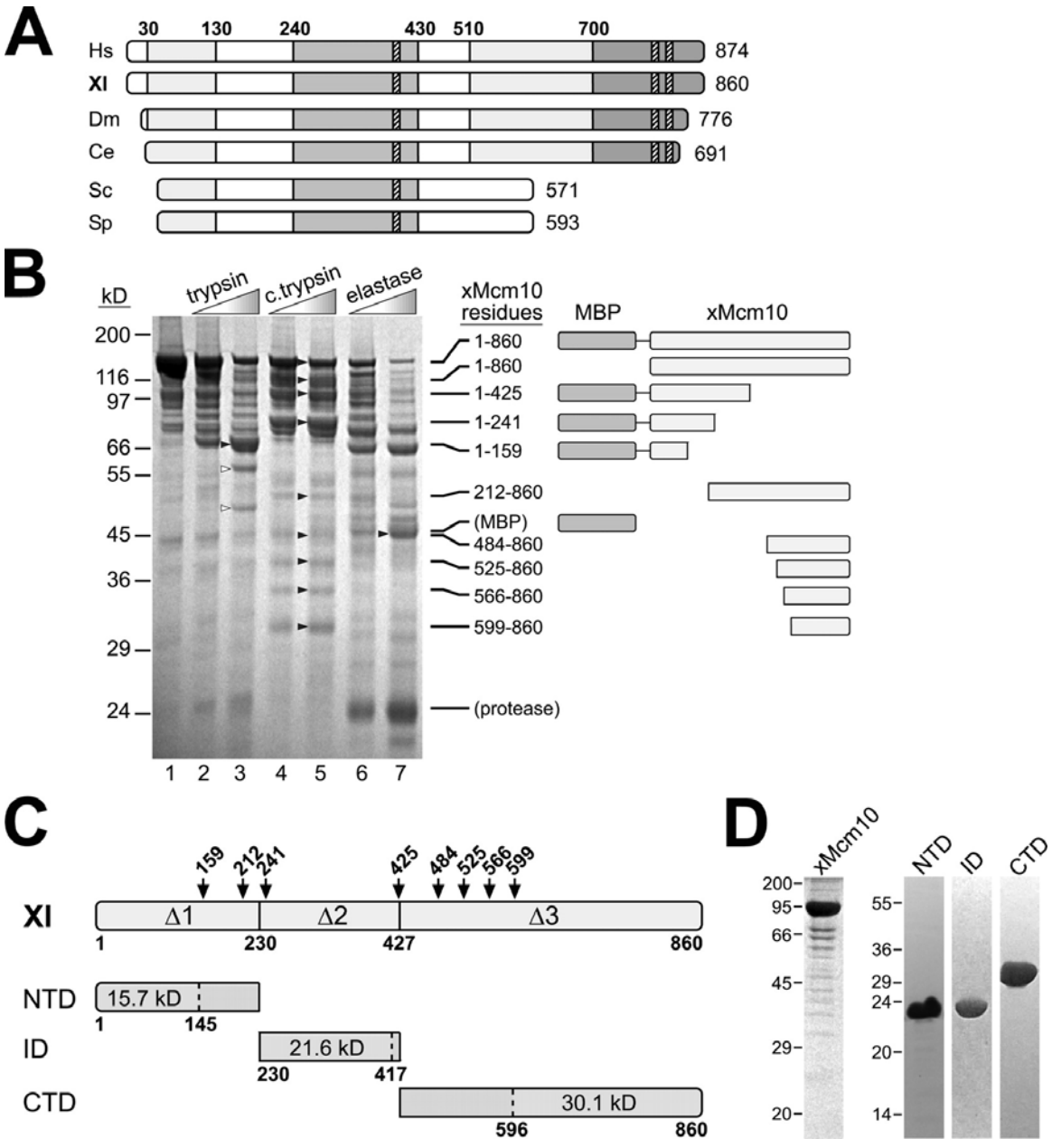
Results

XMcm10 Contains Three Structural Domains

In the current study experiments to characterize the domain architecture of vertebrate Mcm10 were carried out using the *X. laevis* ortholog because of previous investigations of the function of the protein using *Xenopus* egg extracts (Walter et al. 1998; Wohlschlegel et al. 2002). Homology exists in three distinct regions of the protein (Figure 5A, Appendix Figure A1). The internal region (aa 240-430) is highly conserved among all known Mcm10 orthologs, with an overall similarity of 21.3% (39.0% for non-yeast Mcm10). Likewise, the C terminus contains a region of high (~ aa 700-860) and moderate (aa 510-700) similarity among higher eukaryotes. However, this region is not present in the yeast proteins (23.3% similarity for metazoan as compared with 3.6% for all eukaryotes). Moderate sequence similarity also exists at the N terminus (10% similarity for aa 1-130 in non-yeast sequences). This sequence analysis immediately suggested the presence of at least three domains tethered by disordered linkers. Consistent with this, no secondary structure was predicted in regions 130-230 and 575-624 (Appendix Figure A1), and region 130-230 was predicted to be largely disordered.

To experimentally determine the domain organization of Mcm10, the full-length protein was overexpressed in *E. coli* with a cleavable N-terminal MBP tag and a C-terminal His_6 tag. The purified MBP-XMcm10-His_6 protein was subjected to limited proteolytic digestion by trypsin, chymotrypsin, and elastase, and the major proteolytic

Figure 5. Domain architecture of Mcm10. (A) Schematic alignment of Mcm10 sequences from *Homo sapiens* (Hs), *Xenopus laevis* (Xl), *Drosophila melanogaster* (Dm), *Caenorhabditis elegans* (Ce), *Saccharomyces cerevisiae* (Sc), and *Schizosaccharomyces pombe* (Sp). Light and dark gray bars indicate moderate and high sequence conservation, respectively, and hatched boxes represent invariant cysteine/histidine clusters likely involved in zinc coordination. A full sequence alignment is in Appendix Figure A1. (B) Limited proteolytic digestion of XMcm10. 50 pmol MBP-XMcm10 (lane 1) was subjected to proteolysis by trypsin (25 and 100 ng, lanes 2 and 3), chymotrypsin (c.trypsin) (100 and 200 ng, lanes 4-5), and elastase (10 and 25 ng, lanes 6 and 7) and visualized by Coomassie Blue-stained SDS-PAGE. Major proteolytic fragments marked with black arrowheads were unambiguously identified by MALDI-TOF and TOF/TOF tandem mass spectrometry and are shown schematically to the right. Bands marked with white arrowheads contained several co-migrating Mcm10 fragments. The full peptide coverage map used to identify fragment endpoints is in Appendix Figure A2. (C) Three truncation fragments (Δ1, Δ2, Δ3) of XMcm10 were purified and subjected to limited proteolysis to reveal stable domains NTD, ID, and CTD. Proteolytically sensitive sites identified in panel B are highlighted with arrows on top of the full-length protein schematic. Molecular masses and N-terminal sequences shown for each proteolytic fragment were identified by mass spectrometry and Edman degradation. (D) Coomassie-stained SDS-PAGE of purified full-length XMcm10, NTD, ID, and CTD used in this study.


A
30
130
240
430
510
700
Hs
Xl
Dm
Ce
Sc
Sp
874
860
776
691
571
593
B
kD
trypsin
c.trypsin
elastase
xMcm10 residues
MBP
xMcm10
200
116
97
66
55
45
36
29
24
1-860
1-860
1-425
1-241
1-159
212-860
(MBP)
484-860
525-860
566-860
599-860
(protease)
1 2 3 4 5 6 7
C
159
212
241
425
484
525
566
599
Xl
Δ1
Δ2
Δ3
1
230
427
860
NTD
15.7 kD
1
145
ID
21.6 kD
230
417
CTD
30.1 kD
596
860
D
xMcm10
NTD
ID
CTD
200
95
66
45
29
20
55
36
29
24
20
14

fragments were identified by MALDI-TOF MS and MALDI-TOF/TOF tandem MS (Figure 5B). Peptide masses were mapped to the XMcm10 amino acid sequence to define domains (Appendix Figure A2). In most cases the end point regions were defined by peptide ions that were present in the full-length protein but absent in the fragment under study, and in some cases the end point was confirmed with tandem MS on unique peptide(s) that were generated by chymotrypsin cleavage on one side (from limited proteolysis) and trypsin cleavage on the other (from in-gel digestion). Peptides analyzed in this way revealed proteolytic-resistant domains separated by cleavage sites at amino acids 159, 241, 425, 484, 525, 566, and 599 (Figure 5B and Appendix Figure A2).

Using the proteolytically sensitive regions as a guide, three deletion constructs encompassing the entire protein were designed to define the domain boundaries more accurately: $XMcm10^{1-230}$ (Δ1), $XMcm10^{230-427}$ (Δ2), and $XMcm10^{427-860}$ (Δ3). Each of these proteins were expressed in bacteria, purified, and subjected to limited proteolysis by trypsin (Appendix Figure A3). Precise endpoints of tryptic fragments were identified by Edman degradation and MALDI mass spectrometry (Figure 5C). Chymotrypsin, elastase, and endoproteinase-Glu-C digestion was also performed (data not shown). Despite the unique specificities of each protease tested, the resulting cleavage patterns were similar for each Mcm10 deletion mutant. Proteolysis of each deletion mutant revealed the presence of smaller fragments that were resistant to digestion and that were consistent with the cleavage pattern of the full-length protein (Figure 5B) and with regions of sequence conservation (Figure 5A). Cleavage of the C-terminal ends of Δ1 and Δ2 yielded $XMcm10^{1-145}$ and $XMcm10^{230-417}$, respectively. For Δ3, ~170 residues were cleaved from the N terminus, yielding $XMcm10^{596-860}$. The resistance of $XMcm10^{1-145}$,

XMcm10$^{230\text{-}417}$, and XMcm10$^{596\text{-}860}$ to further degradation indicates the presence of stable tertiary folds that stericly preclude protease access to their cleavage sites. To prepare for further characterization, regions 1-145 (NTD), 230-417 (ID), and 596-860 (CTD) were subcloned, overexpressed, and purified (Figure 5D). The anomalous electrophoretic mobility of the NTD can be rationalized on the basis of the predicted pI (4.2) and elongated shape of the protein (see below). The NTD, ID, and CTD were relatively stable to further proteolytic digestion, and circular dichroism spectra confirmed the presence of secondary structure in each domain (data not shown).

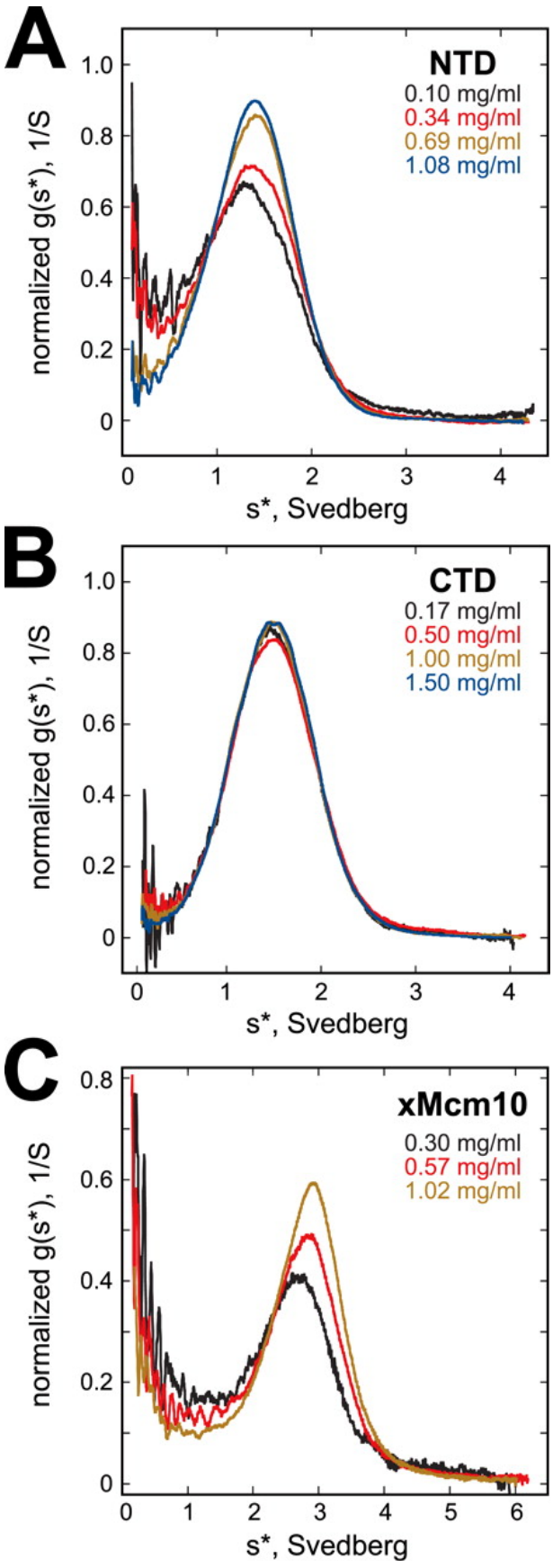
Dimerization of XMcm10-NTD

Purified ScMcm10 and SpMcm10 have been reported to oligomerize in solution (Cook et al. 2003; Lee et al. 2003; Fien et al. 2006), and human Mcm10 was recently reported to form hexameric assemblies (Okorokov et al. 2007). Before a rigorous analysis of XMcm10 oligomerization, we first investigated the hydrodynamic properties of the full-length, NTD, ID, and CTD proteins by gel filtration chromatography (Appendix Figure A4). The elution volumes of full-length and NTD proteins were considerably less than expected for globular, monomeric proteins. Similarly, the CTD showed a modest decrease in retention volume as compared with that of a 30-kDa protein standard. The elution profile of the ID, on the other hand, corresponded exactly to that of a 22-kDa protein, indicating that this domain does not self-associate. These results raised the question of whether XMcm10 oligomerizes in solution or whether the shape of the protein significantly deviates from a globular fold.

The oligomeric states of the NTD, CTD, and full-length proteins were determined using sedimentation velocity experiments (Figure 6). Figure 6A shows an overlay of the normalized g(s*) sedimentation coefficient distributions for four concentrations of the NTD. The distributions shift to the right with increasing concentration, indicating reversible self-association. The best fit to the data were obtained using a monomer-dimer equilibrium model. The sedimentation coefficient for the monomer could not be accurately determined due to the fact that the protein is predominantly dimeric over the concentration range tested. Thus, the sedimentation coefficient ratio s(dimer)/s(monomer) was fixed at 1.45, which is the value predicted for a monomer-dimer system (Garcia de la Torre et al. 1981). The best fit parameters are $s_{20,w}$ (monomer) = 1.22 S, $s_{20,w}$ (dimer) = 1.77 S, a dissociation constant of K_d = 3.1 μM, and a root mean square error of 0.0048 mg/ml. The corrected sedimentation coefficients of the monomer and dimer can be used to calculate frictional ratios, f/f_0, of 1.8 and 2.0, respectively, indicating that the NTD is highly asymmetric.

The normalized g(s*) profiles for the CTD superimpose over the concentration range tested (0.17-1.5 mg/ml), indicating that the system does not undergo reversible association under these conditions. The molecular weight obtained from a global fit of the data to a single species model is 31.0 kDa, which agrees closely with the predicted monomeric value of 30.1 kDa. The frictional ratio (f/f_0) of 1.89 indicates that CTD is also quite asymmetric, consistent with its gel filtration behavior. Figure 5C shows the normalized g(s*) distributions for the full-length enzyme. Like NTD, the distributions shift to the right with increasing concentration, indicating mass-action association. In this case, the presence of lower- and higher-S contaminants precludes further analysis of

Figure 6. Self-association of XMcm10. Shown are overlays of normalized g(s*) plots from sedimentation velocity experiments at different concentrations of XMcm10-NTD (A), CTD (B), and full-length enzyme (C). NTD and CTD were prepared in 20 mM Tris, pH 7.5, 100 mM NaCl, 3.5 mM BME, and 5% glycerol, and full-length enzyme was prepared in 20 mM Tris, pH 7.5, 500 mm NaCl, 1 mM DTT, 5% glycerol. Conditions: rotor speed, 55,000 rpm; temperature, 20 °C; interference optics.


A
NTD
0.10 mg/ml
0.34 mg/ml
0.69 mg/ml
1.08 mg/ml
normalized g(s*), 1/S
s*, Svedberg
B
CTD
0.17 mg/ml
0.50 mg/ml
1.00 mg/ml
1.50 mg/ml
normalized g(s*), 1/S
s*, Svedberg
C
xMcm10
0.30 mg/ml
0.57 mg/ml
1.02 mg/ml
normalized g(s*), 1/S
s*, Svedberg

these data. However, the limiting sedimentation coefficient of ~2.6 S at low concentration indicates that XMcm10 is predominantly monomeric at low concentrations with $f/f_0 \sim 2.2$. Assuming an alternative model where the $s = 2.6$ S species is a dimer yields an unreasonably high $f/f_0 \sim 3.5$.

Zinc-dependent Stability of XMcm10-ID and CTD

Sequence alignments show clusters of highly invariant cysteine and histidine residues in both the ID and CTD (Figure 7A), suggesting that these domains contain zinc binding motifs. Strong evidence has been provided for the presence of a zinc motif in ScMcm10 internal region (Cook et al. 2003), although zinc binding by the CTD has not yet been reported. To verify the presence and determine the stoichiometry of Zn^{2+} in XMcm10 domains, we analyzed each of the domains by GFAA spectroscopy. Molar ratios of Zn^{2+}/XMcm10 for the NTD, ID, and CTD were determined to be 0.16, 1.3 ± 0.3, and 1.8 ± 0.5, respectively (Table 1). As a positive control, 3-methyladenine DNA glycosylase I (TAG), which has been shown previously to contain 1 Zn^{2+}/molecule (Kwon et al. 2003; Metz et al. 2007), was analyzed by GFAA and returned a value of 0.98 Zn^{2+}/TAG. We, therefore, conclude that the NTD, CTD, and ID contain 0, 1, and 2 Zn^{2+} ions, respectively. In support of the GFAA data, x-ray fluorescence emission spectra of XMcm10-ID single crystals, which were grown in the absence of Zn^{2+} in the crystallization buffer, revealed a strong peak at 9.6 keV corresponding to the Zn^{2+} absorption edge (data not shown).

The importance of bound zinc on the tertiary folding of the ID and CTD was investigated by limited proteolysis protection assays. The ID and CTD were subjected to

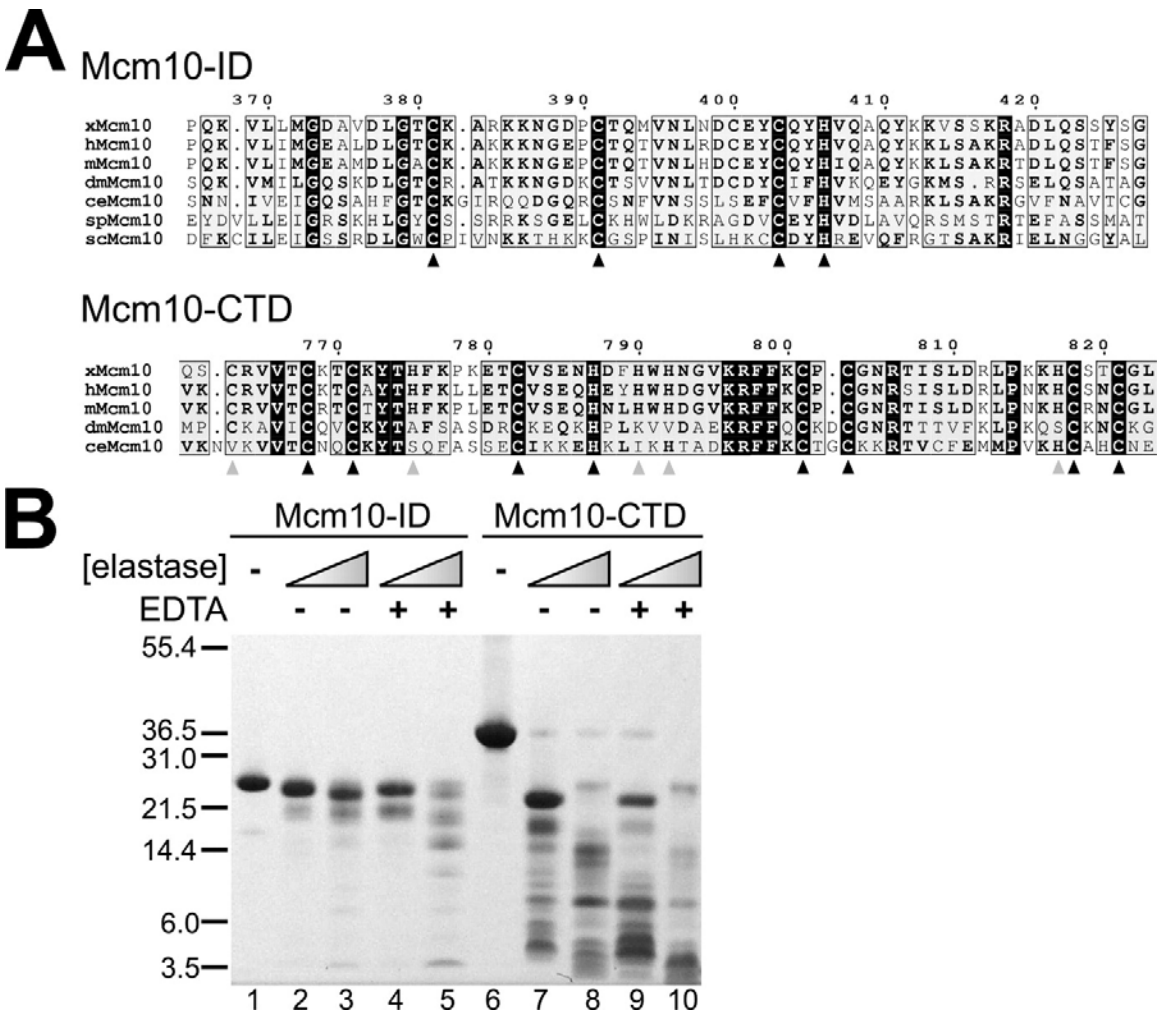


Figure 7. Effect of EDTA on the stability of XMcm10-ID and -CTD. (A) Sequence alignment of ID and CTD regions containing invariant (black triangles) and conserved (gray triangles) cysteine and histidine residues likely involved in Zn^{2+} coordination. (B) SDS-PAGE of elastase-catalyzed proteolysis of ID (lanes 1-5) and CTD (lanes 6-10) in the presence (lanes 4, 5, 9, and 10) and absence (lanes 2, 3, 7, 8) of 10 mM EDTA. 100 pmol of each Mcm10 domain was incubated with 10 ng (lanes 2, 4, 7, and 9) and 100 ng (lanes 3, 5, 8, and 10) elastase.

proteolysis by elastase in the presence and absence of EDTA, a known Zn^{2+} chelator. Both domains were more readily degraded in the presence of EDTA (Figure 7B), suggesting that in the absence of bound Zn^{2+}, the ID and CTD were at least partially unfolded and, thus, more susceptible to protease cleavage. Similarly, when the ID and CTD were incubated at room temperature for 10 days in the presence or absence of EDTA, spontaneous degradation was increased in the presence of EDTA (Appendix Figure A5). These results suggest that the zinc motifs in XMcm10-ID and -CTD play a key role in maintaining the overall structural integrity of these domains.

Table 1. Molar equivalents of Zn^{2+} in XMcm10 domains determined by atomic absorption spectroscopy

Protein	Zn^{2+} / XMcm10
XMcm10-NTD	0.16
XMcm10-ID	1.3 ± 0.3
XMcm10-CTD	1.8 ± 0.5
TAG (control) [a]	0.98

[a] (Kwon et al. 2003; Metz et al. 2007)

XMcm10-ID and CTD Are DNA Binding Domains

To quantitatively characterize the DNA binding activity of purified XMcm10, the change in fluorescence anisotropy was monitored as the protein was added to a fluorescein-labeled 25-mer oligonucleotide (Figure 8). Binding isotherms for MBP-XMcm10-His_6 show that the full-length *Xenopus* protein bound to both ssDNA and dsDNA with the same affinity ($K_d \sim 0.1$ μM) (Figure 8A, Table 2). To determine whether Mcm10 might bind to the replication fork at the ss/dsDNA junction, a forked substrate containing both ssDNA and dsDNA regions was also tested and did not show a difference

in binding affinity (K_d = 0.08 ± 0.06 μM) compared with ssDNA and dsDNA (Table 2). Interestingly, in the presence of EDTA, binding of XMcm10 to dsDNA was abolished, whereas the affinity for ssDNA remained unchanged (Figure 8A, Table 2). The overall anisotropy change for ssDNA binding was different between EDTA and non-EDTA titrations, indicating that a change in the tumbling rate of the complex occurred, likely as a result of EDTA-induced local unfolding of the zinc motifs (Figure 7). These results establish that zinc-dependent structural integrity of XMcm10 is important for the dsDNA binding activity.

Table 2. Dissociation constants for DNA binding [a]

	ssDNA	dsDNA	Fork [b]	Bubble [c]
XMcm10	0.12 ± 0.02	0.09 ± 0.03	0.08 ± 0.06	nd
XMcm10 + 10 mM EDTA	0.14 ± 0.04	≤ 300	nd	nd
XMcm10-NTD	≤ 300	≤ 300	nd	nd
XMcm10-ID	3.39 ± 0.49	7.83 ± 1.44	3.09 ± 0.99	5.21 ± 1.86
XMcm10-CTD	1.41 ± 0.24	2.21 ± 0.20	2.67 ± 0.34	4.77 ± 2.57

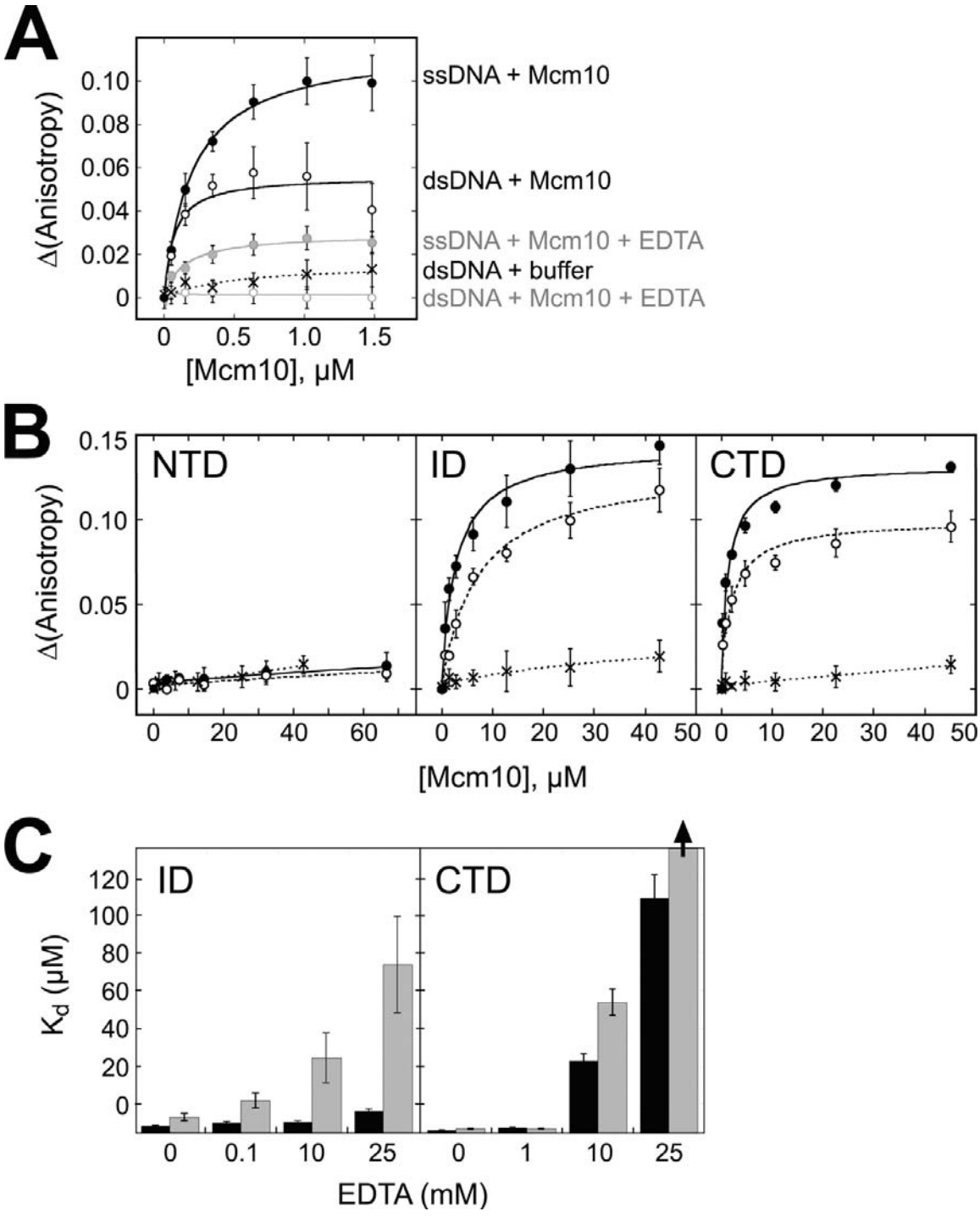
[a] K_d for XMcm10 binding to 25mer oligonucleotides are reported in μM protein. nd, not determined

[b] Forked DNA = $(dsDNA)_{25}$-2x$(ssDNA)_{25}$ for full length and $(dsDNA)_{10}$-2x$(ssDNA)_{15}$

[c] Bubble DNA = $(dsDNA)_{10}$-2x$(ssDNA)_{15}$-$(dsDNA)_{10}$

Binding of DNA to the NTD, ID, and CTD was then measured to determine the DNA binding domain of XMcm10. No anisotropy change was observed in the presence of the NTD, indicating that this domain does not interact with DNA (Figure 8B). Unexpectedly, both the ID and CTD showed robust binding to both ssDNA and dsDNA (Figure 8B). The affinity of each domain for DNA was roughly the same and was an order of magnitude less than that of the full-length protein (Table 2). Unlike full-length XMcm10, the affinity of each domain for ssDNA was ~2-fold greater than for dsDNA.

Figure 8. DNA binding of XMcm10. Binding was monitored as a change in fluorescence anisotropy as full-length (A) and isolated domain (B and C) proteins were titrated into a solution containing fluorescently labeled DNA. Error bars represent the S.D. from the average values from three independent measurements. (A) Binding isotherms for full-length MBP-XMcm10-His6 binding to ssDNA (filled symbols) and dsDNA (open symbols) in the absence (black) and presence (gray) of EDTA. A control in which buffer without protein was added to the DNA is shown as black X's. (B) Binding curves for each XMcm10 domain against ssDNA (closed circles) and dsDNA (open circles), and for buffer-only controls (Xs). (C) The dissociation constants (K_d) for XMcm10-ID and -CTD binding to ssDNA (black bars) and dsDNA (gray bars) derived from the anisotropy data are plotted as a function of EDTA concentration. The K_d for XMcm10-ID/dsDNA binding in 25 mM EDTA is $\geq$300 μM, the limit of detection for this assay.

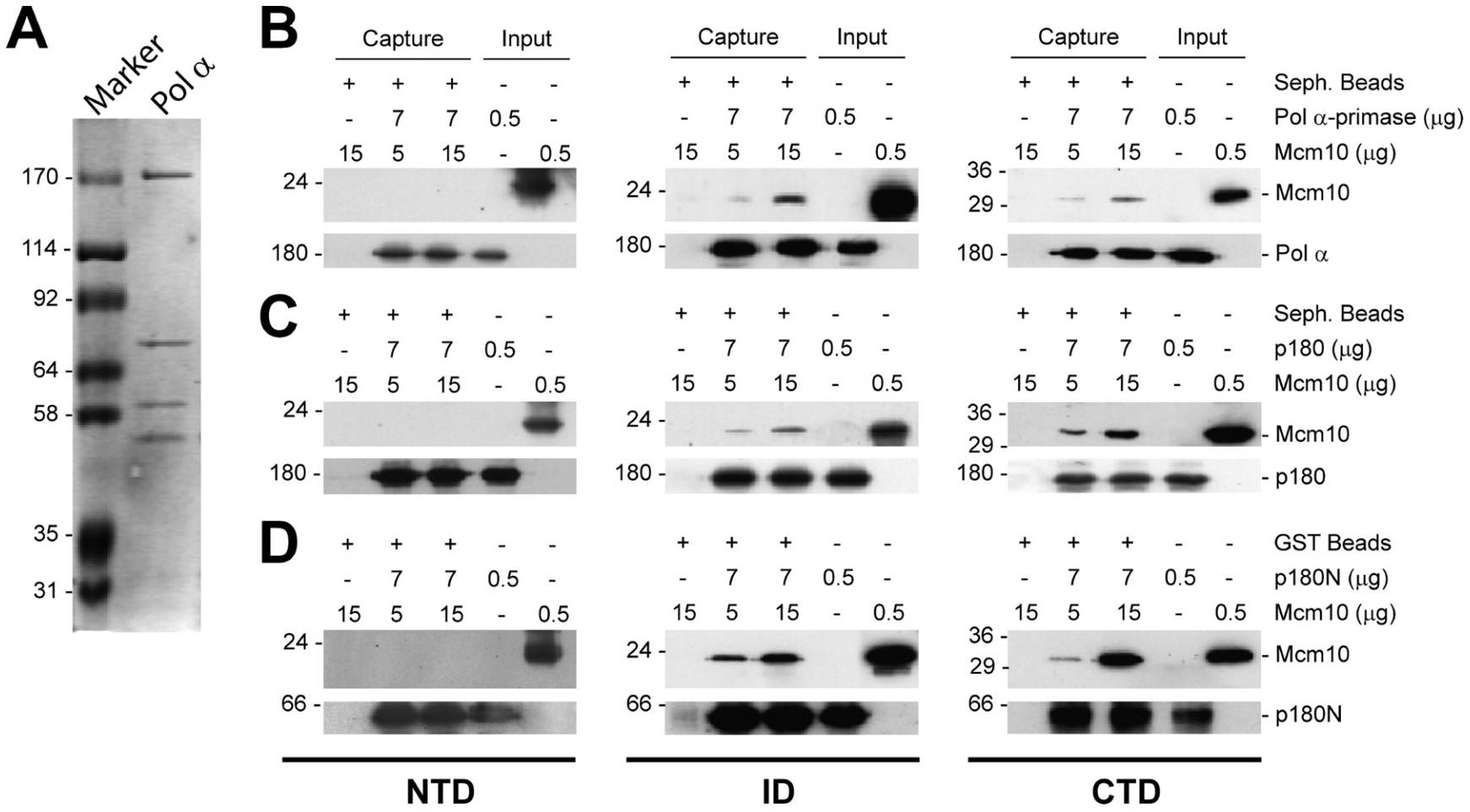

A
Δ(Anisotropy)
0.10
0.08
0.06
0.04
0.02
0
0
0.5
1.0
1.5
[Mcm10], μM
ssDNA + Mcm10
dsDNA + Mcm10
ssDNA + Mcm10 + EDTA
dsDNA + buffer
dsDNA + Mcm10 + EDTA
B
Δ(Anisotropy)
0.15
0.10
0.05
0
NTD
ID
CTD
0
20
40
60
0
10
20
30
40
50
0
10
20
30
40
50
[Mcm10], μM
C
K_d (μM)
120
100
80
60
40
20
0
ID
CTD
0
0.1
10
25
0
1
10
25
EDTA (mM)

To test the effect of the Zn^{2+} motifs, binding experiments for each domain were again carried out in the presence of EDTA (Figure 8C). Both XMcm10-ID and -CTD exhibited a dramatic decrease in dsDNA binding affinity as a function of increasing EDTA concentration, whereas the ssDNA affinity was only moderately affected under the same conditions (Figure 8C). Interestingly, EDTA had a greater affect on ssDNA binding to the CTD than the ID, suggesting that ssDNA is able to bind to the ID in the absence of a folded zinc motif.

XMcm10 Binding to DNA Polymerase α*-Primase Is Localized to the ID and CTD*

We investigated whether vertebrate Mcm10 can undergo direct, physical interactions with pol α, and if so, these interactions can be mapped with the XMcm10 domains. Because purified recombinant human pol α has been shown to substitute functionally for the *X. laevis* protein in *in vitro Xenopus* replication assays (Michael et al. 2000), human pol α was chosen for these experiments (Figure 9A). The first experiment examined the ability of the purified four-subunit human pol α-primase complex immobilized on beads to capture His-tagged XMcm10 domains from solution. After incubation with purified XMcm10-NTD, ID, or CTD and extensive washing, XMcm10 domains remaining bound to the beads were detected by denaturing gel electrophoresis and anti-His Western blot. Figure 9B shows the results of the pol α-Mcm10 affinity capture, in which both the ID and CTD, but not the NTD, bound to the polymerase complex. The experiment was repeated using only the purified catalytic pol α-p180 subunit in the absence of p48, p58, and p68. Again the NTD was not detected in the bound fraction, and both the ID and CTD bound to p180 (Figure 9C). This result

Figure 9. Binding of XMcm10 to the p180 subunit of DNA pol α. (A) Coomassie Blue-stained SDS-PAGE gel of purified pol α. B-D, affinity capture experiments between XMcm10-NTD (left panels), -ID (center panels), or -CTD (right panels) and pol α (B), p180 (C), or p180N (D). Amounts of protein added to each binding reaction are shown above the Western blots. (B) The intact pol α complex was mixed with XMcm10 domains NTD, ID, or CTD as indicated and immunoprecipitated on Sepharose beads coupled to SJK132-20 antibodies against the p180 subunit as indicated. Bound XMcm10 domains were detected by Western blotting with an α-His antibody. (C) The purified catalytic p180 subunit of pol α was mixed with XMcm10 domains and immunoprecipitated as in B.


A
Marker
Pol α
170
114
92
64
58
35
31
B
Capture
Input
Seph. Beads
Pol α-primase (μg)
Mcm10 (μg)
Mcm10
Pol α
C
p180 (μg)
p180
D
GST Beads
p180N (μg)
p180N
NTD
ID
CTD

demonstrates that the p180 subunit is sufficient to bind XMcm10-ID and CTD.

We next sought to map the specific Mcm10-binding region of p180. The p180 subunit has a modular organization with an ~300-residue N-terminal region dispensable for polymerase activity, an extended core region containing the conserved polymerase motifs, and a C-terminal region that complexes with the other subunits (Mizuno et al. 1999). Only the N-terminal region of p180 binds to SV40 T antigen, an interaction essential for viral DNA replication (Dornreiter et al. 1993). Based on this information, an N-terminal construct encompassing p180 residues 1-323 (p180N) was tested. GST-tagged p180N immobilized on glutathione-Sepharose was able to capture both the ID and CTD, but not the NTD, consistent with the pol α-primase and p180 pulldown assays (Figure 9D). Thus, p180N is sufficient for Mcm10 interaction. These results also show that as for binding DNA, the ID and CTD function in a coordinated manner.

XMcm10 Does Not Contain Primase Activity

Based on the recent report that SpMcm10 contains primase activity (Fien et al. 2006), we examined the ability of full-length XMcm10 to synthesize an oligoribonucleotide in the presence of a DNA template. Purified XMcm10 that contained no MBP tag (Figure 5D) was incubated with dT_{50} template and [α-^{32}P]ATP, and product RNA was visualized by denaturing PAGE. No radiolabeled products were apparent when compared with a no-enzyme control reaction (Figure 10). Under identical conditions, pol α-primase showed robust, concentration-dependent formation of oligoribonucleotides ~12 nucleotides in length. This result indicates that a purified preparation of XMcm10 is not capable of priming DNA.

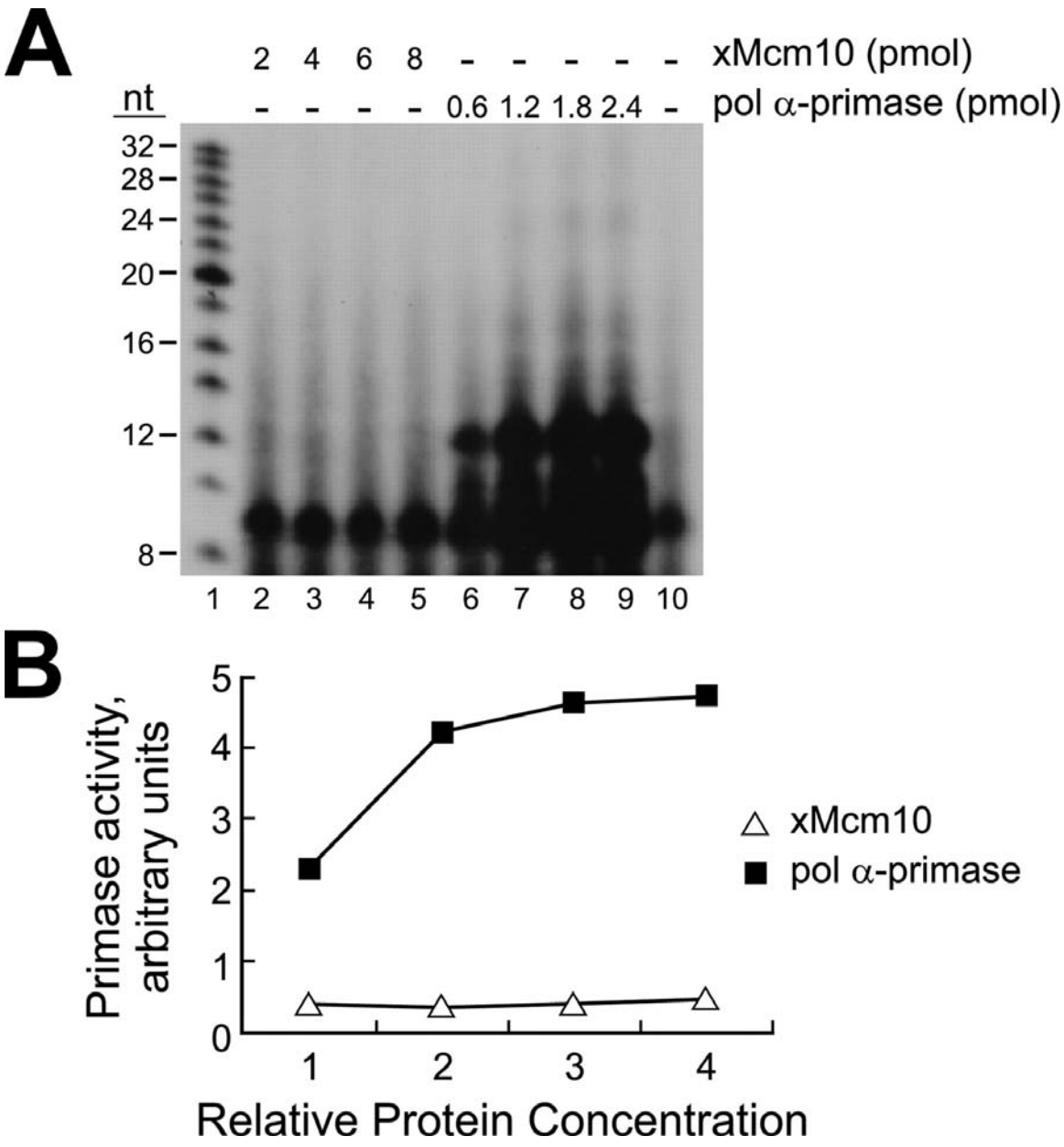


Figure 10. XMcm10 does not contain primase activity. (A) oligoribonucleotide synthesis was assayed in reaction mixtures containing dT50 template, [α-32P]ATP, and increasing amounts of XMcm10 (lanes 2-5) or pol α-primase (lanes 6-9). Lane 10, negative control lacking XMcm10 and pol α-primase. Radiolabeled products were analyzed by electrophoresis on a 25% denaturing polyacrylamide gel containing 7 M urea. (B) quantitation of the autoradiogram shown in A. Primase activity is expressed in arbitrary units, with the reaction containing no protein set to zero. Relative protein concentration corresponds to 0.2, 0.4, 0.6 and 0.8 μM XMcm10 and 0.06, 0.12, 0.18 and 0.24 μM pol α.

Discussion

Modular Architecture of Mcm10

The present work provides new insight into the role of Mcm10 in initiation and elongation complexes by carrying out the first structure-function analysis of the protein. We have determined using limited proteolysis that purified preparations of XMcm10 contain at least three structural domains located from residues 1-145 (NTD), 230-417 (ID), and 596-860 (CTD) (Figure 11). The extreme proteolytic sensitivity of regions 146-230 and 418-596 suggests that these are exposed flexible linkers connecting each independently folded globular domain. It is likely that these flexible regions become more structured or protected from proteolytic cleavage when Mcm10 is part of the larger multiprotein replisome assembly. Nevertheless, the present work suggests that Mcm10 is at least able to adopt multiple conformations in which each globular domain can move relative to the other two. Such a flexible protein architecture would be necessary for the multiple protein and DNA transactions at an inherently dynamic replication fork. Indeed, many replication proteins have evolved modular architectures with distinct domains that are able to act independently or cooperatively to perform a common task (Stauffer et al. 2004; Fanning et al. 2006). For example, separate structural domains often provide multiple binding sites that increase the affinity for one ligand or that enable the protein to contact multiple ligands in a concerted or sequential fashion (Arunkumar et al. 2003).

Structural Features of Mcm10-ID and -CTD

Motifs predicted within the ID and CTD provide a rationale for their interactions with DNA and pol α (Figure 11). The protein structure prediction Protein Homology/analogY Recognition Engine (PHYRE) program (Kelley et al. 2000) and

manual inspection of the XMcm10-CTD primary sequence identified two putative Zn^{2+} binding motifs (aa 692-755 and 768-821) and a three-helical bundle from the winged helix superfamily (aa 692-755) (Appendix Figure A1). These motifs were not identified in yeast Mcm10 proteins. Previously identified motifs in the conserved ID were also found by this method, including an oligonucleotide/oligosaccharide binding fold (aa 286-346) and zinc motif (391-406) (Homesley et al. 2000; Izumi et al. 2000; Ricke et al. 2006). Consistent with the ability of the ID and CTD to bind both DNA and pol α, oligonucleotide/oligosaccharide binding folds, winged helix bundles, and zinc motifs have each been shown to mediate protein-protein interactions in addition to their role in nucleic acid recognition (Leon et al. 2000; Mer et al. 2000; Stauffer et al. 2004).

The zinc binding motifs are essential to the structure and function of Mcm10. Mutations in the putative CCCH-type zinc finger within the conserved ID have been shown to disrupt the association of ScMcm10 with chromatin (Homesley et al. 2000), to cause growth defects in yeast, and to disrupt the NMR chemical shift dispersion of purified ScMcm10 (Cook et al. 2003). Our atomic absorption data show conclusively that 1 molar eq of zinc is present in the ID and reveal two additional zinc atoms bound to the CTD (Table 1). The effect of Zn^{2+} chelation on Mcm10 DNA binding activity and protein stability (Figures 7B and 8A and 8C, Table 2) helps to explain the dissociation of Mcm10 from chromatin in the *S. cerevisiae mcm10-43* (C320Y in the ID) mutant (Solomon et al. 1992; Homesley et al. 2000).

The arrangement of the invariant Cys/His clusters in the XMcm10-CTD into a $\mathbf{C}X_2\mathbf{C}X_{10}\mathbf{C}X_4\mathbf{H}X_{13}\mathbf{C}X\mathbf{C}X_{14}\mathbf{C}X_2\mathbf{C}$ consensus sequence (Figure 7A) raises several possibilities for the precise role of the CTD zinc motifs. On one hand, the sequences of

each CCCH or CCCC cluster do not deviate significantly from the classical DNA sequence-specific **C**X_2**C**X_{12}**H**X_3**H** zinc finger (Klug et al. 1995). However, there was no difference in binding affinities between either the ID or CTD tested against three different oligonucleotide sequences (data not shown), suggesting that Mcm10 does not recognize DNA in a sequence-specific manner. On the other hand, the two tandem cysteine-rich clusters in the CTD are remarkably similar in sequence to LIM domains and RING finger motifs, which provide protein-binding interfaces important for a variety of

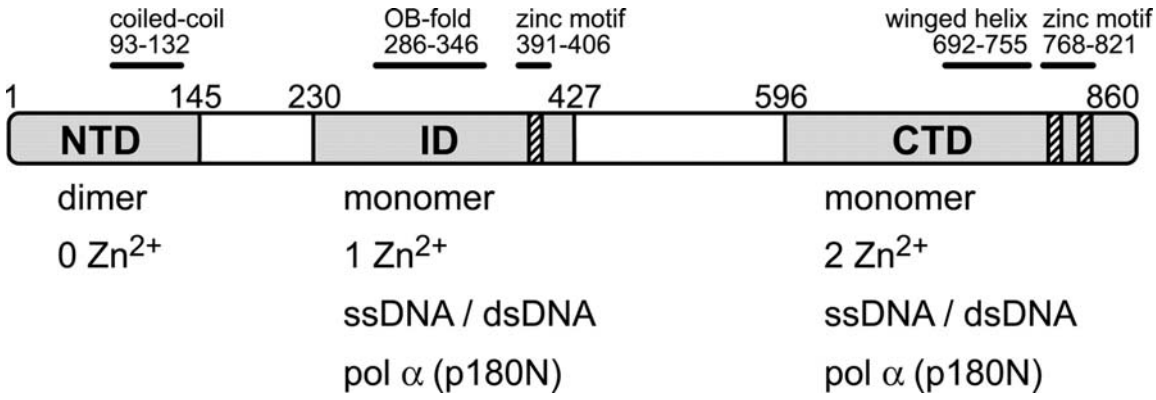


Figure 11. Vertebrate Mcm10. The schematic summarizes the domain organization and functional regions of XMcm10 identified in this study. The NTD, ID, and CTD are shaded gray, and conserved cysteine/histidine clusters predicted to chelate Zn^{2+} are shown as cross-hatched strips. Predicted structural motifs are shown as black bars above the protein. Listed below each domain are the oligomerization states, number of zinc ions bound, and binding partners.

cellular functions (Borden 2000; Kadrmas et al. 2004). It is noteworthy that the CTD zinc motif is immediately adjacent in the primary sequence to a putative winged helical bundle, which was predicted based on its similarity to that of the SCF ubiquitin ligase (Murzin et al. 1995). The globular assembly formed from the RING protein Rbx1, and the winged helical of Cul1 in the SCF complex is an interaction integral to the cullin-

RING ubiquitin ligase machinery (Zheng et al. 2002; Petroski et al. 2005). Thus, the zinc motif in XMcm10-CTD might stabilize the protein fold through a winged helical-RING interaction.

Structural and Functional Differences between Vertebrate and Yeast Mcm10

The lack of sequence conservation within the C-terminal region helps to reconcile differences in DNA binding activities of SpMcm10 and XMcm10. The DNA binding affinity for SpMcm10 N-terminal (1-303) and C-terminal (295-593) fragments, which are truncated between the putative oligonucleotide/oligosaccharide binding fold and zinc finger of the ID, was the same as that of the full-length protein (Fien et al. 2004). Full-length XMcm10, on the other hand, bound to DNA with 10-fold greater affinity than XMcm10-ID or -CTD alone (Table 2). Additionally, SpMcm10 exhibited a 20-fold preference for ssDNA over dsDNA (Fien et al. 2004), whereas XMcm10 bound to ssDNA and dsDNA with the same affinity. Although the domain structure of yeast Mcm10 is unknown, these results are consistent with a second DNA binding domain in vertebrate XMcm10-CTD that is not present in the yeast proteins.

The sequence divergence and different DNA binding activities between vertebrate and yeast Mcm10 suggest that these proteins have evolved subtly different functions. An additional DNA binding domain may have evolved in response to the greater complexity of the genome and the lack of specific nucleotide sequences at origins of replication. Alternatively, the additional DNA and pol α binding domain and the lack of detectable primase activity in XMcm10 suggest that vertebrate Mcm10 evolved a means to recruit pol α-primase in lieu of itself priming DNA. Structural studies will be required to

determine whether the ID and CTD are classical DNA binding domains, or if they form versatile structural scaffolds commonly observed in replication proteins (Shamoo et al. 1995; Mizuno et al. 1999; Lee et al. 2000; Bochkareva et al. 2002).

Perspectives on the Mcm10 Role at the Replication Fork

Structural arrangement of Mcm10 domains together with their macromolecular interactions provides insight into Mcm10 function. Our results are consistent with the notion that Mcm10 recruits pol α to origins of replication (Fien et al. 2004; Ricke et al. 2004; Ricke et al. 2006). With each of two separate domains encompassing DNA and pol α binding activities, Mcm10 might mediate a hand-off mechanism between pol α and DNA. Domain rearrangement to facilitate a handoff between replication proteins and DNA has been proposed for SV40 T antigen-mediated replication protein A loading onto DNA (Jiang et al. 2006).

Evidence is provided here for NTD-mediated dimerization of vertebrate Mcm10 (Figure 6). Analytical ultracentrifugation clearly showed dimerization of the NTD with a K_d of ~3.1 μM. The full-length enzyme is predominantly monomeric at low concentration but also self-associates, and by analogy to NTD it is likely also a monomer-dimer system. We observed that the NTD of mammalian and yeast Mcm10 contains a predicted coiled-coil (Appendix Figure A1), a highly asymmetric motif that would explain protein dimerization and the anomalously short gel filtration retention times of Mcm10 constructs containing the NTD. Indeed, frictional ratios calculated from the sedimentation data are indicative of a highly asymmetric protein. These data are consistent with glycerol gradient sedimentation results showing SpMcm10 dimerization

and suggesting an elongated ScMcm10 structure (Lee et al. 2003) and are intriguing in light of the recent report that human Mcm10 forms a globular homohexameric assembly (Okorokov et al. 2007).

NTD-mediated dimerization raises the interesting possibility that Mcm10 interacts with both leading and lagging strand polymerases at a replication fork. Direct physical interactions between Mcm10 and pol α have now been observed in ScMcm10, SpMcm10, and XMcm10 (Fien et al. 2004; Ricke et al. 2006), and genetic studies raise the possibility that Mcm10 also interacts with replicative polymerases δ and ϵ. The coiled-coil interaction would orient both subunits of the Mcm10 dimer in the same direction and consequently provide the polarity needed for the individual subunits to associate with co-directional leading and lagging strands.

The fact that XMcm10 did not preferentially bind to forked DNA substrates (Table 2) suggests that Mcm10 does not reside directly at the fork but, rather, some distance behind the unwinding DNA. On the other hand, interactions between Mcm10 and Mcm2-7 subunits have been observed by yeast two-hybrid (Izumi et al. 2000). Our data suggest that Mcm10 travels with pol α by association with the N-terminal end of p180. This region is dispensable for polymerase activity of p180 (Mizuno et al. 1999), suggesting that Mcm10 is capable of interacting with pol α during DNA synthesis. The p68 subunit of pol α has been reported to interact with SV40 T antigen, tethering pol α to the viral replication fork (Collins et al. 1993; Ott et al. 2002), but p68 did not interact with XMcm10 (data not shown). In addition, we were unable to detect a direct interaction between XCdc45 and pol α or between XMcm10 and XCdc45 (data not shown). In summary, the structural studies begun here provide a framework for future studies to

elucidate the spatial arrangement of vertebrate Mcm10 and its binding partners and to develop a model for the action of these proteins within the replisome.

Experimental Procedures

Cloning, Expression, and Purification of XMcm10

The cDNAs encoding full-length XMcm10 (FL, 1-860) and deletion fragments 1-145, 1-230, 230-427, 427-860, and 596-860 were PCR-amplified from a previously described plasmid encoding a GST-XMcm10 fusion (Wohlschlegel et al. 2002). The FL-XMcm10 PCR product was ligated into a modified pMAL-c2x vector (New England Biolabs) to generate an maltose-binding protein (MBP)-XMcm10-His_6 fusion protein, and XMcm10 fragments were ligated into a modified pET-32a plasmid (Novagen) to generate N-terminal thioredoxin (Trx)-His_6 fusion proteins. Protein was overexpressed in *Escherichia coli* BL21(DE3) cells in Luria-Bertani medium supplemented with 100 μg/ml ampicillin, 5 μM $ZnSO_4$, and 0.5 mM IPTG. Proteins were overexpressed at 22 °C for 4 h (FL) or at 16 °C for 16 h (fragments). The cells were resuspended in 50 mM Tris-HCl, pH 7.5, 500 mM NaCl, 10% glycerol, and lysed under pressure (25,000 psi) using an EmulsiFlex-C3 homogenizer (Avestin, Inc.). FL-XMcm10 was purified by tandem Ni-NTA and amylose affinity chromatography, cleavage of the MBP tag, and SP-Sepharose cation exchange. Protein was concentrated and stored in FL buffer (20 mM Tris-HCl, pH 7.5, 500 mM NaCl, 1 mM DTT, and 5% glycerol). XMcm10 fragments were purified by Ni-NTA affinity chromatography followed by cleavage of the Trx-His_6 tag. The cleaved proteins were further purified by cation exchange (fragments 230-427, 427-860, 596-860) or anion exchange (1-145 and 1-230) chromatography followed by gel filtration on a

Superdex™ 200 preparative column (GE Healthcare) that had been equilibrated with S-200 buffer (20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 5% glycerol, 4 mM BME. Structural integrity of fragment proteins was verified by circular dichroism spectroscopy.

Limited Proteolysis and Fragment Identification

Proteolysis experiments were carried out in S-200 buffer, in which 5-20 μM XMcm10 was incubated with 1-200 ng of protease (trypsin, α-chymotrypsin, elastase, or endoproteinase-Glu-C) in a 10-μL reaction at 37 °C for 30 min. Proteolysis protection reactions contained 10 mM EDTA. Proteases were inactivated by adding 10 μL of SDS-PAGE sample buffer (63 mM Tris-HCl, pH 6.8, 700 mM BME, 2% w/v SDS, 0.03% w/v bromophenol blue, and 10% glycerol) and heating for 5 min at 95 °C. Proteolytic fragments were separated by SDS-PAGE and visualized by Coomassie Blue staining.

Proteolytic fragments from MBP-XMcm10-His_6 were excised from the SDS-PAGE gel and subjected to in-gel digestion with Trypsin Gold (Promega) using standard procedures (Anumanthan et al. 2006). The resulting peptides were analyzed by matrix-assisted laser desorption/ionization, time-of-flight mass spectrometry (MALDI-TOF MS) and TOF/TOF tandem MS using a Voyager 4700 (Applied Biosciences, Framingham MA). Peptide ion masses (M+H) were accurate to within 20 ppm after internal calibration using either trypsin autolytic peptides or XMcm10-derived peptides confirmed by TOF/TOF MS.

Molecular masses of XMcm10 domains resulting from proteolysis of deletion mutants Δ1, Δ2, and Δ3 were obtained by MALDI-TOF mass spectrometry of the proteolysis reactions before SDS-PAGE. Reactions were concentrated in 0.1%

trifluoroacetic acid, mixed with 3 μL of saturated sinapinic acid in 60:40 (v/v) acetonitrile:1% trifluoroacetic acid/distilled H_2O, and 1 μL was deposited onto a gold 100-well plate. Mass spectra were acquired on a Perceptive Biosystems Voyager Elite TOF spectrometer equipped with a laser desorption ionization source and an extended-path ion reflector. Protein standards from Sigma (MSCAL1-1KT) were used for mass calibration. For N-terminal sequencing of XMcm10 domains, intact proteolytic fragment proteins were transferred from SDS gel to a PVDF membrane, stained with Ponceau S, extracted from the membrane, and subjected to Edman degradation chemistry using an Applied Biosystems Model 492HT Protein/Peptide Sequencer equipped with an on-line phenylthiohydantoin-derivative analyzer.

Zinc Quantitation

Quantitative analysis of zinc bound to XMcm10 was performed using graphite furnace atomic absorption (GFAA) spectroscopy. Analyses were performed using a PerkinElmer Life Sciences HGA SIMAA 6000 graphite furnace equipped with an AAnalyst 800 GFAA/FLAA spectrophotometer. XMcm10 domains were quantified by absorbance spectroscopy at 280 nm using extinction coefficients of 0.092 (NTD), 1.09 (ID), and 0.524 (CTD) $ml \cdot mg^{-1} \cdot cm^{-1}$.

Gel Filtration Chromatography and Analytical Ultracentrifugation

Size exclusion chromatography of FL-XMcm10 was performed on a Superose 6 column (GE Healthcare) equilibrated with 20 mM Tris-HCl, pH 7.5, 500 mM NaCl, 5% glycerol, and 1 mM DTT. XMcm10 domains were eluted from an analytical Superdex™

200 column (GE Healthcare) equilibrated with S-200 buffer. 50-μL solutions of either XMcm10 (~1-2 mg/ml) or molecular weight standards were eluted at 0.5 ml/min. The standard curve was generated from thyroglobulin (670 kDa), aldolase (158 kDa), albumin (67 kDa), chicken ovalbumin (44 kDa), equine myoglobin (17 kDa), and vitamin B_{12} (1.4 kDa).

Sedimentation velocity analysis was conducted at 20 °C and 55,000 rpm using interference optics with a Beckman-Coulter XL-I analytical ultracentrifuge. Double sector synthetic boundary cells equipped with sapphire windows were used to match the sample and reference menisci. FL-XMcm10 was prepared in FL-buffer, and NTD and CTD were prepared in S-200 buffer. The data were initially analyzed using the program DCDT+, which computes the apparent sedimentation coefficient distribution function $g(s^*)$ using the time-derivative method (Stafford 1992; Philo 2000). For CTD, the molecular weight and sedimentation coefficient of the main component was obtained by global fitting of the data sets collected at multiple concentrations to a hybrid discrete-continuous model with Sedphat (Schuck 2003). For NTD, the data were fit to a monomer-dimer equilibrium model using the programs Sedanal (Stafford et al. 2004) and Sedphat. Molecular masses, partial specific volumes, and solvent densities were calculated using the SEDNTERP program (Laue 1992).

Fluorescence Anisotropy

DNA binding was measured by following an increase in fluorescence anisotropy as protein (MBP-XMcm10-His_6, NTD, ID, or CTD) was added to oligonucleotide d(TGACTACTACATGGTTGCCTACCAT) containing a 6-carboxyfluorescein moiety at

the 3′-end either alone (ssDNA) or annealed to an excess of the complementary strand (dsDNA). Forked DNA substrate tested against full-length Mcm10 was generated from two 50-mer deoxyoligonucleotides in which dC_{25} was added to the 3′-end of the sequence above and to the 5′-end of the complementary sequence. For Mcm10-ID and -CTD, forked and bubble DNA substrates were generated from the sequences d(**GGTAGGCACG**AACCATGTAGTAGTA)/d(AACCATGTAGTAGTA**CGTGCCT ACC**) and d(**GGTAGGCACG**AACCATGTAGTAGTA**GGCAATCAGC**)/d(**GCTGATTGC**CAA CCATGTAGTAGTA**CGTGCCTACC**), respectively, in which the boldface denotes duplex regions. Protein was added over the concentration range of 0.05-50 μM to a solution containing 25 nm DNA in S-200 buffer. For EDTA titrations, the buffer was supplemented with 0.1, 1, 10, and 25 mM EDTA. Polarized fluorescence intensities using excitation and emission wavelengths of 495 and 515 nm, respectively, were measured for 30 s (1/*s*) and averaged. Anisotropy (*r*) was calculated using the equation $r = (I_{par} - I_{perp})/(I_{par} + 2I_{perp})$, where I_{par} and I_{perp} are the observed fluorescence intensities recorded through polarizers oriented parallel and perpendicular, respectively, to the direction of vertically polarized light. Dissociation constants (K_d) were derived by fitting a two-state binding model to data from three experiments using Kaleidagraph 3.6 (Synergy).

Mcm10-Pol α Binding Assay

Recombinant DNA polymerase α was purified by immunoaffinity chromatography from extracts of Hi-5 insect cells co-infected with four recombinant baculoviruses as previously described (Voitenleitner et al. 1997). The p180 subunit was

prepared identically except only one recombinant baculovirus was used for infection. p180N (aa 1-323) was amplified by PCR on a cDNA template pBR322-p180 and cloned into the BamHI/EcoRI sites of a pGEX-2T expression vector (GE Healthcare). GST fusion proteins were expressed and purified by glutathione-agarose affinity chromatography as described previously (Smith et al. 1988). For the binding experiments, a total of 7 μg of purified polymerase α or p180 was incubated with SJK132-20 antibodies covalently coupled to Sepharose-4B beads (GE Healthcare), or 7 μg of purified p180N was incubated with glutathione-agarose beads (Sigma-Aldrich) in binding buffer (30 mM HEPES-KOH, pH 7.8, 10 mM KCl, 7 mM $MgCl_2$) containing 2% nonfat dry milk for 1 h at 4 °C with end-over-end rotation. Reactions contained either 5 or 15 μg of Trx-His_6-XMcm10-domain proteins. The beads were washed once with binding buffer, three times with wash buffer (30 mM HEPES-KOH, pH 7.8, 75 mM KCl, 7 mM $MgCl_2$, 0.25% inositol, 0.1% Nonidet P-40), and once with binding buffer (rotated for 10 min during each wash). The beads were resuspended in 30 μL of 2× SDS-PAGE loading buffer and heated at 100 °C for 5 min. Half of each sample was analyzed by 10% SDS-PAGE and immunoblotting with monoclonal antibody 2CT25, specific for the p180 subunit of polymerase α, rabbit anti-GST (Invitrogen) for p180N, and H-15 anti-His (Santa Cruz Biotechnology) for XMcm10 domains. Trx-only control experiments were performed to confirm that pol α, p180, and p180N did not interact with the Trx tag.

DNA Primase Assay

Oligoribonucleotide synthesis activity was measured as previously described for spMcm10 (Fien et al. 2006). Briefly, 2-8 pmol of purified XMcm10 or 0.6-2.4 pmol of purified polymerase α-primase were incubated at 37 °C for 40 min with 1.0 μM dT_{50}, 25

μCi of [α-^{32}P]ATP, and 0.1 mM ATP in a 10 μM reaction containing 40 mM Tris-HCl, pH 7.4, 10 mM magnesium acetate, 1 mM DTT, and 100 μg/ml bovine serum albumin. Reactions were treated with 1 unit of calf intestine phosphatase at 37 °C for 40 min. After the addition of 3 μL of sequencing gel running buffer (98% formamide, 10 mM EDTA at pH 8.0, 0.1% xylene cyanol, 0.1% bromophenol blue), samples were heated to 98 °C for 5 min and separated on a 25% polyacrylamide, 7 M urea gel. RNA was visualized by autoradiography.

CHAPTER III

SOLUTION NMR STRUCTURE OF THE C-TERMINAL DNA BINDING DOMAIN OF MCM10 REVEALS A CONSERVED MCM MOTIF*

Abstract

The eukaryotic DNA replication protein Mcm10 associates with chromatin in early S-phase and is required for assembly and function of the replication fork protein machinery. *Xenopus laevis* (X) Mcm10 binds DNA via a highly conserved internal domain (ID) and a carboxy-terminal domain (CTD) that is unique to higher eukaryotes. Although the structural basis of the interactions of the ID with DNA and polymerase α is known, little information is available for the CTD. We have identified the minimal DNA binding region of the XMcm10-CTD and determined its three-dimensional structure by solution NMR. The CTD contains a globular domain composed of two zinc binding motifs. NMR chemical shift perturbation and mutational analysis show that ssDNA binds only to the N-terminal (CCCH-type) zinc motif, whose structure is unique to Mcm10. The second (CCCC-type) zinc motif is not involved in DNA binding. However, it is structurally similar to the CCCC zinc ribbon in the N-terminal oligomerization domain of eukaryotic and archaeal MCM helicases. NMR analysis of a construct spanning both the ID and CTD reveals that the two DNA binding domains are structurally independent in solution, supporting a modular architecture for vertebrate Mcm10. Our results provide insight in the action of Mcm10 in the replisome and support a model in which it serves as a central scaffold through coupling of interactions with partner proteins and the DNA.

* The work presented in this chapter is currently *in press* at J Biol Chem as Robertson, P. D., B. Chagot, W. J. Chazin, and B. F. Eichman (2010).

Introduction

DNA synthesis at the eukaryotic replication fork requires coordination of enzymatic activities through a network of interactions within the dynamic multiprotein replisome. During replication initiation, the individual components of the replisome are assembled at each origin of replication in a sequential and regulated fashion to ensure that the genome is copied only once and at the proper time during each cell cycle. During G1, each origin is licensed for replication through ORC-dependent loading of a pre-replicative complex (pre-RC), which consists of the origin recognition complex (ORC), Cdc6, Cdt1, and the Mcm2-7 helicase (reviewed in (Bell et al. 2002)). As cells transition into S-phase, Mcm10 is recruited to the origin (Wohlschlegel et al. 2002; Ricke et al. 2004) and the pre-RC is activated through a series of phosphorylation events by cyclin- and Dbf4-dependent kinases (CDK and DDK) (Tanaka et al. 2007; Zegerman et al. 2007). Chromatin association of Mcm10 is required for loading Cdc45 and GINS (Wohlschlegel et al. 2002; Xu et al. 2009), which function with Mcm2-7 as a Cdc45-GINS-Mcm2-7 (CMG) helicase complex (Gambus et al. 2006; Moyer et al. 2006; Ilves et al. 2010). The mechanism of initial denaturation of duplex DNA at the origin is unknown, but is signaled by the presence of single-stranded (ss) DNA binding protein, replication protein A (RPA) (Tanaka et al. 1998; Zou et al. 2000; Moyer et al. 2006; Pacek et al. 2006). DNA synthesis is initiated by DNA polymerase α-primase (pol α), which associates with chromatin and the CMG complex via several factors, including Mcm10, Cdc45, RPA, And-1/Ctf4, and RecQL4 (Walter et al. 2000; Ricke et al. 2004; Zhu et al. 2007; Gambus et al. 2009; Im et al. 2009; Tanaka et al. 2009; Xu et al. 2009). Association of replicative DNA polymerases δ and ε, along with RPC loading of PCNA

completes the replisome and initiates the elongation process (reviewed in (Burgers 2009)).

Mcm10 is essential for the formation of an active replication fork (Merchant et al. 1997), and participates in numerous interactions with components of the replisome (Gambus et al. 2006). Mcm10 interacts with single-stranded and duplex DNA, consistent with a possible role as a protein-DNA mediator during origin melting (Fien et al. 2004; Robertson et al. 2008; Eisenberg et al. 2009). In early S-phase Mcm10 interacts with subunits of Mcm2-7, Cdc45, and is necessary for the assembly of the CMG helicase complex (Merchant et al. 1997; Homesley et al. 2000; Izumi et al. 2001; Christensen et al. 2003; Lee et al. 2003; Gambus et al. 2006; Im et al. 2009). In addition, Mcm10 interacts with DNA pol α, preventing its degradation *in vivo* and possibly serving to recruit pol α to the replisome (Ricke et al. 2004; Yang et al. 2005; Ricke et al. 2006; Chattopadhyay et al. 2007; Robertson et al. 2008). Mcm10 interactions with DNA, Mcm2-7 and pol α suggest it may function as a scaffold to physically link helicase and polymerase machinery within the replisome during the stages of replication initiation and primer elongation (Ricke et al. 2004; Lee et al. 2010).

Despite the identification of Mcm proteins from the same genetic screen for mutants defective in minichromosome maintenance (Maine et al. 1984), Mcm10 is evolutionarily distinct from Mcm2-7 and no sequence or structural homology has been identified between them (Liu et al. 2009). We previously established that the 95-kDa XMcm10 protein contains at least three structured domains—an N-terminal coiled-coil domain believed to facilitate protein oligomerization, and zinc-binding internal (ID) and C-terminal (CTD) domains that both bind DNA and pol α (Robertson et al. 2008).

Interestingly, an electron micrograph structure of human Mcm10 revealed a hexameric ring structure (Okorokov et al. 2007), although no other reports of Mcm10 hexamerization exist in the literature. To date, the only high resolution structural information available for Mcm10 exists for the highly conserved ID (Warren et al. 2008; Warren et al. 2009).

The vertebrate orthologs of Mcm10 contain a second DNA and pol α binding domain at the extreme C-terminus (Robertson et al. 2008). The sequence of the vertebrate CTD, not identifiable in yeast or land plants, contains putative winged-helix and zinc-binding domains that are predicted to facilitate DNA binding. To better understand the role of the vertebrate CTD and its interactions with DNA and other replication proteins, we mapped the DNA binding site to the zinc-coordinating region of XMcm10-CTD and determined the NMR solution structure of a globular domain containing this activity. Structural and mutational data support a separation of function for the two CTD zinc motifs comprising this domain. NMR studies of a tandem ID plus CTD construct revealed the modular architectural organization. These results support a model in which Mcm10 functions as a scaffold through essential protein and DNA interactions within the replisome.

Results

Structural Characterization of the C-terminal Domain of XMcm10

We previously identified residues 596-860 of *Xenopus laevis* Mcm10 as a structured C-terminal domain (CTD) containing two Zn^{2+} ions and encompassing both DNA and pol α binding activities (Robertson et al. 2008). Hence structural studies were

initiated using intact XMcm10$^{596-860}$. However, this and a library of constructs spanning this region proved refractory to crystallization presumably as a result of flexible and/or disordered polypeptide segments within this domain. To test this hypothesis, limited proteolysis experiments were performed on XMcm10$^{596-860}$. A stable ~20-kDa fragment was observed (Appendix Figure B1), which was identified to span residues 690-842 by MALDI-TOF mass spectrometry and Edman degradation. Sequence analysis of this subdomain revealed it contains a putative winged helix (residues 692-755) and two clusters of conserved cysteine and histidine residues (residues 768-821) previously predicted to coordinate Zn^{2+} ions (see Figure 7) (Okorokov et al. 2007; Robertson et al. 2008). No structural motifs and little secondary structure were predicted within residues 596-620. Importantly, purified XMcm10$^{690-842}$ bound ssDNA with roughly the same affinity as the larger XMcm10$^{596-860}$ construct, as measured by a fluorescence polarization assay (see below). Thus, XMcm10$^{690-842}$ appeared to be a stable subdomain of the CTD that retains DNA binding activity.

The structural features of XMcm10$^{690-842}$ were investigated by heteronuclear NMR spectroscopy (Figure 12), and to this end, over 90% of the backbone amide resonances were assigned (Appendix Figure B2). Using these backbone assignments, we first probed the DNA binding region using NMR chemical shift perturbation experiments. DNA binding was monitored by perturbation of the ^{15}N-^{1}H HSQC spectrum as unlabeled ssDNA was titrated into ^{15}N-enriched XMcm10$^{690-842}$. Signals from 19 residues showed a significant perturbation in the HSQC spectra in response to the binding of ssDNA, indicative of a specific binding event (Figure 12A). All of the perturbed residues localized to the putative Zn^{2+} binding region within residues 755-833. No signals

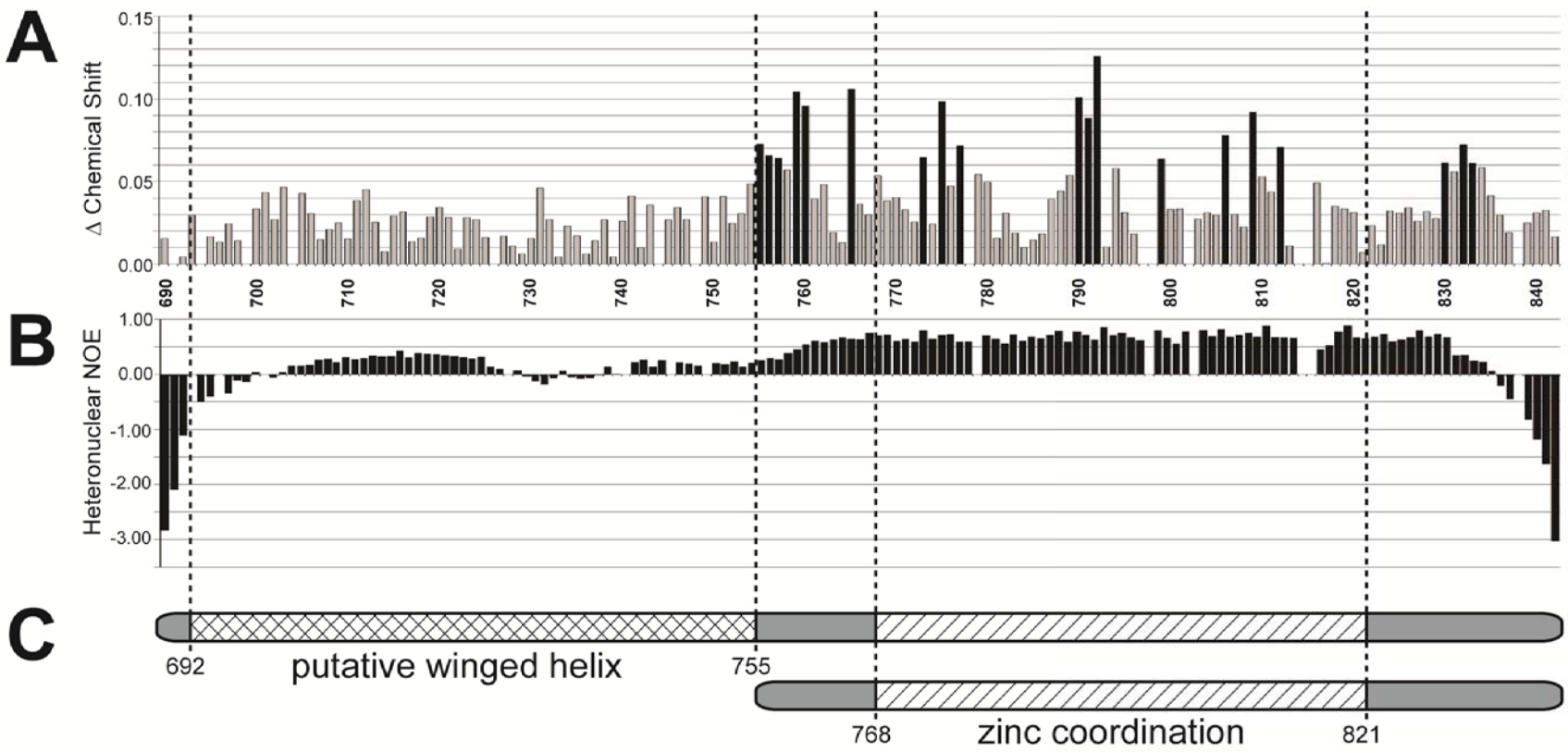


Figure 12. Mcm10-CTD contains a DNA binding sub-domain. (A) Chemical shift differences between ^{15}N-XMcm10$^{690\text{-}842}$ alone and in the presence of a 4-fold molar excess of ssDNA are plotted for each residue. Residue numbers are displayed along the x-axis. Black bars represent chemical shift differences greater than 1 standard deviation from the mean. (B) $\{^{1}H\}$-^{15}N heteronuclear NOE values, which correlate with the degrees of freedom for individual backbone residues. Positive values indicate more structural restraint, while negative values indicate more mobility. (C) Schematic of the two CTD constructs used in structural studies, XMcm10$^{690\text{-}842}$ (top) and XMcm10$^{755\text{-}842}$ (bottom) and the locations of putative motifs.

corresponding to the putative winged helix were affected from addition of even a 4-fold molar excess of ssDNA (Figure 12A). Thus, DNA binding in the CTD is localized exclusively to the zinc binding region.

We next assayed the relative flexibility of the backbone of XMcm10$^{690-842}$ using {^{1}H}-^{15}N heteronuclear NOE experiments (Figure 12B). Strikingly, residues 758-834 encompassing the zinc binding region generated an average NOE value of 0.66, indicative of a well-folded, globular structured domain. In contrast, the putative winged helix region (residues 690-757) had an average NOE of 0.14, indicative of much higher backbone flexibility. Taken together, the high NOE values and chemical shift perturbation data demonstrate that the zinc binding region is a well-folded, DNA binding motif (Figure 12C). The high sequence conservation and eight invariant cysteine and histidine residues in the C-terminal 100 residues of Mcm10 from higher eukaryotes suggest XMcm10$^{690-842}$ contains a functionally important domain suitable for structure determination**.**

Solution NMR Structure of XMcm10$^{755-842}$

The solution structure of XMcm10$^{755-842}$, a construct encompassing the zinc binding region, was determined by multidimensional heteronuclear NMR (Cavanagh et al. 2007). Nearly complete resonance assignments were obtained for this construct using standard double- and triple-resonance experiments, along with an (HB)CB(CGCD)HD spectrum that was acquired to assign sidechain resonances of the aromatic residues. NMR structures were generated using a combination of CYANA distance geometry calculations (Guntert 2004) and restrained molecular dynamics refinements in AMBER

(Case et al. 2005). Due to the large number of long-range restraints identified, a high precision structure was obtained with low total violation energies, no distance violations greater than 0.2 Å, no torsion angle violations greater than 5°, and low molecular energies (Table 1). The 20 conformers with the lowest restraint violation energy were selected for the final representative ensemble and are shown in Figure 13A.

Since zinc had been previously shown to bind to this domain, we examined the NOE-based structures to determine if the zinc coordinating residues could be identified. In fact, the eight invariant cysteine/histidine residues predicted to coordinate zinc ions were positioned into two distinct clusters in the structure (Appendix Figure B3). Moreover, the side chains within each cluster were posed in a tetrahedral geometry consistent with zinc binding. In addition, we observed that proton and carbon chemical shifts for many of these conserved residues deviated from the expected range represented in the Biological Magnetic Resonance Bank. Based on these observations and on our previous identification of two zinc atoms binding in the CTD (Robertson et al. 2008), we incorporated zinc ions and imposed distance restraints in both clusters in the final stages of structure refinement. As anticipated, these additional restraints significantly increased the precision of the structure in and around the Zn^{2+} sites.

The three-dimensional structure of $XMcm10^{755\text{-}842}$ is comprised of two independent zinc motifs tethered closely together in the shape of a "V", with the Zn^{2+} ions bound at the ends of each arm (Figure 13B). Superposition of the individual zinc motifs within the ensemble of NMR structures revealed a high degree of similarity and that each is slightly better defined than the entire globular domain (Appendix Figure B4). The N-terminal CCCH zinc motif (Zn1) spans residues 755-795 and consists of a three-

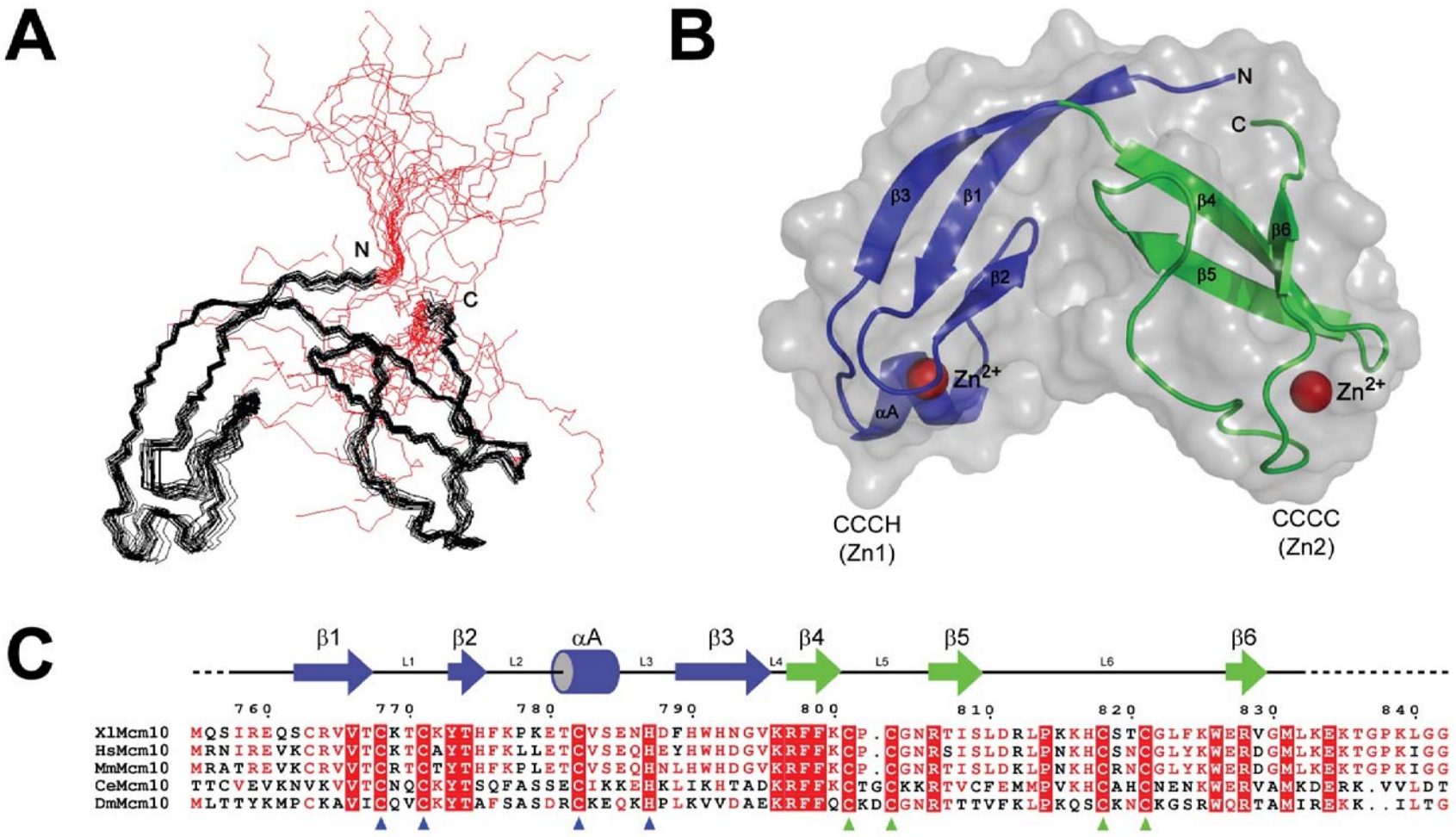
Table 3. Structural Statistics for XMcm10$^{755-842}$

Restraints for calculation	
Total NOE restraints	2436
Intraresidue	487
Sequential	632
Medium range	348
Long range	969
Dihedral angle restraint	28
Constraint violations, mean ±S.D.	
Distance violations	
0.1 Å < *d* < 0.2 Å	1.79 ± 1.03
d > 0.2 Å	0
Average maximum distance violations (Å)	0.14 ± 0.02
Torsion angle violations > 5.0°	0
Average maximum torsion angle violations (°)	0
AMBER energies, mean ±S.D. (kcal mol^{-1})	
Restraint	2.25 ± 0.40
van der Waals	-645 ± 11
Total molecular	-2804 ± 11
Precision, RMSD from mean (Å), ordered region[a]	
Backbone	0.63 ± 0.21
All heavy atoms	1.04 ± 0.15
Ramachandran statistics[b] (%)	
Most favored	86.0
Additionally allowed	13.1
Generously allowed	0.4
Disallowed	0.4

[a] Residues 761-832

[b] PROCHECK nomenclature

Figure 13. The Structure of the Dual Zinc Cluster in XMcm10-CTD. (A) Backbone superposition of the twenty lowest energy NMR structures of $XMcm10^{755-842}$. Regions in red correspond to unrestrained terminal regions, which have been omitted from subsequent figures. (B) $XMcm10^{755-842}$ depicted as a ribbon and superimposed onto a transparent grey molecular surface. Individual CCCH and CCCC zinc motifs are colored blue and green, respectively, and Zn^{2+}ions are depicted as red spheres. (C) Sequence alignment of the C-terminal region of Mcm10 proteins from *Xenopus laevis* (Xl), *Homo sapiens* (Hs), *Mus Musculus* (Mm), *Caenorhabditis elegans* (Ce), and *Drosophila melanogaster* (Dm). The secondary structure is shown schematically above the sequence, and Zn^{2+} coordinating residues are highlighted with triangles below.

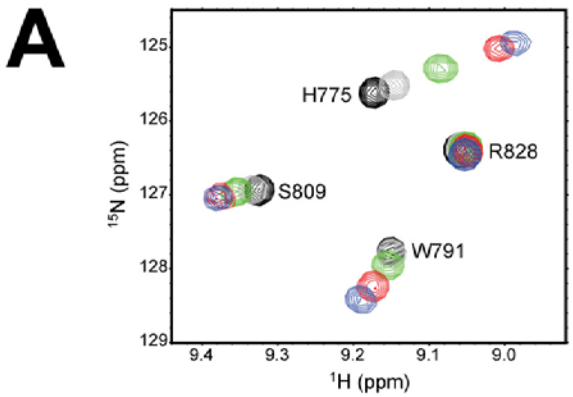

A
N
C
B
N
C
β3
β1
β4
β6
β2
β5
Zn2+
Zn2+
αA
CCCH
(Zn1)
CCCC
(Zn2)
C
β1
β2
αA
β3
β4
β5
β6
L1
L2
L3
L4
L5
L6
760
770
780
790
800
810
820
830
840
XlMcm10 MQSIREQSCRVVTCKTCKYTHFKPKETCVSENHDFHWHNGVKRFFKCP.CGNRTISLDRLPKKHCSTCGLFKWERVGMLKEKTGPKLGG
HsMcm10 MRNIREVKCRVVTCKTCAYTHFKLLETCVSEQHEYHWHDGVKRFFKCP.CGNRSISLDRLPNKHCSNCGLYKWERDGMLKEKTGPKIGG
MmMcm10 MRATREVKCRVVTCRTCTYTHFKPLETCVSEQHNLHWHDGVKRFFKCP.CGNRTISLDKLPNKHCRNCGLYKWERDGMLKEKTGPKIGG
CeMcm10 TTCVEVKNVKVVTCNQCKYTSQFASSECIKKEHKLIKHTADKRFFKCTGCKKRTVCFEMMPVKHCAHCNENKWERVAMKDERK.VVLDT
DmMcm10 MLTTYKMPCKAVICQVCKYTAFSASDRCKEQKHPLKVVDAEKRFFQCKDCGNRTTTVFKLPKQSGKNCKGSRWQRTAMIREKK..ILTG

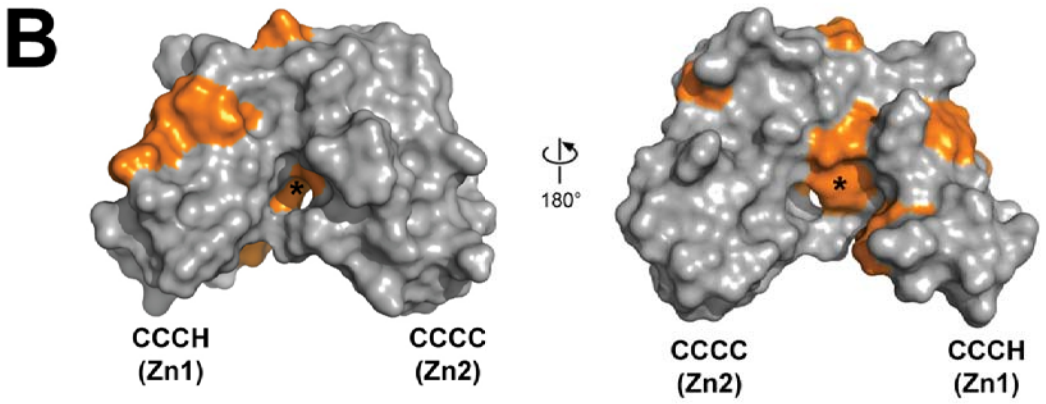
stranded antiparallel β-sheet (β1-β3) capped with a short perpendicular α-helix (αA). The Zn^{2+} ion is coordinated between the helix and sheet by Cys768 and Cys771 on the L1 loop, Cys782 on the αA helix and His787 on the L3 loop. The C-terminal CCCC zinc motif (Zn2) adopts a twisted antiparallel β-sheet (β4-β6) with the zinc coordinated by Cys801 and Cys803 on the short loop between strands β4 and β5, and by Cys818 and Cys821 on the extended loop between β5 and β6. Overall, $XMcm10^{755-842}$ adopts a relatively globular fold as a result of the short linker and side chain interactions between the two zinc motifs.

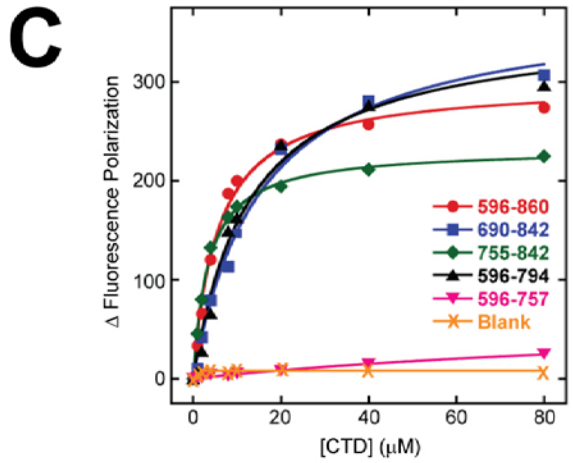
ssDNA Binding is Localized to the CCCH Zinc Motif

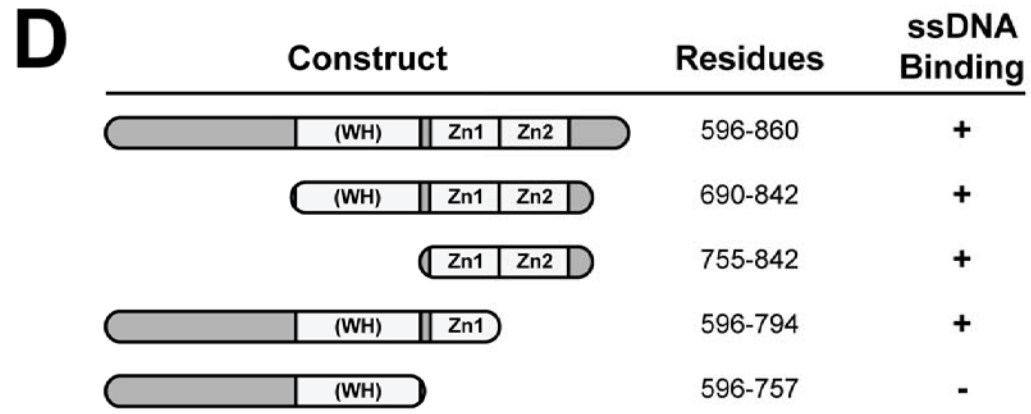
Close inspection of the chemical shift perturbation data shown in Figure 12B suggested that DNA binding to the CTD was dominated by the Zn1 arm. Taking advantage of the complete sequence specific NMR assignments of $XMcm10^{755-842}$ amides, we repeated the NMR titrations with this shorter construct to map the ssDNA binding site onto our structure (Figure 14). DNA binding was determined by monitoring perturbations in the ^{15}N-^{1}H HSQC spectrum as unlabeled ssDNA was titrated into ^{15}N-enriched $XMcm10^{755-842}$ (Figure 14A). Eleven signals shifted significantly in response to DNA, while the others remained unaffected (Appendix Figure B5). Mapping the positions of the perturbed residues onto the structure of $XMcm10^{755-842}$ revealed that ssDNA binding is indeed localized almost exclusively to the Zn1 motif (Figure 14B). The perturbed residues trace a continuous ~35 Å path around the Zn1 arm, raising the possibility that ssDNA partially encircles this motif. In support of this, the optimal length of ssDNA needed to fully engage the CTD was between 10-15 nucleotides, as measured

Figure 14. ssDNA Binding to the CCCH Zinc Motif. (A) Overlays of a representative region of the ^{15}N-^{1}H HSQC spectra of ^{15}N-XMcm10$^{755-842}$ in the absence (black) and presence of 0.5 (grey), 1 (green), 2 (red), and 4 (blue) fold molar excess of ssDNA. (B) Residues perturbed by ssDNA binding from the HSQC titration are colored orange against the molecular surface of XMcm10$^{755-842}$. The asterisk denotes the position of Phe776. (C) ssDNA binding to CTD-deletion constructs was monitored by the change in fluorescence polarization as protein was added to fluorescein (FAM)-labeled d(ATGGTAGGCAACCAT). Addition of buffer only to FAM-DNA is shown as orange Xs. Data shown is from one representative experiment and was reproduced in triplicate. (D) Schematic representation of the data shown in panel C.


A
H775
R828
S809
W791
125
126
127
128
129
9.4
9.3
9.2
9.1
9.0
15N (ppm)
1H (ppm)


B
CCCH
(Zn1)
CCCC
(Zn2)
180°
CCCC
(Zn2)
CCCH
(Zn1)


C
Δ Fluorescence Polarization
300
200
100
0
0
20
40
60
80
[CTD] (μM)
596-860
690-842
755-842
596-794
596-757
Blank


D
Construct
Residues
ssDNA Binding
(WH)
Zn1
Zn2
596-860
+
690-842
+
755-842
+
596-794
+
596-757
-

by the *in vitro* fluorescence polarization assay (Appendix Figure B6). Interestingly, Phe776 at the Zn1-Zn2 interface showed a dramatic resonance shift (Figure 14B and B4), suggesting DNA contacts extend to this region or that an allosteric hinge-like movement of the Zn1 arm accompanies binding of ssDNA.

Given the lack of chemical shift perturbation outside of the Zn1 motif, we examined the contribution of Zn2 and winged helix motifs to DNA binding by mutational analysis. A series of CTD deletion constructs were tested for their ability to bind 15mer ssDNA by fluorescence polarization (Figures 14C and 14D). Addition of the full-length CTD ($XMcm10^{596-860}$) to fluorescein-labeled ssDNA resulted in a robust change in fluorescence polarization and an apparent K_d of 5.2 ± 0.1 μM. Deletion of Zn2 from the CTD ($XMcm10^{596-794}$) had only a modest 2-fold effect on ssDNA binding (K_d = 10.8 ± 1.1 μM). Likewise, $XMcm10^{690-842}$ and $XMcm10^{755-842}$, constructs lacking the region N-terminal to Zn1, bound ssDNA with similar affinity (K_d = 14.4 ± 2.5 and 3.4 ± 0.2 μM, respectively). In contrast, removal of both Zn1 and Zn2 motifs from the CTD ($XMcm10^{596-757}$) completely abrogated DNA binding (Figure 14C). We therefore conclude that Zn1 is necessary for ssDNA binding by XMcm10-CTD and that Zn2 and the putative winged helix motif do not significantly contribute to the binding affinity.

Spatial Separation of Mcm10 DNA binding Motifs

The large separation in DNA binding regions from the ID and CTD raises the question of how the two domains work together to bind DNA with relatively high affinity (Robertson et al. 2008; Warren et al. 2009). We previously suggested that the proteolytic sensitivity and lack of secondary structure in the region between the ID and CTD was the

result of inherent flexibility that may provide Mcm10 with the ability to adapt to different structural states during replisome assembly and progression (Robertson et al. 2008). To gain insight into the extent of the interaction between the two domains, we took advantage of the sequence specific NMR assignments for both the ID and CTD. NMR is a powerful technique for studying protein structural dynamics, and has been applied recently to the highly modular 116-kDa RPA heterotrimer (Brosey et al. 2009). The high protein concentrations required for NMR experiments prevented structural analysis of full-length XMcm10. However, we were able to obtain a high-quality ^{15}N-^{1}H TROSY-HSQC spectrum for XMcm10$^{230-860}$, which encompasses both the ID and CTD as well as the intervening linker region (Figures 15 and B6). The central (~8 ppm) region of the XMcm10$^{230-860}$ spectrum corresponding primarily to residues in random coil and α-helical conformation contains numerous overlapping signals. However, signals outside of this region are well resolved and can be readily compared to signals nearly identical in the spectra of the individual ID and CTD constructs (Warren et al. 2008). We found 94 resonances in nearly identical positions to those in the isolated ID and CTD domains. Since the NMR chemical shift is exclusively sensitive to structural perturbations, these data provide convincing evidence that the structures of the ID and CTD are structurally independent of one another in the XMcm10$^{230-860}$ construct. The conclusion is supported by the absence in chemical shift perturbations of individually ^{15}N-enriched ID and CTD domains when added together in trans (data not shown). NMR line widths and signal intensities imply a modular organization of XMcm10$^{230-860}$. Indeed, three regimes were observed. CTD signals were more intense than those of the larger ID (Appendix Figure B7), consistent with the smaller size of the globular portion of CTD relative to the ID.

The third set of signals corresponded to unassigned resonances with ^{1}H chemical shifts of ~8.2 ppm, which can be attributed to the linker between ID and CTD (Figure 16A). The location of these signals in the 'random coil' region of the spectrum combined with their extraordinarily high intensity implies the linker residues are dynamically disordered. Taken together with the low sequence conservation and high proteolytic sensitivity of the linker (residues 430-595) (Robertson et al. 2008), the NMR studies of XMcm10$^{230-860}$ strongly support the existence of a flexible linker between the two DNA binding domains of XMcm10.

Discussion

A Novel DNA Binding Motif in the Mcm10 C-Terminus

In this study we determined the structure of the zinc cluster within Mcm10-CTD and identified the CCCH zinc motif (Zn1) as the predominant DNA binding region. A search for structural homologs to Zn1 using the Dali server (Holm et al. 1995) returned no results, suggesting Mcm10 is structurally distinct from other replication proteins. Interestingly, no C-terminal zinc motifs are found in yeast Mcm10 sequences, implying that Mcm10 in lower eukaryotes has a different functional architecture and mode of action. Differences between yeast and vertebrate Mcm10 are also evident from the mapping of DNA binding regions. *S. pombe* Mcm10 binds ssDNA with nanomolar affinity through a domain corresponding to the ID and has no affinity in the extreme C-terminal 180 residues (Fien et al. 2004). In contrast, XMcm10 utilizes two relatively low-affinity DNA binding domains to attain nanomolar affinity for the full-length protein

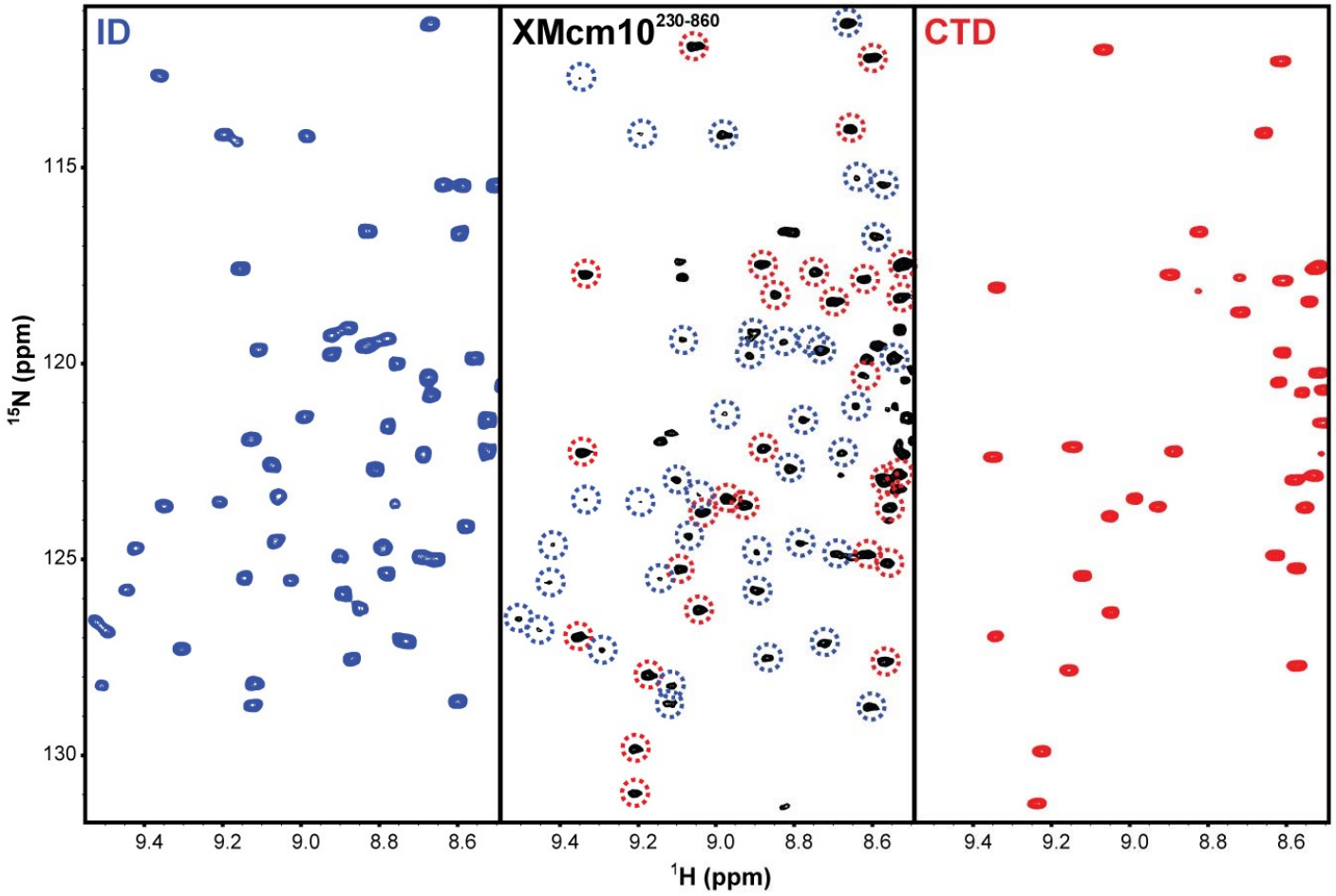


Figure 15. Structural Analysis of XMcm10$^{230\text{-}860}$. ^{15}N-^{1}H HSQC spectra of ^{15}N-XMcm10$^{230\text{-}427}$ (ID, left panel) and ^{15}N-XMcm10$^{755\text{-}842}$ (CTD, right panel), and the ^{15}N-^{1}H TROSY-HSQC spectrum of ^{15}N-XMcm10$^{230\text{-}860}$ (center panel). Signals in the XMcm10$^{230\text{-}860}$ spectrum that correspond to residues in the ID or CTD have been circled in blue and red, respectively.

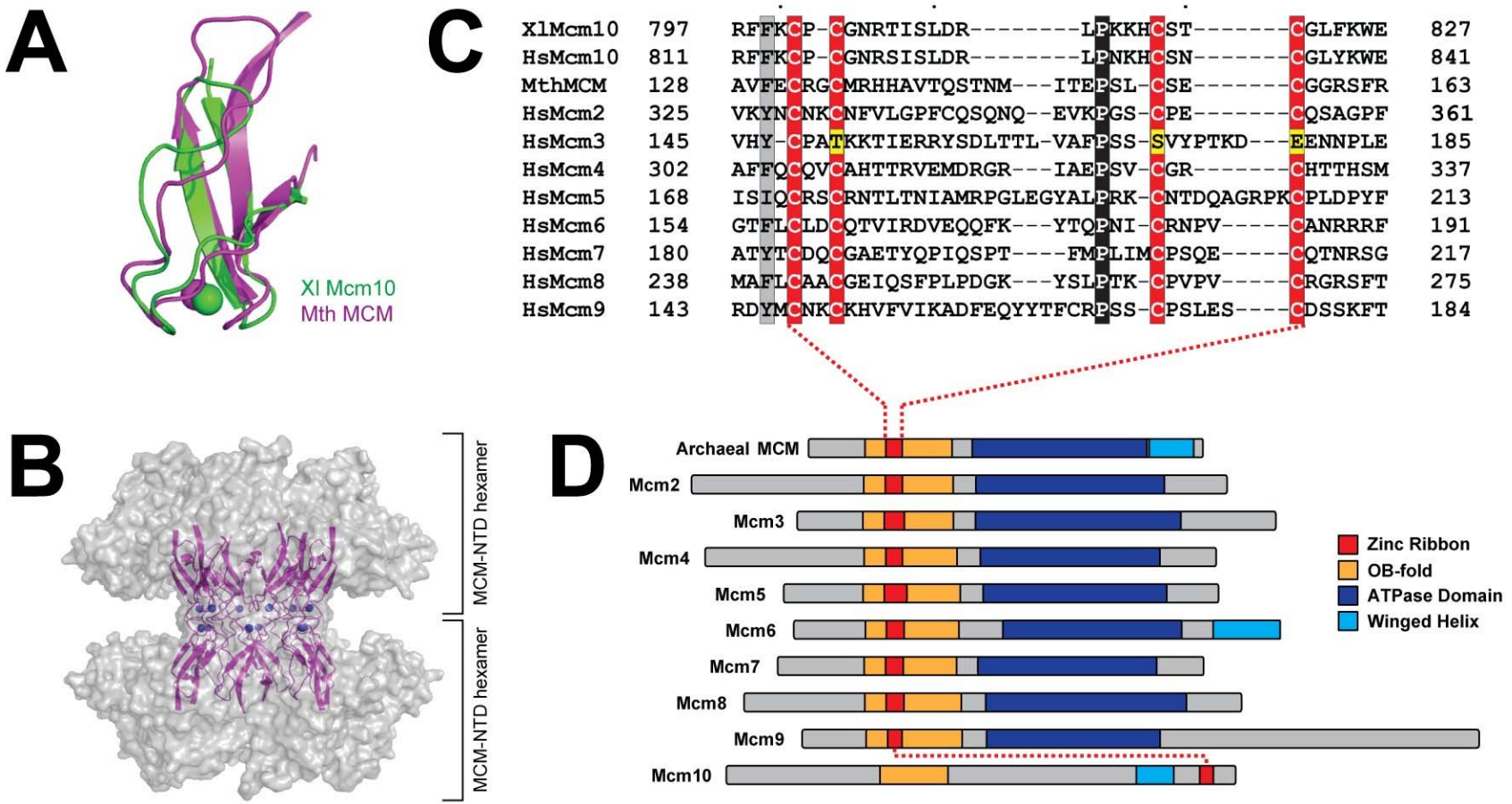
(Warren et al. 2009). The hypothesis that the Mcm10 proteins do not function similarly is consistent with the differences in composition between yeast and vertebrate replisomes and fundamental differences between other replication proteins such as DNA primases.

Conservation of Zinc Motif Sequence and Structure, But Not Function

Unlike the novel Zn1 motif, a Dali search performed on the Zn2 motif identified a clear hit from the crystal structure of *Methanobacterium thermoautotrophicum* MCM helicase N-terminal domain (MthMCM-NTD, PDB: 1LTL) (Fletcher et al. 2003). The zinc motifs in the two structures are very similar, with an RMSD of 2.7 Å for all backbone atoms between MthMCM residues 125-166 and XMcm10 residues 795-830 (Figure 16A). In MthMCM, this zinc motif mediates a head-to-head interaction between two hexameric rings (Figure 16B) and is important for double hexamer formation and proper helicase function (Poplawski et al. 2001; Fletcher et al. 2003; Kasiviswanathan et al. 2004; Fletcher et al. 2005). As shown in Figure 16C, the sequence of this motif is conserved in the six Mcm2-7 subunits (Kearsey et al. 1998; Tye 1999; Poplawski et al. 2001; Fletcher et al. 2003) and in the recently identified Mcm8 and Mcm9 proteins (Gozuacik et al. 2003; Lutzmann et al. 2005; Yoshida 2005). Thus, the structure of the Mcm10 CCCC zinc ribbon presented here reveals a homology between the evolutionarily distinct Mcm10 and Mcm2-9 protein families (Liu, 2008 #1552).

Unlike the CCCC zinc motifs, Mcm10 and MthMCM contain OB-folds that are in the same location in the primary structures (Figure 16D). In the MCM helicase, the CCCC zinc ribbon is inserted into the L12 loop of the OB-fold to form a continuous domain important for helicase activity, likely by stabilizing the overall fold of the NTD

Figure 16. Conservation of the CCCC Zinc Ribbon among MCM proteins. (A) Structural alignment of the CCCC zinc ribbon from *Xenopus laevis* (Xl) Mcm10$^{755\text{-}842}$ (green) and *Methanobacterium thermoautotrophicum* (Mth) MCM (magenta, PDB: 1LTL). (B) The crystal structure of the MthMCM N-terminal domain double hexamer is shown as a transparent grey molecular surface with the head-to-head CCCC motifs at the hexamer-hexamer interface rendered as magenta ribbons and blue Zn^{2+} spheres. (C) Sequence alignment of the CCCC zinc motifs from *Xenopus* and human Mcm10, MthMCM, and human Mcm2-9 proteins. Conserved residues are shaded grey, invariant residues black, and zinc-coordinating residues red and yellow. (D) Schematic alignment of archaeal MCM and eukaryotic Mcm2-10 proteins colored according to domain (AAA^+, blue) or supersecondary motif (CCCC zinc ribbon, red; OB-fold, gold; winged helix, cyan).


A
Xl Mcm10
Mth MCM
C
XlMcm10 797 RFFKCP-CGNRTISLDR--------LPKKHCST-------CGLFKWE 827
HsMcm10 811 RFFKCP-CGNRSISLDR--------LPNKHCSN-------CGLYKWE 841
MthMCM 128 AVFECRGCMRHHAVTQSTNM---ITEPSL-CSE-------CGGRSFR 163
HsMcm2 325 VKYNCNKCNFVLGPFCQSQNQ--EVKPGS-CPE-------CQSAGPF 361
HsMcm3 145 VHY-CPATKKTIERRYSDLTTL-VAFPSS-SVYPTKD---EENNPLE 185
HsMcm4 302 AFFQCQVCAHTTRVEMDRGR---IAEPSV-CGR-------CHTTHSM 337
HsMcm5 168 ISIQCRSCRNTLTNIAMRPGLEGYALPRK-CNTDQAGRPKCPLDPYF 213
HsMcm6 154 GTFLCLDCQTVIRDVEQQFK---YTQPNI-CRNPV-----CANRRRF 191
HsMcm7 180 ATYTCDQCGAETYQPIQSPT----FMPLIMCPSQE-----CQTNRSG 217
HsMcm8 238 MAFLCAACGEIQSFPLPDGK---YSLPTK-CPVPV-----CRGRSFT 275
HsMcm9 143 RDYMCNKCKHVFVIKADFEQYYTFCRPSS-CPSLES----CDSSKFT 184
B
MCM-NTD hexamer
MCM-NTD hexamer
D
Archaeal MCM
Mcm2
Mcm3
Mcm4
Mcm5
Mcm6
Mcm7
Mcm8
Mcm9
Mcm10
Zinc Ribbon
OB-fold
ATPase Domain
Winged Helix

(Appendix Figure B8) (Fletcher et al. 2003; Kasiviswanathan et al. 2004). In contrast, the L12 loop in Mcm10 does not contain a zinc motif and is directly involved in ssDNA binding (Warren et al. 2008; Warren et al. 2009). In fact, the OB-fold in the ID and the zinc motif in the CTD are separated in sequence by over 400 residues and are completely independent (Figure 15 and 16D). Thus, although Mcm10 and Mcm2-7 share common motifs, including putative winged-helix regions (Figure 16D), it is unlikely that the two proteins share a similar architecture.

Implication for Mcm10 Function

Like MthMCM, the yeast Mcm2-7 replicative helicase was recently shown to load onto DNA as a double hexamer connected through its NTD (Remus et al. 2009). Given the conservation of the CCCC zinc motif (Figure 16C), the double hexamer in the eukaryotic helicase likely occurs in a manner to MthMCM. Indeed, mutations within this zinc motif disrupt double hexamer formation in MthMCM (Fletcher et al. 2005) and result in lethality or temperature sensitivity in yeast (Yan et al. 1991; Dalton et al. 1997). Interestingly, replication progression complexes contain only one copy of Mcm4, which suggests that a single Mcm2-7 ring is sufficient to unwind DNA during elongation. Since the NTD zinc motif mediates double hexamer formation, the zinc motif surface would be available to bind other proteins. It is enticing to speculate that Mcm10 engages Mcm2-7 through interactions between zinc motifs. Indeed, Mcm10 has been shown to interact directly with various subunits of the Mcm helicase (Izumi et al. 2001; Lee et al. 2003; Gambus et al. 2006) and has been implicated in physically linking the helicase and pol α (Ricke et al. 2004; Lee et al. 2010). The observation that Mcm10 remains associated with

the polymerases upon uncoupling the helicase-polymerase complex (Pacek et al. 2006) suggests a higher affinity interaction between Mcm10 and pol α. Although poorly folded in the context of the isolated CTD, the putative winged-helix domain may provide an additional protein interaction module for Mcm10. Together, the modular architecture and myriad of protein and DNA interaction sites support a model in which Mcm10 functions as a scaffold, serving to co-localize critical elements of the replisome during the initiation and elongation phases of replication.

Experimental Procedures

Protein Expression and Purification

The cDNA for all CTD-containing fragments were PCR-amplified from a previously described plasmid (Wohlschlegel et al. 2002) and ligated into a modified pET-32a (Novagen) expression vector to generate N-terminal thioredoxin (Trx)-His_6 fusion proteins. $XMcm10^{230-860}$ was purified as previously described (Warren et al. 2009). $XMcm10^{755-842}$, $XMcm10^{690-842}$, $XMcm10^{596-794}$, $XMcm10^{596-757}$, and $XMcm10^{596-860}$ proteins were overexpressed in *E. coli* BL21(DE3) cells in LB medium supplemented with 100 μg/ml ampicillin, 7.5 μM $ZnSO_4$, and 0.5 mM IPTG for 3 hours at 37 °C. Isotopically enriched Mcm10 samples for NMR were overexpressed in M9 minimal medium supplemented with $^{15}NH_4Cl$ and/or [$^{13}C_6$]glucose (Cambridge Isotope Laboratories) as the sole sources of nitrogen and/or carbon. Cells were resuspended in 50 mM Tris-HCl at pH 7.5, 500 mM NaCl, and 10% glycerol (Buffer L) and lysed with an EmulsiFlex-C3 homogenizer (Avestin). Cell lysates were centrifuged at 35,000 x g for 20 minutes. Trx-His_6-Mcm10 proteins were purified from the supernatant by nickel-

nitrilotriacetic (Ni-NTA, Qiagen) affinity chromatography using a Buffer L wash and Buffer L/250 mM imidazole elution. Fractions were visualized by SDS-PAGE and Coomassie Blue staining. Those containing Mcm10 were pooled and incubated overnight at 4 °C with PreScission Protease (GE Healthcare) at a 1:50 protease:Mcm10 mass ratio to remove the Trx-His_6 affinity tag. After cleavage, this mixture was diluted 10-fold in 50 mM Tris-HCl at pH 7.5, and 10% glycerol (Buffer A) and passed through a 5-ml Q Sepharose HP anion exchange column directly onto a 5-ml SP Sepharose HP cation column (GE Healthcare). Proteins were eluted using a linear gradient from Buffer A/0.1 M NaCl to Buffer A/1 M NaCl. Fractions containing Mcm10 proteins were pooled, concentrated using an Amicon spin concentrator (Millipore) and purified over a 320-ml Superdex 200 gel filtration preparative grade column (GE Healthcare) that had been equilibrated in 20 mM Tris-HCl at pH 7.5, 100 mM NaCl, 5 mM β-mercaptoethanol, and 5% glycerol.

NMR Spectroscopy

All NMR data were collected at 25 °C using Bruker DRX600 and DRX800 spectrometers equipped with cryoprobes. Data were processed with Topspin 2.0b and further analyzed using Sparky (Goddard et al. 2006). All XMcm10 samples were buffer exchanged using an Amicon Spin Concentrator (Millipore) into 25 mM NaH_2PO_4 at pH 6.5 in 90% H_2O/10% D_2O and 100 mM NaCl ($XMcm10^{690\text{-}842}$ and $XMcm10^{755\text{-}842}$) or 175 mM NaCl ($XMcm10^{230\text{-}860}$). $XMcm10^{690\text{-}842}$ samples were concentrated to 305 μM for $\{^1H\}$-^{15}N heteronuclear NOE experiments and to 190 μM for ^{15}N-^1H HSQC titrations. $XMcm10^{755\text{-}842}$ was concentrated to 920 μM for structure determination experiments and

to 300 μM for HSQC titrations. XMcm10$^{230-860}$ samples for ^{15}N-^{1}H TROSY-HSQC were at a concentration of 80 μM.

Backbone resonance assignments for XMcm10$^{690-842}$ and XMcm10$^{755-842}$ were obtained using a combination of 2D ^{15}N-^{1}H HSQC and 3D HNCACB, CBCA(CO)NH and HNCO spectra. Side chain aliphatic resonances were assigned using H(CCCO)NH, (H)CC(CO)NH, and HBHANH 3D experiments. Aromatic side chain resonances were assigned using (HB)CB(CGCD)HD and 2D homonuclear COSY, TOCSY and NOESY (T_m = 120 ms) experiments. NOE-based distance restraints were assigned from a homonuclear 2D NOESY experiment, as well as 3D ^{13}C-NOESY-HSQC and ^{15}N-NOESY-HSQC (T_m = 120 ms) experiments using ^{13}C,^{15}N-enriched samples. Steady-state {^{1}H}-^{15}N heteronuclear NOE data were collected with and without 3 sec of ^{1}H saturation and 7 sec of recycle delay. NOEs for 144 resolved resonances were determined using the ratio of signal intensities with and without ^{1}H saturation. Additional details for acquisition parameters are provided in Appendix Table B1. All chemical shifts were deposited in the BMRB.

Structure Calculation

Starting structures were determined using CYANA (Guntert 2004). Seven iterative cycles of calculations were carried out starting with a set of manually assigned NOEs. 100 structures were created per cycle with backbone torsion angle restraints obtained from ^{1}H, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, and ^{15}N chemical shifts using TALOS with a minimum range of ±35°. The 50 structures with lowest values of the CYANA target function were further refined using restrained molecular dynamic simulations in AMBER (Case et al.

2005) following published protocols (Hu et al. 2003). The twenty conformers with the lowest restraint violation energy were selected as the final representative ensemble. PROCHECK-NMR and MolProbity were used to assess the quality of the final ensemble (Laskowski et al. 1996; Davis et al. 2007). The final ensemble and distance restraints have been deposited in the PDB under accession code 2KWQ.

DNA Binding

Mcm10 DNA binding affinities were measured by following the increase in fluorescence polarization as the protein was added to DNA oligonucleotides labeled at the 3′ -end with 6-carboxyfluorescein (FAM, Integrated DNA Technologies). Binding data for CTD deletion mutants shown in Figure 14C were measured using d(ATGGTAGGCAACCAT)-FAM. Oligonucleotides used to determine the length dependence of DNA binding are shown in Appendix Figure B5. Mcm10 proteins were added over a concentration range of 0-20 μM to a solution of 50 nM fluorescein-DNA, and polarized fluorescence intensities were measured at excitation and emission wavelengths of 495 and 538 nm, respectively. The experiments were performed at 25 °C in 25 mM Tris-HCl at pH 7.5, 100 mM NaCl, and 5% glycerol. Each measurement was made in triplicate and the apparent dissociation constants (K_d) were calculated by fitting the binding curve using a two-state binding model using Kaleidagraph 3.6 (Synergy Software).

Chemical shift perturbation experiments with CTD constructs were performed using d(ATGGTAGGCAACCAT) at Mcm10:DNA molar ratios of 1:0, 1:0.5, 1:1, 1:2, and 1:4. Chemical shift perturbations were quantified using the equation $\Delta\delta_{ave} = (((\Delta\delta_{1H})^2$

+ $(\Delta\delta_{15N}/5)^2)/2)^{1/2}$. Values of $\Delta\delta_{ave}$ greater than one standard deviation from the mean were considered significant.

CHAPTER IV

FURTHER STRUCTURAL AND BIOCHEMICAL CHARACTERIZATION OF MCM10 AND ITS C-TERMINAL DOMAIN

In order to expand our understanding of the molecular role Mcm10 plays during DNA replication, this chapter is aimed at further characterizing the functional capabilities of Mcm10 and its domains through structural and biochemical approaches. These studies include additional structural data for the CTD, as well as a more thorough investigation of its DNA binding capabilities. An in-depth analysis of DNA binding by larger Mcm10 constructs is also presented, and the role of the linker region between the ID and CTD was investigated. Finally, recent progress made towards determining the oligomeric state of XMcm10 is presented and discussed in detail.

Structural Studies of XMcm10-CTD

Crystallization Trials

The structure of the zinc-coordinating region of XMcm10-CTD (residues 755-842) presented in chapter 3 helped us to understand the DNA binding capabilities of the CTD, as well as identified a structural link to other MCM proteins. In addition to this region, I have also done extensive work towards determining the 3-dimensional structure of the entire CTD (residues 596-860). To increase the likelihood of obtaining diffraction quality crystals of XMcm10-CTD, a total of seventeen variations of the CTD were subcloned into expression vectors. Each construct was designed to minimize disordered regions using a combination of secondary structure and disorder prediction algorithms, as

well as sequence conservation. Of the seventeen constructs successfully cloned, ten were optimized for bacterial expression, purified, and tested for the presence of secondary structural elements using circular dichroism (CD) (Table 4). After confirming the constructs were folded using limited proteolysis, they were subjected to crystallization screening which included ~1200 conditions in-house and another ~1500 conditions at the Hauptman-Woodward High Throughput screening (HWI) facility in Buffalo, NY (Luft et al. 2003). Screening was completed at multiple protein concentrations, as well as in the presence and absence of glycerol in the setup buffer. In addition, a number of the constructs were also screened in the presence of varying lengths of single- and double-strand DNA, or with the equimolar amounts of purified XMcm10-ID in an attempt to obtain crystals of XMcm10-CTD in complex with a binding partner. Crystals were obtained with $XMcm10^{596-860}$ in a slight excess (1:1.2) of ss22mer DNA in 100 mM MES pH 6.0, 100 mM $CaCl_2$ and 20% PEG 2000 (v/v). However, conditions were optimized to 100 mM MES pH 6.0, 150 mM $CaCl_2$ and 22% PEG 2000 (v/v) to attain maximum crystal size. Crystals were flash frozen and examined for X-ray diffraction at a synchrotron source (APS, SER-CAT). Despite the size and reproducibility of the crystals, none of those screened showed consistent signs of diffraction. Crystallization conditions and DNA length/sequence were further optimized to yield diffraction, however none of the subsequent optimizations were successful. The failure to crystallize was likely attributed to region(s) of flexibility within the domain which could inhibit the formation of consistent crystal lattice contacts.

Table 4. XMcm10-CTD crystallization constructs

Residues	Apo	+ DNA	+ ID	@ HWI
596-860	+/-	+/-	+/-	+/-
623-860	+/-	+	+	+
635-860	+/-	+		+/-
641-860	+/-	+	+	+
685-860	+/-			
690-860	+/-			+
698-860	+/-			
690-842	+/-	+/-	+	+/-
755-842	+/-	+		+/-
755-834	+/-			

Note: +/- corresponds to with (+) and/or without (-) glycerol in the setup buffer

NMR Spectroscopy

While crystallization of XMcm10-CTD was unsuccessful in yielding a high resolution structure, the series of CTD truncation constructs proved useful for analysis by NMR spectroscopy. The full CTD, XMcm10$^{596-860}$, had a molecular mass of ~30 kDa and had spectral properties suggesting it was not suitable for structure determination by NMR. However, XMcm10$^{690-842}$, which contained the putative winged helix and the zinc-coordinating region and had a mass of ~18 kDa, provided excellent spectral properties. Comparison of the ^{1}H-^{15}N HSQC spectra of ^{15}N-XMcm10$^{690-842}$ and ^{15}N-XMcm10$^{596-860}$ revealed no major changes in the positions of the well-dispersed signals (Figure 17), which indicates that there were no major structural changes for XMcm10$^{690-842}$ compared to the fully intact CTD. Multidimensional heteronuclear NMR experiments were employed to assign 92% of all protons. After manual assignment of NOE signals in each NOESY spectra, CYANA was used to generate structures (Figure 18). Unfortunately, the RMSD values obtained from CYANA for these models were poor due to the unrestrained

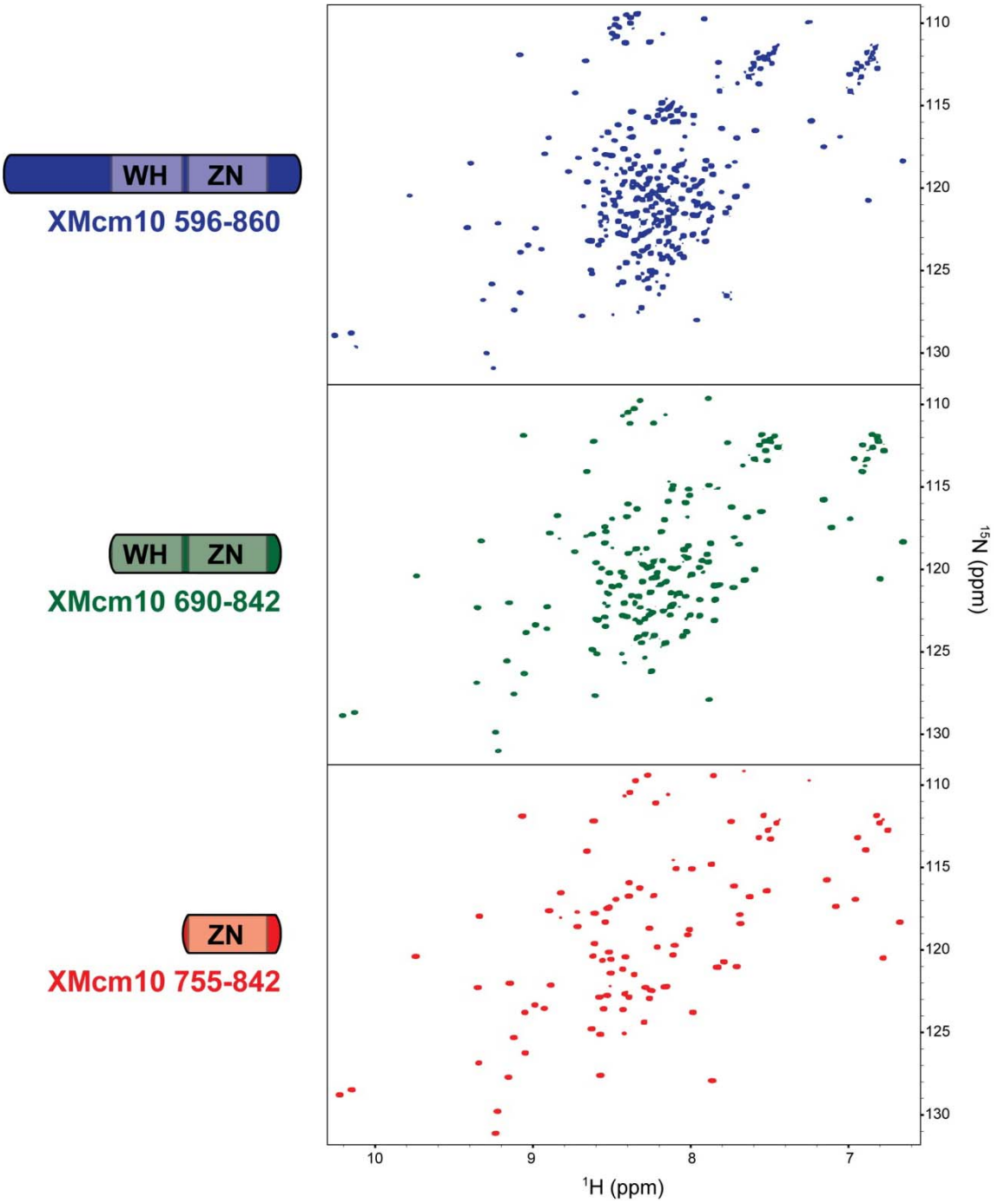


Figure 17. Heteronuclear 2D NMR Analysis of the CTD**. (**Left) Schematics of the three CTD constructs: XMcm10$^{596\text{-}860}$ (top, blue), XMcm10$^{690\text{-}842}$ (middle, green), and XMcm10$^{755\text{-}842}$ (bottom, red). Regions containing the putative winged helix and zinc-coordination are labeled as WH and ZN, respectively. (Right) Individual ^{15}N-^{1}H HSQC spectra for each of the CTD constructs displayed on the left side plotted in the corresponding color.

Table 5. Structural Statistics for XMcm10$^{690-842}$

Residues	690-840	760-840
Distance Limits		
Total	2009	2009
Intra, |i-j|==0	518	518
Sequential, |i-j|==1	686	686
Medium-range, 1<|i-j|<5	367	367
Long-range, |i-j|>=5	438	438
Average assignments/constraint	1.00	1.00
Average target function value	32.12	32.12
Average backbone RMSD to mean	12.84	1.78
Average heavy atom RMSD to mean	12.78	2.36

N-terminal 70 residues of the protein (Table 5). TOCSY and NOESY experiments were repeated and all initial spectra were recollected to try and obtain more assignments and NOE restraints for the dynamic alpha helical region. However, overlapping signals within the spectra prevented us from obtaining the level of assignments necessary to produce a high quality structure. While the structure of XMcm10$^{690-842}$ was not high enough quality for our purposes, it did yield some structural insights. For example, residues 760-835 conform to a well-folded structure, while the mostly α-helical N-terminus was observed to be dynamic in the ensembles. The high level of variability observed for the N-terminal region of the construct could be due to limited NOE restraints. However, another possible explanation is that this region is actually dynamic with respect to the rest of the construct. Heteronuclear NOE data presented in Figure 12A reveals indications of folded structure within the N-terminal half of XMcm10$^{690-842}$. However, the linker between the N- and C-terminal sections has low NOE values, suggesting it is relatively flexible. These data support the hypothesis that the N-terminal region of XMcm10690^{-842}, while it is folded, is highly dynamic in solution. This information was utilized to design the XMcm10$^{755-842}$

construct whose structure was presented in chapter 2. ^{15}N-^{1}H HSQC of XMcm10$^{755-842}$ contained all of the dispersed signals observed in the XMcm10$^{596-860}$ and XMcm10$^{690-842}$ spectra (Figure 17), indicating that the structure of residues 755-842 is consistent in each of the three constructs.

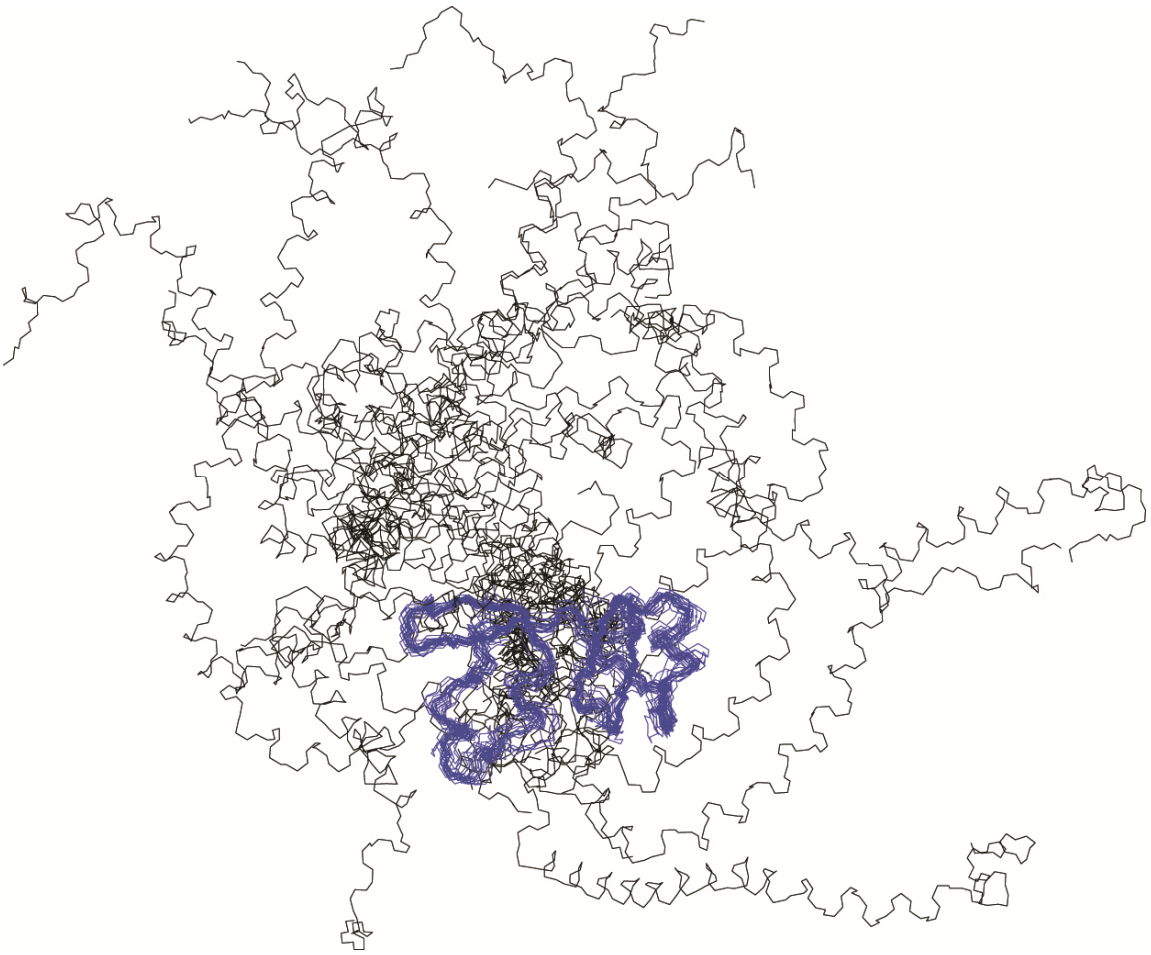


Figure 18. Solution NMR Structure of XMcm10$^{690-842}$. An ensemble of the twenty backbone conformers of XMcm10$^{690-842}$ with the lowest RMSD and target function values (CYANA). Residues 755-835, for which the solution structure was presented in Chapter 3, are colored blue.

Locating the Pol α Binding Site on XMcm10-CTD

Initially, Mcm10 was believed to participate only in pre-RC activation in early initiation. However, subsequent studies revealed that Mcm10 had a second role in DNA replication initiation that involved polymerase α. In addition to physically interacting with pol α, Mcm10 also stimulated its polymerase activity and prevented its degradation *in vivo* (Fien et al. 2004; Ricke et al. 2004; Ricke et al. 2006). The Mcm10 interaction has been mapped to the N-terminal 323 residues of p180, the catalytic DNA polymerase subunit (Fien et al. 2004; Robertson et al. 2008). In addition, the pol α interacting region on Mcm10 has been localized to two sites, one within the ID and a second site within the CTD (Robertson et al. 2008). An in-depth analysis of the ID defined its interaction to the cleft of the Mcm10 OB-fold and residues 286-310 of p180 (Warren et al. 2009). While these results clarified the molecular basis for ID binding to pol α, there has been no further investigation for the interaction with the CTD. In order to fully understand the details of Mcm10-pol α interaction, we pursued in-depth characterization of binding by the CTD.

The affinity chromatography experiments presented in chapter 2 showed XMcm10-CTD is sufficient to interact with pol α, specifically through the N-terminal 323 residues of the p180 (p180N). After the initial characterization of this interaction, a structural investigation of p180N was initiated. The gene fragment was transferred into an expression vector and large amounts of the protein were purified for structural studies. However, the resulting protein product was not homogenous after visualization by SDS-PAGE, sand instead of a single purified product after gel filtration chromatography, four distinct species of similar mass were observed for p180N. These bands were excised from

the gel and analyzed by MALDI-TOF mass spectrometry for identification. All of the bands were confirmed to be p180 fragments, yet none of them represented the intact p180N (residues 1-323). After careful inspection of the mass spectrometry data and comparison with the predicted secondary structure of p180N, a cleavage site was identified around residue 188. As such, two constructs, $p180^{1-188}$ and $p180^{189-323}$, were designed. In addition, studies in the Fanning laboratory suggested the design of a third construct: $p180^{1-145}$. Each of these constructs was purified and tested for interactions with CTD by the Fanning laboratory using affinity chromatography. While $p180^{189-323}$ showed no signs of interaction with the CTD, $p180^{1-188}$ interacted weakly and $p180^{1-145}$ showed a strong interaction. As a result, a further biochemical and structural characterization of the interaction between XMcm10-CTD and $p180^{1-145}$ was initiated.

The $p180^{1-145}$ construct was overexpressed and purified using affinity, ion exchange and gel filtration chromatography. The elution profile for $p180^{1-145}$ from gel filtration column indicated an approximate mass that was double the expected 17 kDa. Since there is no report of a p180 self-interaction in the literature, the most likely interpretation was that $p180^{1-145}$ has an elongated or non-globular shape, which affected its retention time on the gel filtration column. Purified $p180^{1-145}$ was also analyzed by circular dichroism (CD) to identify secondary structural elements. The results showed the construct to be composed mainly of random coil (~70%) and a low amount of alpha helices (~20%), which was in close agreement with secondary structure predictions. Taken together, these data support a mostly random coil, extended structure for $p180^{1-145}$.

To further probe the interaction between the XMcm10-CTD and $p180^{1-145}$, fluorescence anisotropy experiments were performed to determine the binding affinity.

For these experiments, a fluorescent moiety was conjugated to $p180^{1-145}$. The sample was prepared by incubating $p180^{1-145}$ with an excess of MTS-fluorescein (2-[(5-fluoreceinyl) aminocarbonyl] ethyl methanethiosulfonate) for 6 hours at room temperature. Due to the presence of only one cysteine residue, Cys141, in can be assumed that the moiety was linked to this position. Elution from a 1-ml Q-Sepharose column yielded pure fluorescently-labeled $p180^{1-145}$. Next, a change in fluorescence anisotropy was monitored for fluorescein-$p180^{1-145}$ while increasing amounts of $XMcm10^{596-860}$ were titrated in. No significant change in anisotropy was observed with the addition of $XMcm10^{596-860}$, suggesting that the two proteins were not interacting under these conditions. However, the absence of an appropriate positive control for this binding assay makes these results difficult to interpret, and thus an interaction cannot be excluded from these results.

As an alternative approach, isothermal titration calorimetry (ITC) was employed to measure heat changes resulting from interaction between $XMcm10^{596-860}$ and $p180^{1-145}$. Heat change (ΔH) was measured during a titration of buffer into a sample cell of only buffer to establish the heat of buffer mixing (Figure 19A), followed by a titration of $p180^{1-145}$ into a solution containing $XMcm10^{596-860}$ (Figure 19B). During the experimental titration, no significant changes in heat release were, once again indicating no interaction between the two proteins.

While both anisotropy and ITC experiments yielded negative results for an $XMcm10^{596-860}$ and $p180^{1-145}$ interaction, it was possible that the binding affinity of this interaction was outside of the detection limits of these two methods. Fluorescence anisotropy and ITC are most effective at measuring binding affinities in the nanomolar to millimolar range, making it possible that an extremely weak or transient interaction may

go undetected. As a final means to determine if these two constructs interacted, we performed a chemical shift perturbation experiment. Using a similar approach as described in chapter 3 with ssDNA, we observed the chemical shifts of ^{15}N-XMcm10$^{690-842}$ as p180$^{1-145}$ was titrated in. A total of 22 of the 148 signals were perturbed greater than 1 s.d. from the mean, 12 of which were greater than 1.5 s.d. All but one of the signals corresponded to residues within the zinc-coordinating region of XMcm10-CTD, suggesting the putative winged helix region does not participate in binding. Mapping the

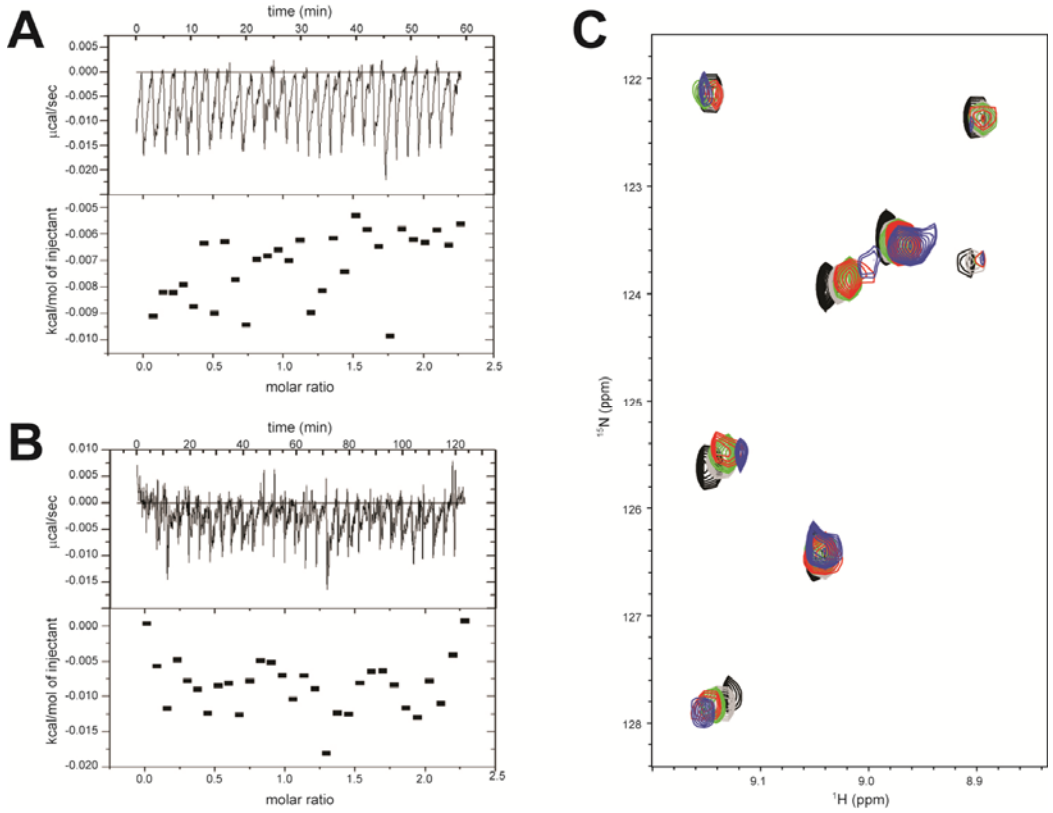


Figure 19. Experiments to detect the interaction of XMcm10-CTD and p180$^{1-145}$. Isothermal titration calorimetry data for (A) buffer into buffer and (B) p180$^{1-145}$ in to XMcm10$^{596-860}$. Upper panels contain the raw calorimetry data for each titration, while lower panels show the integrated heat change. (C) Representative region of the ^{15}N-^{1}H HSQC spectrum overlays of ^{15}N-XMcm10$^{690-842}$ in the presence of 0:1 (black), 0.5:1 (grey), 1:1 (green), 2:1 (red), 4:1 (blue) molar ratios of purified p180$^{1-145}$.

perturbed residues onto the structure of XMcm10$^{755-842}$ revealed that the majority of these peaks are the same as those perturbed by ssDNA binding (Figure 20). This would suggest that p180$^{1-145}$ and ssDNA compete for the same binding site on XMcm10-CTD. A competition with DNA and p1801-145 binding to Mcm10 has been observed with XMcm10-ID, consistent with the highly acidic region of p180 acting as a DNA mimic (Warren et al. 2009). While this hypothesis needs to be tested directly using CTD competition experiments with DNA

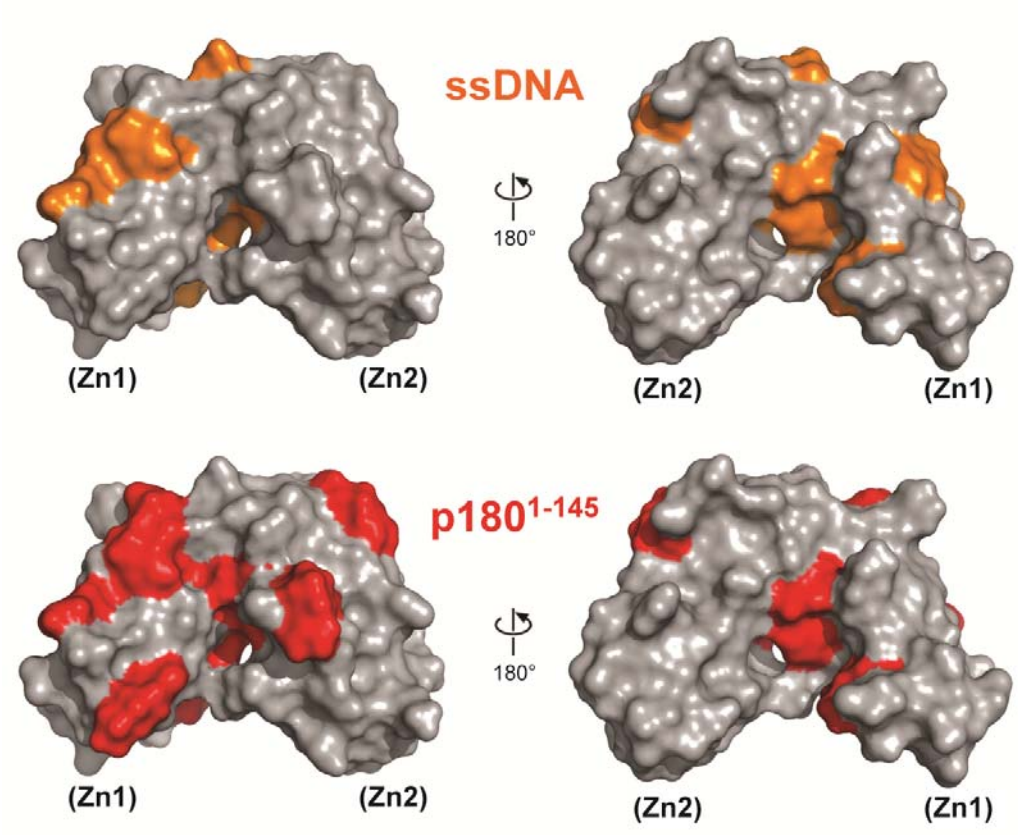


Figure 20. XMcm10-CTD binding sites for ssDNA and pol α. (Top) Residues perturbed by ssDNA binding from HSQC titration data are colored orange against the molecular surface of XMcm10$^{755-842}$. (Bottom) Residues perturbed by p180$^{1-145}$ binding to XMcm10$^{690-842}$ are colored red against the same molecular surface depicted in the top panel.

and pol α, the fact that ssDNA binds with a low micromolar affinity compared to $p180^{1-145}$, which appears to bind with millimolar or greater affinity suggests that DNA would be the preferred substrate in the event of competition. This would be similar to previous competition results observed with the ID (Warren et al. 2009), and may help support the model of Mcm10 exchanging pol α for exposed ssDNA at the replication fork.

Characterization of DNA Binding by XMcm10

Mcm10 associates with the origin at the G1/S transition and dissociates after replication prior to G2 (Izumi et al. 2000; Izumi et al. 2001). While this association is dependent upon prior ORC and Mcm2-7 loading *in vivo*, subsequent *in vitro* studies showed yeast Mcm10 was able to bind single and double-strand DNA independently through the ID (Homesley et al. 2000; Wohlschlegel et al. 2002; Fien et al. 2004). More recently, it has been shown that the yeast Mcm10 displays differential packing on single versus double-stranded DNA (Eisenberg et al. 2009). The authors showed that one molecule of Mcm10 bound 25-50 nt lengths of double-stranded DNA, yet the same span of ssDNA was able to be bound by ~3 Mcm10 molecules. It was suggested that a structural rearrangement or change in oligomerization was occurring with Mcm10 response to binding ssDNA (Eisenberg et al. 2009). While it is unclear as to the reason for this phenomenon, the authors suggest it may be linked to a role in initial origin melting, where duplex DNA is first separated into single strands.

In contrast to single DNA binding domain present in yeast Mcm10, vertebrate homologs are capable of binding DNA through its ID and/or CTD (Robertson et al. 2008). Further analysis of the ID showed ssDNA binds within the OB-fold cleft and

extends across the zinc finger, with a length of 5-10 nt being the preferred substrate as determined by fluorescence anisotropy (Warren et al. 2008). A similar investigation of length preference with the CTD has shown ~15 nt to be sufficient for ssDNA binding, and 20-25 nt to attain maximal binding affinity (Figure 21). In addition to fluorescence assays, DNA binding was also detected by electrophoretic mobility shift assays (EMSA).

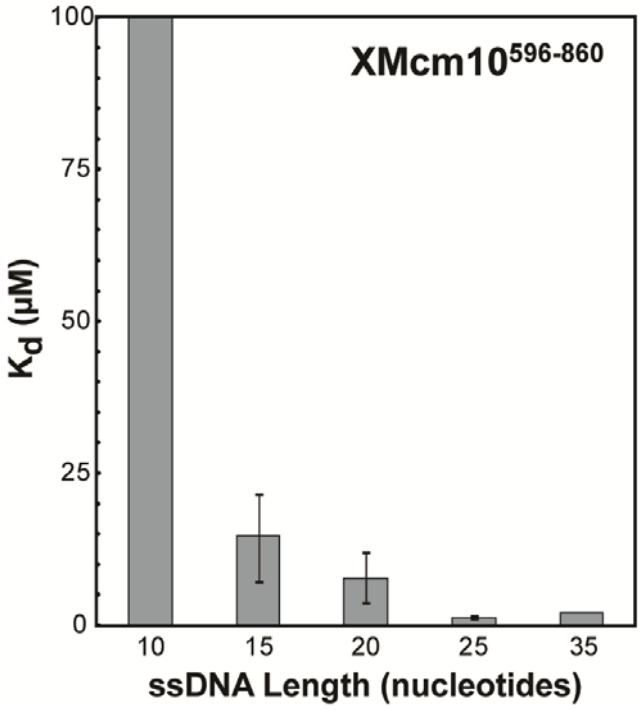


Figure 21. Length dependence for ssDNA binding to XMcm10$^{596-860}$. Plot of dissociation constants (K_d) for XMcm10$^{596-860}$ binding to various lengths of ssDNA measured by fluorescence anisotropy. Values are the averages and standard deviations from three independent measurements. DNA binding was measured using 50 nm FAM-DNA in 25 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM DTT, and 5% glycerol.

Under similar conditions as the ID experiments, CTD bound to radiolabeled 25mer ssDNA with ~3 μM affinity and showed a single super-shifted band, indicating a 1:1 stoichiometry for CTD binding a 25 nt length of DNA. Subsequent fluorescence

anisotropy experiments also confirmed the 1:1 stoichiometry for binding a 25mer. It is interesting to note that the length preference for ssDNA by the CTD is slightly longer compared to XMcm10$^{755-842}$ (Chapter 3). Despite the subtle change in length preference, the similar DNA binding profiles of these two constructs support the hypothesis that residues 596-755 play little to no role in binding DNA.

Having thoroughly characterized the ability of the ID and CTD to bind various DNA substrates individually, we next wanted to investigate how these two domains bound DNA in the context of a larger Mcm10 construct. The ID and CTD are separated by 169 residues in primary structure, which we previously determined to be a flexible linker tethering the two domains together (Chapter 3). In order to determine if this linker region had any effect on DNA binding, two different constructs were tested for ssDNA binding and length dependence. The first construct spanned the ID and CTD with the full linker in between (residues 230-860, XMcm10$^{230-860}$). The other contained the ID and CTD, but the linker region was replaced by a short 12 residue insertion (XMcm10-IC). These two constructs were then tested for the ability to bind different lengths of ssDNA by fluorescence polarization. Interestingly, removing the linker region reduced the maximal ssDNA binding affinity from 0.46 ± 0.2 μM for XMcm10$^{230-860}$ to 2.9 ± 0.3 μM for XMcm10-IC, a ~6-fold reduction. This suggested that the linker region, or the inherent flexibility it creates between the two domains, does affect the overall binding affinity for ssDNA. Full-length XMcm10 has been previously shown to have a 0.12 ± 0.02 μM affinity for ssDNA (Robertson et al. 2008), a modest increase compared to XMcm10$^{230-860}$. While these affinities are similar, suggesting that the NTD does not

affect DNA binding, a more thorough investigation is needed before any conclusions can be made.

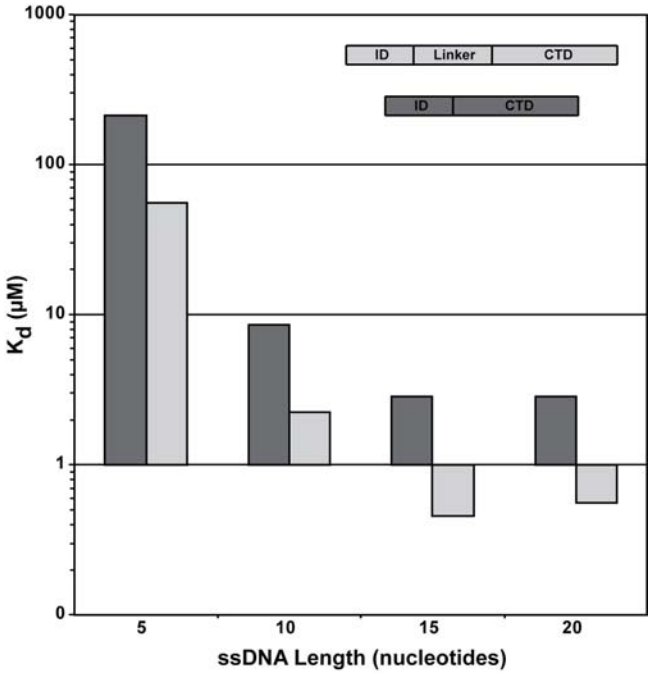


Figure 22. Effect of the linker region between ID and CTD on DNA binding. Plot of dissociation constants (K_d) for XMcm10$^{230-860}$ (light gray) and XMcm10-IC (dark grey) binding to various lengths of ssDNA measured by fluorescence polarization. A logarithmic scale has been used on the y-axis in order to better depict the differences in affinity. Values are the averages from three independent measurements.

Despite a change in affinity, both constructs appeared to attain maximum binding affinity for a length in between 10-15 nt which suggests the minimal length of DNA needed to span both binding sites is similar with and without the linker. It is also interesting that ID and CTD individually prefer 10-15 nt for binding, yet this same length of DNA is able to bind XMcm10$^{230-860}$ with higher affinity (Figure 22). This observation can be explained through one of two possible scenarios (Figure 23). First, the ID and CTD could both bind to the same region of ssDNA, sandwiching the DNA in between the

two binding sites (Figure 23, part A). However, this scenario is unlikely due to steric collisions that would result from these two domains being positioned in such close proximity on a single strand of DNA. The second possibility is that these two domains stack end to end on the ssDNA, perhaps to the point of overlapping (Figure 23, part B). This would allow the DNA to span both domains without causing any steric clashing. In support of this hypothesis, reducing the linker affected the maximum binding affinity of Xmcm10-IC. Perhaps the extended flexible linker is required in order for these two domains to position themselves properly to create a cooperative high affinity binding site.

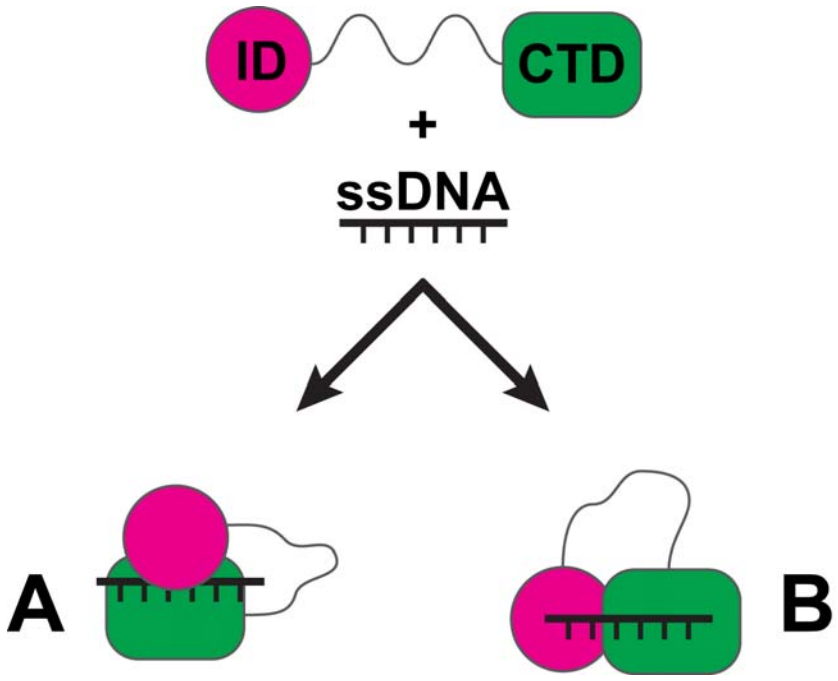
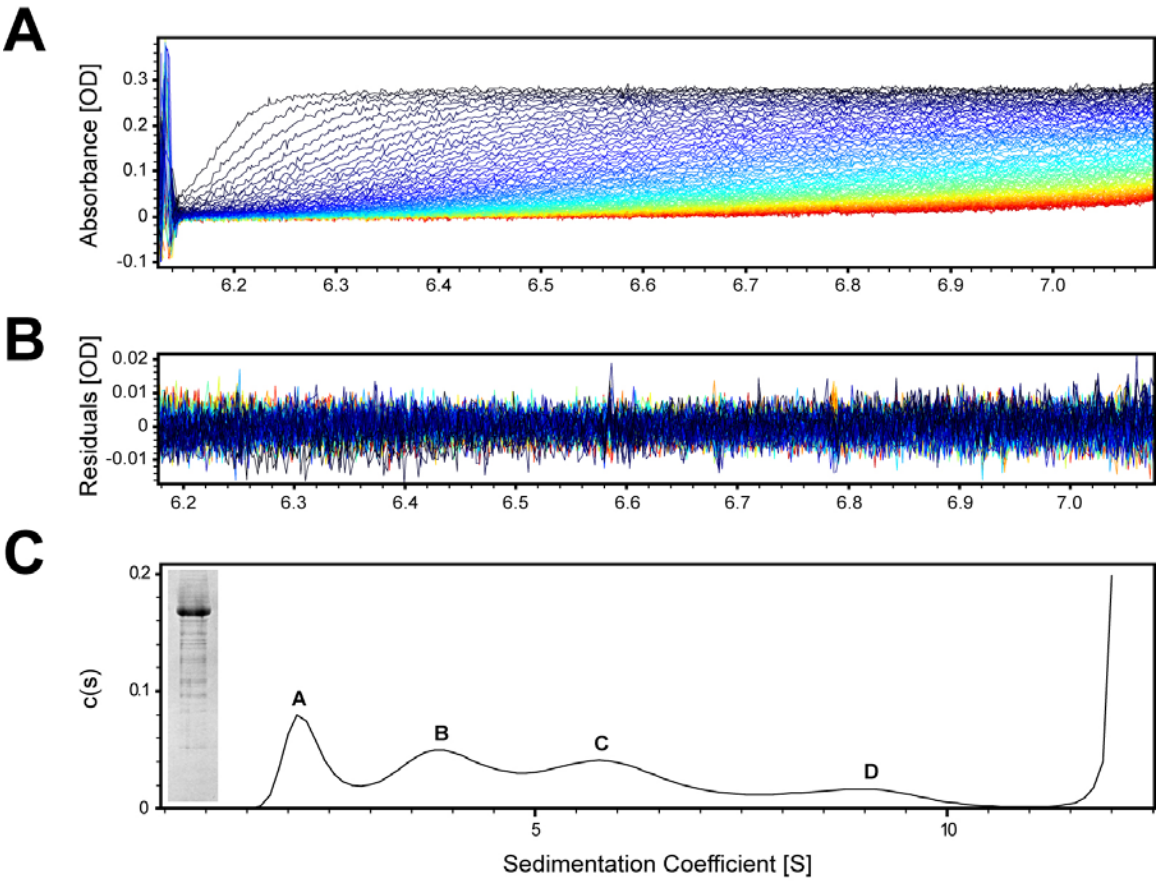


Figure 23. Possible mechanisms for DNA binding by Mcm10-ID and -CTD. Illustration of ssDNA binding by the ID (magenta) and CTD (green) of Mcm10, connected by its flexible linker (grey). Part A shows the two domains sandwiched together with the ssDNA in between. Part B shows the two domains stacked end-to-end with the ssDNA binding across both domains. Domain and linker size are not to scale.

Determining the Oligomerization State of XMcm10

The self-interaction of Mcm10 has been investigated by a number of different studies, yet the oligomeric state still remains controversial. An early study performing pull-downs with *Drosophila* Mcm10 first showed that Mcm10 is capable of self-association (Christensen et al. 2003). This observation was further supported by a subsequent study in *S. cerevisiae* which showed quantitatively that Mcm10 forms a dimer, and that the ID is sufficient for this self-interaction (Cook et al. 2003). A similar study in *S. pombe* also showed Mcm10 to form dimers, but their results showed the interaction region at the N-terminus (Fien et al. 2006). In contrast to yeast data, a separate study using human Mcm10 was able to show the formation of a ring-shaped hexamer by electron microscopy (EM), which was supported with gel filtration data (Okorokov et al. 2007). As presented in chapter 2, our initial biochemical analysis of *Xenopus* Mcm10 showed signs of oligomerization. However, the limited data supported a sedimentation profile most similar to a dimer-sized complex. In addition, the identification of a putative coiled-coil motif within the NTD led to the hypothesis that this was the region of self-interaction.

In order to determine the oligomeric state of XMcm10, we have performed a preliminary analysis by analytical ultracentrifugation (AUC). Previous studies using this approach with full-length protein were limited by poor stability of the sample. Purification from *E. coli* cells yielded low amounts of protein that was very susceptible to spontaneous degradation and required high salt and glycerol to help stabilize the protein. This was not conducive to the need for minimal buffer conditions or the time frame which included shipping the samples to a collaborator for analysis. We have since



D

Peak	A	B	C	D
S Value	2.291	3.989	6.088	8.950
Best-fit MW	54.7	125.7	237.0	422.4
Minimum MW	16.0	36.8	69.4	124.0
% Total Signal	20.3	30.0	31.0	13.4

Figure 24. Determining the oligomeric state of XMcm10 by AUC. (A) Plot of 200 data sets obtained from a sedimentation velocity experiment at 30,000 RPM using absorbance optics and XMcm10 at a concentration of 0.5 O.D. (B) Residuals obtained from the line fits for the data in panel A. (C) Distribution plot showing the sedimentation coefficients for the four predominant species in solution. *Inset*, the XMcm10 sample tested by AUC being visualized by SDS-PAGE. (D) A table listing information for the individual species pertaining to S value, predicted molecular weight (MW) and the percentage of area for that peak compared to the overall plot.

switched to insect cells for overexpression of full-length XMcm10. This has increased the overall protein yield and improved stability issues from our previous prokaryotic expression system. Sedimentation velocity experiments were performed in-house using three different concentrations of XMcm10 in a solution of 25 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5 mM BME. (Figure 24). While this data represents a large improvement in quality compared to previously reported results, it remains insufficient to determine the oligomeric state of XMcm10. Additional optimization of experimental parameters, including buffer components, needs to be completed in order to obtain clear and concise data. While data presented in Figure 24 has no conclusive results, initial interpretation of the data suggests that there were four predominant species in the sample. Although XMcm10 has a predicted molecular weight of 91.5 kDa, no peak was observed corresponding to this mass. However, the data analysis software assumes the proteins are somewhat globular when fitting the data. If Mcm10 adopts an extended, "beads on a string" architecture, this would result in a sedimentation profile far different than those observed with more compact proteins. The high frictional ratios calculated during data analysis support this - the best fits were obtained using values beyond 2.2, which are well beyond the normal range of 1.2-2.0. If the protein is flexible and extended in solution, this would create a large amount of drag during sedimentation, thus causing an increase in the frictional coefficient. In order to confirm this hypothesis, additional analysis by AUC, as well as investigations using small angle x-ray scattering (SAXS) and multiangle light scattering (MALS) will be necessary. The latter two methods will confirm the presence of an extended, flexible structure and provide the overall size and shape of XMcm10. These results can then be used to optimize AUC data collection and analysis

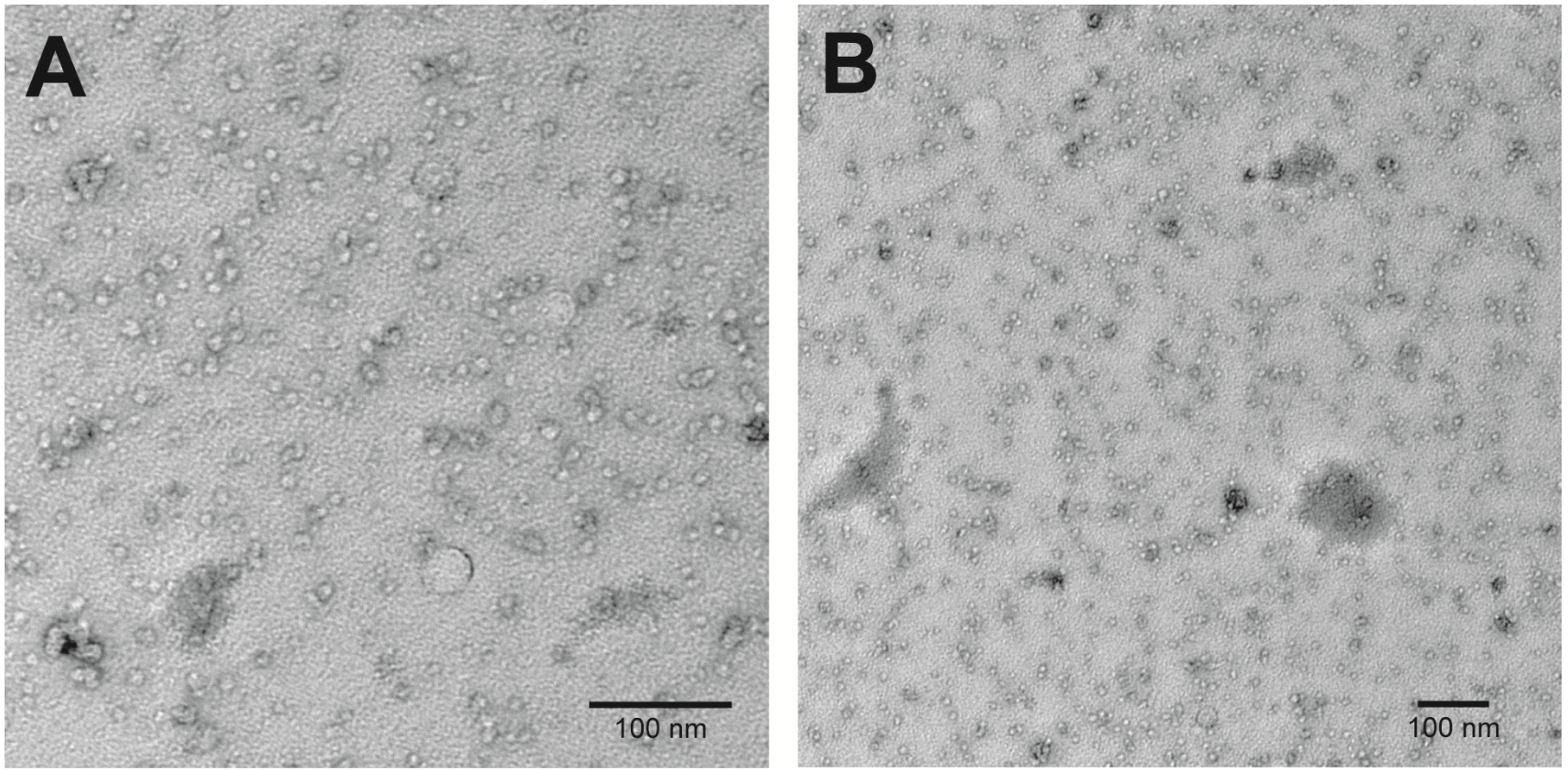


Figure 23. Electron micrographs of XMcm10. Two representative micrographs of XMcm10 obtained by negative stain EM, one view is a close-up (A) and one has been slightly expanded (B). Scale bars for each image are in the bottom right corner.

parameters, allowing for a more exact determination of the oligomeric state. As mentioned previously, human Mcm10 has been shown to form hexameric rings by negative stain EM (Okorokov et al. 2007). These ring structures measured 160 X 120 Å and contained a 25 Å wide pore located in the center of the ring. If XMcm10 formed a similar hexameric complex, it too would be of an appropriate size to be visualized using EM. To investigate whether XMcm10 also forms hexameric rings, we analyzed full-length protein by negative stain EM in collaboration with the Ohi lab using the same buffer as AUC experiments (Figure 25). Most of the identifiable particles were amorphous in shape and had an apparent molecular weight of < 200 kDa. Despite scanning multiple grids with proteins from multiple purifications in a variety of different buffers, no consistent ring structures were observed in any of the images collected. While the lack of rings suggests XMcm10 does not form hexamers, further studies with additional buffers will be required to confirm this observation. In addition, results from AUC analysis will also help to confirm or deny the existence of hexameric species of XMcm10.

Experimental Procedures

XMcm10-IC Expression and Purification

XMcm10-IC was overexpressed in *E. coli* Rosetta cells in LB medium supplemented with 100 µg/ml ampicillin, 7.5 µM $ZnSO_4$, and 0.5 mM IPTG for 18 hours at 16 °C. Cells were resuspended in 50 mM Tris-HCl at pH 7.5, 500 mM NaCl, and 10% glycerol (Buffer L) and lysed with an EmulsiFlex-C3 homogenizer (Avestin). Cell lysates were centrifuged at 35,000 x g for 20 minutes. Trx-His_6-Mcm10 proteins were purified

from the supernatant by nickel-nitrilotriacetic (Ni-NTA, Qiagen) affinity chromatography using a Buffer L wash and Buffer L/250 mM imidazole elution. Fractions were visualized by SDS-PAGE and Coomassie Blue staining. Those containing Mcm10 were pooled and incubated overnight at 4 °C with PreScission Protease (GE Healthcare) at a 1:50 protease:Mcm10 mass ratio to remove the Trx-His_6 affinity tag. After cleavage, this mixture was diluted 5-fold in 50 mM Tris-HCl at pH 7.5, and 10% glycerol (Buffer A) and loaded onto a 5-ml SP Sepharose HP cation exchange column (GE Healthcare). Proteins were eluted using a linear gradient from Buffer A/0.1 M NaCl to Buffer A/1 M NaCl. Fractions containing Mcm10 proteins were pooled, concentrated using an Amicon spin concentrator (Millipore) and purified over a 320-ml Superdex 200 gel filtration preparative grade column (GE Healthcare) that had been equilibrated in 20 mM Tris-HCl at pH 7.5, 200 mM NaCl, 2 mM β-mercaptoethanol, and 10% glycerol.

XMcm10 Expression and Purification

XMcm10-IC was overexpressed using baculovirus infection of Sf-21 insect cells. Cells were resuspended in 50 mM Tris-HCl at pH 7.5, 500 mM NaCl, and 10% glycerol (Buffer L) and lysed with an EmulsiFlex-C3 homogenizer (Avestin). Cell lysates were centrifuged at 35,000 x g for 20 minutes. His_6-XMcm10 was purified from the supernatant by nickel-nitrilotriacetic (Ni-NTA, Qiagen) affinity chromatography using a Buffer L wash and Buffer L/250 mM imidazole elution. Fractions were visualized by SDS-PAGE and Coomassie Blue staining. Those containing Mcm10 were pooled and diluted 3.33-fold in 50 mM Tris-HCl at pH 7.5, and 10% glycerol (Buffer A) and loaded onto a 10-ml Source-Q anion exchange column (GE Healthcare). Proteins were eluted

using a 50mM steps from Buffer A/0.15 M NaCl to Buffer A/0.3 M NaCl, followed by a linear gradient to Buffer A/1 M NaCl. Fractions containing Mcm10 proteins were pooled, concentrated using an Amicon spin concentrator (Millipore) and purified over a 175-ml Superose 6 gel filtration preparative grade column (GE Healthcare) that had been equilibrated in 25 mM Tris-HCl at pH 7.5, 150 mM NaCl, 2 mM DTT, and 5% glycerol.

NMR Spectroscopy

NMR data were collected at 25 °C using Bruker DRX600 and DRX800 spectrometers equipped with cryoprobes. Data were processed with Topspin 2.0b and further analyzed using Sparky (Goddard et al. 2006). Samples were buffer exchanged using an Amicon Spin Concentrator (Millipore) into 25 mM NaH_2PO_4 at pH 6.5 in 90% H_2O/10% D_2O and 100 mM NaCl. $XMcm10^{596-860}$ was concentrated to 205 μM, $XMcm10^{690-842}$ was concentrated to 190 μM, and $XMcm10^{755-842}$ was concentrated to 300 μM for HSQC titrations.

The structures of $XMcm10^{690-842}$ were determined using CYANA (Guntert 2004). Seven iterative cycles of calculations were carried out starting with a set of manually assigned NOEs. 100 structures were created per cycle with backbone torsion angle restraints obtained from ^{1}H, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, and ^{15}N chemical shifts using TALOS with a minimum range of ±35°. The 20 structures with lowest values of the CYANA target function are displayed in Figure 18.

Pol α Binding

Protein samples for ITC analysis were buffer exchanged into 25 mM Tris (pH 7.5) and 100 mM NaCl and concentrated to 74 μM ($XMcm10^{596-860}$) and 741 mM ($p180^{189-323}$). A total of 1.75 ml $XMcm10^{596-860}$ was placed into the sample cell and $p180^{189-323}$ was added introduced using consecutive injections of 6 μl. Data were collected at 25 ºC using a MicroCal VP-ITC.

NMR titrations were performed at 25 °C using a Bruker DRX800 spectrometer equipped with a cryoprobe. Data were processed with Topspin 2.0b and further analyzed using Sparky (Goddard et al. 2006). $XMcm10^{690-842}$ and $p180^{1-145}$ were buffer exchanged using an Amicon Spin Concentrator (Millipore) into 25 mM NaH_2PO_4 at pH 6.5 in 90% H_2O/10% D_2O and 100 mM NaCl. $XMcm10^{690-842}$ was concentrated to 190 μM and $p180^{1-145}$ was concentrated to 940 μM. Chemical shift perturbation experiments were performed using Mcm10:p180 molar ratios of 1:0, 1:0.5, 1:1, 1:2, and 1:4. Perturbations were quantified using the equation $\Delta\delta_{ave} = (((\Delta\delta_{1H})^2 + (\Delta\delta_{15N}/5)^2)/2)^{1/2}$. Values of $\Delta\delta_{ave}$ greater than 1.5 standard deviations from the mean are considered significant.

DNA Binding

Mcm10 DNA binding affinities for were measured by following the increase in fluorescence polarization/anisotropy as the protein was added to DNA oligonucleotides labeled at the 3′- end with 6-carboxyfluorescein (FAM, DY). The procedure used for obtaining $XMcm10^{596-860}$ binding data is described in chapter 3. Oligonucleotides used to determine the length dependence of DNA binding are shown in Appendix Figure B5. Mcm10 proteins were added over a concentration range of 0-20 μM to a solution of 50

nM fluorescein-DNA, and polarized fluorescence intensities were measured at excitation and emission wavelengths of 495 and 538 nm, respectively. Reactions using XMcm10$^{230-860}$ and XMcm10-IC were performed at 25 °C in 25 mM Tris-HCl at pH 7.5, 150 mM NaCl, and 5% glycerol. Each measurement was made in triplicate and the dissociation constants (K_d) were calculated by fitting the binding curve using a two-state binding model using Kaleidagraph 3.6 (Synergy Software).

Analytical Ultracentrifugation

Purified XMcm10 was buffer exchanged using an Amicon Spin Concentrator (Millipore) into 25 mM Tris-HCl at pH 7.5, 150 mM NaCl, 10 mM MgCl, and 5 mM BME. The sample was diluted to 1 O.D. at 280 nm and 390 µL was loaded into one side of a sample cell equipped with quartz lenses, along with 400 µL of buffer in the adjacent compartment. The cell, rotor and detector were equilibrated to 4 °C for ~4 hours and then analyzed overnight at 30,000 rpm using absorbance optics. A total of 251 scans were obtained and analyzed using SEDFIT (Schuck 2000).

Electron Microscopy

Purified XMcm10 was applied to glow-discharged carbon-coated copper grids, stained with uranyl formate (0.7% wt/vol), and analyzed using a Philips Tecnai T12 electron microscope at an acceleration voltage of 80 kV with varied magnifications of x36,000 (Figure 23A) and x18,000 (Figure 23B).

CHAPTER V

DISCUSSION AND FUTURE DIRECTIONS

The Role of Mcm10

Despite decades of research, the biological function of Mcm10 during replication initiation remains elusive. The current compilation of Mcm10 literature reveals important and complex interactions with three other replication factors: the Mcm2-7 helicase, DNA polymerase α, and DNA. Each of these three interactions will be discussed individually prior to a more global perspective on Mcm10 function. As an aid for this discussion, Figure 26 shows a schematic of Mcm10 with all regions of interaction mapped onto it for reference.

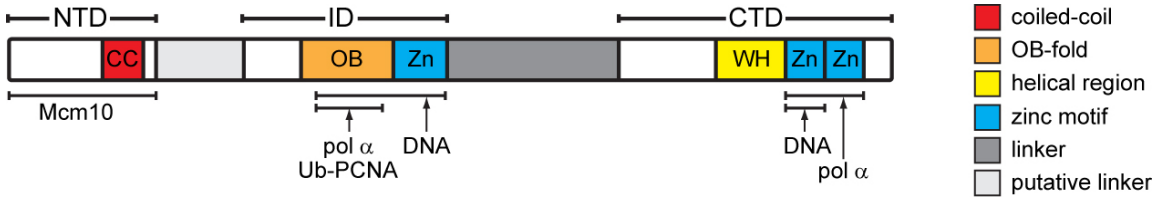


Figure 26. Schematic diagram of vertebrate Mcm10 showing the domain organization and regions of interaction. Domains are labeled on top of the diagram, while regions of known interaction are listed below. Regions of known or predicted structure are labeled and colored according to the legend on the right side. Information shown here is from a combination of my work and additional sources (Das-Bradoo et al. 2006; Warren et al. 2008; Warren et al. 2009).

Initially, interactions with components of the pre-RC and the need for Mcm10 in the loading of downstream replication factors implicated a distinct role for Mcm10 in early DNA replication initiation. The first role of Mcm10 is to mediate an interaction

between Mcm2-7 and DDK, which subsequently results in the phosphorylation of Mcm2-7 at the onset of S phase. In addition, Mcm10 has also been shown to be required for the assembly of the CMG complex, which is hypothesized to occur through its interaction with Mcm2-7 since it has not been shown to interact directly with Cdc45 or GINS. Conversely, the need for Mcm10 in CMG formation may be due to the upstream phosphorylation of Mcm2-7 as opposed to a direct interaction with the complex, however this remains to be investigated. Mcm2-7 loads onto DNA as a head-to-head double hexamer, similar to MCM in archaea. It is hypothesized that the two hexamers dissociate upon origin firing resulting in two bidirectional replication forks. If this hypothesis is correct, then at some point during initiation the Mcm2-7 hexamers will uncouple and thus expose a ssDNA bubble as they separate from one another. Given the interactions between Mcm10 with both Mcm2-7 and ssDNA, this separation of hexamers may be mediated by or require Mcm10. In support of such a mechanism, the interaction between the two hexamers of Mcm2-7 is partly mediated by a zinc ribbon motif reported here to be structurally homologous to the CCCC motif in Mcm10-CTD. It is therefore possible that Mcm10 mimics the zinc ribbon of the adjacent Mcm2-7 hexamer and establishes a very similar interaction. This new interaction could then stimulate hexamer dissociation and spatially allow for the establishment of bidirectional replication forks. While this is an attractive possibility, the lack of this CCCC motif in yeast Mcm10 implies that additional or alternative mechanisms may exist in lower eukaryotes.

In addition to its early interaction with Mcm2-7, Mcm10 is reported to have another role later on in initiation that persists into elongation and involves polymerases. Importantly, Mcm10 interacts with pol α and prevents its degradation by the proteosome,

thus stabilizing it throughout the cell cycle. In addition to stabilization, Mcm10 appears to function as a cofactor for pol α activity, increasing both its affinity for DNA as well as its primase activity. While current data supports the importance of the Mcm10/pol α interaction in DNA replication, a clear picture of how Mcm10 contributes to pol α activity and DNA binding during elongation is still lacking. During elongation, Mcm10 travels with the replication fork. Experiments in which polymerases were uncoupled from the replisome using a small molecule inhibitor did not affect Mcm10's interaction with polymerases. However, it did cause the dissociation of Mcm10 from the helicase complex, suggesting that Mcm10 has a higher affinity for pol α than it does for Mcm2-7. Data presented here refined the Mcm10-pol α interaction to the ID and CTD of Mcm10 and a short N-terminal region of pol α. The dual binding sites on Mcm10, along with its modular architecture described here, suggests two possible mechanisms for pol α loading onto DNA. First, Mcm10 may coordinate pol α loading between its two domains using a hand-off mechanism in which one domain is DNA-bound and the other recruits pol α. The other possibility is that the two independent domains of Mcm10 adopt a more rigid conformation upon binding pol α, aligning its two DNA binding motifs which can then displace pol α for DNA. However, the latter mechanism supports a role for Mcm10 in recruiting pol α to DNA, as opposed to the former which suggests Mcm10 maintains an interaction pol α and DNA. Overall, given the importance of Mcm10 in the activity and stabilization of pol α, the hand-off mechanism is more compatible with the current state of Mcm10 literature.

In addition to pol α, Mcm10 also interacts with the helicase complex. This interaction has led to the hypothesis that Mcm10 acts as a molecular tether, physically

bridging the DNA unwinding with the polymerase activities of the replisome. This role for Mcm10 has classically been viewed as a distinct function, separate from its earlier role in pre-RC activation. However, one may speculate that these roles are more linked than originally thought. It is possible that as Mcm2-7 double hexamers dissociate upon activation, Mcm10 could bind and facilitate this separation (Figure 27). In support of this hypothesis is the observation that Mcm10 directly interacts with DDK (Lee et al. 2003), the kinase that phosphorylates and presumably "activates" the helicase. In addition, Mcm10 interacts directly with Mcm2-7 through its ID and CTD, which points toward two possible mechanisms for this first step. Mcm10 could directly recruit DDK to Mcm2-7, allowing phosphorylation to occur. Or, DDK phosphorylation could expose a binding site for Mcm10 on Mcm2-7. While both are possible, the fact that DDK is recruited to the pre-RC in G1, before Mcm10 which binds in S phase (Izumi et al. 2001; Lei et al. 2001), would suggest that the latter is more likely. However, both of these possibilities would yield a similar result in which the Mcm2-7 complex is phosphorylated and becomes Mcm10-bound. It is interesting to note that loading of the helicase cofactors, Cdc45 and GINS, onto Mcm2-7 is dependent on Mcm10 (Im et al. 2009).

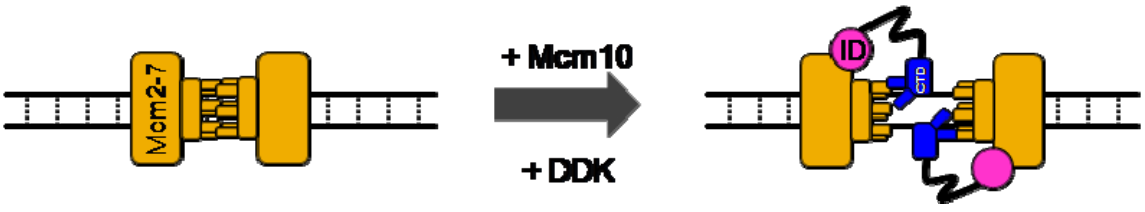


Figure 27. Hypothetical model of disruption of the Mcm2-7 double hexamer by vertebrate Mcm10. Mcm10 and/or phosphorylation by DDK may act to disrupt the association of the Mcm2-7 double hexamer, exposing the NTD CCCC motifs and allowing for an interaction with Mcm10-Zn2.

This suggests that the Mcm10-dependent phosphorylation of Mcm2-7 may induce a conformational change in Mcm2-7 that somehow facilitates Cdc45 and GINS binding. However, it is unclear whether it is the phosphorylation event or Mcm10 binding, or both, that limits CMG formation.

One final player implicated in Mcm10 function during DNA replication is ssDNA. Yeast Mcm10 stays constitutively bound to chromatin throughout the cell cycle, while vertebrate homologs appear to preferentially bind during S phase. Regardless of when binding occurs, this interaction is necessary for proper replication and appears to be a significant component of Mcm10 function. As observed with pol α, DNA binding occurs through both Mcm10-ID and -CTD. While these two domains are structurally distinct and spatially independent of one another, data presented here indicates that flexibility between these domains plays a role in DNA binding. Indeed, we propose that these two domains come together closely and stack end to end on DNA (Figure 23B), aligning their binding clefts to create a cumulative higher affinity site. While the preliminary data presented here supports this model, additional investigation to support this hypothesis is required. While the exact mechanism of DNA binding by these two domains remains unresolved, we firmly establish that Mcm10 is a bona fide ssDNA binding protein which is an important first step in understanding its possible role(s) during DNA replication.

Access to single-strand DNA is limited at the replication fork to prevent damage or breakage, so any interaction with it is likely purposeful. ssDNA first becomes available as the origin is initially unwound. As Mcm10 is the first ssDNA binding protein recruited to origins, it is therefore probable that Mcm10 is first to access these exposed

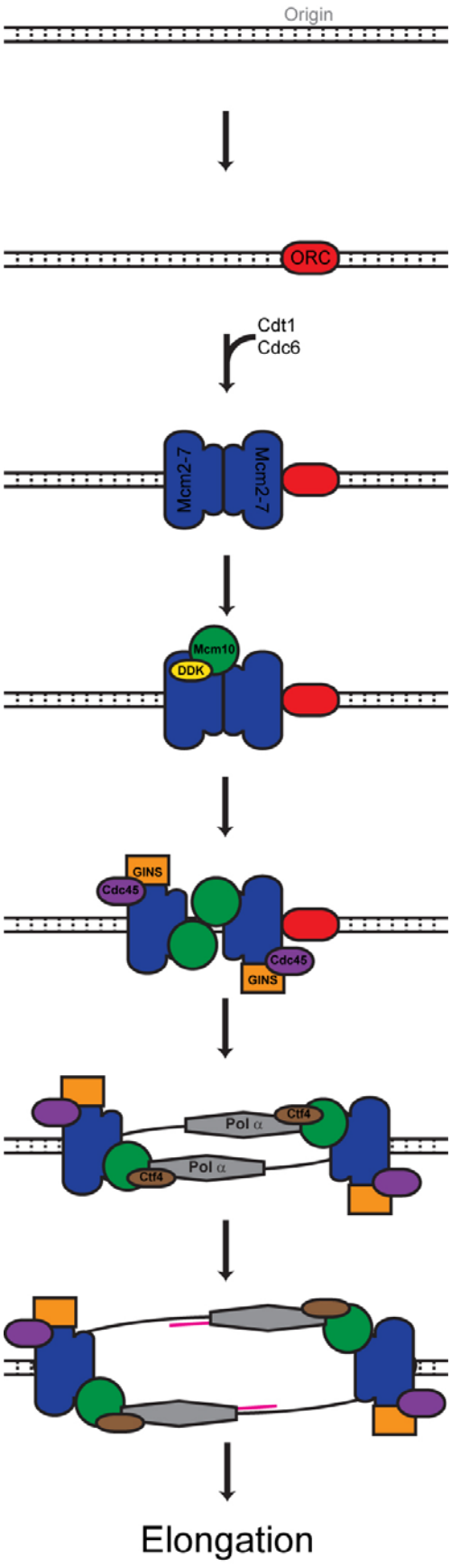


Figure 28. A theoretical model for Mcm10 function during DNA replication initiation. Illustrated is a general order of recruitment of proteins to the replisome.

regions of DNA. The CMG complex has been shown to function singly *in vitro* (Ilves et al. 2010). This suggests that the two Mcm2-7 complexes, although loaded as a double-hexamer, separate and travel in opposing directions after/concurrent with Cdc45 and GINS binding. Once the CMG complex initiates its unwinding activity, the two CMG complexes would move in opposite directions, exposing the single-strand origin DNA in between. This translocation would also spatially allow for Mcm10 binding to ssDNA, as well as establish interactions with Mcm2-7. Once Mcm10 becomes chromatin-bound, recruitment of pol α to the replisome in a manner dependent on Mcm10 and Ctf4 is now possible. It remains unclear how these two proteins cooperate to engage pol α, however both components are necessary for proper recruitment. Data presented in this work suggests that the Mcm10-CTD plays a direct role in this association. Incorporation of pol α into the replisome would then allow for primer synthesis and further elongation to occur. Mcm10's interactions with ssDNA, Mcm2-7 and pol α make it the ideal candidate for facilitating the creation and stabilization of an early replication bubble. Figure 28 is an illustration of the overall mechanism being hypothesized in this section.

Future Directions

The majority of the mechanisms proposed in this chapter are highly speculative. However, all are consistent with the current body of Mcm10 literature. Substantial work remains before the precise molecular function of Mcm10 is defined. The data presented here extends our understanding of Mcm10 by providing detailed biochemical and high-resolution structural data for the C-terminal domain of Mcm10. These data have further clarified the molecular interactions between Mcm10 and other replication components,

ensuring an excellent foundation for future experiments investigating biological function. The data presented in chapters 2 and 3 localized ssDNA and pol α binding to regions of the CTD. Analysis of CTD truncations or site-directed mutants *in vivo* will reveal any phenotypic consequence of losing either or both of these activities. In addition, these mutants may be useful for validating the hand-off mechanism previously described for Mcm10 and pol α loading because Mcm10 would need both functional domains in order to efficiently modulate both pol α and DNA binding.

Another interesting avenue of Mcm10 research is the further characterization of the structural architecture of Mcm10. Techniques like SAXS and MALS can be used to determine a spatial envelope for each of the three domains, as well as full-length Mcm10. Mcm10 is described here to possess a modular architecture, with the ID and CTD being structurally independent of one another. These data can be further verified and extended using SAXS analysis. Similarly, it remains unclear whether the modular architecture of Mcm10 is affected by DNA or pol α binding. Using these techniques, it is possible to determine if the envelope of Mcm10 in the presence of these binding partners changes to a more rigid and globular entity. In contrast, if the envelope is unaffected by ligand binding, this would be more supportive of a modular mechanism where ID and CTD retain their mobility with respect to one another, even when bound to DNA or pol α .

Significant effort has been invested in determining the 3-dimensional structure of Mcm10 by crystallography. Unfortunately, due to stability issues and buffer constraints, these attempts have been unsuccessful. Due to the extensive interaction between Mcm10 and pol α, it is possible that co-expression of these proteins may help to stabilize Mcm10 from degradation and increase its solubility. In addition to screening this complex for

crystallization, this reagent would also prove useful for analysis by EM, SAXS and AUC. Each of these techniques would provide further insight into how these proteins interact. While substantial insight has been gained by dissecting the minimal interacting regions between Mcm10 and pol α, understanding the interactions within the full complex promises to yield the most biologically relevant information. In addition, similar approaches could also be used for studying the Mcm10 and Ctf4 interaction, which is currently poorly understood.

Finally, another interesting avenue of research that will greatly aid in understanding Mcm10 biology is determining its oligomeric state. Currently, numerous discrepancies exist in the literature concerning the oligomerization of Mcm10 in with yeast and vertebrates. Thus, a thorough investigation of XMcm10 oligomerization is warranted. Whether Mcm10 is monomeric, dimeric or hexameric will limit the functional potential for Mcm10 during replication. For example, if Mcm10 is a ring-shaped hexamer, then hypotheses suggesting that it encircles DNA during elongation or establishes a 1:1 interaction with the individual subunits of Mcm2-7 within the replisome, become more plausible. In addition, it is also possible that Mcm10 could adopt different oligomeric states to perform its different functions at the replication fork. While SAXS and MALS will also provided insight into oligomeric state, a continuation of the AUC analysis presented in chapter 4 provide the most definitive answers. In addition, determining whether or not the putative coiled-coil in the NTD has any effect on oligomerization, as hypothesized here, will aid in understanding the means in which Mcm10 self-associates.

As described in this work, significant effort has been put forth to identify the biological function(s) of Mcm10. Specifically, we have investigated the ability of Mcm10 to bind DNA, interact with other replication proteins, as well as form higher order oligomers. In addition, each these functions were mapped to specific regions and/or motifs on the protein. Also included was a high resolution solution structure for the zinc coordinating region of the CTD, which aid in localizing this domains DNA-binding activity. This structure also provided a structural link between Mcm10 and the other members of the Mcm family of proteins, indicating a novel evolutionary connection between these proteins. A further characterization of larger XMcm10 constructs revealed that the region between the ID and CTD acts as a flexible linker, tethering the two domains together. While both of these domains are able to bind DNA independently, preliminary results described here show cooperation between the ID and CTD when full-length Mcm10 binds to DNA. Together, these results begin to place the domain-based analyses into the context of the full-length protein, and thus provide an emerging picture of Mcm10 function. In addition, a clear understanding of the functional potential of Mcm10 *in vitro* will allow better design of *in vivo* experiments aimed at determining the specific role(s) of Mcm10 during replication initiation.

APPENDIX

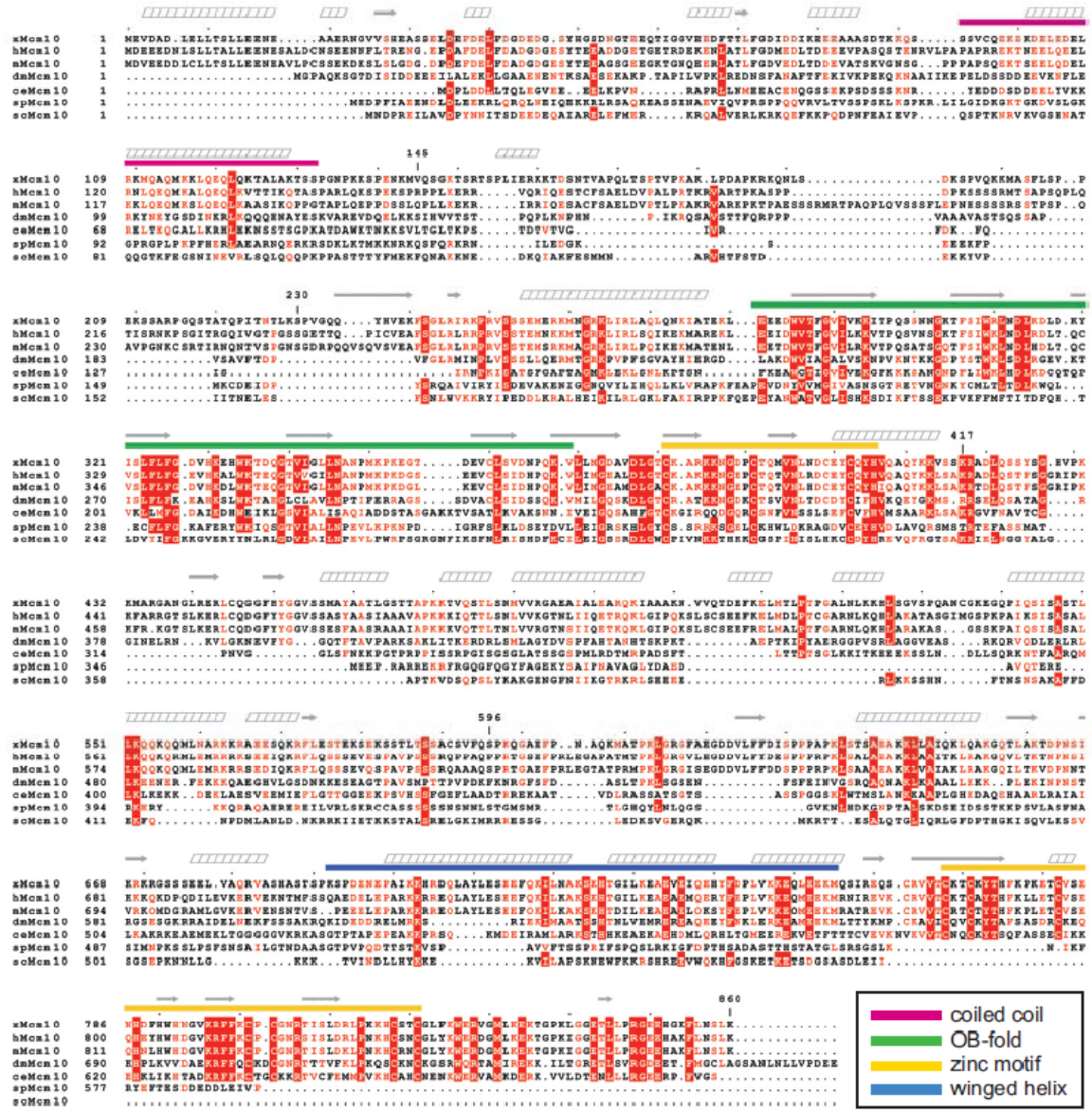


Figure A1. Mcm10 sequence alignment. Primary sequence alignment of Mcm10 proteins from *Xenopus laevis* (x), *Homo sapiens* (h), *Mus musculus* (m), *Drosophila melanogaster* (de), *Caenorhabditis elegans* (ce), *Schizosaccharomyces pombe* (sp), and *Saccharomyces cerevisiae* (sc). Conserved residues are shown in red letters and invariant or strongly conserved residues are highlighted with a red background. Predicted secondary structural elements are shown above the sequence in grey (α-helices, leaning boxes; β-strands, arrows). Predicted structural motifs are shown as colored bars (magenta, coiled coil; green, OB-fold; yellow, zinc motif; blue, winged helix). Sequence alignments were generated with ClustalW (Thompson et al. 1994) and displayed using ESPript (Gouet et al. 1999). Secondary and tertiary structure predictions were carried out using MultiCoil, Phyre, 3D-PSSM (Berger et al. 1995; Wolf et al. 1997; Kelley et al. 2000).

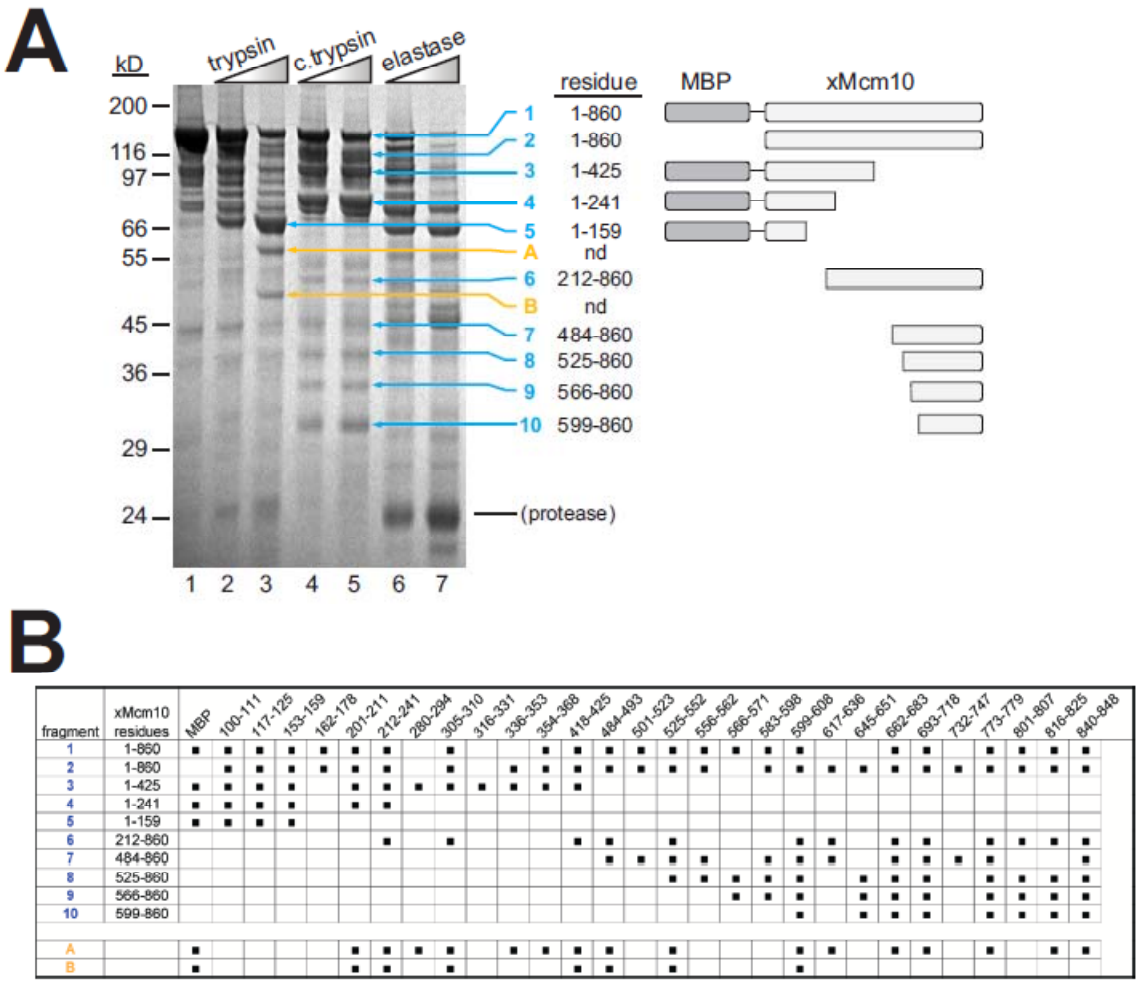


Figure A2. Identification of proteolytically sensitive regions within XMcm10. *A*, Same Coomassie SDS-PAGE gel shown in Figure 4B, with the major proteolytic fragments labeled 1-10, A, and B. Bands 1-10 (blue) were single species that could be unambiguously identified by mass spectrometry, while bands A and B (orange) were mixtures of several co-migrating proteins and thus could not be defined. *B*, Peptide coverage map of fragments shown in panel *A*. Each band 1-10, A, and B was excised from the gel and subjected to complete trypsinolysis and the resulting tryptic peptides (numbers at the top of the chart) were identified by MALDI-TOF and TOF/TOF mass spectrometry. Endpoints of fragments 1-10 could be unambiguously assigned based on the recovered peptides. Peptides spanning MBP and the entire length of Mcm10 were recovered from fragments A and B.

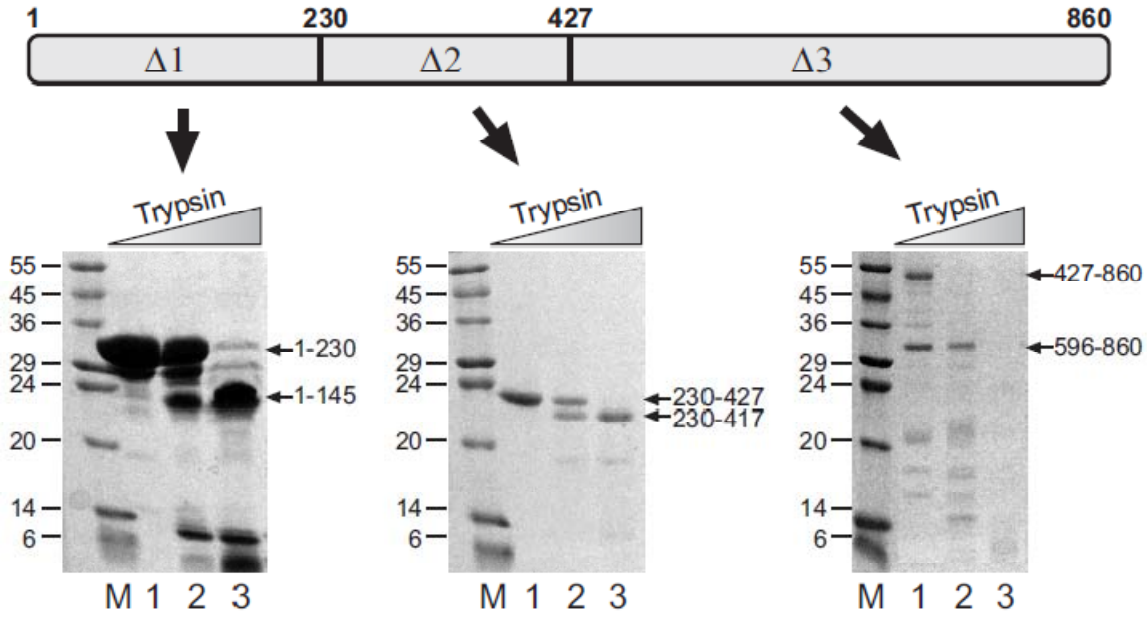


Figure A3. Identification of XMcm10 domains. Three truncation mutant proteins (Δ1, Δ2, Δ3) were purified (lane 1) and subjected to limited proteolytic digestion (lanes 2, and 3) with increasing amounts of trypsin (shown), chymotrypsin, elastase, and endo-GluC. 50-200 pmol XMcm10 mutant was incubated with 1 and 10 ng elastase (lanes 2 and 3, respectively) for 30 min at 37° C. Intact masses of proteolytic products were identified by MALDI mass spectrometry of each reaction mixture. N-terminal sequences were identified by Edman degradation of individual bands from the gel. Lane M, molecular weight standards.

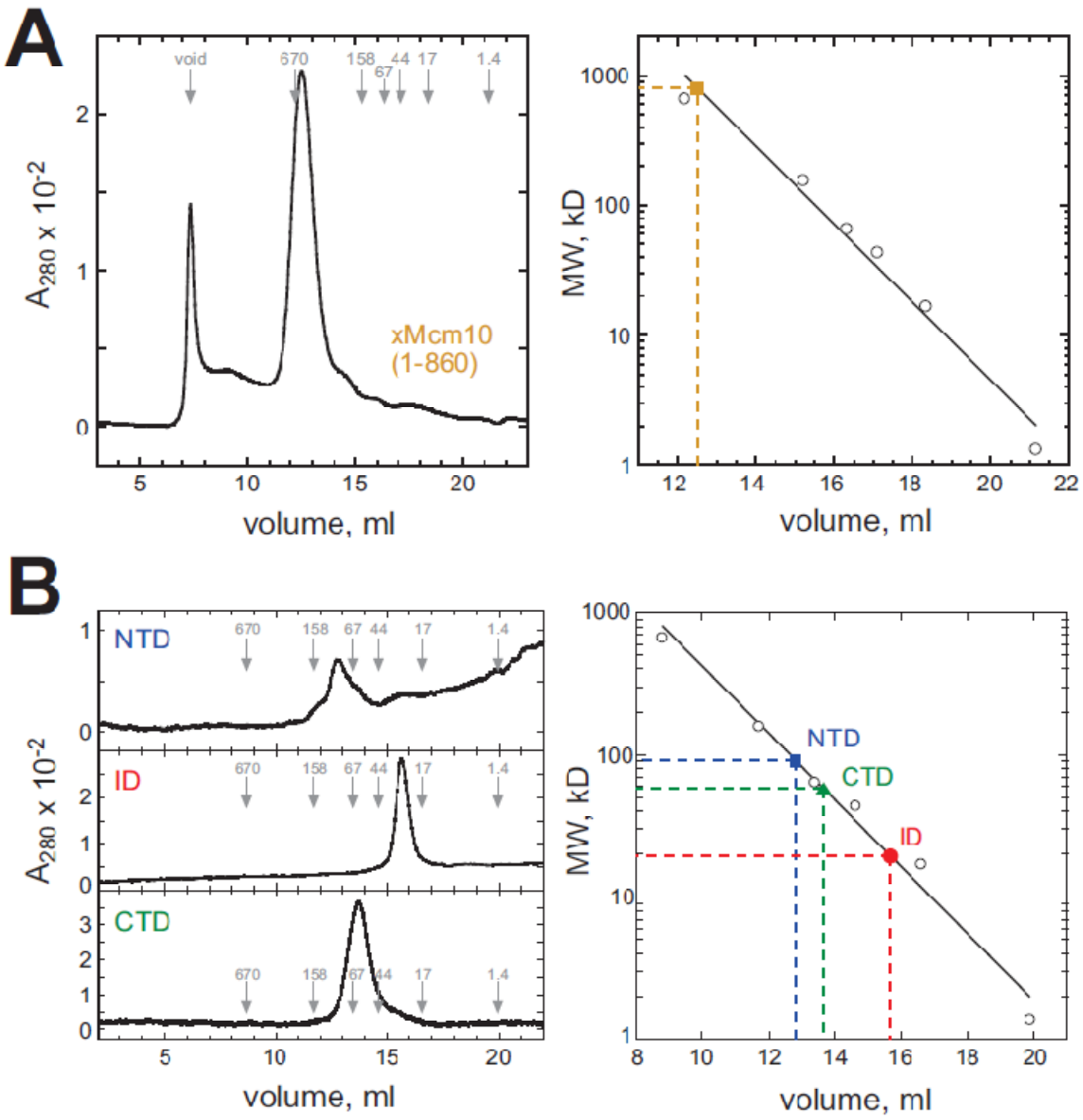


Figure A4. Gel filtration analysis of XMcm10. The left panels show gel filtration chromatograms of full-length XMcm10 (*A*) and individual XMcm10 domains (*B*). Elution volumes of molecular weight standards are marked by gray arrows. The standard curves are shown on the right, with elution volumes for XMcm10 (brown square), XMcm10-NTD (blue square), ID (red circle), and CTD (green triangle) superimposed. Molecular weights calculated from primary sequences are as follows: XMcm10, 95.4 kD; NTD, 16.2 kD; ID, 22.7 kD; CTD, 30.1 kD.

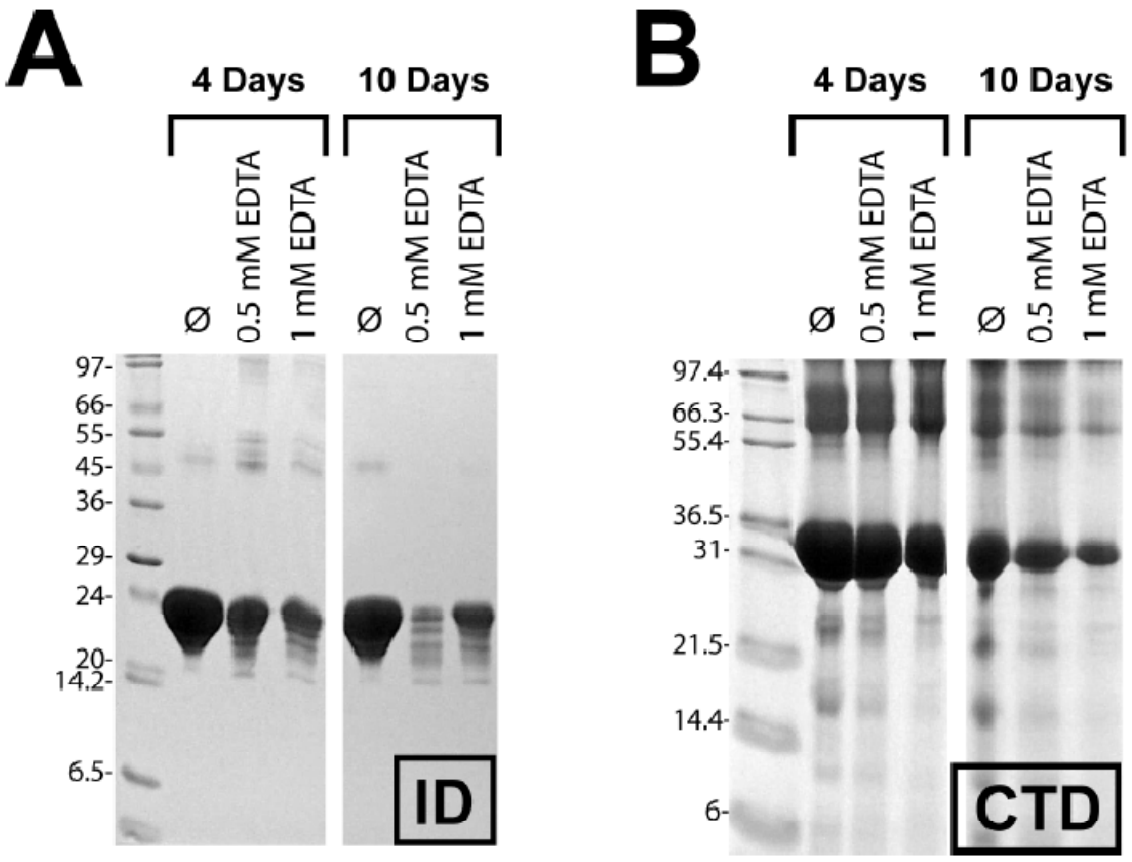


Figure A5. EDTA affects the stability of XMcm10-ID and XMcm10-CTD. The effect of EDTA on the stability of XMcm10-ID (*A*) and -CTD (*B*). Samples were incubated at room temperature for a period of 10 days alone and in the presence of 0.5 mM or 1.0 mM EDTA. Samples were taken on day 4 and frozen at -80ºC to be resolved by SDS-PAGE on day 10.

Table S1. NMR Acquisition Parameters						
Experiment Name	Construct	Dimensions	Scans	Increments (t1×t2×t3)	Magnetic Field (MHz)	Carrier Frequency (ppm)
^{15}N HSQC†	XMcm1$0^{755\text{-}842}$	2D	2	2048×256	600	4.7(H); 120.25(N)
	XMcm1$0^{690\text{-}842}$	2D	32	2048×128	800	4.7(H); 117(N)
	XMcm1$0^{596\text{-}860}$	2D	8	2048×256	600	4.7(H); 120(N)
	XMcm1$0^{230\text{-}860}$	2D	128	2048×256	800	4.7(H); 120(N)
{^{1}H}-^{15}N NOE†	XMcm1$0^{755\text{-}842}$	2D	4	1024×128	600	4.7(H); 120.25(N)
	XMcm1$0^{690\text{-}842}$	2D	16	1024×128	600	4.7(H); 120(N)
HNCA‡	XMcm1$0^{690\text{-}842}$	3D	40	1024×40×64	600	4.7(H); 117(N); 52.5(C)
HNCACB‡	XMcm1$0^{755\text{-}842}$	3D	8	1024×70×180	600	4.7(H); 120.5(N); 48(C)
	XMcm1$0^{690\text{-}842}$	3D	64	1024×70×160	600	4.7(H); 120(N); 40(C)
CBCA(CO)NH‡	XMcm1$0^{755\text{-}842}$	3D	8	1024×70×120	600	4.7(H); 120.25(N); 41(C)
	XMcm1$0^{690\text{-}842}$	3D	8	1024×70×140	600	4.7(H); 120(N); 39(C)
(H)C(CO)NH-TOCSY‡	XMcm1$0^{755\text{-}842}$	3D	16	1024×40×120	600	4.7(H); 120.25(N); 40(C)
	XMcm1$0^{690\text{-}842}$	3D	16	1024×44×140	600	4.7(H); 120(N); 35(C)
H(CC)(CO)NH‡	XMcm1$0^{755\text{-}842}$	3D	8	1024×42×180	600	4.7(H); 120.25(N); 41(C)
	XMcm1$0^{690\text{-}842}$	3D	16	1024×44×140	600	4.7(H); 120(N); 41(C)
HNCO‡	XMcm1$0^{690\text{-}842}$	3D	8	2048×1×128	600	4.7(H); 120(N), 176(C)
HNHA†	XMcm1$0^{755\text{-}842}$	3D	16	1024×140×50	600	4.7(H); 120.25(N)
	XMcm1$0^{690\text{-}842}$	3D	16	1024×128×60	600	4.7(H); 120(N)
HBHANH†	XMcm1$0^{755\text{-}842}$	3D	16	1024×50×150	600	4.7(H); 120.5(N); 41(C)
	XMcm1$0^{690\text{-}842}$	3D	16	2048×48×128	600	4.7(H); 120(N); 39(C)
(HB)CB(CGCD)HD‡	XMcm1$0^{755\text{-}842}$	3D	400	1024×64	600	4.7(H); 119.5(N); 33(C)
^{13}C-NOESY-HSQC‡	XMcm1$0^{755\text{-}842}$	3D	16	1024×50×220	600	4.7(H); 120.25(N); 41(C)
	XMcm1$0^{690\text{-}842}$	3D	16	2048×50×256	800	4.7(H); 116(N); 42(C)
^{15}N-NOESY-HSQC†	XMcm1$0^{755\text{-}842}$	3D	16	1024×80×196	800	4.7(H); 120.25(N); 39.5(C)
	XMcm1$0^{690\text{-}842}$	3D	16	2048×60×240	800	4.7(H); 120(N); 38(C)
COSY	XMcm1$0^{755\text{-}842}$	2D	28	2048×512	800	4.7(H)
	XMcm1$0^{690\text{-}842}$	2D	8	2048×512	600	4.7(H)
NOESY	XMcm1$0^{755\text{-}842}$	2D	32	2048×1024	800	4.7(H)
TOCSY	XMcm1$0^{755\text{-}842}$	2D	28	1024×512	800	4.7(H)
† ^{15}N-enriched sample						
‡ ^{13}C-^{15}N-enriched sample						

Table B1. NMR acquisition parameters.

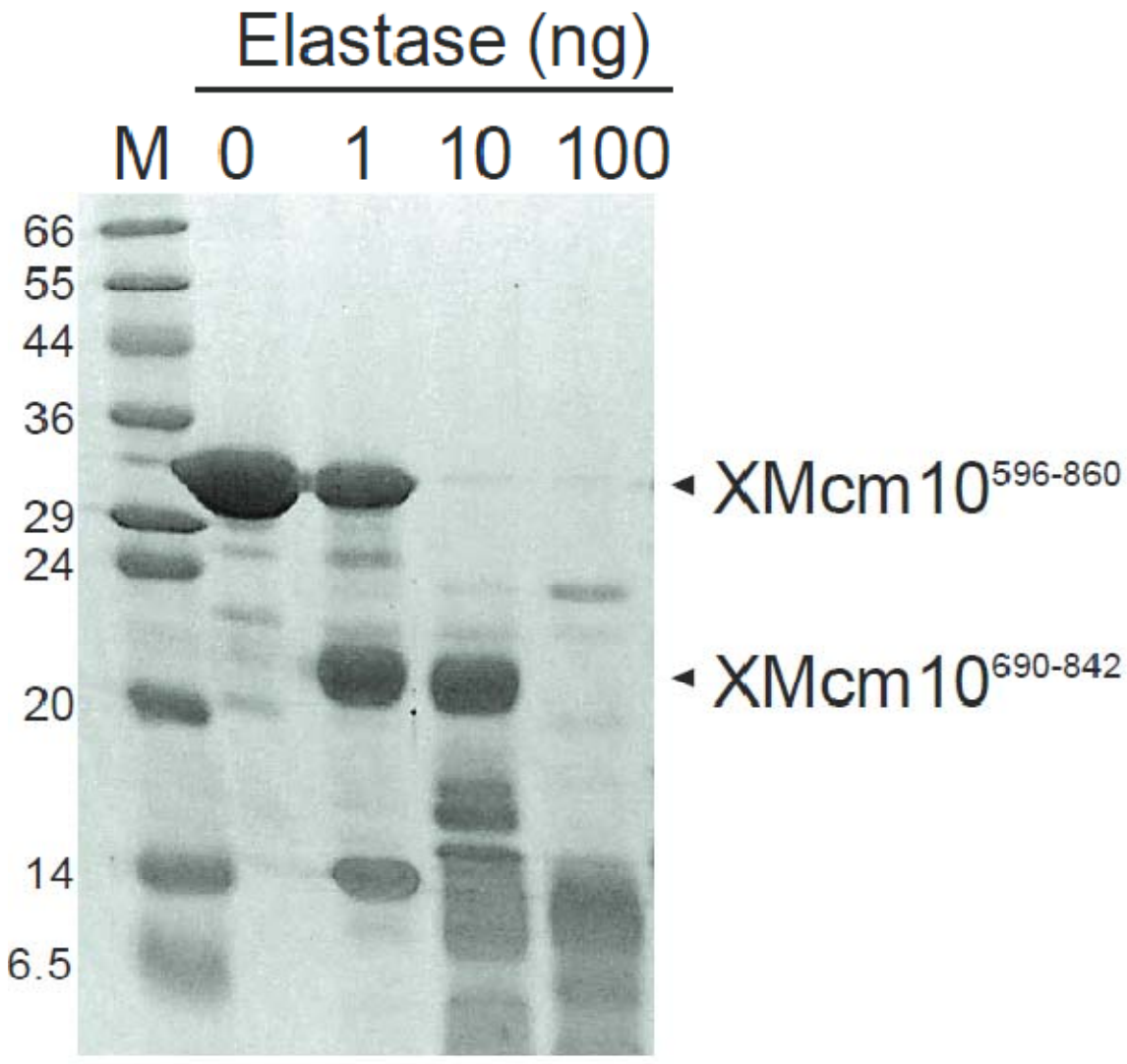


Figure B1. Identification of the XMcm10$^{690-842}$ CTD subdomain. Coomassie Blue stained SDS-PAGE gel of elastase-digested XMcm10$^{596-860}$. Sizes of molecular weight markers (M) in kDa are shown to the left. MALDI-TOF mass spectrometry and Edman degradation identified the prominent 21-kDa molecular weight band as residues 755-842. Proteolysis reactions contained 7.5 μg XMcm10, 1–100 ng elastase, and 25 mM Tris-HCl at pH 7.5, 100 mM NaCl, and 5% glycerol. Reactions were carried out at 25 °C for 30 min and elastase inactivated by the addition of 10 ul of SDS-PAGE sample buffer (63 mM Tris-HCl at pH 6.8, 700 mM B-mercaptoethanol, 2% w/v SDS, 0.03% w/v bromophenol blue, and 10% glycerol) and heat for 5 min at 95 °C.

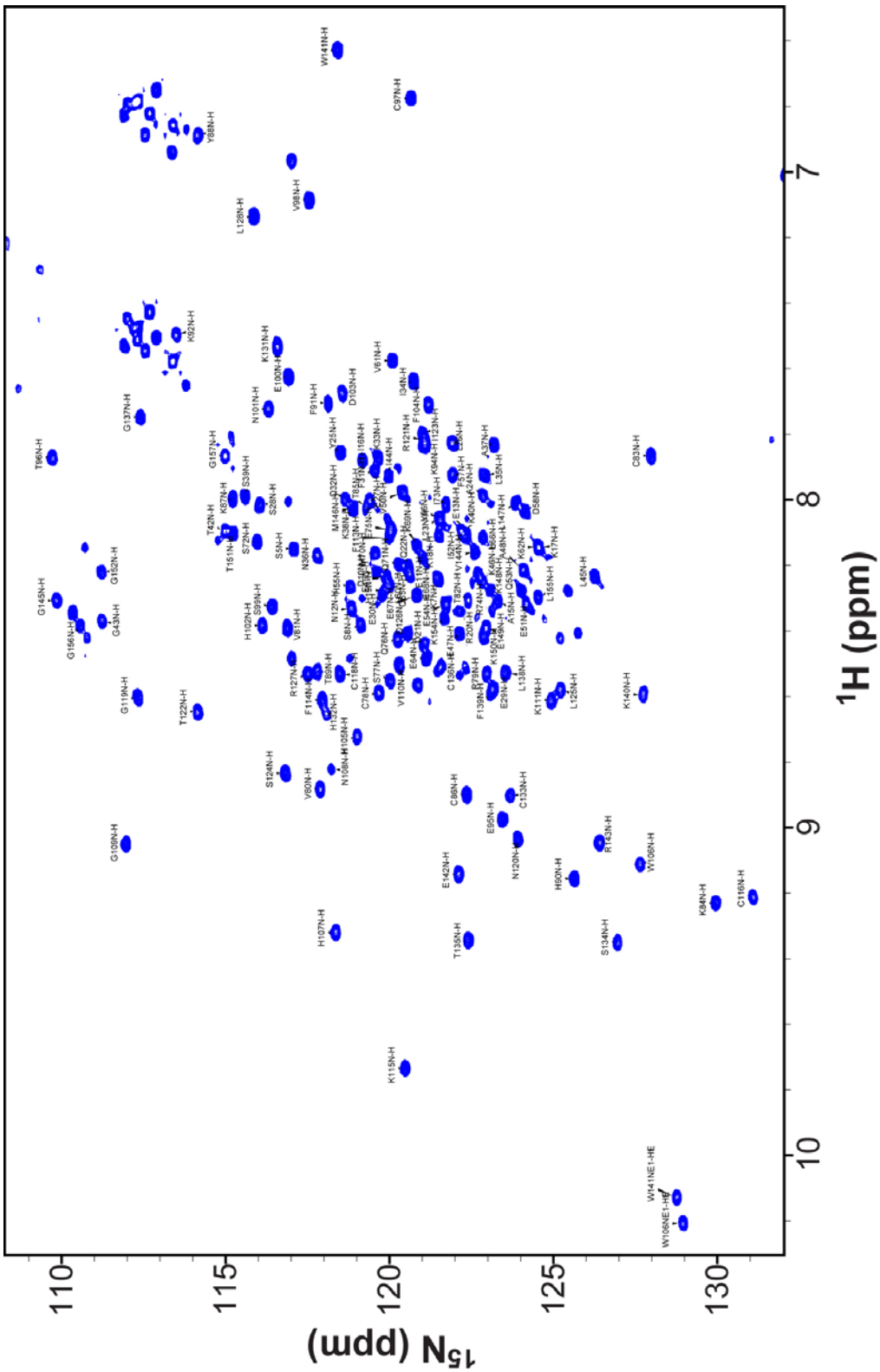


Figure B2. XMcm10$^{690\text{-}842}$ NMR spectrum. ^{15}N-^{1}H HSQC spectrum of ^{15}N-enriched XMcm10$^{690\text{-}842}$ recorded at 800 MHz. All assigned amide signals are labeled with the corresponding residue number.

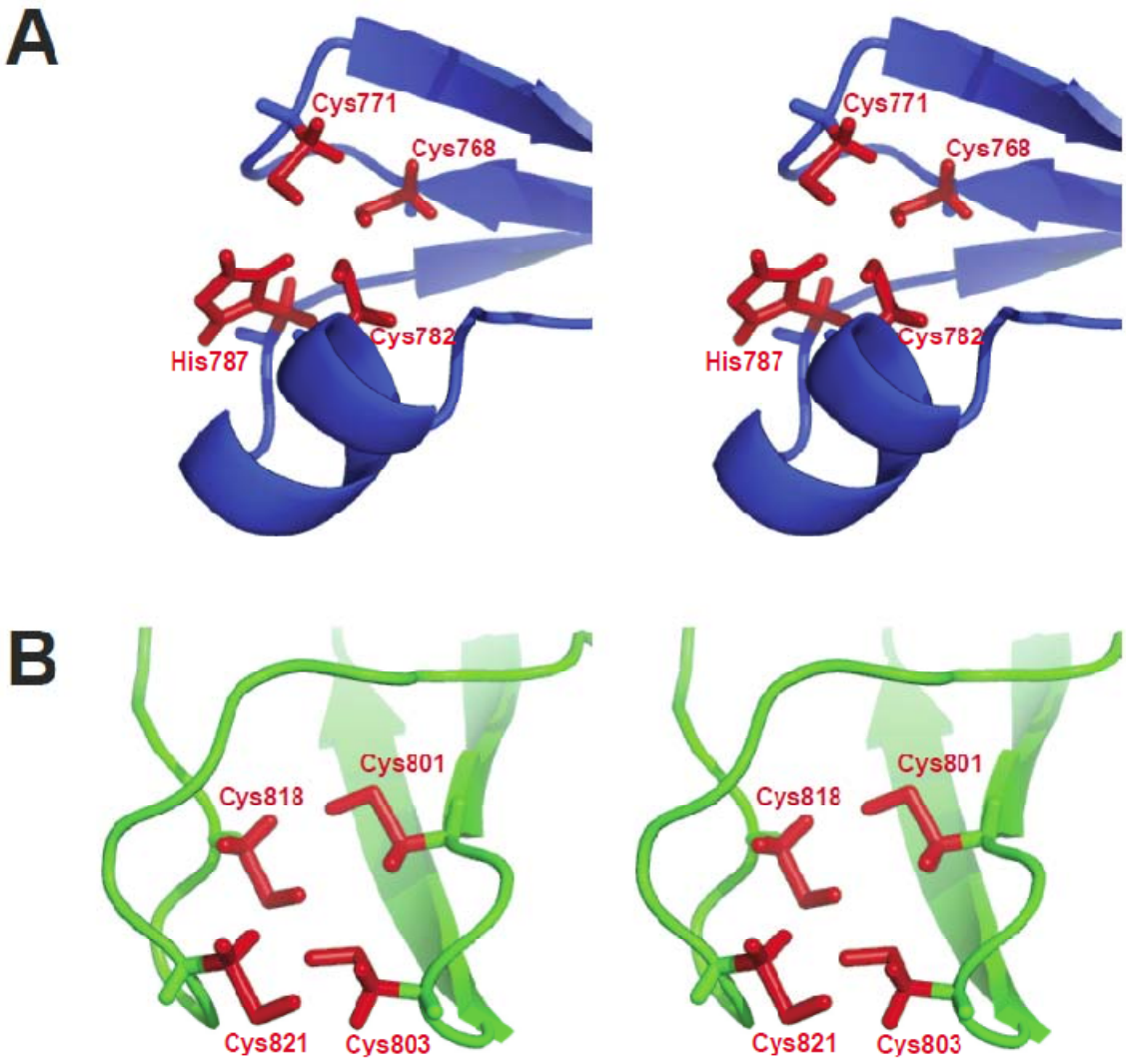


Figure B3. Identification of zinc coordinating residues. Shown are stereodiagrams of the CCCH (A) and CCCC (B) zinc clusters identified in XMcm10$^{755-842}$ prior to imposing restraints specific to zinc coordination. The clusters of side-chains shown in red were the result of energy minimization using only NOE distance restraints.

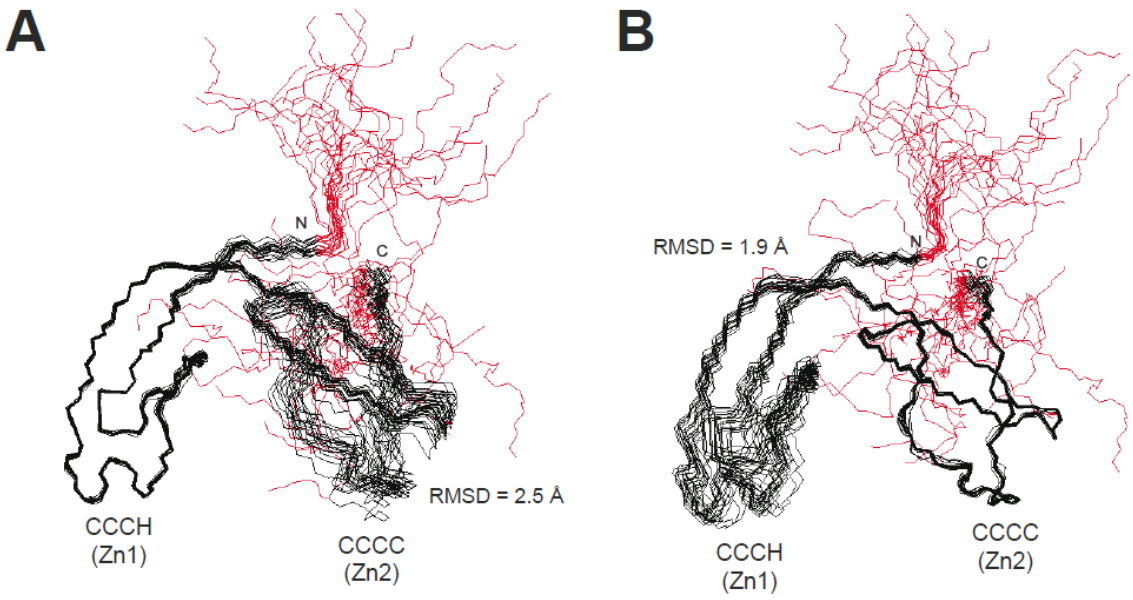


Figure B4. Relative motion between Mcm10-CTD zinc motifs. Backbone superposition of the ensemble of Mcm10$^{755-842}$ structures, aligned at the (A) CCCH motif (residues 765-795) and (B) CCCC motif (residues 795-830). The RMSD for backbone atoms of the unconstrained half of the model was 2.5 Å (A) and 1.9 Å (B).

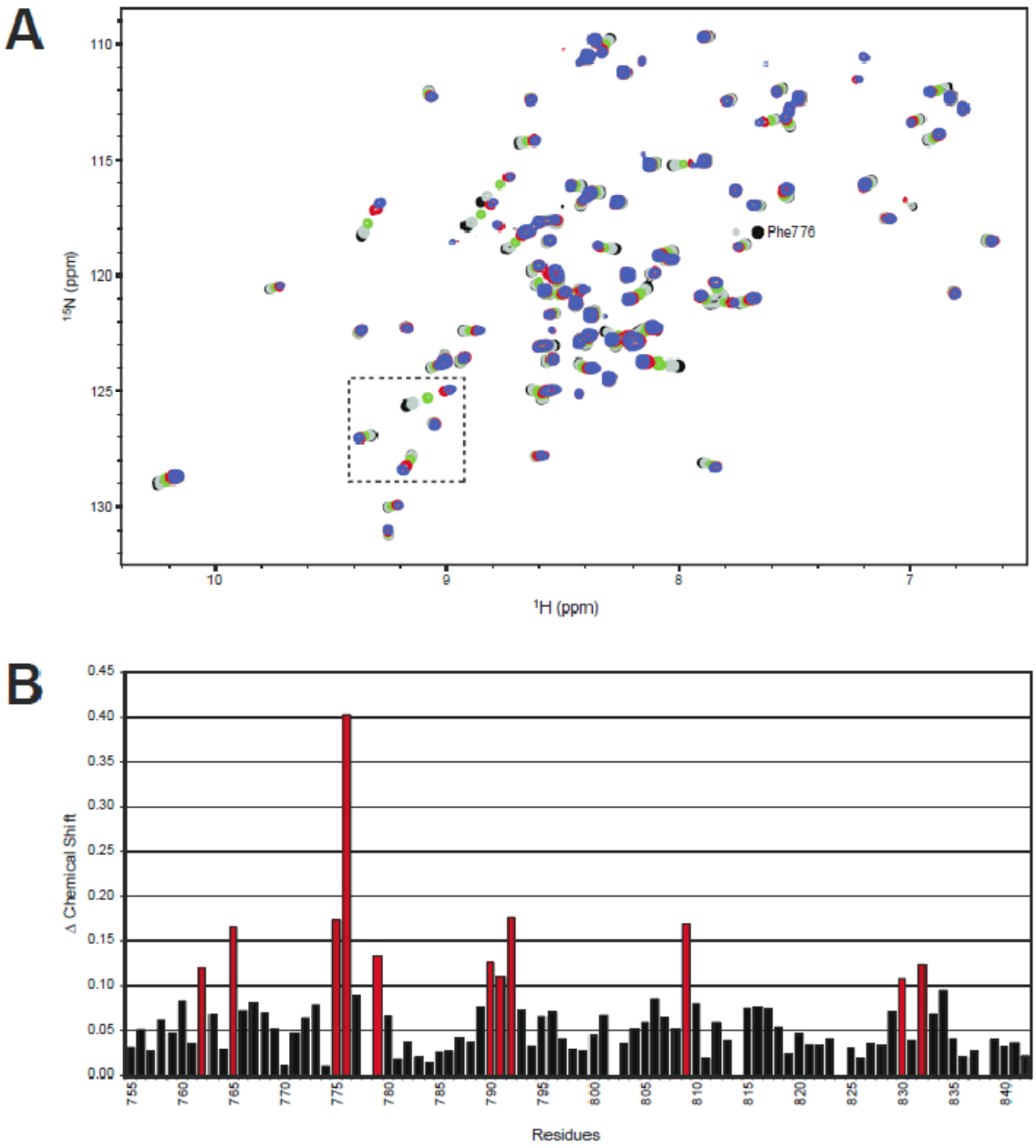
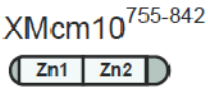


Figure B5. NMR chemical shift perturbation of XMcm10$^{755\text{-}842}$ upon ssDNA binding. (A) Overlays of ^{15}N-^{1}H HSQC spectra of ^{15}N-enriched XMcm10$^{755\text{-}842}$ in the presence of 0 (black), 0.5 (grey), 1 (green), 2 (red), and 4 (blue) fold molar excess of ssDNA. Phe775, which dramatically changes resonance in response to ssDNA, is labeled. (B) Quantitation of chemical shift perturbation between the black and blue spectra shown in panel A, as defined in Experimental Procedures. Red bars represent chemical shift changes greater than one s.d. above the mean.



DNA length (nucleotides)	K_d (μM)	Sequence
25	3.3 ± 0.3	5′-ATGGTAGGCAACCATGTAGTAGTCA-FAM
20	4.1 ± 0.1	5′-AGGCAACCATGTAGTAGTCA-FAM
15	5.6 ± 0.1	5′-ACCATGTAGTAGTCA-FAM
10	17.1 ± 2.3	5′-GTAGTAGTCA-FAM
5	81.2 ± 2.1	5′-AGTCA-FAM

Figure B6. Length dependence for ssDNA binding to XMcm10$^{755-842}$. Dissociation constants (K_d) for XMcm10$^{755-842}$ binding to various lengths of ssDNA were measured by fluorescence polarization as described in Experimental Procedures. Values are the averages and standard deviations from three independent measurements. All oligonucleotides were labeled at the 3′-end with 6-carboxyfluorescein (FAM).

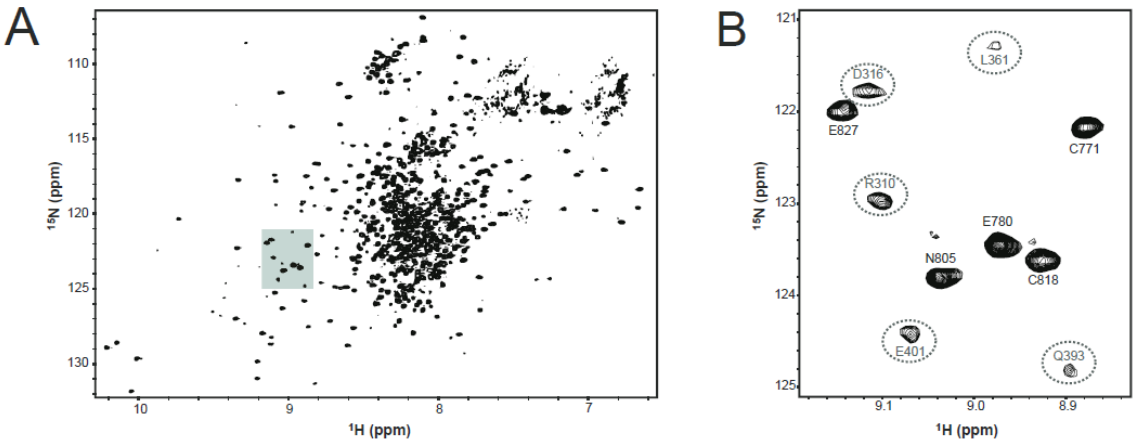


Figure B7. XMcm10$^{230\text{-}860}$ NMR spectrum. ^{15}N-^{1}H TROSY-HSQC spectrum of ^{15}N-enriched XMcm10$^{230\text{-}860}$ recorded at 800 MHz. (B) Expanded view of the grey shaded region of the spectrum displaying signals corresponding to residues from the ID (dotted circles) and the CTD.

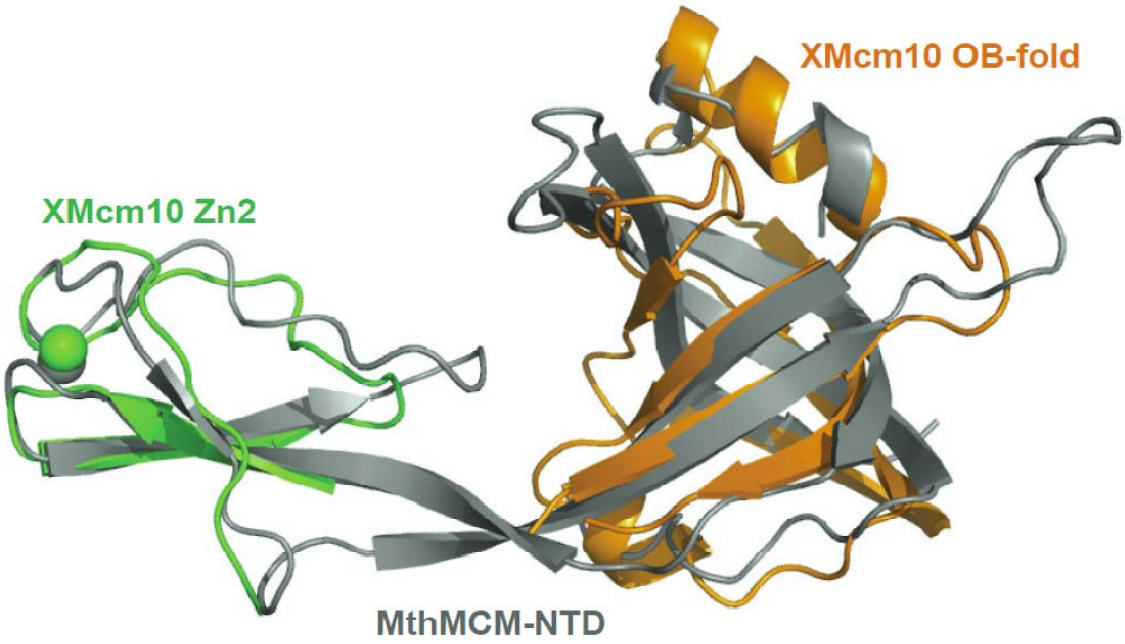


Figure B8. Superposition of XMcm10-ID and -CTD onto MthMCM. Structural alignment of the CCCC zinc ribbon from XMcm10$^{755-842}$ (green), the OB-fold from XMcm10$^{230-427}$ (orange, PDB code 3EBE), and residues 94-242 of MthMCM (magenta, PDB code 1LTL).

REFERENCES

Adachi, Y., J. Usukura and M. Yanagida (1997). "A globular complex formation by Nda1 and the other five members of the MCM protein family in fission yeast." Genes Cells **2**(7): 467-79.

Aladjem, M. I. (2007). "Replication in context: dynamic regulation of DNA replication patterns in metazoans." Nat Rev Genet **8**(8): 588-600.

Anumanthan, G., S. K. Halder, D. B. Friedman and P. K. Datta (2006). "Oncogenic serine-threonine kinase receptor-associated protein modulates the function of Ewing sarcoma protein through a novel mechanism." Cancer Res **66**(22): 10824-32.

Arezi, B. and R. D. Kuchta (2000). "Eukaryotic DNA primase." Trends Biochem Sci **25**(11): 572-6.

Arunkumar, A. I., M. E. Stauffer, E. Bochkareva, A. Bochkarev and W. J. Chazin (2003). "Independent and coordinated functions of replication protein A tandem high affinity single-stranded DNA binding domains." J Biol Chem **278**(42): 41077-82.

Aves, S. J., N. Tongue, A. J. Foster and E. A. Hart (1998). "The essential schizosaccharomyces pombe cdc23 DNA replication gene shares structural and functional homology with the Saccharomyces cerevisiae DNA43 (MCM10) gene." Curr Genet **34**(3): 164-71.

Bae, B., Y. H. Chen, A. Costa, S. Onesti, J. S. Brunzelle, Y. Lin, I. K. Cann and S. K. Nair (2009). "Insights into the architecture of the replicative helicase from the structure of an archaeal MCM homolog." Structure **17**(2): 211-22.

Bell, S. P. and A. Dutta (2002). "DNA replication in eukaryotic cells." Annu Rev Biochem **71**: 333-74.

Berger, B., D. B. Wilson, E. Wolf, T. Tonchev, M. Milla and P. S. Kim (1995). "Predicting coiled coils by use of pairwise residue correlations." Proc Natl Acad Sci U S A **92**(18): 8259-63.

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

Bochkareva, E., S. Korolev, S. P. Lees-Miller and A. Bochkarev (2002). "Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA." Embo J **21**(7): 1855-63.

Bochman, M. L., S. P. Bell and A. Schwacha (2008). "Subunit organization of Mcm2-7 and the unequal role of active sites in ATP hydrolysis and viability." Mol Cell Biol **28**(19): 5865-73.

Bochman, M. L. and A. Schwacha (2007). "Differences in the single-stranded DNA binding activities of MCM2-7 and MCM467: MCM2 and MCM5 define a slow ATP-dependent step." J Biol Chem **282**(46): 33795-804.

Bochman, M. L. and A. Schwacha (2008). "The Mcm2-7 complex has in vitro helicase activity." Mol Cell **31**(2): 287-93.

Bochman, M. L. and A. Schwacha (2009). "The Mcm complex: unwinding the mechanism of a replicative helicase." Microbiol Mol Biol Rev **73**(4): 652-83.

Borden, K. L. (2000). "RING domains: master builders of molecular scaffolds?" J Mol Biol **295**(5): 1103-12.

Brewster, A. S., G. Wang, X. Yu, W. B. Greenleaf, J. M. Carazo, M. Tjajadia, M. G. Klein and X. S. Chen (2008). "Crystal structure of a near-full-length archaeal MCM: functional insights for an AAA+ hexameric helicase." Proc Natl Acad Sci U S A **105**(51): 20191-6.

Broach, J. R., Y. Y. Li, J. Feldman, M. Jayaram, J. Abraham, K. A. Nasmyth and J. B. Hicks (1983). "Localization and sequence analysis of yeast origins of DNA replication." Cold Spring Harb Symp Quant Biol **47 Pt 2**: 1165-73.

Brosey, C. A., M. E. Chagot, M. Ehrhardt, D. I. Pretto, B. E. Weiner and W. J. Chazin (2009). "NMR analysis of the architecture and functional remodeling of a modular multidomain protein, RPA." J Am Chem Soc **131**(18): 6346-7.

Burgers, P. M. (2009). "Polymerase dynamics at the eukaryotic DNA replication fork." J Biol Chem **284**(7): 4041-5.

Carpentieri, F., M. De Felice, M. De Falco, M. Rossi and F. M. Pisani (2002). "Physical and functional interaction between the mini-chromosome maintenance-like DNA helicase and the single-stranded DNA binding protein from the crenarchaeon Sulfolobus solfataricus." J Biol Chem **277**(14): 12118-27.

Caruthers, J. M. and D. B. McKay (2002). "Helicase structure and mechanism." Curr Opin Struct Biol **12**(1): 123-33.

Case, D. A., T. E. Cheatham, 3rd, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang and R. J. Woods (2005). "The Amber biomolecular simulation programs." J Comput Chem **26**(16): 1668-88.

Cavanagh, J., W. J. Fairbrother, A. G. I. Palmer, R. M. and N. J. Skelton (2007). Protein NMR Spectroscopy: Principles and Practice, Academic Press Inc., New York.

Chattopadhyay, S. and A. K. Bielinsky (2007). "Human Mcm10 regulates the catalytic subunit of DNA polymerase-alpha and prevents DNA damage during replication." Mol Biol Cell **18**(10): 4085-95.

Chong, J. P., M. K. Hayashi, M. N. Simon, R. M. Xu and B. Stillman (2000). "A double-hexamer archaeal minichromosome maintenance protein is an ATP-dependent DNA helicase." Proc Natl Acad Sci U S A **97**(4): 1530-5.

Christensen, T. W. and B. K. Tye (2003). "Drosophila MCM10 interacts with members of the prereplication complex and is required for proper chromosome condensation." Mol Biol Cell **14**(6): 2206-15.

Coleman, T. R., P. B. Carpenter and W. G. Dunphy (1996). "The Xenopus Cdc6 protein is essential for the initiation of a single round of DNA replication in cell-free extracts." Cell **87**(1): 53-63.

Collins, K. L., A. A. Russo, B. Y. Tseng and T. J. Kelly (1993). "The role of the 70 kDa subunit of human DNA polymerase alpha in DNA replication." Embo J **12**(12): 4555-66.

Conaway, R. C. and I. R. Lehman (1982). "Synthesis by the DNA primase of Drosophila melanogaster of a primer with a unique chain length." Proc Natl Acad Sci U S A **79**(15): 4585-8.

Cook, C. R., G. Kung, F. C. Peterson, B. F. Volkman and M. Lei (2003). "A novel zinc finger is required for Mcm10 homocomplex assembly." J Biol Chem **278**(38): 36051-8.

Corn, J. E. and J. M. Berger (2006). "Regulation of bacterial priming and daughter strand synthesis through helicase-primase interactions." Nucleic Acids Res **34**(15): 4082-8.

Costa, A., G. van Duinen, B. Medagli, J. Chong, N. Sakakibara, Z. Kelman, S. K. Nair, A. Patwardhan and S. Onesti (2008). "Cryo-electron microscopy reveals a novel DNA-binding site on the MCM helicase." Embo J **27**(16): 2250-8.

Cozzarelli, N. R. (1980). "DNA gyrase and the supercoiling of DNA." Science **207**(4434): 953-60.

Dalgaard, J. Z., T. Eydmann, M. Koulintchenko, S. Sayrac, S. Vengrova and T. Yamada-Inagawa (2009). "Random and site-specific replication termination." Methods Mol Biol **521**: 35-53.

Dalton, S. and B. Hopwood (1997). "Characterization of Cdc47p-minichromosome maintenance complexes in Saccharomyces cerevisiae: identification of Cdc45p as a subunit." Mol Cell Biol **17**(10): 5867-75.

Das-Bradoo, S., R. M. Ricke and A. K. Bielinsky (2006). "Interaction between PCNA and diubiquitinated Mcm10 is essential for cell growth in budding yeast." Mol Cell Biol **26**(13): 4806-17.

Davis, I. W., A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, 3rd, J. Snoeyink, J. S. Richardson and D. C. Richardson (2007). "MolProbity: all-atom contacts and structure validation for proteins and nucleic acids." Nucleic Acids Res **35**(Web Server issue): W375-83.

Dornreiter, I., W. C. Copeland and T. S. Wang (1993). "Initiation of simian virus 40 DNA replication requires the interaction of a specific domain of human DNA polymerase alpha with large T antigen." Mol Cell Biol **13**(2): 809-20.

Dumas, L. B., J. P. Lussky, E. J. McFarland and J. Shampay (1982). "New temperature-sensitive mutants of Saccharomyces cerevisiae affecting DNA replication." Mol Gen Genet **187**(1): 42-6.

Eisenberg, S., G. Korza, J. Carson, I. Liachko and B. K. Tye (2009). "Novel DNA binding properties of the Mcm10 protein from Saccharomyces cerevisiae." J Biol Chem **284**(37): 25412-20.

Fanning, E., V. Klimovich and A. R. Nager (2006). "A dynamic model for replication protein A (RPA) function in DNA processing pathways." Nucleic Acids Res **34**(15): 4126-37.

Fien, K., Y. S. Cho, J. K. Lee, S. Raychaudhuri, I. Tappin and J. Hurwitz (2004). "Primer utilization by DNA polymerase alpha-primase is influenced by its interaction with Mcm10p." J Biol Chem **279**(16): 16144-53.

Fien, K. and J. Hurwitz (2006). "Fission yeast Mcm10p contains primase activity." J Biol Chem **281**(31): 22248-60.

Fletcher, R. J., B. E. Bishop, R. P. Leon, R. A. Sclafani, C. M. Ogata and X. S. Chen (2003). "The structure and function of MCM from archaeal M. Thermoautotrophicum." Nat Struct Biol **10**(3): 160-7.

Fletcher, R. J., J. Shen, Y. Gomez-Llorente, C. S. Martin, J. M. Carazo and X. S. Chen (2005). "Double hexamer disruption and biochemical activities of Methanobacterium thermoautotrophicum MCM." J Biol Chem **280**(51): 42405-10.

Gambus, A., R. C. Jones, A. Sanchez-Diaz, M. Kanemaki, F. van Deursen, R. D. Edmondson and K. Labib (2006). "GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks." Nat Cell Biol **8**(4): 358-66.

Gambus, A., F. van Deursen, D. Polychronopoulos, M. Foltman, R. C. Jones, R. D. Edmondson, A. Calzada and K. Labib (2009). "A key role for Ctf4 in coupling the MCM2-7 helicase to DNA polymerase alpha within the eukaryotic replisome." Embo J **28**(19): 2992-3004.

Garcia de la Torre, J. G. and V. A. Bloomfield (1981). "Hydrodynamic properties of complex, rigid, biological macromolecules: theory and applications." Q Rev Biophys **14**(1): 81-139.

Garg, P., C. M. Stith, J. Majka and P. M. Burgers (2005). "Proliferating cell nuclear antigen promotes translesion synthesis by DNA polymerase zeta." J Biol Chem **280**(25): 23446-50.

Goddard, T. D. and D. G. Kneller (2006). SPARKY 3, University of California, San Francisco.

Gomez-Llorente, Y., R. J. Fletcher, X. S. Chen, J. M. Carazo and C. San Martin (2005). "Polymorphism and double hexamer structure in the archaeal minichromosome maintenance (MCM) helicase from Methanobacterium thermoautotrophicum." J Biol Chem **280**(49): 40909-15.

Gomez, E. B., M. G. Catlett and S. L. Forsburg (2002). "Different phenotypes in vivo are associated with ATPase motif mutations in Schizosaccharomyces pombe minichromosome maintenance proteins." Genetics **160**(4): 1305-18.

Gouet, P., E. Courcelle, D. I. Stuart and F. Metoz (1999). "ESPript: analysis of multiple sequence alignments in PostScript." Bioinformatics **15**(4): 305-8.

Gozuacik, D., M. Chami, D. Lagorce, J. Faivre, Y. Murakami, O. Poch, E. Biermann, R. Knippers, C. Brechot and P. Paterlini-Brechot (2003). "Identification and functional characterization of a new member of the human Mcm protein family: hMcm8." Nucleic Acids Res **31**(2): 570-9.

Gregan, J., K. Lindner, L. Brimage, R. Franklin, M. Namdar, E. A. Hart, S. J. Aves and S. E. Kearsey (2003). "Fission yeast Cdc23/Mcm10 functions after pre-replicative complex formation to promote Cdc45 chromatin binding." Mol Biol Cell **14**(9): 3876-87.

Guntert, P. (2004). "Automated NMR structure calculation with CYANA." Methods Mol Biol **278**: 353-78.

Hayes, T. E., P. Sengupta and B. H. Cochran (1988). "The human c-fos serum response factor and the yeast factors GRM/PRTF have related DNA-binding specificities." Genes Dev **2**(12B): 1713-22.

Hennessy, K. M., A. Lee, E. Chen and D. Botstein (1991). "A group of interacting yeast DNA replication genes." Genes Dev **5**(6): 958-69.

Holm, L. and C. Sander (1995). "Dali: a network tool for protein structure comparison." Trends Biochem Sci **20**(11): 478-80.

Homesley, L., M. Lei, Y. Kawasaki, S. Sawyer, T. Christensen and B. K. Tye (2000). "Mcm10 and the MCM2-7 complex interact to initiate DNA synthesis and to release replication factors from origins." Genes Dev **14**(8): 913-26.

Hu, H. and W. J. Chazin (2003). "Unique features in the C-terminal domain provide caltractin with target specificity." J Mol Biol **330**(3): 473-84.

Ilves, I., T. Petojevic, J. J. Pesavento and M. R. Botchan (2010). "Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins." Mol Cell **37**(2): 247-58.

Im, J. S., S. H. Ki, A. Farina, D. S. Jung, J. Hurwitz and J. K. Lee (2009). "Assembly of the Cdc45-Mcm2-7-GINS complex in human cells requires the Ctf4/And-1, RecQL4, and Mcm10 proteins." Proc Natl Acad Sci U S A **106**(37): 15628-32.

Izumi, M., K. Yanagi, T. Mizuno, M. Yokoi, Y. Kawasaki, K. Y. Moon, J. Hurwitz, F. Yatagai and F. Hanaoka (2000). "The human homolog of Saccharomyces cerevisiae Mcm10 interacts with replication factors and dissociates from nuclease-resistant nuclear structures in G(2) phase." Nucleic Acids Res **28**(23): 4769-77.

Izumi, M., F. Yatagai and F. Hanaoka (2001). "Cell cycle-dependent proteolysis and phosphorylation of human Mcm10." J Biol Chem **276**(51): 48526-31.

Jiang, X., V. Klimovich, A. I. Arunkumar, E. B. Hysinger, Y. Wang, R. D. Ott, G. D. Guler, B. Weiner, W. J. Chazin and E. Fanning (2006). "Structural mechanism of RPA loading on DNA during activation of a simple pre-replication complex." Embo J **25**(23): 5516-26.

Kadrmas, J. L. and M. C. Beckerle (2004). "The LIM domain: from the cytoskeleton to the nucleus." Nat Rev Mol Cell Biol **5**(11): 920-31.

Kasiviswanathan, R., J. H. Shin, E. Melamud and Z. Kelman (2004). "Biochemical characterization of the Methanothermobacter thermautotrophicus minichromosome maintenance (MCM) helicase N-terminal domains." J Biol Chem **279**(27): 28358-66.

Kawasaki, Y., S. Hiraga and A. Sugino (2000). "Interactions between Mcm10p and other replication factors are required for proper initiation and elongation of chromosomal DNA replication in Saccharomyces cerevisiae." Genes Cells **5**(12): 975-89.

Kearsey, S. E. and K. Labib (1998). "MCM proteins: evolution, properties, and role in DNA replication." Biochim Biophys Acta **1398**(2): 113-36.

Kelley, L., R. MacCallum and M. Sternberg (2000). "Enhanced genome annotation using structural profiles in the program 3D-PSSM." J. Mol. Biol **299**: 499-520.

Kelley, L. A., R. M. MacCallum and M. J. Sternberg (2000). "Enhanced genome annotation using structural profiles in the program 3D-PSSM." J Mol Biol **299**(2): 499-520.

Kelman, Z., J. K. Lee and J. Hurwitz (1999). "The single minichromosome maintenance protein of Methanobacterium thermoautotrophicum DeltaH contains DNA helicase activity." Proc Natl Acad Sci U S A **96**(26): 14783-8.

Kinoshita, Y., E. M. Johnson, R. E. Gordon, H. Negri-Bell, M. T. Evans, J. Coolbaugh, Y. Rosario-Peralta, J. Samet, E. Slusser, M. P. Birkenbach and D. C. Daniel (2008). "Colocalization of MCM8 and MCM7 with proteins involved in distinct aspects of DNA replication." Microsc Res Tech **71**(4): 288-97.

Klug, A. and J. W. Schwabe (1995). "Protein motifs 5. Zinc fingers." Faseb J **9**(8): 597-604.

Koonin, E. V. (1993). "A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication." Nucleic Acids Res **21**(11): 2541-7.

Kwon, K., C. Cao and J. T. Stivers (2003). "A novel zinc snap motif conveys structural stability to 3-methyladenine DNA glycosylase I." J Biol Chem **278**(21): 19442-6.

Langston, L. D., C. Indiani and M. O'Donnell (2009). "Whither the replisome: emerging perspectives on the dynamic nature of the DNA replication machinery." Cell Cycle **8**(17): 2686-91.

Laskowski, R. A., J. A. Rullmannn, M. W. MacArthur, R. Kaptein and J. M. Thornton (1996). "AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR." J Biomol NMR **8**(4): 477-86.

Lau, E., T. Tsuji, L. Guo, S. H. Lu and W. Jiang (2007). "The role of pre-replicative complex (pre-RC) components in oncogenesis." Faseb J **21**(14): 3786-94.

Laue, T. M., Shah, B., Ridgeway, T.M., and Pelletier, S.L. (1992). SEDNTERP, Royal Society of Chemistry, UK.

Lee, C., I. Liachko, R. Bouten, Z. Kelman and B. K. Tye (2010). "Alternative mechanisms for coordinating polymerase alpha and MCM helicase." Mol Cell Biol **30**(2): 423-35.

Lee, J. K., Y. S. Seo and J. Hurwitz (2003). "The Cdc23 (Mcm10) protein is required for the phosphorylation of minichromosome maintenance complex by the Dfp1-Hsk1 kinase." Proc Natl Acad Sci U S A **100**(5): 2334-9.

Lee, J. Y., C. Chang, H. K. Song, J. Moon, J. K. Yang, H. K. Kim, S. T. Kwon and S. W. Suh (2000). "Crystal structure of NAD(+)-dependent DNA ligase: modular architecture and functional implications." Embo J **19**(5): 1119-29.

Lei, M., Y. Kawasaki, M. R. Young, M. Kihara, A. Sugino and B. K. Tye (1997). "Mcm2 is a target of regulation by Cdc7-Dbf4 during the initiation of DNA synthesis." Genes Dev **11**(24): 3365-74.

Lei, M. and B. K. Tye (2001). "Initiating DNA synthesis: from recruiting to activating the MCM complex." J Cell Sci **114**(Pt 8): 1447-54.

Leon, O. and M. Roth (2000). "Zinc fingers: DNA binding and protein-protein interactions." Biol Res **33**(1): 21-30.

Leppard, J. B. and J. J. Champoux (2005). "Human DNA topoisomerase I: relaxation, roles, and damage control." Chromosoma **114**(2): 75-85.

Liang, D. T. and S. L. Forsburg (2001). "Characterization of Schizosaccharomyces pombe mcm7(+) and cdc23(+) (MCM10) and interactions with replication checkpoints." Genetics **159**(2): 471-86.

Liu, W., B. Pucci, M. Rossi, F. M. Pisani and R. Ladenstein (2008). "Structural analysis of the Sulfolobus solfataricus MCM protein N-terminal domain." Nucleic Acids Res **36**(10): 3235-43.

Liu, Y., T. A. Richards and S. J. Aves (2009). "Ancient diversification of eukaryotic MCM DNA replication proteins." BMC Evol Biol **9**: 60.

Luft, J. R., R. J. Collins, N. A. Fehrman, A. M. Lauricella, C. K. Veatch and G. T. DeTitta (2003). "A deliberate approach to screening for initial crystallization conditions of biological macromolecules." J Struct Biol **142**(1): 170-9.

Lutzmann, M., D. Maiorano and M. Mechali (2005). "Identification of full genes and proteins of MCM9, a novel, vertebrate-specific member of the MCM2-8 protein family." Gene **362**: 51-6.

Lutzmann, M. and M. Mechali (2008). "MCM9 binds Cdt1 and is required for the assembly of prereplication complexes." Mol Cell **31**(2): 190-200.

Machida, Y. J., J. L. Hamlin and A. Dutta (2005). "Right place, right time, and only once: replication initiation in metazoans." Cell **123**(1): 13-24.

Maine, G. T., P. Sinha and B. K. Tye (1984). "Mutants of S. cerevisiae defective in the maintenance of minichromosomes." Genetics **106**(3): 365-85.

Maiorano, D., O. Cuvier, E. Danis and M. Mechali (2005). "MCM8 is an MCM2-7-related protein that functions as a DNA helicase during replication elongation and not initiation." Cell **120**(3): 315-28.

Maiorano, D., M. Lutzmann and M. Mechali (2006). "MCM proteins and DNA replication." Curr Opin Cell Biol **18**(2): 130-6.

Mer, G., A. Bochkarev, R. Gupta, E. Bochkareva, L. Frappier, C. J. Ingles, A. M. Edwards and W. J. Chazin (2000). "Structural basis for the recognition of DNA repair proteins UNG2, XPA, and RAD52 by replication factor RPA." Cell **103**(3): 449-56.

Merchant, A. M., Y. Kawasaki, Y. Chen, M. Lei and B. K. Tye (1997). "A lesion in the DNA replication initiation factor Mcm10 induces pausing of elongation forks through chromosomal replication origins in Saccharomyces cerevisiae." Mol Cell Biol **17**(6): 3261-71.

Metz, A. H., T. Hollis and B. F. Eichman (2007). "DNA damage recognition and repair by 3-methyladenine DNA glycosylase I (TAG)." Embo J **26**(9): 2411-20.

Michael, W. M., R. Ott, E. Fanning and J. Newport (2000). "Activation of the DNA replication checkpoint through RNA synthesis by primase." Science **289**(5487): 2133-7.

Mimura, S. and H. Takisawa (1998). "Xenopus Cdc45-dependent loading of DNA polymerase alpha onto chromatin under the control of S-phase Cdk." Embo J **17**(19): 5699-707.

Mizuno, T., N. Ito, M. Yokoi, A. Kobayashi, K. Tamai, H. Miyazawa and F. Hanaoka (1998). "The second-largest subunit of the mouse DNA polymerase alpha-primase complex facilitates both production and nuclear translocation of the catalytic subunit of DNA polymerase alpha." Mol Cell Biol **18**(6): 3552-62.

Mizuno, T., K. Yamagishi, H. Miyazawa and F. Hanaoka (1999). "Molecular architecture of the mouse DNA polymerase alpha-primase complex." Mol Cell Biol **19**(11): 7886-96.

Moir, D., S. E. Stewart, B. C. Osmond and D. Botstein (1982). "Cold-sensitive cell-division-cycle mutants of yeast: isolation, properties, and pseudoreversion studies." Genetics **100**(4): 547-63.

Moyer, S. E., P. W. Lewis and M. R. Botchan (2006). "Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase." Proc Natl Acad Sci U S A **103**(27): 10236-41.

Murzin, A. G., S. E. Brenner, T. Hubbard and C. Chothia (1995). "SCOP: a structural classification of proteins database for the investigation of sequences and structures." J Mol Biol **247**(4): 536-40.

Nasmyth, K. and P. Nurse (1981). "Cell division cycle mutants altered in DNA replication and mitosis in the fission yeast Schizosaccharomyces pombe." Mol Gen Genet **182**(1): 119-24.

Nick McElhinny, S. A., D. A. Gordenin, C. M. Stith, P. M. Burgers and T. A. Kunkel (2008). "Division of labor at the eukaryotic replication fork." Mol Cell **30**(2): 137-44.

Okorokov, A. L., A. Waugh, J. Hodgkinson, A. Murthy, H. K. Hong, E. Leo, M. B. Sherman, K. Stoeber, E. V. Orlova and G. H. Williams (2007). "Hexameric ring structure of human MCM10 DNA replication factor." EMBO Rep **8**(10): 925-30.

Ott, R. D., C. Rehfuess, V. N. Podust, J. E. Clark and E. Fanning (2002). "Role of the p68 subunit of human DNA polymerase alpha-primase in simian virus 40 DNA replication." Mol Cell Biol **22**(16): 5669-78.

Pacek, M., A. V. Tutter, Y. Kubota, H. Takisawa and J. C. Walter (2006). "Localization of MCM2-7, Cdc45, and GINS to the site of DNA unwinding during eukaryotic DNA replication." Mol Cell **21**(4): 581-7.

Pacek, M. and J. C. Walter (2004). "A requirement for MCM7 and Cdc45 in chromosome unwinding during eukaryotic DNA replication." Embo J **23**(18): 3667-76.

Passmore, S., R. Elble and B. K. Tye (1989). "A protein involved in minichromosome maintenance in yeast binds a transcriptional enhancer conserved in eukaryotes." Genes Dev **3**(7): 921-35.

Passmore, S., G. T. Maine, R. Elble, C. Christ and B. K. Tye (1988). "Saccharomyces cerevisiae protein involved in plasmid maintenance is necessary for mating of MAT alpha cells." J Mol Biol **204**(3): 593-606.

Petroski, M. D. and R. J. Deshaies (2005). "Function and regulation of cullin-RING ubiquitin ligases." Nat Rev Mol Cell Biol **6**(1): 9-20.

Philo, J. S. (2000). "A method for directly fitting the time derivative of sedimentation velocity data and an alternative algorithm for calculating sedimentation coefficient distribution functions." Analytical Biochemistry **279**(2): 151-163.

Poddar, A., N. Roy and P. Sinha (1999). "MCM21 and MCM22, two novel genes of the yeast Saccharomyces cerevisiae are required for chromosome transmission." Mol Microbiol **31**(1): 349-60.

Poplawski, A., B. Grabowski, S. E. Long and Z. Kelman (2001). "The zinc finger domain of the archaeal minichromosome maintenance protein is required for helicase activity." J Biol Chem **276**(52): 49371-7.

Remus, D., F. Beuron, G. Tolun, J. D. Griffith, E. P. Morris and J. F. Diffley (2009). "Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing." Cell **139**(4): 719-30.

Ricke, R. M. and A. K. Bielinsky (2004). "Mcm10 regulates the stability and chromatin association of DNA polymerase-alpha." Mol Cell **16**(2): 173-85.

Ricke, R. M. and A. K. Bielinsky (2006). "A conserved Hsp10-like domain in Mcm10 is required to stabilize the catalytic subunit of DNA polymerase-alpha in budding yeast." J Biol Chem **281**(27): 18414-25.

Robertson, P. D., E. M. Warren, H. Zhang, D. B. Friedman, J. W. Lary, J. L. Cole, A. V. Tutter, J. C. Walter, E. Fanning and B. F. Eichman (2008). "Domain architecture and biochemical characterization of vertebrate Mcm10." J Biol Chem **283**(6): 3338-48.

Romanowski, P., M. A. Madine, A. Rowles, J. J. Blow and R. A. Laskey (1996). "The Xenopus origin recognition complex is essential for DNA replication and MCM binding to chromatin." Curr Biol **6**(11): 1416-25.

Roy, N., A. Poddar, A. Lohia and P. Sinha (1997). "The mcm17 mutation of yeast shows a size-dependent segregational defect of a mini-chromosome." Curr Genet **32**(3): 182-9.

Schuck, P. (2000). "Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling." Biophys J **78**(3): 1606-19.

Schuck, P. (2003). "On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation." Analytical Biochemistry **320**(1): 104-124.

Schwacha, A. and S. P. Bell (2001). "Interactions between two catalytically distinct MCM subgroups are essential for coordinated ATP hydrolysis and DNA replication." Mol Cell **8**(5): 1093-104.

Shamoo, Y., A. M. Friedman, M. R. Parsons, W. H. Konigsberg and T. A. Steitz (1995). "Crystal structure of a replication fork single-stranded DNA binding protein (T4 gp32) complexed to DNA." Nature **376**(6538): 362-6.

Sheu, Y. J. and B. Stillman (2006). "Cdc7-Dbf4 phosphorylates MCM proteins via a docking site-mediated mechanism to promote S phase progression." Mol Cell **24**(1): 101-13.

Smith, D. B. and K. S. Johnson (1988). "Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase." Gene **67**(1): 31-40.

Solomon, N. A., M. B. Wright, S. Chang, A. M. Buckley, L. B. Dumas and R. F. Gaber (1992). "Genetic and molecular analysis of DNA43 and DNA52: two new cell-cycle genes in Saccharomyces cerevisiae." Yeast **8**(4): 273-89.

Speck, C. and W. Messer (2001). "Mechanism of origin unwinding: sequential binding of DnaA to double- and single-stranded DNA." Embo J **20**(6): 1469-76.

Stafford, W. F. (1992). "Boundary Analysis in Sedimentation Transport Experiments - a Procedure for Obtaining Sedimentation Coefficient Distributions Using the Time Derivative of the Concentration Profile." Analytical Biochemistry **203**(2): 295-301.

Stafford, W. F. and P. J. Sherwood (2004). "Analysis of heterologous interacting systems by sedimentation velocity: curve fitting algorithms for estimation of sedimentation coefficients, equilibrium and kinetic constants." Biophys Chem **108**(1-3): 231-43.

Stauffer, M. E. and W. J. Chazin (2004). "Physical interaction between replication protein A and Rad51 promotes exchange on single-stranded DNA." J Biol Chem **279**(24): 25638-45.

Stauffer, M. E. and W. J. Chazin (2004). "Structural mechanisms of DNA replication, repair, and recombination." J Biol Chem **279**(30): 30915-8.

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

Sugino, A. (1995). "Yeast DNA polymerases and their role at the replication fork." Trends Biochem Sci **20**(8): 319-23.

Takahashi, K., H. Yamada and M. Yanagida (1994). "Fission yeast minichromosome loss mutants mis cause lethal aneuploidy and replication abnormality." Mol Biol Cell **5**(10): 1145-58.

Takayama, Y., Y. Kamimura, M. Okawa, S. Muramatsu, A. Sugino and H. Araki (2003). "GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast." Genes Dev **17**(9): 1153-65.

Tanaka, H., Y. Katou, M. Yagura, K. Saitoh, T. Itoh, H. Araki, M. Bando and K. Shirahige (2009). "Ctf4 coordinates the progression of helicase and DNA polymerase alpha." Genes Cells **14**(7): 807-20.

Tanaka, S., T. Umemori, K. Hirai, S. Muramatsu, Y. Kamimura and H. Araki (2007). "CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast." Nature **445**(7125): 328-32.

Tanaka, T. and K. Nasmyth (1998). "Association of RPA with chromosomal replication origins requires an Mcm protein, and is regulated by Rad53, and cyclin- and Dbf4-dependent kinases." Embo J **17**(17): 5182-91.

Thompson, J. D., D. G. Higgins and T. J. Gibson (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Res **22**(22): 4673-80.

Tye, B. K. (1999). "MCM proteins in DNA replication." Annu Rev Biochem **68**: 649-86.

Tye, B. K. (2000). "Insights into DNA replication from the third domain of life." Proc Natl Acad Sci U S A **97**(6): 2399-401.

Voitenleitner, C., E. Fanning and H. P. Nasheuer (1997). "Phosphorylation of DNA polymerase alpha-primase by cyclin A-dependent kinases regulates initiation of DNA replication in vitro." Oncogene **14**(13): 1611-5.

Volkening, M. and I. Hoffmann (2005). "Involvement of human MCM8 in prereplication complex assembly by recruiting hcdc6 to chromatin." Mol Cell Biol **25**(4): 1560-8.

Walter, J. and J. Newport (2000). "Initiation of eukaryotic DNA replication: origin unwinding and sequential chromatin association of Cdc45, RPA, and DNA polymerase alpha." Mol Cell **5**(4): 617-27.

Walter, J., L. Sun and J. Newport (1998). "Regulated chromosomal DNA replication in the absence of a nucleus." Mol Cell **1**(4): 519-29.

Warren, E. M., H. Huang, E. Fanning, W. J. Chazin and B. F. Eichman (2009). "Physical interactions between Mcm10, DNA, and DNA polymerase alpha." J Biol Chem **284**(36): 24662-72.

Warren, E. M., S. Vaithiyalingam, J. Haworth, B. Greer, A. K. Bielinsky, W. J. Chazin and B. F. Eichman (2008). "Structural basis for DNA binding by replication initiator Mcm10." Structure **16**(12): 1892-901.

Wohlschlegel, J. A., S. K. Dhar, T. A. Prokhorova, A. Dutta and J. C. Walter (2002). "Xenopus Mcm10 binds to origins of DNA replication after Mcm2-7 and stimulates origin binding of Cdc45." Mol Cell **9**(2): 233-40.

Wolf, E., P. S. Kim and B. Berger (1997). "MultiCoil: a program for predicting two- and three-stranded coiled coils." Protein Sci **6**(6): 1179-89.

Wu, P. Y. and P. Nurse (2009). "Establishing the program of origin firing during S phase in fission Yeast." Cell **136**(5): 852-64.

Xu, X., P. J. Rochette, E. A. Feyissa, T. V. Su and Y. Liu (2009). "MCM10 mediates RECQ4 association with MCM2-7 helicase complex during DNA replication." Embo J **28**(19): 3005-14.

Yan, H., S. Gibson and B. K. Tye (1991). "Mcm2 and Mcm3, two proteins important for ARS activity, are related in structure and function." Genes Dev **5**(6): 944-57.

Yang, X., J. Gregan, K. Lindner, H. Young and S. E. Kearsey (2005). "Nuclear distribution and chromatin association of DNA polymerase alpha-primase is affected by TEV protease cleavage of Cdc23 (Mcm10) in fission yeast." BMC Mol Biol **6**: 13.

Yoshida, K. (2005). "Identification of a novel cell-cycle-induced MCM family protein MCM9." Biochem Biophys Res Commun **331**(2): 669-74.

Zegerman, P. and J. F. Diffley (2007). "Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast." Nature **445**(7125): 281-5.

Zheng, N., B. A. Schulman, L. Song, J. J. Miller, P. D. Jeffrey, P. Wang, C. Chu, D. M. Koepp, S. J. Elledge, M. Pagano, R. C. Conaway, J. W. Conaway, J. W. Harper and N. P. Pavletich (2002). "Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex." Nature **416**(6882): 703-9.

Zhu, W., C. Ukomadu, S. Jha, T. Senga, S. K. Dhar, J. A. Wohlschlegel, L. K. Nutt, S. Kornbluth and A. Dutta (2007). "Mcm10 and And-1/CTF4 recruit DNA polymerase alpha to chromatin for initiation of DNA replication." Genes Dev **21**(18): 2288-99.

Zou, L. and B. Stillman (2000). "Assembly of a complex containing Cdc45p, replication protein A, and Mcm2p at replication origins controlled by S-phase cyclin-dependent kinases and Cdc7p-Dbf4p kinase." Mol Cell Biol **20**(9): 3086-96.