

## ABSTRACT

Title of dissertation: STRUCTURAL AND FUNCTIONAL ANALYSIS OF  
DNA REPLICATION INITIATION PROTEINS FROM  
THE ARCHAEON *METHANOTHERMOBACTER  
THERMAUTOTROPHICUS*

Rajesh Kasiviswanathan, Doctor of Philosophy, 2005.

Dissertation directed by: Professor Zvi Kelman  
Program in Molecular and Cell Biology.

The faithful duplication of the chromosome requires the combined efforts of numerous proteins. Cdc6 and MCM are two such proteins involved in the initiation of DNA replication. The genome of the euryarchaeon *Methanothermobacter thermautotrophicus* contains one MCM and two Cdc6 homologues (Cdc6-1 and -2). While MCM is the replicative helicase that unwind the duplex DNA to provide single-stranded DNA substrate for the replicative polymerases, the Cdc6 proteins are presumed to function in origin recognition and helicase assembly at the origin. This thesis elucidates the structure, function and regulation of these archaeal initiation proteins.

The *M. thermautotrophicus* MCM helicase is a dumb-bell shaped double hexamer. Each monomer can be divided into two portions. The C-terminal

catalytic region contains the ATP binding and hydrolysis sites essential for helicase activity. This thesis concentrates its efforts to determine the functional role of the N-terminal region. Using a variety of biochemical approaches it was found that the N-terminal portion of MCM is involved in hexamer/dodecamer formation. The study also identified two structural features at the N-terminus, the zinc- and the  $\beta$ -finger motifs, essential for DNA binding, which in turn is essential for helicase activity. In addition, the N-terminal portion of MCM interacts with both Cdc6 proteins.

The role of the Cdc6-1 and -2 proteins in origin recognition and helicase loading was also elucidated. The results presented in this thesis show that Cdc6-1 has binding specificity to origin DNA sequences suggesting a role for the protein in origin recognition. While both Cdc6 proteins interact with the MCM helicase, Cdc6-2 exhibited tighter binding compared to Cdc6-1 suggesting a role for Cdc6-2 in helicase loading.

Summarizing the observations of this study, a model for the replication initiation process in *M. thermotrophicus* has been proposed, outlining separate role for the two Cdc6 proteins, Cdc6-1 in origin recognition and Cdc6-2 in MCM helicase assembly at the origin.

STRUCTURAL AND FUNCTIONAL ANALYSIS OF DNA REPLICATION  
INITIATION PROTEINS FROM THE ARCHAEON *METHANOTHERMOBACTER*  
*THERMAUTOTROPHICUS*

By

RAJESH KASIVISWANATHAN

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park, in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
2005

Advisory Committee:

Professor Jeffrey DeStefano, Chair  
Professor James Culver  
Professor Douglas Julin  
Professor Zvi Kelman  
Professor David Straney

© Copyright by  
RAJESH KASIVISWANATHAN  
2005

To  
My Wife and Parents

## ACKNOWLEDGEMENTS

I express my profound gratitude to my advisor Dr. Zvi Kelman, whose guidance and constant encouragement have been a great source of inspiration in my professional and personal life. I consider myself very fortunate to have had an opportunity to work with him. His patience and intellectual curiosity will always guide me to work meticulously towards my goals and future endeavors.

My sincere thanks to Dr. Jeffrey DeStefano, the chair of my dissertation advisory committee and committee members Drs. Douglas Julin, James Culver and David Straney for their suggestions and critical evaluation of the work throughout the duration of my graduate study.

I particularly acknowledge Dr. Todd Cooke, the ex-Graduate Director of the Department of Cell Biology and Molecular Genetics, for having confidence in me and allowing me to do a third rotation in Zvi's lab. If not for his patience and understanding, this whole graduate school would have just remained as my unfulfilled dream.

I would like to thank my lab members for their substantial contributions in terms of discussions, critique, development of the projects and publications related to my research. I would like to thank Beatrice for her help during the initial stages of my work, Jae-Ho for meaningful discussions throughout my work and Nozomi, Gun Young, Gyri for their support and encouragement during the later stages of my work. I would also like to thank Eugene for his help with bioinformatics. I would also like to acknowledge Zvi's wife, Lori Kelman, who has

constantly offered help in manuscript corrections for its English and scientific content.

I like to acknowledge Harold for his time and suggestions during the initial stages of the yeast two hybrid experiments and his lab members Madhura, Daya and Antonio for their friendship. Thanks to Jim and Fred, and their lab members and to everyone in the third floor for their insightful discussions during the floor meetings. CARB is a unique environment, which provided me with ample opportunities to socialize. I would like to thank everyone who made me feel at home.

I would like to thank all my friends, especially Dr. Nut, Sai, and Coop for their company, encouragement and for all their financial help during the first two years of my graduate school.

I take this opportunity to thank my parents Lalitha and Kasiviswanathan for letting me go behind my dreams with their incessant support, love and blessings. I also thank my brother, Dinesh for taking care of everything back in India on my behalf in my absence. Also thanks to my in-laws, Lakshmi and Gopalan and family for their belief in me and my career. Special thanks to my grand parents for instilling the desire to learn and be a better person.

This thesis would not have been possible without the never ending motivation, emotional support and sincere prayers of my wife, Mathangi. Thanks a ton, wifey, I am waiting to spend real quality time with you pretty soon.

Finally, I thank almighty for making this learning experience fun and memorable.

# TABLE OF CONTENTS

List of tables	vi
List of figures	vii
List of abbreviations	ix
<b>CHAPTER 1: INTRODUCTION</b>	<b>1</b>
1.1 DNA REPLICATION: AN OVERVIEW	1
1.2 <i>METHANOTHERMOBACTER THERMAUTOTROPHICUS</i>	6
1.3 <i>M. THERMAUTOTROPHICUS</i> INITIATION PROTEINS	7
1.4 ORIGINS OF REPLICATION IN ARCHAEA	8
1.5 ORIGIN RECOGNITION COMPLEX AND Cdc6 PROTEINS	10
1.6 MINICHROMOSOME MAINTENANCE (MCM) HELICASES	18
<b>CHAPTER 2: BIOCHEMICAL AND STRUCTURAL CHARACTERIZATION OF THE MCM N-TERMINAL DOMAINS</b>	<b>28</b>
2.1 ABSTRACT	28
2.2 INTRODUCTION	28
2.3 METHODS	30
2.4 RESULTS	37
2.5 DISCUSSION	49
<b>CHAPTER 3: MUTATIONAL ANALYSIS OF THE MCM HELICASE</b>	<b>56</b>
3.1 ABSTRACT	56
3.2 INTRODUCTION	56
3.3 METHODS	59
3.4 RESULTS AND DISCUSSION	62
<b>CHAPTER 4: INTERACTIONS BETWEEN Cdc6 AND MCM PROTEINS MODULATE THEIR BIOCHEMICAL PROPERTIES</b>	<b>65</b>
4.1 ABSTRACT	65
4.2 INTRODUCTION	65
4.3 METHODS	67
4.4 RESULTS	74
4.5 DISCUSSION	91
<b>CHAPTER 5: DNA BINDING BY THE Cdc6 PROTEIN IS REGULATED BY MCM HELICASE</b>	<b>96</b>
5.1 ABSTRACT	96
5.2 INTRODUCTION	97
5.3 METHODS	98
5.4 RESULTS	101
5.5 DISCUSSION	111
<b>CHAPTER 6: CONCLUSION</b>	<b>113</b>
Appendices	119
References	122



## LIST OF TABLES

	Page
1. A comparison of chromosomal DNA replication in the three domains of life.	2
2. Oligonucleotide used for cloning MCM mutants in pDBLeu and pPC86 vectors.	31
3. Oligonucleotide used for cloning MCM mutants in pET-21a vector.	33
4. Location of mutations in the putative dodecameric region of MCM.	60
5. Positions of oligonucleotides used to construct MCM mutants.	60
6. Sequences of oligonucleotides used for constructing MCM mutants.	61
7. Oligonucleotide used for cloning proteins in pDBLeu, pPC86 and pET-21a vectors.	68
8. Terminology of Cdc6 and MCM proteins used in this study with amino acids and mutant descriptions.	69

## LIST OF FIGURES

	Page
1. Model for initiation of DNA replication.	3
2. Model for the elongation phase of DNA replication.	4
3. Model for the Okazaki fragment maturation.	5
4. Phylogenetic tree of life.	7
5. Three dimensional structure of the ORC/Cdc6 homologue from <i>P.aerophilum</i> .	13
6. Electron microscopic structures of <i>M. thermautotrophicus</i> MCM.	20
7. Three dimensional structures of the N-terminal portion of <i>M. thermautotrophicus</i> MCM.	21
8. Ribbon diagram of the domain structure of the N-terminal MCM.	22
9. PCR cloning strategy and oligonucleotides used to generate MCM constructs.	32
10. Analysis of the MCM multimerization domain.	38
11. Domain C is needed for MCM oligomeric structure.	39
12. Domain C of the N-terminal portion of MCM is multimeric.	41
13. MCM multimerization is not required for ssDNA binding.	43
14. Domain B of MCM is essential for its ssDNA binding activity.	44
15. The N-terminal MCM is needed for DNA-dependent ATPase activity.	46
16. MCM multimerization is needed for its helicase activity.	47
17. An intact N-terminus is essential for dsDNA translocation by MCM.	48

	Page
18. Sequence conservation patterns among archaeal MCM proteins.	53
19. Sequence alignment of the N-terminal MCM.	55
20. X-ray crystal structure of the N-terminal MCM.	59
21. Purified MCM mutant proteins.	63
22. Size determination of MCM mutants.	64
23. MCM interacts with Cdc6-1.	74
24. Cdc6-1 protein interacts with MCM protein in a two-hybrid analysis.	77
25. Cdc6 proteins interact with MCM protein in a Far western analysis.	80
26. Cdc6-1 and -2 proteins interact with MCM in solution.	83
27. Cdc6–MCM interaction is required for the inhibition of MCM translocation along DNA.	84
28. An intact WH domain of Cdc6 is needed for DNA binding.	86
29. Cdc6 autophosphorylation is regulated by MCM binding.	90
30. Cdc6-1 but not Cdc6-2 exhibit preferential binding to origin DNA.	103
31. Cdc6 chimera proteins can not bind DNA.	105
32. The $\beta$ -finger of MCM is essential for ss and dsDNA binding.	107
33. MCM inhibits DNA binding by Cdc6.	109
34. Cdc6-1 reduces, but does not abolish the DNA binding by MCM.	110
35. Models for the helicase assembly at the <i>M. thermautotrophicus</i> origin of replication.	117

## LIST OF ABBREVIATIONS

AAA <sup>+</sup>	ATPases associated with cellular activities
AD	Activation domain
bp	base pair
BSA	Bovine serum albumin
CSM	Complete supplement mixture
DB	DNA binding domain
dsDNA	Double-stranded DNA
EM	Electron microscope
IR	Inverted repeat
kDa	Kilodalton
MBP	Maltose binding protein
MCM	Minichromosome maintenance
OB	Oligonucleotide/oligosaccharide binding
ORB	Origin recognition box
ORC	Origin recognition complex
ORP	Origin recognition protein
PCNA	Proliferating cell nuclear antigen
SDS	Sodium dodecyl sulfate
SDS-PAGE	SDS-Polyacrylamide gel electrophoresis
ssDNA	Single-stranded DNA
WH	Winged-helix

# CHAPTER 1

## INTRODUCTION

### 1.1 DNA REPLICATION: AN OVERVIEW

Chromosomal DNA replication is a complex event involving dozens of proteins and enzymes that ensures the precise and timely duplication of the genetic information. The mechanism of DNA replication is functionally and often structurally conserved in all organisms (1,2). The multi step process can be divided into three stages, namely, initiation, elongation and termination. Much is known about the elongation phase in all three domains of life (bacteria, eukarya and archaea). The events leading to the initiation of replication is well understood in bacteria. In eukarya, although many proteins involved in the process have been identified, not much is known about their function and biochemical properties. In archaea, the replication proteins identified to date are more similar to those in eukarya than to bacteria, based on sequence similarities and biochemical data. Although archaeal DNA replication appears to be similar to that of eukarya, it is a simpler version of it. A comparison of the proteins involved in the chromosomal DNA replication from all three domains of life is summarized in Table 1.

The replication process starts at specific chromosomal region(s), the origin(s) of replication, which acts as a binding site for origin recognition proteins (ORP). During the initiation process (Figure 1), the ORP bind to the origin and locally unwinds the DNA duplex. The ORP then recruits additional initiation

factors to the origin to facilitate the initiation process. Next the helicase is recruited to the DNA and assembled around the DNA at the origin. The helicase couples the hydrolysis of nucleoside triphosphates (NTPases) to nucleic acid unwinding, creating an initial replication bubble which bi-directionally expands away from the origin. The exposed ssDNA, left behind the helicase, is then coated with ssDNA binding protein (SSB) to serve as template for primase, polymerase and other proteins participating in the elongation phase.

**Table 1.** A comparison of chromosomal DNA replication in the three domains of life<sup>1</sup>

	Bacteria	Archaea	Eukarya
Chromosome	Linear or Circular	Circular	Linear
Replication origin(s)	Single	Single or Multiple	Multiple
Origin recognition	DnaA	Cdc6/ORC <sup>2</sup>	ORC
Helicase	DnaB	MCM	MCM
Helicase loader	DnaA and DnaC	Cdc6/ORC <sup>2</sup>	Cdc6 and Cdt <sup>3</sup>
ssDNA binding protein	SSB	SSB or RPA <sup>4</sup>	RPA
Primase	DnaG	Primase	Pol $\alpha$ /Primase
Sliding clamp	$\beta$ -subunit	PCNA <sup>5</sup>	PCNA
Clamp loader	$\gamma$ -complex	RFC	RFC
Polymerase	PolC	PolD and/or PolB <sup>6</sup>	PolB

<sup>1</sup> Bacterial-like features and proteins are in red, eukaryal-like features and proteins are in blue and archaeal-specific proteins are in green.

<sup>2</sup> As eukaryotic Cdc6 is similar to subunits of ORC, and the function of the archaeal homologues are not yet clear, they are referred to as Cdc6/ORC homologues.

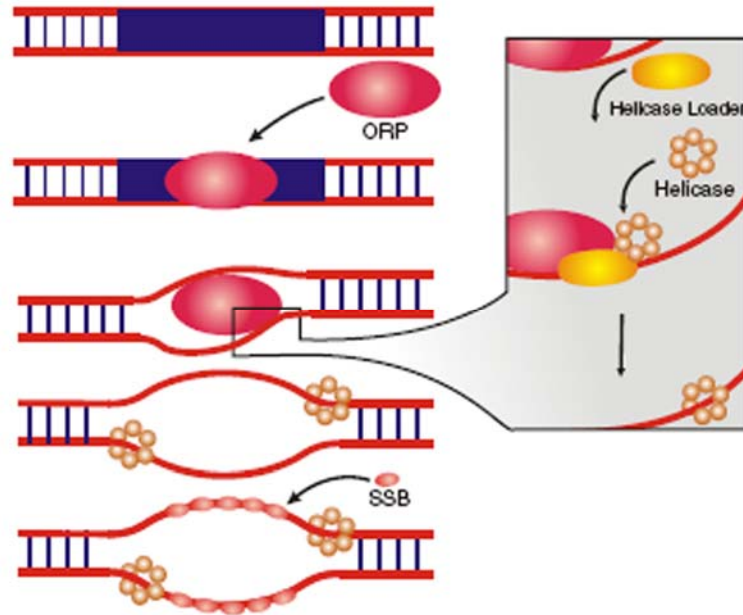
<sup>3</sup> Other factors may also be involved.

<sup>4</sup> Depends on the organism.

<sup>5</sup> In the euryarchaeal species PCNA is a single polypeptide, while in the crenarchaeal species three polypeptides have been identified.

<sup>6</sup> In the euryarchaeal species in which PolD was identified it is not yet clear whether it or PolB is the replicative enzyme.

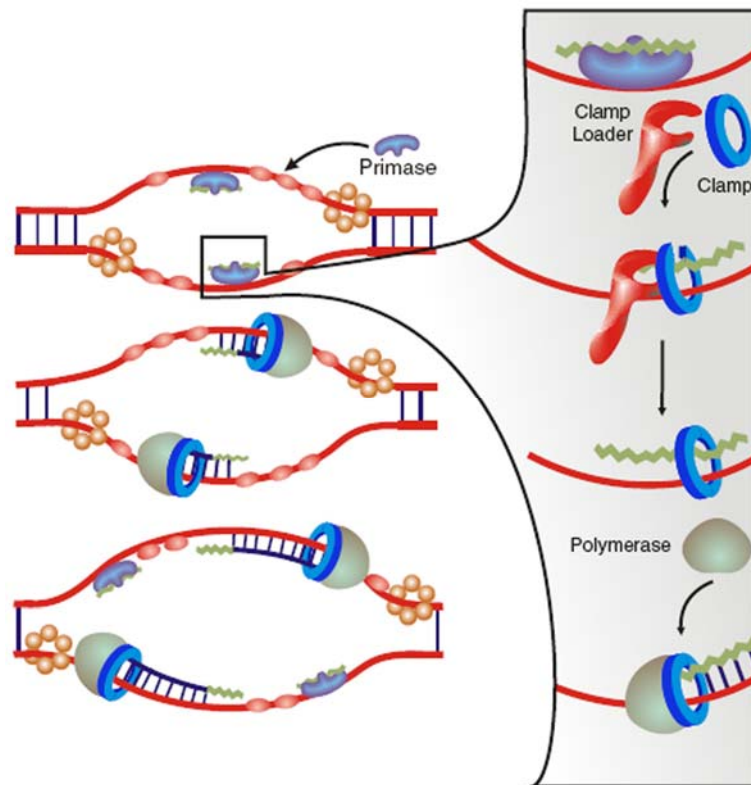
[Adapted from (3)]



**Figure 1. Model for initiation of DNA replication.** The ORP locally unwinds the DNA at the origin. Helicase loader then recruits the helicase at the origin, which unwinds the duplex DNA. Finally, the SSB coats the ssDNA. [Adapted from (1)].

The elongation phase (Figure 2) starts with the assembly of primase to the SSB-DNA complex, which synthesizes short RNA primers. The replicative polymerase then recruited to the replication bubble utilizes the RNA primers to initiate the processive bidirectional DNA synthesis (4). The processivity of the DNA polymerase is ensured by a ring-shaped factor (sliding clamp) that encircles double stranded DNA and binds to the catalytic unit of the polymerase thereby tethering it to the template DNA. Since the sliding clamp has no affinity for DNA, the accessory complex (clamp loader) assembles the sliding clamp on to the DNA. DNA topoisomerases are required to ease the tension created on either side of the replication fork created by the action of replicative helicase,

polymerase and SSB and ensures the effective synthesis of the daughter DNA strand.

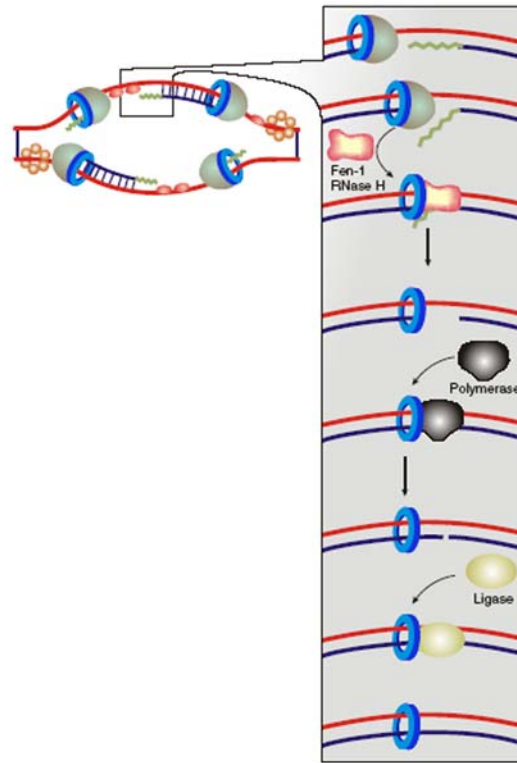


**Figure 2. Model for the elongation phase of DNA replication.** RNA primers are synthesized by primase at the replication fork. The clamp is assembled on to the DNA by the clamp loader, followed by association with the polymerase, which begins processive DNA synthesis at the primed site. [Adapted from (1)].

During chromosomal DNA replication one strand is synthesized continuously (the leading strand) while the other strand (the lagging strand) is synthesized in short stretches called the Okazaki fragments (4). These fragments are joined together through the concerted activity of many enzymes like flap endonuclease 1(Fen-1), RNase H and DNA ligase. Fen-1 and RNase H degrade the RNA fragments from the RNA-DNA hybrid produced by the primase



enzyme during the elongation phase. Finally, the polymerase fills the gaps created by the Fen-1 and RNase H enzymes and the DNA ligase enzyme ligate them to create a mature duplex DNA (Figure 3).



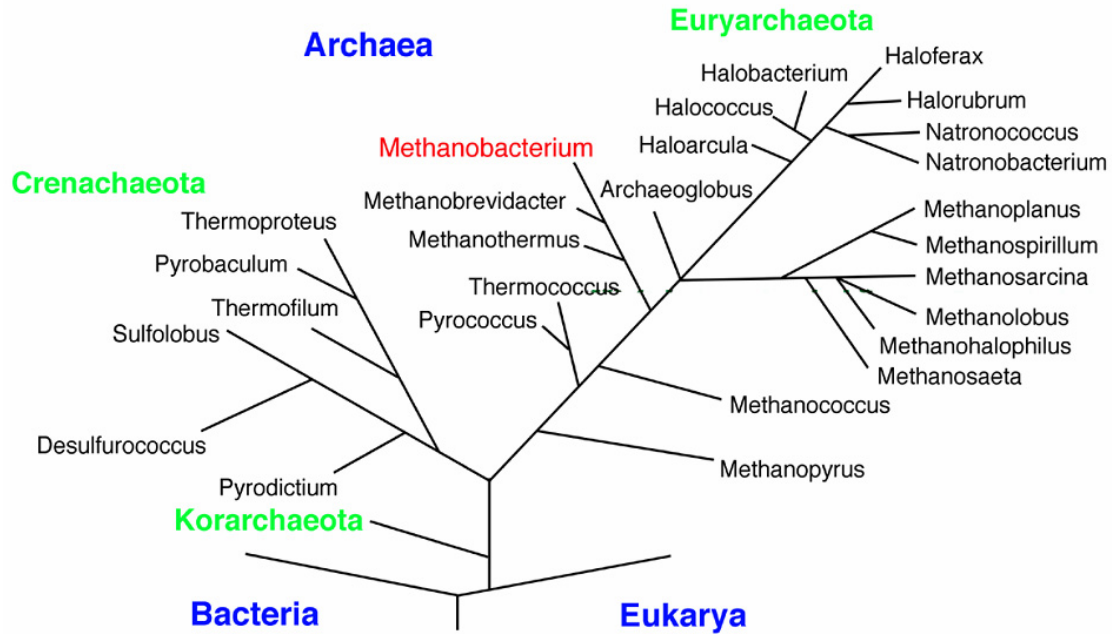
**Figure 3. Model for the Okazaki fragment maturation.** Fen-1 and RNase H remove the RNA primers from the Okazaki fragments. The resulting gaps are then filled and ligated by polymerase and ligase, respectively. [Adapted from (1)].

The initiation phase is a key regulatory stage in replication, as much of the regulation occurs during this step, for example, restricting DNA replication to occur once per cell cycle in order to prevent any deleterious effects of indiscriminate re-replication before the end of the cell division. Hence, it is vital to understand the mechanisms and regulation processes involved in initiating DNA replication.

## **1.2 METHANOTHERMOBACTER THERMAUTOTROPHICUS**

Archaea, considered as the third domain of life, can be divided into three kingdoms: euryarchaeaota, crenarchaeaota and korarchaeaota (Figure 4) (5). To date, euryarchaeaota is the most diverse group and its members include halophiles, hyperthermophiles, methanogens and thermophilic methanogens. The crenarchaeal group is less diverse and consists of members from hyperthermophiles, psychophiles and thermoacidophiles. The extent of diversity of the korarchaeal kingdom is still unclear as only a few members have been identified.

The model organism used in this study to elucidate the mechanism of DNA replication is the euryarchaeon *Methanothermobacter thermautotrophicus* (previously known as *Methanobacterium thermoautotrophicum*  $\Delta$ H). *M. thermautotrophicus* is a lithoautotrophic thermophilic archaeon with an optimal growth temperature of 65°C and a generation time of about 5 hours (6). The complete genome sequence of *M. thermautotrophicus* is 1,751,377 bp with 1855 open reading frames (ORFs) (7). Comparison of the genome sequence with other archaeal, bacterial and eukaryal sequence databases revealed 54% similarity to other archaeal sequences, 42% similarity to bacterial and 13% similarity to eukaryal sequences. However, like most other archaeal proteins, *M. thermautotrophicus* proteins involved in DNA metabolism, transcription and translation are more similar to eukaryal sequences (7).



**Figure 4. Phylogenetic tree of life.** The tree is constructed based on 16S rRNA sequences. The three domains of life are highlighted in blue. Green represents the three kingdoms of the archaeal domain and red, the model organism used in this study [Adapted from (8)].

### 1.3 *M. THERMAUTOTROPHICUS* INITIATION PROTEINS

Homologues of several eukaryotic initiation proteins (e.g., ORC, Cdc6 and MCM) have been identified in the *M. thermautotrophicus* genome. The structure, function and biochemical properties of the two ORC/Cdc6 homologues and one minichromosome maintenance (MCM) homologue identified in the *M. thermautotrophicus* genome is the primary focus of this study. The following sections will summarize the current knowledge about the archaeal initiation proteins in general with an emphasis on *M. thermautotrophicus* proteins. The similarities and differences between *M. thermautotrophicus* initiation proteins with other archaeal, eukaryal and bacterial proteins are also discussed.

The proteins involved in the elongation phase from this organism have been extensively studied and well characterized. However, since elongation phase is beyond the scope of this study, the reader is referred to several reviews on this subject [reviewed in (1,2,9-11)].

#### **1.4 ORIGINS OF REPLICATION IN ARCHAEA**

Although the replication proteins from archaea show little sequence similarities with their bacterial counterparts, they share a similar mode of replication and chromosome structure. However, well-characterized replication origins from all the three domains of life, including viruses, share some similar characteristics (12,13). All origins of replication are A/T rich sequences with stretches of one or more A/T rich regions essential for origin function. A/T rich regions are easier to unwind compared to G/C regions and hence could be more readily bent, which might be appropriate for origin protein binding. Most of the origins also share another common feature, which are the inverted repeat (IR) sequences. These sequences range from a few nucleotides to greater than 100 bases and are thought to assist the binding of initiator protein to the origin. Under supercoiling conditions, like those found *in vivo* these IR sequences may form cruciform structures with a stem and single stranded loops, which might facilitate the local unwinding of duplex DNA for initiating DNA replication.

The predicted archaeal origins from several species using *in silico* analysis revealed a single origin of replication in some archaea like *M. thermautotrophicus*, *Methanosarcina mazei* and *Pyrococcus furiosus*, and

multiple origins in *Halobacterium sp* and *Methanocaldococcus jannaschii* (14-19). However, this analysis failed to detect any origin(s) in some archaea like *Sulfolobus solfataricus*.

Similar to bacteria, in which the gene encoding the origin recognition protein (DnaA) is located in close proximity to the bacterial origin, the *in silico* analysis revealed that in many archaeal species the gene encoding the archaeal homologue of the eukaryotic initiator protein, Cdc6, is also located near the vicinity of the predicted origin sequence (3). Marker frequency analysis of *Archaeoglobus fulgidus* also identified a single putative origin in its genome (20).

The first *in vivo* analysis for determining the origin of replication in archaea, *Pyrococcus abyssi*, by pulse-field gel electrophoresis (21) and two-dimensional gel electrophoresis (22) revealed that replication initiates from a single origin, as previously identified by *in silico* analysis and proceeds bi-directionally to terminate in a chromosomal region located opposite to the origin. However, the same techniques failed to identify an origin for *M. thermautotrophicus*. Recently, the presence of multiple replication origins in the chromosomes of *S. solfataricus* and *Sulfolobus acidocaldarius* have been experimentally identified (23,24). Using two-dimensional gel electrophoresis (24), it was demonstrated that *S. solfataricus* contain two origins of replication, which are located upstream of two of the three Cdc6 homologues (Cdc6-1 and – 3) found in the organism. An independent *in vivo* study using marker frequency analysis and whole-genome microarray analysis (23), revealed three separate origins in both *S. solfataricus* and *S. acidocaldarius*. Both these studies revealed

bi-directional DNA synthesis from the origins, similar to that of the bacterial mode of replication. However, the most striking feature distinct from the bacterial mode of replication is the discovery of multiple origins, which are characteristic of the eukaryotic mode of replication. Hence, although archaea are prokaryotic, they share certain features of eukaryotic DNA replication. The following sections will demonstrate the closeness of archaeal replication initiation proteins/enzymes to their eukaryotic counterpart.

### **1.5 ORIGIN RECOGNITION COMPLEX AND Cdc6 PROTEINS**

In eukarya, the origin binding protein (OBP) is a six-subunit origin recognition complex (ORC), thought to be responsible for initiation (25). The initiator protein, Cdc6, associates with ORC prior to S phase. These proteins along with additional factors regulate the loading of minichromosome maintenance (MCM) helicase onto the origin DNA (25).

The eukaryal Cdc6 protein shares amino acid sequence similarities with three subunits of ORC (Orc1, 4 and 5). Since the function of the archaeal homologue as ORC or Cdc6 remains unknown, it is referred as ORC/Cdc6. In most archaea, with known sequence information, at least one homologue of ORC/Cdc6 has been identified, with most of them containing two homologues (26). However, exceptions do exist. In the case of *M. jannaschii* and *Methanopyrus kandleri*, no clear ORC/Cdc6 homologue was found (26) although a putative homologue has been suggested for *M. jannaschii* (27,28). Other exceptions include *S. solfataricus* with three and *Halobacterium* sp., with ten

homologues. The reason for the differences in the number of ORC/Cdc6 homologues in each genome is currently not known. The different copy numbers of these homologues in different archaea suggest that, these homologues may act as heteromultimers during the initiation process similar to eukaryal ORC (29). Furthermore, in many archaea, at least one of the ORC/Cdc6 homologues is located immediately downstream of the origin sequence, suggesting that they may function as the OBP (23,24,30), similar to the bacterial OBP, DnaA, which is located in close proximity to the origin and forms homomultimers at the origin to form an active pre-initiation complex (2,4). However, it is also possible that archaea may contain some unidentified and unique proteins which can interact with the ORC/Cdc6 proteins to form a functional OBP.

Similar to the bacterial DnaA, three subunits of eukaryal ORC (Orc1, 4 and 5), eukaryal Cdc6 protein, the archaeal ORC/Cdc6 homologues also belong to the AAA<sup>+</sup> superfamily of ATPases (31-33). Members belonging to this superfamily contain a purine nucleoside triphosphate binding site containing the characteristic Walker-A [GXXGXGKT(T/S)] and -B [D(D/E)XX] signature motifs (34). While the Walker-A motif is thought to be involved in ATP binding, the Walker-B motif is thought to be responsible for ATP hydrolysis (35,36).

### **Structures of archaeal ORC/Cdc6 homologues**

The crystal structure of the ORC/Cdc6 homologue from *Pyrobaculum aerophilum* revealed the two expected domains found within the members of the AAA<sup>+</sup> family (33). The structure also revealed a three-domain composition, consisting of the N-terminal domain I, containing a RecA-type fold that includes

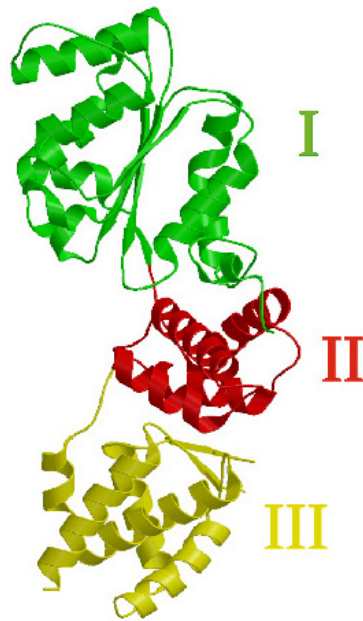
the Walker-A and -B motifs involved in ATP binding and hydrolysis respectively, linked to domain II, consisting mostly of  $\alpha$ -helices. An ATP binding pocket is located between these two domains. The C-terminus portion of the protein contains domain III, which is structurally related to the winged-helix (WH) domain. The WH domain is implicated in double-stranded (ds) binding (28) and has been identified on the basis of sequence similarities, at the C-terminus of Cdc6 and ORC subunits in all organisms studied (28,37).

The structure of another archaeal ORC/Cdc6 homologue from *Aeropyrum pernix* (38), suggested that its overall fold is similar to the homologue of *P. aerophilum* (28). In addition, the crystal structure also revealed the presence of ADP bound to the molecule, similar to the observation in the *P. aerophilum* ORC/Cdc6 structure (28). As suggested previously, this observation might indicate that these proteins form a very tight complex with ADP (39). Finally, the structure also revealed an intact DNA-binding motif at the WH domain at the C-terminus of the protein (28). Gel-mobility shift assays with truncated proteins lacking the C-terminus revealed that the WH domain is essential for DNA binding (38).

The structure of these two archaeal ORC/Cdc6 homologues is strikingly similar to that of the bacterial OBP, DnaA (40). The only notable difference between them lies in the domain responsible for DNA binding. While the archaeal ORC/Cdc6 homologues contain a WH domain responsible for DNA binding, the bacterial DnaA contain a helix-turn-helix (HTH) domain at its C-terminus (40). This difference can be attributed to the different substrate



recognition sequences by these two proteins. While the DnaA protein recognizes specific 9 bp DnaA-box sequences in the bacterial origin (*oriC*), the specific substrates for the ORC/Cdc6 homologues from *P. aerophilum* and *A. pernix* have not yet been determined.



**Figure 5.** Three dimensional structure of the ORC/Cdc6 homologue from *P.aerophilum*. Domain I is shown in green, domain II in red and domain III (winged-helix domain) in yellow. [Adapted from (28)].

Recent studies on the two ORC/Cdc6 homologues from *M. thermautotrophicus* have suggested that one of the homologues specifically recognizes a 13 bp A/T rich IR sequences inside its putative origin [(41), see also chapter 5]. This 13 bp region may be the conserved IR element that has been identified in the origins of several archaea (42). This 13 bp sequence may also be the archaeal specific origin recognition box (ORB), similar to the DnaA box found in bacteria, which may be recognized by the ORC/Cdc6 homologues

through its WH domain. Studies on the DNA binding of ORC/Cdc6 homologues from other archaea to the conserved ORB will determine its role in the archaeal origin.

### **Biochemical properties of archaeal ORC/Cdc6 proteins**

To date, only a limited number of biochemical studies have been reported on the archaeal ORC/Cdc6 homologues. The ORC/Cdc6 proteins from *M. thermautotrophicus* and *P. aerophilum* were found to possess low ATPase activity in the absence of additional factors (Z. Kelman, personal communication). This observation is similar to other replication initiation proteins belonging to the AAA+ family of ATPases like the eukaryal ORC, Cdc6 protein and bacterial DnaA [reviewed in (43)].

The *M. thermautotrophicus* and *P. aerophilum* ORC/Cdc6 homologues were shown to undergo autophosphorylation on Ser residues, which is inhibited in the presence of single stranded (ss) or double stranded (ds) DNA (39). Further, the autophosphorylation is not inhibited by dsDNA when the protein lacks the WH domain suggesting the role of this domain in dsDNA interaction (39). The autophosphorylation appears to be conserved among the eukaryotic proteins too, as the Cdc6 protein from the fission yeast, *Schizosaccharomyces pombe*, was also been suggested to undergo autophosphorylation (39). Though the autophosphorylation of the archaeal ORC/Cdc6 homologues were inhibited by both ss and dsDNA, DNA had no effect on the phosphorylation of the eukaryotic Cdc6 proteins (39). This might be due to the presence of an additional N-terminal extension found in eukaryotic Cdc6 proteins, which is

thought to be important for DNA binding (44). However, the role of autophosphorylation still remains unknown though it was suggested that the phosphorylation may play a regulatory role *in vivo* (39). Studies on ORC/Cdc6 homologue from another archaeon *S. solfataricus* (Cdc6-1) revealed that the protein was capable of autophosphorylation (45).

Biochemical studies on the two ORC/Cdc6 homologues (Cdc6-1 and -2) from *M. thermautotrophicus* suggest that they may also function as helicase loaders [(46,47), see also chapters 4, 5 and 6]. In bacteria, following the transient unwinding of the duplex DNA at the origin by the bacterial OBP, DnaA, the bacterial helicase loader DnaC assembles the replicative helicase DnaB at *oriC* (4,48). The helicase and ATPase activities of DnaB are inhibited when it is complexed with DnaC. The complex is dissociated when the ATP bound to DnaC is hydrolyzed making the DnaB helicase active (49,50). DnaC interaction with DnaB does not require ATP binding by DnaC, but is required for inhibiting the helicase activity of DnaB (51). This inhibition is thought to be an essential regulatory function of DnaC to prevent DnaB to function in an origin independent fashion. The Cdc6 protein in eukaryotes is also presumed to be the helicase loader acting in the same manner like bacterial DnaC loader, although no direct biochemical evidence is available to date (43). If Cdc6 is indeed the helicase loader then it should interact with the MCM helicase and prevent its function in an ATP dependent manner.

Studies conducted with the *M. thermautotrophicus* Cdc6-1 and -2 proteins suggest that both proteins inhibit *M. thermautotrophicus* MCM helicase activity in

an ATP dependent fashion and the WH domain of Cdc6 proteins are essential for this inhibition [(46,47), see also chapter 4]. The study also suggested that the inhibition of MCM activity by Cdc6 proteins may be primarily due to Cdc6-MCM interactions and not Cdc6-DNA binding [(46,47), see also chapter 4]. Furthermore, the results suggested that this inhibition is species specific in archaea as the *P. aerophilum* Cdc6 failed to inhibit *M. thermautotrophicus* MCM helicase activity [(46,47), see also chapter 4].

Studies conducted with the three ORC/Cdc6 homologues from *S. solfataricus* (Cdc6-1, -2 and -3) revealed that Cdc6-1 and -2 interact with *S. solfataricus* MCM and inhibit its ATPase and helicase activities (45,52). While the N-terminus portion of the protein exhibit ATPase activities, the C-terminal WH domain of Cdc6-2 binds DNA and inhibits the helicase activity of MCM (52). Studies on ORC/Cdc6 homologues from *A. fulgidus* show that the Cdc6 homologue could displace the *A. fulgidus* MCM pre-bound to DNA substrates (53). Yeast two hybrid analyses revealed that *M. thermautotrophicus* Cdc6-1 interacts with MCM [(46,47), see also chapter 4]. Further, the study also showed species specific interactions between Cdc6 and MCM. The ORC/Cdc6 homologue from *P. aerophilum* failed to interact with *M. thermautotrophicus* MCM [(46,47), see also chapter 4]. However, a detailed interaction study to determine the domain(s)/amino acids that participate in this protein-protein interaction and interactions between *M. thermautotrophicus* Cdc6-2 and MCM revealed that, while Cdc6-1 interacts with MCM via its WH domain, only full-length Cdc6-2 interacted with MCM [(47), see also chapter 4]. In addition, the study suggested

that this protein-protein interaction regulated Cdc6 autophosphorylation [(47), see also chapter 4].

DNA binding studies on the *M. thermotrophicus* Cdc6-1 and -2 proteins suggest that the only Cdc6-1 protein shows origin sequence specific (the 13 bp IR sequence) interactions with dsDNA [(41), see also chapter 5]. The study also identified two conserved arginine residues present in the DNA recognition helix of WH domain of Cdc6-1 protein essential for DNA binding. A mutation of these two arginine residues to alanine completely abolished the DNA binding [(41), see also chapter 5]. In addition, both Cdc6-1 and -2 proteins bound dsDNA in a cooperative fashion, characteristic of the bacterial DnaA binding to *oriC*. However, the ability of both Cdc6-1 and -2 proteins to bind ssDNA sequences revealed that, while Cdc6-1 had no ssDNA binding, Cdc6-2 bound ssDNA similar to dsDNA (see chapter 5). In addition, the role of Cdc6-MCM interactions in Cdc6-DNA binding suggested that MCM regulates the DNA binding of Cdc6 (see chapter 5).

Gel mobility shift assays for determining the DNA binding of the ORC/Cdc6 homologue from *A. fulgidus* suggest that the protein had preferential binding to bubble and fork substrates compared to ss and dsDNA (53). In the case of *A. pernix*, the ORC/Cdc6 homologue was shown to bind forked DNA through its C-terminal WH domain (38). The study also revealed that the N-terminal domain was completely devoid of DNA binding, while the C-terminus WH domain bound DNA better than the wild-type protein (38). Finally, DNA binding studies on *S. solfataricus* ORC/Cdc6 homologues revealed that Cdc6-1

bound dsDNA better than ssDNA (45). The Cdc6-2 protein in this archaea also showed preferential binding to fork and bubble DNA substrates compared to ssDNA (52,54).

## **1.6 MINICHROMOSOME MAINTENANCE (MCM) HELICASES**

MCM helicases are ring-shaped complexes that play an essential role in archaeal and eukaryal DNA replication by separating the two strands of chromosomal DNA to provide single-stranded substrate for the replicative polymerase. MCM homologues have been identified in all eukaryotes with highly conserved polypeptide sequences [reviewed in (55)] and comprise at least six structurally related proteins, MCM2-7. However, complexes consisting of various combinations of polypeptides can be formed within the cell (56). Genetic and biochemical data suggest that a dimeric complex of MCM4,6,7 heterotrimer is the putative replicative helicase in eukaryotes and that the MCM2, 3 and 5 subunits may play a role in regulation (56,57). However, in contrast to the bacterial DnaB replicative helicase, which has a 5'→3' polarity for DNA unwinding, the MCM4,6,7 heterohexamer possess 3'→5' polarity (56,58). This is similar to that of Simian Virus 40 (SV40) replicative helicase, the large T-antigen (T-Ag) which also has a 3'→5' polarity.

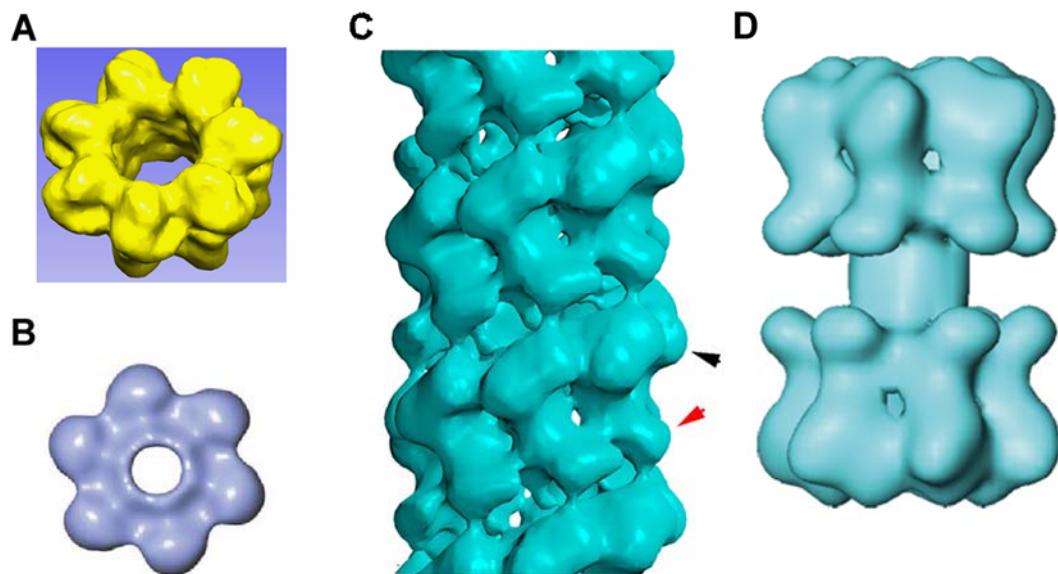
In the case of archaea for which the genomes have been sequenced, at least one homologue of MCM has been identified (26). However, similar to ORC/Cdc6 homologues, the number of MCM homologues also varies in archaea. While most of the archaea contains a single MCM homologue, exceptions do

exist as in the case of *M. jannaschii*, which contains four putative MCM homologues. *In vivo* studies on the MCM homologue from *P. abyssi* revealed that the protein is stable and expressed constitutively throughout the cell cycle (22). The study also estimated 200-400 MCM molecules in rapidly dividing cells and the MCM proteins dissociate from the DNA during the stationary phase or when DNA synthesis is inhibited (22). However, the reason for the dissociation of MCM from DNA during the stationary phase still remains unclear.

### **Structures of archaeal MCM helicases**

The only structural information available to date for any MCM protein either from eukarya or archaea is from the euryarchaea *M. thermautotrophicus*. Several structural studies carried out using various techniques have revealed somewhat different results for the *M. thermautotrophicus* MCM. The first clue for the structure of MCM came from a low resolution electron microscope (EM) image, which suggested that the proteins formed a hexameric ring structure (59). Consistent with the behavior of the protein in solution (59,60), a subset of the images obtained by EM also revealed a double hexameric structure (59). However, a detailed three dimensional reconstruction of the EM images revealed ring shaped heptamers (Fig. 6A) (61). Apparently, this study did not identify double hexamers or double heptamers. In contrast to this observation, an independent study using the three dimensional reconstruction of the EM images, suggested hexameric ring structure for the *M. thermautotrophicus* MCM protein (Fig. 6B) (62). Further, another study conducted from a subset of EM images obtained for heptameric structure for the protein, previously thought to be a stack

of heptameric rings (61), revealed that they were indeed filamentous helical structures (Fig. 6C) (63). These filaments had a 92Å helical pitch containing 7.2 subunits per helical turn (63). The study also suggested a ~24Å inner radius and ~75Å outer radius for the filaments (63). Recently, another three dimensional-EM reconstruction of the full-length MCM protein revealed double-hexameric structure consistent with the oligomeric form of the protein in solution (Fig. 6D) (64). The double hexamers were 182Å in length and 129Å in width (64).

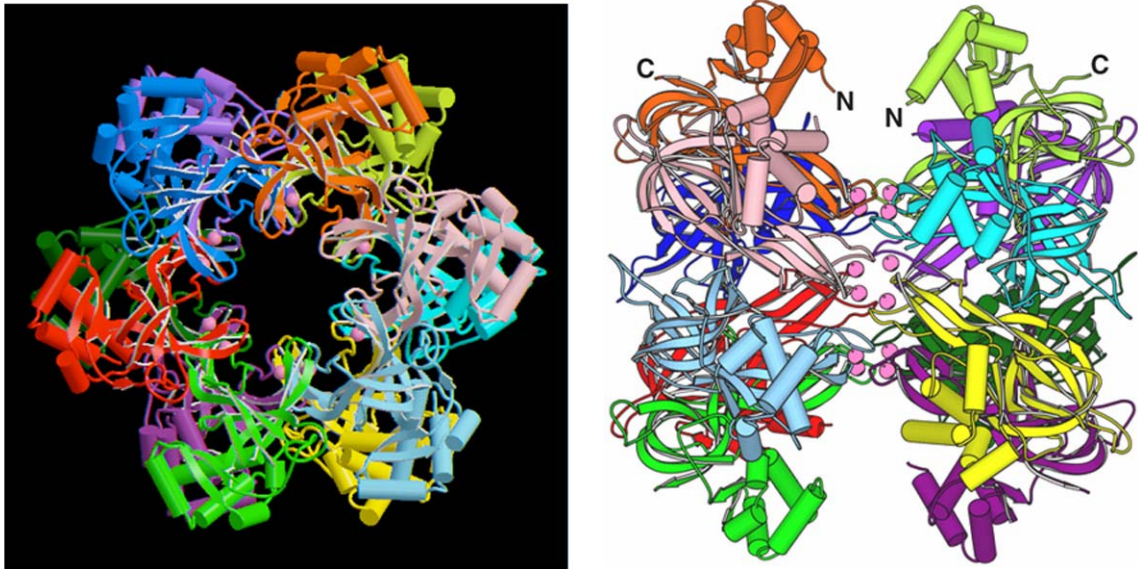


**Figure 6.** Electron microscopic structures of *M. thermautotrophicus* MCM. A, heptamer; B, hexamer; C, helical filaments; D, double hexamers. In panel C, red arrow indicates the N-terminal domain and black arrow indicates the C-terminal domain. [Adapted from (61-64)].

Finally, the high-resolution three dimensional crystal structure of the N-terminal portion of the MCM protein revealed a dumbbell shaped double hexameric structure (Fig. 7) (65) consistent with the observation of the full-length protein in solution (59,60). The crystal structure revealed an unusually long (118Å) and highly positively charged central channel capable of accommodating



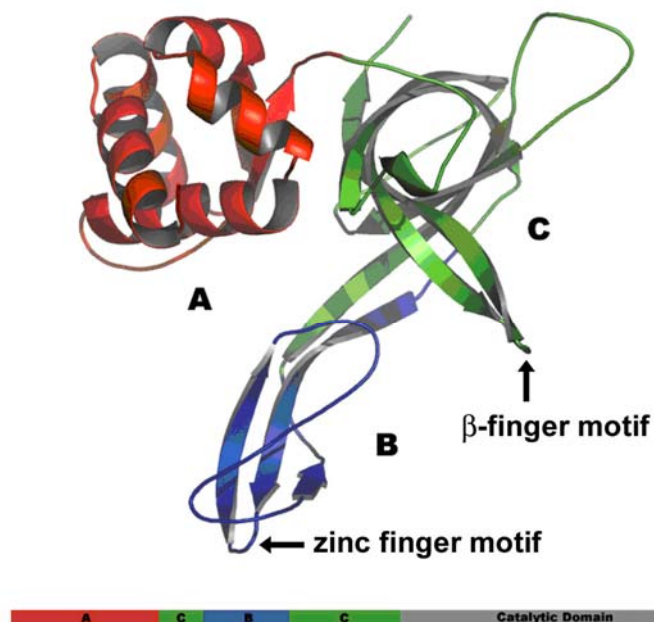
dsDNA (65) and a highly negatively charged outer surface. Structure assisted sequence alignment suggest a structural conservation for all six MCM proteins from eukaryotes in spite of their highly divergent N-terminal sequences, implicating that a similar double-hexameric structure could be formed by the N-terminal portion of eukaryotic MCM (65).



**Figure 7.** Three dimensional structures of the N-terminal portion of *M. thermautotrophicus* MCM. Left panel, top view and right panel, side view of the structure. Pink circles represent zinc atoms. [Adapted from (65)].

The overall structure also revealed a three domain composition for the N-terminal portion of MCM (Figure 8). Domain A, located farthest from the helical axis contains four  $\alpha$ -helices. Domain B contains three anti parallel  $\beta$ -strands and a zinc-finger motif thought to be important for holding the two hexamers intact. Domain C, positioned between domains A and B, comprises of five long  $\beta$ -strands that resembles an oligonucleotide / oligosaccharide binding (OB) fold

characteristic of many ssDNA binding proteins (66). A  $\beta$ -hairpin finger motif is also found in this domain, which projects inside the central cavity of the protein and is suggested to be important for dsDNA binding by the protein (65). This domain also connects the N-terminal part of the protein to the C-terminal catalytic region, which harbors the ATP binding site.



**Figure 8. Ribbon diagram of the domain structure of the N-terminal MCM.** The ribbon diagram was constructed using the Pymol program. Domain A (Red), residues 4-92; domain B (Blue): residues 120-169; and domain C (green), residues 93-119 and 170-242. The linear structure of the protein is shown at the bottom. The locations of the two structural motifs involved in DNA binding are indicated by arrows. (Figure generated by Eugene Melamud).

The structure also suggests that two  $\beta$ -strands from domains B and C mediate the monomer-monomer interactions inside each hexamers. In addition to these  $\beta$ -strands, the study predicted a combination of salt bridges, hydrogen

bonds and hydrophobic interactions involved in the stabilization of these hexamers (65).

Although different structural information is available for *M. thermautotrophicus* MCM, the structure of the protein in its active form still remains unknown. The structural variations observed may be due to the differences in the protein concentration used in each analysis. While the EM study uses relatively low concentration of proteins (0.1-0.2 $\mu$ M as monomer) (62) to view single molecule, the crystal structure and gel filtration experiment uses relatively high concentrations (10 $\mu$ M as monomer) (60). However, the concentration at which the protein oligomerizes to form a functional complex still remains to be elucidated. Hence, more studies that concentrate on the structure of the protein in its active form has to be carried out to reveal the true form of the protein inside the cell.

Gel filtration experiments have revealed that, unlike the *M. thermautotrophicus* MCM, which exist as double hexamers in solution (59,60,67), the *S. solfataricus* and *A. fulgidus* MCM exist as hexamers in solution (53,68). However, it is interesting to note that while the eukaryotic MCM remains a heterohexamer in solution, it can form double hexamers only in the presence of fork DNA structures (57).

### **Biochemical properties of archaeal MCM helicases**

DNA helicases like archaeal and eukaryal MCM helicases, and bacterial DnaB helicase, are motor proteins that transiently catalyze the unwinding of duplex DNA molecules by utilizing the energy derived from the hydrolysis of

NTPs. All known DNA helicases share three common biochemical properties: (i) nucleic acid binding, (ii) NTP/dNTP binding and hydrolysis and (iii) polarity of duplex DNA unwinding that depends on NTP/dNTP hydrolysis. Hence, all DNA helicases possess intrinsic DNA-dependent NTPase activity (4).

Consistent with the hypothesis that all DNA helicases bind one strand of duplex DNA and translocate along it to displace its complementary strand, all archaeal and eukaryal MCM proteins studied to date bind and translocate along ssDNA (53,56,59,60,67,68). Further, mutation studies with *M. thermautotrophicus* MCM has revealed that the N-terminal domain B of the molecule is essential for ssDNA binding [(69), see chapter 2] and the zinc-finger motif present in domain B is involved in ssDNA (70) and dsDNA binding (see chapter 5). The zinc finger motifs were first identified as zinc-binding domains important for DNA-protein interactions (71,72). They contain the conserved Cys and His residues and it is interesting to note that eukaryal MCM4,6,7 and euryarchaeal MCM (*M. thermautotrophicus*) contain a C4 type zinc-finger motif (CXXCX<sub>n</sub>CXXC; X is any amino acid) (70). However, the zinc-finger mutant, although devoid of ssDNA binding, still formed double hexamers in solution (70), suggesting that the zinc-binding domain is not involved in holding the double hexamers as predicted by the crystal structure (65). Further, the ssDNA binding ability of *M. thermautotrophicus* MCM was enhanced in the presence of ATP (70). Recent studies have shown that an arginine residue located in close proximity to the zinc motif is involved in double hexamer formation (73).

In addition to binding ssDNA, the N-terminal portion of *M. thermautotrophicus* MCM also bound dsDNA, with much lower affinity (65,74), similar to many hexameric helicases (75). Recent studies showed the full-length MCM also bound dsDNA (see chapter 5). The dsDNA binding of the protein is mediated by the  $\beta$ -hairpin finger motif present in domain C of the molecule [(65), see also chapter 5]. Further, the DNA binding by MCM showed no substrate specificity as it bound both origin specific and random DNA sequences with the same affinity (see chapter 5).

In eukarya, the abundance of MCM proteins suggest that the proteins could be located at regions of unreplicated DNA and may function away from the replication bubble as a rotary pump, which could unwind the dsDNA at a distant replication fork (76). To achieve this, MCM should be able to translocate duplex DNA. Studies on archaeal and eukaryal MCM complexes suggest that both helicases can translocate dsDNA (77). However, while the eukaryal MCM (*S. pombe*, MCM4,6,7 complex) requires a 3'-ssDNA overhang to load and translocate on the duplex DNA, the archaeal MCM (*M. thermautotrophicus*) translocated dsDNA even in the absence of the 3'-ssDNA overhang (77). Further, during duplex DNA translocation the archaeal MCM unlike its eukaryal counterpart, displaced streptavidin bound to biotinylated duplex DNA (77).

All DNA helicases also bind NTP and exhibit DNA dependent NTPase activity essential for dsDNA unwinding. Both archaeal and eukaryal MCM proteins belong to AAA<sup>+</sup> family of ATPases and contain the characteristic Walker-A and -B motifs responsible for ATP binding and hydrolysis respectively

(55,59,60,67). In addition, similar to the eukaryal MCM, the ATPase activity of *M. thermautotrophicus* MCM is stimulated 13-fold by ssDNA and to a lesser extent by dsDNA (60). However, the ATPase activity of *S. solfataricus* MCM is not stimulated by ssDNA, although its basal ATPase activity (in the absence of DNA) appears to be around 25-fold higher than that of *M. thermautotrophicus* MCM (68).

Unlike the bacterial replicative helicase, DnaB, which unwinds the duplex DNA with a 5'→3' polarity, both archaeal and eukaryal helicases possess a 3'→5' helicase activity (53,56,59,60,67,68). The Walker-A motif present in the C-terminus catalytic domains of archaeal MCM binds ATP and is indispensable for the helicase activity. A mutant MCM protein from *M. thermautotrophicus* and *S. solfataricus* in which the conserved lysine of the Walker-A motif changed to alanine did not possess any DNA helicase activity (68,70). However, the mutant protein bound ssDNA as well as the wild-type enzyme (70). Further, the helicase activity of *S. solfataricus* MCM is stimulated by the SSB protein from the same organism, which also interacts with it (68). Additionally, in *S. solfataricus* MCM, mutation of few positively charged residues, predicted to be on the surface of the protein pointing towards the central channel revealed that the residues are important for the unwinding activity of the enzyme (78).

Another feature of the replicative helicase is its ability to unwind long stretches of duplex DNA with high processivity. Studies on the *M. thermautotrophicus* MCM revealed that the enzyme is highly processive and could displace oligonucleotides of at least 500 bp, which is consistent with it

being a replicative helicase (59,67). Similar studies on the *A. fulgidus* MCM revealed that the full-length enzyme could displace 400 bp, however a mutant form of the enzyme that lacks the N-terminal 112 amino acids was more processive unwinding 1000 bp DNA fragments (53). The reason for the greater processivity of the truncated enzyme could be attributed to its better ATPase activity compared to the full-length enzyme (53). Similar results have also been obtained using the eukaryotic MCM 4,6,7 complex, which forms a double heterohexameric complex on a forked DNA substrate and is processive up to 600 bp of duplex DNA (57). Hence this double heterohexamer is considered as the putative replicative helicase in eukaryotes.

## CHAPTER 2

# BIOCHEMICAL AND STRUCTURAL CHARACTERIZATION OF THE MCM N-TERMINAL DOMAINS

### 2.1 ABSTRACT

The crystal structure of the N-terminal portion of *M. thermautotrophicus* MCM protein revealed a dodecameric structure, with each monomer comprising domains A, B and C. Previous studies identified the N-terminal portion to be involved in multimer formation, single-stranded DNA binding, and may also play a role in regulating the helicase activity. However, the functional roles of each of the three N-terminal domains are not known. This chapter describes a detailed biochemical characterization of the N-terminal region of the MCM helicase. Using biochemical and biophysical analyses it is shown that domain C of the N-terminal portion, located adjacent to the helicase catalytic domains, is required for protein multimerization, and that domain B is the main contact region with ssDNA. It is also shown that while oligomerization is not essential for ssDNA binding and ATPase activity, the presence of domain C is essential for helicase activity.

### 2.2 INTRODUCTION

The MCM complex is thought to function as the replicative helicase of archaea and eukarya (55,56,79,80). In eukarya, MCM is a family of six essential



proteins (MCM2-7) with highly conserved amino acid sequences (55,80). *In vivo* and *in vitro* the proteins form a number of complexes, in addition to the heterohexamer, composed of different combinations of the MCM polypeptides (56,58,81). Biochemical studies with the various complexes have shown that a dimeric complex of the MCM4,6,7 heterotrimer contains 3'→5' DNA helicase, duplex DNA translocation, single-stranded (ss) DNA binding, and DNA-dependent ATPase activity. The interactions of MCM4,6,7 with either MCM2 or MCM3,5 were shown to inhibit helicase activity and therefore were suggested to play regulatory roles (56,58,82).

Most archaeal species examined to date contain a single MCM homologue (1,74). Biochemical studies with the archaeal MCM proteins revealed that the enzymes possess biochemical properties similar to those of the eukaryotic MCM4,6,7 complex (For details see chapter 1).

The *M. thermautotrophicus* MCM, and probably other archaeal MCM enzymes, consist of two main portions. The N-terminal region participates in protein multimerization, ssDNA binding and may also have a regulatory function. The C-terminal portion of the protein contains the helicase catalytic domain(s) (59,65). A high resolution three-dimensional structure of the N-terminal portion revealed a dumbbell-shaped double-hexamer. Each monomer folds into three distinct domains (Fig. 8, chapter 1). Domain A, at the N terminus, is mostly  $\alpha$ -helical. Domain B has three  $\beta$ -strands and contains a zinc-finger motif. Domain C, positioned between domains A and B (Fig. 8, chapter 1), contains five  $\beta$ -

strands that forms an oligonucleotide/oligosaccharide binding (OB) fold. This domain connects the N-terminal portion of the enzyme to the catalytic region.

The domain(s) within the N-terminal region responsible for multimerization has not yet been identified. The three dimensional structure suggested that the zinc-finger motif plays a role in double-hexamer formation (65). However, biochemical analysis of a zinc-finger mutant showed that the mutant protein is impaired in ssDNA binding but not double-hexamer formation (70). In addition, studies with the *A. fulgidus* MCM suggested that domain B, which contains the zinc-finger, is not needed for multimerization (53) but probably for ssDNA binding.

This study was therefore initiated to determine the functional roles of each of the three N-terminal domains in the *M. thermautotrophicus* MCM protein. In this study, it is shown that domain A does not play an essential role in MCM function, as helicase, ATPase, and ssDNA binding activities could be observed in mutant proteins in which the domain had been removed. However, the domain is needed for dsDNA translocation and may play a regulatory role. Domain B is the major contact with ssDNA and domain C is necessary and sufficient for MCM multimerization and is essential for helicase activity.

## 2.3 METHODS

**Generation of MCM mutants:** MCM mutants were generated using a PCR-based approach from plasmid containing the gene encoding the wild-type MCM enzyme (67). The three-dimensional structure of the N-terminal portion of

the molecule (65) served as the guide for the construction of the various mutant proteins. The oligonucleotides used for constructing MCM mutants for two-hybrid analysis are shown in Table 2.

**Table 2. Oligonucleotides used for cloning MCM mutants in pDBLeu and pPC86 vectors**

Name	Sequence (5'→3')
101	CTCC <u>GTCGAC</u> CATGATGAAAACCGTGGATAAGAGCAAGACAC
102	CTCC <u>GACGTCTC</u> CAGACTATCTTAAGGTATCCCC
113	CTCC <u>GTCGAC</u> CATGGTTAAGGAGGCCATAGCCCTCCAGC
114	CTCC <u>GACGTCTC</u> CATCACTCACGGTAGCCGTGTATGGATGGTGC
121	CTCC <u>GTCGAC</u> CCATATGAGCAACGTAATACCCCTCAGGGAGC
122	CCAGGAACTCGGATTCATCGGTCTTCTGACGATGCCATC
123	GAATCCGAGTTCCTGGACACCCAGACACTG
141	CCTTAACTATCCTTGGCCTTATCCCCTGAACCTTATGTTGAGGTCC
142	ATAAGGCCAAGGATAGTTAAGGCC
143	GGAGGGCTATGGCCTCCTTAACTTCGTCTGAAGGAGCCTGAAGG
144	GTTAAGGAGGCCATAGCCCTCCAGC
145	GGAGGGCTATGGCCTCCTTAAACCCGCTGAACCTTATGTTGAGGTCC
146	CTCC <u>GACGTCTC</u> CACCCGCTGAACCTTATGTTGAGG
147	CTCCGTCGACCCATATGGAGATAAGGCCAAGGATAGTTAAGG
149	CTCCGACGTCTCATTCTGCTCTGAAGGAGCCTGAAGG

Sal I restriction sites are underlined; Aat II restriction sites are double underlined.

To generate a mutant protein in which the N- or C- terminal portion was deleted (N-ter, Hel,  $\Delta A$ ,  $\Delta AB$ , AB, AC, BC, A, B and C), a simple PCR reaction was used. For generating the mutant proteins involving deletions of domains within the polypeptide chain ( $\Delta B$ ,  $\Delta C$ ,  $\Delta BC$ ,  $\Delta AC$ ), a two-step PCR strategy was adopted (schematically described in Fig. 9). The first PCR reaction amplified the smaller fragments of the mutants using their respective forward and reverse primers. The reverse primers designed for these PCRs had a 16 – 22 bp sequence complementary to the forward primer of the succeeding domain/fragment at the 5' end. The products of these PCR reactions were then mixed at equimolar concentration and a second PCR reaction was performed

using the forward primer of the upstream fragment (which has a Sall restriction site) and the reverse primer of the downstream fragment (which has an AatII restriction site). This resulted in a product with one or two deleted domains within the wild type gene that was cloned into pDBLeu and pPC86 vectors (Invitrogen) between the Sall and AatII sites, yielding fusion proteins to the GAL4 DNA binding (DB) or activation (AD) domains, respectively. These constructs were used in a two hybrid analysis.

PCR cloning strategy of MCM mutants						PCR product	PCR template
A	C	B	C	Helicase			
101/116 >					< 102/119	wt	wt
101/115 >				< 114/118		N-ter	wt
				113 >	< 102	Hel	wt
	121/115 >				< 102/119	△A	wt
101/116 >		< 122	123 >		< 102/119	△B	wt
101/115 >	< 141	142 >	< 143	144 >	< 102/119	△C	wt
101/115 >	< 145			144 >	< 102/119	△BC	wt
		147 >	< 143	144 >	< 102/119	△AC	wt
	121/115 >				< 102/119	△AB	△B
101 >			< 149			AB	△C
101 >				< 114		AC	△B
	121 >			< 114		BC	wt
101 >	< 146					A	wt
		147 >	< 149			B	wt
	121 >			< 114		C	△B

**Figure 9. PCR cloning strategy and oligonucleotides used to generate MCM constructs.**

The location of the oligonucleotides within the MCM gene is shown. The numbers refer to the oligonucleotides in table 2. The domain lengths are not to scale. Domain A in red, Domain B in blue and Domain C in green.

To generate the *E. coli* expression vectors containing the various MCM mutants, the pDBLeu vectors with the different truncated proteins were used as templates for PCR with forward primers containing NdeI and reverse primers containing XhoI restriction sites (Table 3). The PCR products were cloned into the pET-21a vector (Novagen). All proteins used in the study contain the catalytic domain (except for the N-terminal portion and domain C).

**Table 3.** Oligonucleotides used for cloning MCM mutants in pET-21a vector

Name	Sequence (5'→3')
115	GGCTCC <u>CATATG</u> ATGAAAACCGTGGATAAGAGC
116	GGCTCC <u>CATATG</u> AGCAACGTAATACCCCTCAGGGAGC
118	CCGCTCGAGTCATCA(GTG) <sub>8</sub> CTCACGGTAGCCGTGTATGGATGGTGC
119	CCGCTCGAGTCATCA(GTG) <sub>8</sub> GACTATCTTAAGGTATCCCCTTGC
147	CTCCGTGACCC <u>CATATG</u> GAGATAAGGCCAAGGATAGTTAAGG

Nde I restriction sites are underlined; Xho I restriction sites are double underlined.

**Two hybrid analysis:** For the two hybrid analysis, pDBLeu and pPC86 vectors containing the various MCM mutant genes were generated (see above). Plasmids encoding the AD and DB fusion proteins were co-transformed into yeast MaV203 cells (Invitrogen) according to the manufacturer's protocol. Cells were plated on complete supplement mixture (CSM) plates without Leu and Trp and grown for 2–3 days at 30°C. Colonies were streaked on CSM plates without Leu, Trp, and His and containing 10mM 3-amino-1,2,4-triazole to suppress glycerol phosphate dehydratase, an enzyme involved in histidine biosynthesis. Colonies were also streaked on CSM plates without Leu, Trp, and Ura. Plates were incubated at 30°C for 2–3 days. Growth on these plates indicates that the proteins fused to the AD and DB fusion proteins interact.

**Gel filtration analysis:** One hundred fifty micrograms of the various proteins were applied to a superose-6 (HR10/30, Amersham-Pharmacia) or superdex-200 (HR10/30, Amersham-Pharmacia) gel filtration column equilibrated with buffer containing 20mM Tris-HCl (pH=7.5), 100mM NaCl, 0.5mM EDTA, and 10% Glycerol. Columns were run at 22°C. Fractions (250 µl) were collected and analyzed for the presence of the *M. thermotrophicus* MCM proteins by

separation on a 10% SDS-PAGE (except for domain C which was separated on a 15% SDS-PAGE) and staining with Coomassie brilliant blue (R-250).

**Glycerol gradient sedimentation:** The assays were performed by applying 100 µg of domain C protein to a 5-ml 20-40% glycerol gradient in buffer containing 20mM Tris (pH=7.5), 100mM NaCl and 0.5mM EDTA. After centrifugation at 45,000 rpm (190,000 X g) for 18 hr in a Beckman SW 50.1 rotor at 4°C, fractions (200 µl) were collected from the bottom of the tube. The distribution of the protein was determined by fractionation on a 12% SDS-PAGE and staining with Coomassie brilliant blue (R-250).

**Filter binding assay:** Nitrocellulose filter binding assays were carried out by incubating 0.1, 0.3, 0.9 and 2.7 pmol of proteins (as monomers) at 60°C for 10 min in 20 µl buffer containing 20mM Hepes-NaOH (pH=7.5), 10mM MgCl<sub>2</sub>, 2mM DTT, 100 µg/ml BSA, 1mM ATP, and 50 fmol of 5' <sup>32</sup>P-labeled oligonucleotide N120T; 5'-(GTTT)<sub>10</sub>CGCTGCTCTGCCTCCCGCTGCTCTGCCTCCACTCAGC TCCCTGGCACAGCCTGTCCTGGCACAGGCTGTCCACGTCTGGC-3' (2500 cpm/fmol). After incubation, the mixture was filtered through an alkaline-washed nitrocellulose filter (Millipore, HA 0.45 mm) (83), which was then washed with 20mM Hepes-NaOH (pH=7.5). The radioactivity adsorbed to the filter was measured by liquid scintillation counter.

**Gel mobility shift assay:** Complexes formed between the proteins and ssDNA were detected by a gel mobility shift assay in reaction mixtures (15 µl) containing 25mM Hepes-NaOH (pH=7.5), 50mM sodium acetate, 10mM MgCl<sub>2</sub>, 1mM ATP, 1mM DTT, 100 µg/ml BSA, 20 fmol of <sup>32</sup>P-labeled N120T

oligonucleotide (2500 cpm/fmol), and 1 or 3 pmol of proteins (as monomers). After incubation at 60°C for 10 min, 5 µl of 5X loading buffer (0.1% xylene cyanol, 0.1% bromophenol blue, 50% glycerol) was added to stop the reaction. Aliquots of the reaction mixture were electrophoresed for 1.5 h at 200 V through a 4% polyacrylamide gel containing 6mM magnesium acetate and 5% glycerol in 0.5X TBE.

**ATPase assay:** ATPase activity was measured in reaction mixtures (15 µl) containing 25mM Hepes-NaOH (pH=7.5), 5mM MgCl<sub>2</sub>, 1mM DTT, 100 µg/ml BSA, 1.5 nmol of ATP containing 2.5 µCi of [ $\gamma$ -<sup>32</sup>P]ATP (3000 Ci/mmol, Amersham-Pharmacia), and 0.5 or 1.5 pmol of proteins (as monomers) in the presence or absence of 50ng ss $\phi$ X174 DNA. After incubation at 60°C for 60 min, an aliquot (1 µl) was spotted onto a polyethyleneimine cellulose thin layer plate and ATP and Pi were separated by chromatography in 1 M formic acid + 0.5 M LiCl. The extent of ATP hydrolysis was quantitated by phosphorimager (Molecular Dynamics) analysis.

**DNA helicase assay:** The substrate for helicase assays was made as by annealing a 25-mer oligonucleotide 5'-GCCATCGGGTGCCTGGCCGCAGCGG-3' which was pre-labeled with [ $\gamma$ -<sup>32</sup>P]ATP and T4 polynucleotide kinase, to a 74-mer oligonucleotide 5'-GGGACGCGTCGGCCTGGCACGTCCGGCCGCTGCGGC CAGGCACCCGATGGC(GTTT)<sub>6</sub>-3'.

The substrate for duplex DNA translocation assays was made by annealing a 49-mer oligonucleotide 5'-(TTTG)<sub>6</sub>CCGACGTGCCAGGCCGACG CGTCCC-3' which was pre-labeled with [ $\gamma$ -<sup>32</sup>P]ATP, to two other

oligonucleotides: a 25-mer 5'-CCGACGTGCCAGGCCGACGCGTCCC-3' and a 50-mer 5'-GGGACGCGTCGGCCTGGCACGTCGGCCGCTGCGCCAGGCAC CCGATGGC-3'. The substrates for the helicase and duplex DNA translocation assays were purified as described previously (77).

DNA helicase activity was measured in reaction mixtures (15  $\mu$ l) containing 20mM Tris-HCl (pH=8.5), 10mM MgCl<sub>2</sub>, 2mM DTT, 100  $\mu$ g/ml BSA, 5mM ATP, 10 fmol of <sup>32</sup>P-labeled DNA substrate (3,000 cpm/fmol), and the various MCM mutant proteins as indicated in the figure legend. After incubation at 60°C for 1 hr, reactions were stopped by adding 5  $\mu$ l of 5 X loading buffer (100mM EDTA, 1% SDS, 0.1% xylene cyanol, 0.1% bromophenol blue and 50% glycerol), and aliquots were loaded onto an 8% native polyacrylamide gel in 0.5X TBE (90mM Tris, 90mM boric acid, 1mM EDTA) and electrophoresed for 1.5 hr at 200 V. The helicase activity was visualized and quantitated by phosphorimaging.

**Multiple alignment and surface conservation:** The multiple alignment of the archaeal MCM proteins N-terminal portion was constructed by searching sequence of MCM protein chain A (pdbcode: 1ltl), against the NCBI non-redundant protein database using NCBI Blast PSI-BLAST program. After five rounds of PSI-BLAST, sequence relatives with expectation scores < 0.005 from the archaea domain were pooled and aligned using ClustalW program.

The surface conservation was determined by scoring the relative conservation of a given column compared to all other columns in the multiple alignments. The raw score for each column was calculated using standard



Shannon's information theory entropy formula (84) and converted into normalized Z-score distribution by calculating average entropy and standard deviation for all columns in the alignment. The molecular surfaces and surface potential were built using the Grasp program.

## 2.4 RESULTS

**Domain C of the *M. thermotrophicus* MCM N-terminal region is required for multimer formation.** As a first step in determining the region(s) of the MCM needed for multimerization, two hybrid analyses was performed. Previous studies have shown that the N-terminal portion of the protein participates in hexamer/double hexamer formation (59,65) and a high-resolution three-dimensional structure of the *M. thermotrophicus* MCM N-terminal portion revealed a three domain (A, B, and C) structure (Fig. 8, chapter 1) (65).

Using the three-dimensional structure as a guide, a number of constructs were made to express various deletion mutants of the N-terminal domains. These include the expression of the individual domains and the deletion of single or multiple domains of the N-terminal region. The genes encoding the deletion proteins were generated using a PCR-based approach (see Methods) and cloned into the pDBLeu and pPC86 vectors resulting in fusion proteins with the GAL4 DNA binding (DB) and activation (AD) domains, respectively. The various constructs were analyzed for their ability to interact with the full-length protein, the N-terminal portion of the molecule, and for self-interaction (Fig. 10). As shown in Fig. 10, each protein containing domain C showed interaction with

itself, the full-length, and the N-terminal portion of the MCM protein. Furthermore, domain C is the only intact domain that demonstrated self-interaction.

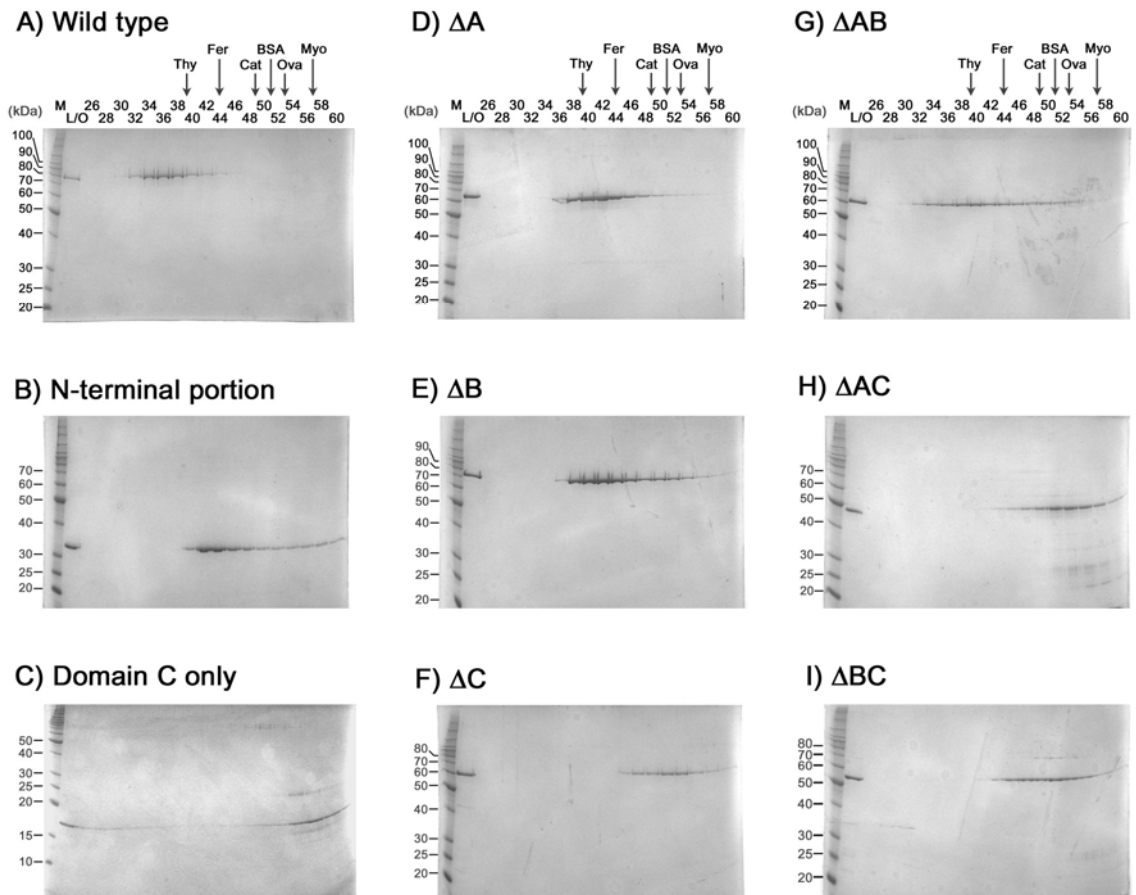
Domain composition of MCM mutants					Protein name	Amino Acids	Molecular mass (kDa)	Interaction with		
A	C	B	C	Helicase				wt	Self	N-ter
■	■	■	■	■	wt	1-666	75.6	+	+	+
■	■	■	■	■	N-ter	1-286	33.2	+	+	+
				■	Hel	287-666	42.4	-	-	-
		■	■	■	△A	92-666	64.9	+	+	+
■	■	■	■	■	△B	1-120, 169-666	70.1	+	+	+
■	■	■	■	■	△C	1-90, 121-168, 287-666	58.6	-	-	-
■	■	■	■	■	△BC	1-90, 287-666	53.2	-	-	-
		■	■	■	△AC	121-168, 287-666	48.0	-	-	-
		■	■	■	△AB	92-120, 287-666	59.5	+	+	+
■	■	■	■	■	AB	1-90, 121-168	16.1	-	-	-
■	■	■	■	■	AC	1-120, 169-286	27.6	+	+	+
■	■	■	■	■	BC	92-286	22.5	+	+	+
■	■	■	■	■	A	1-90	10.6	-	-	-
		■	■	■	B	121-168	5.4	-	-	-
		■	■	■	C	92-120, 169-286	17.1	+	+	+

**Figure 10. Analysis of the MCM multimerization domain.** A summary of the two hybrid analysis of the interactions between the various *M. thermotrophicus* MCM truncated proteins, the full-length and N-terminal portion of the molecule as well as self interactions between the different mutant proteins. '+' indicates interaction and '-' no interaction.

To confirm the observation made with the two hybrid analysis, all deletion proteins containing the catalytic domain, the N-terminal region and domain C alone were expressed and purified from *E. coli* (see appendices) and analyzed on a superose-6 gel-filtration column (Fig.11).

As shown in Figure 11, truncated proteins containing domain C form dodecamers (panels A, B, D, E and G). All the proteins lacking domain C are monomeric (panels F, H and I). Several of the mutant proteins which contain domain C, however, are less stable than the full-length protein as is evident from the presence of two peaks (dodecamers and hexamers) or by the "trailing" of the proteins suggesting that the complex dissociated during fractionation on the

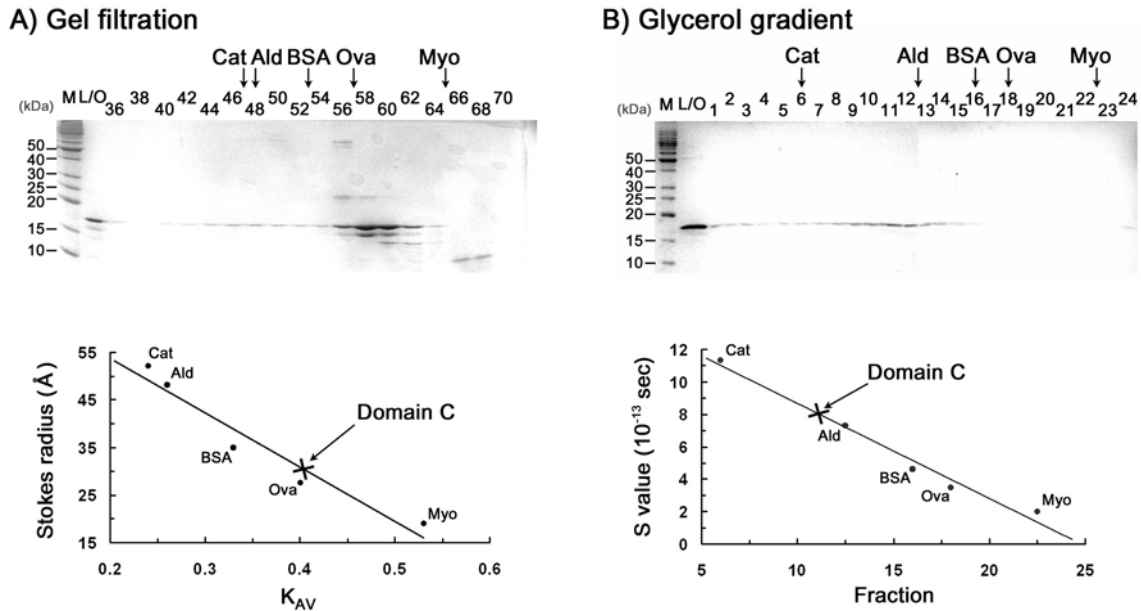
sizing column (e.g. panel G). These observations suggest that either additional domains, besides domain C, are involved in double-hexamer formation or the deletion of the other domains may affect the overall stability of the MCM complex.



**Figure 11. Domain C is needed for MCM oligomeric structure.** One hundred fifty micrograms of purified *M. thermotrophicus* MCM protein (as indicated in the figure) were loaded onto a superose-6 gel filtration column and analyzed as described in Methods. Aliquots (50  $\mu$ l) of each fraction were subjected to 10% (15%, panel C) SDS-PAGE analysis followed by Coomassie blue staining. The peak positions of thyroglobulin (Thy, 669 kDa), ferritin (Fer, 440 kDa), catalase (Cat, 232 kDa), bovine serum albumin (BSA, 67 kDa), ovalbumin (Ova, 43 kDa) and myoglobin (Myo, 17 kDa) are indicated at the top of the figure. M, molecular mass standard; L/O, load on.

The data presented in Figures 10 and 11 suggest that domain C is responsible for multimerization of the *M. thermautotrophicus* MCM. The superose-6 column, however, is not the most suitable one for the analysis of domain C alone, as the domain is only 17.9 kDa in size. Therefore, a superdex-200 column was used to analyze this domain (Fig. 12A). This analysis revealed the presence of a major peak at fractions 58-60. To get a more accurate size determination, a glycerol gradient sedimentation was also performed (Fig. 12B).

In the glycerol gradient a major peak was observed at fractions 11 and 12. Combining the S value and Stokes radius in the mass equation (85) yields a native mass of 103 kDa for the major peak on gel filtration (Fig. 12A). This is in good agreement with the expected size of a hexameric structure (107.8 kDa). However, a minor peak at fractions 46-48 was also observed following the gel filtration (Fig. 12A). This elution position is consistent with a dodecameric structure (216 kDa). In addition, the glycerol gradient sedimentation shows a “trailing” of the proteins from the dodecameric to hexameric structures. These results may suggest that although domain C alone can form double-hexamers, they are not stable and readily dissociate to hexamers during the gel filtration and glycerol gradient analyses.

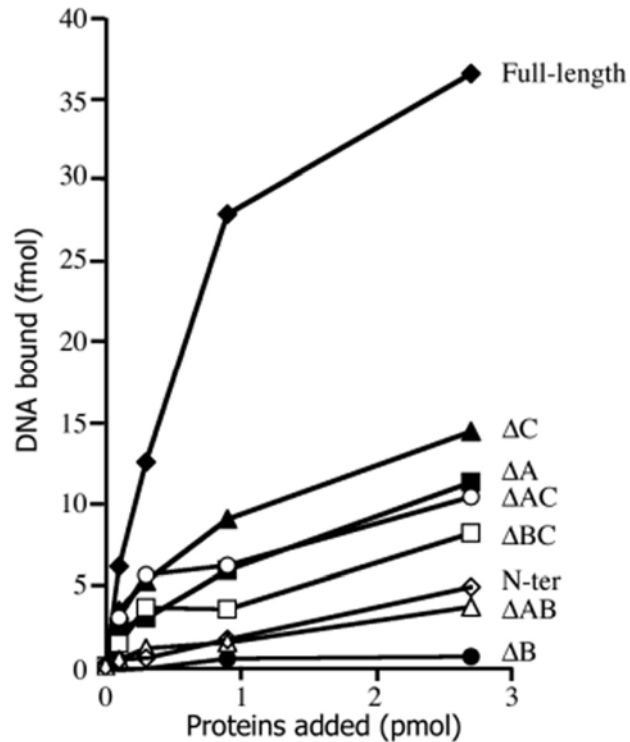


**Figure 12. Domain C of the N-terminal portion of MCM is multimeric.** A) Size exclusion chromatography. One hundred fifty micrograms of purified domain C were loaded onto a superdex-200 gel filtration column and analyzed as described in Methods. Aliquots (50 $\mu$ l) of each fraction were subjected to 15% SDS-PAGE analysis followed by Coomassie blue staining. B) Glycerol gradient sedimentation. One hundred micrograms of purified domain C were fractionated on a 20-40% glycerol gradient as described in Methods. Aliquots (15  $\mu$ l) of each fraction were subjected to 12% SDS-PAGE analysis followed by Coomassie blue staining. In A and B, the peak positions of catalase (Cat, 232 kDa, 52.2 Å, 11.3 S), aldolase (Ald, 158 kDa, 48.1 Å, 7.3 S), bovine serum albumin (BSA, 67 kDa, 34.9 Å, 4.6 S), ovalbumin (Ova, 43 kDa, 27.5 Å, 3.5 S), myoglobin (Myo, 17 kDa, 19.0 Å, 2.0 S) are indicated at the top of the figure. M, molecular mass standard; L/O, load on.

**Domain B of the N-terminal portion of *M. thermautotrophicus* MCM is needed for efficient ssDNA binding.** During DNA unwinding the helicase translocates along one strand of the duplex and displaces the complementary strand. Hence, all DNA helicases, including *M. thermautotrophicus* MCM, have

been shown to interact with ssDNA (75). Previous studies demonstrated that the zinc-finger motif is needed for ssDNA binding by *M. thermautotrophicus* MCM (70) as well as by the eukaryotic helicase (86). The results of these studies, however, suggested that other regions of the proteins are also needed for efficient DNA binding. Thus, the ability of the various mutant proteins to interact with ssDNA was determined using a filter binding assay (Fig. 13). As shown in Figure 13, all truncated proteins show substantially reduced levels of ssDNA binding in comparison to the full-length enzyme. Deletion of domain B has the most severe effect on ssDNA binding, as no ssDNA binding could be detected in a truncated protein in which only domain B ( $\Delta B$ ) was deleted (Fig. 13). This observation suggests that domain B is the major region of the MCM that contacts ssDNA. The zinc-finger motif is located in domain B and thus these results are consistent with the past reports illustrating the need for an intact zinc-finger for efficient DNA binding (70,86). Interestingly, deletion of domain B in conjunction with another domain (either A or C) resulted in a protein with better ssDNA binding in comparison to a protein in which only domain B was removed. It is possible that removing such a large part of the protein exposed charged regions that may show affinity to the negatively charged DNA. Nevertheless, all these mutant proteins show a substantial reduction in ssDNA binding in comparison to the wild-type enzyme. On the other hand, a mutant protein lacking domain C ( $\Delta C$ ) that could not form hexamers (Fig. 11F) retained about 40% of ssDNA binding in comparison to the wild-type enzyme (Fig. 13). These results are consistent with previously reported observations (59,60) which show that

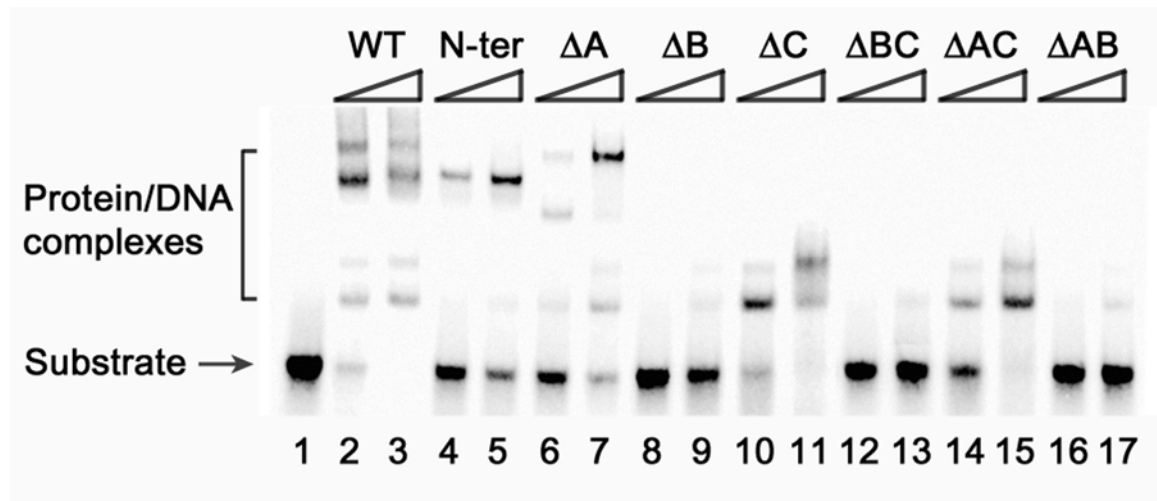
monomeric MCM protein retains detectable ssDNA binding. This indicates that hexamer/dodecamer formation is not needed for ssDNA binding.



**Figure 13. MCM multimerization is not required for ssDNA binding.** Filter binding assays were performed as described in Methods using  $^{32}\text{P}$ -labeled N120T oligonucleotide in the presence of 0.1, 0.3, 0.9 and 2.7 pmol of protein (as monomer). The average result of three experiments is shown.

The ability of the truncated proteins to interact with ssDNA was also determined using a gel mobility shift assay (Fig. 14). The results are similar to those observed with the filter binding assay. Domain B is most critical for ssDNA binding, as proteins lacking the domain ( $\Delta\text{B}$ ,  $\Delta\text{BC}$ , and  $\Delta\text{AB}$ ) bind poorly to ssDNA (lanes 8, 9, 12, 13, 16 and 17) in comparison to the full-length or

truncated proteins containing domain B. Proteins deleted for domain C ( $\Delta C$ ,  $\Delta AC$ ) retained ssDNA binding (lanes 10, 11, 14 and 15). Interestingly, these proteins resulted in slower-migrating bands in comparison to the proteins that include domain C (full-length, N-terminal portion and  $\Delta A$ , lanes 2-7). This is likely due to the different structure of the proteins. The proteins without domain C ( $\Delta C$ ,  $\Delta AC$ ) are monomeric and thus are not expected to result in similar shifts to those created by the hexameric/dodecameric proteins.



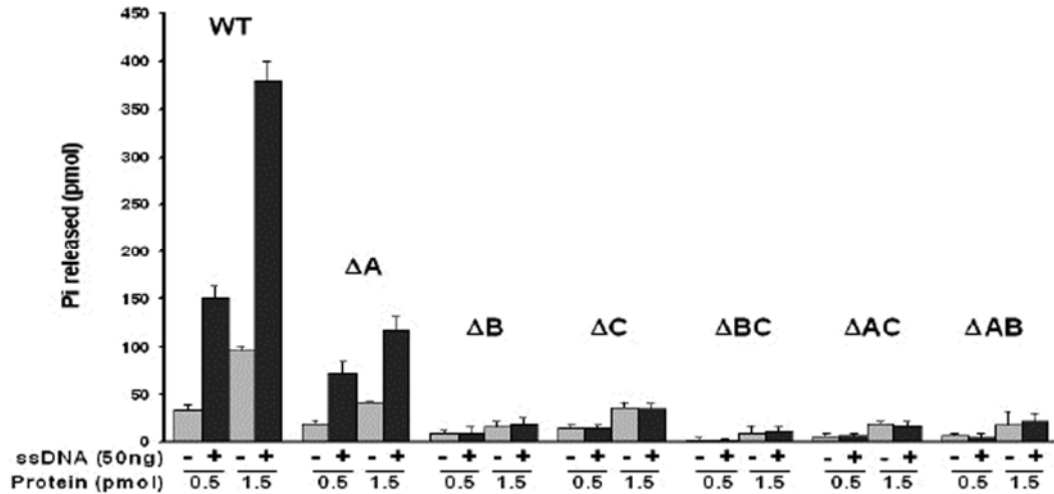
**Figure 14. Domain B of MCM is essential for its ssDNA binding activity.** Gel mobility shift assays were performed as described in Methods using  $^{32}P$ -labeled N120T oligonucleotide and 1 and 3 pmol of proteins (as monomer). Lane 1, substrate only. (Experiment performed by Jae-Ho Shin)

In agreement with previous observations (70) the full-length protein formed a faster and slower migrating band (lanes 2 and 3). Both SV40 Large-T antigen and *S. pombe* MCM helicases were also shown to form a faster-migrating band that was shown to be a result of a hexamer binding to the DNA and a slower-migrating band that was a result of a double-hexamer binding



(57,87). The same explanation has been proposed to the *M. thermautotrophicus* MCM protein in which the faster migrating band contains only one hexamer while the slower one contains a double-hexamer (70).

The *M. thermautotrophicus* MCM ATPase activity was shown to be stimulated in the presence of ssDNA (59,60,67,70). Thus, another indirect approach to demonstrate the domains required for ssDNA binding is by performing ATPase assays in the presence and absence of DNA. Therefore the ATPase activity of the various MCM mutant proteins and stimulation of that activity by ssDNA was examined. As shown in Figure 15, and similar to previously reported observations, the ATPase activity of the full-length enzyme was stimulated about 4-fold in the presence of ssDNA (59,60,67,70). Only the truncated protein lacking domain A ( $\Delta A$ ) shows a similar level of stimulation of activity by ssDNA, though the amount of ATPase activity of the mutant protein is lower than the wild-type enzyme. None of the other truncated proteins showed stimulation of ATPase activity by DNA. Also, while the protein, lacking domain A contained about 30% of the ssDNA binding activity compared to the full-length enzyme, the binding activities of all the other truncated proteins were substantially reduced.

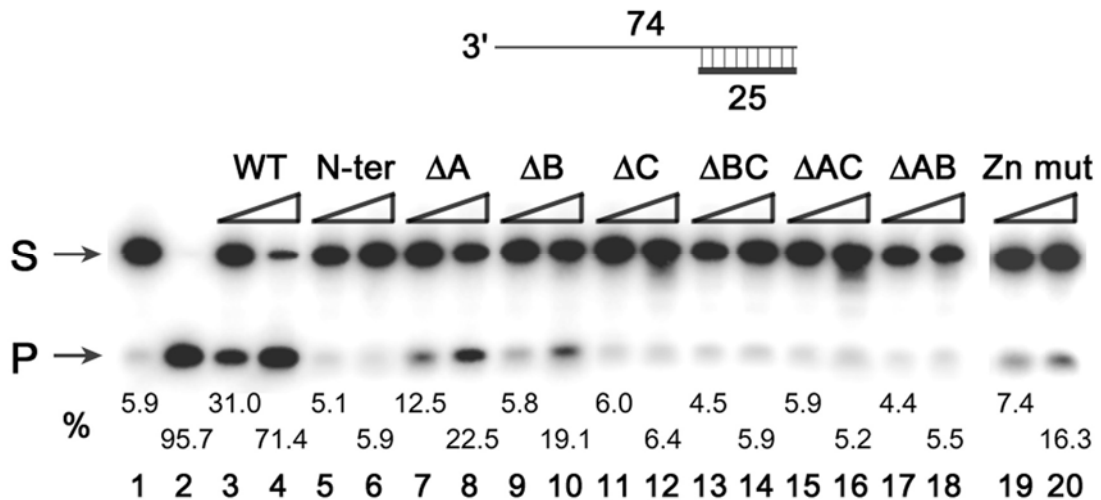


**Figure 15.** The N-terminal MCM is needed for DNA-dependent ATPase activity. The ATPase activity of the wild-type and the various truncated proteins was determined using 0.5 and 1.5 pmol of enzyme (as monomer) in the absence or presence of ssDNA as described in Methods. Error bars represent the standard deviation calculated from three experiments.

***M. thermoautotrophicus* MCM hexamerization is required for helicase activity but only the intact protein can translocate along duplex DNA.** Next, the requirement for the three N-terminal domains (A, B and C) for helicase activity was examined. Based on the data presented above (Figs. 13, 14 and 15) it was anticipated that most of the mutants would have very low, if any, helicase activity. Therefore, the helicase and dsDNA translocation experiments were performed using a higher protein concentration (10-50 fold) than that required to detect helicase activity with the full-length enzyme (46,77).

As shown in Figure 16, only the intact protein has an appreciable helicase activity, as 1 and 3 pmol of enzyme (as monomer) displaced 25% and 66% of the substrate, respectively (lanes 3 and 4). The only truncated proteins with detectable helicase activity are those missing domain A ( $\Delta A$ , lanes 7 and 8) or B

( $\Delta B$ , lane 9 and 10). As these proteins are hexameric/dodecameric (Fig. 10D and E), the results demonstrate the need for ring formation by the proteins for helicase activity. The helicase activity of these mutant proteins, however, is substantially less than the full-length enzyme. The deletion of domain C or deletion of any two domains completely abolished helicase activity. As expected, the N-terminal domain alone showed no detectable helicase activity (lanes 5 and 6).

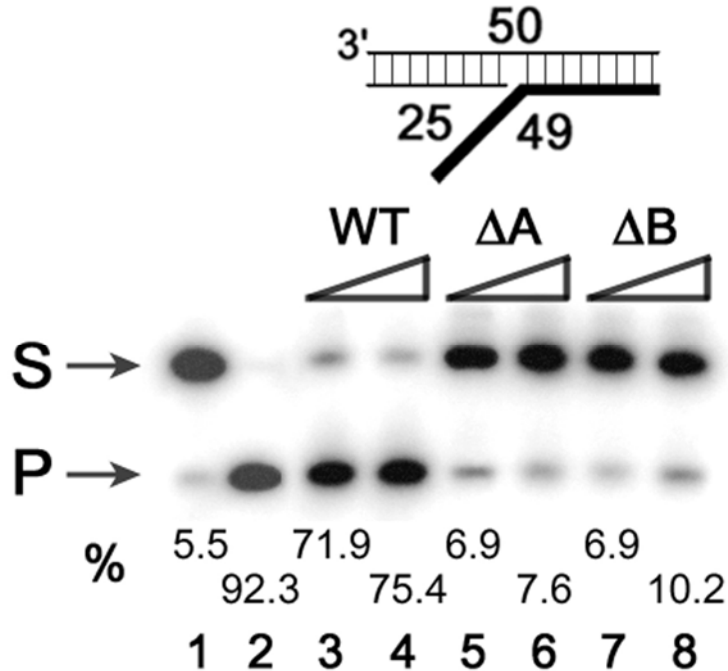


**Figure 16. MCM multimerization is needed for its helicase activity.** DNA helicase assays of the various mutant proteins were performed as described in Methods using 10 fmol of substrate, and 1 and 3 pmol of proteins (as monomer). The percent displacement of the  $^{32}P$ -labeled oligonucleotide from the duplex DNA substrate is indicated as (%). Lane 1, substrate only; lane 2, boiled substrate. (Experiment performed by Jae-Ho Shin)

The reduced helicase activity of the truncated proteins lacking domain B may be attributed to a reduced affinity for ssDNA. It was previously shown that a zinc-finger mutant is impaired in ssDNA binding (70) and domain B deletion results in a similar reduction in ssDNA binding (Figs. 13 and 14). Indeed, the

helicase activity of the protein lacking domain B is similar to that observed with a zinc-finger mutant (compare lanes 9 and 10 to lanes 19 and 20), supporting the idea that the reduction in activity is, in part, due to impaired ability to interact with DNA.

It was recently demonstrated that the *M. thermautotrophicus* and eukaryotic MCM complexes are capable of translocating along duplex DNA (77,88). Thus, the mutant proteins that show some level of helicase activity ( $\Delta A$  and  $\Delta B$ , Fig. 16) were tested for their ability to move along duplex DNA.



**Figure 17. An intact N-terminus is essential for dsDNA translocation by MCM.** Duplex DNA translocation assays were performed as described in Methods using 1 and 3 pmols of proteins (as monomer). The <sup>32</sup>P-labeled oligonucleotide is marked in bold. The percent displacement of the labeled oligonucleotide is indicated as (%). Lane 1, substrate only; lane 2, boiled substrate. (Experiment performed by Jae-Ho Shin)

As shown in Figure 17 neither of the truncated proteins is capable of moving along the duplex (lanes 5-8). At similar protein concentrations, however, the full-length enzyme results in efficient translocation along the duplex (lanes 3 and 4). These observations further suggest that the removal of any domain from the N-terminal portion of MCM impairs its ability to translocate along DNA.

## 2.5 DISCUSSION

The biochemical analysis of the truncated *M. thermautotrophicus* MCM proteins described in this study, together with past observations, revealed that each of the domains in the N-terminal portion of the molecule has a different role in MCM function. Domain A may play a regulatory role; Domain C is necessary and sufficient for protein multimerization; and Domain B is the main contact with ssDNA.

**MCM hexamer formation.** To date, the structures of a number of archaeal MCM complexes have been determined. Although all proteins are homologous, they appear to aggregate differently. While the *M. thermautotrophicus* MCM appears to form double hexamers (59,60,67,70), the enzymes from *S. solfataricus* (68), *A. fulgidus* and *A. pernix* (53) are hexamers. Electron micrograph reconstruction of the *M. thermautotrophicus* MCM revealed ring-shaped hexamers (62) or heptamers (61), but no dodecamers could be observed. The observation that the *M. thermautotrophicus* enzyme forms only single-ring structures at low concentrations (such as those used for EM) may suggest that the double hexamers are formed by non-specific hydrophobic or

ionic interactions involving the two hexamers. These non-specific interactions may be present only in the *M. thermautotrophicus* MCM enzyme, as to date this is the only MCM complex of either archaea or eukarya in which dodecameric structures have been reported. In eukarya, however, the MCM4,6,7 complex was shown to form double hexamers in the presence of a forked DNA substrate (57).

The observation that domain C is involved in hexamer and double hexamer formation is consistent with a previously published observation showing that a truncated *M. thermautotrophicus* MCM protein, in which the first 111 amino acids have been removed, is monomeric (59). Domain C starts at amino acid 92 (65), and thus the N-terminal portion of domain C is deleted in the mutant protein. Taken together, these observations suggest that the N-terminal part of domain C (located N-terminal to domain B in the primary amino acid sequence) may be required for multimerization. The deletion, however, is not likely to affect the overall structure of the molecule, as, similar to the data described here (Figs. 13 and 14), the deleted mutant, although monomeric, retains its ability to bind DNA (59).

MCM homologues from other archaeons may also multimerize via domain C. The region within domain C needed for multimerization, however, may be different than that in *M. thermautotrophicus* MCM. Removal of the N-terminal part of domain C and the entire domain B from the *A. fulgidus* and *A. pernix* MCMs does not affect the hexameric structure of the proteins (53), suggesting that the C-terminal portion of domain C, adjacent to the catalytic domain, is

needed for hexamer formation. Thus, *M. thermautotrophicus* MCM may be different from these enzymes not only by forming dodecamers but also in region involved in multimer formation.

**MCM interactions with ssDNA.** The biochemical analysis of the truncated MCM proteins described here and elsewhere (70) demonstrates that domain B is the main contact with ssDNA, probably via the zinc-finger fold located within the domain. It is possible however that domain C also participates in ssDNA binding, as it contains an OB fold. Interestingly, the three-dimensional structure of domain C is superimposable on the OB2 domain of BRCA2 and domain B is located in a position similar to the Tower domain of BRCA2 molecule (89). DNA binding was shown to occur in the cleft between the OB2 and Tower domains. It is possible that in *M. thermautotrophicus* MCM protein DNA interactions occur in the cleft between domain B and C in addition to the interactions of ssDNA with the zinc-finger located in domain B.

**Does domain A of the N-terminal region of the *M. thermautotrophicus* MCM play a regulatory role?** The data presented in this study revealed that removal of domain A has only limited effects on *M. thermautotrophicus* MCM function, suggesting that the domain may be dispensable for MCM function. Supporting evidence for this notion comes from sequence analysis of archaeal MCM proteins from different species. Several archaeal MCMs do not contain domain A while others have a truncated form of it. The *M. jannaschii*, *Sulfolobus takodeii* and *A. fulgidus* MCM proteins lack domain A. *Halobacterium* sp NRC-1 MCM homologue is missing the first 50 amino acids of domain A. Yet

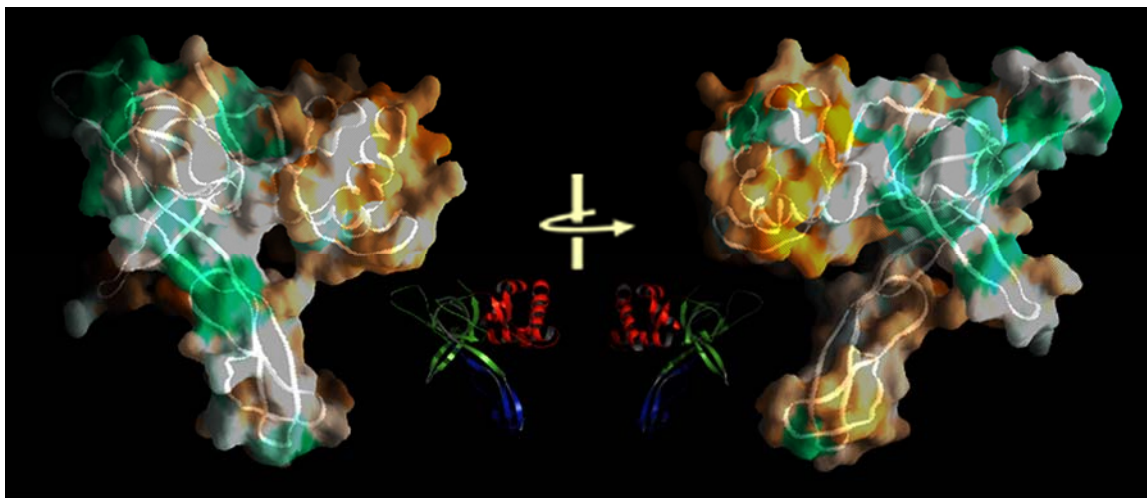
biochemical studies with the *A. fulgidus* protein showed that it possesses DNA helicase activity (53). These observations lead to the proposal that domain A may have a regulatory function(s) *in vivo* (53).

Furthermore, the surface of domain A is less conserved among MCMs of 27 archaeal species studied in comparison to domains B and C (Fig. 18). This observation suggests the possibility that the domain may be needed for protein-protein interactions as a regulatory mechanism and/or during the initiation process. As in different organisms, the structures of proteins interacting with MCM may be different; the MCM will have to adjust its structure and surface residues to facilitate these interactions. One obvious candidate for such interactions is the binding of the archaeal Cdc6 protein to MCM [(45-47), see also chapter 5]. The interactions between the two proteins were shown to regulate the helicase activity of MCM (45,46). In addition, it was demonstrated that the interactions are species specific [(46,47), see also chapter 4] and thus may be mediated by domain A.

The hypothesis that domain A plays a regulatory role may also be supported by the observation that the domain is required for dsDNA translocation (Fig. 17). It was hypothesized that dsDNA translocation by the archaeal, eukaryal and bacterial helicases may play a regulatory role during the process of initiation of DNA replication (76,77,88,90). Supporting evidence for the regulatory role of domain A in regulating MCM helicase activity comes from studies conducted with an eukaryotic Mcm5 mutant (*mcm5-bob1*) (91). In this mutant a conserved Pro is substituted with Lys. Structural and biochemical studies of this



mutant prompted Fletcher et al. (65) to propose a role for the mutation, and thus domain A, in the regulation of MCM activity.



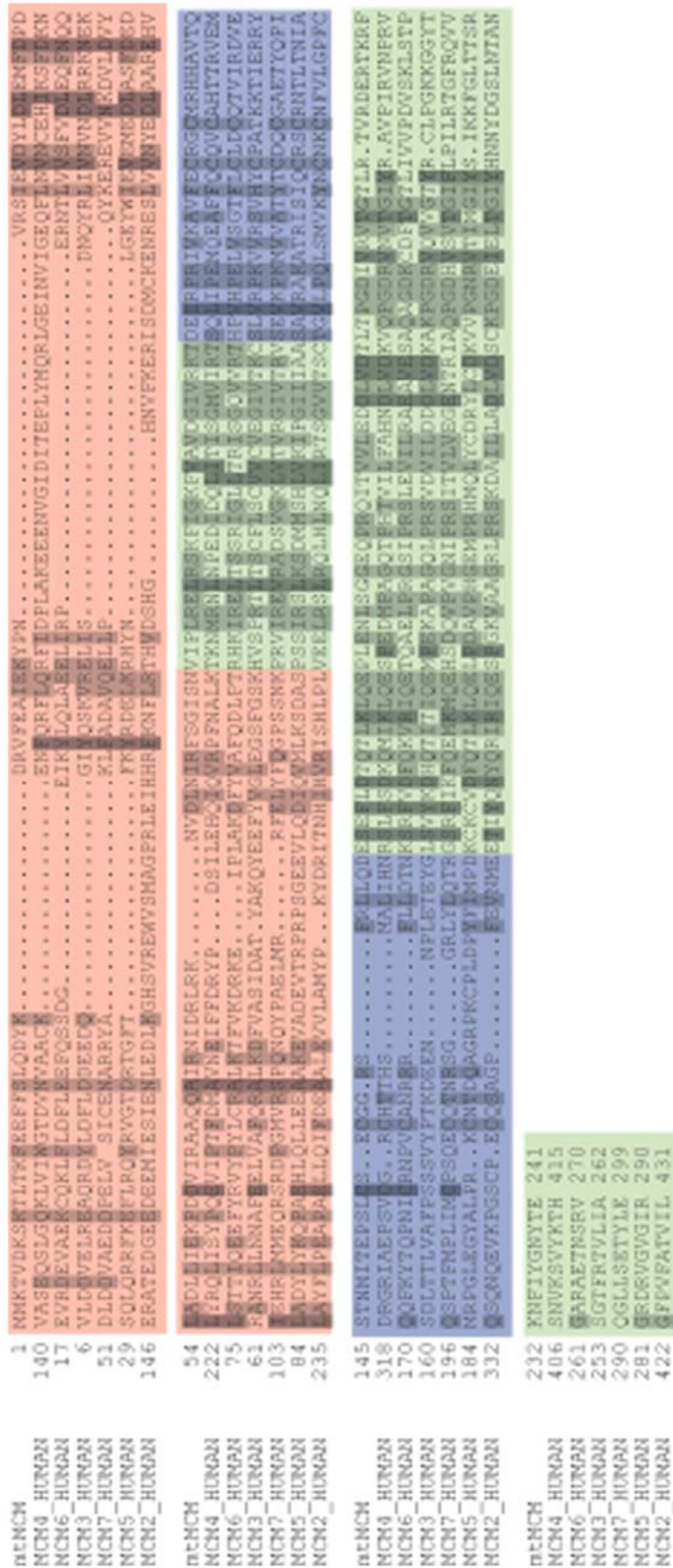
**Figure 18. Sequence conservation patterns among archaeal MCM proteins.** Sequence conservation was calculated as described in Methods and projected onto the surface of the MCM protein (pdbcode: 1ltl, chain A). The figure was generated using GRASP. The color spectrum ranges from orange (least conserved) to white (average conservation) to green (most conserved). (Figure generated by Eugene Melamud).

**Does the N-terminal region of MCM play similar roles in eukarya?** The eukaryotic MCM is a family of six polypeptides with varying sizes but with shared helicase domains within their central part (55,81). Although the proteins were shown to form hexamers in solution, high resolution structures of the individual proteins or the complex have not yet been obtained. Only low resolution electron micrograph structures of the heterohexameric Mcm2-7 (92) and that of MCM4,6,7 (93) complexes are available. These structures revealed a globular shape with a central cavity. The regions and/or domains needed for multimerization, however, are not yet known. It was shown, however, that the N-

terminal 112 residues of murine MCM4 are not required for hexamer formation (86). This region, however, is unique to MCM4 and is not found in other eukaryotic MCM family members or the archaeal homologues (Fig. 19).

The eukaryotic MCM helicase is a hetero-hexamer, which is different from the homo-hexamers of the archaeal enzymes. Therefore, the multimerization domain might be different in eukarya. In order to determine whether similar regions may be involved in multimerization of the archaeal and eukaryal complexes, a sequence alignment (ClustalW) was performed between the N-terminal domains of *M. thermotrophicus* MCM and the human MCM2-7 proteins (Fig. 19).

Similar to the observation in the archaeal proteins (Fig. 18), domain A is the least conserved among the six eukaryotic polypeptides. Domain C, on the other hand, is much more conserved, with about 25% identity among the six eukaryotic family members. Furthermore, domain C does not contain any insertions or deletions within the primary amino acid sequence, suggesting that a structural fold may exist which can not tolerate additional loops and alterations. Hence, these observations may suggest that in eukarya, similar to archaea, the region of MCM located next to the catalytic domain is involved in hexamer formation.



**Figure 19.** Sequence alignment of the N-terminal MCM. Sequence alignments of the N-terminal portion of MCM of the six human MCM (MCM2-7) and the *M. thermautotrophicus* MCM proteins. The colors of the different domains based on the domain structure of MCM, are similar to those shown in Figure 8 (domain A in red, domain B in blue and domain C in green). (Alignment generated by Eugene Melamud).

## CHAPTER 3

### MUTATIONAL ANALYSIS OF THE MCM HELICASE

#### 3.1 ABSTRACT

Chromosomal DNA replication is initiated upon assembly of the replicative helicase at the origin of replication. In eukarya and archaea, the MCM proteins are thought to function as the replicative helicases. The proteins form ring shaped structures that encircle and move along single-stranded DNA and unwind the duplex. The three dimensional structure of the N-terminal portion of MCM from the archaeon *M. thermautotrophicus* revealed a double hexameric structure and suggested that the zinc-finger motif of the molecule participates in hexamer-hexamer interactions. However, biochemical studies, including a variety of mutant proteins lacking the zinc-finger motif, suggested that zinc binding is not required for dodecamer formation. Thus, it was hypothesized that the loop region between  $\beta 7$  and  $\beta 8$  in MCM may participate in double hexamer formation. Here a comprehensive mutational analysis of this region is described. It is shown that all mutant proteins retain their ability to form dodecamers in solution. The possible explanation for this observation is discussed.

#### 3.2 INTRODUCTION

Replicative helicases are molecular motors, which unwinds the duplex DNA to supply single stranded DNA template for the polymerases during

chromosomal DNA replication. In eukarya and archaea, minichromosome maintenance (MCM) proteins are presumed to function as the replicative helicases (55,56,80).

DNA replication in archaea is more similar to those found in eukarya than to those in bacteria [reviewed in (1,2,9,10,94)]. All sequenced archaeal species contains at least one MCM homologue (1,74). Biochemical studies on the MCM proteins from *M. thermautotrophicus*, *S. solfataricus* and *A. fulgidus* revealed that these enzymes possesses 3'→5' helicase activity, single-stranded and double-stranded DNA binding and translocation and DNA-dependent ATPase activity (see chapter 1 for details).

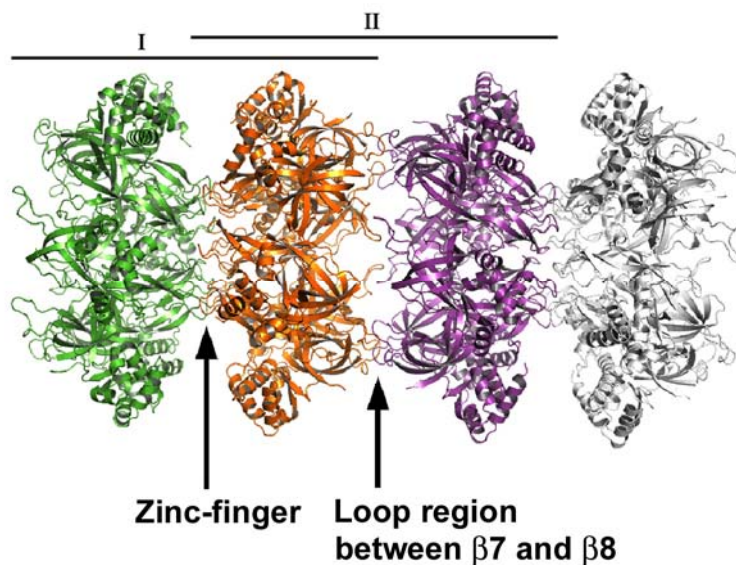
The structure of the archaeal MCM is unclear. The MCM homologues of *S. solfataricus* (68) and *A. fulgidus* (53) form hexamers in solution. The *M. thermautotrophicus* enzyme appears to form dodecamers in solution (59,60,67) and a dodecameric structure was also suggested by the crystal structure of the N-terminal portion of the protein (65) (see also chapter 1). However, electron microscope reconstructions of the full-length *M. thermautotrophicus* enzyme revealed hexameric (62), heptameric (61), filamentous (63) and double hexameric structures (64) (see chapter 1 for details).

Like the eukaryotic MCM, the archaeal enzyme consists of two main parts. The N-terminal region participates in protein multimerization and DNA binding while the C-terminal portion contains the helicase catalytic domain(s) (59,65,69). A high-resolution structure of the N-terminal part of the *M. thermautotrophicus*

molecule revealed a dumbbell-shaped double hexamer (65) (see chapter 1 for details).

The N-terminal part of the *M. thermautotrophicus* MCM crystallized as stacks of hexameric rings (Fig. 20). It was proposed by Fletcher et al. (65) that the active structure of the molecule is dodecameric (double-hexameric). They also proposed that the zinc atoms mediate hexamer-hexamer interactions (I in Fig. 20) as the zinc atoms of each hexamer are in close proximity to the zinc atoms in the other. However, four lines of evidence suggest that this may not be the case. 1) Biochemical studies with a MCM mutant protein devoid of zinc binding retain its ability to form dodecamers in solution (70). 2) A dodecameric structure was also observed with a mutant protein in which domain B, which contains the zinc-finger motif, was deleted [(69), see also chapter 2]. 3) Biochemical studies suggested that domain C is necessary and sufficient for protein multimerization [(69), see also chapter 2]. And 4) all archaeal MCM studied to date show a hexameric and not dodecameric aggregation state (in solution), yet they all contain a zinc-finger motif [e.g. (53,68)].

Thus, it was hypothesized that the loop region between  $\beta 7$  and  $\beta 8$  located within domain C, and is in close proximity between the two hexameric rings in the structure may be required for dodecamer formation (II in Fig. 20) and not zinc binding as previously been proposed. Mutational analysis of this region showed that all mutant proteins form dodecamers in solution.



**Figure 20.** X-ray crystal structure of the N-terminal MCM. The packing of the double hexamers in the crystal show four consecutive hexamers stuck end-to-end against each other. The double hexamer formation proposed by Fletcher et al. (65) is marked as “I” and proposed interactions in this study is marked as “II”.(Figure generated by Eugene Melamud).

### 3.3 METHODS

Generation of MCM mutants: The three dimensional structure of the N-terminal portion of the *M. thermotrophicus* MCM protein was used to identify and construct the mutant proteins. All MCM mutants were generated by a two-step PCR-based strategy (see chapter 2 for details) from a plasmid containing the wild-type MCM gene, which served as template for the PCR reaction. The final PCR products were cloned into the pET-21a vector (Novagen).

The list of mutants used in this study and the location of the mutations are summarized in Table 4.

Mutant	Mutation	Location
Zn	C158S	zinc-finger
L1	E182G, E185G, E191G	loop between $\beta$ 7 and $\beta$ 8
L2	E182G, P183G, L184G, E185G, L187G, E191G, P193G	loop between $\beta$ 7 and $\beta$ 8
$\beta$ 9	I213R, T217S, T219I	$\beta$ 9
ER	E182R, E185I, E191R	loop between $\beta$ 7 and $\beta$ 8
IR	I213R, R215I	$\beta$ 9
TT	T177K, T197E	$\beta$ 7, $\beta$ 8
$\Delta$ 5	deleted amino acids 188-192	loop between $\beta$ 7 and $\beta$ 8
R7	amino acids 188-194 $\rightarrow$ A	loop between $\beta$ 7 and $\beta$ 8
A7	amino acids 188-194 $\rightarrow$ A	loop between $\beta$ 7 and $\beta$ 8

The positions and sequences of oligonucleotides used to generate the mutations are summarized in Tables 5 and 6 respectively.

		MCM mutants
115	← 119	Full-length
115	← 119 207 ← 208	Zn finger
115	← 119 187 ← 188	Loop 1
115	← 119 189 ← 190	Loop 2
115	← 119 191 ← 192	$\beta$ 9
115	← 119 193 ← 194	ER
115	← 119 195 ← 196	IR
115	← 119 197 ← 198	TT
115	← 119 199 ← 200	$\Delta$ 5
115	← 119 201 ← 202	R 7
115	← 119 203 ← 204	A 7

The primers containing the restriction sites NdeI and XhoI are shown in grey. The primers harboring the mutations are shown in black.



<b>Table 6. Sequences of oligonucleotides used for constructing MCM mutants</b>	
<b>Primer</b>	<b>Sequence (5'→3')</b>
115	<u>GGCTCCCATATGATGAAAACCGTCGATAAGAGC</u>
119	<u><u>CCGCTCGAGTCATCAGTGGTGGTGGTGGTGGTGGACTATCTTAAGGTATCCCCT</u></u>
207	GAGCCATCACTCTGCTCAGAGAGTGGTGGGAGATCCTTCAGGC
208	GCCTGAAGGATCTCCCACCACTCTCTGAGCAGAGTGATGGCTC
187	CTGAAACTCCAGGGGCCCTGGGGAACCTTTCCGGTGGGGGACAGCCCCGGC
188	GCCGGGGCTGTCCCCACCGGAAAGGTTCCCCAGGGGCCCTGGAGTTTCAG
189	CAGACACTGAAACTCCAGGGGGGCGGGGGGAACGGTTCGGTGGGGGACAGGGG
190	GTTATCTGCCGGCCCTGTCCCCACCGGAAACGGTTCCCCCGCCCCCTGGAGTT
191	CACACCCGGGGATAGGGTGAGGGTGTCCGGCATCCTCAGGAC
192	GTCCTGAGGATGCCGGACACCCTCACCTATCCCCGGGTGTG
193	CTGAAACTCCAGAGGCCCTGAGGAACCTTTCCGGTGGGAGACAGCCCCGGC
194	GCCGGGGCTGTCTCCCACCGGAAAGGTTCTCAGGGGCCTCTGGAGTTTCAG
195	CTCACACCCGGGGATAGAGTGATTGTGACCGGCACCCTC
196	GAGGGTGCCGGTACAATCACTCTATCCCCGGGTGTGAG
197	GAGTTCCTGGACACCCAGAACTGAACTCCAGGAGCCCCTGGAGAACCTTTCCG GTGGGGAACAGCCCCGGCAGATAGAAGTTGTCCTGGAGGACG
198	CGTCCTCCAGGACAATTCTATCTGCCGGGGCTGTTCCCCACCGGAAAGGTTCTCC AGGGGCTCCTGGAGTTTCAGTTTCTGGGTGTCCAGGAACTC
199	CAGGAGCCCCTGGAGAACCTTCCCCGGCAGATAACAGTTGTCC
200	GGACAACCTGTTATCTGCCGGGGAAGGTTCTCCAGGGGCTCCTG
201	CAGGAGCCCCTGGAGAACCTTCCGGCGGCGGCGGCGGCGGCGGCAGATAACAGTT GTCC
202	GGACAACCTGTTATCTGCCGCCGCCGCCGCCGCCGCCGAAGGTTCTCCAGGGGCTC CTG
203	CAGGAGCCCCTGGAGAACCTTGCCGCCGCCGCCGCCGCCGCCAGATAACAGTT GTCC
204	GGACAACCTGTTATCTGGGCGGCGGCGGCGGCGGCGGCAAGGTTCTCCAGGGGCT CCTG

In primer 115, the restriction site for NdeI is underlined.

In primer 119, the restriction site for XhoI is double underlined.

**Gel Filtration Analysis:** Fifty micrograms of each MCM mutant proteins purified as described in the Appendices were injected to superose-6 (HR10/30, GE Healthcare) or superdex-200 (HR10/30: GE Healthcare) gel filtration column at 22°C pre-equilibrated with buffer containing 20 mM Tris-Cl (pH 7.5), 100 mM NaCl and 10% glycerol. The molecular weight of the wild-type and mutants were estimated by comparing their peak positions to that of the molecular mass

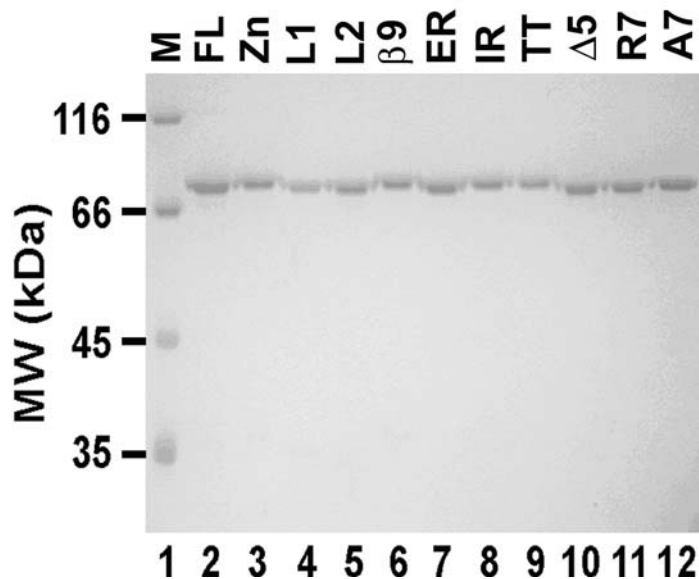
standards (thyroglobulin, ferritin, albumin and ovalbumin) which were also analyzed in the same column under identical conditions.

### 3.4 RESULTS AND DISCUSSION

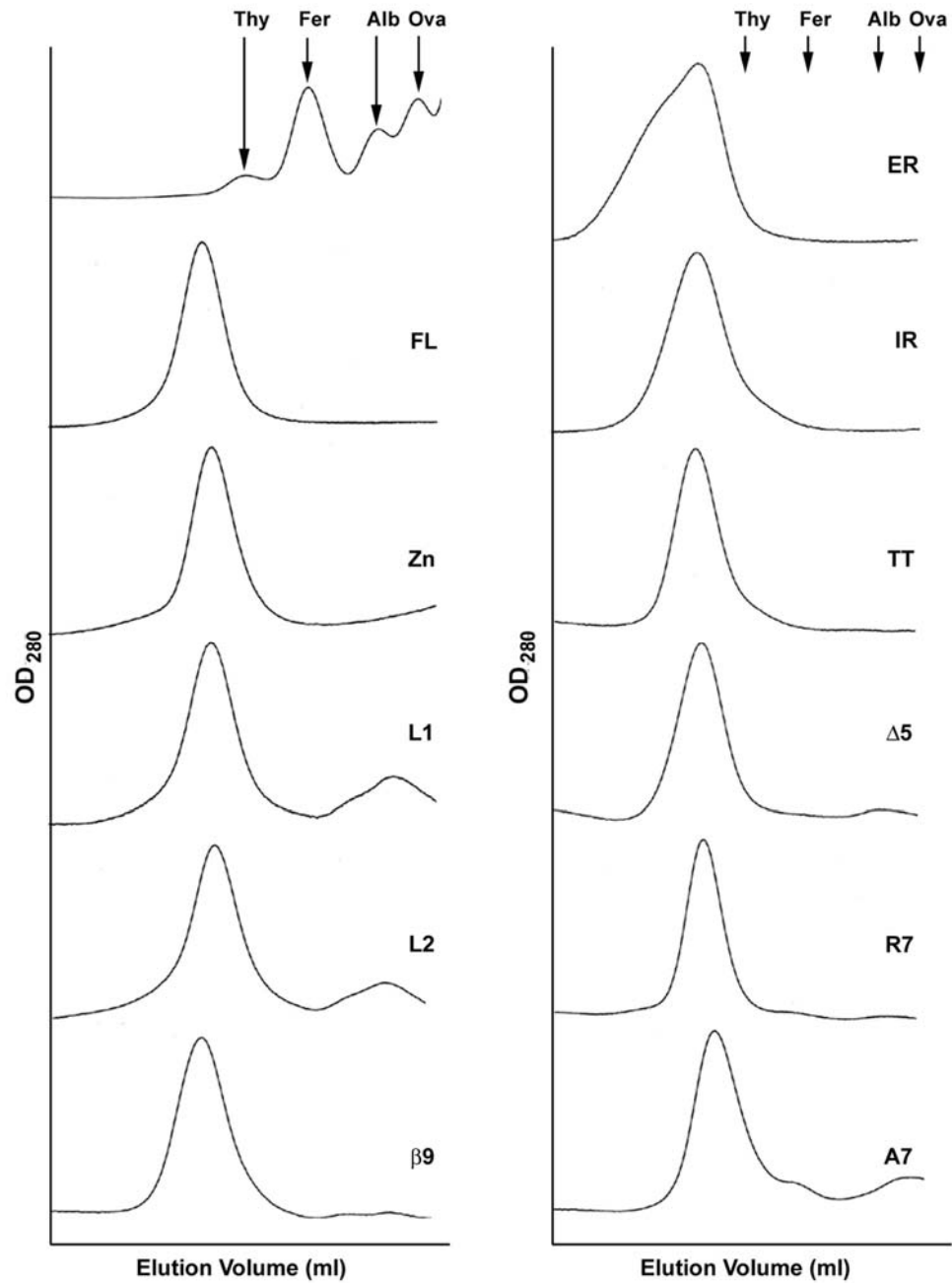
**The loop region between  $\beta 7$  and  $\beta 8$  may be involved in dodecamer formation.** The proposal by Fletcher et al that the zinc is involved in dodecamer formation, is largely due to the fact that the N-terminal MCM crystals were formed as stacks of hexamer rings (65). However, upon a closer look at these crystal stacks, it is obvious that the loop region between  $\beta 7$  and  $\beta 8$  of each monomer of a hexamer also lies in close proximity with the adjacent hexamer molecule (II in Fig. 20). Previous studies have also indicated that domain C of the molecule is essential for double hexamer formation [(69), see also chapter 2]. Since, the loop region between  $\beta 7$  and  $\beta 8$  is present in domain C, this region may be involved in hexamer-hexamer interactions.

Hence several mutations in the loop and adjacent regions of  $\beta 7$  and  $\beta 8$  were generated (Table 4). In addition, there are several residues that are different in *M. thermotrophicus* MCM but conserved in most archaeal hexameric MCM proteins. These residues were also mutated to resemble those in the hexameric complexes (Table 4). All mutant proteins were then overexpressed in *E.coli* as fusion proteins with 6 His tag at the C-terminus and purified using a  $\text{Ni}^{2+}$ -column as described in Appendices section. The subsequent analysis on a SDS-PAGE revealed that all mutants were purified to near homogeneity (Fig. 21). The wild-type and mutant proteins were analyzed for

their solution structure using a superose-6 gel filtration column (Fig. 22). The analysis suggested that all ten mutant proteins remained as double hexamers in solution as their peak positions were located in a similar location as that of the wild-type protein (Fig. 22). The results indicated that the mutated residues either are not sufficient to sever the hexamer-hexamer interactions or that the loop between  $\beta 7$  and  $\beta 8$  does not participate in dodecamer formation. However, it is possible that other residues or a combination of several residues are needed to sever the interactions between the hexamers.



**Figure 21. Purified MCM mutant proteins.** Shown is a Coomassie Blue stained 10% SDS-PAGE, containing 0.5  $\mu\text{g}$  of the various MCM mutants. Lane 1, molecular weight marker; lane 2, full-length MCM; lanes 3-12, various MCM mutant proteins as shown in the top of the gel.



**Figure 22. Size determination of MCM mutants.** Gel-filtration analysis was performed as described in Methods using superose-6 gel-filtration column. The elution profile of the full-length MCM and the mutants are shown. The elution peaks of the molecular weight markers (top, left panel), thyroglobulin (Thy, 670 kDa), ferritin (Fer, 440 kDa), albumin (Alb, 67 kDa) and ovalbumin (Ova, 45 kDa) are indicated by arrows at the top of the figure.

# CHAPTER 4

## INTERACTIONS BETWEEN Cdc6 AND MCM PROTEINS MODULATE THEIR BIOCHEMICAL PROPERTIES

### 4.1 ABSTRACT

The ORC, Cdc6 and the MCM complex play essential roles in eukaryotic initiation of chromosomal DNA replication. Homologues of these proteins may play similar roles in archaeal replication initiation. While the interactions among the eukaryotic initiation proteins are well documented, the protein-protein interactions between the archaeal proteins have not yet been determined. This chapter describes a detailed structural and functional analysis of the interactions between the *M. thermautotrophicus* MCM and the two Cdc6 proteins (Cdc6-1 and -2) identified in the organism. The main contact between Cdc6 and MCM occurs via the N-terminal portion of the MCM protein. It was found that Cdc6-MCM interaction, but not Cdc6-DNA binding, plays the predominant role in regulating MCM helicase activity. In addition, the data show that the interactions with MCM modulate the autophosphorylation of Cdc6-1 and -2. The results also suggest that MCM and DNA may compete for Cdc6-1 protein binding.

### 4.2 INTRODUCTION

Initiation of DNA replication requires the assembly of multiprotein complexes at the origin. In *E. coli*, DnaA protein binds to *oriC* where, aided by

additional proteins, it locally unwinds the origin [reviewed in (4)]. Then ATP-bound DnaC associates with DnaB, the replicative helicase, and recruits it to the origin-DnaA complex to form a prepriming complex. Upon binding to the origin DNA, ATP bound to DnaC is hydrolyzed, releasing DnaC from the complex and activating the helicase (95). *In vitro*, interactions between DnaA and DnaB, and DnaB and DnaC, have been reported but no direct interactions between DnaA and DnaC could be observed (96).

In eukarya, initiation starts with the assembly of a six-subunit origin recognition complex (ORC) at the origin, with ORC serving as a platform on which the pre-replication complex is assembled. The pre-replication complex includes, in addition to ORC, the MCM helicase, Cdc6, Cdt1 and several additional proteins. The release of the helicase and the initiation of DNA synthesis depend on the activity of several proteins including Mcm10, geminin, Cdc45 and cell-cycle dependent kinases [reviewed in (25)]. Using two-hybrid analysis, pull-down experiments, and immunoprecipitation, interactions between many of these proteins were reported including interactions between Cdc6, MCM and ORC [for example see (97)].

The initiation process in archaea is currently unknown. *In silico* analysis suggested that archaeal DNA replication proteins are more similar to those in eukarya than to those found in bacteria. However, the archaeal replication complexes contain fewer subunits than the eukaryotic homologues [reviewed in (1,2)]. Based on primary amino acid sequence analysis it was shown that most archaea contain a single MCM homologue and one or two Cdc6/ORC

homologues (1,2). Some exceptions do exist and up to four MCM and nine Cdc6/ORC homologues have been identified in different archaeons. The eukaryotic Cdc6 protein shows amino acid sequence similarity to subunits of ORC (Orc1, 4, and 5) and it has not yet been determined whether the archaeal Cdc6/ORC homolog functions as ORC, Cdc6, or both. Hereafter, the archaeal Cdc6/ORC proteins will be referred to as Cdc6. Biochemical properties of the archaeal MCM and Cdc6 proteins are reviewed in detail in chapter 1.

The interactions among the archaeal initiation proteins are currently unknown. Therefore, this study determined the interactions between the *M. thermautotrophicus* MCM and the two Cdc6 homologues, Cdc6-1 and -2. The results demonstrate that both Cdc6-1 and -2 proteins interact with the N-terminal portion of MCM. These interactions are required for the regulation of MCM helicase activity by Cdc6 and also modulate the autophosphorylation of Cdc6-1 and -2.

### 4.3 METHODS

**Generation of mutant proteins:** Cdc6 mutants were generated using a PCR-based approach as previously described for the construction of the MCM mutants in chapter 2 from plasmid containing the gene encoding the wild-type *M. thermautotrophicus* Cdc6-1 and -2 proteins (39). The three-dimensional structures of the *P. aerophilum* and *A. pernix* Cdc6 proteins (28,38) served as the guide for the construction of the mutant proteins. The oligonucleotides used to generate the Cdc6 and MCM mutants are shown in Table 7.

<b>Table 7. Oligonucleotides used for cloning proteins in pDBLeu and pPC86 vectors</b>			
<b>Protein</b>	<b>Amino acids</b>	<b>Forward primer (5'→3')<sup>1</sup></b>	<b>Reverse primer (5'→3')<sup>1</sup></b>
<b>Cdc6-1</b>	<b>FL</b>	1-382 CTCCGTCGACCATGAACATTT TTGATG AGATAGGG	CTCCGACGCTTTAAACACCC CAGAGTGAGTCC
	<b>TR</b>	1-286 CTCCGTCGACCATGAACATTT TTGATGAGATAGGG	CTCCGACGCTTTATGTGATC TTGTTGTGTTT
	<b>WH</b>	287-382 CTCCGTCGACCATGGACATAA TACTGACACTACC	CTCCGACGCTTTAAACACCC CAGAGTGAGTCC
	<b>FL mut</b>	1-382 (R <sub>334,335</sub> →A) CTCCGTCGACCATGAACATTT TTGATGAGATAGGG GACTCTGTTTCTACGCAGCA ATATTTGACTTC	GAAGTCAAATATTGCTGCGT AGGAAACAGAGTC CTCCGACGCTTTAAACACCC CAGAGTGAGTCC
	<b>WH mut</b>	287-382 (R <sub>334,335</sub> →A) CTCCGTCGACCATGGACATAA TACTGACACTACC GACTCTGTTTCTACGCGGC GATATTTGACTTC	GAAGTCAAATATCGCCGCG TAGGAAACAGAGTC CTCCGACGCTTTAAACACCC CAGAGTGAGTCC
<b>Cdc6-2</b>	<b>FL</b>	1-379 CTCCGTCGACCATGAAAGGC GATAAGAGGCGC	CTCCGACGCTCAGAGGCC AGGGCACCTTCTG
	<b>TR</b>	1-291 CTCCGTCGACCATGAAAGGC GATAAGAGGCGC	CTCCGACGCTCATATGAGA TTCACGGGTCC
	<b>WH</b>	292-379 CTCCGTCGACCATGCACACG GTGCGAACCTGAAC	CTCCGACGCTCAGAGGCC AGGGCACCTTCTG
	<b>FL mut</b>	1-379 (R <sub>337</sub> →A) CTCCGTCGACCATGAAAGGC GATAAGAGGCGC TACTCATCCTTCAACGCGATC ATTGAGAAACTG	CAGTTTCTCAATGATCGCGT TGAAGGATGAGTA CTCCGACGCTCAGAGGCC AGGGCACCTTCTG
	<b>WH mut</b>	292-379 (R <sub>337</sub> →A) CTCCGTCGACCATGCACACG GTGCGAACCTGAAC TACTCATCCTTCAACGCGATC ATTGAGAAACTG	CAGTTTCTCAATGATCGCGT TGAAGGATGAGTA CTCCGACGCTCAGAGGCC AGGGCACCTTCTG
<b>MCM βFmut</b>	1-666 (R <sub>227</sub> →A, K <sub>229</sub> →A) CTCCGTCGACCATGATGAAAA CCGTGGATAAGAGCAAGACA C CTCAGGACGGTGCGGGACGA GCGCGACAGCGGTTTCAAG	CTTGAAACGCGCTGTCCGG CTCGTCCCGCACCGTCCCTG AG CTCCGACGCTCAGACTATC TTAAGGTATCCCC	
<b>paCdc6</b>	1-389 CTCCGTCGACCATGGCAATTG TTGTTGACGATTTC	CTCCGACGCTCATCTCAAC TCCTCTTAAATTA	
<b>paMCM</b>	1-680 CTCCGTCGACCATGTCTCTTG AAATTGAGTTGG	CTCCGACGCTTATCCACC CGCTTGATATAGC	
<b>Oligonucleotides used for cloning proteins in pET-21a vector</b>			
<b>Protein</b>	<b>Amino acids</b>	<b>Forward primer (5'→3')<sup>1</sup></b>	<b>Reverse primer (5'→3')<sup>1</sup></b>
<b>Cdc6-1</b>	<b>FL</b>	1-382 GGCTCCCATATGCACCACCA CCACCACCACATGAACATTTT TGATGAGATAG	CTCCCTCGAGTTAAACACCC CAGAGTGAGTCC
	<b>FL mut</b>	1-382 (R <sub>334,335</sub> →A) GGCTCCCATATGCACCACCA CCACCACCACATGAACATTTT TGATGAGATAG	CTCCCTCGAGTTATGTGATC TTGTTGTGTTT
	<b>TR</b>	1-286 GGCTCCCATATGCACCACCA CCACCACCACATGAACATTTT TGATGAGATAG	CTCCCTCGAGTTAAACACCC CAGAGTGAGTCC
	<b>WH</b>	287-382 GGCTCCCATATGCACCACCA CCACCACCACATGGACATAAT ACTGACACTAC	CTCCCTCGAGTTAAACACCC CAGAGTGAGTCC
	<b>WH mut</b>	287-382 (R <sub>334,335</sub> →A) GGCTCCCATATGCACCACCA CCACCACCACATGGACATAAT ACTGACACTAC	CTCCCTCGAGTTAAACACCC CAGAGTGAGTCC
<b>Cdc6-2</b>	<b>FL</b>	1-379 GGCTCCCATATGCACCACCA CCACCACCACATGAAAGGCG ATAAGAGGCGC	CTCCCTCGAGTCAGAAGCC AGGGCACCTTCTG
	<b>FL mut</b>	1-379 (R <sub>337</sub> →A) GGCTCCCATATGCACCACCA CCACCACCACATGAAAGGCG ATAAGAGGCGC	CTCCCTCGAGTCATATGAGA TTCACGGGTCC
	<b>TR</b>	1-291 GGCTCCCATATGCACCACCA CCACCACCACATGAAAGGCG ATAAGAGGCGC	CTCCCTCGAGTCAGAAGCC AGGGCACCTTCTG
	<b>WH</b>	292-379 GGCTCCCATATGCACCACCA CCACCACCACATGCACACGG TGCGAACCTGA	CTCCCTCGAGTCATCTCAAC TCCTCTTAAATTA
	<b>WH mut</b>	292-379 (R <sub>337</sub> →A) GGCTCCCATATGCACCACCA CCACCACCACATGCACACGG TGCGAACCTGA	CTCCCTCGAGTCATCTCAAC TCCTCTTAAATTA
<b>paCdc6</b>	1-389 GGCTCCCATATGCACCACCA CCACCACCACATGGCAATTGT TGTTGACGATTTC	CTCCCTCGAGTCATCTCAAC TCCTCTTAAATTA	
<b>Oligonucleotides used for cloning proteins with kinase recognition motif in pET-21a vector</b>			
<b>Protein</b>	<b>Amino acids</b>	<b>Forward primer (5'→3')<sup>1</sup></b>	<b>Reverse primer (5'→3')<sup>1</sup></b>
<b>MCM</b>	<b>WT<sup>2</sup></b>	1-666 GGCTCCCATATGCTTGAAGA GCTTCTGTTATGATGAAAACC GTGGATAAGAGC	CCGCTCGAGTCATCAGTGG TGTTGGTGGTGGTGGACTA TCTTAAGGTATCCCTT GTC
	<b>Δ B<sup>2</sup></b>	1-120, 168- 666 GGCTCCCATATGCTTGAAGA GCTTCTGTTATGATGAAAACC GTGGATAAGAGC	CCGCTCGAGTCATCAGTGG TGTTGGTGGTGGTGGACTA TCTTAAGGTATCCCTT GTC
	<b>Δ C<sup>2</sup></b>	1-91,121- 168, 287-666 GGCTCCCATATGCTTGAAGA GCTTCTGTTATGATGAAAACC GTGGATAAGAGC	CCGCTCGAGTCATCAGTGG TGTTGGTGGTGGTGGACTA TCTTAAGGTATCCCTT GTC
	<b>Δ A</b>	92-666 GGCTCCCATATGCTTGAAGA GCTTCTGTTATGATGAAAACC GTGGATAAGAGC	CCGCTCGAGTCATCAGTGG TGTTGGTGGTGGTGGACTA TCTTAAGGTATCCCTT GTC
	<b>N-ter</b>	1-286 GGCTCCCATATGCTTGAAGA GCTTCTGTTATGATGAAAACC GTGGATAAGAGC	CCGCTCGAGTCATCAGTGG TGTTGGTGGTGGTGGACTA TCTTAAGGTATCCCTT GTC

1. Restriction sites are underlined.

2. For PCR of WT, Δ B and Δ C proteins with kinase recognition motif, the respective proteins cloned into pDBLeu (without the kinase motif) were used as template.



For the two-hybrid analysis the various mutants were cloned into the pDBLeu and pPC86 vectors between the Sall and AatII sites, yielding fusion proteins to the GAL4 DNA binding (DB) or activation (AD) domains, respectively. The mutant proteins and the terminology used throughout the text are shown in Table 8.

<b>Table 8. Terminology of Cdc6 and MCM proteins used in this study with amino acids and mutant descriptions</b>			
<b>Cdc6-1 and -2</b>	<b>Terminology</b>	<b>Amino acids (with mutations)</b>	
		<b>Cdc6-1</b>	<b>Cdc6-2</b>
Full-length	FL	1-382	1-379
Truncated	TR	1-286	1-291
Winged helix	WH	287-382	292-379
Full-length mutant	FLmut	1-382 (R <sub>334,335</sub> →A)	1-379 (R <sub>337</sub> →A)
Winged helix mutant	WHmut	287-382 (R <sub>334,335</sub> →A)	292-379 (R <sub>337</sub> →A)
Walker A mutant	WAmut	1-382 (K <sub>65</sub> →E)	1-379 (K <sub>71</sub> →E)
Walker B mutant	WBmut	1-382 (D <sub>149</sub> →N)	1-379 (D <sub>148</sub> →N)
<i>P. aerophilum</i> Cdc6	paCdc6	1-389	
<b>MCM</b>		<b>Amino acids (with mutations)</b>	
<b>Terminology</b>			
Full-length	FL MCM	1-666	
N-terminus	N-ter	1-286	
C-terminus	C-ter	287-666	
Deletion of domain A	Δ A	92-666	
Deletion of domain B	Δ B	1-120, 168-666	
Deletion of domain C	Δ C	1-91,121-168, 287-666	
Deletion of domain AB	Δ AB	92-120, 169-666	
Deletion of domain AC	Δ AC	121-168, 287-666	
Deletion of domain BC	Δ BC	1-91, 287-666	
Domain A	A	1-91	
Domain B	B	121-168	
Domain C	C	92-120, 169-286	
Domain AB	AB	1-91,121-168	
Domain AC	AC	1-120, 168-286	
Domain BC	BC	92-286	
Walker A mutant	WAmut	1-666 (K <sub>325</sub> →A)	
Zinc-finger mutant	ZFmut	1-666 (C <sub>158</sub> →S)	
β finger mutant	βFmut	1-666 (R <sub>227</sub> →A, K <sub>229</sub> →A)	
<i>P. aerophilum</i> MCM	paMCM	1-680	

For protein expression in *E. coli*, the mutant proteins containing six His residues upstream of the stop codon of MCM and at the N-terminus of Cdc6 were cloned into the pET-21a vector between the NdeI and XhoI sites. The Cdc6 proteins with MBP-tags were generated by cloning the genes into an *E. coli* expression vector containing His<sub>6</sub>-MBP recognition site at the N-terminus (98). Following expression in *E. coli*, the proteins were purified as described in the appendix section. The various MCM mutants containing a cAMP-dependent protein kinase recognition motif were generated by PCR using the mutant genes as template (99).

**Two-hybrid analysis:** For the two-hybrid analysis, pDBLeu and pPC86 vectors containing the various combinations of MCM and Cdc6 mutant genes were generated (see above). The detailed protocol for the two-hybrid analysis is summarized in chapter 2. After yeast transformation, the growth of yeast cells on the selection plates were monitored and scored every 24h for 4 days. Growth indicates that the proteins fused to the AD and DB vectors interact. Interactions were also analyzed using CSM plates without Leu, Trp and Ura as well as with a  $\beta$ -galactosidase assay for LacZ expression.

**Far Western dot-blot assay:** Protein labeling for Far Western analysis was performed as previously described (99) using 1.8nmol of protein (as monomers) in a 100 $\mu$ l reaction containing 20mM Tris-HCl (pH=7.5), 15mM magnesium acetate, 2mM DTT, 100mM NaCl, 15 $\mu$ l (50pmol) [ $\gamma$ -<sup>32</sup>P]ATP (3000 Ci/mmol, GE Bioscience) and 5 U of protein kinase A (Sigma) at 37°C for 60

min. The labeled proteins were purified from the unincorporated nucleotides using a sephadex G-50 gel filtration column equilibrated with reaction buffer.

Far Western dot-blot assays were carried out using Minifold I (Schleier and Schuell) apparatus by blotting 0.05, 0.15, 0.25, 0.5, 1.5 and 2.5nmol or 0.05, 0.15, 0.5 and 1.5nmol of protein (as monomers) onto nitrocellulose membrane (Schleicher and Schuell) pre-washed with 0.5X SSC (75mM NaCl, 7.5mM Sodium citrate). Following blotting, the wells were washed three times with 0.5X SSC. The nitrocellulose membrane was then blocked by incubating in 1X TBST buffer [20mM Tris-HCl (pH=7.6), 137mM NaCl, 3% Tween-20] containing 4% (w/v) non-fat dry milk for 18h at 4°C. The blocked nitrocellulose membrane was washed three times (20 min each) with 50ml of HYB buffer [20mM Hepes-NaOH (pH=7.5), 1mM MgCl<sub>2</sub>, 1mM DTT, 100mM NaCl and 10% glycerol] at 22°C. Hybridization was carried out with 300pmol of <sup>32</sup>P-labeled proteins in 20ml of HYB buffer for 3h at 22°C. The membrane was washed three times (20 min each) with 50ml HYB buffer at 22°C, air-dried and analyzed by phosphorimaging (Molecular Dynamics). The membrane was then cut and the radioactivity adsorbed by each dot in the membrane was measured by liquid scintillation counter. The specific activities of the labeled proteins used were: full-length MCM, 4.5 cpm/fmol; ΔA MCM, 3.5 cpm/fmol; ΔB MCM, 1.3 cpm/fmol; ΔC MCM, 1.8 cpm/fmol; N-ter, 30.6 cpm/fmol; PCNA, 80 cpm/fmol.

**Protein pull-down assay:** The pull-down assays were carried out by binding 2μg of MBP-tagged Cdc6-1 or -2 proteins to 30μl amylose resin (New England Biolabs) washed and equilibrated with 100μl binding buffer [20mM Tris-

HCl (pH=7.5)] at 22°C for 10 min. The reaction was continued at 22°C for 10 min after adding 6µg of His-tagged MCM or PCNA protein. The beads were then washed two times with 500µl wash buffer [20mM Tris-HCl (pH=7.5), 100mM NaCl] and centrifuged at 4,500 rpm for 30 sec. Proteins bound to the beads were eluted with 40µl elution buffer [20mM Tris-HCl (pH=7.5), 100mM NaCl and 50mM maltose]. The samples were then analyzed after adding 10µl of 5X SDS loading buffer [250mM Tris-HCl (pH=6.8), 500mM DTT, 10% SDS, 0.5% Bromophenol blue and 50% glycerol], boiled and separated on 10% SDS-PAGE and visualized by Coomassie blue staining.

**DNA helicase assay:** The substrate for helicase assays was made by annealing a 25-mer oligonucleotide 5'- CCGACGTGCCAGGCCGACGCGTCCC - 3' which was pre-labeled with [ $\gamma$ -<sup>32</sup>P]ATP and T4 polynucleotide kinase, to a 50-mer oligonucleotide 5'- GGGACGCGTCGGCCTGGCACGTCGGCCGCTGCGG CCAGGCACCCGATGGC -3'. The substrate for duplex DNA translocation assays was made by annealing a 61-mer oligonucleotide 5'-(TTTG)<sub>9</sub>CCGACGT GCCAGGCCGACGCGTCCC-3' which was pre-labeled with [ $\gamma$ -<sup>32</sup>P]ATP, to two other oligonucleotides: a 25-mer 5'-CCGACGTGCCAGGCCGACGCGTCCC-3' and a 50-mer 5'-GGGACGCGTCGGCCTGGCACGTCGGCCGCTGCGGCCAG GCACCCGATGGC-3'.

DNA helicase activity was measured as in Chapter 2. To determine the effect of the various Cdc6-1 and -2 mutant proteins on the MCM helicase activity, reaction mixtures containing 0.3 pmols of MCM protein (as monomer) and 0.3, 1.2 and 4.8pmol of the various Cdc6 proteins were added to the helicase assay.

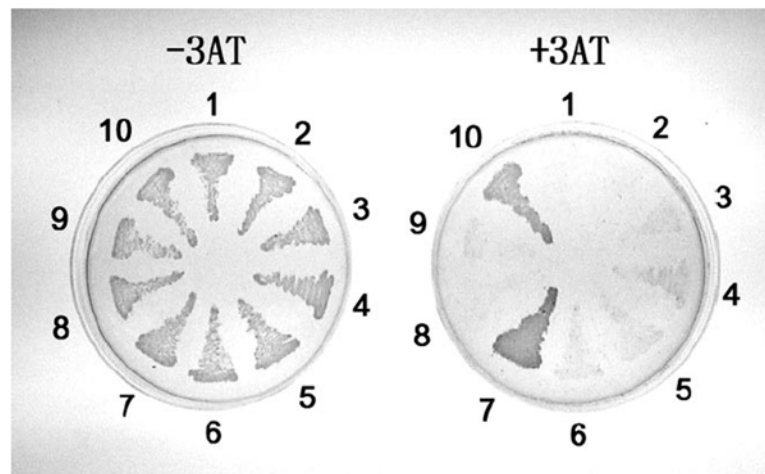
**Filter binding assay:** Filter binding assays were performed as described in chapter 2, with either 45-mer ssDNA (MR163; 5'-CATATGTACATGGGTACAT ATGTACATGGGTACATATGTACAT-3') or with dsDNA generated by annealing the MR163 oligonucleotide to its complementary sequence MR164 (5'-ATGTAC ATATGTACCCATGTACATATGTACCCATGTACATATGTA-3'). The assays were carried out as described in chapter 2 using 0.1, 0.3, 0.9 and 2.7 pmol of protein (as monomers).

**Protein autophosphorylation:** Protein autophosphorylation assays were performed as previously described (39) with 10pmol of Cdc6 proteins in a reaction mixture (15µl) containing 3.3pmol of [ $\gamma$ -<sup>32</sup>P]ATP, 25mM Hepes-NaOH (pH=7.5), 5mM MgCl<sub>2</sub> and 1mM DTT in the presence or absence of 20pmol of MCM (K<sub>325</sub>→A) protein. Following incubation for 20 min at 65°C the reaction was stopped by adding 5µl of 5X SDS loading buffer [250mM Tris-HCl (pH=6.8), 500mM DTT, 10% SDS, 0.5% Bromophenol blue and 50% glycerol], boiled. The proteins were then separated on 10% SDS-PAGE followed by Coomassie blue staining and autoradiography.

## 4.4 RESULTS

**MCM interacts with Cdc6-1 and -2.** *In vitro* studies demonstrated that the *M. thermautotrophicus* Cdc6-1 and -2 proteins inhibit MCM helicase activity and it was therefore suggested that the proteins interact (46). Two-hybrid analyses performed with *M. thermautotrophicus* Cdc6-1 and MCM revealed an

interaction between the two proteins (Fig. 23). As shown in Fig. 22 (sector 7) the *M. thermautotrophicus* Cdc6 interacts with its MCM. The specificity of the interaction was determined in sector 9. In contrast to *M. thermautotrophicus* Cdc6, the Cdc6 homologue from *P. aerophilum* (paCdc6) did not bind to *M. thermautotrophicus* MCM. In addition, the possibility of an indirect interaction via a bridging protein in the two-hybrid screen is unlikely, given the differences between eukaryotic and archaeal proteins. Furthermore, an intact WH domain is needed for the interactions between the proteins, as the truncated form of Cdc6-1 failed to interact with MCM (sector 8). It was shown previously (39) that this domain was important in interactions with dsDNA.



**Figure 23. MCM interacts with Cdc6-1.** Two-hybrid analyses of the interactions between *M. thermautotrophicus* MCM and Cdc6-1 were performed as described in Methods. CSM plate minus Leu, Trp, and His (left panel); CSM plate minus Leu, Trp, and His containing 10 mM 3-amino-1,2,4-triazole(3AT) (right panel). 1, pDBLeu and pPC86; 2, pDBLeu-Cdc6-1 and pPC86; 3, pDBLeu-Cdc6-1TR and pPC86; 4, pDBLeu-paCdc6 and pPC86; 5, pDBLeu- MCM and pPC86; 6, pDBLeu and pPC86-MCM; 7, pDBLeu-Cdc6-1 and pPC86-MCM; 8, pDBLeu-Cdc6-1TR and pPC86-MCM; 9, pDBLeu-paCdc6 and pPC86-MCM; and 10, pDBLeu-MCM and pPC86-MCM.

Next, a detailed interaction study between the MCM and Cdc6-1 and -2 proteins to determine the interaction domains were performed using two-hybrid analysis. The three-dimensional structure of the N-terminal portion of *M. thermotrophicus* MCM (65) and the structures of *P. aerophilum* and *A. pernix* Cdc6 proteins (28,38) were used as a guide for constructing mutant and truncated proteins.

Genes encoding various MCM and Cdc6 derived proteins were generated using a PCR-based approach (see Methods) and cloned into the pDBLeu and pPC86 vectors (Invitrogen) resulting in fusion proteins containing the GAL4 DNA binding (DB) and activation (AD) domains, respectively. Different combinations of Cdc6 and MCM derivatives were analyzed for their ability to interact. Using this assay Cdc6-1 showed clear and strong interactions with MCM (Fig. 23 and 24A). No interactions, however, could be detected between Cdc6-2 and MCM or Cdc6-1 and Cdc6-2 in the two-hybrid assay.

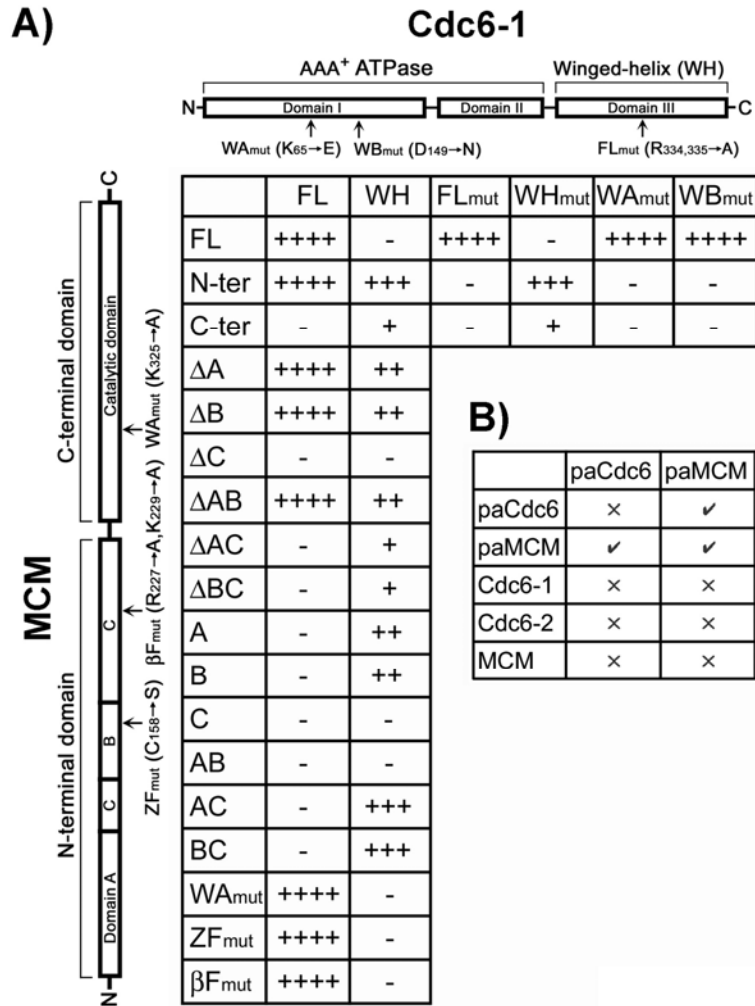
As shown in Figure 24A, the C-terminal catalytic domain of MCM is not required for the interaction with Cdc6-1 as the N-terminal part of the MCM molecule is sufficient for efficient binding. The data also suggest that domain C of the N-terminal region of MCM is required for Cdc6-1 binding, as proteins lacking either domain A ( $\Delta A$ ) or domain B ( $\Delta B$ ) or both domains ( $\Delta AB$ ) are capable of binding to Cdc6-1. When domain C was removed from MCM ( $\Delta C$ ) no interactions with Cdc6 could be detected.

Furthermore, the data presented in Figure 24A suggest that the WH domain of Cdc6-1 is the main contact region to MCM. A truncated protein

containing only the AAA<sup>+</sup> catalytic domains (domains I and II) did not interact with any MCM derivative. In addition, proteins containing mutations in the Walker-A and -B motifs of Cdc6 retain their ability to interact with MCM, illustrating that an active ATPase domain is not required for MCM binding. It was also found that the WH domain alone is capable of interaction with the N-terminal portion of MCM. Interestingly, the WH domain did not interact with the full-length MCM protein. The presence of the catalytic domains of MCM may prevent access of the Cdc6-1 WH domain to the binding site on MCM when the AAA<sup>+</sup> domains of Cdc6 are missing. However, in the full-length Cdc6-1 protein, the WH domain is far from the AD domain of the fused protein and thus may have a better access to the binding site on domain C of MCM.

The two-hybrid analysis also suggests that ATP and DNA binding by MCM and/or Cdc6-1 are not required for interaction between the proteins. The protein with a mutation in the Walker-A motif (K<sub>325</sub>→A) of MCM, which is devoid of helicase activity and ATP binding (59,67), also strongly interacts with Cdc6. The β-finger and zinc-finger mutants of MCM bind Cdc6-1 as well as the intact protein (Fig. 24A). The β-finger mutant cannot bind dsDNA or ssDNA ((65), see also chapter 5), while the zinc-finger mutant is impaired in ssDNA (70) and dsDNA (see chapter 5) binding. In addition, the Cdc6-1 protein with mutation in the WH domain shown to be devoid of DNA binding (41), retained MCM interaction (Fig.24A).





**Figure 24. Cdc6-1 protein interacts with MCM protein in a two-hybrid analysis.** A summary of the two-hybrid analysis of the interactions between the various *M. thermotrophicus* (A) and *P. aerophilum* (B) MCM- and Cdc6-derived proteins, performed as described in Methods. In (A), cell growth observed after 24 h, +++++; 48 h, +++; 72 h, ++; 96 h, + and no growth, - are shown. In (B), 'tick mark' indicates cell growth after 2 days and 'cross' indicates that no growth was observed.

To confirm the observation made with the two-hybrid analysis, and to determine whether Cdc6-2 is capable of binding MCM, several of the MCM and Cdc6-1 and -2 derived proteins were expressed and purified from *E. coli* (see

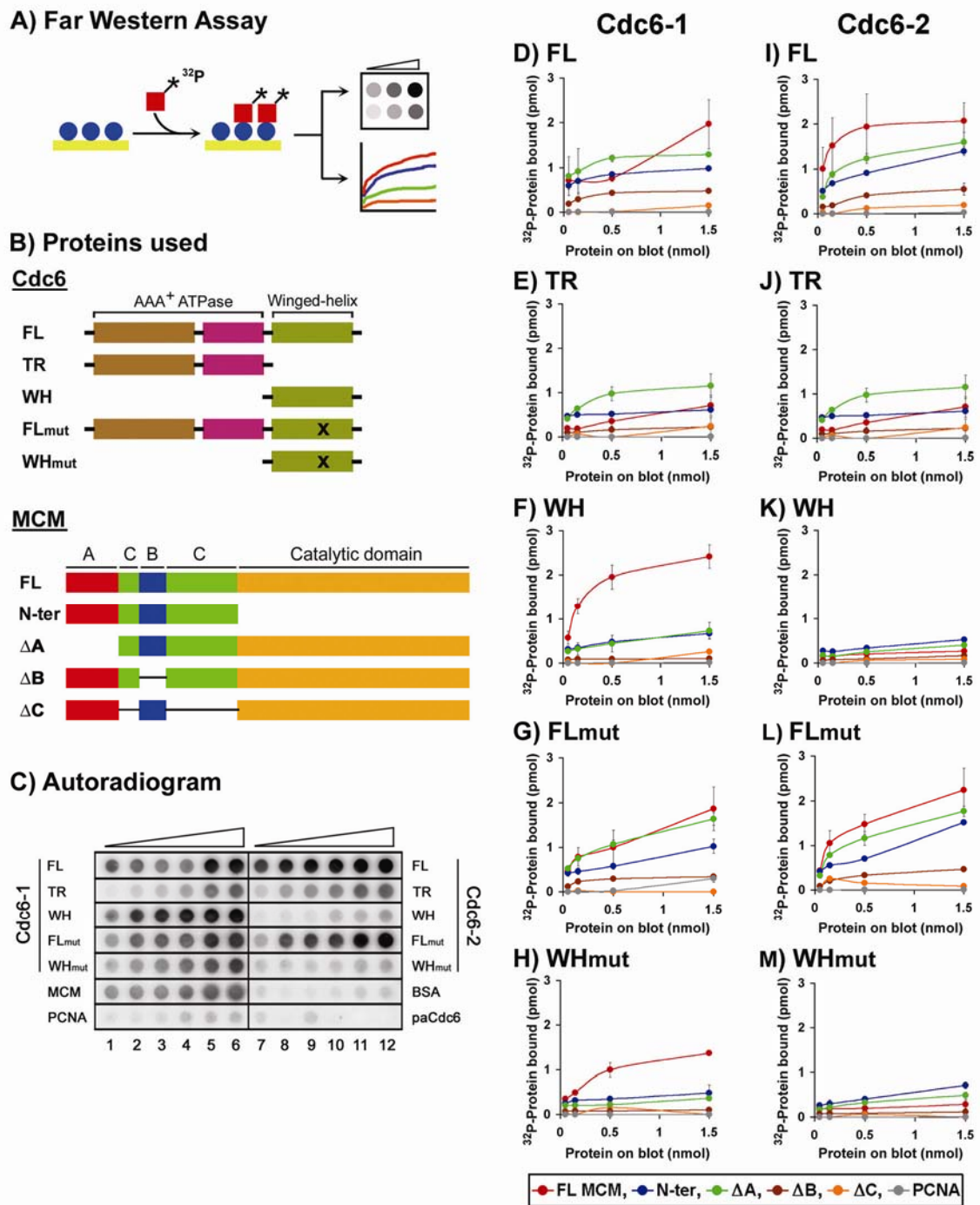
Methods), and their ability to interact was determined using a Far Western experiment (schematically described in Fig. 25A). The various amounts of the Cdc6 proteins (0.05-2.5 nmols) were absorbed onto a nitrocellulose filter and probed with <sup>32</sup>P-labeled full-length MCM or its derivatives at a concentration of 15nM (300 pmol of proteins in 20 ml reaction volume) (see Methods). An example of the results obtained with the full-length enzyme is shown in Figure 25C. The graphs in panels D-M summarizes the results of three independent experiments performed with the different probes. One should point out that the level of interactions is very low as only a few pmols of labeled proteins are interacting with nmol amounts of proteins on the filter. This may be explained by the low concentrations of MCM proteins used in the experiment and/or the low affinity of Cdc6 and MCM. This is supported by the inability of our group and others, to detect direct interactions between the proteins in sizing column or glycerol gradient.

Similar to the observation made with the two-hybrid analysis (Fig. 24A), the full-length MCM protein interacts efficiently with the full-length Cdc6-1 protein (Fig. 25C and D) and its WH domain (Fig. 25F) in the Far Western experiment. In fact, the interaction between MCM and the WH domain of Cdc6-1 is slightly better than the interaction with the full-length Cdc6-1 enzyme. It is possible that the region of interaction on the WH domain is exposed when the catalytic domains are removed from the Cdc6-1 molecule, resulting in better binding.

In the Far Western assay the full-length MCM also interacts with Cdc6-2 (Fig. 25I). The region(s) participating in MCM binding, however, are somewhat

different than those of Cdc6-1. While the Cdc6-1 WH domain binds to MCM, very limited binding could be detected between MCM and the WH domain of Cdc6-2 (Fig. 25 compare panels F and K). In fact, appreciable binding could be detected only between the full-length Cdc6-2 (Fig. 25I) and its mutant form (FLmut, Fig. 25L) and MCM.

Next, several truncated MCM proteins containing the cAMP-dependent protein kinase recognition motif were purified, labeled and analyzed for their ability to interact with the various Cdc6 proteins as described above for the full-length MCM enzyme. As shown in Figure 25 panels D-M, proteins containing only the N-terminal part of MCM, proteins missing domain A ( $\Delta A$ ) or domain B ( $\Delta B$ ), all interact with the full-length Cdc6-1 and -2 proteins and their mutant forms (FLmut) (panels D, G, I and L). The proteins also interact with the truncated and WH domains of Cdc6-1 and -2, but to a lesser extent (panels E, F, J and K). Protein lacking domain C ( $\Delta C$ ), however, did not interact with any of the Cdc6 derived enzymes (panels D-M). Like the two-hybrid analysis, these data show that the N-terminal portion of MCM plays a major role in Cdc6 interaction and that domain C may be required for binding. However, protein containing only the N-terminal domain binds Cdc6 weaker than the full-length enzyme, suggesting a minor role for the catalytic domain of MCM in Cdc6 binding.



**Figure 25. Cdc6 proteins interact with MCM protein in a Far western analysis.** A Far western assay was performed as described in Methods with various concentrations of Cdc6-1 and -2 derived proteins and *M. thermotrophicus* <sup>32</sup>P-labeled proteins as probes. (A) A schematic representation of the Far western dot blot assay. (B) The Cdc6 and MCM proteins used in the

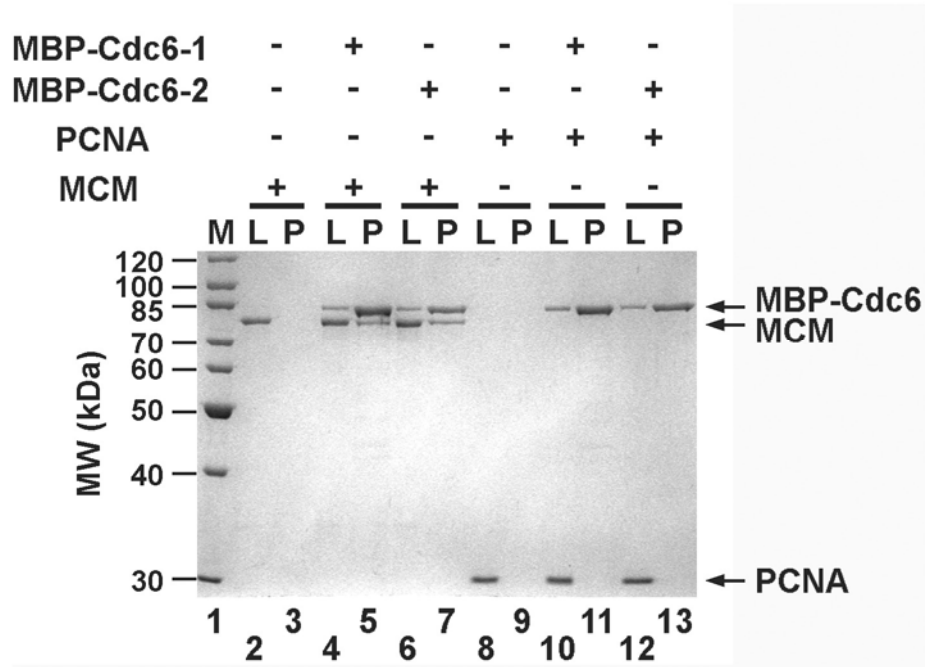
study. 'X', in the FLmut and WHmut of Cdc6 in (B), indicates the position of the WH mutations (Cdc6-1, R<sub>334,335</sub>→A and Cdc6-2, R<sub>337</sub>→A). (C) A representative blot obtained using FLMCM as a probe. The amount of proteins used in the blot is lanes 1 and 7, 0.05 nmol; lanes 2 and 8, 0.15 nmol; lanes 3 and 9, 0.25 nmol; lanes 4 and 10, 0.5 nmol; lanes 5 and 11, 1.5 nmol; lanes 6 and 12, 2.5 nmol. (D–M) The averages of three independent experiments (with error bars) for the amounts of the various probes used bound to the Cdc6-1 and -2 derived proteins. The colors used are red, FLMCM protein; blue, N-ter MCM; green, ΔA MCM; brown, ΔB MCM; orange, ΔC MCM; gray, PCNA. The colors used are also shown at the bottom of the figure.

As described above, the MCM protein lacking domain A (ΔA) binds to Cdc6-1 and -2 as well as the N-terminal part of the molecule (panels D and I). The protein lacking domain B (ΔB), however, binds to both Cdc6-1 and -2 more weakly than either the N-terminal part or the ΔA protein (panels D and I). These results may suggest that the Cdc6 binding site is located in the region connecting domain B and C. This possibility would also explain the observation that the protein lacking domain C cannot bind Cdc6.

Several controls were used to demonstrate that the interactions observed in the Far Western assays are specific for Cdc6 and MCM. BSA did not interact with the full-length MCM (Fig. 25C). In addition, a Cdc6 homologue from *P. aerophilum* also failed to interact with the *M. thermautotrophicus* MCM (Fig. 25C). These observations are consistent with the two-hybrid analysis that also failed to detect interactions between the *M. thermautotrophicus* and *P. aerophilum* proteins (Fig. 23 and 24B). These results demonstrated that although all archaeal Cdc6 proteins are similar in structure and primary amino acid sequences (100), their interactions with MCM are species-specific.

As an additional control, blots with the same Cdc6 proteins were probed with  $^{32}\text{P}$ -labeled *M. thermautotrophicus* PCNA protein (Fig. 25 panels D-M) (101). PCNA is a good negative control for MCM interacting proteins. Both proteins are ring-shaped homomultimers that encircle DNA and both have a similar charge distribution, with positive charged residues in the central cavity and negative charged residues on the outer surface (65,102). As shown in Figure 25 panels D-M, PCNA did not bind to Cdc6-1 and -2, further demonstrating that the interactions between Cdc6 and MCM are specific.

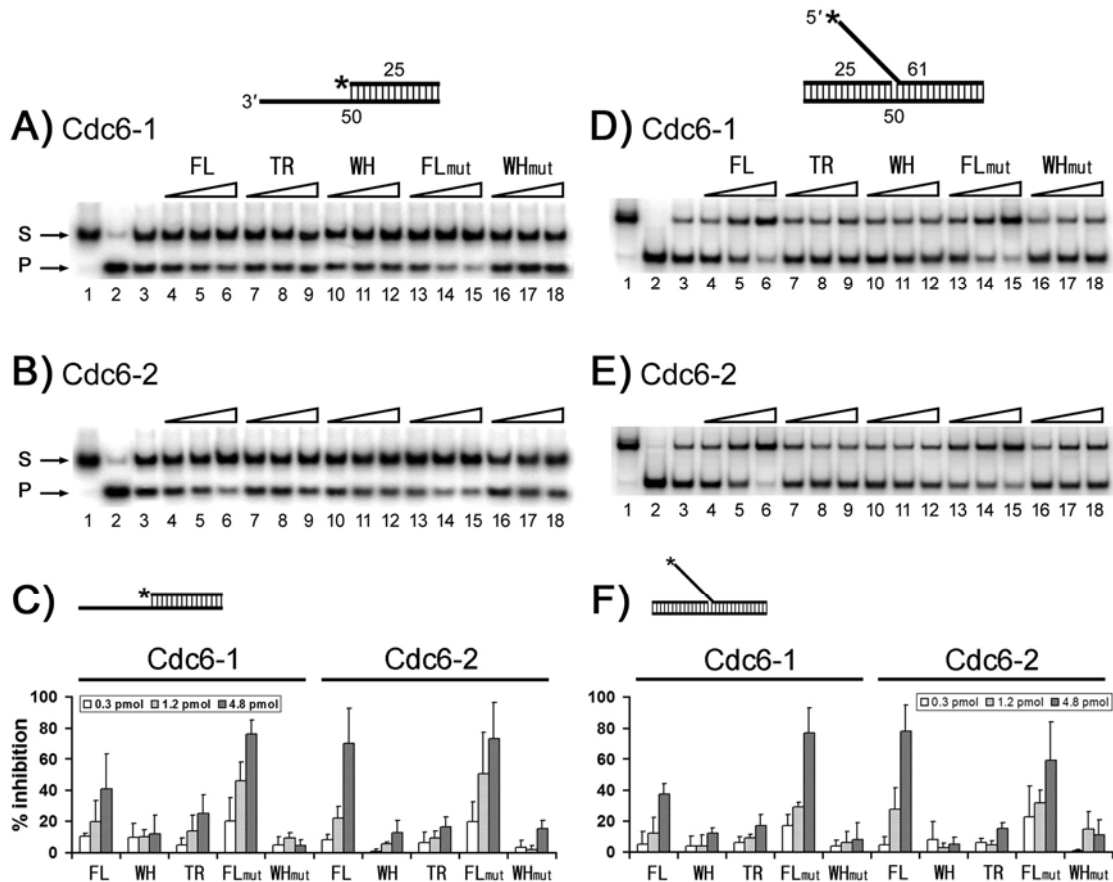
The experiments described in Figures 23, 24 and 25 were performed with either immobilized proteins or in heterologous system. Therefore, in order to demonstrate that purified *M. thermautotrophicus* MCM and Cdc6 interact in solution, pull-down experiments were performed. Un-tagged MCM protein was incubated with MBP-tagged Cdc6-1 and -2 proteins bound to amylose resin (Fig. 26). As shown in Fig. 26, MCM can be pulled-down by its association with either Cdc6-1 (lane 7) or -2 (lane 9). In the absence of Cdc6 proteins no MCM was observed in the pull-down fraction (lane 3) showing that the MCM does not associate with the resin by itself. As an additional control the *M. thermautotrophicus* PCNA protein was used. As shown, PCNA did not associate with either Cdc6-1 or -2 (lanes 11 and 13).



**Figure 26. Cdc6-1 and -2 proteins interact with MCM in solution.** Protein pull down assays were performed as described in Methods by binding 2  $\mu$ g of MBP-tagged Cdc6-1 or -2 proteins to amylose resin in the presence of 6  $\mu$ g of untagged MCM or PCNA proteins. Lane 1, molecular weight marker; lanes 2 and 3, MCM alone; lanes 4 and 5, MCM and MBP-Cdc6-1; lanes 6 and 7, MCM and MBP-Cdc6-2; lanes 8 and 9, PCNA alone; lanes 10 and 11, PCNA and MBP-Cdc6-1; lanes 12 and 13, PCNA and MBP-Cdc6-2. Lanes 2, 4, 6, 8 and 10 contain 10% of the reaction mixture and are marked by 'L'. Lanes 3, 5, 7, 9 and 11 contain the proteins eluted from the amylose resin and are marked by 'P'.

**Cdc6-MCM interactions are required for efficient regulation of MCM helicase activity by Cdc6.** After establishing that Cdc6 and MCM interact, the effect of those interactions on MCM helicase activity was determined. It was previously shown that Cdc6-1 and -2 inhibit MCM helicase activity (46). However, it is not yet clear whether Cdc6-MCM or Cdc6-DNA interactions play the major role in this inhibition. Therefore, the various Cdc6 mutant proteins

were studied for their ability to inhibit MCM helicase activity. As shown in Figure 27 both full-length Cdc6-1 and -2 proteins inhibit helicase activity (panels A and B compare lanes 4-6 to lane 3, see also panel C). As shown previously (46), Cdc6-2 inhibits MCM helicase activity better than Cdc6-1 (compare lanes 4-6 in panels A and B, see also panel C).



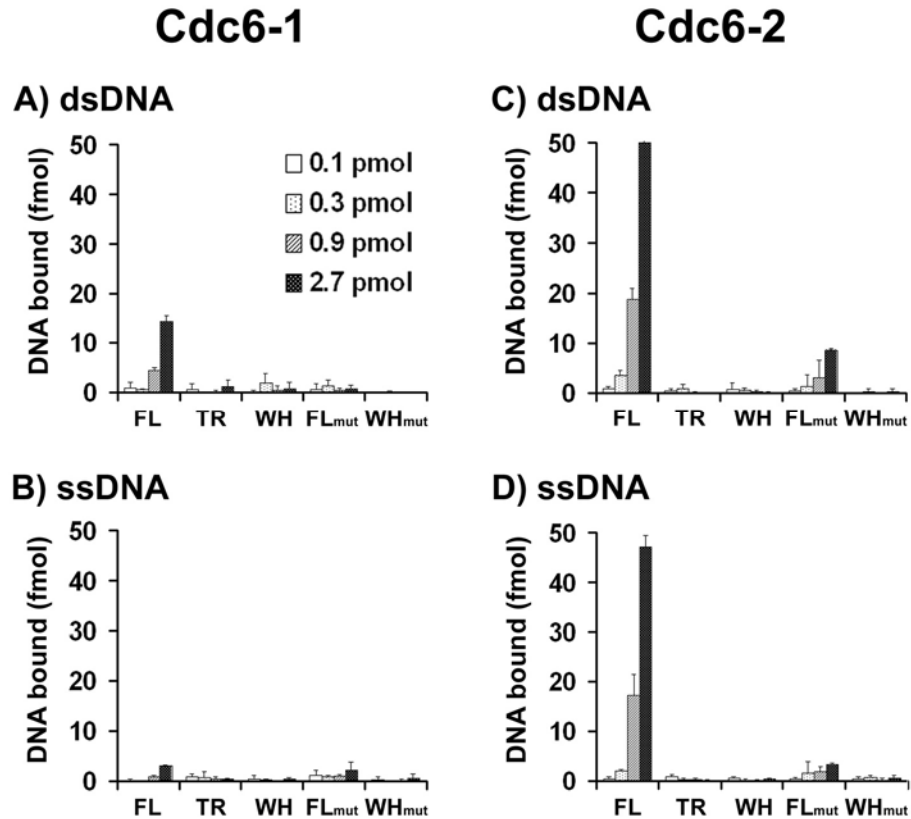
**Figure 27. Cdc6–MCM interaction is required for the inhibition of MCM translocation along DNA.** MCM helicase translocation along ssDNA (A–C) and dsDNA (D–F) was assayed as described in Methods in the presence of 0.3 pmol MCM(as monomer) and increasing amounts of Cdc6-1 and -2 proteins and their derivatives. (A, B, D and F) show representative gels. Lane 1, substrate only; lane 2, boiled substrate; lane 3, no Cdc6; lanes 4, 7, 10, 13 and 16, 0.3 pmol of Cdc6 protein as monomer; lanes 5, 8, 11, 14 and 17, 1.2 pmol of Cdc6 protein as monomer; lanes 6, 9, 12, 15 and 18, 4.8 pmol of Cdc6 protein as monomer. <sup>32</sup>P-labeled strands are marked



by an asterisk. (C and F) summarize the percent inhibition of MCM translocation (C) and duplex translocation (F) observed in the presence of the various Cdc6 proteins. Error bars represent the standard deviation calculated from 3 experiments. (Experiment performed by Jae-Ho Shin).

Is DNA binding by Cdc6 required for MCM inhibition? It was shown that substitution of two Arg residues (Arg<sub>334</sub> and Arg<sub>335</sub>) by Ala in the recognition helix of the WH domain of Cdc6-1 completely abolished dsDNA binding (41). In Cdc6-2 there is only a single Arg residue (Arg<sub>337</sub>) in a similar location and therefore this residue was also replaced by Ala. These mutations were generated in the full length and in the WH domain constructs of Cdc6-1 and -2. The genes encoding these mutant proteins were cloned into pET-21a (Novagen) for protein expression in *E. coli*.

As shown in Figure 28 while Cdc6-2 bind efficiently to both ss and dsDNA (panels C and D) Cdc6-1 bind only weakly to dsDNA and not at all to ssDNA (Fig. 28A and B). Although the WH domain was reported as the main interaction region between the archaeal Cdc6 proteins and dsDNA (39,41), the intact WH domain fails to interact with DNA (Fig. 28A and C). This may be explained by either the domain is misfolded or that the AAA<sup>+</sup> domains are also required for dsDNA binding. The AAA<sup>+</sup> were previously shown to be required for Cdc6 binding to ssDNA (39). As previously reported (41), the Cdc6-1 protein with mutation in the WH domain failed to interact with dsDNA (Fig. 28A). Although a protein with a similar mutation in Cdc6-2 retained some dsDNA binding, it was substantially reduced in comparison to the wild-type enzyme (Fig. 28C).



**Figure 28. An intact WH domain of Cdc6 is needed for DNA binding.** Filter binding assays were performed as described in Methods using 50 fmol of  $^{32}\text{P}$ -labeled ss or dsDNA oligonucleotides in the presence of 0.1, 0.3, 0.9 and 2.7 pmol of proteins (as monomer). The averages with standard deviations of three experiments are shown. (A) dsDNA binding of Cdc6-1 proteins; (B) ssDNA binding of Cdc6-1 proteins; (C) dsDNA binding of Cdc6-2 proteins; (D) ssDNA binding of Cdc6-2 proteins.

To determine whether DNA binding by Cdc6 is required for the inhibition of MCM helicase activity, the WH mutant proteins were studied for their effect on MCM helicase activity. As shown in Figure 27, both Cdc6-1 and -2 full-length enzymes with mutations in the WH motif (FLmut) are capable of inhibiting the helicase (Fig. 27 panels A and B, lanes 13-15; see also panel C) illustrating that

DNA binding is not essential for the inhibition. Interestingly, the mutant Cdc6-1 protein appears to be a better inhibitor than the wild-type enzyme (Fig. 27 panels A, compare lanes 13-15 to lanes 4-6; see also panel C). The WH domains of the Cdc6 proteins were shown to be required for interaction with MCM (Fig. 23, 24 and 25) and for DNA binding (39,41). Therefore, these interactions may compete and thus when Cdc6-1 interaction with DNA is abolished, tighter interactions with MCM can occur, resulting in better inhibition. Such competition may play a role during the initiation process at the origin DNA (see discussion). Furthermore, though the WH protein of Cdc6-1 interacted with the MCM, no efficient helicase inhibition could be observed. This could be because interaction between the full-length Cdc6-1 protein and MCM may dissociate the MCM complex (as was previously suggested (60)) while the WH domain of Cdc6-1 protein may not.

It was shown that the *M. thermautotrophicus* MCM as well as the eukaryotic helicase can translocate along duplex DNA (77,88). It was therefore suggested that dsDNA translocation by the replicative helicase may play a role during the initiation and/or elongation phases of DNA replication (76). As Cdc6 proteins play an essential role in the initiation process in eukarya, and probably in archaea as well, the effects of the interactions between Cdc6 and MCM on duplex translocation by the MCM enzyme were studied.

As shown in Figure 27 panels D-F, and similar to the results with helicase translocation along ssDNA (panels A-C), both full-length Cdc6 proteins inhibit duplex translocation by the helicase. Like inhibition of ssDNA translocation, Cdc6-2 is a better inhibitor in comparison to Cdc6-1 (Fig. 27 compare lanes 4-6

in panels D and E; see also panel F). Cdc6-1 mutant protein, devoid of DNA binding, is a much better inhibitor of duplex translocation by the helicase than the full-length protein (Fig. 27 panel D, compare lanes 13-15 to lanes 4-6; see also panel F).

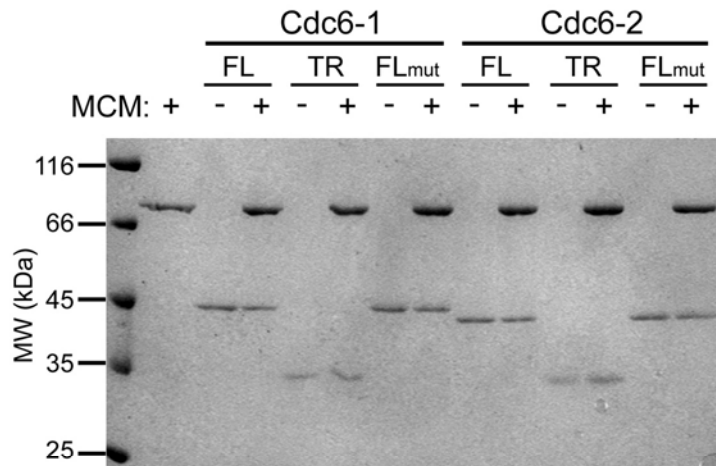
**Cdc6-MCM interaction modulates Cdc6 autophosphorylation.** After demonstrating that the interaction between Cdc6 and MCM is involved in the regulation of the helicase activity, the effect of the interactions on Cdc6 autophosphorylation was determined. To date, the only biochemical activity, besides DNA binding, shown for the archaeal Cdc6 proteins are their ability to undergo autophosphorylation in the presence of  $\gamma$ -ATP (39,45,46,52). Although the level of phosphorylation *in vitro* is very low (39) it has been reported for all archaeal proteins studied. Therefore, it was proposed that the autophosphorylation might play a regulatory role during the initiation process (1,2,39). It was also hypothesized that Cdc6-MCM interactions may regulate the phosphorylation activity during assembly of the helicase around DNA at the origin (2,39). The regulation of the autophosphorylation by DNA binding was previously demonstrated (39).

Thus, the effect of the interactions between Cdc6 and MCM on Cdc6 autophosphorylation was analyzed (Fig. 29). A mutant form of MCM in which Ala replaced Lys325 was used in order to prevent ATP hydrolysis by MCM that would otherwise limit the available ATP for the Cdc6 autophosphorylation reaction. As shown in Figure 29, the presence of MCM modulates the autophosphorylation of both Cdc6 full-length proteins. However, the effect of the

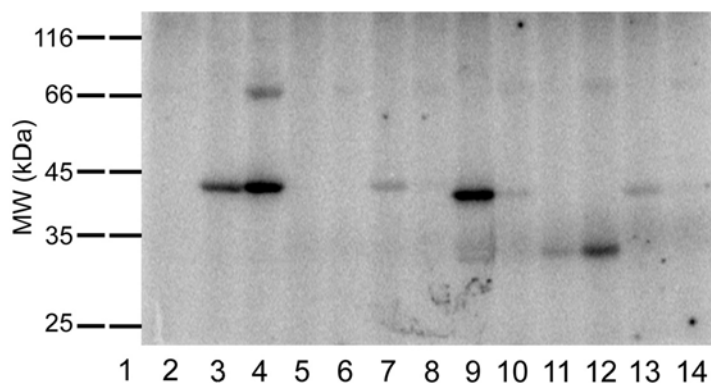
interaction with MCM has opposite effects on Cdc6-1 and -2. While binding of MCM stimulates the phosphorylation of Cdc6-1 (Fig. 29 panels A and B, compare lane 4 to lane 3), the interaction with Cdc6-2 inhibits the autophosphorylation (Fig. 29 panels A and B, compare lane 10 to lane 9). It is possible that the WH domain of Cdc6-1 contains the phosphorylation site, and in solution it is packed against the rest of the molecule and prevents it from being efficiently phosphorylated. MCM binding to the WH domain of Cdc6-1 (Fig. 23, 24 and 25) may expose the residue, resulting in better phosphorylation. In support of this idea, a truncated form of Cdc6-1, missing the WH domain, cannot be phosphorylated (Fig. 29 panels A and B, lanes 5 and 6). This is in contrast to Cdc6-2 in which a truncated protein retains the ability to autophosphorylate (Fig. 29 panels A and B, lanes 11 and 12) (39). In addition, the mutant forms of both full-length proteins phosphorylated to a lesser extent than the unmutated enzymes. It is possible that the mutation affects the WH structure in such a way that the phosphorylation site is not accessible.

The reasons for the opposite effect of MCM on Cdc6-1 and -2 phosphorylation are currently unknown. However, in light of the prevailing hypothesis that one Cdc6 is needed for origin recognition while the other for MCM loading (2) one would expect a different binding between the two proteins and MCM and this may result in different effect on the phosphorylation. When a helicase loading assay is developed this hypothesis could be tested.

## A) Coomassie



## B) Autoradiogram



**Figure 29. Cdc6 autophosphorylation is regulated by MCM binding.** Cdc6 phosphorylation reactions were performed as described in Methods in a reaction mixture (15 ml) containing 10 pmol of Cdc6 protein and 3.3 pmol of  $[\gamma\text{-}^{32}\text{P}]\text{ATP}$  in the absence (lanes 3, 5, 7, 9, 11 and 13) or in the presence (lanes 4, 6, 8, 10, 12 and 14) of 20 pmol MCM. The autophosphorylation reactions were carried out for 20 min at 65°C. Following incubation, the proteins were separated by 10% SDS-PAGE and visualized by Coomassie blue staining (A) and autoradiography (B). Lane 1, molecular mass (kDa); lane 2, MCM alone; lanes 3 and 4, Cdc6-1 full-length protein; lanes 5 and 6, Cdc6-1 truncated proteins; lane 7 and 8, Cdc6-1 full-length protein with mutant WH domain; lanes 9 and 10, Cdc6-2 full-length; lanes 11–12, Cdc6-2 truncated proteins; lane 7 and 8, Cdc6-2 full-length protein with a mutated WH domain.

## 4.5 DISCUSSION

In both eukarya and archaea, the mechanism by which the MCM helicase is assembled around the DNA at the origin is not yet understood. In both systems, however, the prevailing notion is that the Cdc6 protein, in conjunction with ORC (or its functional homolog in archaea), plays an essential role in helicase loading. In addition, it is believed that the mechanism will bear similarity to the assembly of the *E. coli* DnaB helicase at *oriC*.

In the archaeon *M. thermautotrophicus* it was suggested that one of the Cdc6 proteins is the functional homologue of the bacterial DnaA while the other is the functional homologue of DnaC (1,74). In addition, structural similarities between the bacterial DnaA and the archaeal Cdc6 proteins have been reported (40). The study reported here, on the *M. thermautotrophicus* Cdc6 and MCM proteins, demonstrates that the Cdc6 and MCM proteins interact. It is also shown that the association between the two proteins regulates their respective enzymatic properties.

**What is needed for Cdc6-MCM interaction?** Two-hybrid and Far Western analysis demonstrated the interactions between MCM and the two Cdc6 homologues identified in the *M. thermautotrophicus* genome. The study also identified the domains needed for the interactions between them. Although both Cdc6 homologues are very similar in primary amino acid sequence (100) and were suggested to have similar structure and domain organization (28,38), they appear to utilize different regions for MCM binding. While Cdc6-1 binds MCM predominantly via the WH domain, this domain of Cdc6-2 does not interact with

the helicase. Only the full-length Cdc6-2 protein showed appreciable MCM binding. This is similar to the observations made with a Cdc6 homologue from the archaeon *S. solfataricus* in which an indirect assay suggested that the WH domain of one of the three Cdc6 homologues found in the organism is not required for MCM binding (54). In addition, it was shown that although the three Cdc6 homologues of *S. solfataricus* are very similar in primary amino acid sequence, they have different functions (24).

The interaction between Cdc6-2 and MCM may also be similar to that of DnaA and DnaB in *E. coli*. It was shown that, while DnaA binds to the DnaB helicase via a region located at the N-terminal part of the AAA<sup>+</sup> catalytic domains (103), the origin recognition domain is at the C-terminal region of the molecule (40). Similarly, it was demonstrated that the eukaryotic Cdc6 protein also interacts with MCM via the AAA<sup>+</sup> catalytic domains, which are separated from the WH domain (104).

The archaeal MCM proteins can be divided into two main parts, the N-terminal region, needed for protein multimerization and DNA binding, and the C-terminal AAA<sup>+</sup> catalytic domain responsible for catalytic activity (59). The data presented here suggest that the major contact between MCM and Cdc6 is via domain C of the N-terminal portion of MCM. Neither domain A, suggested having a regulatory role, nor domain B, needed for ssDNA binding, are essential for Cdc6 interaction. In addition, neither ss nor dsDNA binding by Cdc6 or MCM are needed for their interactions, as mutant proteins devoid of DNA binding retain



their ability to interact with MCM, and DNA was not present in the Far Western analysis.

**Do the *M. thermautotrophicus* Cdc6-1 and -2 proteins play different roles during the initiation process?** The *M. thermautotrophicus* Cdc6-1 and -2 proteins are proposed to have different functions during the initiation process. It was suggested that one protein is the origin binding protein, and thus is the functional homologue of the eukaryotic ORC and bacterial DnaA, while the other is the functional homologue of the eukaryotic Cdc6 and bacterial DnaC and participates in helicase loading (2).

*In silico* analysis of different archaeal Cdc6 proteins suggests that they belong to two distinct subgroups, referred to as group I and II (38,100). It is not yet clear, however, whether these two groups have different functions. It is also not yet clear whether they have different structures, as the two structures solved to date are of proteins belonging to subgroup II (28,38). As part of the difference between the two subgroups lie in the WH domains, it was suggested that they might bind DNA differently. While the *M. thermautotrophicus* Cdc6-2 (subgroup II, (38)), can bind both ss and dsDNA (Fig. 28, see also chapter 5), Cdc6-1 (subgroup I, (38)), can bind only to dsDNA [(41), Fig. 28, see also chapter 5]. The observation that the two *M. thermautotrophicus* Cdc6 proteins interact differently with DNA substrates and utilize different regions for MCM binding support the hypothesis of different roles for these proteins during the initiation process.

It was proposed that the autophosphorylation of Cdc6 proteins may regulate helicase loading and/or be regulated by the initiation process (39). The observation that Cdc6-1 and -2 autophosphorylation are regulated by MCM supports this hypothesis. Furthermore, the observation that autophosphorylation of one protein is stimulated by MCM binding while that of the other is strongly inhibited may suggest different roles for Cdc6-1 and -2 during the initiation process. Whether this is the case remains to be seen.

**Does a switch mechanism between Cdc6, MCM, and DNA regulate the initiation of DNA replication?** Studies have demonstrated that the archaeal Cdc6 proteins can regulate MCM helicase activity (45,46,52). There are several possible mechanisms for this inhibition. Direct binding of Cdc6 to MCM may prevent helicase movement along the DNA. The interaction may also destabilize the MCM complex or destabilize MCM interactions with DNA. Alternatively, binding of Cdc6 to the ssDNA and/or the duplex part of the DNA substrate may prevent helicase translocation along the DNA. Previous studies using full-length Cdc6 proteins from different archaeons (46) suggested that direct interactions between Cdc6 and MCM are required for the inhibition. These studies illustrated that when the Cdc6 and MCM are from the same organism efficient inhibition could be observed. Only limited inhibition could be detected when Cdc6 and MCM from different species were tested. As all Cdc6 proteins studied retained their DNA binding activity these observations suggested that protein-protein interactions are needed for helicase inhibition.

The use of the large number of mutant proteins described in this study takes this analysis further. The data show that Cdc6-1 and -2 have different inhibitory effects on MCM translocation. While Cdc6-2 is an efficient inhibitor of both ss and dsDNA translocation by the helicase, only about 50% of the inhibition observed with Cdc6-2 could be detected with Cdc6-1. The study also clearly illustrates that DNA binding by Cdc6 is not the predominant factor for the inhibition of MCM activity. Cdc6-DNA interaction may play some role, however. Furthermore, mutant forms of Cdc6-1 and -2, impaired in DNA binding, inhibit MCM helicase activity as well or better than the wild-type enzyme. This is an interesting and potentially important observation. It suggests that MCM and DNA may compete for Cdc6 binding. When Cdc6 cannot bind DNA, it binds MCM tighter and thus efficiently inhibits the helicase activity. It may suggest the possibility of a switch mechanism mediated by Cdc6 occurs during the initiation process in which Cdc6 binds to MCM, brings it to the DNA at the origin and then switches to DNA binding, releasing the helicase, which can then associate with the DNA. It is also possible that only one of the Cdc6 proteins (the helicase loader) is involved in the switch, while the other forms the DNA structure (replication bubble) on which the helicase will be loaded.

## CHAPTER 5

# DNA BINDING BY THE Cdc6 PROTEIN IS REGULATED BY THE MCM HELICASE

### 5.1 ABSTRACT

The initiation of DNA replication depends upon the recognition of the origin of replication by specific proteins. Among those are the *Escherichia coli* DnaA and the eukaryotic origin recognition complex (ORC). In the archaeon *M. thermautotrophicus* two homologues of ORC have been identified which are referred to as Cdc6-1 and -2. It is thought that these archaeal proteins function in origin recognition and helicase loading. Studies have shown that both Cdc6 proteins bind with greater affinity to repeat DNA sequences found within the organism's origin of replication in comparison to random sequences and that both proteins also interact with the replicative helicase MCM (see chapter 4). The mechanism of helicase assembly at the origin, however, is currently unknown. As a step to gain insight into the process, the effect of the interactions between MCM and Cdc6 on their respective DNA binding properties was studied. While Cdc6-2 is capable of binding both single-stranded and double-stranded DNA, Cdc6-1 can bind only dsDNA with preference to origin-derived sequences. It was also found that although the winged-helix (WH) domain at the C-terminus of the Cdc6 proteins is needed for DNA binding, chimeric proteins containing the catalytic domain of one enzyme (homologue) and the WH domain of the other

failed to bind DNA. These observations suggest that only intact Cdc6 proteins can interact with DNA. Using MCM mutant proteins devoid of DNA binding, it was found that MCM interactions with Cdc6 inhibit its DNA binding. Taking together with previously made observations these results provide a basis for a working hypothesis of the initiation of *M. thermautotrophicus* DNA replication.

## 5.2 INTRODUCTION

Origin recognition is a critical event during the initiation of DNA replication during S-phase. This function is carried out by specific origin recognition protein(s) that bind to specific sequences within the origin. In archaea, this activity is carried out by the archaeal homologues of the eukaryotic Cdc6 and the origin recognition complex (ORC) proteins (24,41) referred to in the text as Cdc6.

The three dimensional structure of the archaeal Cdc6 revealed a three-domain composition (28,38). Domain I, with a RecA type fold together with domain II, form the catalytic part of the molecule, where the ATP binding pocket is located between these two domains. The C-terminus portion of the protein has a winged-helix (WH) fold found in a number of DNA binding proteins (see chapter 1 for details).

Only limited studies on the archaeal Cdc6 enzymes have been reported. The proteins from the archaeon *M. thermautotrophicus* were shown to interact and regulate the helicase activity of the MCM helicase [(46,47), see also chapter 4]. The Cdc6 proteins from several organisms were shown to bind ss and dsDNA in a sequence independent manner via the WH domain (39,53,68) and

this nonspecific binding regulate the autophosphorylation of the proteins (39) In addition, origin specific binding was also been reported with the *M. thermautotrophicus* (41) and *S. solfataricus* (24) enzymes. It was also shown that the *M. thermautotrophicus* proteins bind with greater affinity to inverted repeats found within the origin region in comparison to random DNA sequences (41).

The biochemical properties of the archaeal MCM are better understood (summarized in chapter 1). Two structural features in MCM were shown to be involved in DNA binding. A zinc-finger motif was shown to participate in ssDNA binding (70) while a role for the  $\beta$ -finger motif in dsDNA binding was also been suggested (65,105).

Though the Cdc6 proteins are thought to play an essential role in helicase assembly at the origin, the mechanism is currently unknown. As part of the ongoing effort to address this question, the effect of the interactions between MCM and Cdc6 on the DNA binding of the proteins was investigated. It is shown here that while Cdc6-1 binds only dsDNA with preference to origin derived sequences, Cdc6-2 binds both ss and dsDNA with no preferential binding to origin sequences. It is also shown that MCM binding to Cdc6-1 and -2 inhibits their DNA binding.

### 5.3 METHODS

**Construction of Cdc6 chimeric proteins.** Cdc6 chimeric proteins were generated using a PCR-based approach as previously described for the

construction of MCM mutants (see chapter 2) using plasmid containing the gene encoding the wild-type *M. thermotrophicus* Cdc6-1 and -2 proteins (39). The three dimensional structures of the *P. aerophilum* and *A. pernix* Cdc6 proteins (28,38) served as the guide for the construction of the chimeric proteins. To generate the Cdc6-N2C1 chimera, which contains the AAA<sup>+</sup> catalytic domains of Cdc6-2 (amino acids 1-291) and the WH domain of Cdc6-1 (amino acids 287-382) the MR183 (5'-GGACCCGTGAATCTCATAGGGGGGGACATAATACTGACACTACCG-3') and MR184 (5'-CGGTAGTGTCAGTATTATGTCCCCCCTATGAGATTACGGGTCC-3') oligonucleotides were used. To generate the other Cdc6 chimera, Cdc6-N1C2, which contains the AAA<sup>+</sup> catalytic domains of Cdc6-1 (amino acids 1-286) and the WH domain of Cdc6-2 (amino acids 292-379) the MR185 (5'-GAACACAACAAGATCACAGGGGGGCACACGGTGCGAACCCCTGAAC-3') and MR186 (5'-GTTTCAGGGTTCGCACCGTGTGCCCCCCTGTGATCTTGTTGTGTTTC-3') oligonucleotides were used. The proteins were expressed and purified as previously described for the wild-type enzyme (46).

**Preparation of DNA substrates for filter binding assays.** Single stranded DNA substrates for filter binding assays were prepared by labeling the oligonucleotide using [ $\gamma$ -<sup>32</sup>P]ATP and T4 polynucleotide kinase. The labeled oligonucleotides were purified as previously described (77). For dsDNA substrates, the labeled ssDNA was annealed with 2X concentration of its unlabeled complementary strand in the presence of 50mM NaCl and 20mM Hepes-HCl (pH=7.0) by boiling for 5 min at 95°C followed by slow cooling to 30°C followed by 90 min incubation at 30°C. The substrates were then purified as

previously described (77). The oligonucleotides used in this study are origin DNA (5'-TTTACACTTGAAAGGGTTTACACTTGAAAGGGTTTACACTTGAAA-3') and random DNA (5'-TACATATGTACATGGGTACATATGTACATGGGTACATATGTACAT-3') sequences. The origin sequence repeats are underlined and the random sequences are double-underlined. The random sequences were generated by maintaining the base composition as the origin specific sequences but scrambling their order.

**Filter binding assay.** All nitrocellulose DNA filter binding assays were performed with 2.5 nM of either <sup>32</sup>P-labeled origin specific or random, ssDNA or dsDNA. The binding reactions were performed as described in chapter 2.

DNA binding assays performed for determining the substrate specificity of Cdc6, Cdc6 chimeras and MCM proteins, contained 15, 30, 45, 90, 150 and 225 nM of protein (as monomer). The experiments performed to determine the effect of MCM on Cdc6 DNA binding, contained 15, 30, 45, 90 and 150nM of Cdc6 proteins (as monomer) and 300 nM of the  $\beta$ -finger mutant of MCM protein (as monomer). The effect of Cdc6 on MCM DNA binding was performed with 15, 30, 45, 90 and 150 nM MCM (as monomer) and 300 nM of Cdc6-1 protein with mutation in the WH domain.

All DNA binding experiments were repeated three times and their averages with standard deviations are shown in the figures.

**DNA helicase assay.** The substrate for DNA helicase assays was made as previously described (77) by annealing a 61-mer oligonucleotide 5'- (TTTG)<sub>9</sub>CCGACGTGCCAGGCCGACGCGTCCC-3' which was pre-labeled with



[ $\gamma$ - $^{32}$ P]ATP and T4 polynucleotide kinase, to a 74-mer oligonucleotide 5'-GGGACGCGTCGGCCTGGCACGTCGGCCGCTGCGGCCAGGCACCCGATG GC(GTTT) $_6$ -3'. The substrate was purified as described previously (77).

DNA helicase activity was measured in reaction mixtures (15  $\mu$ l) containing 20mM Tris-HCl (pH=8.5), 10mM MgCl $_2$ , 2mM DTT, 100  $\mu$ g/ml BSA, 5mM ATP, 0.67 nM of  $^{32}$ P-labeled DNA substrate (3,000 cpm/fmol), 13.33 nM of MCM protein (as monomer) and 26.67, 80 and 240 nM of Cdc6-1, -2 or their chimera proteins as indicated in the figure legend. The assay was carried out as described in chapter 2.

## 5.4 RESULTS

### ***M. thermautotrophicus* Cdc6-1 and -2 bind differently to DNA.**

Previous studies, using a 282bp DNA fragment derived from the *M. thermautotrophicus* putative origin of replication (42) demonstrated that both Cdc6-1 and -2 bind to origin DNA (41). It was also shown that the proteins bind tighter to DNA fragments which contains three inverted repeats found within the origin region in comparison to a fragment that contain only a single repeat (41).

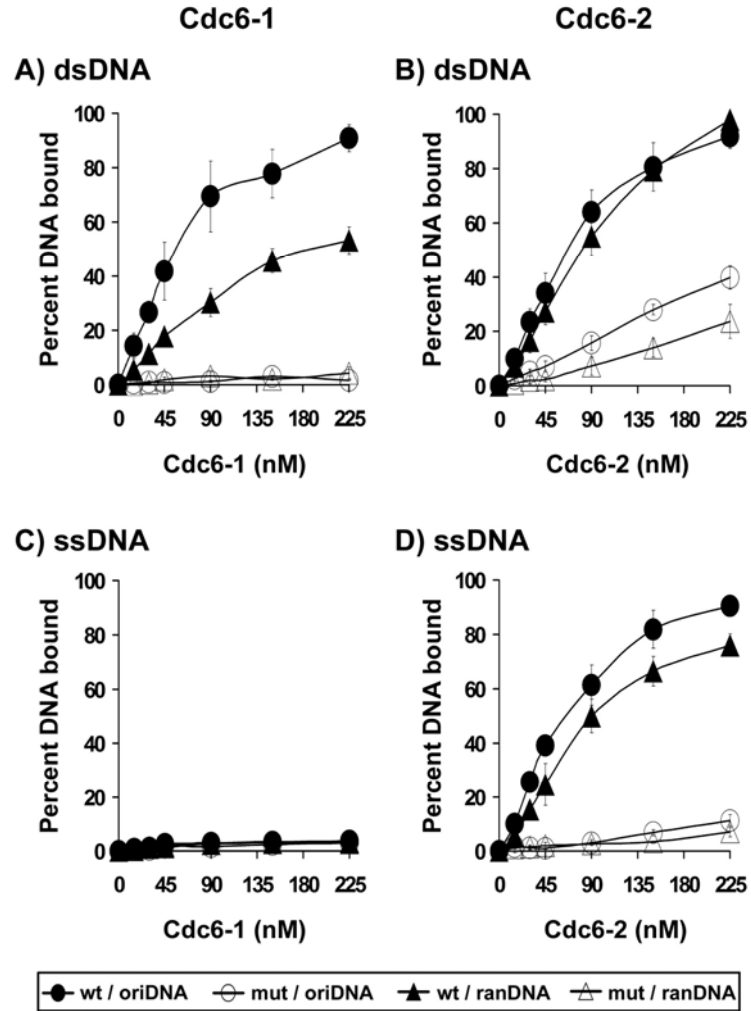
These experiments have been expanded to determine the binding of the proteins to ssDNA and under physiological conditions. The previous studies did not evaluate the ssDNA binding by the Cdc6 proteins and all experiments were performed at 37°C which is much lower than the physiological growth temperature (60°C) of *M. thermautotrophicus*. A number of biochemical studies, including those with Cdc6, MCM, the polymerase and its accessory proteins,

have shown that the enzymes are not active at 37°C (39,106,107). Thus, the binding of Cdc6-1 and -2 to oligonucleotide containing three tandem repeats found within the origin was studied at 60°C, the physiological temperature. As shown in Figure 30, and similar to the work by Capaldi and Berger, both proteins bind to dsDNA containing the origin derived inverted repeats (Fig. 30A and B, closed circle). Cdc6-2 bound efficiently to both specific and random DNA (Fig. 30B, closed circle and closed triangle). Cdc6-1, on the other hand, showed a clear preference to origin specific double-stranded sequences (Fig. 30A, compare closed circle to closed triangle). The results with Cdc6-1 are similar to those previously reported with a longer (282mer) DNA substrate which contain three origin repeats and long random sequence regions between them (41).

Next, the ability of the proteins to bind ssDNA was determined using origin specific and random DNA. While Cdc6-2 binds both ssDNA substrates efficiently (Fig. 30D, closed circle and closed triangle), no binding could be detected with Cdc6-1 (Fig. 30C, closed circle and triangle).

The archaeal Cdc6 proteins contain a WH motif at its C-terminus (28,38). Indirect studies suggested that this domain of Cdc6-2 is required for dsDNA but not ssDNA binding (39). In addition, using a mutant form of Cdc6-1 in which the DNA recognition helices was mutated it was shown that the domain is needed for dsDNA interactions. It is not yet clear if the WH domain also involved in ssDNA binding. Therefore, mutant proteins were generated and their ability to bind DNA was evaluated. It was shown that two Arg residues (Arg<sub>334</sub> and Arg<sub>335</sub>) located in the recognition helix of the WH domain of Cdc6-1 play a major role in dsDNA

binding (41). Thus a protein in which the two Arg residues were replaced by Ala was generated. In Cdc6-2 there is only a single Arg residue (Arg<sub>337</sub>) in a similar, but not identical location and it also was replaced by Ala.



**Figure 30. Cdc6-1 but not Cdc6-2 exhibit preferential binding to origin DNA.** DNA binding analysis of Cdc6-1 (A and C, filled symbols) and Cdc6-2 (B and D, filled symbols) and Cdc6-1 WH mutant (A and C, open symbols) and Cdc6-2 WH mutant (B and D, open symbols) were performed as described in Methods using filter binding assays in the presence of 2.5 nM <sup>32</sup>P-labeled origin specific (circles) or random (triangle) DNA sequences in the presence of 15, 30, 45, 90, 150 and 225 nM of Cdc6 proteins (as monomer). Panels A and B, dsDNA; panels C and D, ssDNA. The average result of three experiments is shown with standard deviations.

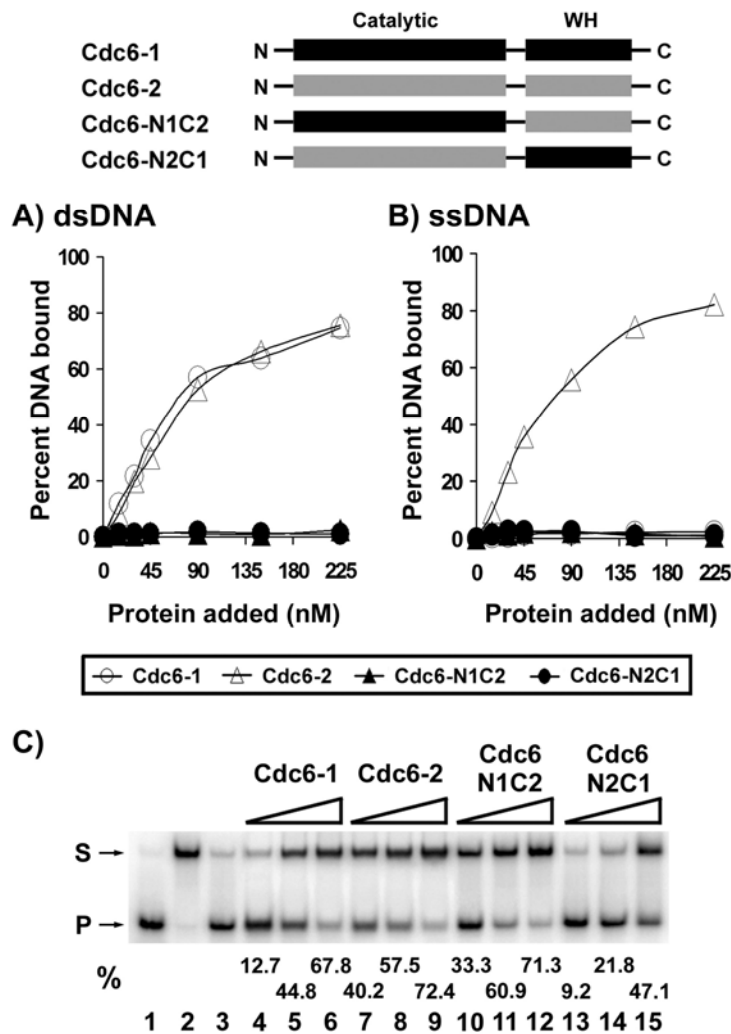
Whether the proteins with the mutation in the WH retain the ability to bind DNA was studied using the origin specific and non-specific substrate (Fig. 30, open circle and open triangle). It was found that while the mutation in Cdc6-1 completely abolished dsDNA binding (Fig. 30A, open circle and open triangle) the mutation in Cdc6-2 reduced dsDNA binding by about 50% (Fig. 30B, open circle and open triangle). The effect on ssDNA binding, however, was much more prominent with about 90% reduction in DNA binding in comparison to the wild-type Cdc6-2 enzyme (Fig. 30D, open circle and open triangle).

The data presented in Figure 30 suggest that the WH domains play a major role in DNA binding by Cdc6. Thus, the ability of the intact WH domains of Cdc6-1 and -2 and the truncated forms of the enzymes which do not have the WH domains were evaluated for their ability to interact with DNA. No DNA binding could be detected with proteins containing only the WH domain or with truncated proteins in which the WH domain was deleted [(47) and see chapter 4].

This observation suggested that a full-length Cdc6 is needed for DNA binding. It is possible however that a protein composed of the catalytic domain of one and the WH domain of the other will retain DNA binding activity. Thus chimera proteins that contain the N-terminal catalytic domains from one Cdc6 and the C-terminal WH domain from the other (Cdc6-N1C2 and Cdc6-N2C1, schematically shown at the top of Fig. 31) were generated and tested for their ability to bind DNA. As shown in Figure 31A and B, neither chimera could interact with DNA (closed circle and closed triangle). These results suggest that not only both the WH and catalytic domains are needed for DNA binding but that

only the WH and catalytic domain from the same protein can bind DNA. The differences in substrate specificity or structural differences between the two Cdc6 proteins may result in these requirements.

It was previously shown that the archaeal Cdc6 proteins inhibit MCM helicase activity (46). It was also been shown that direct protein-protein interactions and not DNA binding by Cdc6 is required for the inhibition [(47), see also chapter 4]. Therefore, this assay was used as a means to demonstrate that the chimeras are properly folded. As shown in Figure 31C, both chimeras inhibit MCM helicase activity suggesting that the chimera proteins are properly folded.

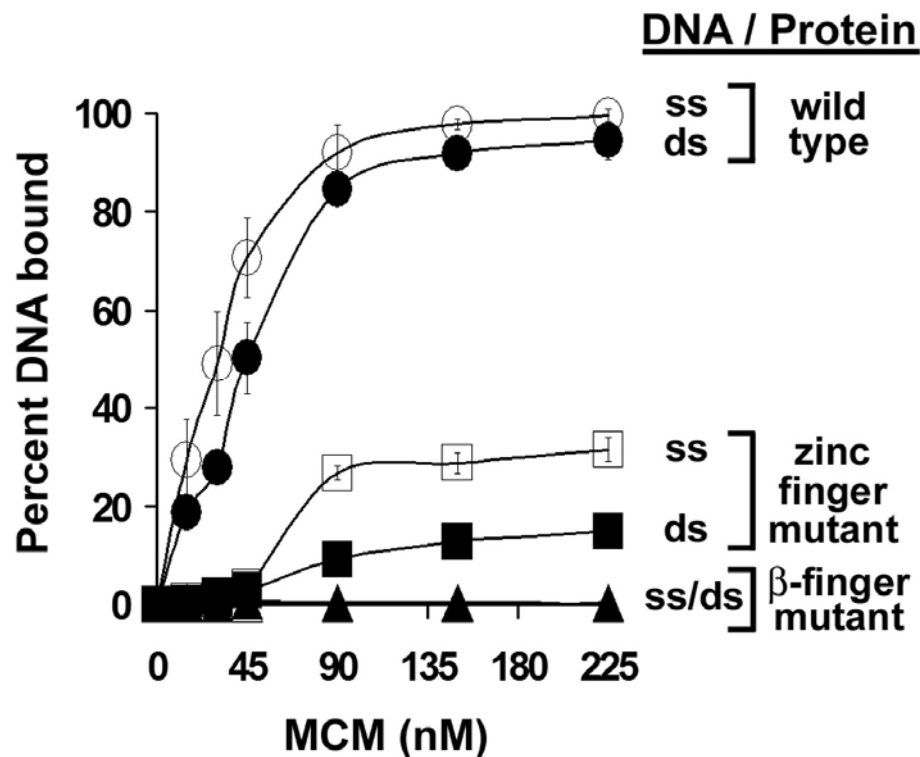


**Figure 31. Cdc6 chimera proteins can not bind DNA.** Top: Schematic representation of the Cdc6 chimera proteins. DNA binding by Cdc6-1 (panels A and B, open circle) and Cdc6-2 (panels A and B, open triangle) and its chimeras, Cdc6-N1C2 (panels A and B, closed triangle) and Cdc6-N2C1 (panels A and B, closed circle) were performed using filter binding assays as described in Methods in the presence of 2.5 nM <sup>32</sup>P-labeled origin specific ssDNA (A) or dsDNA (B) in the presence of 15, 30, 45, 90, 150 and 225 nM of proteins (as monomer). The average result of two experiments is shown. C) The effect of full-length and chimera Cdc6 proteins on the helicase activity of MCM was determined as described in Material and Methods in the presence of 13.33 nM MCM (as monomer) and 26.67 nM lanes 4, 7, 10 and 13, 80 nM lanes 5, 8, 11 and 14, and 240 nM lanes 6, 9, 12, and 15 of the Cdc6 proteins (as monomers). Percent inhibition of helicase activity in comparison to the reaction without Cdc6 (lane 3) is marked by %.

**The  $\beta$ -finger of *M. thermautotrophicus* MCM is required for both ss and dsDNA binding.** It was previously shown that the zinc-finger motif located within domain B of the N-terminal part of MCM play a role in ssDNA binding (70). However, the role of this motif in dsDNA binding remains unknown. In addition, preliminary studies performed with the N-terminal part of the molecule suggested that the protein may also bind dsDNA via a  $\beta$ -finger motif located within domain C (65). The role of the  $\beta$ -finger motif in dsDNA binding in the context of the full-length protein and the role of the motif in ssDNA binding have not yet been determined.

Thus, the ability of a zinc-finger mutant protein in which Cys<sub>158</sub>, which is a part of the zinc-finger, was replaced by Ser (70) and a  $\beta$ -finger mutant protein in which R<sub>227</sub> and K<sub>229</sub> were replaced by Ala were studied for their ability to bind ss and dsDNA. Both mutant proteins form dodecamers in solution [(70) and Z.

Kelman, personal communication]. As shown in Figure 32, the wild-type enzyme binds both ss and dsDNA. The mutant proteins, however, are impaired for DNA binding. The zinc-finger mutant retained the ability to interact with DNA though less efficiently in comparison to the wild-type enzyme. The mutations in the  $\beta$ -finger motif, on the other hand, completely abolished the ability of MCM to interact with DNA (Fig. 32).



**Figure 32.** The  $\beta$ -finger of MCM is essential for ss and dsDNA binding. DNA binding by MCM and its mutants were performed using filter binding assays as described in Methods in the presence of 2.5 nM  $^{32}$ P-labeled random ssDNA (open symbols) or dsDNA (closed symbols) in the presence of 15, 30, 45, 90, 150 and 225 nM of MCM (as monomer). Circle, wild-type enzyme; square, zinc-finger mutant; triangle,  $\beta$ -finger mutant. The average result of three experiments is shown with standard deviations.

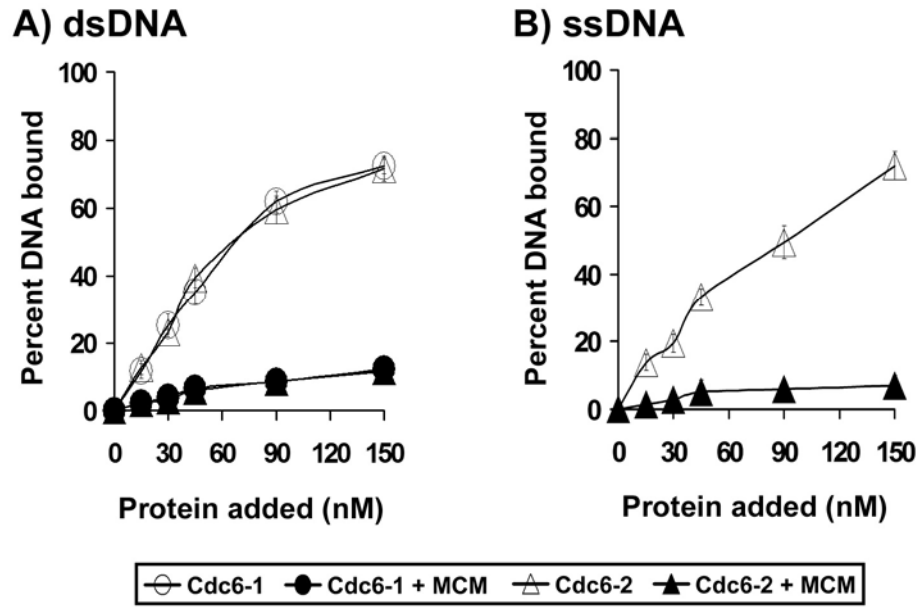
These experiments were repeated with both origin specific and random sequences with no observed difference as expected from a replicative helicase. The  $\beta$ -finger mutant protein did not possess helicase activity on any substrate studied (Z. Kelman, personal communication) while the zinc-finger mutant retained limited activity [(69), see also chapter 2]. The results with the  $\beta$ -finger mutant are different from those reported for the *S. solfataricus* MCM in which similar mutation reduced, but not abolished DNA binding and retained helicase activity (105). These observations add to a growing number of evidence showing that although all archaeal replication systems are similar, differences do exist, in particular between the euryarchaeal and crenarchaeal kingdoms.

**MCM-Cdc6 interactions influence their DNA binding activities.** Cdc6-1 and -2 were shown to interact with MCM [(46,47), see also chapter 4] and these interactions inhibit MCM helicase activity [(47), see also chapter 4]. In order to gain insight into the possible mechanism of helicase loading, the effect of MCM interactions with Cdc6 on the DNA binding activity of Cdc6-1 and -2 were determined. For this study, the  $\beta$ -finger mutant of MCM was used, as it is devoid of any DNA binding activity (Fig. 32) and thus will not create background in the experiment yet it retains the ability to interact with Cdc6-1 and -2 [(47), see also chapter 4]. Hence, any observable DNA binding will solely be due to Cdc6.

These studies revealed that the presence of MCM substantially reduced dsDNA binding by Cdc6-1 and -2 (Fig. 33A, compare closed symbols to open symbols). MCM also had a similar effect on the ssDNA binding by Cdc6-2 (Fig. 33B, compare closed symbols to open symbols). As Cdc6-1 did not bind ssDNA



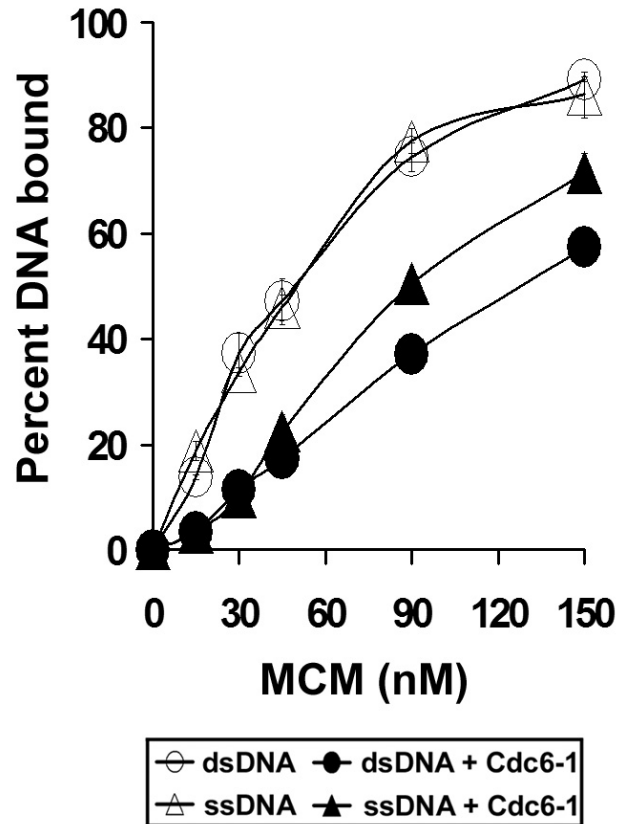
(Fig. 30C) MCM has no effect on this reaction and thus is not shown in Figure 33B. These results show that MCM-Cdc6 interactions affect the DNA binding activity of Cdc6.



**Figure 33. MCM inhibits DNA binding by Cdc6.** DNA binding assays were performed as described in Methods using 2.5 nM  $^{32}$ P-labeled origin specific dsDNA (A) or ssDNA (B) with 15, 30, 45, 90 and 150 nM of Cdc6-1 (circle) and Cdc6-2 (triangle) in the presence (filled symbols) or absence (open symbols) of 300 nM of the  $\beta$ -finger mutant of MCM. The average result of three experiments is shown with standard deviations.

What about the effect of Cdc6 on DNA binding by MCM? To address this question the WH mutant of Cdc6-1 was used to determine its effect on MCM-DNA interactions (Fig. 34). As the mutation in the WH domain of Cdc6-2 did not abolish DNA binding by the protein the effect of this protein was not evaluated. As shown in Figure 34, Cdc6-1 slightly impaired MCM binding to DNA. Cdc6-1

binding to DNA in the absence of MCM under the same experimental condition was subtracted from each of the experiments shown in the Figure 34.



**Figure 34. Cdc6-1 reduces, but does not abolish the DNA binding by MCM.** DNA binding assays were performed as described in Methods using 2.5 nM  $^{32}\text{P}$ -labeled origin specific ssDNA (triangle) or dsDNA (circle) with 15, 30, 45, 90 and 150 nM of MCM in the presence (filled symbols) or absence (open symbols) of 300 nM Cdc6-1 protein with mutation in the WH domain. The average result of three experiments is shown with standard deviations.

These experiments demonstrate that in addition to the effect of MCM-Cdc6 interactions on MCM helicase activity and the autophosphorylation of Cdc6 ((47), see also chapter 4), they also modulate the DNA binding of these proteins.

## DISCUSSION

Many archaeal genomes contain two Cdc6 homologues. It was proposed that one protein is the functional homologue of the eukaryotic Cdc6 while the other is the functional homologue of ORC involving in origin recognition (2). In addition, it was shown that most archaeal origins identified to date are located in the vicinity of a gene encoding for Cdc6 (24,26,108,109) suggesting that the product of this gene is the origin binding protein (21).

In *M. thermautotrophicus* the origin of replication is located upstream of the gene encoding Cdc6-1 (42), and thus it is the prime candidate to function as the origin recognition factor. The data presented here together with past studies (41) support this hypothesis. Cdc6-1 show clear preferential binding to inverted repeats found within the origin region in comparison to random DNA sequences (Fig. 30. and (41)). These observations are similar to those made with the eukaryotic ORC [e.g. (110)] and the bacterial DnaA protein (4) in which these proteins exhibit clear preferential binding to origin DNA. In addition, the archaeal Cdc6 proteins were shown to have structural similarities to the bacterial DnaA (40).

The archaeal Cdc6-1, however show several differences from the eukaryal ORC and the bacterial DnaA. While Cdc6-1 did not bind ssDNA, both ORC and DnaA show such an activity (110,111). For DnaA, it was shown that ATP binding is required for ssDNA binding (111) and for ORC, ssDNA stimulate the ATPase activity of the enzyme (110). ATP had no effect on Cdc6-1 (and Cdc6-2) DNA binding to either ss or dsDNA (Z. kelman, personal communication). It was

shown, however, that recombinant archaeal Cdc6 proteins purified from *E. coli* are tightly bound to ADP (28,38,39). Thus, it is likely that whether ATP is present in the reaction or not the results mimic the situation in an ADP bound form of the enzyme. Thus, even if ATP binding is required for ssDNA binding by Cdc6-1, it cannot be observed.

The archaeal helicase loader has not yet been identified. However, the data presented here in conjunction with other observations, and the similarities to the bacterial and eukaryal systems suggest that Cdc6-2 may play a role in helicase loading. The protein shows amino acid sequence similarities to the eukaryotic Cdc6 (100) which participates in MCM assembly at the eukaryotic origin (112). In bacteria, when the helicase loader, DnaC, associate with the replicative helicase, DnaB, it inhibits its helicase activity which is relieved upon the helicase loading at the origin (113). Similar observations were made with the archaeal Cdc6 proteins which also inhibits the MCM helicase activity [(45-47), see also chapter 4]. In *M. thermautotrophicus* it was shown that Cdc6-2 inhibits MCM helicase activity better than Cdc6-1 [(46,47), see also chapter 4]. In addition, Cdc6-2 does not have any clear DNA binding preference for origin sequence (Fig 30). Thus, taking together, these observations may support the idea that in *M. thermautotrophicus* Cdc6-2 participates in helicase loading.

## CHAPTER 6

### CONCLUSION

The complex task of replicating the genome of an organism requires the coordinated activity of numerous proteins. Studies on archaeal DNA replication was initiated only in the past decade primarily triggered after the completion of the whole genome sequences of several members of this domain. However, much remains to be determined regarding the structural, functional and biochemical properties of the replication proteins from archaea. The whole genome sequences revealed that the genetic information of archaea is present in a bacterial-like circular chromosome, though the proteins involved in DNA replication share sequence similarities to the eukaryal proteins. In addition, sequence comparisons revealed that the replication proteins are less complex in archaea compared to eukarya.

The primary goal of this study was to better understand the mechanisms involved in initiating DNA replication in the archaeon *M. thermautotrophicus*. To achieve this, the biochemical, functional and regulational properties of the proteins involved in the process were determined. *M. thermautotrophicus* genome contains two homologues of Cdc6 proteins and one MCM homologue. While, the Cdc6-1 and -2 proteins are thought to function in origin recognition and/or helicase loading, the MCM homologue is the replicative helicase. Hence the study concentrated on elucidating the roles of these proteins in replication initiation.

The vast structural information available for the *M. thermautotrophicus* MCM helicase was exploited to a large extent in determining the functional roles of the N-terminal domains. The study revealed independent roles for the three N-terminal domains; A, B and C. While, domain A, was suggested to play a regulatory role, domains B and C were shown to be involved in ssDNA binding and protein multimerization respectively. Since, domain C was identified as the multimerization domain, site directed mutagenesis of certain key residues hypothesized to be involved in double-hexamer formation were performed to determine the role of dodecamer vs. hexamer for protein function. However, all mutants remained double-hexamers in solution.

Cdc6-1 and -2 proteins share sequence similarities to the eukaryal Cdc6, which is presumed to be the putative helicase loader. Hence, as a first step to determine the functional role(s) of the two archaeal Cdc6 proteins in replication initiation, Cdc6-MCM interactions were studied. The results revealed an interaction of MCM with the two Cdc6 proteins. In addition, the mapping of domains involved in these protein-protein interactions revealed that while MCM interacts with both Cdc6 proteins via domain C of the N-terminal portion, the interaction domains in Cdc6-1 and -2 to MCM were different suggesting that these proteins may have different function(s) in the initiation process. Additionally, the effects of the Cdc6-MCM interactions on their respective biochemical properties disclosed a regulatory role for the Cdc6 proteins in MCM helicase activity and a modulatory role for the MCM helicase in Cdc6 autophosphorylation.

To further the efforts towards understanding the mechanism of helicase assembly at the origin, the roles of MCM and Cdc6 proteins in DNA binding was determined. The study confirmed that Cdc6-1 bind origin specific inverted repeat sequences with higher affinity compared to random DNA substrates. However, the Cdc6-2 protein had no DNA binding specificity towards origin or random sequences. Further, the study identified the motifs/amino acids involved in DNA binding in all three proteins. The results also revealed that the DNA binding activity of these proteins is regulated by Cdc6-MCM interactions.

But what is the precise function of these two Cdc6 homologues during helicase assembly at the origin? Domain C of MCM is sufficient for its hexamerization and the data also suggest that domain C is also needed for Cdc6-MCM interactions. These observations raise several interesting questions and hypotheses regarding Cdc6 function during the initiation process. Do the interactions with MCM occur via the interfaces of two MCM monomers thereby breaking apart the MCM complex? Or do Cdc6 bind only to the hexameric form of MCM? Or does Cdc6 interact with a region of domain C of MCM that is separated from the multimerization portion of the MCM protein?

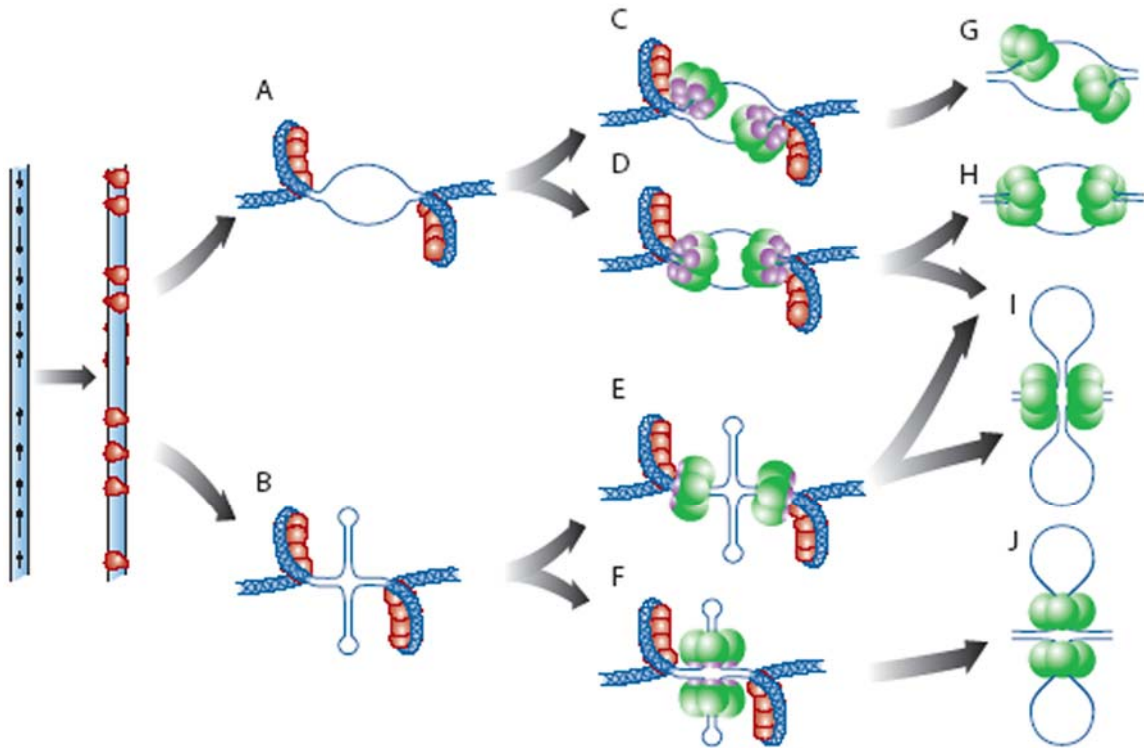
The answers for these questions are currently unknown. The sizing column chromatography and glycerol gradient sedimentation analysis failed to reveal Cdc6-MCM interactions. These observations can be explained by weak or transient interactions between the molecules. However, if successful, these studies could show whether the interactions with Cdc6 dissociate the MCM hexamer complex. Nevertheless, it is tempting to speculate that Cdc6 binds to

the interface between the monomers in the MCM hexamers. This binding may weaken the interaction among the monomers and facilitate helicase assembly at the origin. This is reminiscent of the loading of the *E. coli* processivity factor, in which binding of the clamp loader to the sliding clamp causes an opening at one interface of the dimeric clamp (114,115). Future studies are needed to determine which roles the Cdc6-MCM interactions play during helicase loading.

It is also possible that the Cdc6 proteins alter the conformation of MCM or inhibit conformational changes that may be needed for MCM activity. It was shown that the DnaB hexamers can adopt two different architectures of  $C_3$  and  $C_6$  symmetry (116), but binding of DnaC appears to “freeze” the helicase in the  $C_3$  architecture (117). *M. thermautotrophicus* MCM protein has been shown to form several structures including hexamers, heptamers, dodecamers and filaments. Thus, MCM may adopt different forms for different functions. It was proposed that bacteriophage-T7 helicase forms hexameric rings for ssDNA translocation and heptameric rings for duplex translocation (118). It is possible that Cdc6 binding to MCM “freezes” the complex in one form (an inactive form) and only upon proper MCM loading at the origin does Cdc6 dissociate from the complex and allow MCM to function as a helicase.

The results presented in this thesis provide the basis for various possible models for the initiation of DNA replication in *M. thermautotrophicus* (Fig. 35). Some gaps in the available information were filled based on the similarities between the archaeal and bacterial and eukaryal systems.





**Figure 35. Models for the helicase assembly at the *M. thermautotrophicus* origin of replication.** See text for details.

The *M. thermautotrophicus* origin contains 10 short and two long inverted repeats (black arrow) (42). Upon binding of Cdc6-1 (orange) to the inverted repeats within the origin it aggregates to form the initial replication bubble (A) or a cruciform structure (B), as was suggested as a possible structure for the archaeal origin upon Cdc6 binding (2). Cdc6-2 (purple) then associates with MCM (green rings) and bring it to the Cdc6-1-origin complex forming a Cdc6-1, -2 and MCM ternary complex (C-F). Following MCM assembly at the origin in the correct orientation the Cdc6 proteins dissociate and release the helicase to initiate bi-directional DNA synthesis (H-K). It is likely that ATP binding and/or hydrolysis play a major role in the process as both Cdc6 proteins belong to the AAA<sup>+</sup> family

if ATPases but since there is no data regarding the role of ATP it was not included in the model. It is not yet clear whether MCM translocates along ss or dsDNA while unwinding the chromosomes (119) and thus both possibilities are depicted in the model (ssDNA translocation, H; dsDNA translocation I-K). It is also not yet established whether the two helicase rings are moving along the DNA away from each other to form the two replication forks (H and I) or that the protein remains stationary and the DNA is pulled through it (J and K).

## APPENDICES

### Expression and purification of MCM recombinant proteins

A fresh single colony of *E. coli* strain BL21-CodonPlus(DE3)-IRL cells (Stratagene) harboring the MCM gene in the pET-21a expression vector, was inoculated into 50 ml Luria-Bertani (LB) medium in the presence of 100 µg/ml ampicillin and 25 µg/ml chloramphenicol and incubated overnight at 37°C with shaking at 250 rpm. This was used to inoculate five liters of LB media containing the appropriate antibiotics. When the culture reached an  $A_{600}$  of 0.5, protein expression was induced by the addition of 1mM isopropyl-1-thio-β-D-galactopyranoside. The culture was further grown for 3 h at 37°C, after which the cells were harvested and stored at -80°C.

The cell pellet was thawed and resuspended in 50 ml lysis buffer containing 50mM Tris-HCl (pH 8.0), 0.5M NaCl, and 20% glycerol at 4°C. The cells were then lysed by sonication, after which the cell debris was removed by centrifugation at 4 °C. The lysate was bound to 5 ml of Ni<sup>2+</sup> beads with gentle shaking for 1 h at 4°C. Following binding, the mixture containing the Ni<sup>2+</sup> beads was poured into a column and washed with 50 ml lysis buffer containing 10 mM imidazole. The column was moved to 22°C, and all subsequent elution steps were carried out at that temperature. The column was further washed with 10 column volumes of 50 mM imidazole in elution buffer containing 40mM Tris-OAc (pH 8.0), 0.4M potassium acetate, and 20% glycerol. The proteins were then eluted in 10ml fractions with increasing concentrations of imidazole (100, 150, 200, 250 and 300 mM) in the same elution buffer. The protein in the eluted

fractions were visualized using 10% SDS-PAGE and the peak fractions containing the purified protein samples were pooled for further purification using ion exchange chromatography using Q-Sepharose column.

The pooled proteins were diluted with 10 volumes of 20mM Tris-HCl (pH 8.0), 100mM NaCl and 20% glycerol to bring down the salt concentration. The proteins were then loaded onto Q Sepharose column pre-equilibrated with 20mM Tris-HCl (pH 8.0), 100mM NaCl and 20% glycerol. After protein binding, the column was washed with 10 column volumes of 20mM Tris-HCl (pH 8.0), 300mM NaCl and 20% glycerol. The protein was then eluted using 3 column volumes of 20mM Tris-HCl (pH 8.0), 600mM NaCl and 20% glycerol. The eluted proteins were dialyzed 2 times against 20mM Tris-HCl (pH 8.0), 100mM NaCl and 20% glycerol to bring down the salt concentration of the buffer. Finally, protein concentration was measured by Bradford (Bio-Rad) using BSA as the standard and the samples were flash-frozen in liquid nitrogen and stored at -80 °C.

### **Expression and purification of Cdc6 recombinant proteins**

A fresh single colony of *E. coli* strain DE3 pLysS cells (Novagen) harboring the Cdc6 gene in the pET-16b expression vector, was inoculated into 50 ml Luria-Bertani (LB) medium in the presence of 100 µg/ml ampicillin and 25 µg/ml chloramphenicol and incubated overnight at 37°C with shaking at 250 rpm. This was used to inoculate five liters of LB media containing the appropriate antibiotics. When the culture reached an  $A_{600}$  of 0.5, protein expression was induced by the addition of 1mM isopropyl-1-thio-β-D-galactopyranoside. The

culture was further grown for 16 h at 16°C, after which the cells were harvested and stored at -80°C.

The cell pellet was thawed and resuspended in 50 ml lysis buffer containing 50mM Tris-HCl (pH 8.0), 0.5M NaCl, and 20% glycerol at 4°C. The cells were then lysed by sonication, after which the cell debris was removed by centrifugation at 4 °C. The lysate was bound to 5 ml of Ni<sup>2+</sup> beads with gentle shaking for 1 h at 4°C. Following binding, the mixture was poured into a column and washed with 50 ml lysis buffer containing 10 mM imidazole. The column was moved to 22°C, and all subsequent elution steps were carried out at that temperature. The column was further washed with 10 column volumes of 50 mM imidazole in elution buffer containing 40mM Tris-OAc (pH 8.0), 0.4M potassium acetate, and 20% glycerol. The proteins were then eluted in 10ml fractions with increasing concentrations of imidazole (100, 150, 200, 250 and 300 mM) in the same elution buffer. The protein in the eluted fractions was visualized using 10% SDS-PAGE and protein concentrations were measured by Bradford (Bio-Rad) with bovine serum albumin (BSA) as the standard. The peak fractions containing the purified proteins were flash-frozen in liquid nitrogen and stored at -80 °C.

## REFERENCES

1. Grabowski, B. and Kelman, Z. (2003) ARCHAEAL DNA REPLICATION: Eukaryal Proteins in a Bacterial Context. *Annu Rev Microbiol*, **57**, 487-516.
2. Kelman, L.M. and Kelman, Z. (2003) Archaea: an archetype for replication initiation studies? *Mol Microbiol*, **48**, 605-615.
3. Kelman, L.M. and Kelman, Z. (2004) Multiple origins of replication in archaea. *Trends Microbiol*, **12**, 399-401.
4. Kornberg, A. and Baker, T.A. (1992) *DNA replication*. 2nd / ed. W.H. Freeman, New York.
5. Barns, S.M., Delwiche, C.F., Palmer, J.D. and Pace, N.R. (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc Natl Acad Sci U S A*, **93**, 9188-9193.
6. Zeikus, J.G. and Wolfe, R.S. (1972) *Methanobacterium thermoautotrophicus* sp. n., an anaerobic, autotrophic, extreme thermophile. *J Bacteriol*, **109**, 707-715.
7. Smith, D.R., Doucette-Stamm, L.A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K. *et al.* (1997) Complete genome sequence of *Methanobacterium thermoautotrophicum*  $\Delta$ H: functional analysis and comparative genomics. *J Bacteriol*, **179**, 7135-7155.

8. Sandman, K., Pereira, S.L. and Reeve, J.N. (1998) Diversity of prokaryotic chromosomal proteins and the origin of the nucleosome. *Cell Mol Life Sci*, **54**, 1350-1364.
9. MacNeill, S.A. (2001) Understanding the enzymology of archaeal DNA replication: progress in form and function. *Mol Microbiol*, **40**, 520-529.
10. Ishino, Y. and Cann, I.K. (1998) The euryarchaeotes, a subdomain of Archaea, survive on a single DNA polymerase: fact or farce? *Genes Genet Syst*, **73**, 323-336.
11. Cann, I.K., Ishino, S., Hayashi, I., Komori, K., Toh, H., Morikawa, K. and Ishino, Y. (1999) Functional interactions of a homolog of proliferating cell nuclear antigen with DNA polymerases in *Archaea*. *J Bacteriol*, **181**, 6591-6599.
12. Boulikas, T. (1996) Common structural features of replication origins in all life forms. *J Cell Biochem*, **60**, 297-316.
13. DePamphilis, M.L. (1996) In DePamphilis, M. L. (ed.), *DNA replication in eukaryotic cells*. Cold Spring Harbor Laboratory Press, [Plainview, New York], pp. 45-86.
14. Grigoriev, A. (1998) Analyzing genomes with cumulative skew diagrams. *Nucleic Acids Res*, **26**, 2286-2290.
15. Lobry, J.R. (1996) Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol Biol Evol*, **13**, 660-665.

16. Mrazek, J. and Karlin, S. (1998) Strand compositional asymmetry in bacterial and large viral genomes. *Proc Natl Acad Sci U S A*, **95**, 3720-3725.
17. Salzberg, S.L., Salzberg, A.J., Kerlavage, A.R. and Tomb, J.F. (1998) Skewed oligomers and origins of replication. *Gene*, **217**, 57-67.
18. Zhang, R. and Zhang, C.T. (2004) Identification of replication origins in the genome of the methanogenic archaeon, *Methanocaldococcus jannaschii*. *Extremophiles*, **8**, 253-258.
19. Zhang, R. and Zhang, C.T. (2002) Single replication origin of the archaeon *Methanosarcina mazei* revealed by the Z curve method. *Biochem Biophys Res Commun*, **297**, 396-400.
20. Maisnier-Patin, S., Malandrin, L., Birkeland, N.K. and Bernander, R. (2002) Chromosome replication patterns in the hyperthermophilic euryarchaea *Archaeoglobus fulgidus* and *Methanocaldococcus (Methanococcus) jannaschii*. *Mol Microbiol*, **45**, 1443-1450.
21. Myllykallio, H., Lopez, P., Lopez-Garcia, P., Heilig, R., Saurin, W., Zivanovic, Y., Philippe, H. and Forterre, P. (2000) Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science*, **288**, 2212-2215.
22. Matsunaga, F., Forterre, P., Ishino, Y. and Myllykallio, H. (2001) *In vivo* interactions of archaeal Cdc6/Orc1 and minichromosome maintenance protein with the replication origin. *Proc Natl Acad Sci U S A*, **98**, 11152-11157.



23. Lundgren, M., Andersson, A., Chen, L., Nilsson, P. and Bernander, R. (2004) Three replication origins in *Sulfolobus* species: Synchronous initiation of chromosome replication and asynchronous termination. *Proc Natl Acad Sci U S A*, **101**, 7046-7051.
24. Robinson, N.P., Dionne, I., Lundgren, M., Marsh, V.L., Bernander, R. and Bell, S.D. (2004) Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. *Cell*, **116**, 25-38.
25. Bell, S.P. (2002) The origin recognition complex: from simple origins to complex functions. *Genes Dev*, **16**, 659-672.
26. Myllykallio, H. and Forterre, P. (2000) Mapping of a chromosome replication origin in an archaeon: Response [Response]. *Trends in Microbiol*, **8**, 537-539.
27. Aravind, L. and Koonin, E.V. (1999) DNA-binding proteins and evolution of transcription regulation in the archaea. *Nucleic Acids Res*, **27**, 4658-4670.
28. Liu, J., Smith, C.L., DeRyckere, D., DeAngelis, K., Martin, G.S. and Berger, J.M. (2000) Structure and function of Cdc6/Cdc18: implications for origin recognition and checkpoint control. *Mol Cell*, **6**, 637-648.
29. Moon, K.Y., Kong, D., Lee, J.K., Raychaudhuri, S. and Hurwitz, J. (1999) Identification and reconstitution of the origin recognition complex from *Schizosaccharomyces pombe*. *Proc Natl Acad Sci U S A*, **96**, 12367-12372.
30. Kennedy, S.P., Ng, W.V., Salzberg, S.L., Hood, L. and DasSarma, S. (2001) Understanding the adaptation of *Halobacterium* species NRC-1 to

- its extreme environment through computational analysis of its genome sequence. *Genome Res*, **11**, 1641-1650.
31. Davey, M.J., Jeruzalmi, D., Kuriyan, J. and O'Donnell, M. (2002) Motors and switches: AAA<sup>+</sup> machines within the replisome. *Nat Rev Mol Cell Biol*, **3**, 826-835.
  32. Neuwald, A.F., Aravind, L., Spouge, J.L. and Koonin, E.V. (1999) AAA<sup>+</sup>: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res*, **9**, 27-43.
  33. Ogura, T. and Wilkinson, A.J. (2001) AAA<sup>+</sup> superfamily ATPases: common structure--diverse function. *Genes Cells*, **6**, 575-597.
  34. Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *Embo J*, **1**, 945-951.
  35. Story, R.M. and Steitz, T.A. (1992) Structure of the recA protein-ADP complex. *Nature*, **355**, 374-376.
  36. Subramanya, H.S., Bird, L.E., Brannigan, J.A. and Wigley, D.B. (1996) Crystal structure of a DExx box DNA helicase. *Nature*, **384**, 379-383.
  37. Giraldo, R. and Diaz-Orejas, R. (2001) Similarities between the DNA replication initiators of Gram-negative bacteria plasmids (RepA) and eukaryotes (Orc4p)/archaea (Cdc6p). *Proc Natl Acad Sci U S A*, **98**, 4938-4943.

38. Singleton, M.R., Morales, R., Grainge, I., Cook, N., Isupov, M.N. and Wigley, D.B. (2004) Conformational changes induced by nucleotide binding in Cdc6/ORC from *Aeropyrum pernix*. *J Mol Biol*, **343**, 547-557.
39. Grabowski, B. and Kelman, Z. (2001) Autophosphorylation of the archaeal Cdc6 homologues is regulated by DNA. *J Bacteriol*, **183**, 5459-5464.
40. Erzberger, J.P., Pirruccello, M.M. and Berger, J.M. (2002) The structure of bacterial DnaA: implications for general mechanisms underlying DNA replication initiation. *Embo J*, **21**, 4763-4773.
41. Capaldi, S.A. and Berger, J.M. (2004) Biochemical characterization of Cdc6/Orc1 binding to the replication origin of the euryarchaeon *Methanothermobacter thermoautotrophicus*. *Nucleic Acids Res*, **32**, 4821-4832.
42. Lopez, P., Philippe, H., Myllykallio, H. and Forterre, P. (1999) Identification of putative chromosomal origins of replication in Archaea. *Mol Microbiol*, **32**, 883-886.
43. Lee, D.G. and Bell, S.P. (2000) ATPase switches controlling DNA replication initiation. *Curr Opin Cell Biol*, **12**, 280-285.
44. Feng, L., Wang, B., Driscoll, B. and Jong, A. (2000) Identification and characterization of *Saccharomyces cerevisiae* Cdc6 DNA-binding properties. *Mol Biol Cell*, **11**, 1673-1685.
45. De Felice, M., Esposito, L., Pucci, B., Carpentieri, F., De Falco, M., Rossi, M. and Pisani, F.M. (2003) Biochemical characterization of a CDC6-like

- protein from the crenarchaeon *Sulfolobus solfataricus*. *J Biol Chem*, **278**, 46424-46431.
46. Shin, J.H., Grabowski, B., Kasiviswanathan, R., Bell, S.D. and Kelman, Z. (2003) Regulation of minichromosome maintenance helicase activity by Cdc6. *J Biol Chem*, **278**, 38059-38067.
  47. Kasiviswanathan, R., Shin, J.H. and Kelman, Z. (2005) Interactions between the archaeal Cdc6 and MCM proteins modulate their biochemical properties. *Nucleic Acids Res*, **33**, 4940-4950.
  48. Carr, K.M. and Kaguni, J.M. (2002) *Escherichia coli* DnaA protein loads a single DnaB helicase at a DnaA box hairpin. *J Biol Chem*, **277**, 39815-39822.
  49. Wahle, E., Lasken, R.S. and Kornberg, A. (1989) The dnaB-dnaC replication protein complex of *Escherichia coli*. I. Formation and properties. *J Biol Chem*, **264**, 2463-2468.
  50. Wahle, E., Lasken, R.S. and Kornberg, A. (1989) The dnaB-dnaC replication protein complex of *Escherichia coli*. II. Role of the complex in mobilizing dnaB functions. *J Biol Chem*, **264**, 2469-2475.
  51. Davey, M.J., Fang, L., McInerney, P., Georgescu, R.E. and O'Donnell, M. (2002) The DnaC helicase loader is a dual ATP/ADP switch protein. *Embo J*, **21**, 3148-3159.
  52. De Felice, M., Esposito, L., Pucci, B., De Falco, M., Manco, G., Rossi, M. and Pisani, F.M. (2004) Modular organization of a Cdc6-like protein from the crenarchaeon *Sulfolobus solfataricus*. *Biochem J*, **381**, 645-653.

53. Grainge, I., Scaife, S. and Wigley, D.B. (2003) Biochemical analysis of components of the pre-replication complex of *Archaeoglobus fulgidus*. *Nucleic Acids Res*, **31**, 4888-4898.
54. De Felice, M., Esposito, L., Pucci, B., De Falco, M., Rossi, M. and Pisani, F.M. (2004) A CDC6-like factor from the Archaea *Sulfolobus solfataricus* promotes binding of the mini-chromosome maintenance complex to DNA. *J Biol Chem*, **279**, 43008-43012.
55. Tye, B.K. (1999) MCM proteins in DNA replication. *Annu Rev Biochem*, **68**, 649-686.
56. Ishimi, Y. (1997) A DNA helicase activity is associated with an MCM4, -6, and -7 protein complex [published erratum appears in J Biol Chem 1998 Sep 4;273(36):23616]. *J Biol Chem*, **272**, 24508-24513.
57. Lee, J.-K. and Hurwitz, J. (2001) Processive DNA helicase activity of the minichromosome maintenance proteins 4, 6, and 7 complex requires forked DNA structures. *Proc Natl Acad Sci U S A*, **98**, 54-59.
58. Lee, J.K. and Hurwitz, J. (2000) Isolation and characterization of various complexes of the minichromosome maintenance proteins of *Schizosaccharomyces pombe*. *J Biol Chem*, **275**, 18871-18878.
59. Chong, J.P., Hayashi, M.K., Simon, M.N., Xu, R.M. and Stillman, B. (2000) A double-hexamer archaeal minichromosome maintenance protein is an ATP-dependent DNA helicase. *Proc Natl Acad Sci U S A*, **97**, 1530-1535.
60. Kelman, Z., Lee, J.K. and Hurwitz, J. (1999) The single minichromosome maintenance protein of *Methanobacterium thermoautotrophicum* DH

- contains DNA helicase activity. *Proc Natl Acad Sci U S A*, **96**, 14783-14788.
61. Yu, X., VanLoock, M.S., Poplawski, A., Kelman, Z., Xiang, T., Tye, B.K. and Egelman, E.H. (2002) The *Methanobacterium thermoautotrophicum* MCM protein can form heptameric rings. *EMBO Rep*, **3**, 792-797.
  62. Pape, T., Meka, H., Chen, S., Vicentini, G., van Heel, M. and Onesti, S. (2003) Hexameric ring structure of the full-length archaeal MCM protein complex. *EMBO Rep*, **4**, 1079-1083.
  63. Chen, Y.J., Yu, X., Kasiviswanathan, R., Shin, J.H., Kelman, Z. and Egelman, E.H. (2005) Structural Polymorphism of *Methanothermobacter thermoautotrophicus* MCM. *J Mol Biol*, **346**, 389-394.
  64. Gómez-Llorente, Y., Fletcher, R.J., Chen, X.S., Carazo, J.M. and Martín, C.S. (2005) Polymorphism and double hexamer structure in the archaeal MCM helicase from *Methanobacterium thermoautotrophicum*. *J. Biol. Chem*, **in press**.
  65. Fletcher, R.J., Bishop, B.E., Leon, R.P., Sclafani, R.A., Ogata, C.M. and Chen, X.S. (2003) The structure and Function of MCM from archaeal *M. thermoautotrophicum*. *Nature Struct Biol*, **10**, 160-167.
  66. Bochkarev, A., Bochkareva, E., Frappier, L. and Edwards, A.M. (1999) The crystal structure of the complex of replication protein A subunits RPA32 and RPA14 reveals a mechanism for single-stranded DNA binding. *Embo J*, **18**, 4498-4504.

67. Shechter, D.F., Ying, C.Y. and Gautier, J. (2000) The intrinsic DNA helicase activity of *Methanobacterium thermoautotrophicum*  $\Delta$ H minichromosome maintenance protein. *J Biol Chem*, **275**, 15049-15059.
68. Carpentieri, F., De Felice, M., De Falco, M., Rossi, M. and Pisani, F.M. (2002) Physical and functional interaction between the MCM-like DNA helicase and the single-stranded DNA binding protein from the crenarchaeon *Sulfolobus solfataricus*. *J Biol Chem*, **277**, 12118-12127.
69. Kasiviswanathan, R., Shin, J.H., Melamud, E. and Kelman, Z. (2004) Biochemical Characterization of the *Methanothermobacter thermautotrophicus* Minichromosome Maintenance (MCM) Helicase N-terminal Domains. *J Biol Chem*, **279**, 28358-28366.
70. Poplawski, A., Grabowski, B., Long, S.E. and Kelman, Z. (2001) The zinc finger domain of the archaeal minichromosome maintenance protein is required for helicase activity. *J Biol Chem*, **276**, 49371-49377.
71. Hanas, J.S., Hazuda, D.J., Bogenhagen, D.F., Wu, F.Y. and Wu, C.W. (1983) *Xenopus* transcription factor A requires zinc for binding to the 5 S RNA gene. *J Biol Chem*, **258**, 14120-14125.
72. Miller, J., McLachlan, A.D. and Klug, A. (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *Embo J*, **4**, 1609-1614.
73. Fletcher, R.J., Shen, J., Gómez-Llorente, Y., Martín, C.S., Carazo, J.M. and Chen, X.S. (2005) Double hexamer disruption and biochemical

activities of *Methanobacterium thermoautotrophicum* MCM. *J. Biol. Chem*,  
**in press.**

74. Kelman, Z. and Hurwitz, J. (2003) Structural lessons in DNA replication from the third domain of life. *Nat Struct Biol*, **10**, 148-150.
75. Patel, S.S. and Picha, K.M. (2000) Structure and function of hexameric helicases. *Annu Rev Biochem*, **69**, 651-697.
76. Laskey, R.A. and Madine, M.A. (2003) A rotary pumping model for helicase function of MCM proteins at a distance from replication forks. *EMBO Rep*, **4**, 26-30.
77. Shin, J.H., Jiang, Y., Grabowski, B., Hurwitz, J. and Kelman, Z. (2003) Substrate requirements for duplex DNA translocation by the eukaryal and archaeal minichromosome maintenance helicases. *J Biol Chem*, **278**, 49053-49062.
78. Pucci, B., De Felice, M., Rossi, M., Onesti, S. and Pisani, F.M. (2004) Amino Acids of the *Sulfolobus solfataricus* Mini-chromosome Maintenance-like DNA Helicase Involved in DNA Binding/Remodeling. *J Biol Chem*, **279**, 49222-49228.
79. Tye, B.K. (2000) Insights into DNA replication from the third domain of life. *Proc Natl Acad Sci U S A*, **97**, 2399-2401.
80. Forsburg, S.L. (2004) Eukaryotic MCM Proteins: Beyond Replication Initiation. *Microbiol Mol Biol Rev*, **68**, 109-131.



81. Tye, B.K. and Sawyer, S.L. (2000) The Hexameric Eukaryotic MCM Helicase: Building Symmetry from Nonidentical Parts. *J Biol Chem*, **275**, 34833-34836.
82. Ishimi, Y., Komamura, Y., You, Z. and Kimura, H. (1998) Biochemical function of mouse minichromosome maintenance 2 protein. *J Biol Chem*, **273**, 8369-8375.
83. McEntee, K., Weinstock, G.M. and Lehman, I.R. (1980) recA protein-catalyzed strand assimilation: stimulation by *Escherichia coli* single-stranded DNA-binding protein. *Proc Natl Acad Sci U S A*, **77**, 857-861.
84. Schneider, T.D. and Stephens, R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res*, **18**, 6097-6100.
85. Siegel, L.M. and Monty, K.J. (1966) Determination of molecular weights and frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation. Application to crude preparations of sulfite and hydroxylamine reductases. *Biochim Biophys Acta.*, **112**, 346-362.
86. You, Z., Ishimi, Y., Masai, H. and Hanaoka, F. (2002) Roles of Mcm7 and Mcm4 subunits in the DNA helicase activity of the mouse Mcm4/6/7 complex. *J Biol Chem*, **277**, 42471-42479.
87. Smelkova, N.V. and Borowiec, J.A. (1997) Dimerization of simian virus 40 T-antigen hexamers activates T-antigen DNA helicase activity. *J Virol*, **71**, 8766-8773.

88. Kaplan, D.L., Davey, M.J. and O'Donnell, M. (2003) Mcm4,6,7 Uses a "Pump in Ring" Mechanism to Unwind DNA by Steric Exclusion and Actively Translocate along a Duplex. *J Biol Chem*, **278**, 49171-49182.
89. Yang, H., Jeffrey, P.D., Miller, J., Kinnucan, E., Sun, Y., Thoma, N.H., Zheng, N., Chen, P.L., Lee, W.H. and Pavletich, N.P. (2002) BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. *Science*, **297**, 1837-1848.
90. Kaplan, D.L. (2000) The 3'-tail of a forked-duplex sterically determines whether one or two DNA strands pass through the central channel of a replication-fork helicase. *J Mol Biol*, **301**, 285-299.
91. Hardy, C.F., Dryga, O., Seematter, S., Pahl, P.M. and Sclafani, R.A. (1997) *mcm5/cdc46-bob1* bypasses the requirement for the S phase activator Cdc7p. *Proc Natl Acad Sci U S A*, **94**, 3151-3155.
92. Adachi, Y., Usukura, J. and Yanagida, M. (1997) A globular complex formation by Nda1 and the other five members of the MCM protein family in fission yeast. *Genes Cells*, **2**, 467-479.
93. Sato, M., Gotow, T., You, Z., Komamura-Kohno, Y., Uchiyama, Y., Yabuta, N., Nojima, H. and Ishimi, Y. (2000) Electron microscopic observation and single-stranded DNA binding activity of the Mcm4,6,7 complex. *J Mol Biol*, **300**, 421-431.
94. Cann, I.K. and Ishino, Y. (1999) Archaeal DNA replication. Identifying the pieces to solve a puzzle. *Genetics*, **152**, 1249-1267.

95. Fang, L., Davey, M.J. and O'Donnell, M. (1999) Replisome assembly at *oriC*, the replication origin of *E. coli*, reveals an explanation for initiation sites outside an origin. *Mol Cell*, **4**, 541-553.
96. Marszalek, J. and Kaguni, J.M. (1994) DnaA protein directs the binding of DnaB protein in initiation of DNA replication in *Escherichia coli*. *J Biol Chem*, **269**, 4883-4890.
97. Kneissl, M., Putter, V., Szalay, A.A. and Grummt, F. (2003) Interaction and assembly of murine pre-replicative complex proteins in yeast and mouse cells. *J Mol Biol*, **327**, 111-128.
98. Tropea, J.E., Cherry, S., Nallamsetty, S., Bignon, C. and Waugh, D.S. (2005) A generic method for the production of recombinant proteins in *Escherichia coli* using a dual His6-MBP affinity tag. *Methods. Mol. Biol.*, **in press**.
99. Kelman, Z., Naktinis, V. and O'Donnell, M. (1995) Radiolabeling of proteins for biochemical studies. *Methods Enzymol*, **262**, 430-342.
100. Giraldo, R. (2003) Common domains in the initiators of DNA replication in *Bacteria*, *Archaea* and *Eukarya*: combined structural, functional and phylogenetic perspectives. *FEBS Microbiol. Rev.*, **26**, 533-554.
101. Kelman, Z. and Hurwitz, J. (2000) A unique organization of the protein subunits of the DNA polymerase clamp loader in the archaeon *Methanobacterium thermoautotrophicum* DH. *J Biol Chem*, **275**, 7327-7336.

102. Kelman, Z. and O'Donnell, M. (1995) Structural and functional similarities of prokaryotic and eukaryotic DNA polymerase sliding clamps. *Nucleic Acids Res*, **23**, 3613-3620.
103. Marszalek, J., Zhang, W., Hupp, T.R., Margulies, C., Carr, K.M., Cherry, S. and Kaguni, J.M. (1996) Domains of DnaA protein involved in interaction with DnaB protein, and in unwinding the *Escherichia coli* chromosomal origin. *J Biol Chem*, **271**, 18535-18542.
104. Jang, S.W., Elsasser, S., Campbell, J.L. and Kim, J. (2001) Identification of Cdc6 protein domains involved in interaction with Mcm2 protein and Cdc4 protein in budding yeast cells. *Biochem J*, **354**, 655-661.
105. McGeoch, A.T., Trakselis, M.A., Laskey, R.A. and Bell, S.D. (2005) Organization of the archaeal MCM complex on DNA and implications for the helicase mechanism. *Nat Struct Mol Biol.*, **12**, 756-762.
106. Kelman, Z. and Hurwitz, J. (2000) A unique organization of the protein subunits of the DNA polymerase clamp loader in the archaeon *Methanobacterium thermoautotrophicum* deltaH. *J Biol Chem*, **275**, 7327-7336.
107. Kelman, Z., Lee, J.K. and Hurwitz, J. (1999) The single minichromosome maintenance protein of *Methanobacterium thermoautotrophicum* DeltaH contains DNA helicase activity. *Proc Natl Acad Sci U S A*, **96**, 14783-14788.
108. Kelman, Z. and White, M.F. (2005) Archaeal DNA replication and repair. *Curr Opin Microbiol*, **8**, 669-676.

109. Robinson, N.P. and Bell, S.D. (2005) Origins of DNA replication in the three domains of life. *Febs J*, **272**, 3757-3766.
110. Lee, D.G., Makhov, A.M., Klemm, R.D., Griffith, J.D. and Bell, S.P. (2000) Regulation of origin recognition complex conformation and ATPase activity: differential effects of single-stranded and double-stranded DNA binding. *Embo J*, **19**, 4774-4782.
111. Speck, C. and Messer, W. (2001) Mechanism of origin unwinding: sequential binding of DnaA to double- and single-stranded DNA. *Embo J*, **20**, 1469-1476.
112. Bell, S.P. and Dutta, A. (2002) DNA replication in eukaryotic cells. *Annu Rev Biochem*, **71**, 333-374.
113. Davey, M.J. and O'Donnell, M. (2003) Replicative helicase loaders: ring breakers and ring makers. *Curr Biol*, **13**, R594-596.
114. Jeruzalmi, D., Yurieva, O., Zhao, Y., Young, M., Stewart, J., Hingorani, M., O'Donnell, M. and Kuriyan, J. (2001) Mechanism of Processivity Clamp Opening by the Delta Subunit Wrench of the Clamp Loader Complex of *E. coli* DNA Polymerase III. *Cell*, **106**, 417-428.
115. O'Donnell, M., Jeruzalmi, D. and Kuriyan, J. (2001) Clamp loader structure predicts the architecture of DNA polymerase III holoenzyme and RFC. *Curr Biol*, **11**, R935-946.
116. San Martin, M.C., Stamford, N.P., Dammerova, N., Dixon, N.E. and Carazo, J.M. (1995) A structural model for the *Escherichia coli* DnaB helicase based on electron microscopy data. *J Struct Biol*, **114**, 167-176.

117. Barcena, M., Ruiz, T., Donate, L.E., Brown, S.E., Dixon, N.E., Radermacher, M. and Carazo, J.M. (2001) The DnaB.DnaC complex: a structure based on dimers assembled around an occluded channel. *Embo J*, **20**, 1462-1468.
118. Toth, E.A., Li, Y., Sawaya, M.R., Cheng, Y. and Ellenberger, T. (2003) The crystal structure of the bifunctional primase-helicase of bacteriophage T7. *Mol Cell*, **12**, 1113-1123.
119. Takahashi, T.S., Wigley, D.B. and Walter, J.C. (2005) Pumps, paradoxes and ploughshares: mechanism of the MCM2-7 DNA helicase. *Trends Biochem Sci.*, **30**, 437-444.